Facultad de Traducción y Documentación
Departamento de Traducción e Interpretación

Faculté de traduction et d'interprétation
Département de traitement informatique multilingue (TIM)

Assuring accessibility during web localisation: an empirical investigation on the achievement of appropriate text alternatives for images

Thesis submitted for assessment with a view to obtaining the degrees of:

Doctora en Traducción y mediación intercultural (Universidad de Salamanca)
and
Docteur en traitement informatique multilingue (Université de Genève)

Silvia Rodríguez Vázquez

Supervisors:

Jesús Torres del Rey, Universidad de Salamanca
Pierrette Bouillon, Université de Genève

Salamanca/Genève, March 2016

Version from 30th March 2016

−¡Vaia nevada caeu esta noite! ¡Rafaela, ven! ¡Olla, olla! −gritou Emilio ó sair ó balcón para espreguizarse e respirar fondo o ar dun novo día.

A súa dona, incrédula, achegouse á varanda e observou que os campos estaban tan verdes coma sempre. O sol quentaba con lacazana intensidade os prados, e Rafaela pensou que o seu marido estaba chanceando. Dez minutos despois, Emilio Rodríguez quedaría cego de por vida.

En Fernando Méndez, *Historia dun crime − O caso do metílico*

Abstract

In spite of the considerable advancements made to date in the field of human-computer interaction, the universal goal of granting access to information on the Web for all is not yet a reality. While the rights of people with disabilities concerning their participation in the digital society are acknowledged in many national and international policies, web professionals still fail to create websites that are at least compliant with the web accessibility (WA) guidelines established by the W3C. The first and most basic recommendation for enhancing WA is to add a text alternative to all non-text content that is presented to the user, so that it can be changed into other forms people need, such as braille or speech. In the case of images, the most widely adopted mechanism to do so is to introduce an `alt` attribute for each `<img>` element present in a web page and provide as its value a text equivalent that serves the same purpose (descriptive, functional, decorative) of the image it is associated with. This technique is particularly beneficial for those population groups with a visual impairment, such as the blind community.

This thesis sets out to investigate the role of localisers in the achievement of more accessible multilingual websites for blind users, placing special emphasis on ways to assure access to images in the localised web product. Our work is motivated by the little discussion that there has been thus far in prior work about accessibility within the context of the multilingual Web and the scant attention paid up to the present by both web localisation practitioners and scholars to the needs of disabled people in the target audience. In addition, our interest in the particular case of images is driven by the obstacles still encountered by web professionals, both in monolingual and multilingual websites, to produce high quality text alternatives, namely the limited guidance that exists to assist them in their formulation, and the poor support offered by web accessibility evaluation (WAE) tools for image accessibility assessment, especially with regard to linguistic checks.

To conduct our investigation, we adopt a multi-method empirical approach, combining descriptive and experimental research strategies. First, we carry out a survey on the procedures currently followed by WA experts when checking multilingual websites for accessibility, with a view to providing insight into how they could be more smoothly integrated into the localisation workflow. After obtaining an overview of the state of the art, we narrow the scope of our research to images as the main object of study. In an attempt to cover some of the limitations stated above with regard to image accessibility evaluation, we develop a set of 40 controlled language (CL) rules for assuring the generation of appropriate text alternatives and facilitating a language-based assessment during the web

localisation process. Finally, we conduct an experimental study consisting of a controlled localisation experiment whose output is then subjected to user evaluation, with a view to assessing the effectiveness of the aforementioned CL proposal and with the ultimate goal of examining (i) whether text alternatives are translated by localisers, and (ii) whether having previous knowledge of WA or using accessibility-oriented quality assurance (QA) tools —Acrolinx, a CL checker featuring 10 of the rules we propose, and aDesigner, a general WAE tool— has an impact on the appropriateness of the image text alternatives they produce. Within the framework of this thesis, appropriateness is operationalised as the pertinence and accuracy of the description provided for a given image, according to its function and context within a web page.

The quantitative and qualitative data gathered throughout these three different stages of our work suggest that at present there is no standardised procedure for assessing the accessibility level of multilingual websites, indicating that further research is needed to define how this quality indicator can be included within the web localisation process. Similarly, while WA experts consider that localisers should be held accountable for the level of accessibility achieved in the target web product, in our study we observe a noteworthy degree of variability with regard to the awareness of accessibility issues among these professionals. More specifically, we note that text alternatives are not systematically translated, and that both WA knowledge and the use of accessibility-oriented QA tools are decisive in creating appropriate text alternatives during the localisation process. The outcome of the user evaluation reveals, in fact, that localisers with some WA background produce text alternatives of significantly higher quality than those without, and that Acrolinx performs significantly better than aDesigner in this regard, when only one tool is available. Our findings suggest that CL software can prove beneficial during WA assessments, and they shed light on the need to include accessibility considerations in localisation QA processes if the Web for all goal is to be attained.

Keywords: *web accessibility, web localisation, image accessibility, image localisation, human-computer interaction, controlled language, image text alternatives,* `alt` *attribute, alt text appropriateness, blind users, empirical research.*

Resumen

A pesar de los importantes avances conseguidos hasta la fecha en el campo de la interacción persona-ordenador, el objetivo universal de facilitar el acceso a la información para todos en la web todavía no es una realidad. Si bien numerosas políticas nacionales e internacionales reconocen los derechos de las personas con discapacidad relativos a su participación en la sociedad digital, los profesionales de la web siguen sin crear sitios que satisfagan los criterios de conformidad básicos de accesibilidad web (AW) establecidos por el W3C. Una de las pautas fundamentales que se deben seguir para mejorar la AW consiste en proporcionar alternativas textuales para todo contenido no textual, de modo que éste se pueda convertir a otros formatos que las personas necesiten, tales como braille o voz (por ejemplo, a través de un sintetizador). En el caso de las imágenes, el mecanismo más utilizado para ello es la introducción de un atributo alt en cada uno de los elementos <img> de una página web, proponiendo un texto alternativo que cumpla el mismo valor comunicativo (descriptivo, funcional, decorativo) de la imagen a la que acompaña. Adoptar dicha práctica beneficia de manera especial a los grupos de población con discapacidad visual, como las personas con ceguera.

Esta tesis persigue investigar el papel que pueden desempeñar los localizadores en la consecución de sitios web multilingües más accesibles para usuarios ciegos, prestando especial atención a las distintas formas en las que se puede garantizar un acceso adecuado a las imágenes del producto web localizado. Nuestro trabajo surge a raíz de la falta de atención prestada en la literatura a cuestiones de accesibilidad en sitios web multilingües, así como del bajo nivel de interés que ha suscitado el estudio de las necesidades de las personas con discapacidad en el campo de la localización web, tanto a nivel profesional como académico. Por otro lado, la selección de las imágenes como objeto de estudio principal viene motivada por las dificultades a las que se enfrentan todavía hoy los profesionales de la web a la hora de elaborar alternativas textuales de calidad tanto en sitios web monolingües como multilingües, lo cual se debe especialmente a la falta de un conjunto de pautas claras sobre cómo escribir textos alternativos pertinentes, así como a las deficiencias que presentan en ese sentido las herramientas destinadas a la evaluación de la accesibilidad (herramientas WAE), especialmente en lo que respecta a las correcciones de tipo lingüístico.

Nuestro trabajo, para el cual hemos adoptado un enfoque empírico multimetódico, combina estrategias de investigación de carácter descriptivo y experimental. La primera consiste en identificar, a través de una encuesta, los procedimientos que siguen actualmente los expertos en AW durante la evaluación del nivel de

accesibilidad de sitios web multilingües, con el fin de explorar cómo estos se podrían integrar de forma más eficaz en el proceso de localización. Tras obtener una idea más clara del estado de la cuestión, y como respuesta a algunas de las limitaciones expuestas anteriormente en relación con la accesibilidad de imágenes, proponemos un conjunto de 40 reglas de lenguaje controlado (LC) para ayudar a los localizadores en la elaboración de textos alternativos pertinentes y facilitar su evaluación, sobre todo a nivel lingüístico. Finalmente, llevamos a cabo un estudio dividido en dos fases: un experimento de localización web y una evaluación con usuarios, a través del cual se busca analizar la eficacia de diez de las reglas de LC desarrolladas, así como observar (i) si los localizadores traducen los textos alternativos de las imágenes, y (ii) si el hecho de tener conocimientos previos en materia de accesibilidad o utilizar herramientas de control de calidad (CC) destinadas a la evaluación de la AW (Acrolinx, la herramienta de lenguaje controlado que permite aplicar las reglas que elaboramos, y aDesigner, una herramienta WAE tradicional) afecta de alguna manera la calidad de las alternativas textuales producidas por los localizadores. En el marco de esta tesis, entendemos que dicha calidad depende del grado de pertinencia alcanzado en la descripción proporcionada la imagen, teniendo en cuenta su función y su contexto en la página web.

Los datos cuantitativos y cualitativos obtenidos indican que, en la actualidad, no existe una metodología normalizada para evaluar la accesibilidad de los sitios web multilingües. Ello indica que es necesario llevar a cabo más estudios para definir cómo este indicador de calidad se podría integrar con éxito en el proceso de localización. Asimismo, aunque los expertos consideran los localizadores deben asumir la responsabilidad de garantizar un nivel de accesibilidad aceptable en el producto web meta, nuestros resultados muestran una variabilidad notable en lo que respecta a la concienciación de estos profesionales sobre los problemas de AW. Concretamente, observamos que los textos alternativos no se traducen de forma sistemática, y que tanto tener nociones básicas sobre accesibilidad como usar herramientas destinadas a la evaluación de la AW son factores decisivos para la elaboración de textos alternativos pertinentes durante el proceso de localización. Los resultados de la evaluación con usuarios revelan que, de hecho, la calidad de los textos alternativos producidos por localizadores familiarizados con conceptos básicos en materia de accesibilidad es mayor que la de aquellos que no lo están, y que, cuando solo se utiliza una herramienta, Acrolinx permite conseguir resultados significativamente mejores en este sentido que aDesigner. Nuestro trabajo indica que las herramientas de LC pueden resultar de gran utilidad durante las evaluaciones de AW y, además, apunta a la necesidad de tener en cuenta cuestiones de accesibilidad en la fase de CC del proceso de localización si se pretende conseguir verdaderamente el objetivo de una web para todos.

Palabras clave: *accesibilidad web, localización web, accesibilidad de imágenes, localización de imágenes, interacción persona-ordenador, lenguaje controlado, texto alternativo, alternativa textual, atributo* `alt`*, usuarios ciegos, investigación empírico-experimental.*

Résumé

En dépit des progrès considérables réalisés à ce jour dans le domaine de l'interaction homme-machine, l'objectif universel de faciliter l'accès à l'information sur le Web pour tous n'est pas encore une réalité. Bien que de nombreuses politiques nationales et internationales reconnaissent les droits des personnes handicapées en matière de participation à la société numérique, les professionnels du Web ne parviennent pas encore à créer des sites web qui soient au moins conformes aux règles d'accessibilité du Web (AW) établies par le W3C. La recommandation la plus fondamentale pour améliorer l'AW est de fournir un équivalent textuel à tout contenu non textuel qui est présenté à l'utilisateur, de sorte qu'il puisse être converti en d'autres formes (transcription en braille ou synthèse vocale par exemple) selon les besoins. Dans le cas des images, le mécanisme le plus largement adopté pour ce faire est celui d'insérer un attribut `alt` dans chaque élément `<img>` présent dans une page web et de fournir comme valeur un équivalent textuel servant à expliquer le but de l'image (descriptif, fonctionnel, décoratif) à laquelle il est associé. Cette technique est particulièrement avantageuse pour les groupes de la population qui sont atteints d'un handicap visuel, comme c'est le cas des aveugles.

Cette thèse a pour objectif d'étudier le rôle des localisateurs dans la réalisation de sites web multilingues plus accessibles pour ce groupe d'utilisateurs, en mettant particulièrement l'accent sur les moyens possibles pour assurer l'accès aux images du produit web localisé. Notre travail est motivé par le manque d'études dans la littérature sur l'accessibilité du Web multilingue, ainsi que par le peu d'intérêt que les professionnels de la localisation et le milieu académique étudiant ce domaine de spécialité ont jusqu'à présent montré pour les besoins des personnes handicapées au sein du public cible. En outre, nous prêtons une attention particulière aux images en raison des obstacles que rencontrent encore les professionnels du Web qui travaillent avec des sites monolingues et/ou multilingues lors de la formulation d'équivalents textuels de qualité, notamment du manque de conseils détaillés sur la manière de les écrire et des limites des outils d'évaluation de l'AW permettant de vérifier automatiquement l'accessibilité des images, surtout en ce qui concerne les corrections linguistiques.

Pour mener à bien notre travail, nous adoptons une approche empirique multiméthode, combinant des stratégies de recherche descriptives et expérimentales. Tout d'abord, nous effectuons une enquête sur les procédures actuellement suivies par des experts en AW pendant la vérification de l'accessibilité des sites web multilingues, en vue de mieux comprendre la manière dont elles pourraient être plus facilement intégrées dans le processus de localisation. Après avoir obtenu un

aperçu de l'état de l'art, nous délimitons la portée de notre recherche aux images comme objet central d'étude. Afin de combler certaines des lacunes indiquées ci-dessus en matière d'évaluation de l'accessibilité des images, nous développons un ensemble de quarante règles de langage contrôlé (LC) visant à assurer la production d'équivalents textuels pertinents et à faciliter leur vérification au niveau linguistique au cours du processus de localisation web. Enfin, nous menons une étude expérimentale consistant en une expérience de localisation et une étape d'évaluation par des utilisateurs aveugles, en vue de tester l'efficacité de dix des règles de LC développées, ainsi que d'examiner (i) si les équivalents textuels sont traduits par les localisateurs et (ii) si le fait d'avoir des connaissances préalables sur l'AW ou d'utiliser des outils d'assurance qualité (AQ) destinés au contrôle de l'accessibilité (Acrolinx, un outil de validation fondé sur dix des quarante règles que nous proposons, et aDesigner, un outil général pour la vérification de l'accessibilité) a un effet sur la pertinence des équivalents textuels qu'ils produisent. Dans le cadre de cette thèse, la pertinence est opérationnalisée comme la justesse de la description fournie pour une image donnée, en fonction de sa valeur communicative et de son contexte dans une page web.

Les données quantitatives et qualitatives recueillies au cours de notre travail suggèrent que, à ce jour, il n'y a pas de procédures normalisées pour l'évaluation de l'accessibilité des sites web multilingues. Ceci indique que des travaux de recherche supplémentaires sont nécessaires pour définir la manière dont cet indicateur de qualité peut être intégré dans le processus de localisation. De même, alors que les experts en AW considèrent que les localisateurs devraient être responsables du niveau d'accessibilité du produit web cible, nous observons dans notre étude un degré remarquable de variabilité entre ces professionnels en ce qui concerne la prise de conscience des questions d'accessibilité. Plus précisément, nous constatons que les équivalents textuels ne sont pas systématiquement traduits et que les connaissances sur l'AW et l'utilisation d'outils d'AQ sont déterminantes pour créer des équivalents textuels pertinents au cours du processus de localisation. Les résultats de l'évaluation par des utilisateurs révèlent, en effet, que les localisateurs qui ont des connaissances de base en matière d'AW produisent des équivalents textuels d'une qualité nettement plus élevée que ceux qui n'en ont pas, et qu'Acrolinx est significativement plus performant qu'aDesigner à cet égard, lorsqu'un seul outil est disponible. Notre travail suggère que les logiciels de LC peuvent s'avérer avantageux pour des évaluations d'AW, et il met en lumière la nécessité d'inclure des considérations en matière d'accessibilité dans les processus d'AQ en localisation si l'on veut réaliser le paradigme d'un Web pour tous.

Mots-clés: *accessibilité web, web localisation, accessibilité des images, localisation des images, interaction homme-machine, langage contrôlé, équivalents textuels, attribut `alt`, pertinence, utilisateurs aveugles, recherche empirique.*

Declaration by the author

I hereby declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified. Notions taken over directly or indirectly from other sources have been identified as such. Permission to reproduce and/or adapt the third-party graphical work used in this thesis has been granted by the corresponding authors prior to its publication and appropriate credit and/or reference has been included when needed. This thesis has not previously been presented in identical or similar form to any other examination board.

The work presented in this thesis was carried out within the framework of a Joint Doctoral Degree between the University of Salamanca (USAL), Spain, and the University of Geneva (UNIGE), Switzerland. Following the terms and regulations of the corresponding agreement, signed in August 2012, as well as the requirements to obtain the special mention of *Doctor internacional* by the USAL, two summaries of this research can be found in Appendix L (Spanish) and Appendix M (French).

This thesis was conducted from April 2011 to March 2016, under the supervision of Dr. Jesús Torres del Rey, from the Faculty of Translation and Documentation of the USAL, and Prof. Pierrette Bouillon, from the Faculty of Translation and Interpreting of the UNIGE. It received financial support from the former Rectors' Conference of the Swiss Universities (CRUS), now 'swissuniversities', as per Contract Nº E13/01.

Acknowledgements

They say that a picture paints a thousand words. However, not even with a million pictures could I ever describe how grateful I am to all the people who offered me support and encouragement throughout this journey.

First and foremost, I would like to express my greatest gratitude to my supervisors: to Jesús Torres del Rey, for having sparked in me this passion for localisation and translation technologies more than ten years ago, as well as for his thoroughness and guidance since the very first moment I decided to embark on this adventure; and to Pierrette Bouillon, for having welcomed me into her team and making me discover new interesting areas of research, which certainly have enriched my investigation. They have both generously shared their precious time, knowledge and expertise with me over the last years, and always gave me the freedom to develop my own ideas. Thank you for your trust, as well as the endless opportunities and support you have provided me with.

Pursuing a double doctorate degree across countries and universities has been sometimes challenging but, in the end, I can only see it as a rewarding experience both at a personal and a professional level. It has allowed me to grow as a researcher and work with very talented people along the way. I am thankful to Emilio Rodríguez Vázquez de Aldana, who has helped me solve many of the technical challenges that emerged during the exploitation of the web corpora I used for this thesis, and with whom, besides two surnames, I share the enthusiasm of being part of the Cod.eX research group. Great thanks are also due to the members of the Department of translation at the USAL, who taught me everything I know about the translation profession; a big thank you goes to Rosario Martín Ruano, who was the first person to encourage me to pursue an academic career and wisely advised me when preparing the first draft of my thesis project, and to Joaquín García Palacios and Teresa Fuentes Morán, who warmly welcomed me during my multiple stays in Salamanca.

I would also like to thank all the colleagues from the FTI in Geneva who have contributed to creating an incomparable working environment over the last five years. A special mention goes to Victoria and Johanna for their help during the CL rule development process; to Jean-Pierre, Philippe and Jesús for the technical support offered while preparing the experimental part of my work; to Marianne for her eternal smile, strength and encouraging words on the most difficult days; to Irene, for helping me spread our shared dream of making life easier for those who need it; and to Julie, Lucile, Mélanie and Véronique, who made themselves always available to revise my written French. Most of all, I owe my officemate and friend, Véronique, a huge debt of gratitude for her patience, her kindness and the interesting discussions we had during our breaks in those long and stressful days of work. My PhD years would not have been the same if she was not there. I cannot but mention how fully indebted I am as well to Lucía, not only for the insightful talks about localisation practice and empirical research methods, but also for being such a good soul and for her unconditional friendship. Even before meeting in person, she had always been thoughtful and supportive. Thank you for your understanding and invaluable help throughout all these years. Similarly, I also want to acknowledge the support provided by Marc-Olivier Boldi, from the UNIGE, who helped me find my way among the "data chaos" and who made statistics seem more accessible; and by Jen Conroy, who accepted to proofread my English writing, sometimes on very short notice.

My research work has been positively influenced by many experts in the accessibility field. Above all, I want to express my sincere gratitude to Markus Reich, former director of the Access for all Foundation, and to Anton Bolfing, in charge of the Research and development Department of the same institution, for inviting me to Zürich to present my research ideas and for allowing me to use their data for the purposes of my thesis. I am also thoroughly grateful to the ASSETS and W4A communities for having taught me so much about the needs of people with disabilities and how to do research to empower them in this ever evolving technologised society. A special thanks goes to the ASSETS 2012 Doctoral Consortium (DC) mentors for their enlightening advice at the early stages of my PhD, and to my peer mentees; in particular, to Jessica, Anke and Hernisa for the inspirational conversations and the good times shared ever since.

On a personal level, I would like to say a big thank you to Paula, Alba, Ángel and Thomas, for their friendship and for holding up to our promise to keep trying to see each other, no matter where, for our annual 'eurogathering'. The simple thought of meeting you again (at least) each May gave me the strength to keep advancing with my research. Many thanks go as well to all my friends in Spain, who always tried to find a moment to catch up during my short visits to the country, and to the other many good friends I made in Geneva and those spread around the world who have always had encouraging words for me. I am equally grateful to Sorin, for being so considerate (and for the biscuits!), and to Ica, who wished perhaps more than anyone for me to finish this PhD dissertation and had to leave before I could give her a copy.

This thesis would not have been a reality without the affection of my grandparents Joaquín and Nieves, my mother Isabel, my father José Antonio and my brother Cristian, who have always given me all the support I needed to pursue my professional career, even if that meant for me to be far from what I will always call home. I owe them everything I am today, and their hard work and determination are an example for me to follow. Finally, the completion of this thesis

has been possible in large part thanks to the patience and understanding of my partner, who kept me going and took care of me, especially during this last period, without ever asking anything in return. Thank you, Florin, for your love, gentleness and positive-thinking, but most of all, for bringing out the *Alice* in me and making me believe in a dream-like world.

A los que se han ido y a aquellos que están por venir

Contents

List of figures

List of tables

List of abbreviations

AEM	Accessibility Evaluation Method
AET	Accessibility Evaluation Toolbar (by John Gunderson)
ALI	Acrolinx Linguistic Integrated Development Environment
ANOVA	Analysis of Variance
ARIA	Accessible Rich Internet Applications
ASCII	American Standard Code for Information Interchange
AT	Assistive Technology
ATAG	Authoring Tool Accessibility Guidelines
AUI	Aural User Interface
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CAT	Computer-Assisted Translation
CFP	Call for Participation
CL	Controlled Language
CMS	Content Management System
CSD	Context Segment Definition
CSS	Cascading Style Sheet
CSV	Comma-Separated Values
DTP	Desktop Publishing
DV	Dependent Variable
E2R	Easy to Read
EOWG	Education and Outreach Working Group (W3C context)
FAE	Functional Accessibility Evaluator
GIF	Graphic Interchange format
GLM	Generalised Linear Model

GUI	Graphical User Interface
HCI	Human-Computer Interaction
HOCL	Human-Oriented Controlled Language
HSD	Honest Significant Difference (Tukey's test)
HTML	Hypertext Markup Language
ICT	Information and Communication Technology
IDE	Integrated Development Environment
IDML	InDesign Markup Language
IE	Internet Explorer
ISO	International Organisation for Standardization
IV	Independent Variable
JPEG	Joint Photographic Experts Group
KWIC	KeyWord In Context
LISA	Localisation Industry Standards Association
LMM	Linear Mixed Model
LRC	Localisation Research Centre
Max	Maximum
Min	Minimum
MOCL	Machine-Oriented Controlled Language
MT	Machine Translation
NLP	Natural Language Processing
OCR	Optical Character Recognition
OR	Odds Ratio
PDF	Portable Document Format
PL	Plain Language
PNG	Portable Network Graphic
POS	Part of Speech
POUR	Perceivable, Operable, Understandable, Robust (WCAG 2.0 principles)
PSD	PhotoShop Document

Q^1	First Quartile
Q^3	Third Quartile
QA	Quality Assurance
RDWG	Research and Development Working Group (W3C context)
SC	Success Criteria
SD	Standard Deviation
SE	Standard Error
SEO	Search Engine Optimisation
SVG	Scalable Vector Graphics
TIFF	Tagged Image Format File
UAAG	User Agent Accessibility Guidelines
UCD	User-Centred Design
UX	User eXperience
W3C	Word Wide Web Consortium
WA	Web Accessibility
WAE	Web Accessibility Evaluation
WAI	Web Accessibility Initiative
WAIM	Web Accessibility Integration Model
WAT	Web Accessibility Toolbar (by The Paciello Group)
WCAG	Web Content Accessibility Guidelines
WCAG-EM	Website Accessibility Conformance Evaluation Methodology
WIMP	Windows, Icon, Mouse, Pointer
WYSIWYG	What You See Is What You Get
XHTML	Extensible Hypertext Markup Language
XLIFF	XML Localisation Interchange File Format
XML	Extensible Markup Language

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is the use of a book," thought Alice, "without pictures or conversations?"

In Lewis Carrol, *Alice's Adventures in Wonderland*

Chapter 1

Introduction

This thesis deals with the challenge of assuring access to digital information by blind people within the context of the multilingual Web, with a particular focus on image accessibility. This chapter provides an overview of the work we conducted on this area by describing the research background, arguing the reasons that motivated our work, stating the goals and the methodology followed to attain them and introducing how the thesis content is structured.

1.1 Background

Information and Communication Technologies (ICTs) permeate almost every aspect of our daily lives. They have shaped the way in which knowledge is produced and shared within multiple contexts, including the professional, political, educational, economic and social spheres. In this ever-increasing technologised era where digital literacy has become the norm (Folaron 2012), the World Wide Web (the Web) is now seen as an indispensable commodity. As its inventor puts it, the power of the Web —itself built on the open Internet— derives from the fact that documents are put online in standard form and then linked together, and "the universality and flexibility of this linking architecture has a unique capacity to break down boundaries of distance, language, and domains of knowledge" (Berners-Lee 2007).

The World Wide Web Consortium (W3C) is the international body that develops protocols and guidelines to ensure the long-term growth of the Web under the vision of *One Web,* which in turn relies on two main design principles: Web on Everything and Web for All.[1] While the former is grounded on fostering the development of technologies that enable web access anywhere, anytime, using any device, the latter aims at pursuing the full potential of the Web as a social construct. The rationale behind the Web for All paradigm is that the information contained on the Web should be available to all people, regardless of the hardware and software they use, their native language or culture, their geographical location, or their sensory or interaction modes —which may depend on the person's physical or mental ability.

[1] https://www.w3.org/Consortium/mission.html#principles Last access: 2nd February 2016.

The advancements achieved up to present with regard to human-computer interaction (HCI) technologies for people with disabilities (see section 1.1.1), together with the proliferation of web accessibility (WA) standards (see section 1.1.2) and the growing demand of internationalised and localised web content (see section 1.1.3), have been key to the achievement of a more inclusive Web for All over the last 30 years. By adopting an interdisciplinary perspective, this thesis builds upon the three aforementioned milestones of the information society to advocate for a higher level of accessibility awareness in the production of multilingual websites and, more specifically, in the manipulation of graphical content that occurs throughout the process (see section 1.2).

1.1.1 Access to the Web by people with disabilities

Disability is part of the human condition. The Word Health Organisation (WHO) predicts that almost everyone will be temporarily or permanently impaired at some point in life (WHO 2011, 3). This statement is grounded on the fact that disability is understood not only as a feature of the person, directly caused by disease, trauma or other health condition, but also as a socially-created problem that is not an attribute of an individual (WHO 2002, 8–9), but that rather depends on the overall environment in which the person lives. When such definition is applied to the HCI context, one could argue that users are handicapped not only by the constrained modalities in which they operate with digital content, but also by the technology used to convey that information, and the technology they depend on to perceive it (Harper and Yesilada 2008a, xv).

Within the framework of this thesis, we are particularly interested in the interaction between web-related technology and people who cannot rely at all on their visual modality to perceive digital information.

1.1.1.1 People with visual impairments and the Web

According to the World Health Organisation, there are four levels of visual function: normal vision, moderate visual impairment, severe visual impairment and blindness. The second and third level are often grouped under the category "low vision" which, together with blindness[2], are referred to as visual impairments (WHO 2014). Worldwide, 39 million people are estimated to be blind, and 246 million to have low vision (ibid). In Spain, the latest report published on this issue indicates that around one million people are registered as visually impaired, 71,000 of whom are blind (Ernst & Young 2012, 28). Proportionally, the numbers in Switzerland are also relatively high: from the 325,000 who have reported a visual

[2] The legal definition of blindness may vary across countries. The World Health Organisation states that people are blind when they have a 3/60 or lower visual acuity, as per the Snellen Test (WHO 2016).

impairment in the country, approximately 10,000 Swiss citizens suffer from a blindness condition (UCBA 2012, 4).

Prior to the 1980s, the primary information resources for blind people were limited to a few books, some as paper Braille and a few recorded on audio tapes (Asakawa 2014). With the advent of personal computers as a “structured dynamic communication medium” (Winograd and Flores 1987, 176) and the Web as a “global information universe” (Berners-Lee 1992), individuals with vision disabilities became more autonomous, especially blind people: without a sighted person's assistance, blind users can now carry out everyday activities on their own, such as shopping online, reading newspapers, performing banking transactions or simply engaging in correspondence with friends. Today, the Web is for many of them an essential source of information, employment and entertainment (Harper and Yesilada 2008a, 1).

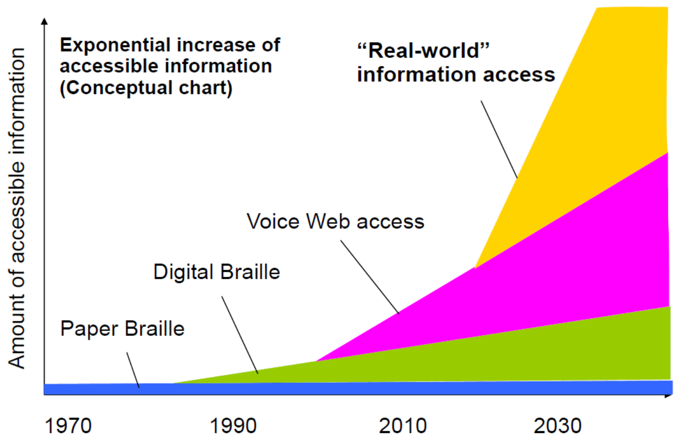


Figure 1.1. Exponential growth of information resources for blind people (Asakawa 2014)

All the above has been made possible thanks to the significant efforts devoted since the early 1990s by both the industry and academia to facilitate the access of blind users to computers and the Web. These have mainly focused on the conversion of visual-based information into one or more different media taking into account the sensorial capacities of the users in order to help them augment their cognition and to increase the comfort and accuracy of information transactions (Freitas 2010, 272). The Web being a visual medium par excellence, the challenge was to provide blind users with a similar HCI experience to those who could see the screen, by allowing them to intuitively navigate through web pages and enjoy surfing the Internet without using their eyes (Asakawa 2005). In this context, most approaches aimed at presenting output information from the computer through the auditory channel and through tactile devices (Barreto 2008, 10). All these efforts resulted in two significant technological innovations: Braille digitalisation and voice-based web access (Asakawa 2014). While the next major challenge in making information accessible to the blind is to develop advanced cognitive systems that understand people, objects, and the surrounding environment and present these concepts for them (see Figure 1.1), traditional assistive technologies as the ones we will present in the following section remain, to date, the most widely used resources for non-visual rendering of digital content.

1.1.1.2 Assistive technologies used by blind users

The term assistive technology (AT) is used within the context of HCI Studies to refer to any piece of software or hardware designed to facilitate the use of computers by people with impairments (DRC 2004, 1) that cannot use the conventional technologies needed (known as WIMP: Windows, Icon, Mouse, Pointer) to access a graphical user interface (GUI). Functionality provided by assistive technology includes alternative presentations (e.g., synthesised speech or magnified content), alternative input methods (e.g., voice), additional navigation or orientation mechanisms and content transformations (Caldwell et al. 2008). As in the case of web browsers, media players and other programs that help in retrieving, rendering and interacting with web content, ATs act as user agents (ibid).

The difficulties experienced by people with low vision in terms of web access are different to those of blind users. While the former need ways of enhancing the visual output to accommodate their level of vision, the latter need non-visual alternatives to the content displayed on the screen (Edwards 2008, 150–151). Partially sighted users tend to solely rely on the assistance provided by software applications that increase the size of text or images, commonly known as screen magnifiers (Paciello 2000, 71). Blind people, however, use a wider variety of ATs which include (but are not limited to) refreshable Braille displays, voice browsers and screen readers. A refreshable Braille display is a tactile device that displays Braille characters and is used to read text output. Voice browsers and screen readers are software that create a vocal rendering of the contents shown on the screen of a device (e.g. a computer or a smartphone). It is worth noting that the use of these ATs is not mutually exclusive; for instance, blind users can use both a Braille display and a screen reader at the same time when interacting with a computer, or switch from one system to the other at their convenience.

The main drawbacks of refreshable Braille displays are that they are extremely expensive, have a limited character capacity and require the user to be proficient at reading Braille (Barreto 2008, 10). Similarly, voice browsers (also referred to as speech or talking browsers) have important limitations. Since they can only interpret content structured using HyperText Markup Language (HTML), they lack the flexibility of screen readers, which can make a whole range of software accessible (Edwards 2008, 154). In addition, since talking browsers are self-voicing, they cannot be used in parallel with a screen reader, whose use is fundamental to access the computer in the first place and launch the browser itself (ibid). For all these reasons, screen readers are the most popular text-to-speech software among blind users.

Screen readers present the information that would be displayed to a sighted user as synthesised speech, providing a number of features for speeding and simplifying the search of relevant information in the source being explored (Barreto 2008, 10). Visual design relies on the fact that a sighted reader can choose which portion of the text to focus on at any time (Edwards 2008, 155). Screen readers replace this functionality through a shortcut-driven interface that allows users to navigate between webpage elements in a serial manner. For example, to access elements of certain types, the user can request, through a keyboard command, to open an auxiliary window containing a list of all elements of that type[3] (Borodin et al. 2010). This way, screen reader users can get a mental or "off-screen" model of a given page by just hearing, for instance, how many headers and links the page has (Theofanos and Redish 2006). Therefore, for blind users, being able to browse the Web is entirely dependent on the page having a suitable semantic structure for the AT to use (Connor 2012, 37). The robustness of the page (i.e. how well it is designed for continuous operation with a very low failure rate) will depend, in turn, on the abilities of its creator (or the capabilities of the tool they use) to adequately implement web design best practices. The interdependency between and complementarity of these and other components of web development will be further discussed in section 1.1.2.

1.1.1.3 Non-visual access to web graphical content

The significant human capabilities in terms of visual acuity, contrast sensitivity, field of view and colour vision, along with the increasing performance characteristics of computer displays, have encouraged designers and developers to fully exploit the Web as a visually-rich interface (Barreto 2008, 8). A study conducted by Asakawa (2005) revealed that the number of images on the Web increased by more than four times the amount between 1996 and 2005, a fact that was confirmed later by Chen and Harper (2008), who observed a significant growth in the usage of graphical formats in web documents over a ten-year period (1999-2008). In a Web increasingly populated with graphical content, the screen-reading process poses some challenges.

Screen readers convert a two-dimensional document to a one dimensional text string, transforming the traditional GUI into an aural user interface (AUI) and making audio the dominant medium (Freitas 2010, 273). This "linearisation" of the

[3] There are different screen readers available on the market −the most popular ones are JAWS, Window-Eyes, the freeware NVDA and VoiceOver, the built-in screen reader of Mac OS X− so there is no standard command to activate this element retrieval feature. For instance, to display a list of all headings of one page, JAWS users would press Insert+F6 (Connor 2012, 38), while Window-Eyes users would need to first use Ins+Tab and then press Alt+H (Thatcher et al. 2006, 111).

web page (Thatcher et al. 2006, 105), which implies a change of channel —visual into auditory—, can result in a considerable loss of information if graphical content is not associated with a non-visual alternative representation that serves the equivalent purpose. The accuracy and effectiveness of the communication act between the blind user and the computer thus depends (i) on the existence of such alternative content, and (ii) on the ability of screen reading software to identify it as such and convey it to the user.

For blind people to perceive graphical content such as images, these must be associated with a text equivalent. The HTML `alt` attribute, introduced for the first time in 1995 as part of the HTML 2.0 specification to provide this functionality (Berners-Lee and Connolly 1995) is, since then, the most widely adopted technique to render meaningful images on the Web accessible. Images are usually inserted in a web page by means of the `<img>` element. When a screen reader gives focus to this element, the software indicates the presence of the image and reads the corresponding `alt` attribute value.[4] For instance, should a blind user find the image depicted in Figure 1.2[5] when consulting a page, they would listen to the following information: "Graphic: The white sandy beach of Rodas in the Cíes islands, located in Galicia, Spain, on a summer day with a clear blue sky." If the `alt` attribute was missing, the screen reader would still announce that an image exists, but no textual replacement would be provided, with the resulting downgrade of the web browsing experience. Similarly, the interaction between the Web and the blind user would suffer if the alt text was misleading, i.e. if it did not represent what is truly shown on the image.

```
<body>
...
<img
src="myweb-en/images/praia-rodas-
cies.jpg" alt="The white sandy beach
of Rodas in the Cíes islands, located
in Galicia, Spain, on a summer day
with a clear blue sky.">
...
</body>
```

Figure 1.2. Visual and non-visual rendering (in HTML) of an image

[4] The value of this attribute is often referred to as "alt text". Henceforth in this thesis, we will use the terms "alt text" and "text alternative" indistinctively when referring to the `alt` attribute content. Other synonyms that may appear include "text equivalent" or "alternative text". These terms are generally preferred over others such as "alternate text", "alternate textual descriptions" or "alternative text descriptors", which will only be used when appearing within a citation. When referring to the HTML `alt` attribute itself, the word `alt` will be highlighted using a different font.

[5] The HTML code excerpt shown in this figure is intentionally presented in a different font from the one used in the main body of the text. Similarly, emphasis was added by the author for easier identification of HTML elements and attributes. These writing conventions will be maintained throughout the entire thesis.

In this thesis, text alternatives are considered the gateway to image accessibility and represent our main object of study (see section 1.2.2). Considerations about their implementation and quality evaluation will be addressed in detail in Chapter 3.

1.1.2 Web accessibility conformance

Understanding the needs of people with disabilities and how they interact with the Web is crucial for designing accessible websites, but it might not be sufficient to guarantee a smoother human-computer interaction for this population group. Web accessibility is also determined by technical aspects, such as conformance with W3C technical specifications and WA best practices. These are the primary focus of the activities carried out by the W3C Web Accessibility Initiative (WAI), launched in 1997 to develop guidelines and resources to help make the Web accessible.

1.1.2.1 Defining web accessibility

As can be inferred from what has been discussed thus far, web accessibility means that people with disabilities can perceive, understand, navigate, interact with and contribute to the Web (Lawton Henry 2005b). According to the WAI, web accessibility depends on different technical and human components working together (Chisholm and Henry 2005, 32). The former include the user agents used by end users, the authoring and evaluation tools used by professionals involved in the web life cycle, and the technical specifications that describe the features of languages that are used to create and deliver web content. The latter (human components) encompass not only the end users, but also the content producers: those who design, code, create and edit web content (ibid). Thatcher et al. (2006, 14–15) use image text alternatives to illustrate the interdependencies between all these components, as summarised below:

- Technical specifications provide a standard for marking up text alternatives for images (i.e. the `alt` attribute).
- Authoring tools enable, facilitate and promote the use of the `alt` attribute.
- Evaluation tools allow content producers to check that alt texts exist and help them determine if they are appropriate.
- User agents provide human and machine interfaces to the alt text.
- Content producers implement text alternatives via the `alt` attribute, often using authoring tools and evaluation tools.
- Users perceive alt texts using their assistive technologies and a browser.

When an accessibility feature, such as the alt text, is effectively implemented in or by one sub-component, the other ones are more likely to implement it (ibid). In an attempt to promote this vision and enhance the interaction between all the above-mentioned components, the WAI published, between years 1999 and 2000, a series of accessibility guidelines based on the fundamental technical specifications of the Web (see Figure 1.3) [6] : the Authoring Tool Accessibility Guidelines (ATAG), the User Agent Accessibility Guidelines (UAAG) and the Web Content Accessibility Guidelines (WCAG). Within the framework of this thesis, we are interested in the last set (the WCAG), as they prescribe how to make web content accessible (including images) to people with disabilities.

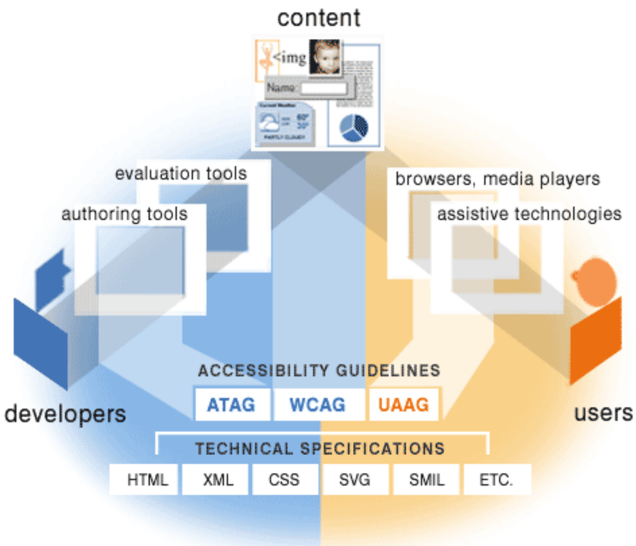


Figure 1.3. Essential Components (and Interaction) of Web Accessibility

1.1.2.2 The Web Content Accessibility Guidelines (WCAG) 2.0 documents

The first version (WCAG 1.0) was published as a W3C Recommendation in May 1999 (Chisholm et al. 1999). After a revision process of five years, version 2.0 was launched in May 2008 (Caldwell et al. 2008) with significant changes and is considered nowadays a seminal reference work –it became an ISO/IEC standard (40500:2012) in 2012 (International Organization for Standardization 2012a).

The first difference worth highlighting in the WCAG 2.0 with respect to WCAG 1.0 is the organisation of the content. While the latter consisted of 14 guidelines, the former revolves around four principles, which form the acronym POUR:

[6] Image by Michael Duffy, from: Essential Components of Web Accessibility. S.L. Henry, ed. W3C (MIT, ERCIM, Keio). Status: Updated August 2005. www.w3.org/WAI/intro/components Last access: 2nd February 2016.

1. Content must be *Perceivable* (P); 2. Interface elements must be *Operable* (O); 3. Content and controls must be *Understandable* (U); and 4. Content should be *Robust* (R) to work with current and future web technologies. In total, there are now 12 guidelines associated with these principles, which in turn include 61 success criteria (SC) that serve to determine the degree to which each guideline is met.[7] Success is measured up to three levels: A, AA, and AAA. The more As, the higher the level of accessibility, with AA as the typical threshold level for accessibility certification. The novelty of the WCAG 2.0 is that these SC are defined in the form of testable statements, thus solving a major criticism of the previous version.[8]

In addition, it should be mentioned that WCAG 2.0 were designed to be more informative and instructive than the WCAG 1.0. To this end, not only did the W3C include a wide variety of techniques as an extra layer of guidance in the WCAG 2.0 document itself to help content producers achieve conformance, but they also developed a comprehensive set of documentation materials to support the correct implementation of the WCAG 2.0 (Lawton Henry 2008) (see Figure 1.4):

- *How to Meet WCAG 2.0* - A customisable quick reference to WCAG 2.0 that includes all of the guidelines, success criteria, and techniques for content producers to use as they are developing and evaluating web content.
- *Understanding WCAG 2.0* - A guide to understanding and implementing each guideline and success criterion in WCAG 2.0.
- *Techniques for WCAG 2.0* - A collection of techniques and common failures, each in a separate document that includes a description, examples, code and tests.

At a broader level, this thesis attempts to argue why localisation professionals, in their role as content producers involved in the multilingual Web development cycle (see section 1.1.3), should also be responsible for achieving an acceptable level of accessibility conformance in the target Web product, not only by following the WCAG 2.0 and related documents, but also by appropriately using accessibility-oriented evaluation tools. At a more specific level, we seek to study the implementation and evaluation of WCAG 2.0 Guideline 1.1 with regard to images, which prescribes the following: *Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language* (Caldwell et al. 2008). As indicated

[7] The list of guidelines associated with each one of this principles can be found in Chapter 2, Table 2.1.

[8] Considerations about the testability of the WCAG 2.0 success criteria will be further addressed in Chapter 2, section 2.3.

earlier, this can be achieved through the appropriate use of the `alt` attribute, both in monolingual and multilingual websites.

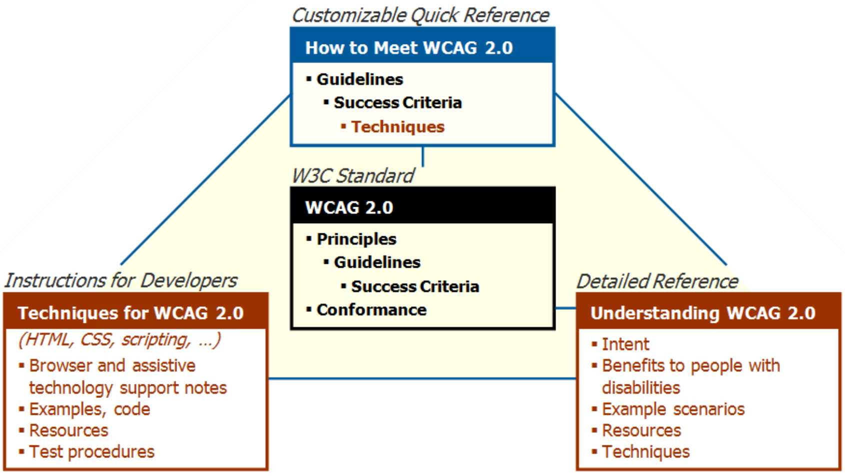


Figure 1.4. Documentation material provided by the W3C to achieve web accessibility (Lawton Henry 2008)

1.1.3 Understanding the production of the multilingual Web

In general terms, web localisation can be described as the translation, engineering and testing of web content (Esselink 2000, 3). By ultimately offering a native language web experience, localisation contributes to the Design for All principle of the W3C and represents the central activity in the production of a multilingual website. While considerations about the localisation process will be addressed in a more comprehensive manner in Chapter 2, we find it relevant to introduce the topic here and lay the foundation for later describing the motivation of our work (section 1.2).

1.1.3.1 The GILT model

In an increasingly globalised digital society, the multilingualisation of the Web has not gone unnoticed. Behind are the early days of the Internet era, when users had to read English online content because there were few other alternatives. Today, websites adapt to the languages spoken by digital information and services consumers worldwide. This phenomenon originates in the wider paradigm of globalisation, understood as "the transformation of business and processes to support customers around the world, in whatever language, country, or culture they require" (LISA in Jiménez Crespo 2013, 25). Globalisation efforts unavoidably trigger a succession of interrelated processes commonly known as GILT (Cadieux and Esselink 2002): the internationalisation, localisation and translation of digital products, such as websites.

According to Pym (2014, 119), internationalisation is “the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design”. When internationalisation is considered as a mindset, and not an afterthought, it contributes to making localisation smoother. During the development phase, the aim is to generate a “neutral enough” website, so that it can be easily adapted later, during the actual localisation process, to make it attractive to users in particular locales[9] (Pym 2011, 413). This can achieved, for instance, by offering support for international natural language character sets or by facilitating the identification and modification of translatable strings (e.g. avoiding hard-coded text in images), to name just a few techniques (Esselink 2006, 23).

After the internationalisation of the source Wwb product, the localisation phase would cover not only the translation of text, but also the adaptation of other non-textual elements (e.g. images, colours) in order to meet the requirements of the target audience (Fernández Costales 2009). Although it is not entirely uncommon for one person to be entirely responsible for the localisation task (Gouadec 2007, 43), in the GILT model localisation is considered as an independent step from the translation itself, covering the preparation and management of the web content to be translated, the post-translation engineering tasks (such as image editing or the adaptation of HTML code) and the final quality testing (Jiménez Crespo 2013, 26). We argue that the extent to which localisation and translation activities are separated is determined by a complex array of factors, including the size and scope of the project, its technical complexity and the human and financial resources available.

This thesis moves away from the hierarchical and linear structure initially imposed by the GILT model (Montalt i Resurrecció 2003) in two ways. Firstly, translation and localisation engineering tasks will be treated as a seamless endeavour that can be carried out by a single person, the localiser,[10] who would be expected to have the necessary linguistic and instrumental competences to successfully deliver a (culturally and technically) functional target web product. The inevitable overlap between localisation and translation, that has been already discussed by multiple scholars in the field (Sandrini 2005; Nauert 2007; Fernández Costales 2009; Schäler 2010), especially when addressing the multimodal essence of the Web, will also be central in our research. Secondly, following Jiménez Crespo's

[9] In the localisation industry, the term locale refers to a group of people that share the same language and cultural conventions. The group may or may not be in the same physical location, hence the definition of locales by language-country pairs (Cadieux and Esselink 2002). For instance, the locale French-France (FR-FR) is different from the locale French-Switzerland (FR-CH).

[10] In our work, the localiser accomodates the combined roles of translator and localisation engineer. Other authors, like Jiménez Crespo (2013, 26) or Mata Pastor (2005, 247) have used the forms “translator-localiser” or “traductor/localizador” to designate such a function.

proposal (2013, 27), we intend to advocate for a more interactive GILT cycle, in which communication and knowledge exchange between the different actors involved in the website development process and the localiser occurs in both directions (top-bottom and bottom-up).

1.1.3.2 Web localisation as user-centred process

Away from considering it as a merely business-driven activity to reach international markets, as was the case when it emerged in the 1980s (Dunne 2015), we see localisation as a process where the context of reception and the end user play a critical role. Within Translation Studies, researchers have often explained this target-oriented nature of the localisation task by adopting a functionalist approach (Jiménez Crespo 2009b; Fernández Costales 2009). Functionalism overshadows the strict notion of equivalence and the overriding relevance traditionally given to the source text, establishes that the purpose or *Skopos* of the translational action is of utmost importance, and assumes that functionality is achieved when the translation meets the receiver's expectations, needs, previous knowledge and situational conditions (Nord 1997). In the context of this thesis, the concept of functionality is still linked to the end user's environment, but it is expanded to also cover the effectiveness of the user interactions with web content from a more technical perspective.

In this sense, we embrace the definition of localisation proposed by Sandrini (2008, 9), who describes it as “the process of modifying an existing website to make it accessible, usable and culturally suitable to a target audience”; in other words, we agree with the fact that the ultimate goal of localisation is to assure the proper functioning of the resulting target website. By placing the accent on the end user, Sandrini states that the aim of the localisation task should be that people from a specified locale can use the target product without any difficulty in their own language (ibid). In our work, we understand that users from the same locale share the same language and similar cultural conventions, but they do not necessarily have the same physical and sensory abilities, which affect the way in which they interact with web content. Furthermore, we believe that difficulties encountered by users when browsing a localised website can be associated not only with problems in terms of linguistic and cultural adequacy, but also with functionality-related obstacles that the commissioner of the task failed to identify in the source and/or that the localiser could not amend in the final target product. When extrapolated to our main object of study, difficulties might occur if a localised website includes images that (i) lack an `alt` attribute, (ii) contain untranslated alt texts, or (iii) have inappropriate text alternatives which do not convey the same meaning or purpose of the images they are associated with.

1.2 Motivation

After more than 15 years since the publication of the first Web Content Accessibility Guidelines (WCAG), studies show that very few websites are fully compliant (Lopes et al. 2010; Harper and Chen 2012; Power et al. 2012). Although over the last two decades the Web has experienced some improvements in terms of accessibility, researchers believe that not all advances can be attributed to a higher focus on accessibility per se by web professionals, but rather to (i) changes in coding styles to attain both better cross-browser consistency and cross-device compatibility, as well as to (ii) the exploitation of new browser features to enhance page layout and design (Richards et al. 2012; Hanson and Richards 2013). Webmasters, web developers and web designers are well aware of the societal benefits of embracing web accessibility (Yesilada et al. 2012; Putnam et al. 2012), but they often lack the time and even training to implement it (Lazar et al. 2004; Trewin et al. 2010). Some also allege that accessibility guidelines are difficult to comprehend and that web accessibility evaluation (WAE) tools do not offer the adequate support needed (ibid).

This thesis is motivated by the incentive of understanding whether these difficulties have also been reported in the case of the multilingual Web and whether localisers have a role to play therein. Ideally, accessibility considerations should be built into the everyday practices across the full web product life-cycle (Cooper et al. 2012), from conception and development, to delivery, maintenance and localisation. In our work, we approach this topic first from a general perspective, looking into the accessibility measures adopted in multilingual sites (see section 1.2.1), to then move to a more concrete area of study: the accessibility of images on the Web and its localisation (see section 1.2.2).

1.2.1 Accessibility of the multilingual Web

Up to now, the study of accessibility implementation and evaluation of multilingual websites has not received too much attention in the literature, despite the fact that, as Folaron proposes (2012, 25), the Web is now a "space of translation" par excellence, where translation practices play a key role in keeping global and local networks alive, maintaining a fluid communication among their users.

The W3C indicates that when websites have multiple versions that are independent of one another in use (for instance, a website in different languages, with different URLs), each version should be assessed for accessibility independently (Velleman and Abou-Zahra 2014). It is thus assumed that in the case of highly localised websites or culturally customised websites[11] (Singh and Pereira 2005), the developers of the source web products are not necessarily

[11] The different levels of website localisation are reviewed in Chapter 2 (section 2.4.1).

responsible for the accessibility of the target ones. Similarly, in the WCAG 2.0 document, one of the few references made to accessibility conformance in a multilingual web context is the possibility of acknowledging that a website or page is only partially compliant with the guidelines when only content in one of the languages available has been checked (Caldwell et al. 2008). Other than that, and apart from the work done by the W3C Internationalisation (I18n) Activity groups to foster the Design for All principle, no official W3C document or group explicitly addresses how accessibility can be assured in multilingual websites. Similarly, no research-oriented WA studies have been reported to date that provide empirical or theoretical evidence about this issue.

Within the field of localisation, accessibility has been pointed to as one of the paradigms that nurtures the interdisciplinarity of Localisation Studies (see Figure 1.5). Some authors have discussed the shared interests between web localisation and accessibility, claiming that when an accessible website is rendered multilingual, web localisation professionals are expected to ensure that accessibility achievements are maintained across the different website language versions they are working on (Gutiérrez y Restrepo and Martínez Normand 2010; Tercedor Sánchez 2010). Furthermore, it has been claimed that the localised version should never be less accessible than the original, and that depending on how much freedom is given to the localisation team, the localised version of a website could even provide more functionality than the source to accommodate the needs of a wider range of users (ibid).

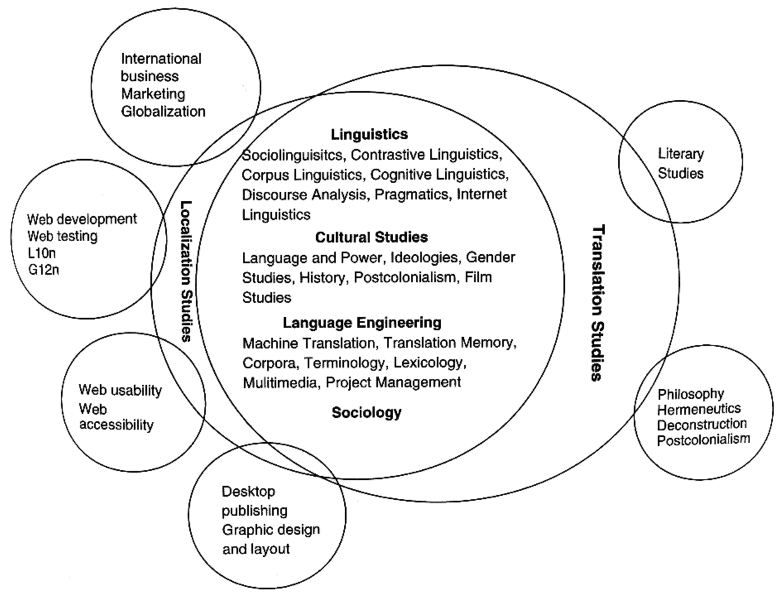


Figure 1.5. Map of Localisation Studies depicting the interdisciplinarity of the field, as proposed by Jiménez Crespo (2013, 135)

Pym (2011, 424) suggests that accessibility is one of the ethical problems that website localisation is facing today. The need for the localiser to tackle the issues that make it difficult for a person to interact with the Web has been acknowledged by certain scholars interested in the field (Tercedor Sánchez 2010; Jiménez Crespo 2009a), but only a few have considered it as a necessary step of the quality assurance process (Gibb and Matthaiakis 2007; Jiménez Crespo 2013). After observing the success of a series of seminars on web accessibility taught to localisation students (Rodríguez Vázquez 2014), we are determined to raise awareness on accessibility-related issues among the web localisation community, which is still believed to be very low (Ó Broin 2004).

1.2.2 Image accessibility and its localisation

The importance of the Web for information dissemination is unquestionable, but the dominance of visual design on the Web leaves visually impaired people at a disadvantage (Harper and Chen 2012). Blind users experience distress and frustration when using the Web due to accessibility barriers, leading them to lose, on average, 30.4% of the time they spend on the computer (Lazar et al. 2007). One of the major causes of frustration for this population group is the inaccessibility of images (Paciello 2000; Asakawa 2005; Petrie et al. 2005; Lazar et al. 2007). The work presented in this thesis reflects our willingness to contribute to the elimination of this barrier, and is motivated by (i) the problems that still exist to generate appropriate text alternatives for images, and (ii) the paucity of empirical research focusing specifically on how this issue is addressed during web localisation.

1.2.2.1 Production of image text alternatives

Despite the fact that ensuring the existence of text alternatives to non-text content is crucial for blind individuals to successfully participate in the information society, recent work still reports low conformance rates concerning image accessibility (Access for All 2011; Hanson and Richards 2013). Although the ratio for missing `alt` attributes has been decreasing consistently in the Web over the last few years (Asakawa 2005; Richards et al. 2012), the main problem is now the presence of uninformative text alternatives (ibid). We argue that this might be for three main reasons:

Limited guidance for appropriate text alternatives formulation

While adding an `alt` attribute to an `<img>` element when introducing images in a web page appears to be a simple task, creating an appropriate text equivalent might not be as straight-forward. The value of the `alt` attribute should communicate the purpose of the image or explain its meaning, that is, a text alternative should serve as a replacement for the image, and not as a mere description of its content. Elaborating an appropriate text alternative requires therefore not only analytical and language skills, but also a significant time investment. Relevant guidance on what information to include in the `alt` attribute (i.e. the alt text composition) usually comes in the form of long and hard-to-understand official documents —for instance, the ISO Technical Specification 20071-11 (International Organization for Standardization 2012b)— by which web professionals, with a more technical-oriented profile, might feel overwhelmed. In addition, language-oriented tips about how to write appropriate text alternatives (i.e. hints about their formulation) are scarce and hardly any references can be found in the literature with regard to preferred lexicon or which syntactic structures to use.

Poor automated support for image accessibility assessment

When content producers lack the time or training needed to put existing guidance on how to render images accessible into practice, they may rely on automated evaluation solutions to look for a quick solution or simply bridge the knowledge gap. When verifying web content against WCAG 2.0 Guideline 1.1, generic web accessibility evaluation (WAE) tools present `<img>` elements that lack an `alt` attribute as problematic, but little feedback is provided about the appropriateness of the alt texts contained in the page or any tips to improve it. Warnings related to image accessibility commonly remain too vague (e.g. "Ensure that the alternative text conveys the content of the image") and tend to be considered unhelpful (Petrie et al. 2005). Prior work has attempted to overcome the limitations of current WAE tools (Bigham et al. 2006; Bigham 2007; Olsen et al. 2010), but most studies have focused on the detection of unmeaningful alt texts (e.g. strings of alphanumeric characters, file extensions, file size information) instead of proposing a solution looking into the quality of alt texts formulated using natural language constructions.

Accountability for image accessibility

Apart from the lack of a complete set of guidelines about alt text formulation and the non-specificity[12] of current evaluation tools with regard to the verification of the appropriateness of text alternatives, the third stumbling-block to image accessibility is that there is uncertainty as to who should be held responsible for providing quality alt texts for images. Actors involved in the web life cycle range from web commissioners and engineers to graphic designers and content authors. In theory, if we take into account the knowledge and know-how of the different stakeholders involved in web production, it is the developers who should make sure that the mechanism for adding a text alternative exist (i.e. the `alt` attribute),

[12] Specificity is described by Brajnik (2004, 7) as the number of different possible issues that can be detected and described by a tool. The larger this set, the more capable the tool is of providing specic warnings and suggestions, and therefore the more useful it is.

whereas editors should be the ones generating an informative alt text to insert within the said attribute. In practice, however, following such a strict workflow is challenging, since images might be added, modified or deleted at different web development stages. While this situation perfectly applies to the construction of monolingual websites, it can get worse if web localisation is requested.

1.2.2.2 Translation and adaptation of image text alternatives

The value of the `alt` attribute is usually considered as a text string that needs to be translated during the web localisation process (Gibb and Matthaiakis 2007, 668; Mata Pastor 2009b, 552; Roturier 2015, 88). The importance of such action for the achievement of an enhanced web browsing experience by blind users has even been recognised by experts in the accessibility field (Clark 2002, 99). Nonetheless, to the best of our knowledge, only two localisation studies have investigated the presence of alt texts in translated websites.

Adopting a corpus-based descriptive approach, Jiménez Crespo (2008) analysed a Spanish parallel corpus, comprising both original (N=172) and localised (N=95) websites from private sector companies. He found that the translated sub-corpus contained a higher number of `alt` attributes, indicating that this was probably due to an increasing use of Content Management Systems (CMS), which insert them automatically in the page. Apart from reporting the average of words per alt text, the presence of non-alphabetical characters (spaces, *, /) and the predominance of noun forms, no further data was revealed with regard to the composition and formulation of alt texts. Similarly, Fernández Costales (2010) conducted a large scale study to investigate the localisation of university websites from 27 countries of the European Union. As part of a comprehensive qualitative analysis of different textual, non-verbal and cultural elements in the web document collection gathered, he concluded that the translation of text alternatives receives scant attention during the localisation process of this type of institutional website.

Despite the valuable contributions made by these two studies, the appropriateness of alt texts was not closely examined, probably because the observation of text alternatives represented only one of the multiple aspects analysed in these authors' work. Within Translation Studies, considerations about the composition and formulation of textual descriptions for images have only been explored with reference to technical texts (Prieto Velasco 2009). While some practical criteria have been extrapolated to the particular case of images on the Web (Tercedor Sánchez and Prieto Velasco 2009, 84), no further studies were conducted to explore, for instance, whether these guidelines could be applied by localisation professionals or whether their implementation could be facilitated through automated support.

1.3 Thesis goals

All the above considered, the broader objective of this thesis is to elaborate on the role of web localisation professionals as contributors to the accessibility of the multilingual Web. The scope of our investigation is delimited by three more specific goals related to accessibility assurance, presented below in increasing order of importance:

> **Goal 1:** To investigate whether established procedures exist to ensure the accessibility of multilingual websites in general.
>
> **Goal 2:** To propose an automated testing solution based on controlled language (CL)[13] technology for assuring the generation of appropriate text alternatives for images on the Web during the localisation process, with a view to covering some of the limitations of existing guidance and tools designed for image accessibility evaluation.
>
> **Goal 3:** To determine the extent to which localisers are capable of assuring that an acceptable level of image accessibility is achieved in the target web product.

In our work, we assume that accessibility evaluation should be part of the quality assurance measures taken during the Web localisation process to guarantee a functional target web product. Following Saldanha and O'Brien's definition (2014, 95), we understand quality assurance (QA) as the sum of all systems and processes used to help create or maintain quality. In this sense, QA encompasses all the actions undergone or put in place to ensure (prospectively) and to assess (retrospectively) accessibility.[14] It is worth mentioning that this thesis does not attempt to explore in depth the QA models used in the localisation industry or theorise about the more complex notion of quality in translation. We seek instead to support with empirical evidence (see section 1.4) that accessibility-related issues, such as the appropriateness of localised text alternatives, can be seen by end users as a determining factor of the quality of the target web product.

1.4 Methodology and research questions

To reach the three specific goals set forth in the previous section, we have adopted a multi-method empirical approach, combining descriptive and experimental investigation strategies. In the context of HCI studies, descriptive investigations are often the first step of a research project; they enable researchers

[13] A CL is a subset of a natural language that has been specifically restricted with regard to the lexicon, syntax, and style it uses (Roturier 2006, 47).

[14] This distinction is inspired by the discussion on differents concepts of translational quality in House (2013).

to construct an accurate description of what is happening in a given area of study, and serve as a basis on which to ground subsequent research stages. Experimental investigations, on the other hand, provide an opportunity to explore causal relations (Lazar et al. 2010, 20–22). This thesis has relied mainly on the former to address the first two specific goals (Goal 1 and Goal 2) and fundamentally on the latter to achieve Goal 3, which constitutes the central contribution of our research work to the fields of accessibility and localisation. In what follows, we briefly explain the methodological triangulation of our investigation, depicted in Figure 1.6.

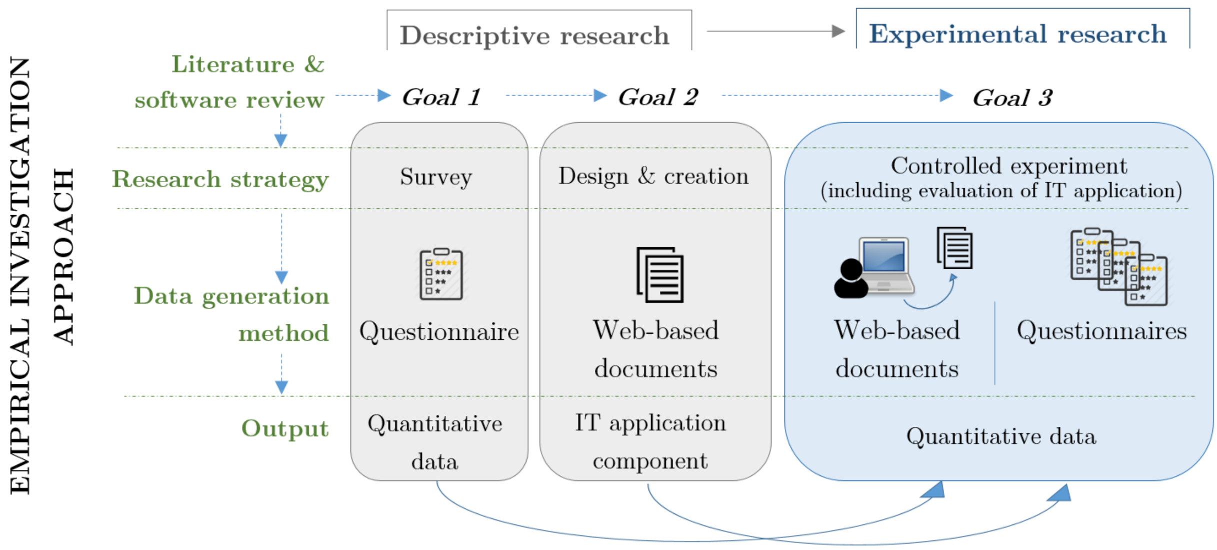


Figure 1.6. Overview of the methodological approach adopted in this thesis

To attain **Goal 1**, we have reviewed the existing literature about the implementation and assessment of web accessibility, as well as the most relevant steps of the web localisation process. The study of the literature has been complemented with a survey of 67 WA experts about (i) the specific procedures they follow in multilingual web accessibility assessment tasks and (ii) their perceptions regarding the potential contribution of localisation professionals to the achievement of a higher degree of accessibility in multilingual websites. Quantitative data about these two concerns were produced using an online accessible questionnaire as the main data generation method.

After a comprehensive literature and software review with regard to image accessibility issues, **Goal 2** was achieved through the development of 40 controlled language rules (CL) for French using a declarative error-based formalism (Bredenkamp et al. 2000) to facilitate the identification of inappropriate text alternatives. The rules in question, whose definition was based on the analysis of two collections of documents (existing guidance on how to write text alternatives and a third-party web corpus) can be applied by means of Acrolinx, a state-of-the-art CL checker. By developing a new IT application component, we followed a research strategy that in the field of information systems and computing is known as “design and creation” (Oates 2005, 107). Oates indicates that, when this strategy is adopted, the IT product “can be the main focus of the research, a vehicle for something else, or a tangible end-product of a project where the focus is on the development process” (ibid, 109). In our work, the CL rules represent an analytical tool for the localiser to address the appropriateness of alt texts, i.e. they are a vehicle for these web professionals to achieve image accessibility. The evaluation of the IT component developed was integrated into the research strategy followed to attain **Goal 3.**

In order to determine the extent to which localisers are capable of assuring that an acceptable level of image accessibility is achieved in the target web product (**Goal 3**), we conducted an experimental study that consisted of two stages: (i) a controlled web localisation experiment (Stage 1) for which we recruited 28 localisers (14 of whom had some previous background on web accessibility); and (ii) a user evaluation (Stage 2), that involved the participation of seven blind users. The study sought to answer the following research questions:

R1. Are image text alternatives considered by localisers as translatable elements during the web localisation process?

R2. Does the use of accessibility-oriented QA tools during the web localisation process result in more appropriate text alternatives for images?

R3. Does knowledge of web accessibility help localisers to produce more appropriate text alternatives for images?

During the experiment, localisers were first asked to localise a three-page website containing 130 images, and then to assess the resulting localised website for image accessibility using two QA tools: (i) Acrolinx, the CL checker for which we developed the CL rules, and (ii) aDesigner, a WAE tool (Asakawa 2005). The order in which tools were used was counterbalanced, in order to avoid any potential bias, so two different QA scenarios were tested: aDesigner-Acrolinx (QA scenario A) and Acrolinx-aDesigner (QA scenario B). The appropriateness of the localised text alternatives was later rated by the screen reader users via a questionnaire.

With a view to answering the aforementioned research questions, we analysed the quantitative data gathered in both stages of the study to measure the effect of having WA knowledge and of using accessibility-oriented QA tools (primary independent variables) on the translation and the appropriateness of the text alternatives produced by the localisers (dependent variables). The impact of other secondary independent variables on the localisation outcome, such as the QA scenario followed by participants, was also measured.[15] The experimental work presented in this thesis focuses, therefore, on the *translation product* (Saldanha and O'Brien 2014, 5). More detailed information about the methodology adopted to achieve **Goal 3** can be found in Chapter 5. The hypotheses and sub-hypotheses associated with our research questions are included in the same chapter, under section 5.4.

1.5 Thesis structure

The content of this thesis is organised as follows:

Chapter 2 revolves around Goal 1, addressing both web accessibility and localisation from a general perspective. First, it offers an overview of how accessibility is implemented and assessed throughout the web development cycle, highlighting the benefits of web accessibility and the challenges that web professionals still face when trying to meet the W3C accessibility standards. Second, it reviews the key aspects of the web localisation process and argues why the localiser should be involved in the achievement of a more accessible Web for all. The chapter also presents the survey conducted with WA experts, introducing the methodology followed and reporting the main findings.

Chapter 3 provides an in-depth description of how images on the Web can be made accessible and localised, which is a key step to attain Goals 2 and 3. First, it reviews the mechanisms that exist to introduce image text alternatives in web documents, it presents which are the most common strategies to write them appropriately, and explores how tools support the image accessibility assessment task. Second, it offers an overview of the topic of image localisation and explains what is needed to assure image accessibility during the localisation process. Both Chapter 2 and Chapter 3 cover the literature and software review that lay the foundation of our investigation.

Chapter 4 presents the outcome of Goal 2. It first examines how controlled language has proved beneficial in prior translation work, as well as the different forms of CL that have been previously applied to the Web. Then, it reports on 40 CL rules aimed at detecting inappropriate text alternatives in French. It describes the rule development procedure, the three rule sub-sets defined and how these rules can be applied using Acrolinx. It also indicates how and which rules were finally selected to be evaluated in the experimental study.

Chapter 5 describes the methodological framework designed to achieve Goal 3. It provides a thorough explanation of all the aspects related to the participants,

[15] Apart from the QA scenario, six other secondary independent variables were also explored. A summary of all the variables considered during data analysis can be found in Table 6.1.

materials, tasks and procedures of the two stages of the experimental study: the controlled web localisation experiment (Stage 1) and the user evaluation (Stage 2). It also emphasises the research questions already presented in this introduction and sets forth the hypotheses and sub-hypotheses associated with the study.

Chapter 6 examines the research questions related to the experimental study by testing our hypotheses and sub-hypotheses and concluding whether they can be supported or need to be rejected. More specifically, it focuses on the analysis of the effect of the two primary independent variables (WA knowledge and use of accessibility-oriented QA tools) and the QA scenario (secondary independent variable) on the translation and appropriateness of localised alt texts.

Chapter 7 presents and discusses the additional findings of the experimental study. First, it covers the results associated with the impact of six other secondary independent variables on the appropriateness of the localised text alternatives, according to the evaluation carried out by the screen reader users. Then, it reports on the subjective data gathered through the questionnaires distributed to localisers after the localisation and QA tasks performed during the controlled experiment.

Chapter 8 summarises our investigation by presenting the main achievements of this thesis with regard to the three specific goals established and the overall contribution to the fields of accessibility and localisation. It also discusses the main limitations of our work and provides ideas for new future research directions.

1.6 Published work

Some of the discussions, methodological aspects and results presented in this thesis have been partially published in the form of short conference papers and scientific journal articles. The key findings from the survey conducted with WA experts presented in Chapter 2 are covered in Rodríguez Vázquez and Bolfing (2013) and Rodríguez Vázquez (2013). Similarly, we have advocated for the empowerment of localisers as facilitators of both web and image accessibility in Rodríguez Vázquez and Torres del Rey (2012) and Torres del Rey and Rodríguez Vázquez (2013). Information about the different CL rule development phases and the preliminary analyses of their impact on the appropriateness of text alternatives for images can be found in Rodríguez Vázquez et al. (2014) and Rodríguez Vázquez (2015a). Finally, the outcome of the experimental study reported in Chapters 5 to 7 has been partially published in Rodríguez Vázquez (2015c; 2016).

Chapter 2

Web accessibility, assessment and localisation

This chapter sets the context of this thesis by discussing the fundamental aspects related to the accessible development of multilingual websites. To this end, we review the key notions related to web accessibility, and how it can be achieved and evaluated. The chapter also describes the most relevant steps of the localisation process to then support, from both a theoretical and an evidence-based perspective, the consideration of accessibility issues as an additional responsibility of localisation professionals.

2.1 Overview

The evolutionary nature of the Web is indisputable. Every day, it continues to expand as more of the information needed to conduct personal, professional, and public facets of our lives moves partially or wholly online (Folaron 2010, 448). This 'digital explosion' in terms of information and services available on the vast World Wide Web has inevitably led to a dramatic increase in the number of Internet users worldwide, with a tenfold scale-up since the beginning of the current century until present.[1] From 2000 and 2015, global Internet penetration grew from 6.5% to 43%, resulting in a total of 3.2 billion web users by the end of 2015 (International Telecommunication Union 2015).

The technological progress that has driven such unprecedented growth in Information and Communication Technology (ICT) access has also enabled a higher representation of languages and cultures on the Web. The emergence of localisation in the 1980s as the practice of "combining language and technology to produce a product that can cross cultural and language barriers" (Esselink 2006, 21) marked the beginning of a market-led international expansion of multilingual digital products. Nowadays, if you are a global business or want to go global, you need locally consumable web content. The 2015 State of Web Localisation Survey conducted by Lionbridge (2015), one of the world leaders in localisation services, revealed that 93% of more than 200 global companies surveyed translate product and services-related web pages, and that almost 40% of those enterprises need content localised weekly. The popular notion of English as the main language of

[1] http://www.internetlivestats.com/internet-users/ Last access: 25th February 2016.

business is being dismantled as studies confirm that users are more inclined to purchase products online when related information is presented in their own language, and when they know that native-language support will be provided after buying (DePalma and Sargent 2014).

Publishing web content in a language that users can read and understand does not only bring benefits to industries and potential customers, but also to governments —in the form of greater community participation and a more inclusive policy environment— and to individuals in general, since access to information in their mother tongue can save lives, preserve people's liberty and support their education and financial well-being (Anastasiou and Schäler 2010). For multilingual web content to bring such economic and social prosperity, however, its access by everyone, irrespective of the people's (dis)abilities, needs to be guaranteed.

This chapter seeks to examine which measures can be taken to achieve the goal of assuring the creation of accessible multilingual websites. This necessarily involves understanding why web content should be made accessible in the first place, regardless of the number of language versions it is available in. Similarly, it requires the examination of web accessibility (WA) best practices and the attitudes towards accessibility of the web professionals who are responsible for implementing them. Whereas section 2.2 covers these issues and discusses which obstacles can hinder the endorsement of web accessibility recommendations, section 2.3 describes how web content can be assessed for compliance with accessibility standards, along with the limitations of the most popular WA evaluation methods. Throughout these two sections, we shall highlight whether the particularities of multilingual websites have been considered so far in the related literature. The second part of the chapter is dedicated to (i) describe the most important aspects of the web localisation process and understand its different relationships with web content accessibility (section 2.4), and to (ii) present a survey on accessibility assessment practices currently followed by web accessibility experts when dealing with websites available in two or more languages, as well as the perceptions of these professionals with regard to the localisers' involvement in WA (section 2.5).

2.2 Rendering web content accessible

2.2.1 Web accessibility, usability and universal design

The most recent definition of web accessibility that can be found in the literature has been proposed by Petrie et al. (2015) and reads as follows:

> "All people, particularly disabled and older people, can use websites in a range of contexts of use, including mainstreams and assistive technologies;

> to achieve this, websites need to be designed and developed to support usability across these contexts."

Before suggesting this definition, its creators reviewed the WA-oriented work published between 1996 and 2014 and collected 50 representative accessibility definitions for inspection,[2] including the one used by the W3C (Lawton Henry 2005b) presented in Chapter 1 (see section 1.1.2.1). Upon a qualitative analysis of the 50 definitions, six core concepts were identified and put together into the aforementioned definition proposal. In what follows, and prior to focusing on the topic of web accessibility more in detail, we want to comment on some of them, as the notions they refer to sometimes get mixed up with accessibility even if they are often reflected implicitly and/or explicitly in accessibility definitions: *universal design* and *usability*.

The concept that was referred to most in the definitions studied by Petrie et al. (2015) (98%, N=49 out of 50) concerned the groups of users that are affected by accessibility and their characteristics, and it was finally covered in their proposition as "all people, particularly disabled and older people". This formulation echoes the notion of universality present in one of the most famous quotes from the inventor of the Web, Tim Berners-Lee: "The power of the Web is in its universality. Access by everyone regardless of disability, is an essential aspect".[3] The idea of the Web as a universal platform (Berners-Lee 2007) is complemented with the notion of universal design when Petrie et al. state that people should be able to "use websites in a range of contexts of use". *Universal design* refers to the process of creating products that are usable by people with the widest possible range of abilities, operating within the widest possible range of situations, without the need for adaptation or specialised design (Lawton Henry and Abou-Zahra 2014; Connor 2012). In this sense, it is important to note that accessibility represents only one aspect of the universal design principle, by primarily referring to the design of products for people with disabilities (ibid).

Another concept that was repeated across the accessibility definitions inspected by Petrie et al. (2015) consisted of the characteristics of any potentially accessible website, the most recurrent of which were related to the notion of usability —hence the last part of their final definition proposal. *Usability* refers to how quickly people can learn to use something, how efficient they are while using it, how memorable it is, how error-prone it is and how much users like using it (Nielsen and Loranger 2006). A usable web product features, therefore, five dimensions: easy to learn, efficient, effective, error tolerant and engaging (Quesenbery 2008). The 5 E's of

[2] The sample included definitions from standards documents, academic papers and books for practitioners, with authors coming from 21 countries. The list of definitions can be consulted in the Appendix of the cited publication (Petrie et al. 2015).

[3] https://www.w3.org/standards/webdesign/accessibility Last access: 25th February 2016.

usability are interdependent and some dimensions might be prioritised over others according to the ultimate goal of the website.

It is often claimed that usability and accessibility focus on different aspects of web usage: while the former refers to the general quality of usage, the latter is more concerned with access to information (Vieritz et al. 2010, 336). The most common approach is to consider accessibility issues as a sub-set of usability problems, based on the simplistic principle that to use a website you need first to be able to have access to it.[4] Thatcher et al. (2006, 26–27) share this belief —acknowledging that there might be a significant overlap between the two— and argue that usability problems impact all users equally, while accessibility problems put people with disabilities at a disadvantage relative to individuals without disabilities. Petrie and Kheir (2007) prefer, however, to distinguish between three categories: (i) "pure accessibility" problems, which only affect people with disabilities; (ii) "pure usability" problems, which only affect non-disabled people; and (iii) "universal usability" problems, which affect both people with and without disabilities. Yet, in the study they conducted with six disabled (blind) and six non-disabled (sighted) people, they found that the problems encountered by the two groups comprised two intersecting sets, with approximately 15% overlap, which suggests that "universal usability" issues are not that common.

Our position is that creating web content following universal design principles increases the likelihood of achieving accessible and usable websites or, in other words, it fosters universal usability[5] (Horton and Leventhal 2008). Some WA experts (Connor 2012) have even suggested that there are some similarities between the 7 Principles of Universal Design (Center for Universal Design 1997) and the WCAG 2.0,[6] which will be later reviewed in section 2.2.3. Similarly, in agreement with practitioners like Paciello (2000, 55), we believe that a website that is designed accessibly often enhances the usability of that web product (Paciello 2000, 55). This is particularly relevant for the population group we are interested in: *the blind community*, as previously stated in the Introduction of this thesis. web usability is about three times better for sighted users than for users with visual impairments (Pernice and Nielsen 2001), hence the need for supporting the implementation of accessibility technical recommendations in addition to usability guidelines.

[4] Nevertheless, the opposite might also be true: having access to a website does not necessarily guarantee a high degree of usability.

[5] Universal usability can be defined as having more than 90% of all households as successful users of information and communications services at least once a week (Shneiderman 2000).

[6] For instance, Guideline 4b *Provide adequate contrast between essential information and its surroundings*, associated to Principle 4 of Universal Design: *Perceptive information,* is closely related to Guideline 1.4 *Distinguishable: Make it easier for users to see and hear content, including separating foreground from background,* which is linked to WCAG 2.0 Principle 1. *Perceivable.*

All in all, it should be mentioned that promoting web accessibility is not only beneficial for the sake of usability, but also for universal design. In the next section (2.2.2), we outline other factors that may drive web professionals to embrace the W3C standards on web accessibility.

2.2.2 Why advocate for web accessibility?

Since the foundation of the W3C Web Accessibility Initiative (WAI), multiple researchers and practitioners have highlighted the benefits of web accessibility based on anecdotal and empirical evidence (Paciello 2000; Clark 2002; Thatcher et al. 2006; Sloan 2006; Connor 2012; Yesilada et al. 2012). This section intends to summarise the political, technical, financial and social arguments that have been put forward so far for providing web accessibility.

2.2.2.1 Legal and policy reasons

One of the factors that might influence the adoption of WA best practices is simply the need to comply with the laws and policies established both at an international and a national level. Over the last few decades, we have seen the proliferation of pieces of legislation focusing on the needs and rights of individuals with disabilities with a view to supporting the principles of equality, non-discrimination and social inclusion.

The Convention on the Rights of Persons with Disabilities,[7] adopted on 13 December 2006 by the United Nations, has been ratified not only by the European Union (EU), but also by individual countries (for instance, Spain ratified it 2011 and adapted its legislation accordingly, while Switzerland did so in 2014). On the issue of accessibility (article 9), the Convention prescribes that State parties shall take the appropriate measures to "promote access for persons with disabilities to new information and communications technologies and systems, including the Internet." In the same vein, the EU has launched several programmes on the subject of ICT access —such as the European *i2010 initiative* on e-Inclusion from 2007—[8] and the recently adopted proposal for a directive on the accessibility of the public sector bodies' websites.[9]

Significant national efforts have also been devoted in numerous countries to promoting the enforcement of web accessibility best practices, particularly after the publication of the first W3C Web Content Accessibility Guidelines (WCAG) 1.0 in 1998. The example of the United States if often referred to in the literature, as

[7] http://www.un.org/disabilities/default.asp?id=269 Last access: 25th February 2016.

[8] http://eur-lex.europa.eu/legal-content/GA/TXT/?uri=uriserv:c11328 Last access: 25th February 2016.

[9] https://ec.europa.eu/digital-agenda/en/news/proposal-directive-european-parliament-and-council-accessibility-public-sector-bodies-websites Last access: 25th February 2016.

a large number of policies and standards have been introduced since the early 1990s to advocate for a more inclusive society. Concretely, the most influential piece of legislation in the country is the mandate for Electronic and Information Technology Accessibility Standards, commonly known as Section 508.[10] This document requires federal agencies that develop, procure, maintain or use ICTs to ensure accessibility to that technology for people with disabilities —both those who are federal employees and other members of the general public. It also specifies that, when involved, industry is responsible for delivering an accessible solution (Paciello 2000, 34).

Similar legislative requirements exist in European countries. In Spain,[11] for example, it is worth mentioning (i) the Royal Decree 1494/2007 by which the Regulations on the basic requirements to ensure that people with disabilities have access to technologies, goods and services related to the information society and social communication media are approved, and (ii) Law 49/2007, which defines the regime of offences and sanctions in the field of equal opportunities, non-discrimination and universal accessibility for people with disabilities, which can receive a fine of up to €90,000 for serious offences. Similarly, Switzerland advocates for the non-discrimination of disabled people through the federal law *LHand* from 2002, as well as other cantonal pieces of legislation.[12] More specifically, websites of the different bodies of the federal government must be compliant with WCAG 2.0 (level AA), according to the federal guidelines for web design P028. In both countries, an accessibility standard exists (UNE 139803 and eCH-0059 respectively), but only in the case of Spain can it be enforced in the private sector.

2.2.2.2 Technical reasons

The implementation of web accessibility best practices should not be observed only as a legal or political obligation, but also as an opportunity to enhance technical performance. As Clark notes (2002, 21) "standards compliance is a form of programming maturity", and accessibility is often seen as a clear attribute of high-quality websites (Thatcher et al. 2006, 42). In addition, increased web accessibility improves interoperability, as it helps enable content in different configurations and promotes designing for device independence. Designing websites with accessibility in mind also decreases site development and maintenance time

[10] http://www.section508.gov/content/learn Last access: 25th February 2016. More detailed information about U.S. WA-related law can be found in Chapter 16 of Thatcher et al. (2006).

[11] A summary of Spanish laws and accessibility-oriented standards is available at: http://www.euroblind.org/convention/article-21--freedom-of-expression-and-access-to-information/nr/1331 Last access: 25th February 2016.

[12] A summary of Swiss laws and accessibility-oriented standards is available at: http://www.euroblind.org/convention/article-21--freedom-of-expression-and-access-to-information/nr/1322 Last access: 25th February 2016.

(Connor 2012, 19). For example, if content presentation style (such as font size, background colour or page layout) is appropriately defined in Cascading Style Sheets (CSS) and not through HTML, time and technical efforts to change the visual rendering of a website over time are reduced. Interestingly enough, this is also well-known internationalisation good practice, as we will discuss later in this chapter (see section 2.4).

It is not uncommon to believe that, despite the technical benefits we have just outlined, accessibility prevents web developers and designers from building technically advanced and visually attractive websites. Nevertheless, some research studies have demonstrated that accessibility does not constrain visual design. For example, after an evaluation of 100 websites by 51 disabled users with a wide range of impairments, Petrie et al. (2004) found that some of the most accessible websites were the ones that had, in fact, complex visual designs, including elaborated graphics and complex layouts. Furthermore, usability studies have shown that visual complexity can equally have a negative effect on the web browsing experience of sighted users (Michailidou et al. 2008) and require a higher cognitive load (Harper et al. 2009), two facts that downplay even more the "accessibility means boring" myth.

2.2.2.3 Economic reasons

The technical and legal motivations explained thus far can make of accessibility a good business case when combined with other financial benefits, such as an enhanced Search Engine Optimisation (SEO) (Gómez et al. 2007). Smith (2011) argues that accessibility and SEO are both about getting relevant content to users and highlights that best practices in both areas are somehow aligned, such as using proper alternative text for images, providing descriptive link texts (i.e. avoiding "click here" or "read more" links) or ensuring that page titles are informative and succinct (ibid). Moreover, rendering web content accessible can result in high return on investment thanks to increase website use rates. If a website is accessibility compliant, a wider range of individuals will be able to use it, thus covering more market segments. Empirical studies have shown, in fact, that people with disabilities, such as blind users, are more likely to interact with websites that are accessible than with those that are not (Bigham et al. 2007; Bigham 2009).

Similarly, companies that put accessibility standards into practice implicitly demonstrate a higher commitment towards the social principle of non-discrimination, which can also bring in financial rewards. The importance given to customer satisfaction can improve the image of the company, which in turn may provoke competitive advantage due to differentiation from direct competitors (Leitner et al. 2014). Last but not least, implementing web accessibility guidelines can lead to cost savings, most of which are the fruit of the technical benefits discussed before. Thatcher et al. (2006, 46–47) point out that accessibility (i) decreases personnel costs that may be derived from website maintenance over the long term, (ii) reduces server costs when accessibility contributes to decrease server loading, (iii) decreases the need for creating multiple versions of a site for different devices and (iv) reduces potential high legal expenses due to non-compliance with web accessibility requirements.

2.2.2.4 Ethic-moral reasons

As alluded to earlier, building accessible websites contributes to the removal of interaction and communication barriers. Greater web accessibility provides equal opportunities for people with and without disabilities to enjoy the large benefits of the Internet at a personal[13] and professional level. A survey conducted with 300 people from the industry and the academia working on WA-related issues revealed that the primary motivations of most respondents to embrace web accessibility were precisely "being inclusive" and "being ethical" (Yesilada et al. 2012). This suggests that, when considering accessibility, individuals are driven by societal factors rather than legal or financial incentives. While one would expect corporations to have a vested interest in the latter, studies have indicated that social aspects, such as equality, ethical behaviour, social commitment and responsible attitude towards society also represent the main drivers for web accessibility implementation among complex organisations in the private sector (Leitner and Strauss 2010; Leitner et al. 2014).

All the above is strongly related to how individuals should be motivated to create accessible web content. When people gain some perspective of individuals with disabilities, they are more likely to want to make their content more accessible to those who may have difficulties accessing it (WebAIM 2013). The notions of 'guilt', 'punishment' or 'requirement' associated with legal factors often overlook the human element of accessibility, while 'enlightening' or 'inspirational' factors, such as fighting against social inclusion, can make web professionals, commissioners and policy-makers understand that embracing accessibility is the right thing to do (Smith 2013). Finally, we cannot forget that, overall, accessibility brings benefits to everyone. People can experience a temporary disability or simply 'be handicapped' by the technology they use or the environment in which they live. As Harper and Yesilada assert (2008a, xvii) "to understand the needs of disabled users is to understand the needs of everyone."

[13] A study has shown that users with visual impairments are in a better mood when they access accessible Web content (Pascual et al. 2014). Advocating for accessibility can thus have a profound influence on the well-being of people with disabilities.

2.2.3 Web accessibility implementation

Throughout the last sections, we have insisted on the importance of building websites with accessibility in mind, but we have not yet closely looked at which best practices can be followed, who is expected to apply them or whether there are still any obstacles to overcome during WA implementation. The following sections cover these aspects and precede the discussion on web accessibility assessment.

2.2.3.1 The fundamentals

As announced in the Introduction of this thesis, the set of accessibility recommendations we are most interested in is the one provided by the W3C, the voluntary international body dedicated to the standardisation of web technologies and led by Tim Berners-Lee, the inventor of the Web. More specifically, we want to focus on the most recent version (2.0) of the Web Content Accessibility Guidelines (WCAG), "developed through the W3C process in cooperation with individuals and organizations around the world, with a goal of providing a shared standard for web content accessibility that meets the needs of individuals, organizations, and governments internationally" (Caldwell et al. 2008).

The POUR principles, briefly introduced in Chapter 1 (section 1.1.2.2) lay the foundation necessary for anyone to access and use web content. They establish that users must be able to (i) *perceive* the information being presented in a website (i.e. information cannot be invisible to all of their senses), (ii) *operate* the interface (i.e. it cannot require interaction that a user cannot perform), (iii) *understand* the information contained in the website as well as the operation of the user interface and (iv) access content that is *robust* enough to be interpreted reliably by a wide variety of user agents, including assistive technologies (Cooper et al. 2015b). Additionally, 12 guidelines and 61 success criteria (SC) were defined to address these principles. While the former are announced as recommendations, the latter describe specifically what must be achieved in order to conform to each guideline. Table 2.1 offers an overview of the organisation of the guidelines, by principle. Conformance to WCAG 2.0 can be achieved (i) by making a website compliant with specific success criteria or (ii) by providing a conforming alternate version.[14] It is worth noting that, when the second option is chosen for websites which have multiple language versions, Caldwell et al. (2008) indicate that conforming alternate versions are then required for each language. To the best of our knowledge, this is one of the few occasions, together with the ones cited in Chapter 1 with regard to WA assessment (section 1.2.1), where reference is made in the WCAG 2.0 to accessibility implementation in the context of multilingual websites.

[14] Caldwell et al. (2008) describe it as a version that, while providing all of the same information and functionality in the same human language as the non-conforming website, conforms at the designated level (A, AA or AAA).

Table 2.1. Web Content Accessibility Guidelines (WCAG) 2.0

No.	Guideline
Principle 1: Perceivable	
1.1	Text alternatives: Provide text alternatives for any non-text content.
1.2	Time-based Media: Provide alternatives for time-based media.
1.3	Adaptable: Create content that can be presented in simpler ways without losing information or structure.
1.4	Distinguishable: Make it easier for users to see and hear content, including separating foreground from background.
Principle 2: Operable	
2.1	Keyboard Accessible: Make all functionalities available from a keyboard.
2.2	Enough Time: Provide users enough time to read and use content.
2.3	Seizures: Do not design content in a way that is known to cause seizures.
2.4	Navigable: Provide ways to help users navigate, find content, and determine where they are.
Principle 3: Understandable	
3.1	Readable: Make text content readable and understandable.
3.2	Predictable: Make web pages appear and operate in predictable ways.
3.3	Input Assistance: Help users avoid and correct mistakes.
Principle 4: Robust	
4.1	Compatible: Maximize compatibility with current and future user agents, including assistive technologies.

For a website (or its alternate version) to reach the minimum level of conformance established by the W3C, it has to meet all the SC labelled as Level A, which are the most important in number (a total of 25). To achieve Level AA conformance, a website needs to satisfy all the Level A and Level AA success criteria (i.e. 25 + 13). The maximum degree of conformance that can be attained is for web content to meet all 61 SC, including the 23 considered as Level AAA. When web accessibility standards want to be satisfied, it has become commonplace to try to create AA-conformant websites, as the W3C states that "it is not recommended that Level AAA conformance be required as a general policy because it is not possible to satisfy all Level AAA success criteria for some content" (Caldwell et al. 2008). This statement could be illustrated with the SC associated to Guideline 2.1. *Keyboard Accessible: Make all functionality available from a keyboard.* Most actions carried out by a mouse can also be done from the keyboard (e.g. clicking, selecting, moving, sizing). However, there is a small class of input that is path-dependant and can only be performed with a pointing device. Examples given in Cooper et al. (2015b) for this type of input include free hand drawing, watercolour painting or flying a helicopter through an obstacle course. Success criterion 2.1.3 *Keyboard (No Exception)* (Level AAA) requires that all page functionality is available using the keyboard, which can be seen as extremely restrictive if website commissioners or web developers want to feature the aforementioned functions in their website. In these scenarios, conformance could be claimed with regard to SC 2.1.1 *Keyboard* (Level A), which suggests the same recommendation as SC 2.1.3 but considers path-dependent functions as an exception.

WCAG 2.0 guidelines and SC have been created to be technology-independent, that is, they do not include details about any technologies in particular,[15] so that they can be applicable to both current and future technologies (Lawton Henry 2009). This flexibility is one of the improvements claimed by the W3C (ibid) with respect to WCAG 1.0 (Chisholm et al. 1999). Since the WCAG 2.0 are 'technologically agnostic' (Connor 2012, 20), specific guidance on how one can satisfy each success criterion is to be found in the WCAG 2.0 supporting materials. These include a wide variety of examples, recommended techniques, common failures that show which practices to avoid, and explanations about how to interpret the different layers of guidance (Figure 1.4 in Chapter 1 summarises the content of this documentation).

Although, as Harper and Yesilada (2008b, 76) acknowledge, guidelines are useful, they are only one part of the overall process of supporting web accessibility. For accessibility features to be effectively implemented, it is not sufficient for content producers to understand the WCAG 2.0 and consult the related resources provided by the W3C. Authoring tools need to support the techniques recommended or those that web creators want to implement, as do user agents —including browsers and assistive technologies used by people with disabilities. Chisholm and Henry (2005) believe that the interdependencies between these different components have historically fed a "chicken and egg" problem of who goes first in implementing an accessibility feature, hence the lack of a starting point in the implementation cycle shown in Figure 2.1.

Lawton Henry (2005a) declares that if one component in the WA implementation process has poor accessibility support, sometimes other components can compensate through 'work-arounds' that require much more effort and are not beneficial for accessibility overall. For example, let us assume that content producers are aware of the necessary measures that need to be taken to satisfy SC 1.3.4 *Abbreviations* (Level AAA) —e.g. they know they can use the element `<abbr>` (Cooper et al. 2015a, sec. HTML and XHTML Techniques, H67) as a mechanism for introducing the expanded form or meaning of an abbreviation present in the main body of a web page. If the authoring tools they are using do not offer the necessary support, they could code this mark-up 'by hand' using an advanced text editor. Similarly,

[15] In the example we have just presented, SC related to Guideline 2.1 do not prescribe how to technically make content operable through a keyboard interface, but rather what functionality is needed for users, i.e. to be able to successfully use a website by only using keyboard input.

the opposite can also occur. The technologies which the different human components of the WA implementation cycle rely on may support the use of `<abbr>`, but it can be the case that the content producer does not completely understand its functionality and therefore decides not to implement it or does it wrongly.

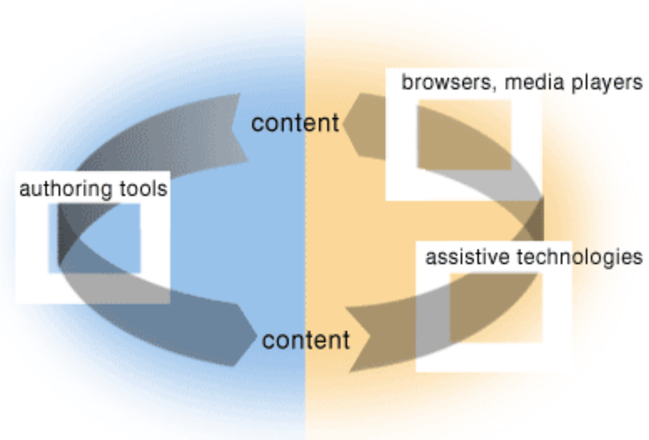


Figure 2.1. Implementation cycle of accessibility features (Lawton Henry 2005a)

In order to prevent such situations, it is necessary to develop and install appropriate procedures to ensure the accessibility of the web content produced on an on-going basis, for example, by means of peer-review or publication workflow processes (Abou-Zahra 2008, 84). In this sense, it has been claimed that it is particularly effective when content producers can address the relevant accessibility requirements themselves, rather than compensate with a comprehensive quality assurance step at a later stage (ibid). In the following section, with a view to understanding whether such on-going evaluation really occurs, we review previously published work about the attitudes of web professionals towards WA implementation.

2.2.3.2 Accountability for web accessibility

Of all the essential accessibility components that we have mentioned thus far (both in the Introduction and in the previous sections of the present chapter), we consider web content producers to be the main gateway to a more inclusive Web for all. On one hand, those who have the necessary knowledge and know-how to create websites that are compliant with accessibility standards have the power not only to apply recommended techniques, but also to promote best practices among their peers, highlight flaws in user agents to tool developers, and educate clients and upper management teams on accessibility benefits. On the other hand, content producers who are not aware of accessibility issues have always had the possibility, given their human nature, of consulting training materials and learning about the way in which people with disabilities access and use the Web. Nevertheless, a number of surveys have revealed rather disappointing results when it comes to the activities and perceptions related to web accessibility by these professionals, especially concerning the coverage of the needs of visually-impaired individuals.[16]

Lazar et al. (2004) reported on the data collected from 175 webmasters through an international survey aimed at understanding their knowledge on accessibility issues and the reasons for their actions related to WA. Thirty-six per cent (36%, N=63) of the respondents indicated that they were not familiar with W3C accessibility standards, while 47 webmasters (30%) acknowledged that they had never created accessible websites. In the same vein, when asked whether the websites they were working on by that time were accessible for blind users, 38 respondents (22%) said that they were not and another 38 (22%) stated that they were not sure. Paradoxically, 166 out of the 175 webmasters (around 95%) replied yes to the question ‘Do you consider ethics in planning and/or updating your websites?’, despite the fact that only 103 (59%) indicated that their organisations were planning to have accessible websites in the future.

The Disability Rights Commission (DRC) in the UK conducted interviews with 25 website developers and found that only 9% claimed any sort of accessibility expertise (DRC 2004). In parallel, they also launched a questionnaire targeting website commissioners and web development agencies to investigate their attitudes towards WA. Although the questionnaire was sent to over 1,000 potential participants, the response rate was very low (approximately 7.5%), which denotes, according to the researchers, a relatively low level of interest in accessibility issues (ibid). The final data gathered came from respondents belonging both to large organisations with more than 250 employees and small companies with less than 250 employees. The difference between both groups was surprising: most respondents from the former group (97%) claimed to be aware of accessibility as an important issue, and 68% asserted that they took accessibility into account when developing a site. Nevertheless, only 29% respondents from small business said they actually built websites with accessibility in mind. It is also worth highlighting that although 58% of website design agencies claimed to discuss accessibility issues with

[16] As a complementary approach, researchers in the field have suggested that people with disabilities should not only be regarded as Web consumers but also as accessible content producers, especially in the context of the Web 2.0 paradigm. For instance, Bigham (2009) explored the potential of blind users to both collaboratively improve WA for themselves and partner with Web developers to improve their content. Among other solutions, he proposed Accessmonkey, a tool for blind users to create and inject scripts across many platforms to improve WA (Bigham and Ladner 2007). Although this is an interesting approach, within the framework of this thesis, visually-impaired people are primarily considered as end users.

their clients, only 31% of the clients showed a positive attitude towards it (DRC 2004, 37).

Another national survey, this time carried out in Brazil, returned similar results in terms of accessibility awareness (Freire et al. 2008). It covered 613 subjects from academy, industry and government; only 19.9% of whom declared to consider accessibility in their web projects, mainly for social responsibility-based personal reasons —the existence of organisation and customer requirements were the lowest rated motivations (17.21% and 24.59% respectively). We find it relevant to add that, from all the participants, a large number (45.19%) stated that they knew technologies to make the Web accessible by the blind, but they did not know how to create web pages accessible for these users (ibid). In the same line, around 40% of the respondents reported to have no WCAG-related skills and said that they had never heard of the Brazilian accessibility law (ibid).

Petrie et al. (2011) conducted interviews with 47 professionals from three key groups of stakeholders in the value chain for WA: website commissioners (N=26), web developers (N=7) and web accessibility experts (N=14). Here, we are interested in the results concerning the first group, as the findings reported from the other two deal with WA assessment issues and will be discussed later in section 2.3.3. From the 26 web commissioners surveyed, only 11.5% spontaneously mentioned disabled and older people as potential audiences for their websites, which means, according to the authors, that accessibility is often omitted from their agenda. Actually, only 7.7% of the participants from this group made reference to the WCAG 2.0 when talking about their websites' accessibility conformance level. This contrasts with the data collected when asked whether their organisations had a policy on web accessibility, since 61.5% of the respondents said they did.

Finally, we deem it relevant to summarise the results from the survey carried out by Putnam et al. (2012), who retrieved data about how 185 user experience (UX) and human-computer interaction (HCI) professionals considered accessibility. They found that 11% (N=21) of the respondents attributed no importance to accessibility issues in their work. Furthermore, 19% (N=35) of the web professionals surveyed replied the same but expressing regret or being apologetic. While, after coding the open-ended responses, the researchers concluded that most participants (70%, N=129) had shown some sort of positive attitude towards WA, only 16% of them really reported a high level of accessibility consideration.

All these studies have contributed to the broader discussion about who is responsible for creating and advocating for an accessible Web. Two major conclusions could be derived from the data collected throughout these questionnaires and interviews. On one hand, their findings suggest that there is still a widespread lack of awareness about WA-related issues among website commissioners, clients and content producers, both in terms of (i) the legislation that exists about the matter and (ii) the knowledge needed to implement accessibility features when designing and developing web content (e.g. knowing the recommendations covered in the WCAG 2.0). On the other hand, these studies confirm what we have observed earlier when reviewing the W3C accessibility standards: that up to the present, scarce attention has been paid to accessibility considerations in the context of the multilingual Web. In these surveys, the focus was mostly placed on the early stages of web content development or its later maintenance (by webmasters), but no reference was made to the implementation of accessibility features in a potential localisation phase, nor were localisers ever considered as stakeholders of the web production cycle. Inspired by this prior work, we conducted a survey to address these and other issues related to multilingual WA assessment, the results of which will be reported in section 2.5 of the present chapter.

2.2.3.3 Challenges in web accessibility implementation

In spite of the efforts made by the W3C to publish a set of accessibility guidelines along with supporting materials to understand and implement them, and despite the measures taken at a national and international level to pass regulations that demand organisations from both public and private sectors to render web content accessible, the goal of achieving an inclusive Web for all is far from being met. Multiple studies have demonstrated that accessibility conformance levels are still very low. The DRC (2004) checked 1,000 websites for accessibility and found that 81% of them failed to satisfy even the most basic accessibility success criteria. Lopes et al. (2010) declared that, after verifying 30 million web pages as part of a large scale WA study, the number of good quality pages was minimal. In the same vein, a ten-year longitudinal study comprising approximately 6,000 home pages revealed that in 90% of the cases accessibility guidelines were mostly ignored (Harper and Chen 2012). Findings from the latest known large scale study covering a 14-year period from 1999 to 2012 over 100 top-traffic and government websites (Hanson and Richards 2013) are consistent with previous investigations, as the web documents analysed exhibited generally low conformance with WA indicators.

The 'Web Accessibility Integration Model' (WAIM) of Lazar et al. (2004), which highlights the multiple points within the web life cycle where accessibility can be incorporated or forgotten, will serve us to explain why web accessibility is not yet a reality. This model considers there to be three interrelated categories of influences on the accessibility (or inaccessibility) of a website: (i) societal foundations, (ii) stakeholder perceptions and (iii) web development practices. The authors (ibid) claim that societal factors such as education, training, government policy and accessibility statistics in the news can all help form the perceptions of commissioners, content producers and clients. They argue that these stakeholders influence the actual web development process (both the initial design and subsequent re-design), which in turn is affected by guidelines and tools. As the authors (ibid, 272) put it, “guidelines and tools help not only content producers with guidance, but also these guidelines and tools help provide the current ‘working definition’ for web accessibility.”

The major roadblocks to accessibility cited in the studies reviewed and other related work are associated to the three aforementioned categories. The following brief discussion on this matter is summarised in Table 2.2. Freire et al. (2008) indicated that two of the main reasons reported by web professionals for not developing accessible websites were the poor spread of accessibility law and the lack of training on the matter. This was also the case in Lazar et al. (2004), where webmasters indicated as well other factors such as lack of time, and managerial and client support. In the same line, some respondents from Putnam et al. (2012) stated that, while aware of accessibility concerns, they had little or no control over how accessibility was considered at a higher level in their workplace and discussed sacrificing accessibility for time, budget and client/company needs. Flaws in guidelines and accessibility evaluation tools were also highlighted as obstacles during WA implementation. Since the latter will be explored in more detail in the next section (2.3) and in Chapter 3 with regard to image accessibility (section 3.4), we would like to focus here on the former.

Harper and Yesilada (2008b) argue that the fact that WCAG 2.0 cover more than 200 pages and are complemented with an additional 200-page ‘how-to’ annex renders their implementation reasonably complicated: a downside that had been previously been highlighted with regard to the WCAG 1.0 (Sloan et al. 2006). This might explain why practitioners condemn these guidelines as impractical (Lazar et al. 2004) and claim that they would prefer a quantitative bullet-point list of accessibility problems in ‘clear language’ (Petrie et al. 2011). In addition, Harper and Yesilada believe that some user groups such as people with cognitive, language and learning disabilities are not fully addressed by the WCAG 2.0.[17] This goes in agreement with the opinion of authors like Fernandes et al. (2014) who consider that the W3C guidelines are primarily dedicated to address the needs of visually-impaired people, leaving other disabled communities unattended. Similarly, it corresponds to the results reported in Power et al. (2012), who observed that only 50% of the problems encountered by 32 users during a task-based web accessibility evaluation study were covered by the success criteria of the WCAG 2.0. To the constraints already highlighted in the literature, we may add that the diverse nature of these guidelines and SC requires a wide variety of skills for their implementation to be effective. Certain aspects of the WCAG 2.0 deal with language-related web components which are not directly linked to the instrumental and technical knowledge typical of web developers, designers or

[17] The WCAG 2.0 editors themselves acknowledge that the guidelines are not able to address the needs of people with all types, degrees, and combinations of disability (Caldwell et al. 2008).

webmasters. For instance, it is likely that these professionals, who have a rather technical-oriented profile, will find recommendations related to the provision of textual alternatives to visual content or other SC under Principle 3. *Understandable* difficult to implement, unless they can rely on the help of other actors in the web development chain, such as content editors.

Table 2.2. Challenges in WA implementation

WAIM categories	Roadblocks to web accessibility	Source
Societal foundations	Lack of training	*Lazar et al. 2004, Freire et al. 2008, Petrie et al. 2011*
	Poor spread of accessibility law	*Freire et al. 2008*
Stakeholders perceptions	Lack of client support	*Lazar et al. 2004, DRC 2004, Putnam et al. 2012*
	Lack of managerial support	*Lazar et al. 2004, Petrie et al. 2011, Putnam et al. 2012*
	Lack of knowledge by content producers	*Lazar et al. 2004, Freire et al. 2008*
Web development process	Limitations in budget	*Putnam et al. 2012*
	Lack of time	*Lazar et al. 2004, Freire et al. 2008, Putnam et al. 2012*
	Confusing/impractical accessibility guidelines	*Lazar et al. 2004, Petrie et al. 2011, Harper and Yesilada 2008b*
	Incomplete accessibility guidelines	*Harper and Yesilada 2008b*
	Limitations in technology	*Lazar et al. 2004, Petrie et al. 2011*

We hypothesise that, while all these challenges can —and probably do— equally arise during the production of multilingual websites, the participation of other more multifaceted stakeholders, such as web localisers, could prove beneficial to overcome some of the accessibility barriers mentioned above (especially if they have an appropriate WA background). As Lazar et al. (2004, 280) point out, accessibility is not an individual effort and people at all levels, through the development and maintenance of a website, should be involved. Before further considering the role of localisers in the achievement of a more accessible multilingual Web in sections 2.4 and 2.5, we deem it important to examine the nature of the measures that currently exist to guarantee that accessibility is attained in the first place, as well as if they have been effective thus far. This will allow us to better understand the relevance of localisers joining the call for a more accessible Web for all.

2.3 Assessing web accessibility

As we suggested earlier, accessibility considerations should not be left for the end of web projects, as is often the case (Thatcher et al. 2006, 20), but rather taken into account from the early stages of the web development cycle. However, this does not mean that web accessibility evaluation (WAE) −understood as an assessment of how well web content can be used by people with disabilities (Abou-Zahra 2008, 79)− should be limited to the web production stages. Accessibility testing is essential for assessing and monitoring websites throughout their entire lifetime and, in many cases, this on-going evaluation is more important that the initial ones (ibid). Vigo (2009, 24) refers to these two different scenarios as *proactive evaluation* and *reactive evaluation*. While the former is embedded in an accessibility-aware iterative design, the latter refers to the accessibility evaluation task carried out after the web page or site has been released. Proactive evaluations are *formative*, i.e. they support an iterative accessible development process by helping content producers identify lists of problems to fix. Reactive evaluations, on the other hand, are *summative*, i.e. they are conducted to assess the accessibility level of an interface, to validate it or to compare one interface against different versions or different systems (Brajnik 2008b). In the context of this thesis, our interest lies in *reactive evaluation* processes, as can be inferred from the goals stated in Chapter 1 (see section 1.3).

One of the measures taken by the W3C to facilitate the assessment of websites for accessibility once the content is ready was to write all WCAG 2.0 SC as testable statements (Caldwell et al. 2008). According to the W3C, testability can be either machine-based or human-based. More specifically, Cooper (2005) establishes that a success criterion is 'machine testable' if there is a known algorithm that will determine, with complete reliability, whether the technique used to satisfy it has been implemented or not. In turn, SC are considered as 'reliably human testable' when the technique(s) chosen to achieve them can be tested by human inspection and it is believed that at least 80% of knowledgeable human evaluators would agree on the conclusion. Nevertheless, no information is provided in the WCAG 2.0 document (or any of the supporting materials) with regard to the likelihood that an 80% agreement level can be achieved, neither do they indicate which success criterion falls into which category (Brajnik et al. 2012). As we will discuss later, this is in fact one of the challenges of WAE (see section 2.3.3).

While the SC offer instructions for evaluating web content feature by feature, a broader approach is commonly needed to plan and execute a WAE. The Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0[18] proposes a

[18] WCAG-EM 1.0 is one of the documents belonging to the "Evaluating Websites for Accessibility" resource suite of the WAI: https://www.w3.org/WAI/eval/ Last access: 1st March 2016. Details about other evaluation methodologies defined prior to WCAG-EM 1.0 can be found Abou-Zahra (2008, 91).

more general procedure to evaluate how well websites conform to the WCAG 2.0 (Velleman and Abou-Zahra 2014). To this end, five different steps (which are not necessarily sequential) are described: (i) *1. Define the evaluation scope*, determining which versions of the website will be assessed (mobile, one or all language versions) and selecting a conformance level (A, AA or AAA); (ii) *2. Explore the website to be evaluated* in order to identify its common Web pages, its essential functionality and the web technologies relied upon to provide the websites (HTML, CSS, PDF); (iii) *3. Select a representative sample*, taking into account different factors such as the size of the website or the variety of functionality and technologies used; (iv) *4. Audit the selected sample* according to the conformance requirements previously established; and v) *5. Report the findings.* In the following sections, the focus will be placed on step 4, as it represents the central stage of every evaluation procedure and has received considerable attention in the WA-related literature.[19]

2.3.1 Accessibility evaluation methods

When auditing a website, a reduced sample of web pages of the same website or specific content within a given page, the evaluator ―be it a person, a team of people or any other entity responsible for carrying out the evaluation (Velleman and Abou-Zahra 2014)― may choose one or several accessibility evaluation methods (AEMs). An AEM is a procedure aimed at finding accessibility problems, such as guideline violations, and possibly assessing levels of accessibility, that prescribes which criteria should be used under which conditions, so that the said accessibility problems can be detected (Brajnik 2008b, 69). AEMs can be categorised according to three basic types of testing techniques: (i) “user testing”, which is carried out by end users ―generally with some type of disability― in informal or formal settings; (ii) “manual testing”, that is conducted by human evaluators, who could be experts or novices; and (iii) “automated testing”, which is carried out by means of software tools (Abou-Zahra 2008, 84). In what follows, we will review the most popular AEMs within each category and highlight some of their major advantages and disadvantages. In doing so, we will take as the main reference the taxonomy proposed by Giorgio Brajnik in prior related work (Brajnik 2008a; Brajnik 2008b; Brajnik et al. 2011; Brajnik et al. 2012).

2.3.1.1 User testing

User testing is the generic term used to describe a process where formal or informal experiments are set up with real users, who are individually asked to

[19] Step 3, which concerns page sampling in WAE, has also been extensively studied in prior work (Brajnik et al. 2007; Velleman and van der Geest 2013; Harper et al. 2015). However, the scope of this thesis is limited to only one dimension of the accessibility evaluation procedure, which is the actual testing strategies followed, as will be argued in section 2.3.1.

perform goal-free or goal-oriented navigation on a website (Brajnik et al. 2012). It is often considered as a technique that complements other ones, such as manual and/or automatic testing (Abou-Zahra 2008). Common user evaluation paradigms include, among others, performance measures, such as taking into account the time required by the user to complete a task or the number of incorrect link choices, to name just a few; logging user actions; questionnaires; observations or think-aloud (Jay et al. 2008). While extremely useful, one of the major issues with user testing is the difficulty to filter out personal bias and preferences, and identify the actual issues; for instance, different accessibility barriers may arise during a user testing session depending on the users' expertise in the use of a computer or of assistive technology (Abou-Zahra 2008, 87). In order to prevent this from happening, formal checks, carried out by professionals who follow sound procedures, are often preferred over informal checks, such as asking individual persons like friends or colleagues for their opinions (ibid).

What Brajnik (2008b, 74–76) calls s*ubjective assessments* could be regarded as a form of informal checks. Subjective assessment is a process where an evaluator hires a panel of users who are asked to explore/use a website in full autonomy and send back their opinions; the evaluator then collects such feedback to determine the level of accessibility achieved. Contrary to this assessment modality, which is more appropriate in the case of formative evaluations, *formal user testing* settings require the participation of experienced evaluators who directly observe users while performing given tasks and who, based on the data collected, can generate a list of problems and assign severity levels. Similarly, during this type of testing, users are asked to use applications and assistive technologies they are familiar with, and they are screened according to their level of experience in using these tools (ibid).

Remote user testing is considered by Brajnik et al. (2011) to be a useful variant that can ease the typical logistical difficulties of laboratory and field user testing. Petrie et al. (2006) investigated this type of method —where users work asynchronously at their home or workplace on given tasks without the physical presence of the evaluator— by comparing it against local user evaluations. Based on the results obtained, they concluded that remote user testing would not be appropriate for formative evaluations but it could be suitable in the case of summative evaluations, where the purpose of the assessment is simply to know whether participants can perform certain actions or to see how they rate the accessibility level of the web content presented to them according to a predefined set of parameters. For these reasons, remote user testing was the AEM chosen for the second stage of the experimental study described in Chapter 5 (see section 5.3.4).

2.3.1.2 Manual testing

Manual testing is the process by which the primary responsibility for making the final decisions concerning the results of an accessibility assessment is held by human evaluators (Abou-Zahra 2008, 86). Manual tests may cover a broad range of accessibility provisions, imply varying degrees of software tool support, and demand different requirements with regard to the skills and knowledge of the human evaluators. Therefore, Abou-Zahra (ibid) suggests differentiating between the following types of manual testing: (i) non-technical checks, that can be carried out by non-technical evaluators such as content authors —the author gives, as an example, the task of determining whether a text alternative of an image describes it appropriately; (ii) technical checks, conducted by evaluators with technical skills to address, among others, markup code and document structure issues; and (iii) expert checks, carried out by evaluators with knowledge of how people with disabilities use the Web and who can identify issues that relate to user interaction.

Screening techniques are precisely one form of expert checks. They are based on using a website in a way that some sensory, motor, or cognitive capabilities of the evaluator are artificially reduced (Brajnik et al. 2012, 8:3). After carefully selecting the screening conditions so that they match the characteristics of the target population (e.g. using a website through a screen reader with the monitor turned off to simulate how blind people browse the Web), the evaluator explores the website and tries to accomplish selected goals (Brajnik 2008b, 74). However, this method is not recommended for summative evaluations as it is not systematic. Furthermore, it has been claimed that it normally shows low effectiveness since it depends heavily on the experience level of the evaluator in using the assistive technology, which rarely would match the experience of users (ibid).

In this sense, *inspection methods* are largely more popular, given that they can involve all types of checks (technical, non-technical and expert). In general terms, they consist of an evaluator inspecting a website, web page or content element for its accessibility (Brajnik et al. 2011, 249). The *barrier walkthrough* method is an analytical inspection technique which relies on expert checks and is based on a heuristic walkthrough, where the context of website usage is explicitly considered (Brajnik 2008a, 114). When this method is applied, an evaluator has to assess a number of predefined barriers that are linked to user characteristics, user activities, and situation patterns so that conclusions about user effectiveness, productivity and satisfaction can be drawn, and appropriate severity scores can be consequently derived (ibid). Still, as in the case of screening techniques, this method is extremely dependent on evaluators' experience and their interpretation of the different accessibility guidelines and principles.

Although studies have shown that the barrier walkthrough is highly effective in reducing false positives because it forces evaluators to focus on constrained usage scenarios, it has been demonstrated that other inspection methods, such as *conformance review*, can help find more accessibility problems (Brajnik 2008a). During conformance review, the evaluator uses a set of accessibility guidelines (e.g. the WCAG 2.0) to decide if a page or website complies with the requirements specified therein (Brajnik et al. 2011, 249). Since this method does not prescribe how to choose scenarios nor how to rate the problems found, it is often implemented in combination with a given WA metric, whose validity, reliability, sensitivity and complexity tend to have a direct impact on the results of the assessment.[20] Similarly, as in every method that relies on a human evaluator, experience in accessibility issues can also affect the final evaluation verdict, as we shall further comment on in section 2.3.3. All in all, despite these and other disadvantages highlighted in the literature such as the impracticality of the technique when a large number of pages needs to be assessed (Brajnik 2008b, 73), conformance review is not only the most widely used inspection method in summative evaluations, but also one of the most popular AEMs together with automatic testing (Brajnik 2008b, 71; Brajnik et al. 2012).

2.3.1.3 Automated testing

Abou-Zahra (2008, 85) states that "automated testing is carried out without the need for human intervention." However, our understanding of this technique is more in agreement with the description provided by Brajnik et al. (2011, 249), who indicates that *automated testing* does actually involve the participation of an evaluator and implies using an automated accessibility tool to check conformance of a website, a web page or an element within a web page against the accessibility tests encoded in that tool. This technique is generally well accepted among evaluators because the use of web accessibility evaluation (WAE) tools presents numerous advantages with respect to other non-automated approaches: it enables an objective assessment of web content, it offers a broad site coverage (i.e. WAE tools can check multiple web pages simultaneously) thus enhancing productivity by significantly reducing time effort, and it is normally cost effective (Ivory 2003, 124).

Harper and Yesilada (2008b, 65–66) consider that these types of tools are important for web accessibility as they provide a medium for content producers to validate their pages against published guidelines without actually reading and manually applying them. Indeed, this is the case when WAE software is only used for determining the conformance of websites to accessibility checks which can be fully executed automatically, such as analysing the syntactic structure of web

[20] Vigo and Brajnik (2011) offer a comprehensive review of Web accessibility metrics. Reference to some of the metrics they consider in their paper will be made in Chapter 5 (section 5.3.1.2) when discussing the operationalisation of one of the dependent variables of this thesis' main experimental study.

content. For example, WAE tools can look for the presence of language attributes in the `<html>` element to decide whether a web page satisfies WCAG 2.0 SC 3.1.1 *Language of page*: *The default human language of each web page can be programmatically determined* under guideline 3.1. *Make text content readable and understandable*. Nevertheless, this kind of markup syntax checks only the addresses of a minor subset of accessibility recommendations and rarely covers other accessibility provisions which deal, for instance, with qualitative aspects related to the usability of the user interface or the natural language used (Abou-Zahra 2008, 85).

The above does not necessarily mean that WAE tools leave these other issues completely unattended. As Abou-Zahra (2005) indicates, WAE software can also assist evaluators in performing accessibility checks which need to be assessed manually. They do so by generating specific warnings that either (i) recommend their users look at particular accessibility guidelines or (ii) ask them to confirm potential accessibility problems. For example, aDesigner, the WAE tool chosen for the experimental study of this thesis,[21] includes the following warning in every check report: "Make the text easy to read". This message points to WCAG 2.0 SC 3.1.5 (Level AAA), which establishes that text in a web page should not require a reading ability more advanced than the lower secondary education level, but the tool does not refer to any text excerpt in particular. On the other hand, when an anchor element (`<a>`) contains the attribute `target="_blank"`, aDesigner identifies it, highlights it to evaluators and then asks "Is the user made aware that there will be pop-up windows or changes in the active window?", i.e. the tool requires the confirmation of the user to know whether SC 3.2.2 (Level A) is respected. This success criterion prescribes that users need to be advised on any change of context before they interact with a given user interface component (e.g. a link).

Over the last few years, the WAI has made an effort to gather together a list[22] of WAE tools including validation, repair and transformation software (Eggert and Abou-Zahra 2014b). While *validation and repair tools* analyse pages against accessibility guidelines and return a report (which can be complemented or not with in-page feedback), *transformation tools* modify the appearance of websites to help identify conceptual design issues with regard to web accessibility (for instance,

[21] More detailed information about this tool is provided in Chapter 3 (section 3.4.1) and Chapter 5 (section 5.2.3.3).

[22] At the time of writing, the list contained 70 different tools that could be filtered by language, type of license, tool modality (plugin, online, standalone) and set of guidelines addressed. However, it should not be regarded as a definitive list, as some tools featured there are no longer traceable or maintained by their owners: https://www.w3.org/WAI/ER/tools/ Last access: 5th March 2016.

by presenting web content without colour).[23] Examples of these types of tools, as well as other information concerning their main features, will be provided in Chapter 3 (section 3.4.1) within the particular context of automated image accessibility assessment.

Given the benefits of WAE tools alluded to thus far, it goes without saying that automated approaches can considerably reduce the effort required to carry out evaluations. Nonetheless, practitioners and WA researchers have generally subscribed to the belief that using only one automated tool alone or solely relying on automated testing is not by itself a viable solution to web accessibility evaluation (Abou-Zahra 2008; Brajnik 2008b). Thatcher et al. (2006, 40) have wisely considered spell checkers as an analogy for WAE tools: for them to be useful and effective, a human with the appropriate skills and knowledge must always determine the validity of the results they yield. The effectiveness of this precise combination of knowledge and tools will be investigated in depth in Chapters 6 and 7 when discussing the results of our experimental study.

2.3.2 Previous studies on web accessibility assessment

In broad terms, investigations in the area of accessibility evaluation have been conducted with a view to either (i) drawing conclusions about the current state of the Web with regard to accessibility (departing from both small and large scale studies on a restricted sample of web pages) or (ii) exploring the quality of the AEMs presented in the previous section. Since reference to the studies addressing the former is being gradually made throughout this thesis, we will focus now on the latter. More specifically, within the particular framework of our research, we are primarily interested in the studies concerning *conformance review* and *automated testing* for three main reasons: first, they are extensively used by practitioners; second, they are the two techniques that have received more attention by scholars in the field; and third, the WA knowledge required to carried out conformance reviews and the tools needed to perform an accessibility evaluation based on automated testing are two variables that, as we have just announced, will be later examined within the context of this thesis' experimental study (see Chapter 5). Yet, reference to other AEMs such as user testing or the barrier

[23] Page transformation tools should not be confounded with transcoding tools. These represent a category of technologies that transform inaccessible Web content into accessible content on the fly (Asakawa and Takagi 2008). Multiple studies have been devoted to investigate transcoding processes (Takagi and Asakawa 2000; Akpınar and Yesilada 2015; Yesilada et al. 2007) but discussing their results falls beyond the scope of this thesis.

walkthrough will be made when appropriate, especially if these approaches were studied in conjunction with the former.[24]

This is precisely the case of the study carried out by Brajnik (2008a), who sought to compare conformance review and barrier walkthrough methods to assess their effectiveness. The author describes this quality as the "extent to which the method can be used to yield desired results with appropriate levels of accuracy and completeness." Furthermore, he notes that effectiveness relies on considerations of *validity* (in terms of correctness or precision —the percentage of reported problems that are true problems— and sensitivity or recall —the percentage of the true problems that are reported) and *reliability*, i.e. the extent to which independent evaluations produce the same results (ibid). The experiment, which involved the participation of 12 novice accessibility evaluators, showed that while the barrier walkthrough method led to a higher level of precision, conformance review returned better results with regard to sensitivity and reliability. The relatively worse outcome of the barrier walkthrough was attributed by the author to the low level of WA expertise of the participants, who were actually students. This assumption was confirmed in a later walkthrough study with 19 expert and 51 non-expert evaluators by Yesilada et al. (2009), who discovered that both effectiveness and reliability of the former group were significantly higher than the latter. Brajnik et al. (2011) provided additional data about the same study and indicated that they had also observed a negative monotonic relationship between the number of judges and reproducibility or reliability of the evaluation results: with one non-expert, only about 50% of the true problems are caught, three experts reveal about 80% of the true barriers, and to cover all of them 14 non-experts are needed (ibid).

The evaluators' level of expertise[25] represents, in fact, one of the key aspects studied in prior investigations that concentrated solely on conformance review. Brajnik et al. (2010) compared the performance of 22 experts and 27 non-experts when assessing four web pages against the W3C accessibility standards. The researchers wanted to particularly ascertain the testability of the 61 WCAG 2.0 success criteria, achieved when 80% of the evaluators agree on whether the they have been met or not, as we explained earlier in this chapter. The experiment showed that 50% of the SC failed to meet the 80% threshold, with experts missing 32% of the true problems and non-experts 49%. After adding the data of two

[24] In addition, it should be noted that this section does not intend to examine all the relevant prior work done on the matter, but to review some of the most substantial studies to then better understand the major remaining challenges in WAE.

[25] Expertise can be characterised in terms of (i) the practice in using a specific evaluation method, which could also involve knowledge of a set of guidelines, and (ii) the knowledge, practice, and skill in accessibility in general (i.e. the experience on assistive technologies, typical accessibility problems, typical user behaviors, or user preferences) and in the underlying Web technologies (Brajnik et al. 2011, 251).

additional experienced evaluators to get a more balanced number of audits, conclusions reached were almost the same: average agreement was at the 70–75% mark, while the error rate was around 29% (Brajnik et al. 2012). As far as the number of evaluators needed to capture all of the true problems is concerned, findings were akin to those of barrier walkthrough studies: one needs to hire four independent experienced evaluators or, equivalently, six novices (ibid). These results are in agreement with other conformance review experiments involving participants without a high level of WA expertise, where only a very low number of SC appeared to be reliably human testable (Brajnik 2009; Alonso et al. 2010).

Within the context of automated testing, the quality of the method relies on the effectiveness of the tools used. To the aforementioned characteristics into which this property is refined —correctness or precision and sensitivity or recall (sometimes also named as completeness in the case of evaluation software)— Brajnik (2004, 19) adds specificity, a notion that we will also refer to in subsequent chapters of this thesis and that is understood as "the level of detail that the tool is able to use when describing a potential problem." The first study aimed at investigating the effectiveness of WAE tools was conducted by Ivory and Chevalier (2002), who empirically examined three tools: WatchFire Bobby, the W3C HTML Validator and UsableNet LIFT.[26] The study consisted of several stages: a survey and an experiment with both developers and end-users. A total of 137 out of the 169 web professionals who replied to the survey reported using WAE tools, around 20% of whom said that they were mostly using them after finishing a website. Nine of these participants were later recruited for an experiment during which they were asked to improve the accessibility of a website, first manually and then with the help of the above-mentioned tools. The results of the experiment, assessed by 22 end users, revealed that although the tools helped designers to identify a large number of problems (which means that a higher level of completeness was registered in the tool condition), these professionals were not effective in interpreting and applying the guidelines. This observation was also made by Mankoff et al. (2005), who compared automated testing (based on the use of Bobby) not only to conformance review, but also to other AEMs such as screening techniques and user testing. Their results indicated that (i) no single evaluator nor tool could be counted on to find a high percentage of accessibility problems, and that (ii) the effectiveness of evaluators using a screen reader was comparable to that of those who had performed an accessibility conformance review.

The inability of tools to check every aspect of web accessibility has also been endorsed by more recent studies. Trewin et al. (2010) surveyed 49 IBM developers,

[26] Bobby was the most popular WAE tool until 2005, when it was acquired by IBM. It is believed that its functionalities are now included as part of IBM's Rational Policy Tester suite: https://goo.gl/q5Xfhp Last access: 4th March 2016. Further insight about the other two tools will be provided in Chapter 3, section 3.4.1.

69% of whom reported to use both a WAE tool that flags accessibility problems and an assistive technology during summative evaluations. Although respondents had access to accessibility-related information and training, WAE was referred to as the most difficult and most time consuming aspect of producing an accessible product. The vague explanations given by automated tools about the errors flagged and the high volume of false positives they return were among the reasons provided by developers to explain the low level of trust they placed on this type of software. As far as low correctness is concerned (i.e. high levels of false positives), Vigo et al. (2013) found that tools that show this pattern are often those that exhibit higher completeness (i.e. recall) scores. Yet, after measuring the effectiveness of six state-of-the-art tools with regard to WCAG 2.0 conformance, they concluded that completeness only ranged between 14% and 38% and that, despite the technological advances experienced since the time when the first study on WAE tools was published, SC coverage was still very narrow (at most, 50% of the WCAG 2.0 success criteria were considered by the tools).

As we have illustrated thus far in this chapter, the literature is populated with a large amount of studies about web accessibility assessment. There is no doubt that, up to the present, considerable research efforts have been dedicated to exploring not only the general accessibility level of the Web, but also how this can be assessed and which methods are more appropriate to do it. Nonetheless, far too little attention has been paid to accessibility issues in the multilingual Web. In contrast to the extensive research work that exists on accessibility evaluation of monolingual websites, there is much less information about which assessment procedures to follow when two or more language versions of the same website are available. In the same vein, nothing has been written about how the different AEMs discussed in this chapter can be applied in the context of multilingual WAE or whether one or other of these techniques are more suited to the particularities of localised websites. As we announced in Chapter 1 (section 1.2.1), only the W3C provides two broad indications on the matter: (i) the different language versions of a given website ‒if they have totally independent URLs‒ can be considered as individual websites for evaluation (Velleman and Abou-Zahra 2014), and (ii) a "Statement of Partial Conformance" can be included in a multilingual website when accessibility support does not exist for all language versions (Caldwell et al. 2008). The present thesis covers this gap in the literature by offering, to the best of our knowledge, the first research work that provides a theoretical insight and empirical evidence ‒based on a survey and an experimental study‒ about the topic of multilingual WAE.

2.3.3 Challenges in web accessibility evaluation

Irrespective of the number of language versions that a website may have, assessing its accessibility is not a trivial task. At a general level, it could be stated that existing barriers to a successful web accessibility evaluation are similar in nature to the ones encountered during accessibility implementation in the first place, such as lack of time and WA awareness or the limitations of the WCAG 2.0 highlighted so far (see section 2.2.3.3). At a more specific level, however, remaining challenges in WAE are also directly related to current accessibility evaluation methods and techniques. We would like to put emphasis on the two major threats to web accessibility assessment that we have identified in the literature:

1) The evaluator or expertise effect

As we have seen, studies suggest that the 80% agreement between evaluators expected by the W3C is rarely or never reached when WCAG 2.0 conformance is tested by human inspection. While, as Brajnik et al. (2012, 8:26) acknowledge, there could be a lot of different causes for the low reliability of the WCAG 2.0, including the complexity of the document that describes them or the supporting materials, the guidelines being technology agnostic, the variability of disabilities they try to address or the large and growing number of techniques to implement them, practitioners and scholars have primarily attributed the blame to what is now well-known as the “expertise effect” or the “evaluator effect”. Extensively studied in the field of web usability (Jacobsen et al. 1998; Hertzum and Jacobsen 2001; Hornbæk and Frøkjær 2008), this phenomenon refers to the differences in the accessibility problems identified by evaluators under similar conditions. These differences depend, in turn, on their level of expertise about WA ─in the case of conformance reviews or any other inspection method─ and/or their experience in the use of assistive technologies ─when the assessment relies on screening techniques, subjective assessment or formal user testing.

The evaluator effect can have a strong impact on the effectiveness of the AEM used as it influences the kind of problems that are detected, at which level of abstraction and how they are rated for severity (Brajnik 2008b). When comparing the performance of experts and non-experts during accessibility audits, studies have demonstrated that expert evaluators are more judgemental, need less time to complete the assessment task and rate themselves as more productive and confident than non-experts, who tend to underestimate the severity of accessibility barriers (Yesilada et al. 2009; Brajnik et al. 2011). Furthermore, it has been claimed that the variability among experts is smaller and that expertise improves the ability to avoid false positives (Brajnik et al. 2012). It seems therefore that not only do web professionals without the necessary WA background face challenges in testing for accessibility problems, but their opinions, which represent the main output of analytical AEMs, can also lead to inaccurate accessibility conclusions about the web content being audited. In this sense, we agree with the belief of many researchers in the field that training is crucial to reduce the expertise gap (Lazar et al. 2004; Abou-Zahra 2008; Brajnik et al. 2011).

Together with the level of WA knowledge and know-how, we argue that the *evaluator effect* might be also exacerbated by the overall set of profession-related skills of the person who performs the accessibility assessment. For instance, we hypothesise that the reliability of language-based SC will be higher if the evaluation panel consists of content editors with a linguistic background rather than web developers or designers alone. In fact, prior work has shown that guidelines and SC related to linguistic issues are often considered as the most difficult to assess by professionals with a technical-oriented profile (Brajnik et al. 2012, 8:19) or left out during the evaluation task (Hanson and Richards 2013). The interpretation of certain WCAG 2.0 success criteria bears some degree of subjectivity and it is likely that some of the concepts behind them will be harder to understand by one or another group of web professionals. In this sense, we embrace the proposal of Abou-Zahra (2008, 104), who asserts that the responsibility for evaluating web content should be ideally distributed in a review team formed of evaluators that have complementary roles and skills. This idea gains even more importance in the context of the multilingual Web and is key to the discussion about the accountability for accessibility within the web localisation process that will be further developed in section 2.4.3.

2) Limitations of web accessibility evaluation technology

Tools could be seen as an essential part of the solution to most of the challenges we outlined with regard to accessibility implementation and evaluation. As Brajnik et al. (2010) indicate, thanks to WAE tools, the possibility to perform accessibility assessments is now practically given to anybody. Web content producers with and without accessibility expertise look to this kind of software with the hope of automating as many checks as possible and finding an explanation to the errors flagged. In other words, they see automated testing not only as a technique for error detection, but also as a means to receive quick accessibility training. Nevertheless, in an ideal WAE scenario, evaluators are normally advised against solely relying on automated tools since, compared to other methods, they still perform poorly in terms of coverage and completeness (Vigo et al. 2013).

As far as WAE tools' effectiveness is concerned, we are currently at an inflection point where, despite the technological advances achieved, it is believed that more than half of the provisions in most accessibility standards are as yet disregarded in automated evaluation software, thus unavoidably requiring manual inspection. Additionally, usability issues have also been highlighted by content producers who use them. Among the complaints found in the literature, it is worth noting the following: (i) the length and detailed nature of the check reports (not in terms of error repair guidance, but with regard to the multiple references made to W3C technical documents), render them unhelpful and difficult to interpret, particularly for non-expert evaluators; (ii) example solutions are not always provided; and (iii) specificity is rather low, i.e. when errors are flagged or warnings are included, tools do not offer explanations as to why a specific code fragment or text content might cause problems and to whom (Mankoff et al. 2005; Abou-Zahra 2008; Trewin et al. 2010; Petrie et al. 2011).

Abou-Zahra (2008, 84) states that optimal accessibility results can be achieved by combining different approaches to benefit from each of their specific advantages. Vigo et al. (2013, sec. 5) echo this belief by applying it to the context of automated testing and suggest that one possible solution to improve the low effectiveness of WAE tools reported in prior work would be to use multiple software: "tools could be employed on those SC they show higher effectiveness levels in order to maximise coverage, completeness and correctness." From this perspective, we argue that general evaluation tools, which offer a more sound platform for syntactic and heuristic checks (Abou-Zahra 2008, 85), could be complemented with specialised automated tools from other domains of knowledge in an attempt to cover those SC which currently suffer from a lack of automated support. For example, software used in the area of natural language processing (NLP) could be implemented along with general WAE tools with a view to tackling the linguistic aspects of the WCAG 2.0. In this thesis, we put this idea into practice by proposing the use of a controlled language (CL) checker to reach higher levels of specificity when testing web content for WCAG 2.0 SC 1.1.1 *Non-text Content: All non-text content that is presented to the user has a text alternative that serves the equivalent purpose.* This somehow follows Lawton Henry and Abou-Zahra's vision (2014), as they recently suggested that NLP and WA were areas that could benefit from collaboration in research and development. Chapter 4 will be dedicated to present our CL proposal, which will be then evaluated in the experimental study introduced in Chapter 5.

2.4 Addressing web accessibility during web localisation

Thus far in this chapter, we have explored two crucial axes of the Web for All paradigm: the implementation of accessibility best practices, and the methods that scholars and practitioners normally adopt to check whether existing WA standards have been correctly followed. While doing so, we have observed that a monolingual website perspective has generally been adopted to examine WA-related issues. In other words, a distinction between monolingual and multilingual (or localised) websites has hardly ever been drawn when dealing with aspects associated with accessibility implementation and evaluation techniques, or when debating about accountability for accessibility throughout the Web life cycle.

Precisely motivated by the little discussion that there has been in the WA literature about the multilingualism of the Web and how the processes that make it happen may have an impact on the degree of accessibility achieved, we have deemed it relevant to investigate whether prior work in the area of web localisation has paid attention to accessibility concerns. This literature review, covered in section 2.4.2, along with the brief introduction about what localisation entails included in section 2.4.1 below, will serve us to better frame our approach to multilingual web accessibility, which is based on the belief that localisers should be held accountable for the proper functioning of the web product they deliver by assuring that the necessary accessibility-related measures have been put in place (see section 2.4.3).

2.4.1 General considerations about the localisation process

In Chapter 1 of this thesis (see section 1.1.3.1), we have already presented the sequence of processes which the multilingual Web rely on by touching upon the different steps of the GILT model. While in the next section (2.4.2) certain references will be made to internationalisation when discussing the shared interests between accessibility and the development of multilingual websites, emphasis will be henceforth placed on the last two components of the aforementioned acronym: localisation and translation. Similarly, it should be noted that our intent is not to provide a thorough account of how web localisation has been studied up to the present,[27] but to highlight the most important aspects of the process with a view to facilitating the understanding of subsequent discussions on the topic within the context of this thesis.

As we announced in the Introduction, localisation is quite often seen as an umbrella term that refers to “the processes whereby digital content and products developed in one locale are adapted for sale and use in one or more other locales” (Dunne 2015, 550). Nevertheless, significant efforts have been devoted in the field to strictly differentiate localisation from traditional translation. In this regard, Schäler (2008, 196) argues that the latter does not necessarily deal with digital material, whereas the former is always happening in the digital world. This vision is in agreement with Dunne's idea of localisation as “translation on the computer, for the computer” (2015, 558) and has important implications in different areas of

[27] The history and evolution of localisation is nicely summarised in Esselink (2006), Schäler (2008), Sin-wai (2013) and Dunne (2015). In the literature, it is also possible to find comprehensive discussions about localisation from an industry perspective (Yunker 2003) and from the point of view of Translation Studies (O’Hagan and Ashworth 2002; Pym 2004; Jiménez Crespo 2013). More pragmatical approaches to localisation are adopted in Esselink (2000) and Roturier (2015), to name just a few.

the process, which have been outlined in the literature as follows: localisation differs from general translation with respect to (i) how translatable elements are identified, (ii) their non-linearity, (iii) their multimodal nature, (iv) the tools needed to render them, (v) the way in which the translation process is prepared and coordinated and (vi) the extent of the changes that may be introduced in the target product (Schäler 2010; Pym 2011, 411; Jiménez Crespo 2013, 19).

Several of these localisation hallmarks are of special interest for our research work. First, it is important to mention that the material localisation professionals work with —regardless of the digital product they are manipulating, be it a web document or a piece of software—comes as text, graphics, audio or video and can be stored in a large variety of file formats (Schäler 2010, 210). Localising this material involves not only identifying the translatable strings therein but also being sensitive to and addressing all semiotic and culture-related elements that the overall product may convey, such as colours, icons, flags, currencies, date formats and so on (Esselink 2000; Yunker 2003; Fernández Costales 2009). In addition, given the digital nature of the task, adapting all the above elements to the language and culture of the target audience necessarily involves some engineering before delivering the final localised product. In the case of websites, this would imply, for instance, not only the manipulation of source code (to adapt hyperlinks or file path locations of elements that might have been replaced), but also in some occasions the restructuration of the macrostructure to accommodate the new language versions, including the insertion of a language selection mechanism (e.g. a global language gateway) (Mata Pastor 2005). This technical aspect leads us to another key factor in localisation: the way in which the process is managed.

Quah (2006, 114–116) suggests that, in broad terms, the typical localisation process[28] involves three primary stages, namely (i) the project preparation —which he also refers to as the 'localisation-enablement' phase, giving as an example the extraction from graphics of the text which is to be translated), (ii) the translation proper and (iii) the quality assurance phase. In projects with such a configuration, it is generally assumed that the different localisation tasks will be distributed among several actors, leaving the first stage (and sometimes also the third) in the hands of the localisation engineer. According to Esselink (2003, 75), this professional —who does not need to be a programmer or a developer— is responsible for all the technical work that might be part of the localisation project, such as the web engineering steps mentioned above. Locating and identifying translatable elements and putting them back once the translation step is finished would be also

[28] In large localisation projects, these three phases are often supervised by a project manager, who is in charge of scheduling all activities, handling finances and communicating with the client, among other managerial tasks (Esselink 2003, 75). This dimension of the localisation process is covered in Mata Pastor (2005, 236–246), who provides a more fine-grained description of the different areas, steps and tasks of the Web localisation process.

part of their duties (Esselink 2002). While the figure of the localisation engineer might be common in large corporations, in other contexts it can be also the case that a single agent performs all the tasks: a web localisation process that Jiménez Crespo (2013, 29) names “individual localisation”. As we have indicated in Chapter 1 (see section 1.1.3.1), we will refer to this agent as the localiser, who could be defined as a translator that possesses strong managerial, instrumental and technical skills in addition to traditional translation and domain expertise (Folaron 2006, 213–216; Jiménez Crespo 2013, 165–179; Dunne 2015, 553; Torres del Rey and Morado Vázquez 2015).

Within the context of web localisation, the nature of the project (time, funding and resources available, requirements of the client/commissioner) might not only have an impact on the team involved in it, but also on the level of localisation ultimately achieved. One of the most well-known categorisations is the one proposed by Singh and Pereira (2005, 10–15), which is based on the role of cultural adaptations and is briefly summarised below:

- *Standardised websites*, in which the same web content is provided for both domestic and international users (one language for all countries/markets).[29]
- *Semi-localised websites*, in which the only web content adapted is the contact page.
- *Localised websites*, in which most content and pages are localised and contained within the same URL, but the original functionalities and back-end are not generally modified.
- *Highly localised websites,* in which all content, site structure and functionalities are fully adapted to the target locale (they usually offer country-specific URLs, for instance www.amazon.fr (France) and www.amazon.es (Spain).
- *Culturally customised websites*, which according to the authors is rare, as it implies a total immersion in the target locale at a cultural level, in terms of the more complex notions of perception, symbolism and behaviour.[30]

[29] These differ from international websites —also referred to in the literature as internationally accessible websites (McDonough 2006). An international website is adapted to the global market but, while it seems like a national website, it has been changed and extended in subtle ways that ideally make it usable in other parts of the word (for instance, in a commercial website, by allowing multiple checkout or registration procedures) (Schade and Nielsen 2013, 14).

[30] Although discussing culture-related aspects lies beyond the scope of this thesis, we deem it important to provide at least the basic definitions of these concepts. Extensively grounded on semiotic studies, the authors described perception as“ the process by which individuals select, filter, organize and interpret information to create a meaningful picture of world”, symbolism as “the system of representations and symbols”, and behaviour as “the cultural-embedded norms and values that make us react the way we do” (Singh and Pereira 2005, chap. 2).

This differentiation is of considerable importance for our investigation because it clarifies the extent of the references made in W3C accessibility-related documents to multilingual websites in the context of WAE, outlined in section 2.3.2. If we take into account Singh and Pereira's categorisation, only highly localised websites and culturally customised websites would actually enter under Velleman and Abou-Zahra's (2014) understanding of the notion of "website in multiple versions", thus overlooking the case of accessibility in localised websites, which are much more common:

> "Some websites are available in multiple versions that are independent of one another in use, that is, using one version does not require or depend on using another version of the website. For example, a website may have a mobile version and there may be versions of a website in different languages that meet this characteristic. Usually each such website version has a different set of URIs. Such website versions can be considered as individual websites for evaluation."

Within the framework of this thesis, when discussing the topic of localisation, we will mostly refer to localised websites as described above, i.e. websites that contain, under the same primary URL, two or more mirrored language versions, to which the user can have access from any page of the website through a language selector. With regard to this type of site, Jiménez Crespo (2013, 34–35) distinguishes, from a more technical perspective, between *lower localisation* levels and *higher localisation* levels. While the former only require the translation of the surface structure by means of replacing the textual strings in the website, the latter also involves adaptations and re-engineering in the underlying structure or deep structure, which contains the programming or tagging. As will be emphasised in the following sections, we argue that accessibility implementation needs to occur at both levels (front-end and back-end), if an operable, understandable, comprehensible and robust (i.e. fully functional) localised website is to be achieved, regardless of the degree of cultural adaptation envisaged.

2.4.2 Web localisation and accessibility: prior related work

Within the growing body of literature dedicated to the study of the wide range of factors influencing the web localisation process, we have observed an incipient interest in the end users' requirements and expectations. As mentioned in Chapter 1 (section 1.1.3.2), certain scholars have studied the perception of localised websites by end users on the basis of functionalist theories (Nord 1997; 2010), which propose that the translator should not focus only on the function of the target document in the socio-cultural environment of reception, but be loyal to both the *Skopos* (or purpose) of the translation and the requests of the commissioner, as stated in the original translation brief.[31] Instead of focusing on Translation Studies approaches to web localisation where the target user plays a critical role, we are more interested here in reviewing prior localisation-related work in which a direct path has been drawn towards the notion of user-centred design (UCD), popular within Web Studies, or towards accessibility, the central topic of our research.

It would not be unreasonable to state that the UCD area that has been more often referred to in previous web localisation publications —probably due to the shared target-oriented perspective of both areas of expertise, as acknowledged by Jiménez Crespo (2013, 36)— is web usability, a concept that we commented on earlier in section 2.2.1. For instance, it has been claimed that badly written sentences, mistranslations, terminology inconsistencies and language drop-down menus in multilingual websites can damage their final usability (Yunker 2003), while the appropriate adaptation of cultural markers, such as icons, colours or other country-specific symbols, can significantly enhance it (Singh and Pereira 2005, 40-45). In the same vein, scholars have put together web localisation and usability concerns with regard to (i) the use of linguistic style guides, asserting that writing adaptation for on-screen reading is one of the many skills professional localisers should possess (Jiménez Crespo 2010), and (ii) how the equal importance in terms of quality of both the design of a website and its linguistic content can affect the localisation process (Pym 2011). Finally, it is also worth highlighting that usability and user experience (UX)[32] considerations have received some attention in the literature regarding the reception by end-users of raw machine translations (Doherty and O'Brien 2012), as well as of controlled text content (Bowker 2015).

When inspecting prior work for accessibility-related concerns within Translation Studies, we have observed that they have traditionally concentrated on audio-visual translation practices, particularly on subtitling, audio description and re-speaking (see, for instance, the Media for All monographs[33] in Brill's Approaches to Translation Studies series). In the localisation arena, accessibility has gained increasing attention in the field of videogames (Mangiron et al. 2014), but the web dimension still remains a highly underexplored topic.

The first reference made to WA in a translation journal is attributed to Fuertes Castro and Martínez Normand (2007), who offered a general overview of the most relevant accessibility implementation and evaluation aspects. However, they never framed these practices within the web localisation process, as Ó Broin (2004) had

[31] See, for instance, Colina (2008), Jiménez Crespo (2009b; 2013) or Fernández Costales (2009).

[32] While the former is more focused on the product (ease of learning, engagement), the latter encompasses issues such as aesthetics, fun and pleasure that define users' emotions rather than the quality of the product. An interesting discussion on the topic can be found in a recent publication by Suojanen et al. (2015), who advocate for a user-centred translation approach to create usable translations.

[33] http://www.brill.com/search?search_title=media%20for%20all Last access: 10th March 2016.

done three years before in *MultiLingual*, a well-known professional localisation magazine. In his article, he stated that we could see localisation as a form of accessibility in its own right, since it takes into consideration a specific target audience and its communicative reality and needs.[34] In addition, he organised the WCAG 2.0 around presentation, structure and authoring categories, highlighting the shared interests of the latter with localisation best practices (e.g. separating content from presentation, facilitating language detection, or writing clear and understandable text). Taking a step further, Gutiérrez y Restrepo and Martínez Normand (2010) later analysed each of the guidelines independently, placing emphasis on the SC they believe to be most relevant to web content localisation (a total of 31 out of 61). As way of illustration we include here some examples: for instance, they suggest that reading sequences are not the same in all languages. Since addressing this issue as a necessary step of the web localisation process, localisers would at the same time satisfy the WCAG 2.0 related success criterion 1.3.2 *Meaningful Sequence* (Caldwell et al. 2008). Similarly, they state that localisation practitioners have to pay special attention to adequately adapting the descriptions contained in headings, labels and sections, which is related to SC 2.4.6 *Headings and Labels* and 2.4.10 *Section Headings* (ibid). Although somehow superficial and descriptive in nature, their review of the WCAG 2.0 in parallel with localisation concerns is, to date, one of the most comprehensive accounts of how accessibility and web localisation complement one another. Additionally, they have been among the first to acknowledge that, for a website to be correctly localised, the new language versions should, at least, be as accessible as the original product. Apart from sharing their opinion in this regard, we also embrace their vision of localisers as professionals who could also improve the accessibility level achieved in the source when creating the target website, as we will later emphasise in the next section (2.4.3).

Yet, before concluding the current one, it is worth mentioning that other scholars have also alluded to accessibility issues related to specific web content elements,[35] or contributed to the discussion about accessibility implementation in localisation from a more holistic point of view. For example, when dealing with navigational aspects in the context of web localisation, Jiménez Crespo (2009a) has noted that terminology inconsistencies associated with navigation menus or links can cause frustration to and mislead visually-impaired users (e.g. referring to the same page as ‘Help’ or ‘Frequently Asked Questions’ within the same website). Tercedor Sánchez (2010), on the other hand, suggests taking accessibility and design best practices beyond the surface level by analysing the way in which all types of existing

[34] Interestingly enough, this idea has been also shared by scholars in the field with regard to translation itself (Schiller 2006; Suojanen et al. 2015, 57).

[35] The localisation literature touching on the particular topic of image accessibility will be reviewed in Chapter 3, section 3.5.2, so we will not refer to it in the present chapter.

information —verbal and nonverbal— may jointly contribute towards the intended meaning or function of the website. According to the author, this could help translators and localisers reduce the cognitive-functional load for all kinds of users by making multimodal information and action prompts cohere and feed one another.[36]

Indeed, as Folaron (2012, 12) proposes, the target users of localised websites have now broadened to include not only those expected from a given locale, but also functionally diverse visitors using different computer operating systems, browsers, and devices to connect to the Internet and to access sites. In this sense, Prieto Velasco (2009, 178) asserts that specialised content producers —as it is in the case of localisers— should be at least familiar with general multimedia content accessibility recommendations, with a view to reaching the widest possible range of users. Accessibility is not, however, (just) a material feature that can be added to the digital product and then be kept or transferred automatically into the target version during the web localisation process. Instead, as can be inferred from the discussions thus far in this chapter, it should be seen as an inclusion-oriented design and communication principle influenced by political, cultural, social, technical and other contextual factors that localisers need to bear in mind when participating in the development of multilingual sites. While designing with internationalisation in mind contributes not only to localisability but also to accessibility,[37] we believe it is not enough to achieve the broader goal of e-inclusion and access to information for all. For this to occur, localisation professionals need to take action, as we argue in the next section.

2.4.3 Localisers' accountability for web accessibility

Jiménez Crespo (2010) considers that for localisers, producing the most usable, effective target web content means not only fully meeting the client's expectations, but also improving end-users' experiences while interacting with localised products. As set forth in Chapter 1 (section 1.1.3.2), within the framework of this thesis, we understand that users from the target audience may be functionally diverse, and that such improvement can also be achieved by assuring that the target website does not fail any of the accessibility requirements established by the WCAG 2.0, as suggested by Gutiérrez y Restrepo and Martínez Normand (2010). This may imply making modifications that go beyond the standard feature set offered by the

[36] A related theoretical discussion about accessibility and localisation has been further developed in Torres del Rey and Rodríguez Vázquez (2013).

[37] Take for instance, the principle that the text-processing language, direction and encoding of a website should be properly declared. This helps not only good rendering by browsers but also by assistive technologies (e.g. screen readers) that need to pronounce the text correctly. Another example would be the appropriate use of styling and layout, based on correct separation of semantics, content and form (e.g. by means of CSS) (Ishida 2015).

base website, a fact that has been already regarded as highly likely during the localisation process when the expectations of the target community are to be met (Roturier 2015, 173). Such access-enabling approach to localisation contrasts with the thought of localisation as a progressive dehumanisation of communication (Pym 2003), and supports the categorisation of professional translation not only as a form of human-computer interaction (O'Brien 2012b), but also as a vehicle to enhance it. In what follows, we further develop these ideas by looking at accessibility as a localisation quality indicator (section 2.4.3.1) and at the localiser as a key actor within multilingual WAE (section 2.4.3.2).

2.4.3.1 Accessibility as a localisation quality indicator

The industry sees quality as one of the most important aspects of the localisation process, but also the most challenging (Lionbridge 2015). It is not surprising, therefore, to observe that the study of quality management procedures throughout the entire web localisation process has received considerable attention in the literature (Dunne 2006; Matis 2011; Dunne 2011). Instead of addressing the topic from a project management perspective, it is our intent to narrow the scope of the discussion to a brief examination of the quality assurance (QA) phase, the aim of which is to check, before delivering the target web product, that optimal quality has been achieved.[38] In this sense, we will focus on which elements or aspects are normally inspected for potential threats to quality and how the identification of these can be automated.

In the context of localisation, the QA process can be broken down into three main areas: (i) linguistic testing, (ii) cosmetic testing and (ii) functionality testing (Esselink 2000). The first focuses on the language dimension of a localised website, such as terminology and style coherence and consistency, grammar, spelling and translation adequacy. The second deals with visual aspects and aims at ensuring the correct visualisation of different elements of the localised product (e.g. the correct display of images, special characters, menu items). The third, in turn, as its names denotes, concentrates on verifying that interaction with the final product is indeed possible (e.g. checking that forms and links work). Roturier (2015, 92) also includes in this list compliance testing (with norms or standards in the target market) and compatibility testing (e.g. to check whether web content is functional in different browsers), which could be regarded as sub-categories of functionality testing. While this separation of QA steps seems convenient at first sight, it can however be misleading, as in the context of digital products it is not always

[38] In this thesis, quality is addressed from a rather pragmatic dimension. To see a thorough theoretical review of industry and academic approaches to quality in Web localisation, see Chapter 5 in Jiménez Crespo (2013).

straightforward to determine where the difference lies between linguistic, cosmetic and functional issues (ibid, 141).[39]

The complexity of the QA localisation phase, where ideally the testing process should be approached from a more holistic perspective, is exacerbated by the limitations in terms of coverage of the QA tools regularly used. These types of tools are normally oriented to automate the linguistic testing phase, as they focus on the identification of (i) formal errors (e.g. punctuation, mistakes in number values, capitalisation, spacing); (ii) translation-related errors, which are regularly checked at a segment-level (i.e. based on bilingual files), such as forgotten or incomplete translations; and (iii) style issues (e.g. terminology consistency) (Makoushina 2007; Debove et al. 2011; Depraetere and Vackier 2011). In recent years, they have also introduced regular expressions functionalities to check exact pattern matches (e.g. special characters, duplicate words, triple letters) and accommodated tag checks, given the markup nature of certain file formats like HTML or XML, with a view to ensuring equivalence and correct order of tags in the target text with regard to the source (ibid).

The few studies that have been published on the use of such tools reveal that, as in the case of WAE tools (see section 2.3.1.3) human verification is always needed to decide whether errors reported are genuine or constitute false positives, the use of localisation QA tools is still crucial and highly relevant, especially because the increasing complexity of the localiser' working environment (variety of tools and files used, high volume of materials to handle, time constraints) makes it more error-prone (Debove et al. 2011, 161). In addition, they are well seen by the industry, given that they can have a positive impact both (i) in the QA process, as they help save time by automating formatting verification tasks that are monotonous and boring (Makoushina 2007), and (ii) in the final quality of the product, as they assist with the identification of formal errors that the human eye has difficulties in spotting (Depraetere and Vackier 2011). We see, nevertheless, several limitations in relying only on localisation QA tools: first, apart from terminology checking against a given database, they do not offer any kind of automation to check the more linguistic aspects of translated content (meaning, textual adequacy); second, they strongly depend on the source document (e.g. if a tag is misplaced both in the

[39] Our main object of study, images on the Web, is a clear example. An image may be correctly displayed on the screen, thus passing both the cosmetic and functional testing. Similarly, its text alternative might have been translated and deemed correct in linguistic terms, therefore not raising any alarm during linguistic testing. Nevertheless, only when examining these three aspects in conjunction, further functionality issues related to image accessibility might be spotted (e.g. inappropriateness of the text equivalent according to the image purpose within the page). This topic will be discussed in detail in Chapter 3.

source and target file, the error might not be flagged), and third, they do not provide any hint about potential functionality issues.[40]

Given these constraints, it has become commonplace to look to QA models, in the forms of checklists or evaluation forms, which an evaluator or reviewer uses primarily as a reference both for linguistic and functionality testing. Perhaps the most widely used in the industry is the LISA QA model, which includes multiple error types (e.g. mistranslation, accuracy, dialect errors), with different weightings and penalties, which mostly revolves around language-related issues. Andreu-Vall and Marcos (2012) offer a more pragmatic approach in this regard, as they propose a heuristic evaluation method based on 24 guidelines with three severity levels that concentrate on rather technical issues, such as the verification of source code aspects (e.g. verification of character set used, declaration of the web page language) or SEO concerns,[41] highlighting the importance of translating page description and keywords, as well as the content of the `<title>` element. Still, these error-driven systems do not provide a complete platform to deal with more serious inadequacies at macro-textual and pragmatic-communicative levels (Jiménez Crespo 2013, 111), nor do they consider other aspects such as end user requirements or the context of reception, which have been put forward as key factors to include in more dynamic QA models (O'Brien 2012a).

This growing interest in the needs of end-users and how they perceive the final localised product is clearly reflected in Jiménez Crespo's (2013, 126–131) proposal of a "real-world model of web localisation quality". The author suggests that localisation QA should be based not only on linguistic and textual approaches —typical of the evaluation of other translation types— but also on other important aspects related to UCD, such as functionality, usability and accessibility. In his view, the former are part of the localised website's internal quality, while the latter, together with the purpose and specifications detailed in the localisation brief, are components of a website's external quality.[42] Figure 2.2 depicts the overlap between the different aforementioned quality indicators.

---

[40] Throught the subsequent chapters of this thesis, we will see how, in the particular context of accessibility QA during the Web localisation process, these limitations could be covered by the combined use of localisation QA tools with other specialised software, such as CL checkers and WAE tools.

[41] Addressing SEO aspects during Web localisation has been put forward in recent work as an efficient approach to communicate with potential target customers/consumers? through the usage of search engines (Lakó 2014).

[42] According to Gouadec (2010), external or extrinsic quality relates to the way a translation product satisfies the expected requirements in terms of purpose, users needs (including accessibility, readability, usability) and functionality, while internal or intrinsic quality refers to the content itself and its form (format, medium).

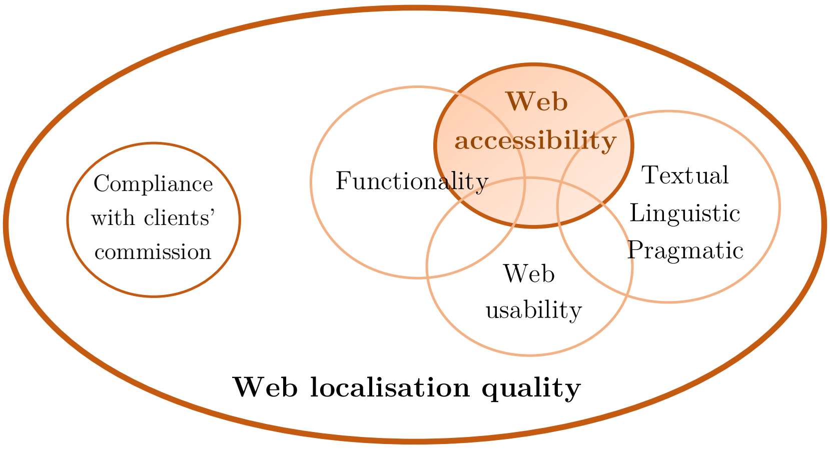


Figure 2.2. Web localisation quality indicators (adapted from Jiménez Crespo 2013, 127)

It should be mentioned here that, in contrast to Jiménez Crespo's (2013, 127) initial proposal —where accessibility received a relatively low level of importance within the quality framework and was believed to share concerns only with functionality and language-related issues— we suggest that the endorsement of WA standards is strongly intertwined with the other three components, as it serves to make sure (i) that the website works, both in technical, cultural and pragmatic terms; (ii) that its content is perceivable and understandable; and (iii) that higher levels of user satisfaction and efficiency are met. Similarly, we believe that given their interrelatedness, equal attention should be paid to all four quality indicators during the web localisation process, and that the observance of the client's or commissioner's specifications should be flexible enough to accommodate potential alterations with regard to the initial objectives when these are made for the sake of inclusiveness, i.e. with a view to adapting the localised product to the different physical and sensory abilities of the target audience. We could therefore expand the functionalist approach that Jiménez Crespo (ibid, 122) adopts when saying that a localised website accomplishes its purpose if (i) it is received as a locally made one, (ii) its web usability is not compromised and (iii) it is functional by adding a fourth condition: if (iv) its content can be rendered through any sensory modality (visual, auditory or tactile) to match the needs of any user.

The consideration of WA as a strong quality indicator of localised websites unavoidably has direct implications on the profile of localisation professionals. At a competence level, they should be familiar with W3C accessibility recommendations, have the technical skills needed to implement them (or at least identify the issues they cover), be aware of how individuals with disabilities interact with the Web and have the instrumental knowledge required to use accessibility-oriented QA tools in combination with localisation-related software.

In addition, joining the call for a more inclusive Web for all necessarily requires a higher degree of involvement of localisers in the multilingual web development cycle in two dimensions: first, it demands a social commitment on their side to boost awareness about WA issues among other actors participating in the web localisation process, including project managers, clients or web commissioners, who might not always be familiar with what is entailed to satisfy the target locale's conventions, legislation, standards and end-users expectations (Dunne 2006, 100–101). Second, we argue that, together with the added-value of having WA knowledge, the wide range of technical, linguistic and strategic skills that lay the foundation of localisation competence can prove beneficial during multilingual WAE processes, a belief that is further developed in the section below.

2.4.3.2 Localisers as contributors to web accessibility audits

As Folaron (2006, 196) states, localisation has been collaborative in nature since its inception. This might be derived from the fact that, as the same scholar puts it, “localisation practice reflects a unique convergence of disciplines: foreign languages, linguistics, translation, computer science, desktop publishing, graphic design and layout, and international business, to name but a few” (ibid, 206). In this sense, it seems reasonable to believe that when a new language version is commissioned, the localisation team should work closely with the original development team, assess what the expectations are and, ideally, address any potential problems in source materials before starting the localisation process itself (Dunne 2006, 114). We argue that this collaborative environment could be maintained on an on-going basis and serve to also discuss accessibility concerns, always taking into account that requirements may vary from one target culture to another, or that legislations may impose different accessibility conformance levels across countries.

Once the website is ready to be operational again —this time as a multilingual site— an integral accessibility assessment should be carried out, based on the same aforementioned knowledge exchange between all actors involved in the web production workflow. This could be of particular benefit for both teams. On one side, if well trained on accessibility parameters, localisers’ intervention could be of added value and contribute to a better informed evaluation, given their background and knowledge of the target audience. Furthermore, as Dunne emphasises (2006, 115), although it is generally assumed that functionality drives user experience, “the vast majority of Human-Computer Interaction, at its most basic level, is governed, and indeed made possible, by language”. On the other side, the technical expertise of developers and designers could be of benefit for the localisation team, whose members might not have a comprehensive understanding of all web technologies needed to correctly implement the most complex accessibility features.

In order to visually represent this working environment, we have taken Abou-Zahra’s (2008, 92) distribution of roles in WA assurance and complemented it by (i) extending it to cover the case of multilingual websites, and (ii) accounting for the reciprocal collaboration among different actors that we have just proposed (see Figure 2.3). Abou-Zahra's vision of the optimal WAE management solution, already alluded to earlier in this chapter (see section 2.2.3), resides on the distribution of responsibilities among multiple web professionals to ensure that the different accessibility requirements covered by the WCAG 2.0 are met. This model is based on the complementarity of the varied skills and background of the actors involved in the production of web content, as well as on the figure of a WA expert or champion, who would lead the evaluation process and offer advice should further improvements be needed (ibid).

While this situation would be the ideal, Abou-Zahra acknowledges that in smaller organisations this "WA master" might not exist, thus implying that the rest of the team members should be responsible for assuring that at least the accessibility specifications related to the web content they manipulate are followed. In the case of the multilingual web production cycle, this would mean that the localiser (or the translator and localisation engineer in larger corporations) might be held accountable for the ultimate level of accessibility achieved in the localised website. In the last section of this chapter, we support this assumption with quantitative and qualitative data collected from a survey on the topic.

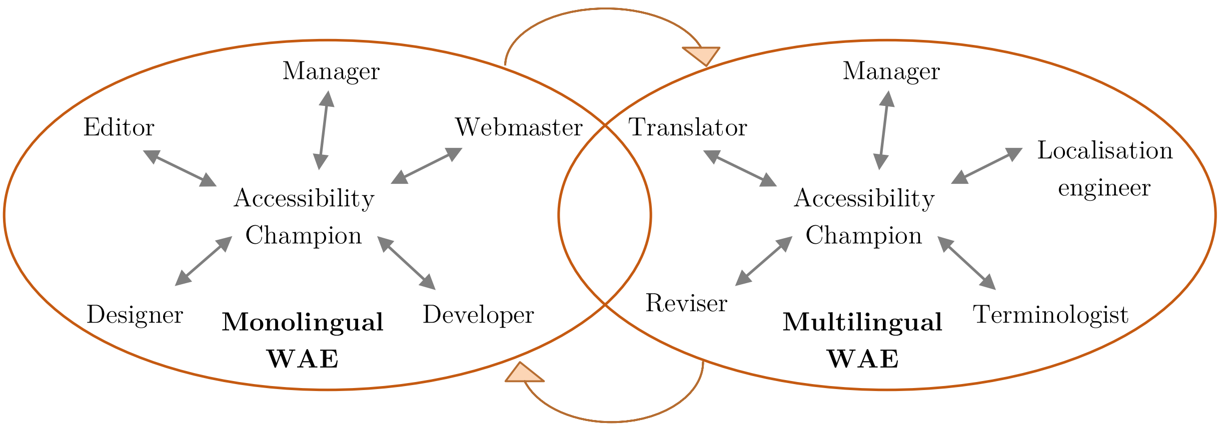


Figure 2.3. Distribution of accessibility-related responsibilities in WAE (adapted from Abou-Zahra 2008, 92)

2.5 Multilingual website assessment for accessibility: a survey

The collaborative environment described in the previous section represents an ideal multilingual web accessibility evaluation scenario. Nevertheless, such a proposal has never been empirically validated, nor have alternative solutions been studied or suggested in the related literature. In fact, up to the present, research on WAE practices has particularly focused on defining and refining the methods applied (see section 2.3.1), as well as on testing the effectiveness of such techniques (see sections 2.3.2 and 2.3.3). However, to the best of our knowledge, no direct path has been drawn towards the analysis of the particularities of multilingual websites and their potential implication on WA achievement and assessment procedures.

In light of the above, we conducted a survey with a view to responding to the lack of attention multilingual website accessibility has received so far. Our intention was to gain insights into the evaluators' behaviour while performing an accessibility assessment job on websites with content in at least two languages. The survey presented in this section differs from prior work in that (i) it makes a clear distinction between monolingual and multilingual websites in the context of WAE, and (ii) for the first time in WA-related research, it considers localisers as web content producers and accessibility stakeholders whose actions might be decisive to attain a more inclusive Web. In what follows, we describe the survey design (section 2.5.1), we present the profile of the respondents (section 2.5.2), we report the main findings (section 2.5.3) and summarise the main conclusions drawn from the study (section 2.5.4).

2.5.1 Design methodology

The goal of this study was two-fold: it aimed at (i) gathering knowledge about the procedures followed in multilingual web accessibility assessment tasks, and at (ii) exploring the perceptions of web professionals about the potential contribution of localisers to the achievement of a higher degree of accessibility in multilingual websites. The survey, which was based on an online questionnaire, addressed web professionals with at least two years of experience in the field of web accessibility, a profile indicator that was later selected as a baseline for the relevance of responses. A 'snowball' sampling method (Oates 2005, 98) was used to recruit targeted respondents online. The survey information and URL was distributed to relevant mailing list servers, such as the WebAIM discussion list, as well as via Twitter and LinkedIn. Before distribution, the survey was checked for content validity (Oates 2005, 211). The questionnaire was deployed via SurveyMonkey, an accessible online survey development and administration platform. It was launched on 16th April 2013 and potential respondents were given six weeks to complete it.

The language of the questionnaire was English. In order to avoid potential misunderstandings, a short explanation about the meaning of localisation was included in the introductory page of the questionnaire. It included a set of screening questions about demographics and expertise, and two larger sections aimed at covering the two main aspects of the above-mentioned survey objective. The questionnaire consisted of both close-ended and open-ended questions (a total of 30), some of which were grounded in conversations with experts and users at earlier stages of this research. Responses to open-ended questions were coded following an inductive approach and allowed us to verify participants' consistency in their responses to multiple-choice questions. The questionnaire in full can be found in Appendix D.1.

2.5.2 Respondents' profiles

Approximately 100 respondents accessed the survey but only 79 completed it. Taking into account the two-year WA experience threshold, 67 usable responses were finally collected from participants coming from Europe (53.73%, N=36,), North America (38.81%, N=26), South America (4.48%, N=3), Asia (1.49%, N=1) and Australia (1.49%, N=1). Fifty-eight respondents were employed and nine reported to be students at the time of the study (N=67, aged between 66 and 22, $\bar{x}$=44, sd=10; 33 male, 34 female). About 68% of the respondents (N=46 of 67) were fluent in at least one language different to their mother tongue, with English being the most popular both as first and second language, followed by French, Spanish and German.

Most professionals had simultaneous jobs, with categories including Web consultants (57%), researchers in the web accessibility field (32%), web developers (28%), web designers (23%) and others (12%, mainly project leaders and webmasters). Around 90% of respondents (N=62 out of 67) stated to have a high degree of expertise in the domain and reported to have assessed between 10 and 50 websites for accessibility in the past year. Among respondents, there were nine (N=9) screen reader users and two (N=2) people daily using speech recognition software.

2.5.3 Main findings

In this section, we report and discuss the survey results concerning the two primary axes of our questionnaire: (i) the actions taken by web professionals when checking multilingual websites for accessibility (section 2.5.3.1) and (ii) their different attitudes towards the responsibility of localisers with regard to the accessibility level of multilingual websites (2.5.3.2). It is worth noting that before surveying respondents about these two aspects, we asked them several questions about their general WA assessment practices, irrespective of the type of website (monolingual, multilingual) they were evaluating. We will include the related information gathered below, prior to the presentation of the main findings, as some data are also particularly relevant for subsequent discussions.

When requested to report on the accessibility evaluation methods they used on a regular basis, most professionals pointed to inspection techniques (85%) and automated testing (70%), followed by user testing, screening techniques and subjective assessment (see Table 2.3). This seems to be consistent with the observations made in prior work about the most popular AEMs among practitioners (see sections 2.3.1 and 2.3.2), which somehow contributes to increase the representativeness of our survey's results despite having relied on a non-probabilistic sampling technique. Of the 67 respondents, 55 (82%) said that they combined two or more AEMs when performing a WA assessment task: 10 (18%) stated that they did so sometimes, 23 (42%) often, and 20 (36%) always, while 2 people did not indicate any frequency. The most repeated combination was conformance review and automated testing (N=38 out of 55, 69%).

Table 2.3. Response count and percent (%) for question 14 (D.1): AEMs generally used

Which method do you use for web accessibility assessment? (you can select more than one option)

Response*	**Total**	%
Inspection methods	57	85%
Automated testing	47	70%
Screening techniques	40	60%
Subjective assessment	20	30%
User testing	42	63%

*Simplified (see Appendix D.1, question 14, to check options given in full)

Through this more generic part of the questionnaire, we also sought to understand the degree of importance attributed to the accessibility of certain elements whose quality assurance would not normally be considered a direct responsibility of web professionals with a more technical-oriented profile, such as those who replied to our questionnaire. We first asked respondents how much time they usually spent on textual or language accessibility-related issues (such as those under WCAG 2.0 principle 3). Findings suggest that the tendency is to not spend more than 25% of the time available for the task on text content. Interestingly, three respondents said that they were normally in charge of verifying only language accessibility aspects, while two replied that they would only look at them if the client requested it (see Figure 2.4 and Table 2.4). We also asked them about the consideration given to culture related elements as potential accessibility barriers (e.g., symbols, shapes, colours, signs). From the 67 respondents, only 21% of the respondents (N=14) answered that they always verify them (21% never, 45% sometimes, 7% often, and 6% did not want to reply to the question).

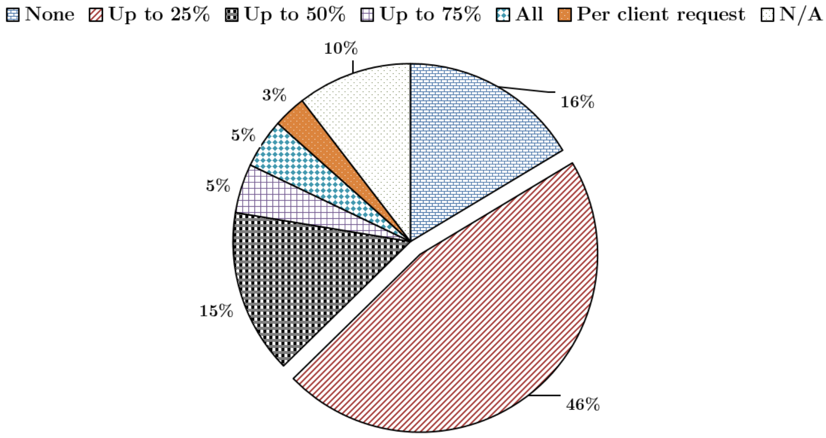


Figure 2.4. Time spent (%) on text-related issues during WAE

Table 2.4. Response count and percent (%) for question 18 (D.1): textual accessibility

When performing a WA assessment task, how much time do you spend on textual accessibility-related issues?

Response	Total	%
I do not assess textual accessibility	11	16%
I spend up to 25% of the time	31	46%
I spend up to 50% of the time	10	15%
I spend up to 75% of the time	3	4%
I only check the website for textual accessibility	3	4%
Other (per client request)	2	3%
N/A	7	10%

2.5.3.1 Accessibility evaluation procedures for multilingual websites

After the two first sections of the questionnaire, which aimed at collecting data on the respondents' demographics and getting an insight into their WAE habits from a broader perspective, we included several questions about the accessibility assessment procedures they followed when dealing with multilingual websites. As a measure to guarantee the validity of the data gathered through this part of the survey, we asked respondents whether they actually had any experience on the matter. Of the 67 participants, 15 (22%) reported that they never assess multilingual websites in their daily practices; 29 (43%) indicated that up to 25% of the websites they evaluate on a regular basis are multilingual; 7 (11%) up to 50%; 4 (6%) up to 75%; 5 (7%) more than 75%; and 7 (11%) stated that all the websites they usually assess have more than one language version. Therefore, the findings we report below are only based on the responses of 52 web professionals.

First, we were concerned with knowing if all language versions in a multilingual website are checked for accessibility. Most respondents (N=35, 67%) preferred to verify all language versions (see Table 2.5). Those who said they normally assessed just one version (N=17, 33%) argued the following: ‘I only speak one language’ (N=4, 23%); ‘Although I speak other languages; I do not feel comfortable assessing websites that are not in my mother tongue’ (N=3, 18%); and ‘I do not have time’ (N=3, 18%). Only one respondent (N=1, 6%) considered that it was not necessary to assess the accessibility of all language versions, and the remaining respondents (N=6, 35%) gave reasons for not doing it that were mostly related to the AEM used. For instance:

> R9: “While doing user testing, we can only assess the version of the end users’ mother tongue” [user testing]

> R43: “Technical difficulties: we do not have screen readers handling non-Latin languages” [screening techniques]

Table 2.5. Response count and percent (%) for question 21 (D.1): language versions

When performing a WA assessment task of a multilingual website, do you test just one version of the website?

Response	**Total**	**%**
Yes	17	33%
No	35	67%

It was interesting to realise that multiple professionals saw multilingual websites as problematic in terms of accessibility only because of text-based issues, and expressed no or little concern ―at least spontaneously― about other elements that can be changed during the web localisation process, such as images, links or the general content structure, to name just a few. Similarly, we observed that it was often assumed that only few accessibility standards referred to language barriers, which may explain why not too much time is dedicated to assess them, as we pointed out earlier. Notice, for example, the following comments:

> R6: “I test the features to make the website accessible and Communications [section] is in charge of ensuring the use of plain language.”

> R62: “I do spot checking in other languages to ensure the content has not been altered. Since most accessibility issues are not text-based this seems like an adequate approach. I can think of scenarios where this wouldn't suffice (where content is not a direct translation but alters based on other criteria, for instance), and I may request an assessment in the other language.”

Additionally, respondents were asked to estimate how often the following statements described their WAE procedure on multilingual websites: (1) *As I check a web element or functionality, I recheck it in the other language versions;*

(2) *I first check one language version, and then I check the others;* (3) *I just check my native language version, and if I find a major accessibility problem, I check if it is present in the other versions too;* (4) *I just check my native language version and I assume that the others have the same degree of accessibility as the one I checked.* Around half of the respondents (N=27, 52%) concluded that procedure (4) was not appropriate, which means that 25 professionals acknowledged that at least in some occasions they only checked the language version that corresponded to their mother tongue. This contrasts somehow with the data from question 21 discussed earlier, which seemed to indicate that more than half of the professionals surveyed asserted to normally verify all language versions of a multilingual website. As far as the other three procedures are concerned, no major agreement was found among respondents (see Table 2.6 and Figure 2.5).[43] It appears as though the sequences depicted in (1) and (2) are considerably more popular, as the former was the least followed procedure and the latter was the one consistently used by the highest number of respondents (N=12, 23%).

Table 2.6. Response count and percent (%) for question 23 (D.1): multilingual WAE procedure

Please estimate how often the following statements describe your web accessibility assessment procedure of multilingual websites

Response*	**1. Never**	**2. Sometimes**	**3. Often**	**4. Always**	**N/A**
Procedure 1	14 (27%)	17 (33%)	11 (21%)	8 (15%)	2 (4%)
Procedure 2	15 (29%)	11 (21%)	11 (21%)	12 (23%)	3 (6%)
Procedure 3	19 (36%)	12 (24%)	12 (23%)	3 (6%)	6 (11%)
Procedure 4	27 (52%)	10 (19%)	8 (15%)	3 (6%)	4 (8%)

*Simplified (see Appendix D.1, question 23, to check options given in full)

Despite the lack of experience on the particularities of multilingual websites of the 15 people who reported to have never assessed any for accessibility, we considered that their opinions about more general issues, not directly related to their recent activity, could be of interest to our research and were therefore taken into account during the data analysis of the remaining questions. For instance, we asked respondents whether different approaches should be adopted when testing monolingual and multilingual websites against accessibility standards. Overall, 57% of respondents considered that they should not be assessed for accessibility differently, 37% stated the contrary, and 6% preferred not to answer the question (see Table 2.7). Those who would apply a different approach (N=25 out of 67) suggested they would look at the following elements separately (they could select

[43] To better interpret the gray scale colour coding used both in this figure and in Figure 2.6, notice that lowest value (Never) is always at the bottom of the graphic (i.e. lowest part of the y-axis) and highest (Always, full black) is at the top.

more than one option): textual content (80%), multimedia content (72%), graphical content (56%), navigation and hyperlinks (44%), semantic structure (40%) and presentation layout (36%). Again, and as expected, text appears to be regarded by WA experts as the type of web content that is more likely to change across language versions. However, when being prompted with other web elements,[44] respondents gave also high consideration to other aspects.

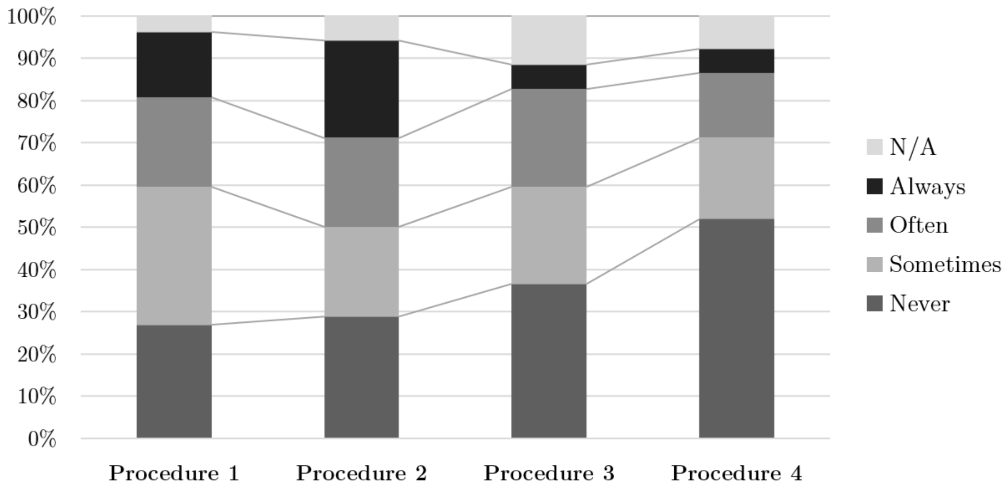


Figure 2.5. Preferred procedures during accessibility assessment of multilingual websites

Table 2.7. Response count and percent (%) for question 25 (D.1): monolingual and multilingual websites

Do you considered that multilingual websites and monolingual websites should be tested for accessibility differently?

Response	**Total**	**%**
Yes	25	37%
No	38	57%
No answer	4	6%

Since a great part of the survey was conformed of multiple-choice questions, respondents were given the opportunity at the end to openly express their point of view about what the ideal accessibility assessment procedure to follow in the case of multilingual websites would be. We have decided to summarise here the most representative opinions found in the qualitative data gathered from the 67 participants.[45] A general tendency was observed with regard to language expertise:

[44] Question 26 was close-ended. See Appendix D.1 to check the responses proposed in full.

[45] It should be noted that this was the last question of the survey, so Web professionals had already replied to the three questions reported in the next section (2.5.3.2) where direct reference to localisers was made.

most respondents explained that a collaborative team of native speaker experts and end-users for each language should be involved in accessibility evaluations of localised websites. In addition, data shows that culturally diverse and/or multilingual experts would be preferred for the task, but all seemed to acknowledge the challenge behind this. In particular, one respondent from a bilingual country supported this opinion from a government website perspective:

> R59: “It's not easy to find accessibility evaluators that are simply bilingual. The result now is that the accessibility is excellent in the language of the evaluator and very uneven in the other language. A solution would be to have localisation professionals trained to evaluate only the ‘challenging multilingual components’ of accessibility (main text, images, table captions, form field IDs…).”

Opinions on the level of importance and priority given during WAE to the native language version of the site versus localised pages are considerably unbalanced. An interesting approach that was repeatedly suggested was to choose a set of representative pages of all versions available for assessment —taking into account the factors mentioned at the beginning of section 2.3 of this chapter— as well as a random selection of individual pages for each language version, regardless of which is the original source. Then, as respondent R62 puts it, the optimal strategy would be to “just check the website feature by feature, first one language, then the other.” Nonetheless, lack of time, limited funding and lack of client support were often highlighted as potential drawbacks to such a procedure. For instance, respondent P46 said: “Anything is possible, as long as you can find a client who is willing to pay for such detailed assessments.” Overall, 64% of the usable responses to this last open-ended question (N=40, 62 out of 67 answered it) pointed directly or indirectly to the relevance of development teams working closely with localisers, or to the responsibility of localisation professionals as contributors to multilingual WA. Some examples are provided below:

> *The ideal multilingual accessibility assessment procedure would be...*
>
> R5: “1) Localisation expert gives initial input; 2) development team with requirements analyst and product manager give input; 3) QA team does initial testing; 4) QA sub-specialist does accessibility and multilingual testing; 5) localisation expert does another pass to see if we got it right.”
>
> R16: “1) Web content prepared by web localisation professional; 2) implemented by web developer; 3) checked for accessibility by both developer and localisation professional.”
>
> R53: “Act collaboratively: designers, developers, web localisation professionals... since many components have to be considered for assuring accessibility of multilingual websites.”

> R62: "Establish how internationalisation/localisation is handled, and plan for special accessibility considerations accordingly."

2.5.3.2 Perceptions of localisation professionals

The data that we have just presented has already provided some insight into the level of involvement that, according to WA experts, localisers should undertake to render multilingual websites accessible. However, before they could freely express their opinions, we asked respondents three specific questions directly linked to web localisation professionals. We will report the findings from these questions before drawing any further conclusions.

Answers to question 24 (see Appendix D.1) *If you were assessing a multilingual website for accessibility, would you appreciate receiving also the feedback of the localisation professional(s) who adapted the web to the other language(s) available?* illustrate that WA experts seem to agree that the participation of localisers in the accessibility evaluation process could contribute to a better informed assessment: 66% (N=44) of them replied 'Yes' and 16% (N=11) said 'No'. Seven percent of the respondents (N=5) did not answer the question and the remaining 11% (N=7) indicated that it depended on the web element or functionality they were assessing (e.g., images, text, multimedia content, etc.), placing special emphasis on text content. Many of those who welcomed the idea of receiving feedback from localisers suggested that the presence of these professionals should be continuous throughout the whole WAE. The comments below illustrate the two last statements made:

> R44: "If there is a noticeable difference in the accessibility or textual layout, I would (accessibility is very concerned with plain, concise and understandable text, something that has to be maintained when content is translated)."

> R17: "There is not only a need for automated tests dealing with multilingual issues, but also for a localiser with basic accessibility knowledge sitting next to the accessibility auditor."

Furthermore, and inspired by a similar question in prior research (Lazar et al. 2004, 280), WA experts were directly asked about who should be responsible for making the Web accessible. This time, however, we made explicit reference to the case of multilingual websites. The tendency observed was the same as in other open-ended questions already discussed: respondents appeared to embrace the belief that multilingual web accessibility is a joint commitment among webmasters and content producers, including editors, designers, web developers and localisers. Respondent R5, for instance, stated: "This is a group effort, and having more than one opinion and slant is important." More specifically, from the list of actors provided, respondents considered, almost unanimously, that localisers have a considerably higher level of responsibility on the matter than webmasters, holding the former nearly as equally accountable for accessibility as web developers and editors (see Figure 2.6 and Table 2.8). Project managers and User Experience (UX) designers were mentioned by six respondents as two professional groups who should be also involved in implementing and assuring an accessible website.

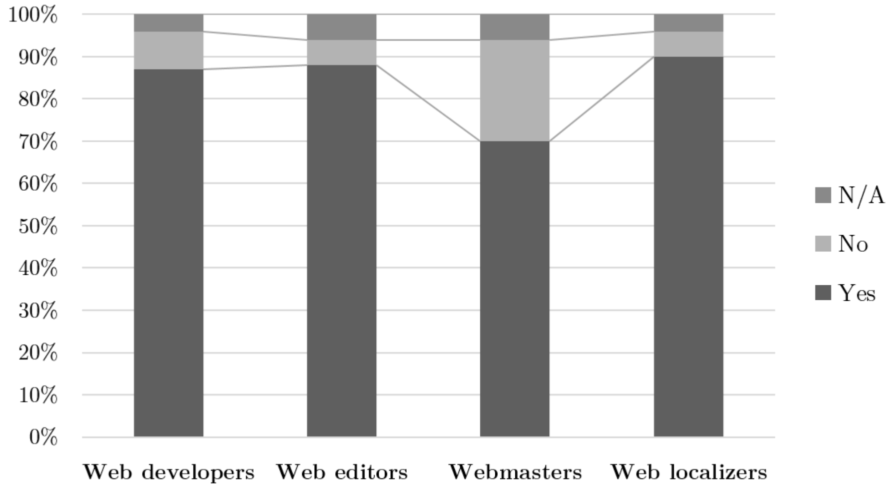


Figure 2.6. Professional groups who should be responsible for creating accessible multilingual websites, according to web experts (% of votes)

Table 2.8. Response count and percent (%) for question 28 (D.1): Accountability for accessibility in multilingual websites

Who do you think should be responsible for making a multilingual website accessible? (you can select more than one option)

Response	Yes	No	N/A
Web developers	58 (87%)	6 (9%)	3 (4%)
Web editors	59 (88%)	4 (6%)	4 (6%)
Webmasters	47 (70%)	16 (24%)	4 (6%)
Web localisers	60 (90%)	4 (6%)	3 (4%)

Given that, by the time the questionnaire was administered, we knew that our research work would later focus on image accessibility, we included one specific question about this type of content (question number 27 in Appendix D.1). As we had anticipated, most respondents declared they would also appreciate localisation professionals' help in the particular case of image accessibility assessment. They could select among different statements to explain their reasons, and they were allowed to choose more than one option. The response distribution was as follows: (1) 'Images in the language version I am assessing might contain culture-related information that should have been taken into account in the localised versions' (69%, N=46 out of 67); (2) 'Images might vary from one version to another and I cannot assess the quality of their alternative text in other languages different to my mother tongue' (51%, N=34); (3) 'Some images might have been removed or changed in certain language versions and I do not understand why' (31%, N=21). Only nine respondents (N=9, 13%) believed that asking the localisation professional for feedback about the accessibility of images was not necessary. As can be inferred from these numbers, and contrary to expectations, WA experts seemed to acknowledge the localiser's role regarding image adaptation. Respondent R65 pointed out the following: “Images may themselves contain text that would somehow need to be localised”, while respondent R67 put forward a concrete case:

> R65: “The images that are significant on our site are very technical. Localisation professionals would need to decide based on their knowledge of their local audience whether to generate different images or not.”

Similarly, it was interesting to observe that more than half of the WA experts surveyed implicitly approved the need to translate and adapt text equivalents for images, as well as the importance of such actions for accessibility: a topic that will be addressed in more detail in section 3.5 (Chapter 3). From an anecdotal perspective, one of the most remarkable comments left by respondents with regard to the issue of image accessibility was that of R44, who suggested that WAE tools should support the verification of image text alternatives in multiple languages and should be able to flag uninformative ones. As was announced in the Introduction of this thesis, we took up this challenge by proposing a language-based solution to precisely provide those functionalities. Since such a proposal will be examined later on (see Chapter 4), we close this section with the call to action issued by the aforementioned respondent:

> R44: “It would be important to develop automated web accessibility checkers that check for alt text phrases such as (picture) (click here) and other non-descriptive phrases in more than just one language. I think a database of these phrases could be constructed and used by such checkers.”

2.5.4 Conclusions

Our survey has offered a valuable first insight into the current accessibility assessment practices followed by WA experts who work with web content available in more than one language. In addition, it has provided a snapshot of these professionals' perceptions around the potential of localisers as key actors in the achievement of multilingual web accessibility. In spite of the limitations to this study —including the small sample size and other potential threats to validity such as the fact that the questionnaire was only administered in English— we deem it relevant to highlight that, above all, respondents found it was thought-provoking and invited further discussion on the topic. In what follows, we highlight the key takeaways from this study.

Multilingual WAE

Data gathered seems to indicate that, overall, little consideration is given to culture-embedded elements and textual content during WA assessment tasks. While particularly worrisome, as this type of content permeates all the Web, these findings are in accord with recent studies claiming that language-related accessibility issues are often the most difficult to assess (Brajnik et al. 2012, 8:19) or left out during the WAE (Hanson and Richards 2013). At a more specific level, evidence found suggests that currently no standardised assessment procedure exists when checking multilingual websites for accessibility issues. This could be interpreted as a direct consequence of the scarce attention this topic has received not only from academia, but also from standards bodies like the W3C, as we have pointed out throughout the present chapter. In addition, the high level of discrepancies observed denotes that the evaluator effect, the phenomenon discussed in section 2.3.3., could be exacerbated in the case of multilingual websites if a consistent evaluation methodology and distribution of tasks is not put in place. Irrespective of the lack of a clear agreement as to which procedure to follow, it appears that verifying every language version is seen by many WA experts as the ideal solution to assure accessibility in multilingual websites. However, according to these professionals, lack of time, human resources, funding, additional language skills and client support makes it difficult for this strategy to be adopted in real life situations.

As a result of the above, during multilingual WA audits, a considerable number of accessibility experts seem to focus on the site version that corresponds to the language they are more fluent in. In our opinion, the accessibility of multilingual websites can be severely compromised by such practice, as the parallelism across different language versions should not be taken for granted. Yet, if they could, WA experts reported that they would inspect certain elements separately (i.e. per language version) for accessibility, such as text, graphics, hyperlinks or page layout. Interestingly enough, these represent the pillars of any multilingual website, where localisers play a critical role.

Localisers' accountability for accessibility

Indeed, the involvement of web localisation professionals in accessibility implementation and evaluation was both implicitly and explicitly referred to by respondents of our survey. WA experts seemed to agree that accessibility should become a primary goal for all people involved in the web development cycle, regardless of the number of language versions available in a given website. Still, in the case of web multilingualism, localisation practitioners' feedback was considered as particularly relevant during accessibility assessment tasks, especially for textual and graphical content. As regards accountability on multilingual web accessibility, localisers ranked #1, followed by web editors, developers and webmasters. Overall, the evidence gathered in our questionnaire supports the arguments that we have gradually put forward in this chapter to advocate for accessibility as a localisation quality indicator, as well as for a higher participation of localisers in multilingual WA audits. Given the constraints previously outlined with regard to the adoption of a sound WAE procedure for multilingual websites, it seems reasonable for localisers to acquire the necessary competences and get further involved in the achievement of a more accessible Web for all.

Although to the best of our knowledge, this study is the most comprehensive account of accessibility in the multilingual Web produced so far, it should be only regarded as a starting point for further research on the matter. For instance, in light of the answers to our survey, it would be appropriate to conduct multilingual website assessment experiments, with both expert evaluators and end-users, in order to contrast and complement the data obtained with direct observations of real-case scenarios. This would enable us to identify in a more reliable manner which are the major flaws of the procedures currently followed. Similarly, results from these empirical studies could be used as a basis to define a set of specific guidelines for multilingual WAE and to better understand how tasks could be more efficiently distributed among review team members.

2.6 Summary

This chapter has laid the groundwork of our research while seeking to cover, at the same time, the first goal of this thesis: Goal 1 (see section 1.3, Chapter 1). We have gone through the most important concepts in the field of web accessibility with a view to providing a framework in which to start and develop the discussion on the achievement of a more accessible multilingual Web for people with disabilities. To this end, we have focused in particular on the existing literature about two determining steps of the Web for All paradigm: accessibility implementation and evaluation.

In section 2.2, the interdependent relationship between accessibility, usability and universal design was briefly discussed, along with the general benefits of embracing web accessibility. Concretely, we have reviewed some of the legal and policy-related measures taken at a national and international level to support the principles of equality, non-discrimination and social inclusion; we have outlined the technical and financial reasons that could motivate web stakeholders to apply accessibility best practices (e.g. enhanced SEO, higher customer satisfaction, reduced site maintenance efforts); and we have depicted WA as a social commitment that should be shared by everyone for the sake of removing interaction and communication barriers. The last part of section 2.2 was dedicated to presenting the WCAG 2.0 defined by the Web Accessibility Initiative of the W3C as well as to examining prior work about their implementation. This has allowed us to identify the major obstacles that still hamper the creation of accessible websites (see section 2.2.3.3). This includes, among others, the poor spread of accessibility law, the lack of the necessary WA knowledge and training by the main professionals responsible for the Web, the flaws in current guidelines and the limitations of automated tools in effectively assisting content producers with their implementation.

Section 2.3 offered an overview of the most popular accessibility evaluation methods that are normally put in place to assess how well Web content can be used by people with disabilities. While user testing techniques imply asking end users to perform goal-free navigation or goal-oriented tasks on a website, manual and automated testing methods require the direct participation of evaluators. During the former, web content is manually inspected for its accessibility based on a set of predefined guidelines, whereas the latter relies on the use of WAE tools. Previous studies have confirmed that these two techniques ―automated testing and manual testing (especially conformance review)― are the most widely used among practitioners, despite two primary disadvantages (see section 2.3.3): (i) the *evaluator effect*, a phenomenon that is caused by differences among evaluators regarding their level of WA expertise and that can negatively affect the reliability of the WAE; and (ii) the *lack of effectiveness* of automated accessibility evaluation software, which usually performs poorly in terms of coverage, completeness and specificity. Given their importance, these two variables (WA knowledge and use of WAE tools) will be further studied in the core experimental study of this thesis (see Chapters 5 to 7).

The comprehensive literature review we performed to investigate how accessibility is implemented and assessed today revealed that, to the best of our knowledge, no particular indications exist with regard to the assurance of accessibility in multilingual websites. The W3C only establishes that when a website contains multiple language versions, these should be assessed for accessibility individually and that only a partial statement of conformance can be issued if only one language version has been subjected to an accessibility audit. This has encouraged us to explore in section 2.4 whether accessibility considerations had received attention within Localisation Studies. While shared concerns between internationalisation and accessibility, as well as between localisation and user experience (UX) have been put forward in the past, we found that the direct relationship between accessibility and localisation has been, to date, a rather underexplored topic. At the end of the section, we indicated why (i) accessibility should be considered as a localisation quality indicator and (ii) localisers should play a more important role in multilingual WAE.

Finally, in section 2.5, we presented the results of the first survey conducted with a view to gaining insight into the evaluators' behaviour while performing an accessibility assessment job on websites with different language versions. Answers from 67 web professionals with at least two years of experience in the field of WA suggest that no standardised assessment procedure exists when checking multilingual websites for accessibility issues, and that localisers are held accountable by the WA community for assuring that accessibility standards are met in the multilingual Web. Grounded on this last statement, this thesis seeks, in subsequent chapters, to further examine the importance of localisers bearing in mind web accessibility challenges while carrying out localisation tasks. To do so, from now on, the focus will be shifted to image accessibility issues.

Chapter 3

Image accessibility, assessment and localisation

Images represent the main object of study of this thesis. In this chapter, we explore how they can be rendered accessible on the Web. We describe the strategies, tools and existing challenges associated with this task, and we discuss how image accessibility should be addressed during web localisation with a view to encouraging a more accessible multilingual Web for all.

3.1 Overview

As discussed in Chapter 2, internationalisation, localisation and accessibility have always been closely linked, if only because their ultimate goal is to make a product accessible to a wider range of users than originally designed for. In addition, they support one another in a reciprocal manner, hence the need for a further collaboration between all the actors participating in the multilingual web development cycle. This need was also acknowledged by web accessibility experts participating in the survey we just presented in the previous chapter, who considered that web localisers should be involved in the achievement of a more accessible Web for all. In this thesis, we have decided to explore the importance of this collaboration when it comes to facilitating the access to images on the Web by blind users.

The accessibility of graphics and multimedia is perhaps the most challenging area in accessibility research (Regan and Kirkpatrick 2008). For screen reader users to be able to perceive graphical content such as images —for instance, via tactile or audio output—, a text equivalent must exist. While this might seem a trivial and simple mechanism to implement, prior work has shown that the non-use and the incorrect use of image text alternatives still emerge as the most frequent, basic accessibility errors (McEwan and Weerts 2007; Access for All 2011), as well as the most common accessibility complaint among visually-impaired users (Petrie et al. 2005). We argue that images still represent a major stumbling block to web accessibility due, in great part, to the limitations of *existing guidance* on how to render them accessible, as well as to poor *automated support.* A recent large-scale web accessibility study has suggested that the generation of inappropriate text alternatives for images may also be linked to the lack of *awareness about their functionality by web professionals* (Hanson and Richards 2013).

In this chapter, we aim at examining these three axes in the following manner: in section 3.3, we present the existing mechanisms that the literature prescribes and/or recommends to introduce and write appropriate image text alternatives in a web document, and we highlight the reasons why, in our opinion, they are incomplete. In section 3.4, we examine how web accessibility evaluation (WAE) tools support the image accessibility assessment task and whether the limitations observed in the existing guidance can be covered or not by the use of these tools. Finally, in section 3.5, we bring to light what exactly the technical challenges of the image localisation process are and we investigate whether the notion of image accessibility has received sufficient attention in this field, both by scholars and localisation professionals. It should be noted that, when discussing the topic of image accessibility, we will adopt a general perspective, that is, we will not focus on any image type in particular.[1] Instead, special importance will be given to the functional role that images have within a web page, as described in the following section (3.2).

3.2 Images on the Web

The amount of images available on the Web is countless. A popular video[2] from 2012 estimated that there were over 350 billion images stored in the vast World Wide Web, while according to Mary Meeker's Internet Trends report[3], only in 2014 people uploaded an average of 1.8 billion digital images every day (which annually represents almost double the estimated number of images on the Internet in 2012). These estimates primarily take into account those images individually uploaded to web applications for graphical content storage and sharing, such as image repositories (e.g. Yahoo's Flickr service) and social media networks (e.g. Twitter or Instagram). In this thesis, the focus is rather placed on images that are embedded in HTML documents (websites) and whose interpretation depends both on their interaction with the surrounding content and with the end users themselves.

Previous studies have claimed that the amount of images present in an average homepage ranges between 28 and 63 (Yunker 2003, 296; Petrie et al. 2005). Most of these images often play a key function within the web document and identifying this function is crucial to determine which is the most appropriate method for rendering them accessible, as will be later discussed in section 3.3. Hence, within

[1] In the literature, it is possible to find accessibility studies about particular image types, such as map plans (Madugalla 2015), chemical diagrams (Sorge et al. 2015) and complex charts (Sharif et al. 2015), among others.

[2] https://vimeo.com/41408616 Last access: 16th December 2015.

[3] http://www.kpcb.com/internet-trends Last access: 16th December 2015.

web accessibility studies, it has become commonplace to classify images taking into account their **content** (what is shown in the image) and/or their **purpose** (the reason why the image exists in that document or, in other words, why the image is shown to the user). Since both criteria are intrinsically interlinked, scholars and web-oriented institutions have often combined them when attempting to put forward their own image classifications. This fact, together with the inherent subjectivity linked to the interpretation of images, might be one of the reasons why there is not a unique image categorisation that is used systematically across the literature. In what follows, we aim at illustrating this heterogeneity by examining those classifications that have been most reiteratively referred to in prior web accessibility work.

Vorburger (1999, 7) proposes a three-level image categorisation: *illustrations*, understood as images which carry information and graphically explain or interpret some information already contained in the text; *navigation aid images,* described as "graphical buttons and similar images which appear inside a link"; and *images for presentation and decoration*, which are used to render the web page visually attractive and "do not usually contain any valuable information" (ibid). The first category has a content-related name, but its description is focused on the purpose. Conversely, the name of the second category denotes the purpose of the images that fall therein, but the definition is merely based on the images content and composition. The third category appears to cover two different purposes (presentation and decoration), yet its definition is somewhat fuzzy and the examples provided to illustrate it (graphical rules, bullets or transparent images) are only considered by the author as presentational, and not decorative.

In her doctoral work, Tang (2012, 23–24) proposes a clearer classification since she aims at focusing only on image purpose. She distinguishes between *informative images*, whose goal are to communicate concrete information to the user, and *control images*, which allow the user to perform specific actions (such as navigate to another web page or submit a form). In contrast to Vorburger, Tang treats images for presentation and decoration individually. She asserts that, apart from adding visual appeal to the document, *decorative images* can also serve to evoke an emotion or create a mood or atmosphere for the user. *Formatting images*, however, only provide structure to the document. These would correspond to Vorburger's presentational images, which according to both authors are often used by professional web developers and designers for layout purposes.

Connor (2012) and accessibility-oriented initiatives such as the WAI-W3C and WebAIM (2015b) all refer to control images (as per Tang's terminology) as *functional images*, defining them as "images used to initiate actions rather than to convey information" (Eggert and Abou-Zahra 2014c). They also agree on the need for differentiating these from *decorative images*, which "have no real functional aspect but are useful only as a purely visual embellishment" (Connor 2012, 180).

Specifically, Eggert and Abou-Zahra (2014c) indicate that images may be decorative when they are:

- Visual styling such as borders, spacers, and corners;
- Supplementary to link text to improve its appearance or increase the clickable area;
- Illustrative of adjacent text but not contributing information ("eye-candy");
- Identified and described by surrounding text.

A major source of disagreement seems to appear when authors and institutions try to classify images whose purpose is not clearly functional or decorative. To these two categories, Connor (2012) adds another four, namely (i) *visually-rich images*, such as photos, drawings or paintings; (ii) *graphs and charts*; (iii) *images of text*, that is, images where text is presented to the user in graphical format; and (iv) *icons*, defined by the author as "images used as a visual clue and are part of a link". Given this last definition, one may wonder why icons are treated separately from functional images. Similarly, one might argue that graphs and charts could be safely considered as visually-rich images for the sake of simplicity.

Eggert and Abou-Zahra (2014c), main editors of the WAI-W3C web accessibility tutorials, also consider images of text as a separate category, together with functional and decorative images. To these three categories, they add (i) *image maps*, that is, images divided into selectable regions ("hotspots") —usually linked to other pages— that allow user interaction; (ii) *informative images,* which "convey a simple concept or information"; (iii) and *complex images*, which "contain substantial information", including graphs and charts, diagrams and illustrations where the page text relies on the user being able to understand the image, or maps showing locations or other information such as weather systems. The misleading combination of content-driven and purpose-driven categories observed in Connor (2012) is present as well in Eggert and Abou-Zahra's classification, where image maps are treated individually, despite having a functional value. In the same vein, while complex images are also informative, they seem to be considered as a separate item, probably because of their high information load.

Finally, it is worth looking at WebAIM image classification, founded on three categories: decorative images, functional images and *advanced images* (2015b). Under this last category, for which no concise definition is provided, WebAIM includes buttons, image maps, background images, logos and complex images, such as charts, graphs and maps. Although at first sight one could infer that these examples are treated separately because they could be either considered as decorative or functional, it would be extremely rare to find, for instance, a button which has a merely decorative role and is not associated with an action (e.g. "Submit" or "Search" buttons). Similarly, background images often play a decorative role in a web document and when they do not, they should be considered as informative or referential. In an attempt to cope with the terminology (and even semantic) inconsistencies observed in the multiple image classifications existing in the literature, we have organised the categories discussed above around what we consider are the main communicative values that an image can represent on a web page: descriptive or referential, functional, and decorative or aesthetical (see Table 3.1).

Table 3.1. Reorganisation of image categories found in the literature as per images broader communicative value

Descriptive/Referential	Functional	Decorative/Aesthetical
Illustrations *(Vorburger 1999)*	**Navigation aid images** *(Vorburger 1999)*	**Decorative images** *(Tang 2012, Connor 2012, Eggert and Abou-Zahra 2014, WebAIM 2015)*
Informative images *(Tang 2012, Eggert and Abou-Zahra 2014)*	**Presentational images** *(Vorburger 1999)*	**Advanced images** —background images— *(WebAIM 2015)*
Graphs and charts (*Connor 2012)*	**Control images** *(Tang 2012)*	
Visually-rich images *Connor 2012)*	**Formatting images** *(Tang 2012)*	
Complex images *(Eggert and Abou-Zahra 2014)*	**Functional images** *(Eggert and Abou-Zahra 2014, WebAIM 2015)*	
Advanced images —charts, graphs, maps— *(WebAIM 2015)*	**Image maps** (*Eggert and Abou-Zahra 2014)*	
	Advanced images —image maps, buttons— *(WebAIM 2015)*	
Icons *(Connor 2012)*		
Images of text *(Connor 2012, Eggert and Abou-Zahra 2014)*		
Advanced images —logos— *(WebAIM 2015)*		

A **descriptive or referential image** is included in a web document to convey new information that has not been presented yet on it, to depict a particular concept or to complement an idea that has been already described elsewhere on the page. The interaction that occurs between the user and referential images is mostly static and their content often receives more importance than their purpose, which is primarily to inform. Conversely, the content of **functional images** is not as critical to their role within the page. Images with a functional value serve to (i) visually organise and structure web content or to (ii) facilitate a dynamic interaction between the web page and the user (i.e. trigger an action). **Decorative or aesthetical images** may be placed on a web page to elicit emotion or simply to render the document more attractive to sighted users. We believe that the communicative value of icons, logos and images of text may vary according to whether they are linked or unlinked elements,[4] hence the decision to not include them under any specific category (see Table 3.1). Henceforth in this thesis, when discussing image accessibility issues, we will often refer to the general image classification that has just been presented. Concrete visual and code-based examples about each category will be gradually presented throughout this chapter.

3.3 Rendering images on the Web accessible

While the complementarity of textual and graphical content is a key aspect in web design, the major challenge encountered by web developers and designers is to use both in a way that serves the needs of a wide variety of users at the same time (Regan and Kirkpatrick 2008, 289). The use of images can dramatically enhance the access to textual information on the Web for people with cognitive disabilities. However, because the blind or low vision user is likely to be using a text-based alternative output (e.g. a screen reader) when browsing a website, images might well become an accessibility barrier for this population group.

To prevent this from happening, the Guideline 1.1 from the W3C Web Content Accessibility Guidelines (WCAG) 2.0 recommends: *Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language* (Caldwell et al. 2008). More specifically, its associated success criterion (SC) 1.1.1 (level A)[5] indicates: *All non-text content that is presented to the user has a text alternative that serves the equivalent purpose* (ibid).[6] By non-text content, the W3C understands not only images, but also pre-recorded or live video-only and sound-only files, as well as form input elements such as radio or image-type buttons, check boxes and text fields. Nevertheless, in this thesis, we will use the term "text alternative" to exclusively refer to text equivalent for images.

While this accessibility recommendation is crucial for screen reader users to perceive images on the Web, providing text alternatives also benefits, as highlighted by Hickson et al. (2014, sec. 4.8.4.1.1), any other user who might:

---

[4] Arguably, this assertion could also apply in the case of photographs, illustrations, drawings, graphs or charts. However, this is much less frequent and chances are that only in poorly designed websites would one find these images as linked elements.

[5] See Chapter 2 (section 2.2.3.1) for a detailed explanation about the structure of WCAG 2.0.

[6] The exceptions included under this success criterion exclusively referring to images (and not to other forms of non-text content) are based on the communicative purpose of the image and will be discussed more comprehensively in section 4.3.2.

a) Have a very slow connection and need to browse a website with images disabled.
b) Have a cognitive impairment and use text to speech software.
c) Be using a text-only browser.
d) Be listening to the page being read out by a voice web browser.
e) Have images disabled to save on download costs.
f) Have problems loading images or be browsing a website where the source of an image is wrong.

Figure 3.1 shows an example of a web page with a graphic-based design. The web region captured in the figure is mainly composed of images of text with a functional value that, through an embedded link, enable users to visit different sections of the website of the University of Salamanca, as well as to switch between the English and Portuguese language versions of the site (see small square icons in blue and green at the top). Figure 3.2 illustrates how this same web page would look in the situations previously described in a), c), e) and f) if every image had a text alternative. Should a screen reader user be browsing this page, the text alternatives replacing the images would be read out loud. The same would occur in scenarios b) and d). Regardless of the situation the user may be in when landing on this web page, the situation of not being able to visualise images on the screen would not prevent them from having access to the information offered on the website.

Figure 3.1. Homepage of the website of the University of Salamanca with image-loading turned on

For users to perceive text alternatives for images on the Web, irrespective of the sensory modality they rely on to do so (visual, auditory, tactile), web developers and content authors have to go through two important steps before: first, choose which HTML technique is the most convenient according to existing W3C standards to offer this functionality; and second, write an appropriate text alternative to act as a replacement of the image, based on its communicative value and the surrounding context. Both aspects shall be discussed in sections 3.3.1 and 3.3.2 respectively.

Figure 3.2. Website shown in Figure 3.1 displayed in a graphical browser with image-loading turned off

3.3.1 HTML technical requirements

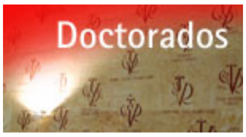
In what follows, we describe how images can be inserted within an HTML document, as well as the different mechanisms that the W3C offers to describe them. Once these have been introduced, we shall explain why we decided to focus our research on the use of the alt attribute as the main technique to render images accessible to screen reader users.

3.3.1.1 HTML basics: the <img> element and the alt, title and longdesc attributes

Most images in HTML documents are introduced through the <img> element. The most commonly adopted technique to meet SC 1.1.1 of the WCAG 2.0 is to add to this element a text alternative using the alt attribute (Cooper et al. 2015a, sec. HTML and XHTML Techniques: H37). According to the last HTML specification[7], "the value of the alt attribute provides equivalent content for those who cannot process images or who have image loading disabled (i.e. it is the <img> element's fallback content: content that is to be used when the external resource cannot be used)" (Pieters et al. 2015). Figure 3.3 shows an HTML code excerpt

[7] At the time of this thesis writing, the status of the HTML 5.1 Specification was "W3C working draft", dating from 8th October 2015. The last official HTML specification is HTML 5 (Hickson et al. 2014) and was approved as a W3C Recommendation on 28th October 2014.

illustrating how a text alternative is added to an image element (`<img>`) on a web document.




```
<a href="http://www.usal.es/webusal/usal_doctorado_
repositorio" class="mc0103">
<img src="Universidad-de-Salamanca_files/doctorados-
peq.jpg" alt="Doctorados"></a>
```

Figure 3.3. HTML code excerpt corresponding to one of the images of texts depicted in Figure 3.1

In both HTML 4 and 5 W3C Recommendations, it is stated that the `alt` attribute must be specified for every `<img>` and `<area>` element contained in a HTML document, while it is optional for elements `input` or `applet`. Nevertheless, unlike other `<img>` attributes such as `src` —which indicates the location of the resource—, its presence is dispensable for an image to be correctly visualised on the browser when images are enabled. For instance, the web page shown in Figure 3.1 would look exactly the same if images carried an `alt` attribute or not. Yet, the absence of an `alt` attribute can render the information conveyed through images inaccessible for sighted users when visual content cannot be displayed. In such a scenario, most browsers would show a bounding box as an image replacement when the alt text is missing, as depicted in Figure 3.4.

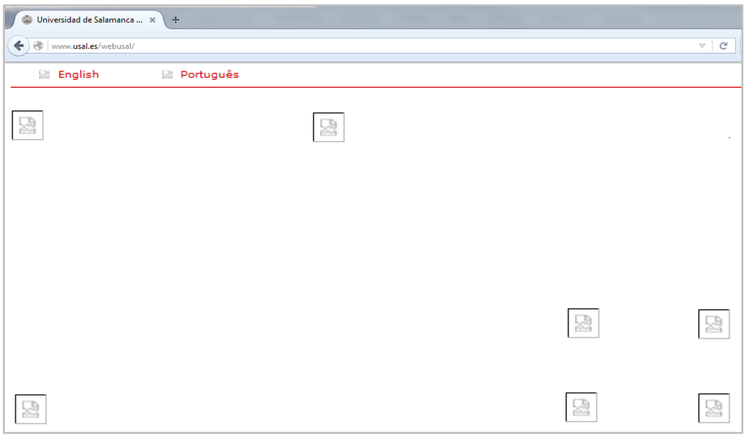


Figure 3.4. Website shown in Figure 3.1 with images lacking an `alt` attribute, as it would be displayed in a graphical browser with image-loading turned-off

Not using the `alt` attribute also degrades the web browsing experience of screen reader users, such as the visually-impaired. When a screen reader gives focus to an `<img>` element on a web page, the software announces the presence of an image to the user. Most popular screen readers such as JAWS or NVDA do so by reading out the word "graphic". If the `<img>` element contains an alt text, the software reads it immediately after that word. When an `alt` attribute is not present, screen readers may (i) pronounce the word "blank" or (ii) look for any other supporting information and announce it to the user, such as the content of the `src` attribute. Pernice and Nielsen (2001, 45) and Connor (2012, 176) suggest that, if the file name is informative, it might help the screen reader user to get an idea of what the purpose of the image is (e.g. logout.png). Nevertheless, the value of the `src` attribute includes the full file location path and the file name is only read out at the end. In addition, the screen reader output can be meaningless and even cause frustration to the user when the `src` value contains non-alphanumeric characters (Thatcher et al. 2006, 137–137). For instance, an image with no `alt` attribute located in `src="images0132_small/moodle-files/logout.png"` would be announced by JAWS as "graphic images 0 1 3 2 underscore small slash moodle dash files slash logout".

In the absence of an `alt` attribute, screen readers may also try to look for a `title` attribute if the user requests it, although this functionality varies across screen readers and screen reader versions (Clark 2002, 67). The HTML5 Specification indicates that this attribute "represents advisory information" for the element it modifies, and suggests that it could be used to include the image credit or simply an image description within an `<img>` element (Hickson et al. 2014). Although the W3C discourages its use because many user agents do not expose the attribute in an accessible manner, the definition provided in the Specification is problematic, since it could lead web developers or content authors to use the `title` attribute alone as a means to provide a text alternative.[8]

While it has been claimed that the `title` attribute can be used to expand on an alt text, inform about helpful but optional details about an image (e.g. image size) or to warn about potential browser behaviour (e.g. "zoomed version opens in a new window") (Clark 2002), Faulkner (2012) recommends not to use it for any text that all users should have access to. This might be derived from the fact that the `title` attribute is displayed in some browsers in the form of a tooltip when hovering over the image (or any other element it may be associated with, such as a link). The W3C understands that this constitutes an accessibility barrier, given that requiring a pointing device such as a mouse to cause a tooltip to appear excludes keyboard-only users and touch-only users, such as anyone with a modern phone or tablet (Hickson et al. 2014).[9]

[8] In an article posted on The Paciello Group website, Faulkner (2015) affirms that in the final version of new HTML 5.1 specification, currently a working draft, the W3C intends to include an statement indicating that the use of the `title` attribute in place of an `alt` attribute on an `<img>` element or for an image caption is prohibited. https://www.paciellogroup.com/blog/2010/11/using-the-html-title-attribute/ Last access: 10th January 2016.

[9] The `title` attribute, like the `alt` attribute, can be used within HTML elements other than `<img>`. Nevertheless, discussing its applicability in such contexts is out of the scope of this thesis.

Over the last years, the use of the longdesc attribute has been equally controversial, even if it is listed as a recommended technique to meet WCAG 2.0 SC 1.1.1 (Cooper et al. 2015a, sec. HTML and XHTML Techniques: H45). As in the case of the alt and the title attributes, longdesc serves to provide information about an image in textual form. The difference relies in that this attribute, which can be only inserted within an <img> element, specifies a link to a long description of the image, but it does not contain the description text itself. According to the HTML4 Specification, the longdesc content should supplement the short description provided using the alt attribute (Raggett et al. 1999), which is restricted in terms of length (see section 3.3.2). However, longdesc is no longer listed as a valid <img> attribute within the HTML5 Specification. Instead, the W3C suggests including a link below any image that needs to be further explained. The link would take the user to a longer description available either within the same page (see Figure 3.5[10], option a) or in a separate web document (see Figure 3.5, option b).

```
[option a]
<img src="teix-aytart.jpeg"
alt="Abstract painting by Galician art-
ist Mingos Teixeira entitled Water and
Earth.">
  <p><a href="#des">Description of the
painting</a>.</p>

[option b]
<img src="teix-aytart.jpeg"
alt="Abstract painting by Mingos
Teixeira entitled Water and Earth.">
  <p><a href="http://bit.ly/xYz">About
this painting</a>.</p>
```

```
[option c]
<img src="teix-aytart.jpeg" alt="Abstract painting by Mingos Teixeira
entitled Water and Earth." longdesc="#des"><p id="des">The painting is
divided into two sections that feature different textures and...</p>

[option d]
<img src="teix-aytart.jpeg" alt="Abstract painting by Mingos Teixeira
entitled Water and Earth." longdesc="http://bit.ly/xYz">
```

Figure 3.5. HTML code excerpts illustrating how provide a long description for an image (e.g. an abstract painting)

Several reasons might have motivated this decision. On one hand, the longdesc attribute was rarely implemented. Connor (2012, 190) points out that it was little-used outside of academia. This assertion is supported by the results obtained from a large-scale web accessibility study carried out by Hanson and Richards, who reported that was never used in any of the 952 government websites

[10] The image of this painting, as well as those shown in Figure 3.7, has been used in this chapter with the consent of its author (all rights reserved to (c) Mingos Teixeira).

they examined over a 14-year period (2013, 2:14). Similarly, they found evidence that its purpose was not correctly understood on multiple occasions, with it being used to give the URL of the image source (ibid). On the other hand, the `longdesc` attribute might have been dismissed due to its poor support by user agents. Numerous authors have indicated that it was not strictly well supported by most popular browsers (Paciello 2000, 100; Clark 2002, 68; Connor 2012, 190) and a recent study by WebAIM revealed that this was also the case for multiple screen readers, such as JAWS, NVDA or Windows-Eyes (WebAIM 2015a). This study took place some months after the W3C had published a new specification reintroducing the "`longdesc` attribute (based on the `longdesc` attribute of HTML 4) to link descriptions to images in HTML5 content" (Nevile and Sadecki 2015). If the HTML 5.1 Specification was to include this extension as normative in the near future, a `longdesc` attribute may be introduced within an `<img>` as shown in Figure 3.5, options (c) or (d).

3.3.1.2 Other HTML image-related elements and mechanisms to provide image descriptions

As mentioned earlier, the `alt` attribute must also be specified for `<area>` elements, needed in image maps. These type of images can be introduced in a HTML document by using the `usemap` attribute within an `<img>` element. The value of this attribute often references a `<map>` element, which specifies areas of the image that will be considered as hotspots, using `<area>` elements that have `href` attributes. In these cases, the `<img>` is considered to be interactive content (Pieters et al. 2015). Image maps are often called "client-side" because the browser (the client) must figure out how to handle the result of a click on the image, that is, it needs to understand whether or not the coordinates of that click fell inside one of the regions specified as an `<area>` of the client-side map (Thatcher et al. 2006, 201).

For client-side image maps to be accessible, both the `<img>` and the different `<area>` elements must contain an `alt` attribute. Client-side image maps are often images of text that would not otherwise be accessible if this attribute was not specified. This is the case of the example shown in Figure 3.6. The image shows the list of faculties of the University of Geneva and their corresponding colours. The HTML code excerpt below refers to the content highlighted within a blue square to the left. Screen readers such as JAWS and Windows Eyes treat image map hotspots with alt text as if they were links (Thatcher et al. 2006, 140). If they gave focus to this image map, they would (ii) read the alt text of the `<img>` element first, and then (ii) announce the presence of three links by reading the corresponding text alternatives as well.


```
<img src="/unige-faculties.jpeg" alt="List of faculties of the Uni-
versity of Geneva and their representing colours" usemap="#facul-
ties"/><map name="faculties">
  <area shape="rect" coords="3,33,76,72"
href="http://www.unige.ch/sciences/index.html" alt="Faculty of Sci-
ences (green)"/>
   <area shape="rect" coords="5,78,88,117"
href="http://www.unige.ch/medecine/index.html" alt="Faculty of Medi-
cine (magenta)"/>
  <area shape="rect" coords="5,153,59,172"
href="http://www.unige.ch/lettres/fr/" alt="Faculty of Humanities
(blue)"/></map>
```

Figure 3.6. HTML code excerpt illustrating the combined use of <img>, <map> and <area> elements

Another element that, like <map> and <area>, is related to the presentation of graphical content within a HTML document is the <figure> element. Introduced in HTML5 for better structural purposes, it "represents some flow content that is self-contained (like a complete sentence) and is typically referenced as a single unit from the main flow of the document" (Hickson et al. 2014). In other words, this element acts as a container within which it is possible to include other elements (e.g. <img>, <table>, <video>, <blockquote>) that will then be treated as a unique block of information. Some of its advantages include that it can contain more than one element, and that it can be directly associated with a caption to describe the unit of information it contains through its child <figcaption> element. Figure 3.7 illustrates how these two HTML5 elements can be used to nest a series of complementary images so that they are semantically marked up with a single caption. Notice that each image also includes an individual alt text.

While it has been suggested that, if <figcaption> is present when <figure> only contains one single image, the `alt` attribute may be omitted to avoid redundancy and duplication of content (WebAIM 2015b), the W3C indicates that such cases are to be kept to an absolute minimum (Pieters et al. 2015). Concretely, the HTML5 Specification states that "[i]f there is even the slightest possibility of the author having the ability to provide real alternative text, then it would not be acceptable to omit the `alt` attribute." In addition, the use of the <figure> element —without which <figcaption> cannot be introduced— is not recommended for every type of image (Hickson et al. 2014, sec. 4.4.11). The <figure> element should only be used for "content that is part of the surrounding flow" (ibid). In the case of images, this means that it would be useful for annotating descriptive images —as per our definition in section 3.2— such as illustrations, photos or diagrams. However, it would not be not appropriate for decorative images and even certain small-size functional images such as icons or logos. Finally, it should be mentioned that <figcaption> is not supported by the latest version (11) of Internet Explorer (IE) (The Paciello Group 2015), which is the most popular browser among screen reader users.

With the advent of HTML5, it also became possible to add accessibility information to HTML elements using the Accessible Rich Internet Applications specification (WAI-ARIA). This specification “provides an ontology of roles, states, and properties that define accessible user interface elements and can be used to improve the accessibility and interoperability of web content and applications” (Diggs et al. 2015). Regarding WCAG 2.0 SC 1.1.1, the W3C lists the use of two aria-* properties as sufficient techniques to provide a text alternative for non-text content: aria-labelledby and aria-describedby. The former associates an element with text that is visible elsewhere on the page by using an ID reference value that matches the ID attribute of the labelling element; yet it is only recommended as an appropriate technique for elements which need a text alternative but do not support the alt attribute; e. g. <a> or <div> (Cooper et al. 2015a, sec. ARIA Techniques, ARIA10). However, the latter (aria-describedby) can be safely used in conjunction with it.

```
<figure><img src="/teix-aytart.jpeg" alt="Whater and Earth">
  <img src="/teix-coordart" alt="Forbidden evasion">
  <img src="/teix-geoart.jpg" alt="Intertwined coordinates">
  <figcaption>Abstract paintings from Galician artist Mingos Teixeira
in the Metamorphosis series.</figcaption></figure>
```

Figure 3.7. HTML code excerpt illustrating how to use the elements <figure> and <figcaption> to nest semantically-related images

The W3C indicates that this WAI-ARIA property is similar to longdesc in that both are useful for providing additional information —complementing the alt text— to help users understand complex images (Cooper et al. 2015a, sec. ARIA Techniques, ARIA15). Nevertheless, in contrast to longdesc, the description referenced must appear within the same page (see example in Figure 3.8). This is often seen as an advantage, since every user would benefit from it. Additionally, it has been claimed that aria-describedby is currently better supported than longdesc (ibid). The combined used of the alt attribute and aria-describedby can, therefore, represent a convenient compromise to render images accessible, given that at least one alternative text would be provided to screen reader users either way, regardless of the compatibility of the user agents they might be using —older ones might only have access to the alt attribute content, while newer ones that support both mechanisms would offer a richer textual equivalent (Connor 2012, 184).

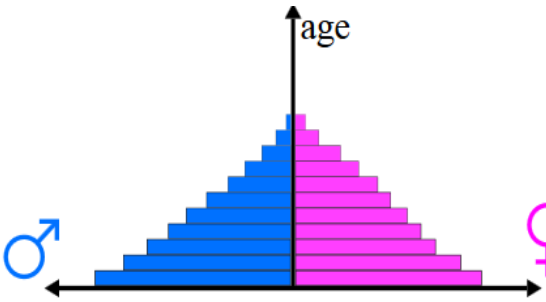
```
<img src="teix-aytart.jpeg"
   alt="Abstract painting by Galician
   artist Mingos Teixeira entitled Water
   and Earth."
   aria-describedby="waterandearth">

<p id="waterandearth">Water and Earth
(canvas, 161 x 130, acrylic) is the
first abstract painting in the artist's
Metamorphosis series. It combines
painting knife techniques with...</p>
```

Figure 3.8. HTML code excerpt illustrating the use of aria-describedby attribute

The last element from the HTML5 Specification worth highlighting in this section is <svg>, which allows to include Scalable Vector Graphics (SVG) directly in the HTML document. SVG is a language for describing two-dimensional graphics in XML (Dahlström et al. 2011). Since the particularities of vector graphics will be further discussed in section 3.5.2, we will focus here on SVG language mechanisms to provide textual alternatives. Two elements can be specified for this purpose: <title> and <desc>. The former can be used in simple graphics to introduce a short description (which may be rendered by web browsers as a tooltip), while the latter is recommended for more complex graphics, since it offers the possibility of providing a longer description (Andersen 2010). These two SVG elements play a similar role as the pair alt - longdesc (or aria-describedby) in the case of the <img> element.

From an accessibility perspective, one of the main advantage of SVG images with respect to other file formats is precisely the possibility of embedding alt text within the file XML-based encoding (Donney et al. 2008). In this sense, one would expect that this would facilitate the access to this information by screen reader users. Nevertheless, it has been claimed that SVG1.1 accessibility support is limited both in browsers and screen readers (Andersen 2010; Watson 2013). This might explain why, even if the SVG format became a W3C recommendation in 2003, studies have shown poor SVG adoption rates on the Web (Chen and Harper 2008, 25). In an attempt to find an alternative solution to render SVG graphics more accessible to screen reader users, web accessibility experts have suggested the use ARIA-based techniques (Watson 2013), as illustrated in Figure 3.9.[11]




```
<body> ...<svg
xmlns=http://www.w3.org/2000/svg
role="img" aria-labelledby="title
desc">
   <title id="title">Figure 3.9.
Population pyramid</title>
   <desc id="desc">Two back-to-back
bar graphs, with the population
plotted on the X-axis and age on the
Y-axis, showing...</desc>
... </svg>
... </body>
```

Figure 3.9. HTML code excerpt illustrating how to render a SVG graphic accessible

The first recommendation is to use the `role` attribute with the value `img`, which usually serves as a container for a collection of elements that form an image (Diggs et al. 2015). When introduced within a `<svg>` element, it ensures that the element is indeed identified as a graphic. The second technique proposed is to use the `aria-labelledby` attribute referencing the `id` values of the `<title>` and `<desc>` elements, so that their content can be announced to screen reader users. Watson (2013) argues that `aria-labelledby` is preferred in the case of SVG over `aria-describedby` due to support reasons and because the practical output is the same. It is expected that with the new SVG 2 specification (Andronikos et al. 2015), currently in working progress, clearer recommendations regarding the use of an ARIA enhanced SVG will be presented.

3.3.1.3 Particular techniques for decorative images

In the sections above we have described different mechanisms for authors to provide text alternatives or longer descriptions for images which need to be identified as such by assistive technology (e.g. screen readers). This would cover images whose meaning is relevant for a successful user interaction, like those which carry a descriptive or functional communicative value. However, the accessibilisation of decorative images follows a different approach. Rendering images purely used for aesthetical purposes identifiable by screen readers can cause discomfort among users, since listening to uninformative text alternatives can slow down their browsing experience or divert them from the main content of the web page. Hence, WCAG 2.0 SC 1.1.1 indicates that decorative images should be "implemented in a way that it can be ignored by assistive technology" (Caldwell et al. 2008).

[11] This image was made available for public use under the Creative Commons CC0 1.0 Universal Public Domain Dedication. Source: https://goo.gl/eGboV1 Last access: 14th January 2016.

The W3C recommends two main techniques to achieve this. The first one is to use an empty `alt` attribute (`alt=""`), often also referred to as “null” `alt` attribute or zero-length alt text (Cooper et al. 2015a, sec. HTML and XHTML Techniques, H67). Screen readers ignore images that contain a zero-length alt text, regardless of the image position or size, not announcing their presence to the user. While some user agents may also interpret images with an `alt` value representing a space character as decorative (`alt=" "`), the general —and safest— rule to follow is to avoid introducing a space within the `alt` attribute (Korpela 2012; Connor 2012).

The second technique suggested in Cooper et al. (2015a) is to include decorative images in Cascading Style Sheets (CSS) (CSS Techniques, C9). CSS is a mechanism for adding style (e.g. fonts, colours, spacing) to web documents. Screen readers can ignore or turn off CSS at the user's request. When this occurs, all the information contained therein is not presented to the user. An image that needs to be hidden from screen reader users, such as a spacer.gif intended to position content or simply a background image to embellish the web page, can be introduced in the CSS file through the `background-image` property. The declaration would be as follows:

```
p {background-image:url('decorative-graph.png');}
```

where `p` would represent the element to which you want to apply the image (Connor 2012).

While the W3C does not specifically propose the use of WAI-ARIA techniques to render decorative images accessible, certain authors suggest that another way to hide an image from a screen reader may be to use `role="presentation"` (Connor 2012, 179). This WAI-ARIA role removes the semantics from the element it is on (Faulkner et al. 2015), so if used within an `<img>` element, it would remove its role of graphic. Yet, (Connor 2012, 179) indicates that the other two approaches (techniques H67 and C9) are better supported by current and older user agents. In addition, according to the last HTML5 specification, the presence of `role="presentation"` does not make missing `alt` conforming, that is, HTML5 validators would flag an `<img>` element without an `alt` attribute as an error, even if it contains the ARIA role (Faulkner 2011).

3.3.1.4 Summary of techniques for providing text alternatives for images

Undoubtedly, the W3C has made significant efforts to introduce, through different specifications, a wide variety of mechanisms to improve the accessibility of images on the Web. Table 3.2 provides an overview of those that have been described in this chapter. Still, the extent to which these end up being successful remains strongly dependent on the support offered by user agents. In this thesis, we will focus exclusively on the `alt` attribute of the `<img>` element as the main mechanism to provide short text alternative for images.

Table 3.2. Overview of HTML elements, attributes and ARIA-based mechanisms for providing text alternatives and long descriptions for images

Element	**Attribute/Property**	**Specification**	**Support**	**Mandatory**
`<img>`				
	`alt` (empty/non-empty)	HTML4 & 5	Full	Yes
	`title`	HTML4 & 5	Partial	No
	`longdesc`	HTML4 (5 unclear)	Low	No
`<area>`	`alt`	HTML4 & 5	Full	Yes
`<figure>`				
`<figcaption>`	n/a	HMTL5	Partial	No
`<svg>`				
`<title>`	n/a	HMTL5, SVG 1.1	Low	No
`<desc>`	n/a	HMTL5, SVG 1.1	Low	No
	`aria-describedby`	ARIA 1.1	Partial	n/a
	`aria-labelledby`	ARIA 1.1	Partial	n/a
	`role="img"`	ARIA 1.1	Partial	n/a
	`role="presentation"`	ARIA 1.1	Partial	n/a

This decision was motivated by the fact that, to the best of our knowledge, the use of the `alt` attribute is the only recommended technique that guarantees a robust support. It was maintained in the new HTML5 specification, and it is considered by the W3C as a very important accessibility attribute (Hickson et al. 2014; Pieters et al. 2015). In addition, its use is mandatory —with very few exceptions—[12] unlike other complementary attributes such as `title` or `longdesc`, whose implementation is still controversial. On the other hand, the use of the `<img>` element is the most widely extended mechanism to introduce static graphical content within an HTML document. As previously discussed, the `<area>` element is becoming less popular on the Web, the SVG format does not enjoy fully accessibility support and the `<figure>` element is not appropriate for every type of image.

As far as WAI-ARIA techniques is concerned, authors should not solely rely on this technology to render images accessible. Whereas most of the latest web browsers and screen readers can already interpret WAI-ARIA roles and properties correctly, the specification dictates that the use of native HTML elements or attributes should be prioritised over ARIA techniques whenever possible (Diggs et al. 2015). Furthermore, the `aria-labelledby` and `aria-describedby` properties refer to text visible on the screen. While this might be observed as an advantage for sighted users as well, there may be graphic-intense web pages where authors

[12] For instance, when a text alternative is not available at the time of publication and the `<figure>` and `<figcaption>` elements need to be used instead (Hickson et al. 2014, sec. 4.7.1.1.16)

might prefer to hide text descriptions due to space restrictions or simply for aesthetical purposes.

3.3.2 Writing appropriate text alternatives

Adding an `alt` attribute within an `<img>` element to enhance image accessibility might seem like a trivial task that can be easily accomplished. Nevertheless, providing it with an appropriate value may not be as straightforward. In fact, it has been claimed that, despite being the first principle of web accessibility, it is one of the most difficult to properly implement (WebAIM 2015b), with some authors considering an art in itself (Bigham et al. 2006; Tang 2012). In an attempt to facilitate this task for web developers and content authors, scholars and institutional bodies have devoted considerable efforts to provide guidance on two main aspects associated with the generation of text alternatives: their *composition*, that is, what information to include in the alt text (content perspective), and their *formulation*, i.e. how to present the information (language perspective). As announced in the Introduction of this thesis, we are mainly concerned with the latter and how it is addressed in web accessibility evaluation tools (see section 3.4 of the present chapter). However, given that alt text formulation is closely related to the former and both are equally important to achieve image accessibility, we consider it pertinent to address them simultaneously when reviewing the existing literature on how to produce appropriate text alternatives.

According to the W3C, "[a]uthoring useful `alt` attribute content requires the author to carefully consider the context in which the image appears and the function that image may have in that context" (Pieters et al. 2015). In the literature, two types of support can be found to help authors in such endeavour: (i) guidance on how to analyse the content of an image which is grounded on theoretical principles of visual information representation, and (ii) experience-based practical guidance to identify the purpose of an image by means of cues —usually introduced in the form of queries— about its potential content and context.

To date, the most complete protocol created to ensure that all the important information conveyed through an image is identified before formulating a text alternative is the ISO Technical Specification ISO/IEC TS 20071-11:2012 (International Organization for Standardization 2012b), which is grounded on the doctoral work of Tang (2012). Tang and Carter (2011, 394) argue that, once the purpose of the image has been understood by the author, they may need further guidance on (i) how identify which is the most crucial information that needs to be communicated to the user, (ii) how this information should be rated, and (iii) how to transform it into comprehensible prose. Through the investigation of research areas including library science, image indexing, captioning, audio description, art description and tactile representation, Tang (2012) designed a procedure aimed primarily at covering the two first needs declared by Tang and Carter. It consists of six steps:

Step 1. Identify the purpose that the image served within the document.
Step 2. Identify the image components within the image.
Step 3. Identify the image (or image component) content.
Step 4. Elaborate on the image (or image component) content.
Step 5. Organise the identified information into text alternatives.
Step 6. Evaluate the resulting text alternatives.

As part of the procedure, Tang (2012) proposes various techniques to extract information from the image, such as answering to the seven *Wh-* questions *−Why, What, Who, Where, When, How Much,* and *How−;* or classifying the resulting pieces of information according to different levels of importance (essential, significant, helpful, not important). Once all the data has been collected, the procedure suggests to remove redundant information and organise it in a logical, readable order (ideally according to the aforementioned degrees of relevance). Since the intent of the procedure is to identify as much raw material about the image as possible before creating the text alternatives (Tang 2012, 28), few references are made to the formulation of alt texts. The only recommendations made are (i) to be succinct and concise, and (ii) to make use of vivid words.

Despite the major contribution made by Tang with the extension of the ISO/IEC 20071, this protocol has several drawbacks. First, its content is independent of the choice of implementation mechanism or of electronic document type (International Organization for Standardization 2012b, 1). This means that it can be used to produce text alternatives that would be later included within the `alt` attribute in a web document, but also text equivalents for images contained in other document formats, such as PDF. The requirements to introduce alt texts might differ across document types (e.g. different markup languages) and authors would thus need to rely on complementary resources to finish the task. Second, the protocol seems more suitable for images with a high information load, but we believe that not all the steps (e.g. 3, 4 or 5) would be needed in the case of functional images on Web. Third, it might be too complex and time-consuming to be used on a daily basis. Through different experiments, Tang (2012) found that it takes, on average, three minutes to generate a text alternative without any guidance support, 15 minutes with the protocol in a document format, and 30 minutes when the procedure is implemented by means of TATI (Text alternatives Tool for Images), a guidance tool that takes authors through every step of the protocol (USERLab 2012). In this sense, it is more likely that actors involved in the web development cycle would prefer to have more pragmatic guidance at their disposal.

Individuals, scholars and institutions with extensive web accessibility experience have built over recent years different resources to help authors understand the nature of the images contained within an HTML document. These are either presented in the form of (i) decision trees, such as the ones created by Carreras Montoto (2014) or the W3C (Eggert and Abou-Zahra 2014c), or simply as (ii) a set of individual recommendations, often organised around image types, which may also include image and `alt` attribute samples. Together, these guidance documents represent a wide body of easy-to-use support tools where practical information about the *composition* of the text alternative is sometimes combined with hints about the *formulation* of the alt text.

In the sections that follow, we aim to summarise these sets of recommendations, highlighting which language-oriented tips have been put forward both from academia and web-oriented institutions. The different sources consulted in the literature provide recommendations that can be either applied to all kinds of images or those that are specific to certain types of images. In the context of this thesis, we will review them in a similar order: general recommendations (3.3.2.1) and recommendations based on image type (3.3.2.2). However, for ease of presentation, guidance within this last section will be organised around two of the main images categories presented in section 3.2: descriptive and functional. Decorative images will not be further addressed, given that we have already introduced the technical requirements that need to be met to render them accessible, the underlying principle being precisely to avoid their description.

3.3.2.1 General recommendations

In the HTML 5 specification, it is stated that "there is no single 'right' or 'correct' piece of alt text for any particular image", given that what the author includes in the `alt` attribute value is both subjective to the context where the image is used and the page author's writing style (Hickson et al. 2014). It is generally understood that an appropriate alt text will communicate the content or purpose of the image, but not its appearance (Craven 2006), as we have previously announced in this chapter. The W3C supports this assertion by specifying that "every image has a reason for being on a page, because it provides useful information, performs a function, labels an interactive element, enhances aesthetics or is purely decorative." (Pieters et al. 2015) Either way, as regards the *composition* of the alt text, the general advice is to be **accurate** and **equivalent** in presenting the same content and function of the image, as well as **not to be redundant** –for instance, by providing through the `alt` attribute the same information presented in the text that surrounds the image– (WebAIM 2015b).

As far as the *formulation* of the text alternative is concerned, the literature suggests authors to be **succinct** (Slatin 2001; Pernice and Nielsen 2001; Craven 2006; WebAIM 2015b). Succinctness implies, for example, avoiding superfluous information in the text alternative, such as redundant expressions like "graphic of" or "image of" (Craven 2006; Nyong 2009; Eggert and Abou-Zahra 2014c). The reasons provided are that sighted users would know this information already, as screen readers announce the presence of an image. Exceptions to this rule may be accepted when the fact that an image is a photograph, a picture or an illustration is important content to be conveyed to the user (WebAIM 2015b). Although the W3C indicates that it is best to avoid the more generic use of these terms (Eggert and Abou-Zahra 2014c).

Related to this recommendation is the order in which information is placed in the alt text and the language complexity level. After a series of interviews with blind users, Petrie et al. (2005) found that the **order** of the words in a text alternative was observed as crucial, with general agreement from the interviewees that the most important information should (i) be placed at the beginning and (ii) be in **simple language**. These findings are also supported by the W3C, that suggests to prioritise information within the text alternative (Eggert and Abou-Zahra 2014c), and by Nyong (2009), who advocates for the "use of the clearest and simplest language appropriate for the website content". Regarding language simplicity, Korpela (2012) adds that it is best to use normal prose instead of a "telegram style", to avoid the use of acronyms and to be extremely careful concerning spelling errors, since their presence may lead to a degraded performance of speech synthesizers. Additionally, Tercedor Sánchez and Prieto Velasco (2009, 84) recommend using the present tense, the active voice, and simple syntax.

The major source of discrepancies within the literature regarding alt text formulation is the **length.** The W3C HTML specifications and other supporting documents do not give any indications about this aspect of text alternatives, so the theoretical upper limit would be over 65,000 characters (Slatin 2001, 78). However, most authors suggest considering user agents' restrictions when deciding how long an alt text should be. Browsers may not display very long text alternatives when image loading is turned off due to space restrictions, and assistive technologies such as screen readers may impose limitations regarding the number of characters that will be read to the user. Clark (2002, 63) states that, by convention, authors should limit alt texts to 1,024 characters (1K) or less. Korpela (2012) recommends to write the `alt` attribute so that its value is at most 50 characters long. Slatin (2001) proposes a maximum of 150 characters instead, indicating that JAWS screen reader can cause problems if the text alternative exceeds this length, e.g. the software could end up dividing the alt text into distinct chunks, as if they belonged to different images. Irrespective of the number of characters, studies have shown that, overall, users prefer more information rather than less, although they believe that there should always be a balance of quality versus quantity (Petrie et al. 2005).

Questions of **punctuation** have also been addressed in the literature. Tercedor Sánchez and Prieto Velasco (2009) suggest that a text alternative should always end with a full-stop, so that screen readers can adapt the intonation accordingly. They also recommend starting the text alternative with a capital letter, as in any other beginning of sentence (ibid). In addition, Korpela (2012) proposes the use of brackets to distinguish image descriptions from replacements (i.e. simple text alternatives), as well as the use of parenthesis to remark that the text contained in it expresses some of the content. For instance, if a complex chart is shown on a web page, the alt text would be `alt="[Pie chart of our sales in 2002 (showing small increases)]"`, while the actual description of its content would be either presented in the surrounding text or within a complementary attribute (e.g. `longdesc`, `aria-describedby`, etc.). While it might seem like a good strategy, no evidence exists that indicates that end users would understand the meaning of this writing convention when spelled out. In the same vein, the announcement of brackets and parentheses could be seen as frustrating and time-consuming. In relation to this last aspect, the literature also recommends to avoid ASCII graphics and HTML character entities, given that it is unclear how user agents, especially screen readers, would respond and announce the `alt` content to the user (Korpela 2012). In the case of emoticons, which are also considered as ASCII art too, it is recommended to describe its meaning (e.g. for a smiley emoticon, an appropriate alt text could be `alt="(I feel happy)"` or `alt="(I am joking)"`).

Finally, it is worth mentioning that multiple sources have put emphasis on what would be considered as a 'wrong' or 'incorrect' `alt` attribute value, rather than on best practices to write appropriate ones. Among the examples of **uninformative content** that have been highlighted in the literature (Clark 2002; Petrie et al. 2005; Craven 2006; Korpela 2012), the most frequently listed are (i) file names (e.g. `alt="myprofilepic.jpg"`); (ii) file formats (e.g. `alt="GIF"`); and (iii) information about file size (e.g. `alt="(23 Kb)"`). Other examples include meaningless placeholders such as “short description of image”, “alt text”, “loading image” or “turn images on”, and vague one-word alt texts such as “image”, “home” or “text”. According to some web studies, these are often the most common alt texts found on the Web (Craven 2006). This might be due to the adoption of unreliable alt text proposals made by authoring tools, including Content Management Systems (CMS). To encourage the use of alternative text, some of these tools have built-in mechanisms for alternative text prompting (Paciello 2000, 77). Some may enforce the insertion of an `alt` attribute and automatically introduce placeholders that remain unchanged when the web page or site is published.

3.3.2.2 Recommendations based on image type

Web developers and content professionals can find guidance on how to compose a text alternative according to the image type directly in the HTML 5 specification (Hickson et al. 2014, sec. 4.8.4.1 *Requirements for providing text to act as an alternative for images*). The W3C provides code and image examples for different scenarios, from which it is possible to infer certain alt text formulation best practices. Other than that, explicit language-related recommendations based on the purpose or communicative value of the image are rather scarce if not non-existent within the document. Complete guidance in this regard is also lacking within academic sources. However, through a comprehensive review of the literature, we have been able to retrieve some useful advice concerning specific image types, such as photographs, linked images, icons and logos.

In what follows, we will combine the information gathered from these sources to summarise the existing recommendations for the composition and formulation of alt texts for descriptive and functional images.

Descriptive images

With regard to the *composition* of alt texts for descriptive images such as **drawings or photographs**, some manuals indicate that authors should not aim at capturing every detail from the image, but at textually conveying its spirit (Connor 2012, 185). Others prescribe, however, that "if there are certain elements of the image that are important, these should be spelled out" (Cunningham 2012, 8). In this sense, studies have revealed that blind users believe that the following elements would require description in the majority of cases: objects, buildings and people shown in the image; what is happening in the image; colours used in the image; location depicted in the image; and the emotion or atmosphere transmitted through the image (Petrie et al. 2005). Tercedor Sánchez and Prieto Velasco (2009, 82–83), as well as Clark (2002, 72–73), point out that colours, shapes and textures may also seek to trigger a special response or emotion on the user, so they should be referred to in the alt when appropriate. They argue that blind and visually-impaired people understand the social significance of colours and are able to picture both animated and non-animated images in their minds when good descriptions are provided, hence the relevance of mentioning as well, if possible, the point of view or position of the camera or observer, even implicitly (ibid).

Tips as regards the *formulation* of such text alternatives include the use of "tight, evocative adjectives" when referring of colours, shapes, textures or connotative concepts conveyed through the image (Clark 2002; Tercedor Sánchez and Prieto Velasco 2009). In addition, Tercedor et al. (2009) provide some examples on how to explicitly refer to the location of elements in the image ("in the centre", "in the top left hand area", "to the right") and suggest minimising orality markers, such as "we can see..." or "in the image we find...".

Specific advice for images other than drawings, photographs or paintings is limited. The general rule concerning the composition of alt texts for **images of text** is to provide the same text in the `alt` attribute as is in the image (Hickson et al. 2014). For **graphical representations** (charts, diagrams, graphs), the HTML 5 specification indicates that "in cases where the text alternative is lengthy,

[... authors should] provide a brief description or label using the `alt` attribute, and an associated text alternative", for instance, using a link that points to a longer description (ibid). In the examples provided, the type of graphic is always announced at the beginning of the alt text, followed by a colon, e.g. `alt="Flowchart: how to identify the purpose of an image"`. This technique could also be used for announcing other kinds of descriptive images, such as **webcam images** (e.g. `alt="Webcam: Avoriaz ski slopes at 9.00 am, 23 January 2016"`).

Logos have received special attention in the literature, with mixed opinions about which is the most adequate approach to provide them with a text equivalent. Clark (2002, 89) considers that there is no downside to including the word 'logo' after the name of the company or institution it belongs to, as shown in Figure 3.10 (option a), and the HTML 5 specification (Hickson et al. 2014) suggests that this indication could be provided between brackets (see Figure 3.10, option b).[13] However, WebAIM (2015b) states that the use of this word is not necessary (see Figure 3.10, option c). This debate is still open and largely subjective in nature in the case of logos with a referential value. Nevertheless, agreement is often reached with regard to which information to include in functional logos.

Functional images

As discussed earlier, functional images may facilitate the dynamic interaction between the web page and the user (i.e. trigger an action). This is often achieved by embedding an image within a hyperlink. Concerning the *composition* of **linked images**, the HTML 5 specification points out that "when an `<a>` element that is a hyperlink, or a button element, has no text content but contains one or more images, [authors should] include text in the `alt` attribute(s) that together convey the purpose of the link or button" (Hickson et al. 2014). In general terms, by "purpose of the link" we understand "the link's destination". If the links points to an external site, then it is commonplace to indicate the name plus the word 'website', while if the link is relative (i.e. there is no change of website involved), it is recommended to indicate the state which page or location is the link taking the user to (see Figure 3.10, options d) and e) respectively). Apart from which keywords to employ, other *formulation* suggestions include: (i) avoiding the use of redundant expressions, such as "link to" or "click here to" (Nyong 2009), as well as advertising strategies, e.g. "Please visit..." (Clark 2002, 78); and (ii) warning visitors when a new window will open upon interaction with the image, for instance: `alt="Sunrise in Playa América, Galicia (larger image opens in a new window)"` (Clark 2002, 78).

[13] Other authors have propose this technique to mark images used for advertising purposes (e.g. `alt="Evian - fresh water from the Alps [advertisement]"`) (Clark 2002).

```
[option a]
<img src="unlogo.jpeg" alt="United Nations logo">

[option b]
<img src="unlogo.jpeg" alt="United Nations [logo]">

[option c]
<img src="unlogo.jpeg" alt="United Nations">

[option d]
<a target="blank" href="www.un.org"><img
src="unlogo.jpeg" alt="United Nations website"></a>

[option e]
<a href="..\index.html"><<img src="unlogo.jpeg"
alt="United Nations homepage"></a>
```

Figure 3.10. HTML code excerpt showing different alt text proposals for a logo

According to Korpela (2012), links express relations and verbs express actions, hence a linked image should rarely contain an imperative verb. Yet, this assertion does not apply for all linked image cases. WebAIM (2015b), for instance, states that the alternative text of a **button**, which is another form of linked image, should describe what the button will do when selected, and gives the following examples: "Submit search", "Place the order", "Accept terms and conditions". A formulation tip that could be inferred from these examples is that the alt text should start with a verb when a specific event occurs upon a click.

The other subcategory within functional images we have previously referred to in this chapter are images used for style or formatting purposes. Some web designers use little graphics in place of true typographic **bullets** or **dashes** in itemised lists (Clark 2002, 82), as well as images of **vertical bars** or **horizontal lines** to mark the beginning and the end of different web sections. While this kind of information should be introduced via HTML elements or the CSS, web designers may use images instead to achieve a certain appearance effect. Guidance found in the literature on the *composition* and *formulation* of alt texts for these images addresses the same aspect: whether the text alternative should contain a denotative (content) or connotative (function) description of the image. Thatcher et al. (2006, 146) and Korpela (2012) consider that both options are feasible. For instance, for a bullet image, both `alt="bullet"` and `alt="new item"` would be correct. They also agree that when a literal description is chosen, no further details about the image should be provided (e.g. it would not be appropriate to say `alt="small and round red bullet"`). Other authors, such as Jiménez Crespo (2008, 428) believe that only meaningful alt texts should be accepted as valid (i.e. interpretations of the image function; for instance `alt="end of the section"`). We agree with this point of view, given that the speech synthesizer offers the same reading for punctuation marks (e.g `alt="- - -"`) and their names (e.g `alt =" dash dash dash"`). Furthermore, studies have revealed that users find the announcement of punctuation marks frustrating and distracting (Petrie et al. 2005).

Advice about the formulation of text alternatives for other types of functional images is almost non-existent. With regard to the composition, it is worth highlighting the recommendations made by the W3C for image maps and CAPTCHA images. In the case of **image maps**, the "`alt` attribute on each of the `area` elements provides text describing the content of the target page of each linked region" (see Figure 3.6 earlier in this chapter). As for **CAPTCHA**, "the text alternative in the `alt` attribute provides instructions for a user in the case where she cannot access the image content" (e.g. `alt="If you cannot view this image an audio challenge is provided.">`) (Hickson et al. 2014).

3.3.2.3 Summary of existing guidance for text alternatives formulation

In the previous sections, we have reviewed the recommendations offered in the literature to compose and formulate appropriate text alternatives for images. Authors seem to be consistent when giving advice as to which should be the general language used (succinct, simple) and which information should never qualify as valid within and `alt` attribute (file name, file size, file format, meaningless placeholders). However, we have observed certain disagreements regarding the optimum alt text length, and some disparities when the matter was addressed from an image-type perspective. This might be derived from the fact that advice given is often based on individual experience and, as WebAIM indicates (2015b), determining appropriate, equivalent, alternative text for an image is often a matter of personal interpretation. Research-grounded recommendations, such as the ones presented in Petrie et al. (2005) are valuable, but scarce at present.

Similarly, there is a lack of complete language-oriented guidance. Hardly any references have been found with regard to preferred lexicon or syntactic structures (noun-based versus verb-based), apart from those provided by Tercedor Sánchez and Prieto Velasco (2009). In addition, the examples given in the literature are mostly founded on the English language, thus restricting the accessibility of existing resources to English speaking web actors. Furthermore, they are not presented explicitly but rather hidden in complex documents of technical nature, such as the W3C specifications. While web designers and developers may be used to browsing this type of documentation, people from more language-oriented professions, such as content editors or translators, might feel discouraged and even overwhelmed by the abundant technical jargon.

Table 3.3 summarises the most significant recommendations for alt text formulation found, which image category they apply to, and what level of agreement is reached among the different sources consulted. The literature review presented in this chapter with regard to how to write appropriate text alternatives for images played an important role in the development of a controlled language-based proposal for alt text checking, described in Chapter 4 (see sections 4.3.1 and 4.3.2).

Table 3.3. Overview of recommendations for alt text formulation

Recommendation	Image purpose	Agreement
Length		
>100 characters	all	low
100-150 characters	all	high
<150 characters	all	low
Redundant expressions		
image/icon/picture of	all	high
link to, click here to	functional	high
logo	descriptive, functional	medium
Placeholders		
insert content here, alt text, image, loading	all	high
Punctuation marks		
full stop (.)	descriptive	medium
use of brackets [ ] or parentheses ()	functional, descriptive	medium
other (– • |)	functional	low
ASCII, HTML entities	all	high
Image metadata		
file format	all	high
file name	all	high
file size	all	high
General language		
succinct, simple language	all	high
accurate	all	high
use of vivid adjectives	descriptive	medium
present tense, active voice	descriptive	low
use of verbs	functional (buttons)	low-medium

3.4 Assessing accessibility of images on the Web

The achievement of appropriate text alternative for images is highly dependent on the level of accessibility awareness of the person who is generating them, as well as on their analytical and language skills. In some web development contexts, time restrictions may also be imposed, making it sometimes impossible for people involved in the web production chain who lack any knowledge of image accessibility to check relevant guidance on the subject. In addition, as we have previously demonstrated, advice can be only retrieved at present from disparate academic and industry-based sources, since no set of complete guidelines exist that can be easily consulted without requiring a significant time investment. In such scenarios, people may rely on automated evaluation tools to look for a quick solution or simply to reassure themselves that they have taken the right decisions.

Together with inspection methods, such as guidelines conformance review, automatic testing is one of the main techniques used for web accessibility evaluation (WAE). Earlier in this thesis (see sections 2.2.1.3 and 2.3.3, Chapter 2), we have already discussed, based on previous studies, the advantages and disadvantages of such technique. In this section, it is our intent to exclusively focus on existing solutions to automatically assess the image accessibility level achieved on a web page or website. Concretely, we are interested in the support they provide for web content to be compliant with WCAG 2.0 SC 1.1.1, which requires images to have a text alternative that serves the equivalent purpose. To this end, we will take into account the two main aspects involved in image accessibility assessment: the *algorithmic testing* and the *judgment testing*. *Algorithmic testing* generally verifies the presence of a valid HTML element or attribute, such as the `alt`. *Judgement*, in turn, comes in when evaluating the quality of the text alternative contained within that element or attribute, that is, whether the text string appropriately serves as a textual equivalent to the image in question (Thatcher et al. 2006, 408-410).

In what follows, we aim at exploring if and to what extent both testing levels are covered in current state-of-the-art tools. Specifically, we will first look at the image accessibility verification capabilities of general web accessibility evaluation (WAE) tools (3.4.1), and then examine prior work done in the development of innovative tools exclusively designed to assist evaluators in image accessibility assessment (3.4.2). After the tool and literature review, we will set forth the advantages and limitations of existing automated solutions, and argue the potential benefits of using other complementary tools, such as controlled language checkers (3.4.3).

3.4.1 Image accessibility coverage in general WAE tools

According to the literature, the features of an accessibility evaluation tool can be presented from different perspectives: (i) *the resources* that can be evaluated (e.g. full web content, only text, only markup formats, only media content); (ii) *the testing functionalities*, including the selection of the guideline(s) against which to check the selected resources, and the test modes available (automatic, manual, semiautomatic);[14] (iii) the *error reporting capabilities* (i.e. how errors are presented

[14] Test modes are explained by Velasco and Abou-Zahra (2014) as follows: (i) *automatic*: the test is carried out automatically by the tool without any human intervention; (ii) *semiautomatic:* the test is partially carried out by the tool, but human input or judgment is still required to decide or help decide the outcome of the test; and (iii) *manual:* the test is carried out by human evaluators (this

to the user and the quality of the error repair guidance offered); and/or (iv) other *tool usage* characteristics, such as the tool interface or how well the tool integrates into authoring or evaluation workflows (Velasco and Abou-Zahra 2014).

In the context of this thesis, our interest is placed on tools that enable the verification of HTML documents and include the possibility of selecting WCAG 2.0 (and ideally level A, SC 1.1.1) as the baseline to perform the test. Similarly, we expect tools to be as specific as possible concerning conformance with SC 1.1.1. Specificity can be defined as “the level of detail that the tool is able to use when describing a potential problem” or in other words, as “the number of different possible issues that can be detected and described by a tool” (Brajnik 2004, 5–7). The larger this set of possible issues, the more capable the tool is of providing specific warnings and suggestions, and therefore the more useful it is (ibid). In the case of text alternatives, a tool would be very specific if (i) it detects the absence of the `alt` attribute, (ii) it detects inappropriate `alt` attribute values, and (iii) it distinguishes between different instances of invalid text alternatives by reporting on language-related problems or their adequateness according to the image purpose (e.g. if it distinguishes incorrect alt texts for functional images from the rest). Last but not least, tools should aim at the maximum level of automation (automatic mode as opposed to manual or semiautomatic, see footnote 14 of this chapter). Finally, if a high degree of specificity cannot be reached, we expect tools to offer enough error repair guidance for those who have little or no knowledge about image accessibility (e.g. by providing examples, clear explanations of the problem found, or links to WCAG 2.0 sufficient techniques and other online helpful resources).

The last decade has seen a growing trend towards the use of web-based WAE tools, probably because they are platform independent and often updated on a regular basis.[15] In particular, web browser extensions, such as the **Web Accessibility Toolbar (WAT)** for Internet Explorer developed by The Paciello Group,[16] and the **Accessibility Evaluation Toolbar (AET)** for Mozilla Firefox developed by Jon Gunderson,[17] have been increasingly gaining popularity. These resourceful add-ons offer access to easy-to-use functionalities, such as the possibility of identifying different elements of a web page or of changing the way in which the

---

includes cases where thy are aided by instructions or guidance provided by tools, but where the evaluators carry out the actual test procedure).

[15] The Web Accessibility Initiative (WAI) website features more than 70 tools, at least half of which are no longer traceable or maintained by their owners (the list dates from 2005 and has not been updated since). https://www.w3.org/WAI/ut3/ER/existingtools.html Last access: 25th January 2016.

[16] At the time of writing, the lastest release was from May 2015:
https://www.paciellogroup.com/resources/wat/ Last access: 25th January 2016.

[17] At the time of writing, the lastest release (veresion 1.5.7.1.1) was from August 2011:
https://addons.mozilla.org/en-GB/firefox/addon/accessibility-evaluation-toolb/?src=userprofile
Last access: 25th January 2016.

page content is visualised (e.g. by removing the CSS). When used for image accessibility evaluation purposes, they can help evaluators to retrieve a list of all the images present on a web page, or hide both inline and CSS images in order to check whether a text alternative is displayed instead (see Figure 3.11). However, for a more advanced assessment, one needs to use other third-party WAE tools.

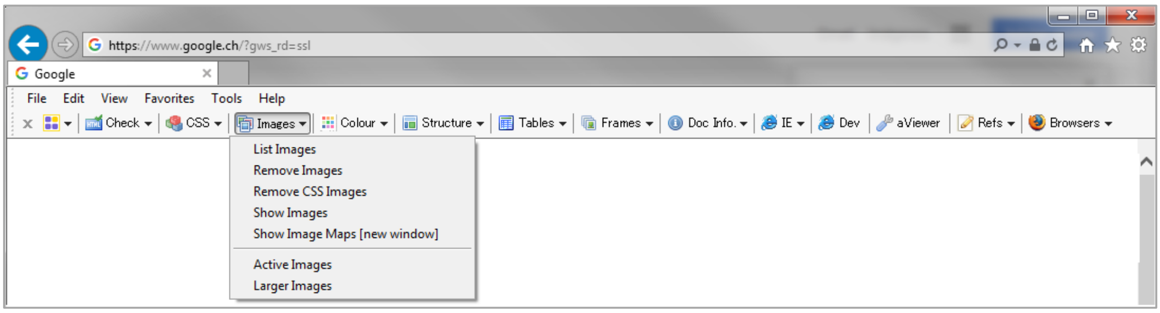


Figure 3.11. Web Accessibility Toolbar (WAT) by The Paciello Group for Internet Explorer

Both the WAT and the AET toolbars provide direct access, among others, to the W3C HTML validator, the WebAIM web accessibility evaluation tool (WAVE), and the Functional Accessibility Evaluator (FAE). The AET toolbar also suggests users check web pages against CynthiaSays, AChecker and T.A.W, all of them online accessibility validators. By way of illustration, we will mainly focus on the image accessibility testing functionalities of the first three and make reference to other tools' features for comparison purposes when appropriate.

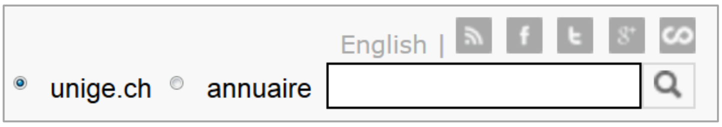
The **W3C HTML validator**[18] enables evaluators to carry out algorithmic tests by checking the robustness of the markup (HTML, XHTML, SVG) used in web documents (either online or locally stored). It automatically detects the absence of `alt` attributes, showing the message "required attribute `alt` not specified" and indicating the line of the source code in which the error appears. However, there is no direct reference to the WCAG 2.0., non-empty `alt` attribute values are not highlighted, and no further guidance is offered about how to assess alt text quality.

The **WAVE tool**,[19] developed by WebAIM, is a free online accessibility validator that allows its users to check web pages available online against WCAG 2.0 level A and AA guidelines, as well as Section 508. After running a check, results are organised in the report around four categories: "errors", "alerts", "features" and "structural elements". The first category corresponds to errors detected in automatic mode, whereas the other three include information to support evaluators when performing semiautomatic or manual tests. Algorithmic testing for conformance with WCAG 2.0 SC 1.1.1 is fully supported, since WAVE lists missing

[18] https://validator.w3.org/ Last access: 25^{th} January 2016.

[19] http://wave.webaim.org/ Last access: 25^{th} January 2016. The tool can be also installed as an extension in Google Chrome browser.

`alt` attributes as errors. Besides, it distinguishes between linked and non-linked images in the error report and provides a different explanation for each problem, as well as brief tips about how to fix them. The tool also points to the corresponding sections of the WCAG 2.0 online (see Figure 3.13, content in green box to the left).




```
<a href="http://www.unige.ch/feed/rss" title="S'abonner au flux
RSS"><img alt="RSS"
src="//www.unige.ch/themes/unige/images/icon_rss.png"/></a>

<a href="http://www.facebook.com/unigeneve" title="Facebook">
<img alt="Facebook"
src="//www.unige.ch/themes/unige/images/icon_facebook.png"/></a>

<a href="http://twitter.com/UNIGEnews" title="Twitter"><img
alt="Twitter"
src="//www.unige.ch/themes/unige/images/icon_twitter.png"/></a>
```

Figure 3.12. HTML code excerpt corresponding to linked social media icons in the University of Geneva website

Figure 3.13. Feedback about image text alternatives in WAVE

Nevertheless, judgment testing support is limited. WAVE only marks images with an alt text in-context and provides general instructions about their evaluation, such as "Ensure that the alternative text presents the content of the image and/or the function of the link." For instance, in the example presented in Figure 3.12, the first three linked icons shown in the image have “RSS”, “Facebook” and “Twitter” as alt texts respectively. These are not completely appropriate, given that they do not provide enough information about the destination of the link (e.g. is it the University of Geneva profile on Facebook that will appear upon clicking on the Facebook icon, or will the user be redirected to his own Facebook page to share the University of Geneva website on his wall?). Yet, WAVE does not offer any specific warning about the quality of these alt texts, but simply highlights them to the user (see Figure 3.13, content in red box to the right). Other tools such as **T.A.W**[20] and **AChecker**[21] adopt a similar approach: while the former marks the aforementioned alt texts as “warnings”, the latter identifies them as “potential problems” and provides two contradictory indications: “`<img>` element may require a long description” and “alt text is not empty and image may be decorative”.

To the best of our knowledge, the **Functional Accessibility Evaluator 2.0 (FAE)**[22] is, to date, the most complete general WAE tool in terms of image accessibility testing functionalities available online. As the other tools presented thus far, it analyses web pages online for requirements defined by the WCAG 2.0 level A and AA Success Criteria. However, unlike others, it supports coding practices suggested in the latest accessibility technologies, like ARIA 1.1 and the HTML 5 specification. Once the check is finished, results can be filtered by type of element (e.g. forms, tables, links or images). Within each element report, flags are divided into four categories: “violations” (V), “warnings” (W), “manual checks” (MC), and “passed” (P).

Performing an automatic algorithmic testing is possible with FAE, since it detects all missing `alt` attributes in a web page or site, listing them as “violations”. In addition, certain judgment testing aspects can also be automated through the tool, although the rules cannot be customised: FAE identifies alt texts which are more than 100 characters long and text alternatives that include the name of the image file. If these rules are not violated, FAE marks them as “passed” (see Figure 3.14, green label). Other type of alt texts, such as the ones shown in Figure 3.12, are simply presented under the rule “Alt text must summarize purpose” (i.e. in semiautomatic testing mode). Guidance offered to understand this rule is more comprehensive than in the case of the other tools showcased before. It gives hints

[20] T.A.W. is an online WAE tool of free access developed by CTIC, Spain. The date of the last update is unknown. http://www.tawdis.net/ Last access: 25th January 2016.

[21] AChecker is an open source WAE tool developed within the framework of the Government of Ontario's Enabling Change Program. It was last updated in 2011. http://www.achecker.ca/checker/index.php Last access: 25th January 2016.

[22] FAE is a WAE tool developed by the Open Accessibility Alliance and the OpenAjax Accessibility Task Force. Checks can be performed through a simple online form or the AInspector Sidebar add-on for Firefox. The last version (0.9.10) dates from August 2015. http://fae20.cita.illinois.edu/ Last access: 25th January 2016.

about how to identify the purpose of an image with examples, it explains how the `alt`, `title` and related ARIA attributes can be used, and it points the evaluator to failures and sufficient techniques associated to WCAG 2.0 SC 1.1.1 (Michael Cooper, Kirkpatrick, and O Connor 2015a). **CynthiaSays**[23] also offers very detailed information about alt text-related failures and techniques, but it is less intuitive and does not provide any type of automatic support for judgment testing.

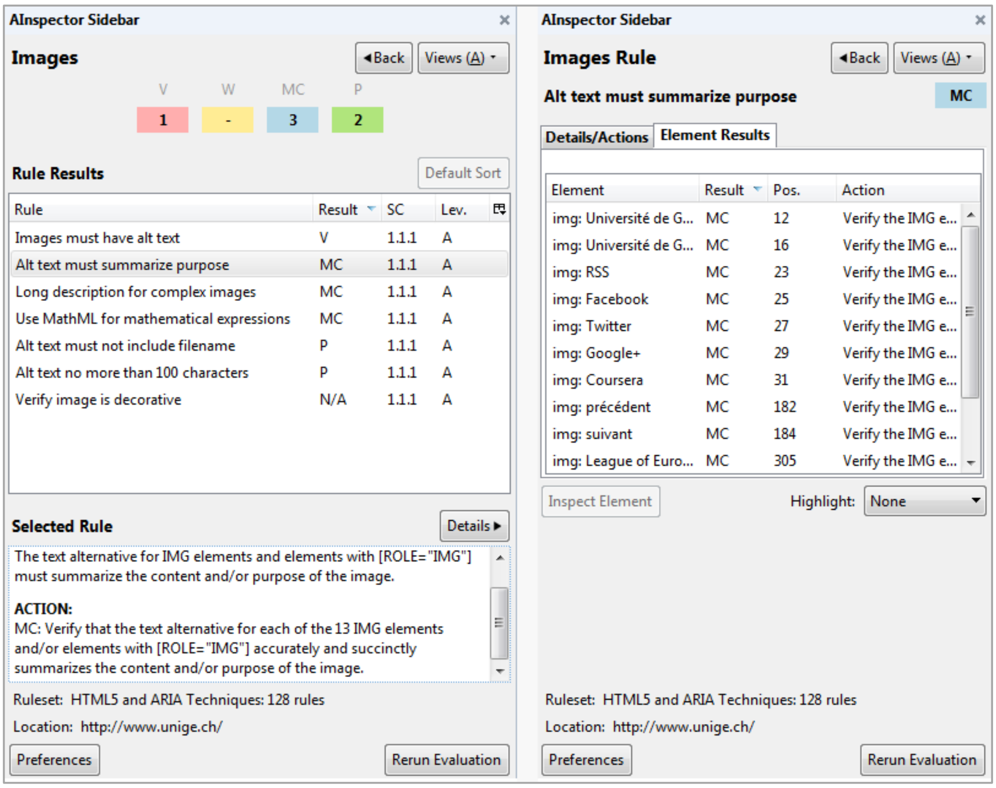


Figure 3.14. Feedback about image text alternatives in FAE AInspector Sidebar

Only certain stand-alone desktop WAE tools, such as A-Prompt and aDesigner, are comparable to FAE in terms of alt text quality verification features. **A-Prompt** was developed at the University of Toronto, Canada, and the TRACE Center at the University of Wisconsin, USA. Although its last release dates from 2008,[24] the tool is still rated as one of the most powerful WAE tools in most recent accessibility-related literature (Connor 2012, 309–311). Apart from detecting missing `alt` attributes, it marks alt texts as “suspicious” when:

- they exceed 150 characters (10-12 words).

[23] This portal is a joint education and outreach project of Cryptzone, ICDRI, and the Internet Society Disability and Special Needs Chapter, USA. http://www.cynthiasays.com/ Last access: 25th January 2016.

[24] http://www.softpedia.com/get/Internet/Other-Internet-Related/A-Prompt.shtml Last access: 25th January 2016.

- they include file size information (e.g. `alt="1675 bytes"`).
- they include an image's filename (e.g. `alt="xmas_tree.gif"`) or an image file extension (e.g. `alt="JPEG"`)
- they are used as placeholder text (e.g. `alt="insert image here"`).

Error repair guidance offered by A-Prompt is, however, scarce. Yet, this lack of support is somehow compensated by two repair features. First, A-Prompt allows the user to enter new or amended alt texts, and replaces them automatically in the web page when confirmed. Second, all text alternatives introduced within the A-Prompt repair environment are stored in a database and made available later via a dropdown menu for reuse.[25] While these features might be advantageous for people responsible for websites who are already familiar with image accessibility best practices, we believe they could lead to undesired alt text problems if not carefully used by users with no accessibility background (for instance, the database could be gradually fed with incorrect text alternatives and serve as a source for error propagation).

aDesigner offers a similar performance to that of A-Prompt as regards the automation of algorithmic and judgment testing for image accessibility compliance. To the list of "suspicious" alt texts provided by A-Prompt, aDesigner adds text alternatives of one-character length (such as `alt="*"`). These are identified by the tool as "probably errors" in the check report, while other non-empty alt texts are considered as items that need "human check". aDesigner presents two other advantages when compared with A-Prompt. Firstly, besides being a WAE tool to validate web pages against WCAG 2.0, it also serves as a visual disability simulator. When a web page (online or locally stored) is checked, the evaluator can see a styled version of the page to the left and a non-styled version to the right, simulating the order in which content would be presented to the blind user (see Figure 3.15). This non-styled version shows as well the errors flagged by the tool in context. Secondly, aDesigner's error repair guidance includes not only a brief description of the problem, but also a list of associated failures and sufficient techniques.[26]

To the best of our knowledge, the other tools currently available on the market do not offer further image accessibility checking capabilities than the ones we have presented in this section. Previous web accessibility research studies involving the use or the assessment of general WAE tools have occasionally made general assertions about their lack of automated support for image accessibility judgement

[25] Older WAE tools such as LIFT or InFocus, which have now disappeared from the market or been incorporated into other commercial web authoring and auditing suites, shared some of A-Prompt functionalities. While LIFT also enabled the identification of the same type of "suspicious" alt texts, InFocus test settings included the possibility of adding list of stop words that should be detected as invalid (had to be an exact match) within `alt` attribute values (Thatcher et al. 2006, 424–434).

[26] More information about aDesigner can be found in Chapter 5, section 5.2.3.3.

testing, but have provided no or little empirical evidence (Ivory and Chevalier 2002; Brajnik 2004; Mankoff et al. 2005; Vigo et al. 2013; Klein et al. 2014). In the following section, we shall look at tools or systems specifically designed for image accessibility enhancement or evaluation.

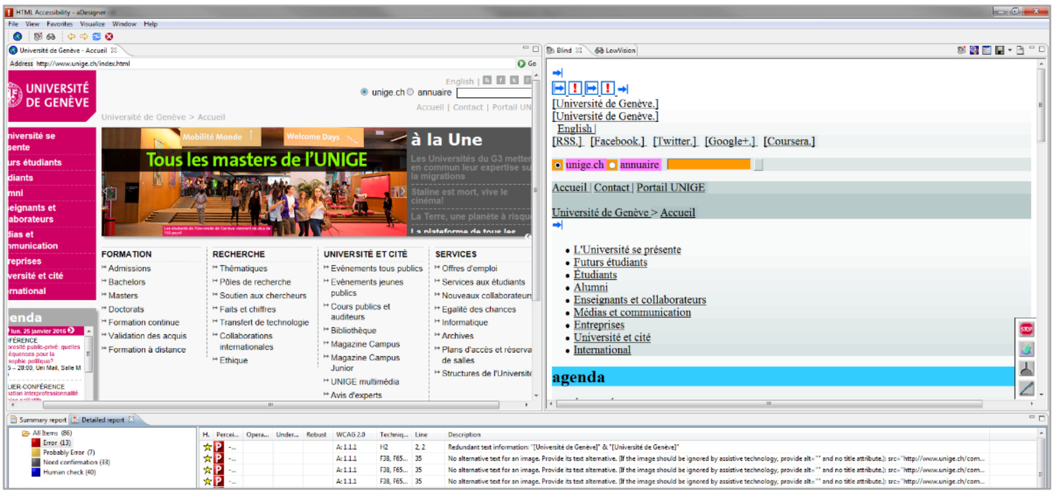


Figure 3.15. aDesigner interface showing a styled and non-styled version of the web page checked (top), and the error report (bottom)

3.4.2 Studies on and tools for automatic evaluation of image accessibility

Prior research work in this area has sought to cover general WAE tools' flaws concerning the verification of image accessibility by focusing on the improvement of automation techniques for: (i) the identification of incorrect text alternatives, and (ii) the generation of alt texts when these were missing within the HTML document.

ALTifier was the first tool developed to assist web developers and content authors in the generation of text alternatives (Vorburger 1999). The tool would scan a web document and provide its user with a list of images lacking an `alt` attribute. Then, based on a series of parameters, the user would be presented with an alt text proposal. The heuristics used by ALTifier to guess a text alternative include: (i) looking at an image's height and width to identify small images such as bullets and rulers, and detect spacer images; (ii) looking at the file name and then replacing '_' and '-' by a blank space (in the hope that files were given meaningful names); and (iii) analysing hypertext links for extraction of document titles in the case of linked images (ibid). However, no studies were conducted to assess the performance of the tool.

Bigham developed an image classifier to "differentiate between good and bad alternative text" (2007, 350) using similar heuristics to the ones chosen by

Vorburger. The system relies upon different features derived from the surrounding context (e.g. how close is the alt text content to the content on the web page where the alternative text is found), as well as on dictionary-based word search; that is, the classifier would treat (i) alt texts as valid if the individual words they are composed off exist in a dictionary or on the web, and (ii) alt texts with arbitrary strings, sequences of punctuation and complex filenames as incorrect. In addition, to determine whether an image should have an alt text, Bigham used a heuristic that stated that an image should be labelled as “must have an alt text” if it was multi-coloured and larger than 10 pixels in both dimensions or if it has a link (2007, 3251).[27] During the evaluation of the tool, the system failed to label 7,999 out of 8,350 images. From the remaining 351, 239 were correctly classified as “good” or “bad”, achieving an 86.3% accuracy.

As Bigham, Olsen et al. (2010) used pattern recognition techniques to develop a system for classifying alt texts as descriptive or undescriptive. Image text alternatives would be categorised as undescriptive if the following features were present: file type abbreviations such as GIF or JPEG, none alphabet characters such as “*”, image size, HTML entities, and “words that are known to cause accessibility barriers such as 'read more', 'click here' or 'title'” (2010, 428). While experiments with this system showed a higher accuracy rate (90%) than in the case of Bigham's image classifier, both tools fail to offer a solution to assess the quality of the alt texts marked as “good” or “descriptive”. It is likely that not all text alternatives which do not include the aforementioned recognisable features are, in fact, appropriate.

In this sense, the automatic generation of alt texts combined with later human intervention could be observed as a fair compromise. Bigham et al. (2006) proposed a system called WebInSight that uses optical character recognition (OCR) and web context labelling to automatically calculate a text alternative. More specifically, WebInSight relies on link-based labelling for linked images and OCR for those images that contained text. To formulate alt text for linked images, the link is explored by the system and a candidate alternative text is returned from the title and headings present on the linked page. Images whose alt text cannot be generated automatically are sent to human labelling services (ibid). This approach is interesting for images which lack an `alt` attribute, but would be more powerful if combined with alt text classifiers like the ones presented before. Besides, it has other major drawbacks: first, alt text generation is limited to linked images and

[27] Considering image dimensions for image classification is quite controversial, since images might as well be large in size and yet be simply used for decorative purposes. Nevertheless, this parameter, together with the presence of colour, has also been used in other web accessibility research studies (Hanson and Richards 2013), as well as in prior work done to facilitate automatic image categorisation for web page summarisation (Hu and Bagga 2003) and image indexing purposes (Paek and Smith 1998).

images of or with text, and assumes that both the title and headings of the linked page, as well as the text present in the images, are meaningful. Second, human labelling services, while optional in the system, required additional costs.

The ESP Game (von Ahn and Dabbish 2004) is a game-based solution for image description that is also based on human-labelling, but that follows a crowdsourcing approach. The first game requires users to label images with isolated words, such as “small”, “brown” and “bag”. Then, users rate the appropriateness of these labels and once the image gets enough approval, it is removed from the to-be-tagged image list. This model, like the system proposed by Keysers et al. (2007), does not lead to the generation of natural language descriptions but simply one-word tags, which might not be regarded as sufficient by end users who expect more elaborated upon text alternatives. Phetch (von Ahn et al. 2006), another game-based tool, solves this inconvenience by requesting users to provide explanatory descriptions of images. Yet, while highly motivational and enjoyable, both computer games simply offer a platform in which to introduce potential alt texts to images that, in fact, are presented to players outside of their web context.

The work of Vinyals et al. (2014) is, to date, the most advanced approach to automatic generation of image descriptions. The Neural Image Caption Generator (NIC) is capable of producing perfectly structured natural sentences based on a neural and probabilistic network system. Although this tool is extremely promising, it was primary meant for large descriptive images and, as in the case of the aforementioned games, it does not take into account the image purpose within a given web document. Moreover, it shares two important limitations with previous systems alike: it only provides descriptions in plain English, and it offers no parallel solution to assess the quality of text alternatives which may have been already written by somebody else.

3.4.3 Advantages and limitations of current automated solutions

Through the use of current state-of-the-art evaluation tools, web developers and content authors can significantly speed up the image accessibility verification process. Almost by default, general WAE tools facilitate, through a simple mouse click, the identification of `<img>` elements without an `alt` attribute, and provide at least a brief explanation of why this attribute should not be missing. Furthermore, most of them offer the possibility of examining errors in context, which can be extremely helpful to understand the purpose of an image and later generate an alt text accordingly. Extraordinarily, WAE tools may even enable users to correct meaningless text alternatives within the same tool environment, as in the case of A-Prompt. In addition, reports generated by these tools often point to relevant W3C sources and occasionally give illustrative examples. Nevertheless, the main weakness of current WAE tools is that they do not offer enough automated support for judgement testing. Only a few exceptions are capable of detecting extremely long text alternatives, or uninformative alt texts including senseless placeholders, image metadata information or isolated non-alphabetical characters. Table 3.4 provides an overview of the tools reviewed in previous sections, indicating which testing modes are supported, according to the level of testing (algorithmic or judgement).

Table 3.4. Image accessibility testing support by general WAE tools

Tool	**Algorithmic testing**		**Judgment testing**	
	Possible?	*Testing mode*	*Possible?*	*Testing mode*
W3C Validator	yes	automatic	no	n/a
WAT toolbar	no	n/a	yes	manual
AET toolbar	no	n/a	yes	manual
WAVE	yes	automatic	yes	semiautomatic
T.A.W.	yes	automatic	yes	semiautomatic
AChecker	yes	automatic	yes	semiautomatic
FAE	yes	automatic	yes	partially automatic
CynthiaSays	yes	automatic	yes	manual
A-Prompt	yes	automatic	yes	partially automatic
LIFT	yes	automatic	yes	partially automatic
InFocus	yes	automatic	yes	semiautomatic
aDesigner	yes	automatic	yes	partially automatic

Additionally, error repair guidance in WAE tools suffers from considerable limitations. On one hand, it is generally oriented to web professionals with a technical background. WCAG 2.0 failures and sufficient techniques are sometimes referred to, but these can be overwhelming for content authors such as editors or translators. In fact, studies have demonstrated that even developers prefer to receive detailed explanations of each problem in clear human language (Petrie et al. 2011). On the other hand, when human input is needed to revise the appropriateness of alt texts retrieved during the check, tools provide too vague instructions, such as "check that all non-text content that is presented to the user has a text alternative that serves the equivalent purpose", rarely including suggestions on how to decipher which is the ultimate communicative value of the image.

Similarly, language-oriented recommendations are almost non-existent. No tool presents the user with tips on how to formulate an alt text (for instance, taking into consideration the recommendations reviewed in section 3.3.2). While, as Lawton Henry points out, evaluation tools can identify accessibility issues but cannot determine whether a product is ultimately accessible (2007, 97), one would expect tools to at least provide users with proper guidance on how to achieve it.

We think that this lack of linguistic guidance is due, in great part, to the low level of specificity of general WAE tools when it comes to detecting image-related problems. In this sense, one may believe that tools specifically designed for the improvement of image accessibility would identify a wider variety of inappropriate alt texts. However, most solutions only differentiate descriptive from uninformative text alternatives, 'good' from 'bad' ones, leaving no room for a more fine-grained sub-categorisation of problematic alt texts within each category. Finally, to the best of our knowledge, the tools that rely upon feature extraction techniques to give feedback to the user only take into account English text. This renders them less effective to detect, for instance, placeholders or one-word alt texts in other languages such as French or Spanish.

For all the reasons set out above, we believe that a controlled language (CL) solution, such as the one that will be presented in the next chapter (4), could prove advantageous in several ways. To begin with, it would allow for a higher level of customisation. Checks would be run taking into account the language of the page in which the images are presented. Since CL checkers feature a part-of-speech (POS) tagger and morphological disambiguator, its use would also allow for a larger flexibility in terms of feature extraction. Furthermore, CL technology could serve the needs of a larger audience. It could be a complementary guidance tool both for (i) web professionals who are less experienced in content editing, such as developers or designers, and for (ii) content authors, such as web translators and localisers, who may not be aware of image accessibility issues and might feel more comfortable with a tool that also provides linguistic feedback.

Before further developing our controlled language approach to alt text evaluation in Chapter 4, we will discuss if and how image accessibility concerns are being taken into account in current image localisation strategies.

3.5 Addressing image accessibility during web localisation

Thus far in the present chapter, we have examined the communicative value that images have within a web document and how this information can be transmitted to visually-impaired users through the use of HTML markup and the insertion of an appropriate text equivalent. Furthermore, we have explored how tools may support people responsible for websites in ensuring that image accessibility-oriented measures have been taken. In doing so, we have adopted a general monolingual Web perspective, that is, we have not considered the possibility of images having to be localised during the development of a multilingual website.

In this section, we aim at analysing the most important aspects of the image localisation process, with a view to understanding (i) which are the current challenges, and (ii) whether special emphasis has been placed by the industry and academia on accessibility-related issues, particularly in the translation, adaptation and/or generation of text alternatives.

3.5.1 General considerations about image localisation

Within the context of the production of multilingual content, images used in the source product often need to be manipulated (and sometimes even replaced or completely deleted) so that the new document in which they appear adheres to the expectations of the target community (Mata Pastor 2009a, 514). According to Ó Broin (2003), image localisation revolves around two key areas: the *global appropriateness* of an image and its *localisability*. The former concerns the suitability of the image for all target locales; in other words, it entails the study of the source image from a cultural perspective. The latter, however, deals with the technical efforts associated with the image localisation task: the ease of localisation —that is usually dependent on the image composition and the file format in which it has been saved—, and how tools can serve the localiser to automate the process (ibid).

While references to culture-related aspects may be occasionally made, this thesis primary addresses the topic of image localisation from a *localisability* point of view, as understood by Ó Broin, with a special interest in functionality-related aspects (i.e. the images' communicative value). In addition, it should be noted that (i) we will not cover the particularities of image localisation in the context of digital products other than web pages (for instance, we shall not address the challenges related to the adaptation of screenshots in software applications or help files), and that (ii) the focus will be placed on the localisation of static and linked images, as opposed to complex animated images (e.g. Adobe Flash animations).

3.5.1.1 What to localise about an image?

Image localisation can be requested as a separate service or as part of a larger web localisation assignment, to an individual localiser or to a localisation team. It might be also the case that it is not explicitly demanded by the client, but it is necessary for the proper functioning of the product in the target culture and, therefore, the localiser proposes it as part of the budget. Regardless of the scenario in which image localisation takes place, it usually involves modifications at two different levels: *visual* and *functional*. Changes that affect the visual result of an image include the adaptation of cultural elements and the translation of text that is graphically represented on the image, when applicable. Functional modifications, however, are needed for the image to be correctly perceived by the end user —via any sensory modality— and imply the adaptation of the image internal structure (HTML markup).

Visual level

Using images on the Web that are globally usable and locally acceptable is one of the key principles of internationalisation. However, when a culturally sensitive design approach is not adopted, localisers should consider whether country-specific graphical content would be natural for the target audience if kept exactly as in the source product. This would include looking at the appropriateness of flags, images of people, images showing specific body parts and actions, visual puns, icons and the use of colour within images, among others (Ó Broin 2003).[28] When any of these graphical elements does not satisfy the end culture expectations, the ideal approach would be to adapt it or replace it with another one which is deemed acceptable in the target culture, an approach also referred to as transcreation (Roturier 2015, 174–175). Yet, while this step is a priori recognised as one of the localiser's primary tasks (Gouadec 2007, 42–45), certain authors suggest that the localiser might not always be requested (or allowed) to make these changes himself, leaving the final decisions to the client or to other actors involved in the multilingual web production chain (e.g. local marketing teams) (Ó Broin 2003; Mata Pastor 2009a).

Translating the text that is visible on the image is, however, a must do for localisers. Text extraction techniques depend heavily on the image format, which in turn determines the ease of localisation of text-embedded images. Raster or bitmapped images are created as a pattern of pixels or 'picture elements', each of which is assigned a code that designates its colour (Savage and Vogel 2014, 125). When bitmapped images −usually saved in GIF, JPEG or PNG format− contain text, it is made up of pixels and not by editable alphanumeric characters. This means that localisable text strings are hard-coded in the image, and thus they cannot be accessed by standard localisation tools (Schäler 2010). In such cases, text can be extracted automatically by means of OCR systems, or manually, which would imply copying or retyping the text in a text editor (Klopman 2016). Although localisers could also try to manipulate the image using a graphics editing tool and translate the text directly in this environment, this practice is not recommended (Mata Pastor 2009a, 522). Text translation should be considered as an independent task, and only once the translation process has concluded, should the localiser (or, when applicable, the designer or the DTP specialist) reintroduce the translated text into the original image file (ibid).

[28] Although entering into considerations about the treatment of such elements during the image localisation process and the implications of overlooking them fall outside the scope of this thesis, they have received considerable attention in the localisation-related literature (Yunker 2003; Schäler 2003; Singh and Pereira 2005; Bolaños Medina et al. 2005; Tercedor Sánchez 2005; Schäler 2007). Their importance in terms of recognisability and memorability, especially in the case of colour, symbols and icons, has been also recognised in web design and usability studies (Marcus 1996; Nielsen 2014; Bedford 2014).

When the localiser is provided with images in the native format in which the bitmapped version was created or with vector graphics file formats, text extraction and its localisation is simpler. In general, native formats such as Adobe Photoshop .PSD, cannot be used by other applications but they present one important advantage: text is stored in a different layer from the rest of the image. If localisers own the native tools, they can inspect the file and quickly access the text. If not, they can use certain tools, as we will see in the next section, that allow for the automatic extraction of the text layers for ease of localisation purposes. Another possibility might be to export the image from the native tool into a vector graphic, such as SVG (Klopman 2016). As we have seen in section 3.3.1.2, SVG files are built on XML, which facilitates their localisation. Text strings that are visible on the image are contained in elements such as `<text>` or `<textPath>` (Dahlström et al. 2011), and can be either located and translated manually by the localiser without investing too much technical and time efforts, or extracted automatically with the help of translation and localisation tools.

Scalable vector graphics are considered to be more advantageous than bitmapped graphics for multiple reasons. Apart from the fact that vector graphics are meant to be more accessible, as we have already discussed, they allow for a smooth scaling and reshaping, their file size is considerably lower than in the case of raster images, and they contribute to a higher level of localisability, given that text is conveniently separated from other non-text elements present in the image (Ó Broin 2003; Donney et al. 2008). Nevertheless, its use is still uncommon, both in terms of implementation by authoring tools (Wassmer 2003) and by web professionals (Chen and Harper 2008).

Functional level

Once the localiser gets the desired visual appearance in the target image, they need to insert it in the target HTML document, so that it can be presented to the end user. This step implies inspecting the code and localising the necessary attribute values of the HTML element used to insert the image in question, for instance, the `<img>` element. Mata Pastor (2005, 215–219) lists the values of the following `<img>` attributes as localisable: (i) `src`, that would need to be modified if the location of the new target image within the website macrostructure has changed with respect to the original location, and (ii) `alt` and other similar attributes such as `longdesc`, whose textual content would need to be translated.[29]

If the image is embedded within an `<a>` element, the `href` value would have to be adapted as well. In the real example shown in Figure 3.16, we can observe some of the image localisation procedures discussed so far. In the Spanish version of the

[29] To this list, we may add the `<title>` and `<desc>` elements in the case of SVG graphics, which, although not visible in the final image, contain translatable text strings.

University of Salamanca's website (to the left), images provide access to information on the bachelor, master and doctorate programmes offered, as well as to the virtual learning platform and some information about the 800th anniversary of the University. In the English version (to the right), the content has been localised to accommodate the needs of foreign students, including a section with general information about international programmes, and another one listing five reasons to study at USAL. Images have been localised too. Notice how, for instance, the first image at the top that reads 'estudia en la USAL' in the Spanish version has been visually and functionally localised. The text has been translated and the picture shows now an international student smiling (as in two other images on the page), probably to depict the USAL as a welcoming institution. Yet, when inspecting the corresponding code, we observe that the `href` and the `src` have been modified, but the alt text has been left untranslated (and so were the ones for the other images on the page), which echoes the observations made by Fernández Costales (2010) after analysing a large corpus of localised websites from European universities. Overlooking the translation of text alternatives during the localisation process represents a threat to the accessibility of the page, since any potential English speaking screen reader will not be able to understand which type of information this website is offering.


```
ES:
<a href="/webusal/futuros_estudiantes" class="mc0101"><img
src="themes/usal/img/estudiaenlausal4.jpg" alt="Estudia en la USAL"></a>
EN:
<a href="/webusal/en/futuros_estudiantes" class="mc0101"><img
src="themes/usal/img/estudiaenlausal_en.jpg" alt="Estudia en la USAL"></a>
```

Figure 3.16. Spanish (left) and English (right) homepage of the University of Salamanca website, with HTML code excerpts for the first image 'estudia en la USAL'

The importance, from an accessibility perspective, of using and localising the `alt` attribute has been mentioned by authors involved both in localisation research (Yunker 2003, 349; Ó Broin 2003; Tercedor Sánchez et al. 2006; Tercedor Sánchez and Jiménez Crespo 2007; Mata Pastor 2009b, 552; Fernández Costales 2010, 200–

202; Roturier 2015, 88), and the web accessibility field (Clark 2002, 99). Our approach to image accessibility localisation, however, is more in agreement with that of Jiménez Crespo (2008, 430), who suggests that, irrespective of the translation task, the localiser should also assess whether the source alt text is appropriate, taking into account the function of the image it is associated to. Furthermore, he considers the possibility of the localiser having to add new text alternatives if needed:

> [D]urante el proceso de localización, [...] el traductor se enfrenta a cuestiones como la identificación de la función de la imagen en el texto, su valor denotativo, su valor connotativo, y cómo potencial la accesibilidad a la imagen en entornos web. En este contexto, se puede requerir que el traductor realice tareas de redacción técnica al crear o adaptar este texto alternativo para imágenes. (ibid)

We argue that this could occur in the following situations: (i) when images present in the source web document lack an `alt` attribute, (ii) when images are replaced in the target web page for cultural adequacy reasons, or (iii) when the alt text provided in the source web document is not appropriate. In addition, we believe that localising image functionality might imply warning the user about a change of language in the `alt` attribute. It might be the case that there are solid reasons to leave the alt text in a language different from the one used in the target web document. In such scenario, the HTML attribute `lang` or the XML attribute `xml:lang` (or both) would need to be inserted within the `<img>` element by the localiser (Clark 2002; Korpela 2012).

3.5.1.2 Tools for image localisation

The visual localisation of images remains, to a large extent, a highly manual task (Ó Broin 2003). Depending on the source image complexity and the localiser's skills, different tools may be used to visually modify an image: from simple image editors, such as the Microsoft Paint application or the freeware GIMP, to professional graphic design software like Adobe Photoshop (Mata Pastor 2005, 226). As mentioned earlier, the file format might also impose the use of different text extraction tools: from OCR software in the case of bitmap images to other more localisation-oriented tools for certain layer-based formats created by proprietary tools like Adobe Photoshop. Rainbow, for instance, enables the extraction of text layers from PSD files and prepares them for translation by wrapping the extracted strings into an XLIFF or RTF file (Wassmer 2003; Mata Pastor 2009a).

In the literature on image localisation, particular attention has been given to file analysis tools with budgeting and quality assurance (QA) features, Globalization Image Assistant (GIA) being the most widely cited.[30] GIA enables localisers to automatically examine a large number of images in order to retrieve information

[30] While very popular in the early 2000s, this tool seems to be no longer maintained by its creators.

such as file size, dimensions, number of layers and file type. In the case of raster images, the user can transcribe the source text into a Unicode text box to facilitate wordcount and translation tasks, but the manipulation of graphics is not possible within the tool environment. Once the localiser has made the necessary modifications to the image files, they can reimport them into GIA to perform a QA check, which basically consists of a side-by-side comparison of source and target files to detect inconsistencies concerning file size, type or dimensions (Haddad 2003). Similar tools include WebBudget[31] which, like GIA, is charged, and the freeware Image Localization Manager tool developed by Muñoz Sánchez (2008).

Nevertheless, none of the aforementioned tools are suitable for localising image functionality in websites. In order to process the markup language-based documents where the images appear, such as HTML (or XML in the case of SVG files), the most convenient and effective approach would be to use a Computer-Assisted Translation (CAT) tool and, if needed, make any additional changes by means of an advanced text or web editor.[32] Both Esselink (2000, 213–219) and Mata Pastor (2005, 228–234) indicate that the former should be preferred over the latter as the primary option because the risk of translatable text being overlooked is minimised, given that CAT tools protect the non-editable strings and generally propose to the users only those that need to be translated. Furthermore, with CAT tools, localisers have the possibility of using terminology databases, translation memories and other language-oriented features such as spellcheckers that are less common in text or web editors, if not non-existent (ibid). It is likely, however, that CAT tools will not treat attributes like `src`, `href` or even `alt` as modifiable by default. This would require the localiser to be aware of their meaning and purpose, and to customise the tool accordingly (see section 3.5.2.2.).

Depending on the context of the web localisation assignment, as well as on the HTML and related computing skills of the localiser, other tools may be used to localise image functionality. If the localiser is requested to transform a monolingual website into a multilingual one, the use of web design and authoring tools could prove useful in recreating the website macrostructure in order to render the target website fully functional—a process which is also referred to as website cloning or pseudo-cloning (Gouadec 2007, 42)—. Similarly, in the case of dynamic websites, the localiser might be asked to work directly in a Content Management System (CMS), such as Drupal or Joomla, if they have not been previously provided with the web content to be localised in a CAT tool-supported file (e.g. in XML, XLIFF or CSV) (Torres del Rey and Rodríguez V. de Aldana 2014, 164–171).

[31] http://www.webbudget.com/ Last access: 30th January 2016.

[32] In the case of SVG files, an alternative to text or web editors might be to use tools designed for scalable vector graphics production and editing, such as the freeware Inkscape: https://inkscape.org. Last access: 30th January 2016.

3.5.1.3 Main challenges of image localisation

Irrespective of the tool at their disposal, localisers still face important challenges when dealing with image localisation. Flat or single-layer images with embedded text are the most commonly used in websites, presumably for two reasons: one, raster images are preferred over vector graphics by webmasters and developers to facilitate image loading (Donney et al. 2008); and two, designers often prefer to use raster images to ensure that the text appears exactly as intended in a web browser (Yunker 2003, 146). Overcoming the technical obstacles derived from the lack of adherence to W3C standards and internationalisation best practices, such as separating text from non-text content, requires an additional time investment for localisers and extra costs for the client. As highlighted by Savourel (2001), SVG offers all the advantages that XML technology provides (Unicode support, easy access to translatable text), but unfortunately this format has lost the battle against GIF and JPEG in terms of popularity on the Web (Chen and Harper 2008, 24–30).

Furthermore, working with non-flexible image files can lead to undesired localisation errors, especially if it is not the localiser who is in charge of producing the final version of the target image. While ideally source images should allow expansion space for localised text, so that there is no need to redraw a graphic or resize a text label (Ó Broin 2003), space restrictions in the target product might be imposed. In such scenarios, localisers need to adopt coping strategies such as changing the font size or type, or using text abbreviation techniques for text legibility not to suffer (Mata Pastor 2009b, 536).

Last but not least, localisers do not always receive the support needed by the client to successfully carry out their task. As Torres del Rey and Rodríguez V. de Aldana emphasise (2013, 5), it is these professionals who are in the best position to “interpret objects, texts and meanings, not for their own sake but for other people, users, for whom translation and localisation is performed, for whom the translator is ethically responsible”, and who have been trained on how to deal with the technical aspects related to website and image localisation. And yet, localisers need to invest continuous efforts to educate the client, not only about their needs in terms of resources (e.g. provision of native file formats, more interaction with other members of the multilingual web production team), but also with regard to their knowledge and abilities.

Finally, it is our belief that not enough measures exist ‒either on the localiser's side or on the client's side‒ to ensure that image functionality is fully achieved in the target web product. It has been claimed that localisers invest more efforts in adapting elements that are essential for the correct functioning of the localised web page or site at a superficial level (e.g. verifying that links are not broken or take the user to the wrong page) than in localising other elements that are part of the back-end or internal structure of the page, such as the content of the `<meta>` element (Jiménez Crespo 2008, 419–429). More specifically, as far as image localisation is concerned, Jiménez Crespo's study suggests that less importance is given to the translation of image-related text strings which are invisible on the website front-end, such as the alt text (ibid, 427). As we have illustrated in Figure 3.16, more attention is paid instead to ensuring the visualisation of the image in the right position within the page or the correct adaptation of the `href` value, in the case of linked images. We argue that this might be due to a lack of awareness of the accessibility benefits of image text alternatives by web localisers and/or to the failure of translation aids in pointing out —implicitly or explicitly— the importance of alt texts to localisers.

3.5.2 Localising images with accessibility in mind: what is needed?

In this last section of the chapter, we shed light on the studies that have put emphasis on fostering an accessibility-aware image localisation process and we present to what extent state-of-the-art localisation-oriented tools support this endeavour.

3.5.2.1 Awareness about alt text functionality

As mentioned earlier, the `alt` attribute value has previously been identified in the literature as a translatable text string. So far, however, there has been little discussion about the importance of raising awareness about the ultimate purpose of alt texts among localisers, or the relevance of producing new appropriate text alternatives, if needed, during the localisation process.

Between 2005 and 2009, the research team from the group *Imagen y Texto* (Image and Text) at the University of Granada, Spain, worked, to the best of our knowledge, on the single largest research project that included web image description studies applied to a translation domain. Its broader goal was the inclusion of multimedia material in the technical and scientific translation classroom (Tercedor Sánchez and Jiménez Crespo 2007). Adopting a cognitive and pedagogical approach, they focused on the image as the object of translation, as well as a visual support to text comprehension. The major observational study involved 70 translation students, who were asked to provide both a denotative and a connotative description to a series of images on the Web, with a view to assessing their knowledge and awareness about accessibility (ibid). The experiment demonstrated that most of them were aware of the importance of adapting image-related textual information to accommodate the various needs of the end users. It was later complemented with a corpus-based descriptive research that investigated the presence of image text alternatives in a Spanish parallel corpus, comprising both original and translated corporate websites (Jiménez Crespo 2008).

While the resulting analysis offers general descriptive statistics, such as the number of `alt` attributes found per sub-corpus with and without a textual value (Tercedor Sánchez and Jiménez Crespo 2007; Jiménez Crespo 2008), the actual appropriateness of the text alternatives collected in both studies was not further assessed.[33] In addition, student image descriptions and the presence or absence of `alt` attributes in the corpus were observed independently from the translation and web localisation process, leaving behind technical considerations such as the potential impact of having used specialised tools to process or check the HTML files produced, or the skills that students would have needed to manipulate the web page source code, should the `alt` attribute be missing.

To our understanding, two of the most noteworthy contributions of this research to the field under study are the well-grounded theoretical framework developed (Prieto Velasco 2009) and the practical criteria defined for image description that we have already referred to in section 3.3.2 (Tercedor Sánchez and Prieto Velasco 2009). However, both seem to have been designed for and rely on the particular nature of technical and scientific texts, where complex images play an essential role in knowledge representation, and then extrapolated to touch audiovisual translation and web localisation interests. Similarly, while the studies conducted with translation students have brought to the forefront the importance of producing quality image descriptions, further empirical studies are needed to investigate how image accessibility is addressed by translation professionals in a web localisation context.

3.5.2.2 Translation aids support

According to Prieto Velasco (2009, 198), the ideal tool for image localisation should allow users, among other requirements, to translate not only the text that is visible on the image (if any), but also the `alt` attribute value. We have already seen that the translation of the former remains, to a large extent, a manual task, for which only partial automated support is possible. The translation and/or adaptation of the latter, however, can be easily accomplished through the use of Computer-Assisted Translation (CAT) tools.

CAT tools enable localisation professionals to directly receive source HTML files —instead of decontextualized translatable strings in plain text format— process them with this software, perform the translation task, and return the automatically-generated target language HTML files, sometimes without even touching the code, especially if no localisation engineering tasks are requested. CAT

[33] As highlighted in Chapter 1 (see section 1.2.2.2), this property was not analysed either in Fernández Costales' qualitative localisation study (2010), which showed that alt texts were not often translated in multilingual university websites.

tools include built-in terminology databases and translation memories which allow localisers to reuse previously translated content when matches are found in the text being processed. In addition, they isolate translatable content from the document markup, rendering the latter non-editable. Nevertheless, text extraction patterns are often customisable, and not all CAT tool providers establish the same extraction rules by default.

Failure to translate the `alt` attributes when using this type of software can be due to an incorrect configuration of the segmentation or parsing rules defined to handle HTML files. As pointed out by Mata Pastor (2005, 217), localisers should not blindly trust CAT tools to retrieve all localisable strings for them. Instead, localisation professionals should have the necessary HTML skills to first assess which elements and attributes require their intervention, as well as an advanced knowledge of the CAT tools they are using to then customise the HTML parsing rules accordingly. For instance, Figure 3.17 shows the options available to establish how the `<img>` element is processed in SDL Studio 2015. Notice that the `alt` attribute has been marked as translatable ('True'), but the `src` attribute hasn't ('False') (see content within the red box on the top).

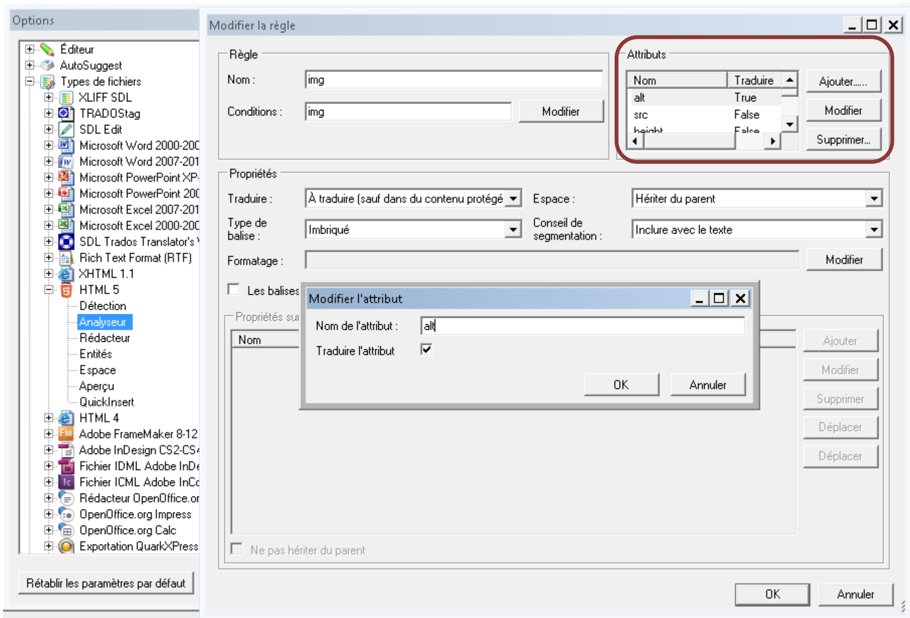


Figure 3.17. Text extraction rule options for HTML filesin SDL Studio 2015

Identifying an alt text as such within the translation editing environment is equally or even more important than translating the text alternative itself. In general, CAT tools provide hints about the provenance of each source segment. This allows localisers to check the different text strings in context (e.g. by previewing the web document on a graphical browser or even within the same CAT tool environment[34]) before making any translation decisions. In the case of image text alternatives, this is crucial, given that the appropriateness of alt texts is dependent on the function of the image and its surrounding context. The way in which information about alt text is presented to the user varies across CAT tools. Some just provide the user with the plain text to translate and, within the segment metadata, they indicate which HTML element and attribute the text has been extracted from (see Figure 3.18). Others hide this information behind 'tags' which actually protect the non-editable content, as in the example shown in Figure 3.19. It is only when this information is displayed that the user can see the HTML structure (see window 'Etiqueta' in Figure 3.19). Either way, localisers need to be able to interpret this information appropriately and take it into account when inserting a translation proposal.

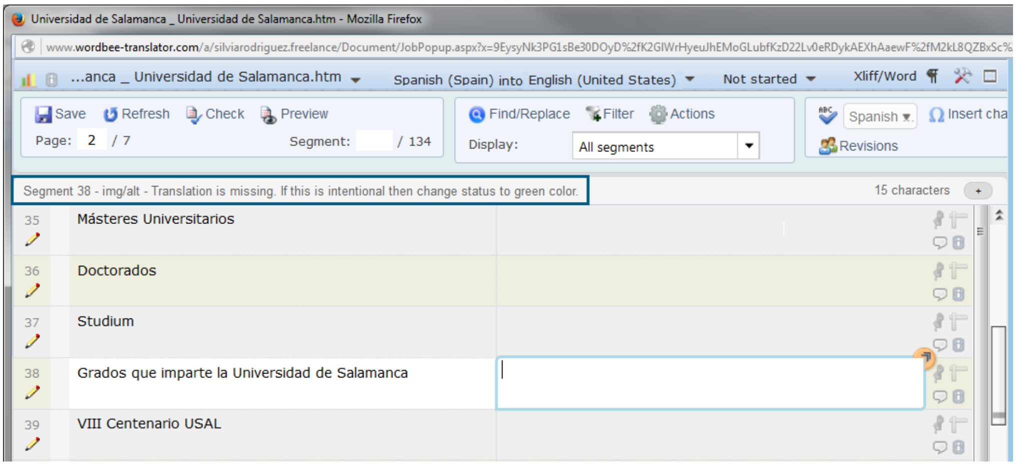


Figure 3.18. Alt text handling by Wordbee

Once the translation phase has been completed, the resulting web page should be verified and tested (Gouadec 2007; Sandrini 2008). Localisers need to make sure that all the source text content has been adapted into the target language. The latter can be done manually by loading the website in different browsers and visually inspecting each page translated, or automatically, through the use of quality assurance (QA) tools. Current CAT tools feature text-related QA functionalities, such as looking for untranslated segments, as well as punctuation, formatting, terminology and inconsistencies. Stand-alone QA tools, such as QA

[34] Regardless of the source file type being processed, most CAT tools present the text to translate in the same table-like structure (see the example shown in Figure 3.18). Some offer a WYSIWYG mode as well, where users can preview how the target document would look once the translation is finished. This functionality is, however, only available for certain types of files (often those which are non-proprietary).

Distiller[35], ApSIC Xbench[36] or Verifika[37], also offer the possibility, among others, of defining regular expressions to search for pattern matches or common typing mistakes, such as duplicated words or spacing errors (Debove et al. 2011). Potential mistranslations can also be automatically flagged by comparing source and target sentences' length (*ibid*). However, all the above does not fully guarantee a quality web localisation job. During the localisation process, changes made may have altered the website's layout or functionality, sometimes leading to encoding problems, broken links or truncated strings due to the new text length, which must be corrected (Sandrini 2008). Existing translation-oriented QA solutions primarily focus on language-related issues, but they do not verify the robustness of the target HMTL documents, nor do they provide any hints about potential accessibility problems which may have been transferred from the source document or emerged when producing the target web product.

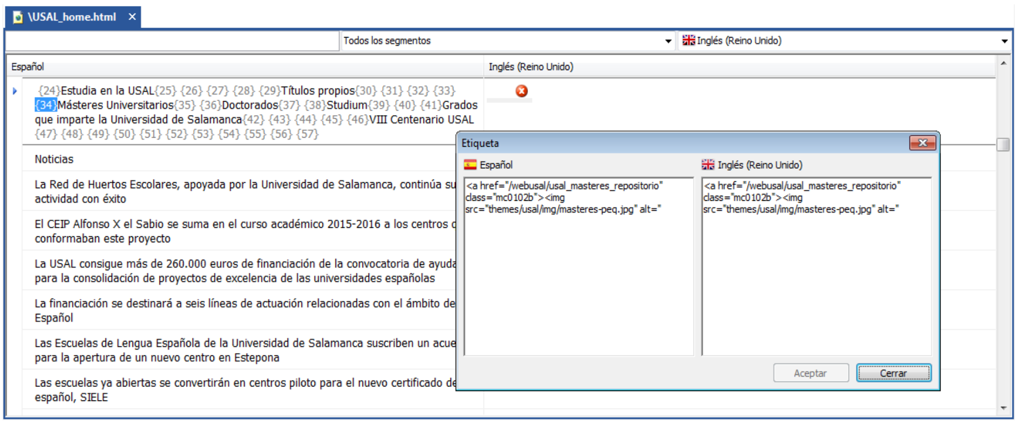


Figure 3.19. Alt text handling by DVX3

Findings from an exploratory study carried out with screen reader users have suggested that two of the main obstacles they experience particularly when browsing the multilingual Web are the inaccessibility of language selectors and the presence of untranslated content, including image text alternatives (Rodríguez Vázquez 2015b). Adding an accessibility testing phase to the web localisation workflow could help to solve such problems, as well as to increase awareness about screen reader users' needs and browsing behaviour within the localisation community: a need that has already been identified in Chapter 2. Since expertise in web accessibility matters has not traditionally been observed as a requirement for localisation professionals, automated solutions could serve to bridge this gap.

[35] http://www.qa-distiller.com/en Last access: 30th January 2016.

[36] http://www.xbench.net/ Last access: 30th January 2016.

[37] https://e-verifika.com/ The latest version (2.5) dates from August 2015. Last access: 30th January 2016.

As we have seen in section 3.4 and 3.5, the achievement of image accessibility is only partially supported by existing related guidance and automated support. While complementing CAT tools with current web accessibility evaluation software is necessary, localisers would encounter the same obstacles as other web professionals when trying to generate appropriate text alternatives. In this sense, the CL-based solution, already suggested in section 3.4.3, could prove advantageous for localisers as well. In fact, the use of linguistic-intelligent authoring programs, through which controlled languages can be implemented, has already been proposed in prior work as an alternative way to reinforce translation quality (Rösener 2010). Applying purpose-driven CL rules can partially relieve content authors and translators of going through time-consuming and hard-to-understand style guides and easily ensure, for example, that client-specific writing conventions are met. In the case of image accessibility assurance processes, a CL checker could help localisers to automate the application of existing recommendations (and potential new ones) with regard to alt text formulation. In the next chapter, we will explore this proposal further in detail.

3.6 Summary

We began this chapter by considering the communicative value of images on the Web from a web accessibility studies perspective (section 3.2). In section 3.3, we described how descriptive, functional and decorative images can be rendered accessible to screen-reader users, and presented the `alt` attribute as the most robust and well-supported mechanism to provide a text alternative for images when these cannot be visualised. Then, existing guidance on how to write appropriate alt texts was reviewed. The main limitation observed was the lack of complete language-oriented guidance about their formulation. In section 3.4, we have demonstrated that this limitation also exists in current web accessibility evaluation tools and that state-of-the-art image accessibility assessment solutions have not yet fully addressed this issue. Finally, in section 3.5, we advocated for the empowerment of localisers as facilitators of image accessibility in the target web product. While the importance of translating text alternatives has been highlighted in the literature, no studies have been conducted to date that investigate whether this type of content is in fact translated during the localisation process, or whether the use of translation aids or QA tools, including WAE and CL software, has an impact on the level of image accessibility achieved in the localised web product.

Chapter 4

A controlled language-based approach to assure image accessibility

This chapter presents a controlled language (CL) proposal consisting of 40 rules designed for creating appropriate image text alternatives in French. We describe how these rules and their corresponding documentation were developed, how they can be applied using Acrolinx technology, and which rules were finally selected for evaluation in a later experimental study (see Chapter 5).

4.1 Overview

Among other important considerations related to image accessibility, in the previous chapter we have brought to the forefront the major shortcomings of existing guidance, evaluation and repair tools regarding the assessment of image text alternatives. From this perspective, we have claimed that a more language-focused automated evaluation approach would be needed to reduce suspicious silence and meaningless noise rates in the check reports yielded by these tools. Similarly, we have argued that better results in terms of alt texts' quality could be achieved if users of accessibility checkers (whether web creators, content editors or localisers) were provided with pertinent feedback on the linguistic appropriateness of the text alternatives they produce, which in turn should be dependent on the pragmatic value of the images they describe.

In this chapter, and with a view to attaining Goal 2 of this thesis (see Chapter 1, section 1.3) we put forward a controlled language-based (CL) proposal based on 40 CL rules that could lend itself particularly well to offer that kind of guidance and thus bridge the existing gap. Since none of the currently available web accessibility evaluation (WAE) technologies includes a robust in-depth linguistic analysis functionality through which we could apply the 40 rules we defined, it was decided to formalise and implement them using Acrolinx technology, one of the most popular commercial automated solutions for controlled language checking. The definition of a new CL proposal and its formalisation using Acrolinx technology represent the design and creation research strategy (Oates 2005, 108–114) presented in the Introduction of this thesis.

The structure of this chapter is as follows: section 4.2 briefly reviews how the application of CLs has proved beneficial in previous research work, and introduces the CL proposal that we believe could successfully complement current image accessibility evaluation techniques. In section 4.3, we report the development methodology adopted to define the aforementioned set of CL rules, aimed at producing appropriate text alternatives in French (section 4.3). After describing the resulting 40 rules (section 4.4), we present how text alternatives can be checked against these rules by means of one of Acrolinx's clients for content verification: the Acrolinx Batch Checker (section 4.5). In the same section, we also introduce the Acrolinx's rule documentation feature, which will be particularly relevant for the experimental study carried out in the context of this thesis (see Chapters 5 to 7). Finally, in section 4.6, we indicate which rules were evaluated during the said experimental study and which procedure was followed to select them.

4.2 Controlled language applications

4.2.1 General definition and CL classifications

One of the most recent definitions of controlled language (CL) that can be found in the literature states that “a CL is a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax, and/or semantics, while preserving most of its natural properties” (Kuhn 2014, 123). This constructed language is often organised around a set of rules that impose certain language constraints which, in turn, are the result of well-thought-out choices (Roturier 2006, 47). This feature is what differentiates CLs from sublanguages, which are not artificially created, but rather naturally arise when experts who share specialised knowledge about a restricted semantic domain communicate about it in a recurrent situation (Kittredge in Kuhn 2014, 124).

From a historical perspective, lexicon, grammar and style-related restrictions in CLs have been primary oriented to text simplification purposes with a view to attacking natural language's inherent complexity (Huijsen 1998), the ultimate goal being the facilitation of human communication. This was precisely the main objective of the first broadly recognised CL, *Basic English*, developed back in the 1930s to help non-native speakers learn English by solely relying on the use of 850 words (Ogden 1930). Since then, over 100 CLs have been created both within academic and industry contexts. Although most of them are based on the English language (Kuhn 2014), controlled varieties of Esperanto, French, German, Greek, Japanese, Mandarin, Spanish and Swedish have also been surveyed (Pool 2006).[1]

[1] When combined, Pool (2006) and Kuhn (2014) offer, to the best of our knowledge, the most comprehensive list of CLs developed up to the present. Given that we have defined our own CL rule set, based on text alternatives-related literature and a purpose-built corpus, without relying on any of the already defined CLs (see section 4.3 of the present chapter), a detailed examination of the CLs compiled by Pool and Kuhn falls outside the scope of this thesis.

The research community has successfully faced the challenge of bringing order to such a wide array of CLs by offering classification proposals taking into account either the CL purpose or the main CL properties. The very first attempt was made by Huijsen (1998), who argued that controlled languages could be human-oriented (HOCL) or machine-oriented (MOCL). While the former seek to improve text comprehension by humans (increased readability and comprehensibility), the latter aim at improving text comprehension by computers (reduced ambiguity and complexity), hence the need for these to be precise and computationally tractable (ibid.). This distinction has been widely used since then, with authors referring sometimes to HOCLs and MOCL as naturalistic and formalistic CLs respectively (Pool 2006).

Schwitter (2015) recently suggested categorising CLs according to the broader purpose they have been designed for, instead of specifically focusing on the problems they were trying to address, as he had initially suggested more than ten years ago (Schwitter 2002). In his new CL classification proposal, he reviews four main groups: (i) CLs for human communication, (ii) CLs for technical documentation, (iii) CLs for machine translation, and (iv) CLs for semantic systems. If we were to follow Huijsen's binary distinction, the first two groups would correspond to HOCLs and the last two groups would be classified as MOCLs.

Other authors have suggested that the purpose is only one of the multiple defining properties of CLs. Wyner et al. (2010) consider that the purpose of a controlled language is as important as who the intended users are and whether the language is domain dependent or independent. In addition, they argue that, in order to properly define a CL, it is necessary to look at around 40 other linguistic, design and application properties. Nevertheless, according to Kuhn (2014), some of these are somewhat fuzzy and do not allow for a strict CL classification. In order to categorise a selection of 100 CLs, he collected instead only nine properties to describe their application environment: the goal (comprehensibility, translation, formal representation); the form (written, spoken); the origin (academia, industry, government) and the domain (specific narrow domain).

Kuhn (2014) found that more than 90% of all controlled languages reviewed were intended to be written. Data concerning domain-specificity were balanced, with half of the languages being designed for a specific and narrow domain, and the other half covering a more general scope. This balance was also observed with regard to the CLs origin (academia versus industry) and the CLs purpose: almost half of the surveyed languages aimed at increasing comprehensibility, while formal representation was the goal of the other half (ibid). Interestingly, Kuhn claimed that 22% of all languages had translatability as their primary objective, pointing at a large overlap between these CLs and those intended to improve text comprehension.

This finding is understandable, given that, as suggested by Nyberg et al. (2003) the advantages of CLs include not only the improvement of readability and comprehensibility levels, but also the consistency and reusability of the source text. Taken together, these benefits render text manipulation easier for both human and computers: they can lead to a smoother human translation, as well as a more effective use of translation aids and machine translation. The aforementioned overlap between CLs for readability and CLs for translatability is also reflected in the attention that CL has received by the translation research community, as will be discussed in the next section.

4.2.2 Controlled language studies in translation research

Over the last decade, we have experienced an increased interest towards the study of CLs within translation and localisation research. The vast majority of studies carried out during that period of time have focused on technical texts, with a strong emphasis on the implications of combining CL approaches with machine translation (MT) for the quality of the final translation product. However, before reviewing those, it is worth mentioning a few others which were primarily devoted to assess the impact of implementing CL rules for the sake of readability.

One example is the Localisation Research Centre (LRC)-awarded work of Cadwell (2008). He showed that controlled versions of three technical texts identified using a readability formula as “very challenging to read”, “somewhat difficult” and “very easy” (i) were thought to be easier to read by both experts and non-expert readers, and (ii) encouraged better retention of keywords. By applying a different methodology, O'Brien (2010) later supported – albeit only partially – Cadwell's findings. By considering only two categories of texts (very difficult and very easy to read) and using eye-tracking methods, she found that the positive effects of CL were just limited to more complex texts. When examining the readability and acceptability of the machine-translated French, Spanish, and Simplified Chinese versions of the same controlled and uncontrolled source English texts, she came to the conclusion, however, that CL had only a small positive effect and varied across languages.

One could say that O'Brien's MT-related results are in line with those from Roturier (2006), who had previously claimed that MT output can sometimes (i.e., depending on the CL rules applied, the type of texts and the language pair) be regarded as useful, comprehensible and acceptable when the source content is controlled. Doherty (2012) also examined the effects of CL on the reading and comprehension of machine translated texts using a mixed-methods approach. The studies conducted revealed that subjective measures (involving humans in the evaluation process) pointed at significantly higher levels of readability and comprehensibility, while objective measures, such as readability formulas, did not.

The different nuances in the outcome of all these studies demonstrate the difficulty of assessing the impact of controlled languages, often acknowledged in the literature (Huijsen 1998; Hartley and Paris 2001; Nyberg et al. 2003)

The effectiveness of applying CL rules on the quality of MT output has also been explored by measuring different aspects related to the post-editing task, with some studies highlighting that, in fact, certain CL rules can both benefit text comprehensibility and reduce the post-editing effort (O'Brien and Roturier 2007). In O'Brien (2006), it has been empirically shown that controlling the input to MT leads to faster post-editing rates, while Aikawa et al. (2007) have suggested that using CL does not only result in higher MT quality, but also in greater post-editing productivity. Similarly, in the study conducted by Gerlach et al. (2013), it has been claimed that the post-editing effort in terms of time and keystrokes was clearly lower for machine-generated translations of pre-edited sentences.

As can be inferred from all the above, most of these studies have limited the application of CLs to the MT input, that is, CL rules have been used to pre-edit the source text, prior to its machine translation. A recent study by Porro et al. (2014) has shown that rule-based automated post-processing of MT output can also be beneficial both in terms of readability and usefulness for subsequent manual post-editing. Nevertheless, few other studies exist that have described or assessed CL rules for MT post-editing purposes. In the same vein, little empirical research has been conducted on the implications of applying CL rules (either manually or automatically) during human translation quality assurance processes. Our intention in this thesis is to contribute to this particular area of study.

It is equally worth noting that much of the available literature on translation and controlled language deals, as mentioned earlier, with technical texts (mainly from the IT domain), written by experts. It is normally expected that the production of this type of texts, regardless of the form in which they are presented (in print or on the screen), will be easily controllable due to their conventionalised linguistic configuration (e.g., short and often imperative sentences, restricted but specialised lexicon, use of passive voice, etc.), hence its recurrent use in CL studies. Although recent work has considered content produced by forum users instead of experts (Gerlach 2015), the texts that served to develop CL rules in her study were addressing topics from specific and narrow domains.

The novelty of the CL research presented in this thesis is that it does not deal with technical texts. The focus is shifted to a particular text type which is rather domain independent but, at the same time, cannot exist outside a hypertextual structure, since it is embedded in the HTML markup. Additionally, it relies on the multimodal nature of the Web, as its ultimate function is to reproduce information initially presented to the user in visual form. The next section aims at precisely exploring the use of controlled languages in the Web as a specific medium of communication.

4.2.3 Forms of controlled language in (and for) the Web

Thus far we have been referring to the concept of readability without further elaborating on its meaning. Zamanian and Heydari (2012) believe Dale and Chall's definition, dating from 1949, to be the most comprehensive: "The sum total (including all the *interactions*) of all those *elements* within a given piece of printed material that affect the success a *group of readers* have with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting" (ibid). More than 60 years later, it would suffice to replace the word 'printed' with 'digital' to get a definition close enough to what we currently understand by Easy to Read on the Web.

The notion of Easy to Read on the Web also builds upon the interaction of different content elements, putting emphasis on how these are accessed and perceived by web users. However, it goes beyond the text to encompass aspects related to information architecture and web design, thus contributing as well to the more general concept of usability (Miesenberger and Petz 2014). Some authors argue that this is one of the most notable differences between the concepts of easy to read and plain language (PL), both of which could be considered as forms of controlled language themselves (Matausch and Nietzio 2012).

While the latter mainly focuses on linguistic guidelines at a word, sentence and paragraph level (PLAIN 2011), the former may not only cover microstructure and macrostructure language-oriented recommendations, but also involve text customisation techniques (font size, line length, word spacing). Similarly, in the E2R literature, it is encouraged to complement textual content with other enriching elements, such as images or multimedia, with a view to illustrating the message that is being transmitted through the text (Inclusion Europe 2009; WebAIM 2010; Cooper et al. 2015b, sec. Understanding Guideline 3.1; Insieme 2015).

To support the readability and comprehensibility aspects of E2R, the research community has devoted considerable efforts to develop and test Natural Language Processing (NLP)-based techniques, such as text annotation (Topac and Stoicu-Tivadar 2012), lexical and syntactical simplification (Saggion et al. 2011; Drndarevic et al. 2012) or text summarisation (Aluísio and Gasperin 2010; Watanabe et al. 2010). Nevertheless, scholars and international bodies have always been confronted with two major challenges: determining who the recipients and beneficiaries of E2R web content are, and how to transform E2R main principles into easily applicable and testable guidelines.

According to the W3C/WAI Research and Development Working Group (RDWG) (2014), globalisation demands information that is usable and readable across borders and cultures, as well as understood by the largest possible user group. The notion of Easy to Read on the Web expands this demand to also take into account the needs of people who have problems with the language in use and information presentation on the Web (ibid). This vast target group may include not only users with intellectual or developmental disabilities, but also non-native speakers and language learners. This audience diversity has led researchers to develop more targeted E2R guidelines for specific user populations, such as dyslexics; who may benefit from the use of concrete font sizes, font and background colours, and spacing constraints (Rello et al. 2012); or people with autism spectrum disorders, for whom figurative or non-literal language such as metaphors, idioms or sarcasm poses a major obstacle (Štajner et al. 2012).

As briefly outlined earlier, the second challenge of building and assessing easy-to-read websites is that many of the essential E2R principles, such as using simple grammatical structures, short sentences and a controlled vocabulary are difficult to state as simple or unambiguous rules (Älli 2012). The most successful attempt to prescribe and implement some of these E2R principles can be observed in the Simple English Wikipedia[2], whose articles try to follow the pillars of *Simple English*, a modified version of Odgen's controlled language *Basic English*, which mostly insists on reducing syntactic complexity (Schwitter 2015).

Efforts to formulate a more definite set of E2R guidelines have been made within multiple international organisations, including the World Institute on Disability (1999), the UK Disability Rights Commission (2004), Inclusion Europe (2009), WebAIM (2010) or the W3C (Cooper et al. 2015a, sec. G153: Making the text easier to read; Cooper et al. 2015b, sec. Understanding Guideline 3.1 and Success Criterion 3.1.5), to name just a few. A review of their work reveals that most sources agree on the following recommendations:

General guidelines:

- Divide blocks of information into manageable units
- Place main idea before exceptions and conditions
- Keep language as simple and clear as possible
- Provide practical examples to clarify content
- Use bulleted or numbered lists instead of long paragraphs
- Check spelling and grammar

Sentence-level guidelines:

- Use sentences no longer than 25 words
- Use the simplest sentence form (subject-verb-object)
- Avoid the passive voice
- Avoid multiple negatives
- Use sentences that contain no more than two conjunctions
- Avoid complex sentences with dependent clauses
- Where possible, use the present tense

[2] https://simple.wikipedia.org/wiki/Main_Page Last access: 26th November 2015.

- When needed, give direct instructions

Word-level guidelines:

- Avoid contractions
- Avoid slang and jargon
- Avoid the use of metaphors
- Avoid unfamiliar or long words
- Avoid words in a different language from that of the main content
- Use highly descriptive words as hypertext anchors
- Avoid acronyms and abbreviations
- Use positive terms
- Avoid weak verbs (excessive use of 'to be' or 'to have')
- Avoid hidden verbs (verbs converted into a noun)

While some of these guidelines seem to be more concrete in nature than those established by the Simple English Wikipedia (which, for instance, does not describe the exact length of a simple sentence or give any particular hint about which verbs to use), others still remain too vague. This can be problematic when trying to apply them manually, as well as difficult to implement them automatically. A possible solution to reduce the ambiguity of some E2R on the Web rules is to limit their area of application.

Mosquera et al. (2012), for example, have shown their concern about the accessibility of short texts used in Web 2.0 applications, which often feature emoticons (e.g. :-) or :=D), non-standard abbreviations (e.g. RT for 'Retweet', TT for 'Trending Topic') and an excessive use of interjections and letter repetitions (e.g. yeaaaahh!!!!). Within this specific context, rules such as “Avoid the use of abbreviations” or “Avoid slang forms” become easier to embrace as well as to formalise. To render this type of web content accessible, Mosquera et al. (2012) propose a NLP-based automated tool capable of transforming these texts into their canonical form. Controlled language solutions have also proved beneficial for the particular case of producing simple RSS feeds, as demonstrated by Schwitter and Tilbrook (2006). Our CL proposal follows a similar approach, as it restricts some of the E2R common guidelines, also found in the literature on how to produce appropriate image text alternatives (see section 3.3.2, Chapter 3), to a particular area of application on the Web: `<img> alt` attributes.

Miesenberg and Petz (2014) hold the view that assistive technologies for blind and visually-impaired people, such as screen readers, also contribute to Easy to Read on the Web by accommodating the display of text. To this assertion we may add that screen readers also contribute to Easy to Read on the Web by accommodating the display of images through the presentation of a textual equivalent. Images might act independently from the textual content or, as we discussed earlier, be there to support or complement an idea present in the text. In either of these situations, text alternatives are crucial for screen reader users to reach and get hold of the information, an inherent goal of E2R on the Web (W3C/RDWG 2014).

In this thesis, we propose a controlled language approach to assure that these text alternatives transmit appropriately the information conveyed through the images they describe. This challenges the classic understanding of CL as a technique that renders texts easier to read and understand (Nyberg et al. 2003), and expands it to also facilitate the perceivability of images on the Web.

4.2.4 Defining a CL proposal for image accessibility on the Web

If we tie together the defining properties of our CL proposal, gradually introduced throughout the preceding sections, we would characterise it as a *human-oriented written controlled language* whose purpose is to assist in the production of appropriate text alternatives for images. The ultimate goal of implementing such a CL proposal is the achievement of a higher level of image accessibility on the Web.

One could argue that, since text alternatives are inherent to digital genres – as opposed to textual genres, as per Jiménez Crespo (2013, chap. 4) – and need to be processed by assistive technologies before being presented to the blind user, our CL proposal should rather be defined as machine oriented or dual-oriented, a term coined by O'Brien (2006, 6) to frame CLs intended for both human- and machine-processing. However, in this thesis, we depart from the assumption that the set of CL rules we propose are intended for *human comprehension*. This is based on the premise that appropriate text alternatives do not only facilitate an enhanced web user experience for blind people, but also contribute to an improved access to and understandability of visual information being consulted on the Web.

If we adhere to Wyner et al. list of controlled language generic properties (2010), it is important to mention that the intended users of our CL proposal are all actors involved in the web production cycle, from content editors to web developers and localisers. The rule set that will be presented in section 4.4 could be both implemented during content authoring or evaluation processes. In other words, it could serve to write appropriate text alternatives in the first place, or to check their appropriateness, after the production phase.

This distinction applies for the development cycle of both monolingual websites and multilingual websites, that is, our CL proposal could prove pertinent in the authoring and evaluation of source web content, as well as in the equivalent phases of web localisation (translation and quality assurance). Within the framework of this thesis, and unless specified otherwise, the focus is placed on the implementation of our CL proposal as a complementary localisation quality assurance measure. We will thus refer to it as a CL for text alternatives verification, rather than for text alternatives authoring, and refer to web localisers as the intended users of the CL rules and the CL checking technology chosen to implement them (see section 4.5).

Similarly, it should be noted that our CL proposal is domain-independent and language-dependent. Although English forms were occasionally considered during the rule development phase (see section 4.3.2), the CL proposal is intended for verifying French textual content. The language choice is motivated by the corpus of web documents on which the CL is fundamentally based (see section 4.3.1). Taking into consideration all the above, it was decided to test the CL proposal within the context of a web localisation experimental study, in which the target language should be French (see Chapter 5).

4.3 CL development methodology

The development methodology followed in order to define a CL proposal to assure image accessibility consisted of two phases: corpora exploitation (section 4.3.1) and rule definition and formalisation (section 4.3.2).

4.3.1 Material selection

According to Hartley and Paris (2001), the construction of new controlled languages is grounded in the analysis of a corpus of documents. This corpus is used to judge which linguistic forms to include in the language (Wyner et al. 2010). In our particular case, two different corpora were built to develop a CL proposal for the verification of image text alternatives.

The first corpus consisted of the existing literature concerning image accessibility on the Web, which was systematically and comprehensively reviewed with a view to identifying language-related recommendations on how to create appropriate text alternatives (see Chapter 3). The examination of this collection of documents was crucial to understand what type of description should be provided depending on the image purpose. It also enabled us to gather, when available, hints about syntactical and lexical preferences.

However, this corpus alone was not considered sufficient to build concrete CL rules because it did not provide enough representative examples of each one of the guidelines suggested. Therefore, a complementary corpus was needed to serve as a training data set for the rule development. The only requirement we sought was that it should be a web corpus, given the nature of the texts we wanted to observe. This matches the recent trends observed in CL-oriented research, where the Web has been increasingly used as training data (Kilgarriff and Grefenstette 2003). The material finally selected for that purpose is presented in the next sections.

4.3.1.1 Training web and alt text corpora

The web document collection chosen as the source of training data initially consisted of 100 websites assessed within the context of a web accessibility study, commissioned by the Swiss authorities and conducted by the Swiss Access for All Foundation (Access for All 2011). The aim of the study was two-fold: (i) to know whether these websites were compliant with WCAG 2.0 up to level AA, (ii) and to observe whether improvements in terms of accessibility had been achieved since 2007, according to the results obtained in the first accessibility study carried out that year by the Foundation. The WAE methods followed were conformance review and user testing. The study findings revealed that, while advances had been achieved from 2007 to 2011, major challenges still remained. Interestingly, image text alternatives were ranked as the most important accessibility barrier yet to be solved (Access for All 2011). We thus assumed that the corpus would contain enough evidence of non-appropriate text alternatives for images upon which to build our CL proposal.

The 100 websites were predominantly Swiss, half of which were mainly governmental and cantonal sites, including those of the ten biggest cities of the country (50%). Others were related to national public transport (10%); state services, including education and postal mail (10%); national media (20%); and private services of potential interest for people with special needs (job portals, shopping online, social media) (10%). Governmental and public administration websites are a common object of study in web accessibility research work (Hong et al. 2008; Bühler et al. 2008; Gambino et al. 2014) because it is normally expected that they will be compliant with existing national and/or international web accessibility guidelines. Since over 70% of the sites included in the Access for All study belonged to that category, we deemed it reasonable to ground our research on this Swiss national corpus.

This choice imposed some constraints in terms of language. Most of the syntactical and lexical preferences identified in the first corpus of image accessibility literature were based on English (see Chapter 3). However, very few Swiss websites have English versions and, when they do, only a simplified localised version is offered to the user. Similarly, not all 100 websites were available in the three official Swiss languages (French, German and Italian). Of these three, only French was among the working languages of the researcher, so websites in languages other than French were dismissed. This decision reduced the initial web document collection size down to 59 websites. Two global websites, namely Facebook and Wikipedia, were also removed, in order to maintain a certain heterogeneity within the training data corpus.

The web accessibility study carried out by the Access for all Foundation took place between April and July 2011. In an attempt to work on similar website versions, Scrapbook (version 1.5)[3] was used to retrieve from the Internet Archive (Kahle 2002) all websites' pages within a distance of one link from the homepage (also included). This technique allowed us to locally store all the web documents and manipulate them easier during the analysis. From the reduced sample (57 websites), five more websites were eliminated due to the impossibility to locate them in the Internet Archive.[4] As a result of all the above, the corpus (hereinafter referred to as 'web corpus I') was finally comprised of 52 French websites. A detailed summary of the contents of web corpus I, including URLs, total number of pages and total number of images retrieved per website, can be found in Appendix A.1.

4.3.1.2 Content extraction and post-processing

The raw web corpus consisted of 2,236 HTML files (see Table 4.1). After an exhaustive clean up, the number of web pages went down to 1,938. Files removed included pages generated automatically by the Internet Archive or Scrapbook during the websites retrieval, as well as pages of less than 5KB size. Since we were mainly interested in observing the instances of image text alternatives available in the web corpus, a content extraction procedure was then designed to isolate `<img>` elements from the rest of the content.

Table 4.1. Descriptive summary of web corpus I built for rule development (training data set)

Main corpus	Pages	`<img>`	No `alt`	Empty `alt`	Non-empty `alt`
Raw	2,236	50,633	9,241	12,238	29,154
Clean	1,938	12,778	1,267	3,566	**7,945**

The 1,938 HTML files were first converted into XLIFF (version 1.2) using Rainbow (version 6.0.26). Rainbow is an Okapi application[5] that is used to perform multiple localisation-related tasks, including file conversion, batch translation or terminology extraction. Rainbow also enables the user to create filters to trigger search and replace actions. Using this feature, a set of regular expressions were applied to each XLIFF file in order to (i) automatically extract the desired information per web page, and (ii) insert this information in a separate CSV file. More specifically, each generated CSV file (one per web page) contained:

- All `<img>` elements of the page and their corresponding `alt` attribute values. Using regular expressions, we were also able to code `<img>` elements

[3] Scrapbook is a Firefox extension to save web pages and manage the resulting collection. http://www.xuldev.org/scrapbook/ Last access: 30th November 2015.

[4] For the last five websites dismissed, no complete version had been stored in the Internet Archive (https://archive.org/) during the period April - July 2011.

[5] http://okapi.sourceforge.net/downloads.html Last Access: 30th November 2015.

according to the alt text type: images with an `alt` attribute, images with no `alt` attribute, or images with an empty `alt` attribute
- The source file of each image (e.g. `./images/landscape.png`)
- The source page from which each `<img>` element was extracted (e.g. contact.html)

The resulting image set was also cleaned up, leading to a final usable corpus of 12,778 `<img>` elements (see Table 4.1). During the data cleaning process, we eliminated images that had been automatically added by the Internet Archive during the websites retrieval. Similarly, images with identical source files and `alt` attribute values used in a similar context within a given web page (or website) were removed. For instance, sometimes a logo with an embedded link to the homepage was placed at the top of each one of the pages retrieved from a website. In such cases, only one instance of the corresponding alt text was kept in the corpus and the rest were discarded since we considered them as duplicates. However, if the same logo was used in the same website multiple times but with other purposes, all alt texts were preserved in the corpus.

Finally, we created a Python script to merge all CSV files into a single .txt file per website to facilitate its analysis during rule development. As a result of the process described in this section, we obtained an alt text training data set containing a total of **7,945 unique text alternatives**. Table 4.1 gives an overview of corpus content before and after it was cleaned up. In Appendix A.1., the same information can be consulted per website.

4.3.2 Rule development procedure

This section describes the methodology we followed to define specific CL rules for text alternatives verification (section 4.3.2.1) and the technology we used to transform them into a set of machine-testable rules (section 4.3.2.2).

4.3.2.1 CL rules definition

The examination of the two corpora built to construct the CL proposal for alt texts verification described in the previous section resulted in two data samples: a general list of recommendations on how to write appropriate text alternatives (see Chapter 3, section 3.3.2), and a set of 7,945 unique text alternatives extracted from a French web corpus.

During the rule development phase, both data samples were jointly consulted, following what scholars working on corpus-driven translation research have called an abductive approach (Saldanha and O'Brien 2014, 62–63). This term describes an analysis process that moves continuously to and from the theory (general statements) and the more concrete level of data (particular instances), in a sort of combination between induction- and deduction-based approaches (ibid).

When applied to the context of our research, this methodology led to the following procedure: for each recommendation found in the literature, the alt text training data set was explored as a source of appropriate and inappropriate image text alternatives examples by means of AntConc[6] (version 3.4.1), a freeware corpus analysis toolkit. The objective was to retrieve the highest number of alt text instances sharing similar linguistic patterns and then observe the web corpus to see if their context of use was also the same (i.e. if the images shared the same function). When this was the case, the instances found were gathered manually and used later to define a corresponding CL rule.

The alt text training data set was primary explored through AntConc's concordance tool, which shows search results in a KWIC (KeyWord In Context) format. Searches performed were often based on either isolated uninflected words or compounds with no more than three elements (mostly prepositional or verbal phrases). When the search yielded unexpected results, the list of recommendations from the literature was checked again in an attempt to look for a reasonable explanation. If discrepancies were found between both data sets, the researcher would rely both on the results of the Access for all accessibility study (Access for All 2011) and on her own expertise on the subject to decide (i) whether the instances of alt text retrieved should be considered as appropriate or inappropriate alt text examples, and (ii) whether a new rule, not covered in the literature, should be created. Therefore, it could be stated that CL rules were inferred from the joint interpretation of existing recommendations and the instances found in the web corpus, or directly from the study of the web corpus alone.

As discussed earlier in this thesis, one of the most popular recommendation in the literature regarding the appropriateness of alt texts is to avoid the use of the word 'image' when describing one, such as 'Image of...', 'The image illustrates...' or 'Image showing...'. It is considered redundant because screen readers already announce the presence of an image to the user. Let it be used here as an illustrative example of the development procedure that has just been described.

To look for instances of alt texts contravening this recommendation in French, the starting point was to search for the word 'image' in the corpus. AntConc was configured to highlight both the word searched and the one immediately after (see Figure 4.1). Notice that the tool also allowed us to observe the word in different positions within the alt text. Searches were performed in an iterative way, as many times as needed during each rule definition process.

Hits emerging from each search were then examined to see if they should all be considered as instances of the same rule. In the search example shown in Figure 4.1, we obtained alt text hits with the following constructions:

a) Image + noun + conjunction + noun: e.g. *Image Air et Terre*

[6] http://www.laurenceanthony.net/software/antconc/ Last access: 30th November 2015.

b) Image + noun: e.g. *Image Symbole*
c) Image + (noun) + verb: e.g. *Image radar montrant…, Image illustrant…*
d) Image + preposition + noun + (punctuation sign): e.g. *Image d'en-tête du site, Image de synthèse: le bâtiment de la halle*
e) Image + preposition + article + noun: e.g. *Image d'une tirelire, Image du bouton permalien*
f) Image + adjective + (punctuation sign): e.g. *Image représentatif, Image promotionnelle: deux femmes face à face…*
g) Article + image + preposition + noun: e.g. *Une image de synthèse de l'avion…*
h) Verb + preposition + article + image + preposition + verb: eg. *Cliquer sur l'image pour ouvrir le lien*

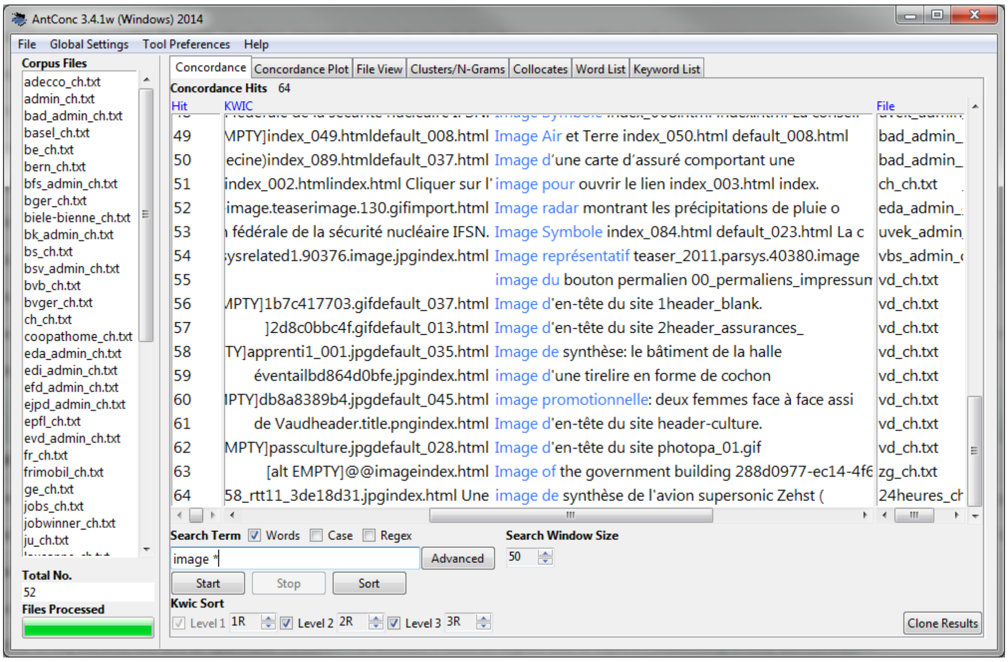


Figure 4.1. AntConc concordance functionality

Alt text instances following the syntactical structures presented in a), b), c) or e) are undoubtedly examples of inappropriate text alternatives for images with a referential value and violate the recommendation already present in the literature. Therefore, they were used to define the corresponding CL rule: *Succinctness in alt text implies the omission of redundant expressions such as "image of", "image illustrating…"* [7] Nevertheless, the redundancy that stems from some of the other hits goes beyond the word level.

[7] In this section we only mention some of the CL rules to illustrate the rule development process. For a more comprehensive description of the final rule set developed, see section 4.4 of the present chapter.

“Image d'en-tête du site”, for instance, is not only redundant because of the use of 'image', but also because it describes a decorative image, which should instead have had an empty `alt` attribute in the first place. Similarly, in “Cliquer sur l'image pour ouvrir le lien”, which is clearly a text equivalent for a functional image, the whole structure itself is redundant. Screen readers already announce to the user that there is an embedded link in the image. What users do not know (at least straight away) is to where the link is taking them, which is precisely the information that should have been included in the alt text.

While the use of the word 'image' is still redundant in these two last cases of alt text inappropriateness, the images they were trying to describe did not have a referential value, but decorative and functional respectively. Hence, the corresponding constructions and alt text instances were taken into consideration during the definition of two other CL rules: 1) *If an image has an embedded link, redundant instructions for operating content should be avoided* and 2) *Decorative images should not be described, unless they have a connotative message that needs to be conveyed to the user.* It thus follows that, for the definition of the CL rules, we were guided not only by the syntactical and lexical cues emerging from the exploration of the alt text training data set, but also by the images value or purpose.

Finally, it is worth mentioning that, while exploiting the French web corpus, we found instances of alt texts in other languages, such as German or English (see, for instance, hit 63 in Figure 4.1: “Image of the government building”). This suggests that text alternatives are sometimes left untranslated or simply introduced in a language different from that of the website. Our CL proposal is based on the premise that text alternatives are written in the language of the page or website their corresponding images appear on. However, for certain rules, English-based constructions and lexicon were also taken into account. Further details on this matter will be given in section 4.4.

4.3.2.2 Acrolinx technology and its rule formalism

Once the rules defined and all the evidence gathered, we used Acrolinx technology to transform them into a machine-processable format. Our ultimate goal was to propose a robust solution that would enable web localisers to check text alternatives against these rules automatically.

Acrolinx is a state-of-the-art checking and authoring support tool. Originally developed at the German Research Center for Artificial Intelligence, this software helps users ensure content readability by checking spelling and grammar adequacy, as well as style and terminology consistency. It is based on a phenomena (i.e. error type) oriented approach to language checking under which an input text is systematically scanned for error candidates (Bredenkamp et al. 2000). This scanning relies on the interpretation of language resources which, in turn, are subdivided into rule files. For each type of error or phenomenon that should be checked, one rule file (.rul) needs to be created. Phenomena are described using a specific proprietary formalism which, as stated by its creators

> [...] permits the specification of error phenomena in terms of regular expressions over complex linguistic objects, represented as feature structures. These feature structures denote the linguistic annotations provided by different underlying NLP components, such as POS and morphological taggers. (ibid)

To translate the CL rules for alt text alternatives verification that we had previously defined while exploring the corpora into Acrolinx rule formalism,[8] we used the Acrolinx Linguistic Integrated Development Environment (ALI), running on the open-source platform Eclipse.[9] Each rule file contained, apart from metadata information (version, author, character encoding, rule ID), two main elements: *object definitions* and *trigger rules* (Acrolinx GmbH 2011). *Objects*, identified with the symbol @, are defined at a word level and can be very precise (e.g. a specific inflection of specific verbs) or very general (e.g. any word identified as a preposition by the POS tagger).

```
@SemanticMeaning_verb   ::= [ MORPH.LEMMA
                        "^(indiquer|symboliser|représenter|
                        décrire|illustrer|présenter|montrer)$"
                        READING.INFLECTION [ tense "present"
                        mode "indicative" ]];

@prep                   ::= [ POS "^(PREP)" ];
```

Triggers specify the patterns (strict sequences of objects or fixed strings) that should be flagged during a check. For instance, if the following trigger rule would be applied, an alt text such as “Image qui montre le coucher du soleil” would be identified by the tool.

```
@image^1 'qui'^2 @SemanticMeaning_verb^3 []*^4
```

Acrolinx formalism also allows the developer to add exceptions (called 'negative evidence') to the trigger rules defined, that is, it provides a mechanism to reduce the chances of the tool returning false positives. For instance, in the following rule file excerpt, it is indicated that Acrolinx should not flag the construction “image + preposition + noun”, presumably specified previously in a trigger rule, when the instance found refers to images from a surveillance camera (e.g. “Des vraies images de caméra surveillance au service de la prévention”).

```
NEG_EV(40)== @image @de 'camera' 'surveillance';
```

[8] This section only outlines the primary defining elements of Acrolinx rule formalism. A more comprehensive description can be found in Gerlach (2015, 39–44).

[9] https://eclipse.org/downloads/ Last access: 30th November 2015.

Similarly, trigger rules can be followed by correction suggestions that will then be proposed to the user in the check report (see section 4.5). The following rule file excerpt suggests that what is referred to by the objects `$prep`, `$obj`, `$image` and `$punct` should be deleted, and that what appears immediately after should start with an initial upper case (e.g. if this suggest rule was to be applied, the alt text "Sur cette image: un coucher de soleil" would be flagged and the suggestion would be "Un coucher de soleil").

```
SUGGEST(10) == $prep $obj $image $punct $tok ->
{suggest: $prep -> '',
  $obj -> '',
  $image -> '',
  $punct -> '',
  $rest -> $rest/initialUpperCase; }
```

All these Acrolinx rule file elements served to formalise the syntactical and lexical restrictions we had identified for each CL rule after the corpora exploitation. During the Acrolinx rule development process, we adopted a trial and error approach: rules were tested on a list of alt texts extracted from the training data set; if any of the inappropriate text alternatives were not flagged, objects, trigger rules and exceptions to the rules were refined until obtaining the desired number of inappropriate alt texts was identified. The same method was applied to propose and improve suggestion rules. This process lead to the definition and formalisation of 40 rules, which is the average number of rules covered by some of the most well-known controlled languages (O'Brien 2003; Pool 2006).

The fact that it is impossible to predict what will be represented in an image posed major challenges during the development of Acrolinx rules. The level of formalisation achieved using Acrolinx technology thus varied depending on the nature of each CL rule. Certain rules could not be categorically associated to specific lexical or syntactical restrictions because, as will be explained in the next section (4.4), they were based on the alt text (and image) function, a parameter which cannot be easily deciphered by CL checking technology (O'Brien 2006, 23). In these cases, lexical cues were defined as objects and then used as triggers in Acrolinx rule files, while examples of appropriate alt texts were added as exceptions.

An example of such a rule could be *If social media and mailing logos have an embedded link, a description of the related action should be provided.* This type of image can trigger multiple actions, such as open an e-mail client to send a message, visit the social media page of an institution or share the information being consulted by the user on his own social media page. The strategy adopted to render this rule machine-testable was to gather and formalise all the alt text examples of inappropriate alt texts for social media and mailing logos found in the corpora. These included the names of social media or information sharing applications (e.g. "Facebook", "Pinterest", "e-mail", "Fil RSS", etc.) and vague or incomplete action descriptions such as “Envoyer à”, “Partager”, “Partager cette page” or “Abonemment RSS”. To compensate for the low level of precision achieved in the trigger rules, two measures were adopted: on one hand, appropriate alt texts containing the aforementioned lexical cues were included in the rule file as negative evidence (e.g. “Partager cette page sur Facebook” or “S'abonner aux flux RSS”). On the other hand, the rule was complemented with comprehensive support information about how to interpret the flag and applied the rule. This technique will be explained in more detail in section 4.5, once we have presented the 40 rules and its categorisation (section 4.4).

4.4 Rule sub-sets for text alternatives verification

In section 4.2 we have discussed the classification and properties of CLs, but little mention has been made of the categorisation of the rules of which they are composed of. In general terms, Nyberg et al. (2003, 252) distinguish between proscriptive and prescriptive rules: the former describes which structures are not allowed, whereas the latter explicitly describes those which are allowed. As we have implicitly shown in the previous section, the rules we have formalised using Acrolinx technology are designed to flag incorrect image text alternatives rather than highlighting to the user which alt texts are appropriate. In this sense, we could say that our set mainly includes proscriptive rules. Nevertheless, some of them are formulated in a prescriptive manner, hence the need of a more fine-grained rule categorisation.

By adopting a more linguistic perspective, O'Brien (2003) puts forward three main categories for CL rule classification, namely *lexical*, *syntactic* and *textual*, with two sub-categories for the latter: *text structure* and *pragmatic*. The author indicates that this classification, inspired by the study of previous work and the analysis of eight Controlled English rule sets, is grounded on the primary functionality of the rule. *Lexical rules* seek to influence word selection or to influence meaning by word selection (e.g. rules covering vocabulary or abbreviation usage, spelling), and *syntactic rules* aim at influencing syntax (e.g. rules specifying which pronouns, quantifiers, conjunctions can be used; date and number formatting, word combination restrictions). Lexical and syntactic rules are very popular within MOCLs because they help reduce ambiguity and textual complexity. However, these are less common in HOCLs, which often have a higher percentage of textual rules. We argue that this is also the case in our CL proposal.

O'Brien (2003) divides textual rules in two sub-categories. She suggests that rules whose purpose is to influence the information load in the text or the text graphical layout should be considered as *text structure rules*. These would include, among others, rules specifying sentence or paragraph length, paragraph or information structure, or indications on capitalisation. Nonetheless, if the primary function of a rule is to influence text purpose or reader response to the text, then it is classified as a *pragmatic rule.*

Of special interest to us is one of the sub-categories for pragmatic rules that O'Brien proposes: *text purpose*, where rules specifying that particular sub-structure are written for one purpose and not another. This notion of purpose or intention (i.e. what the communicative goal of the text is) leads Hartley and Paris (2001, 313) to make a further distinction within this type of pragmatic rules and refer to (i) rules that apply to *procedures*, where the goal is to instruct, and (ii) rules that apply to *descriptions*, where the goal is to inform.

Our CL rule classification is in line with Hartley and Paris' proposal, since it is primary grounded on the communicative goal of the image that the text alternative is acting as a substitute for. It was thus decided to organise rules around two main sub-sets: *descriptive* (section 4.4.1) and *functional* (section 4.4.2). A third sub-set was created to cover rules aimed at detecting *uninformative* alt texts which do not necessarily provide any hints about the image purpose (section 4.4.3).

4.4.1 Sub-set A: Descriptive content

Sub-set A contains 14 rules that advocate the proper accessibilitation of images with a referential or descriptive value within a web page. Table 4.2 shows a brief description for each rule, together with an example of an appropriate text alternative and an example of an alt text that would be flagged by Acrolinx. The 14 rules under this sub-set are organised around three categories, based on the actions implicitly required to the alt text creator:

a) *Provide more information about the image* (5 rules)

These rules suggest alt text creators to add content to descriptive text alternatives for the sake of specificity, as well as for a better assistive technology performance. The two first rules (DA1-screenshot and DA2-complex) indicate that screenshots and complex images (e.g. charts, maps) should be identified as such in the alt text and then described comprehensively. A third rule (DA3-galery) urges alt text creators to describe images in a slide show or an image gallery as if they would be inserted into the main body of the text. A common error in these three cases is to include single-word alt texts such as “capture d'écran”, “diagramme à barres” or “galerie photo”.

Two more rules belong to this category. Inserting abbreviations and acronyms in alt texts should be avoided and their canonical form should be used instead (DA4-acronym). If this is not possible due to length restrictions, acronyms should be written with a space in between letters so that the assistive technology can spell it correctly. Similarly, descriptive text alternatives should end with a full-stop

(DA5-punct) for the screen reader to use the appropriate tone of voice and not link the text alternative with the textual content that appears immediately after.

b) *Delete or transform content to avoid redundancy* (5 rules)

Five rules in this sub-set aim at simplifying the text to prevent blind users from listening to repetitive information when encountering the presence of an image, which is announced out loud by the screen reader as “graphic” or “image”. The five rules are concerned with the use of the words “icône” (DDR1-icon), “image” (DDR2-image), “logo” (DDR3-logo), “photo” (DDR4-photo) and “symbole” (DDR5-symbol) respectively, which in most cases are dispensable. Although one might argue that these rules could have been regrouped into a single one, it is worth noting that these words are not always interchangeable within alt texts (see Chapter 3, section 3.3.2). It was thus preferred to formalise them individually in order to facilitate the definition of negative evidence (exceptions) in Acrolinx rules, as well as to develop more specific written guidance (rule help) for each redundancy case.

c) *Delete or transform content in favour of text type adequacy* (4 rules)

The last four rules of Sub-set A are intended to modify the value of `alt` attributes so that they adhere to the general conventions of this particular text type. The first rule aims at preventing alt text creators from introducing image copyright information in the alt text (DDA1-credits). Developers are often obliged to add this information somewhere in the page and do not consider it necessary or aesthetically appealing to render it visible on the screen. However, alt texts are not meant for such purpose. Image and photo credits offer complementary information that should also be accessible to non users of assistive technology. The most recent recommendation suggests adding this information in the HTML 5 `<figcaption>` element.[10]

Rules DDA2-maj and DDA3-length cover two conventions related to alt texts structure and match two of the textual rules' sub-categories put forward in O'Brien (2003): capitalisation and sentence length. The former specifies that the `alt` attribute value should not be fully written in upper case. The latter establishes that alt texts should not exceed 100 characters. Text length has not only been largely discussed in text alternatives related literature (see Chapter 3, section 3.3.2), but it is also one of the most popular restrictions in E2R on the Web guidelines, as well as in CLs designed for readability and translatability purposes (O’Brien 2006,

[10] This suggestion emerged from an informal conversation through WebAIM's mailing list between the researcher and one of the editors of the HTML 5 W3C Recommendation (Hickson et al. 2014). The message thread can be consulted at:
http://webaim.org/discussion/mail_thread?thread=6667&id=26930#26930 Last access: 2nd December 2015.

24). It is generally suggested that descriptive sentences should be 25 words long and instructional sentences should not contain more than 20 words (ibid). An alt text length restriction of 100 characters is equivalent to approximately 20 words. Finally, the last rule of sub-set A indicates that alt text creators should avoid orality markers (DDA4-oral), such as “In the photo there is a...” or “In this image we observe...”

Table 4.2. Sub-set A: List, descriptions and examples (fr) of the rules (14) to identify inappropriate descriptive content

Category	Rule ID	Description	Example of appropriate alt text	Example of alt text flagged in Acrolinx
Addition: description of content (5 rules)	**DA1-screenshot**	Screenshots should be identified as such and described comprehensively.	*Capture d'écran : Acrolinx Batch Checker 2.7.0 - Options de document.*	*Capture*
	DA2-complex	Complex images (e.g. charts, maps) should be described comprehensively.	*Diagramme de flux : Processus de localisation web.*	*Diagramme de flux*
	DA3-galery	Images in a slide show, carousel or image gallery should be described as if they were placed in the main body of the page.	*Un avion prêt à décoler.*	*Galerie*
	DA4-acronym	Abbreviations and acronyms should not be included in the text alternative. If necessary, write spaces in between letters (e.g. C L instead of CL).	*Pannier avec des fruits exotiques comme la goyave.*	*Pannier avec des fruits exotiques, ex. la goyave*
	DA5-punct	Descriptive text alternative should end with a full-stop.	*Un avion prêt à décoler.*	*Un avion prêt à décoler*
Deletion: redundancy (5 rules)	**DDR1-icon**	Expressions including the word “icon” should be used with care.	*Ressources multimédia*	*Icône multimédia*
	DDR2-image	Succinctness in alt text implies the omission of redundant expressions such as “image of”, “image illustrating…”	*Un avion prêt à décoler.*	*Une image d'un avion prêt à décoler*
	DDR3-logo	An excessive use of the word “logo” should be avoided.	*Word Wide Web Consortium*	*W3C logo*
	DDR4-photo	Succinctness in alt text implies the omission of redundant expressions such as “Photo:”, “Photo of...”	*Un avion prêt à décoler.*	*Photo : un avion prêt à décoler*

Table 4.2. Sub-set A (continued)

Category	Rule ID	Description	Example of appropriate alt text	Example of alt text flagged in Acrolinx
	DDR5-symbol	Expressions including the word "symbol" should be used with care.	*Énergies renouvelables.*	*Symbole énergies renouvelables*
Deletion: adecuacy (4 rules)	**DDA1-credits**	Image copyright information should not be included in the alt text.	*Un avion prêt à décoler.*	*© Studio Photos*
	DDA2-maj	The use of upper-case should be avoided.	*Un avion prêt à décoler.*	*UN AVION PRÊT À DÉCOLER*
	DDA3-legth	Text alternatives that are too long to be retrieved by assistive technology should be avoided (max. 100 characters).	---	----
	DDA4-oral	Orality markers such as "in the image we can see" should be avoided.	*Un avion prêt à décoler.*	*On peut voir un avion prêt à décoler*

4.4.2 Sub-set B: Functional Content

Sub-set B contains 18 rules that seek to offer guidance on how to write more appropriate text alternatives for images with a functional value within a web page. As discussed in Chapter 3 (section 3.2), by functional images we understand (i) images used to initiate actions rather than to convey information, as well as (ii) images that provide structure (e.g. indicate a thematic break) to the web document.

Table 4.3 shows a brief description for each rule, together with an example of an appropriate text alternative and an example of an alt text that would be flagged by Acrolinx. The 18 rules under this sub-set are organised around four categories which are based, as in the case of Sub-set A, on the actions implicitly required to the alt text creator:

a) *Provide more information about the image function* (5 rules)

Sometimes interactive images are identified as such, but their alt texts remain vague or incomplete. We have defined four rules to urge alt text creators to provide more detailed information about the content that is presented to the user once he clicks on an image. Users may be taken to (i) a different website or web page within the same website being browsed, (ii) a new section or element within a list, (iii) an increased version of the image, or (iv) an expanded version of a given piece of content. Rules FAC1-accueil, FAC3-more, FAC4-zoom and FAC5-nav and cover the identification of incomplete alt texts in these four contexts of use (see examples in Table 4.3). Within this category, we also included a special rule on how to render CAPTCHA tests accessible. The rule FAC2-captcha specifies that text alternatives for CAPTCHA images cannot reproduce the textual content shown in the image, and suggests describing instead that which is the accessible alternative to this visual test.

b) *Properly describe actions related to the image function* (4 rules)

Images that trigger a specific action when clicked upon should be described in an informative and efficient manner. This implies providing direct instructions, for which verbal forms are often preferred against noun phrases.[1] The rule FAA1-files specifies instructional patterns that should be used in alt texts for images enabling users to save or download content. The rule FAA4-print is based on the same principle and instructs alt text creators to use verbal forms in order to describe actions related to printing. In both cases, the rules also recommend specifying which type of content (e.g. title, format) is being saved, downloaded or printed.

[1] Our rules prescribe the use of the infinitive as an imperative (e.g. “Télécharger la liste de produits” instead of the conventional imperative form (“Téléchargez la liste de produits”). This is also the preferred form in GIFAS Rationalised French CL (Barthe et al. 1999, 223), the French equivalent to AECMA Simplified English, for giving instructions because it is impersonal.

The rule FAA3-social, already introduced as an example in section 4.3.2.2, is intended to help alt text creators adequately describe the action initiated when interacting with logos of social media or information exchange applications (e.g. subscribing to an RSS feed). As in rules FAA1-files and FAA4-print, the use of noun clusters is ruled out in favour of verbal forms (e.g. “S'abonner aux flux RSS” vs. “Abonemment flux RSS”). Finally, the rule FAA2-window specifies that not only should users be informed through the alt text about a change of web context, but also that this change should be explained in full to avoid any misunderstanding. This information needs to be inserted between brackets after the description of the action initiated through the functional image.

c) *Delete or transform content in favour of text type adequacy* (7 rules)

The first two rules within this category of Sub-set B concern the adequacy of the link destination's description provided in the alt text of a functional image. On one hand, rule FDA1-lang indicates that text alternatives of images that serve as language selectors should not include a denotative description of the image (often the flag of one of the countries in which the language is spoken) but just a brief indication of the website's language version the user is being taken to (ideally, the name of the language in which the content will be presented). On the other hand, FDA2-url proscribes the insertion of full URLs in the alt text. This practice should be avoided because web addresses are not always representative of the type of content or services that they offer and can be misleading to the user.

Another two rules within this category aim at improving the appropriateness of images used to visually accentuate the web content macrostructure, such as horizontal or vertical lines to divide the web page into different sections. Images used as list bullets are also functional, since they serve to organise the content in a list. The rule FDA3-structure aims at avoiding denotative descriptions of structural images, while rule FDA4-punct proscribes the sole use of punctuation signs as text equivalents for the said images (e.g. if images representing bullet points are included in front of each element in a list, the alt should read “new item in the list” and not “bullet point” or “ • ”).

As indicated in W3C documentation referring to Success Criteria 1.3.3 of WCAG 2.0 (Caldwell et al. 2008), people who are blind or have low vision may not be able to understand information if it is conveyed by shape and/or location. Sometimes this information is presented as an image of text or a button. The last three rules of this category urge alt text creators to avoid overreliance on sensory characteristics to instruct users on how to interact with web content. Rules FDA5-lieu, FDA6-forme, FDA7-colour are intended to rule out references in text alternatives to position, shape or size, and colour respectively (see examples in Table 4.3).

d) *Delete or transform content to avoid redundancy* (2 rules)

The last two rules of Sub-set B aim at simplifying the alt text to prevent blind users from listening to superfluous information related to functional images. The rule FDR1-action proscribes the use of repetitive instructions on how to interact with the image. Typically, these include verbal constructions such as “Click here for + verb” or “Go to + name of the website”. The screen reader already announces the presence of a form button or a link to the user, so there is no need to insert redundant expressions in the alt text. For this same reason, the rule FDR2-link specifies that the existence of an embedded link in the image should not be indicated in the text alternative. Examples illustrating both rules are included in Table 4.3.

4.4.3 Sub-set C: Uninformative content

Rules aimed at identifying uninformative image text equivalents are compiled under sub-set C. Table 4.4 shows a brief description of each rule, together with an example of an appropriate text alternative and an example of an alt text that would be flagged by Acrolinx. Although not directly associated with the communicative function of the image, the eight rules from this third sub-set allow alt text creators to quickly detect inappropriate description proposals, and invite them to propose a new text equivalent taking into account the ultimate goal of the image.

A common mistake observed both in the literature and in the training web corpus was the presence of image metadata in the text alternative. The rule U1-format specifies that the `alt` attribute value of an image should not contain the image file name or solely indicate the image file extension. Similarly, indicating the image file size is proscribed by rule U5-taille. File size indications should only be included in the alt text as complementary information if the image triggers a save and/or download action.

Non-alphabetical characters should be used with care within text alternatives. Non meaningful character sequences (rule U3-chain), ASCII art and text emoticons (rule U7-emoticon) and character entities (rule U8-entities) should be avoided for two reasons: first, they do not provide purposeful information about the image; and second, they can lead to an undesired degraded performance of assistive technology. Another form of uninformative alt text is covered in rule U6-space, which specifies that an `alt` attribute value should not be exclusively composed of a space or a sequence of spaces.[2]

Rule U2-placeholder rules out the use of isolated undescriptive words or word sets serving as placeholders, often introduced automatically by Content

[2] See Chapter 3 (sections 3.3.1.3) for a more detailed discussion about the use of spaces in empty `alt` attributes.

Management Systems (CMS). For this reason, in our attempt to transform this guideline into a machine-testable rule through Acrolinx, we have not only considered French words or expressions, such as “image inexistante” or “insérer du contenu ici”, but also their English equivalents.

Finally, a general rule specifying that decorative images should not be described was included under this third sub-set. To define Acrolinx objects and trigger rules for this guideline, we followed a similar approach to the one that was presented in section 4.3.2.2 for FAA3-social. We gathered and formalised all the alt text examples of inappropriate alt texts for decorative images present in the two corpora explored, and reinforced the rule with rule exceptions and interpretation support. As in rule U2-placeholder, both English and French constructions were considered. These include, among others, “thumbshot”, “rubrique”, “cover photo”, “teaser”, “aguichage”, “bannière”, “miniature”, “vignette” or “banner”.

Table 4.3. Sub-set B: List, descriptions and examples (fr) of the rules (18) to identify inappropriate functional content

Category	Rule ID	Description	Example of appropriate alt text	Example of alt text flagged in Acrolinx
Addition: description of content (5 rules)	**FAC1-accueil**	If the purpose of an image with an embedded link is to take the user to a specific page or website, the name of the website or web page owner should be included.	*Page d'accueil du site de l'Université de Genève*	*Page d'accueil*
	FAC2-captcha	An accessible alternative to CAPTCHA should be described in the alt text.	*Écouter le code de vérification si vous ne pouvez pas voir l'image*	*mots test CAPTCHA*
	FAC3-more	Alt text for images enabling the user to expand information should be specific.	*Savoir plus sur les événements à Genève en été*	*En savoir plus*
	FAC4-zoom	If, upon an image click, zooming is allowed, a short image description and instructions on how to get back to normal position should be provided.	*Zoom sur le diagramme de flux (touche ESC pour fermer la fenêtre)*	*Zoom sur l'image*
	FAC5-nav	When a navigational symbol is included, the sense of the navigation should be specified, as well as the items through which the user is browsing.	*Lire l'article suivant*	*Suivant*
Addition: description of action (4 rules)	**FAA1-files**	Text alternatives of images which allow the user to perform an action related to a file should describe it using a verbal form instead of just indicating document type or name.	*Télécharger l'horaire en PDF*	*fichier PDF*
	FAA2-window	If information about change of web context appears in the alt text, it should be explained in full.	*Consulter l'horaire en ligne (la page s'ouvrira dans une nouvelle fenêtre)*	*Horaire (nouvelle fenêtre)*

Table 4.3. Sub-set B (continued)

Category	Rule ID	Description	Example of appropriate alt text	Example of alt text flagged in Acrolinx
Addition: description of action (4 rules)	**FAA3-social**	If social media and mailing logos have an embedded link, a description of the related action should be provided.	*Partager cette page sur Facebook*	*Facebook*
	FAA4-print	Text alternatives of functional images which allow the user to perform an action related to printing should describe it using a verbal form and indicate the document type or name (e.g. Print bus schedule).	*Imprimer l'horaire de la bibliothèque*	*Version imprimable*
Deletion: adequacy (7 rules)	**FDA1-lang**	Alt text of language selectors should include just the name of the language and not a denotative description of the image.	*Español*	*Drapeau de l'Espagne*
	FDA2-url	Full URLs should not be included in the alt text.	*Site internet du World Wide Web Consortium*	*www.w3c.org*
	FDA3-structure	Images that provide structure to the web document should not include a denotative description. Instead, they should either have an empty alt or reflect the macrostructure of the web content.	*Nouvelle section : partenaires du projet*	*barre verticale*
	FDA4-punct	Images that provide structure to the web document should not be described with punctuation signs. Instead, they should either have an empty alt or reflect the macrostructure of the web content.	*Nouvelle section : partenaires du projet*	-------------------
	FDA5-lieu	Instructions provided in the alt text for understanding position of content should not rely solely on sensory characteristics.	*Consulter l'information sous la rubrique 'Horaires'*	*Voir colonne à droite*

Table 4.3. Sub-set B (continued)

Category	Rule ID	Description	Example of appropriate alt text	Example of alt text flagged in Acrolinx
Deletion: adequacy (7 rules)	**FDA6-forme**	Instructions provided in the alt text for operating content should not rely solely on sensory characteristics of components, such as shape or size.	*Accepter*	*Appuyer sur le button rond*
	FDA7-colour	Instructions provided in the alt text for operating content should not rely solely on sensory characteristics of components, such as colour.	*Accepter*	*Cliquer sur le button rouge*
Deletion: redundancy (2 rules)	**FDR1-action**	If an image has an embedded link, redundant instructions for operating content should be avoided.	*Accepter*	*Cliquer pour accepter*
	FDR2-link	If an image has an embedded link, redundant expressions including the word "link" should be avoided.	*Site internet de la Faculté de traduction et d'interprétation*	*Lien : site de la FTI*

Table 4.4. Sub-set C: List, descriptions and examples (fr) of the rules (8) to identify uninformative content

Rule ID	Description	Example of appropriate alt text	Example of alt text flagged in Acrolinx
U1-format	Image file name and format should not be repeated in the alt text.	*Un avion prêt à décoler.*	*image.jpg*
U2-placeholder	Isolated undescriptive words (e.g. image, page) or word sets serving as placeholders should be avoided.	*Un avion prêt à décoler.*	*Insérer description ici*
U3-chain	Strings of non-alphabetical characters (such as numbers or punctuation signs) combined with letters that are not presented in a meaningful way should be avoided.	*Un avion prêt à décoler.*	*degD87_dieo23*
U4-decor	Decorative images should not be described, unless they have a connotative message that needs to be conveyed to the user.	*[empty `alt` attribute]*	*spirale*
U5-taille	Images or downloadable files' size information should not be the only descriptive text included in the alt text.	*Télécharger l'horaire en PDF (345 Mb)*	*345 Mb*
U6-space	A space or double spaces alone should be avoided.	*[empty `alt` attribute]*	
U7-emoticon	ASCII art and text emoticons to represent the information depicted on the image should be avoided.	*Lire l'article suivant*	*==>*
U8-entity	Character entities should be avoided.	*Accepter et envoyer*	*Accepter envoyer*

4.5 Rule application

Thus far we have described Acrolinx rule formalism and the 40 CL rules defined for alt texts verification. This section aims to explain how web localisers would check image text alternatives against these rules by means of Acrolinx technology. To this end, we first present the Acrolinx client for websites verification used in the experimental study of this thesis, Acrolinx Batch Checker (section 4.5.1). Then, we describe the Acrolinx rule implementation support developed to complement Acrolinx rules (section 4.5.2).

4.5.1 Acrolinx Batch Checker

As we announced earlier in this chapter, Acrolinx software helps authors find inconsistencies in writing style, spelling, grammar and terminology by looking at the linguistics of language. To do so, it relies on a server-client architecture. A central server hosts (i) the main checking engine, which manages the NLP components used for the linguistic annotations, and (ii) the language resources (i.e. the rule files) (Gerlach 2015, 27).

Acrolinx checking functionality can then be integrated through plug-ins in multiple authoring tools, including among others GoogleDocs and GoogleSheets; Microsoft Word, PowerPoint and Excel; and Adobe FrameMaker and InDesign. Through these plug-ins, authors are allowed to make corrections interactively: the checking client flags the words or constructions requiring attention and, for each flag, a contextual menu provides the user with more information about the error as well as correction suggestions, if available (ibid).

The Acrolinx IQ Batch Checker is a program designed to check text, HTML, XML or PDF files directly in a directory or a directory structure, without opening them in an editor. As opposed to other Acrolinx clients, the Batch Checker allows users to check multiple documents in a single process for compliance with all defined rules. Within the framework of this thesis, we found that the Acrolinx Batch Checker was the most convenient client to serve as a complementary localisation quality assurance tool because (i) it can be used as an independent desktop application, and (ii) it allows users to check both online websites and locally stored HTML files in batch mode.

Figure 4.2 shows the main Acrolinx Batch Checker interface, where the user can define the checking options. In the *Document Options* tab, the user can indicate if he wants to check online or offline files, the file type, the character encoding and the Context Segment Definition (CSD) (see Figure 4.2, left part). Acrolinx CSD files specify sentence-break elements, elements to include or exclude when a check is performed, and the attributes that have values which should be sent to the server for checking (Acrolinx GmbH 2012, 44–45). This feature is especially useful in localising the applicability of CL rules to a certain section of a document. Since we were interested in checking HTML files, the CSD provided by default for this file type was thus customised to only focus on the `alt` attribute values of `<img>` elements. This enabled us to introduce greater context sensitivity (Hartley and Paris 2001, 214), particularly relevant in the semi-automation of pragmatic rules checking.

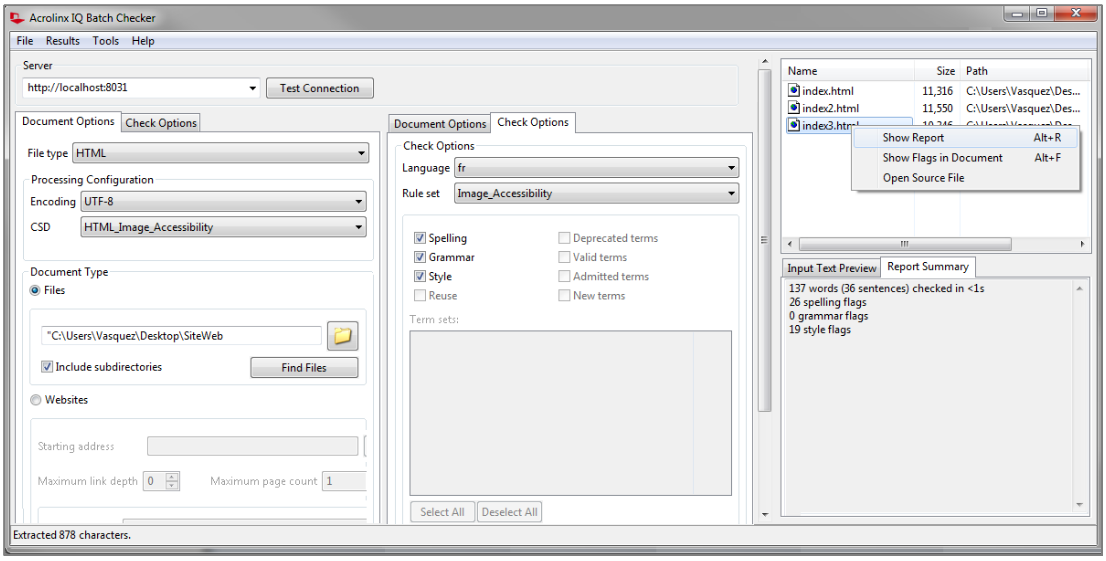


Figure 4.2. Acrolinx Batch Checker interface, where the user can define the checking options (two tabs merged in one single figure)

The *Check Options* tab allows the user to indicate the language of the document(s) that will be checked, the rule set that should be used and which language areas the check should focus on: spelling, grammar, style or use of terminology (see Figure 4.2, middle part). The Acrolinx rule set developed for alt text verification was included within the style category. Spelling, grammar and terminology features were not used in the context of this thesis so they will not be described further.

The Acrolinx Batch Checker can identify different rule violations within the same text alternative. For example, an alt text that reads "Twitter logo.jpg" would be flagged for contravening rules FAA3-social and U1-format. In this sense, it is worth mentioning that the checker was configured to only verify non-empty `alt` attributes for rule compliance. Since most web accessibility evaluation tools already identify the absence of `alt` attributes or warn the user about the presence of empty `alt` attributes (see Chapter 3, section 3.4), it was decided that Acrolinx technology should be presented as a complement to more technical tools, which currently lack a deeper linguistic analysis functionality.[1]

[1] It could have been possible to develop a style rule to flag zero-length `alt` attributes (Acrolinx GmbH 2012, 55). However, in order to identify `<img>` elements without an `alt` attribute, a different

After running a batch check, the results per file are presented to the user in the form of a report summary (see Figure 4.2, right part) and an extensive report in XML format that can be easily visualised in the web browser. As shown in Figure 4.3,[2] this more comprehensive report provides a summary of the rules that have been violated, indicating the number of errors found per rule. Similarly, each flag reported is highlighted individually to the user, together with the name of the rule contravened.

As indicated by Nyberg et al. (2003, 252), in addition to pointing out violations of the CL, a checker may offer help to the user on three levels: (i) provision of fully automatic correction, (ii) specific propositions for correction to be selected and confirmed by the user, and (iii) general advice on making text conform to the rules. The Acrolinx Batch Checker includes a 'Check and Apply' feature that allows the user to automatically correct certain kinds of style issues. Although Acrolinx automatic rule application capability has been successfully implemented in previous research work (Gerlach 2015), this possibility was ruled out from our study because most of the CL rules we proposed are pragmatic in nature and therefore require human judgement. In addition, the Acrolinx Batch Checker 'Check and Apply' feature is available for text, XML and GLM files, but not for HTML-based documents.

In the context of this thesis, we have relied on the other two forms of correction support to offer guidance to web localisers: suggestions and general advice. As announced earlier in section 4.3.2.2 and illustrated in Figure 4.3, we were only able to provide propositions for correction for certain rules. Some suggestions can be implemented right away (e.g. the elimination of the word 'logo' in “Logo VNU (Volontaires des Nations Unies)”), while others require some intervention from the user (e.g. replacing a placeholder with the appropriate piece of content, as in the suggestion “Partager cette page sur [nom du réseau social]”). General advice on how to interpret the rules and adapt the text flagged is accessible to the web localiser directly from the checking report. More information about this feature is provided in the next section.

CSD should have been developed. Having two CSDs means that web localisers would have needed to analyse each web page twice, thus rendering the checking process less smooth and efficient.

[2] Figure 4.3 shows a simplified version of the XML report, where only information concerning style rules is presented. The extended report would also normally include administrative and checking information, such as the date and time of the check and a summary of the checking options chosen prior to the analysis.

Acrolinx Checking Report

acrolinx

Style ▼ Show | Hide

Summary

Avoid describing decorative images	5
Describe actions related to social media and mailing appropriately	5
Include the name of the website or web page owner	2
Avoid using the word 'logo' in the image text alternative	3
Avoid describing images used for structuring web page content	4
Avoid copyright information in the image text alternative	3
Avoid using the word 'image' in the image text alternative	2
Avoid using placeholders in the image text alternative	3
Describe the action related to these files appropriately	3
Describe the action 'Print' appropriately	3

Details

Order by Rule

Avoid describing decorative images
*Photo **Thumb***

Describe actions related to social media and mailing appropriately
Partager → ***Partager cette page sur [nom du réseau social]***

Include the name of the website or web page owner
Accueil → ***Accueil de/du/de la [proprietaire du site internet]***

Avoid using the word 'logo' in the image text alternative
***Logo** VNU (Volontaires des Nations Unies)* → *VNU (Volontaires des Nations Unies)*

Avoid describing images used for structuring web page content
ligne de séparation

Avoid copyright information in the image text alternative
***Photographe :** Jason Alden/Bloomberg*

Avoid using the word 'image' in the image text alternative
Image d'un bus DT → ***Un bus DT***

Avoid using placeholders in the image text alternative
#

Describe the action related to these files appropriately
*Rapport du comité ad-hoc **(.PDF)*** → *Rapport du comité ad-hoc **(Télécharger PDF)***

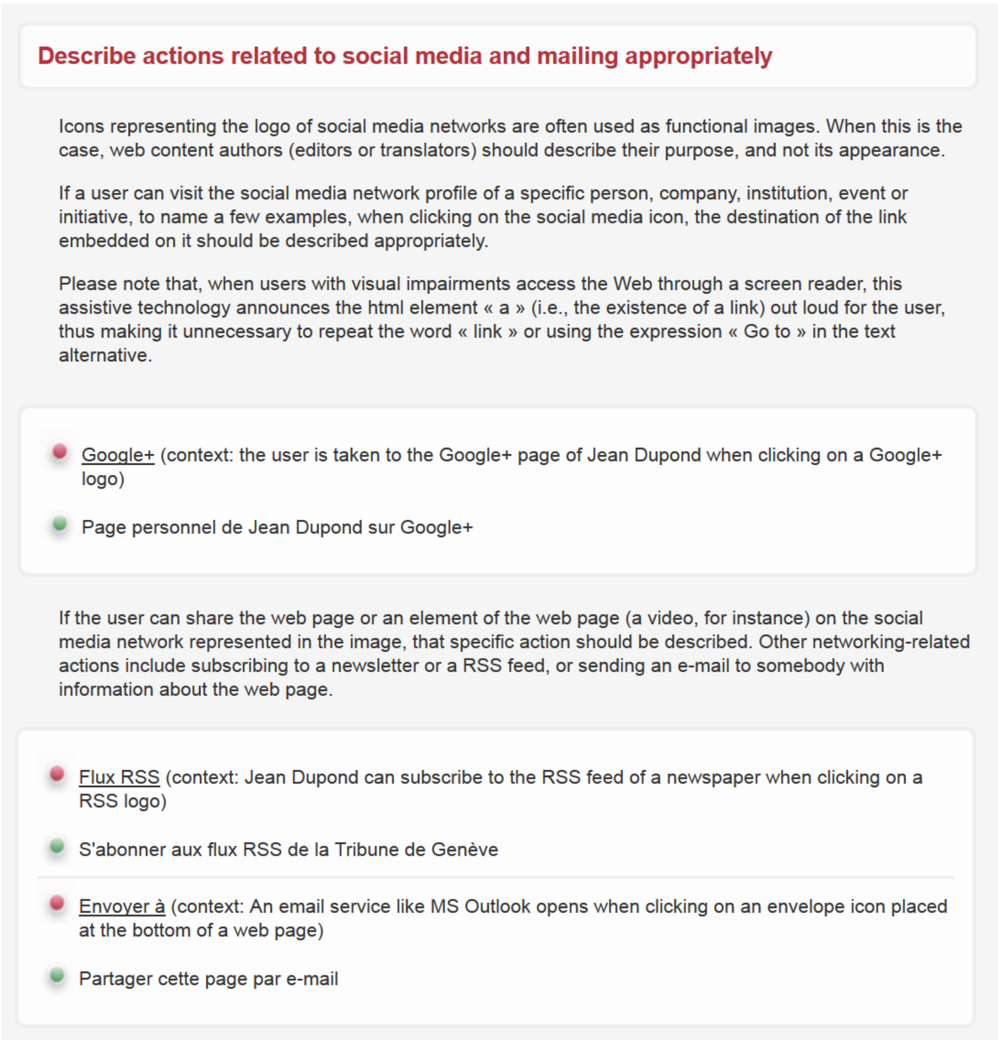
Describe the action 'Print' appropriately
format d'impression → ***Imprimer ce/cette document/page***

Figure 4.3. Example of an Acrolinx Checking Report

4.5.2 Acrolinx rule help

Web accessibility evaluation (WAE) tools present limitations not only in terms of the depth and completeness of the alt text analysis they carry out, but also with regard to error documentation. Both limitations can be successfully covered when using CL software, which focuses on two content dimensions: *the linguistic content*, that is, the rules and their application, and the *didactic content*, how errors are shown and explained to the user (Reuther and Schmidt-Wigger 2000).

In the event that web localisers are requested to apply a certain set of rules they have not been previously trained on, they should be supported in the learning process. When using Acrolinx, this can be made possible through rule help files: customisable XML-based documents that are linked to each rule file and can be comfortably visualised in the web browser. We shall use the example presented in Figure 4.4 to explain the type of information contained therein.



Describe actions related to social media and mailing appropriately

Icons representing the logo of social media networks are often used as functional images. When this is the case, web content authors (editors or translators) should describe their purpose, and not its appearance.

If a user can visit the social media network profile of a specific person, company, institution, event or initiative, to name a few examples, when clicking on the social media icon, the destination of the link embedded on it should be described appropriately.

Please note that, when users with visual impairments access the Web through a screen reader, this assistive technology announces the html element « a » (i.e., the existence of a link) out loud for the user, thus making it unnecessary to repeat the word « link » or using the expression « Go to » in the text alternative.

- Google+ (context: the user is taken to the Google+ page of Jean Dupond when clicking on a Google+ logo)
- Page personnel de Jean Dupond sur Google+

If the user can share the web page or an element of the web page (a video, for instance) on the social media network represented in the image, that specific action should be described. Other networking-related actions include subscribing to a newsletter or a RSS feed, or sending an e-mail to somebody with information about the web page.

- Flux RSS (context: Jean Dupond can subscribe to the RSS feed of a newspaper when clicking on a RSS logo)
- S'abonner aux flux RSS de la Tribune de Genève
- Envoyer à (context: An email service like MS Outlook opens when clicking on an envelope icon placed at the bottom of a web page)
- Partager cette page par e-mail

Figure 4.4. Example of an Acrolinx rule help file as shown in a web browser

REMINDER: Image text alternatives ("alt text") in a nutshell

As noted by W3C (2014a), text alternatives ("alt text") convey the purpose of an image. Text alternatives are used by people who cannot see the image. For example, people who are blind and use screen readers can hear the alt text read out; and people who have turned off images to speed download or save bandwidth can see the alt text.

The text should be functional and provide an **equivalent user experience**, not necessarily describe the image.

You don't usually see the alt text on a web page: it is in the web page markup, but **every image should include it**. It is as easy as adding the attribute "alt" to an existing "img" element.

This alt decision tree, also developed by the W3C (2014b), can be of help to analyse the purpose of an image before adding its text alternative.

- If an image is just decorative and people don't need to know about the image, then it should have null alt (alt=""), without space between the inverted commas.
- If an image conveys information useful for interacting with or understanding the web page content, then it needs alternative text.

References

W3C (2014a). *Easy Checks - A First Review of Web Accessibility* (Available online. Last visited: 30 November 2014).

W3C (2014b). *Web Accessibility Tutorials : An alt Decision Tree* (Available online. Last visited: 29 November 2014).

Figure 4.4. (continued)

Rule help files were developed to provide web localisers with more comprehensive information about the pragmatic value of the images they are trying to render accessible. They serve to complement the short rule description that appears in the checking report and normally consist of three main sections. As illustrated in Figure 4.4, the rule help file starts with a brief categorisation of the type of image the localiser is dealing with (in this case, a functional image), followed by a thorough explanation of the problem detected. When applicable, brief descriptions of how blind users interact with screen readers are also provided.

In addition, for each image accessibility failure case described, we have included illustrative examples of appropriate (marked with a green sign) and non-appropriate (marked with a red sign) text alternatives related to the rule violated. Since rule help files do not contain any images, the context in which the sample alt texts appear is also briefly explained (see Figure 4.4).

The underlying assumption we have made when describing the problem flagged by Acrolinx is that web localisers are aware of what a text alternative is. Yet, we have deemed convenient to include a section in the rule documentation which provides a general introduction to image accessibility and summarises the purpose of an alt text. Every rule help file includes at the end two useful references on the subject: a list of easy alt text checks that can be performed in the web browser, and the W3C alt decision tree (Eggert and Abou-Zahra 2014a).

4.6 Rule selection for evaluation

As outlined in the Introduction, one of the secondary goals of this thesis is to investigate whether the rules we defined and the context of application that we just presented can result in localised websites with an acceptable degree of image accessibility. We are interested, therefore, not only in the impact of using the Acrolinx Batch Checker on the final appropriateness of the translated alt texts, but also in the individual effect of each rule covered by the tool. According to Nyberg et al. (2003, 257), knowing what the contribution of each individual rule is to the overall effect of the CL evaluated is one of the major difficulties encountered when assessing the impact of the use of a HOCL.

In an attempt to address this issue, it was decided to only evaluate the effectiveness of the most popular rules from the 40 rule set defined. To select the rules that would be later assessed in the framework of this thesis' web localisation experimental study (see Chapter 5), we first built a French web corpus (hereinafter referred to as 'web corpus III') similar in nature to the web corpus which served as training data during the rule development process. This new French corpus included governmental, public services and national media web sites from Belgium and Canada, whose content in terms of number of pages and images can be consulted in Appendix A.3. After the websites' retrieval using Scrapbook (version 1.5), the corpus was post-processed following the same criteria explained in section 4.3.2.1 for data clean-up. The final corpus accounted for a total of 5,377 pages and 14,017 instances of unique non-empty `alt` attributes (see Table 4.5).

Table 4.5. Descriptive summary of web corpus III built to select ten rules for evaluation

Sub-corpus	Pages	<img>	No alt	Empty alt	Non-empty alt
Belgium	2,236	10,214	537	2,320	7,357
Canada	3,141	11,857	1,874	3,323	6,660
Total	5,377	22,071	2,411	5,643	14,017

Once the web corpus was ready, we performed a batch check of all web pages per sub-corpus (20 websites each) using the Acrolinx Batch Checker client. The analysis yielded one XML checking report per page as the one previously shown in Figure 4.3. To facilitate the analysis of the 5,377 XML files, we applied a Perl script developed by Acrolinx to extract from these reports all flags registered per rule. This file processing was done through a C# Windows forms application developed by Gerlach (2015). As a result, we obtained a spreadsheet per rule containing all the errors found in web corpus III.

The analysis continued as follows: we counted the total number of flags per rule and then manually calculated the amount of true and false positives. Two main rule selection criteria were adopted: rules that yielded fewer than 50 flags should be dismissed, and the selection should be made proportionally, per sub-set. Taking these into account, we estimated the precision per rule, understood as the proportion of the number of correctly flagged errors to the total number of errors (Nyberg et al. 2003, 258), and sorted rules first per level of precision and then per number of true positives. The process described above led to the selection of ten rules: three from Sub-set A, five from Sub-set B and two from Sub-set C (see Table 4.6).

Table 4.6. List of ten rules selected for evaluation, by sub-set

Sub-set	Rules	Total flags	Precision
A: Descriptive content (3/14 rules)	DDR3-logo	913	99.01%
	DDR2-image	89	98.88%
	DDA1-credits	387	95.87%
B: Functional content (5/18 rules)	FAA3-social	2048	100%
	FAA4-print	292	100%
	FAA1-files	172	100%
	FDA3-structure	75	98.67%
	FAC1-accueil	154	92.21%
C: Uninformative content (2/8 rules)	U2-placeholder	286	100%
	U4-decor	208	99.04%

Once the rule selection process was completed, we were able to better tackle some of the other challenges highlighted by Nyberg et al. (2003) in the assessment of the effect of HOCL rules. The authors bring special attention to carefully considering, among others, important variables such as the number of texts included in the evaluation. In this sense, having a limited number of rules facilitated the choice of more targeted alt texts (10 inappropriate alt texts per rule) to be included in the study. Other variables relevant to the evaluation of a CL are the number of test persons participating in it, or the degree to which the test persons are familiar with the subject matter. The approach adopted to define these variables, as well as the selection of alt texts, will be further discussed in the next chapter (5).

4.7 Summary

In this chapter we have presented a CL rule-based approach to assure image accessibility during the web localisation process. Our proposal brings the attention back to human translation quality in controlled language-oriented research and therefore to human-oriented CL rules, where, as stated by Nyberg et al. (2003), “there is a clear need for more empirical evaluation”.

We began the chapter by offering a broad overview of the evolution of controlled languages, from Ogden's *Basic English* to one of the most recent forms of CL: the Easy to Read on the Web. This literature review led us to define our proposal as a written human-oriented controlled language whose purpose is to assist in the production of appropriate French text alternatives for images (section 4.2).

The CL development methodology followed was described in section 4.3. It consisted of the exploitation of two corpora: firstly, the collection of existing guidelines on alt text writing available in the literature and, secondly, a collection of French web documents previously analysed in the context of a Swiss accessibility study. The joint exploration of both corpora led to the definition of 40 rules, most of which are pragmatic in nature and are associated with the ultimate communication goal of the image and, consequently, of the alt text (section 4.4).

Within the framework of this thesis, the intended users of these rules are web localisers, who would apply them as a complementary localisation quality assurance measure. In section 4.5, we explained how Acrolinx technology could support them in such an endeavour by automating the alt text checking process. To facilitate a more in-depth analysis of the impact of applying each rule individually, we have selected ten rules (the most precise ones per sub-set) for evaluation (section 4.6) within the framework of a larger web localisation experimental study, which will be presented in the next chapter (5).

Chapter 5

Image accessibility assurance in web localisation: an experimental study

This chapter describes the experimental plan designed for this thesis' main study and how the evaluation of the controlled language proposal for text alternatives verification presented in Chapter 4 was integrated therein.

5.1 Overview

As announced in Chapter 1 (see section 1.4), this thesis combines a descriptive research approach with experimental research methods. Within Translation Studies, observations gathered through empirical methods normally focus on the product of translation, the translation process, the participants involved therein, or the context in which translations are produced and received (Saldanha and O'Brien 2014, 5). To determine the extent to which localisers are capable of assuring that an acceptable level of image accessibility is achieved in the target web product (Goal 3, see section 1.3, Chapter 1), we have adopted a product-based approach. Therefore, in our work, emphasis is primarily placed on the appropriateness of the image text alternatives produced, as well as on whether these are translated (*translation outcome*), but not on how image text alternatives are localised (*translation process*). However, as pointed out by Saldanha & O'Brien (*ibid.*), investigating the former is not possible without considering certain aspects of the latter. This unavoidable (yet probably desirable) overlap is reflected in the three more specific research questions we have defined prior to developing the experimental framework:

- **R1.** Are image text alternatives considered by localisers as translatable elements during the web localisation process?
- **R2.** Does the use of accessibility-oriented QA tools during the web localisation process result in more appropriate text alternatives for images?
- **R3.** Does knowledge of web accessibility help localisers produce more appropriate text alternatives for images?

To address these questions, we conducted an experimental study which consisted of two stages: a controlled web localisation experiment (Stage 1) whose output was then subjected to user evaluation (Stage 2). This experimental procedure (translation study followed by the evaluation of its output) has been successfully implemented in previous work when researchers aimed at investigating the effect of a given factor on the quality of the translation output, such as the use of machine translation (MT) technology (Guerberof Arenas 2012) or computer-assisted translation (CAT) tools (Morado Vázquez 2012). The novelty of our study relies on the factors manipulated during the experiment (see section 5.2.1) and the translation quality criteria defined prior to the user evaluation (se section 5.3.1.2). In what follows, we provide a thorough explanation of the methodological aspects and research validity considerations taken into account when designing Stage 1 (section 5.2) and Stage 2 (section 5.3) of our experimental plan. The chapter concludes with the presentation of the hypotheses we aimed at testing through the analysis of the data collected (5.4) and a brief chapter summary (5.5).

5.2 Stage 1: Controlled web localisation experiment

The objective of the web localisation experiment was twofold: firstly, to observe the extent to which translators with different awareness and knowledge levels regarding accessibility take into account web accessibility best practices during the web localisation process, with a particular focus on image accessibility; and secondly, to evaluate the impact of applying our rule set proposal by means of an accessibility-oriented QA tool (see Chapter 4) on the achievement of appropriate image text alternatives in the localised web product. Throughout this first stage of the experimental study, we collected all data relevant to answer research question **R1** and the first data set needed to investigate questions **R2** and **R3**. The experiment was carried out from December 2014 to January 2015 (see Appendix H). A total of 28 translators were asked, (i) to localise a website comprising three web pages from English into French (localisation task); and, once this first task completed, (ii) to check the three web pages for image accessibility (quality assurance task). To begin with, in section 5.2.1 we describe how the localisation experiment was designed and which variables were measured. Next, detailed information about the participants' recruitment process and profiles is presented in section 5.2.2. The last two sections account for the experiment material acquisition (5.2.3) and logistics (5.2.4).

5.2.1 Experimental Design

Conducting controlled experiments provides the opportunity to explore cause and effect relationships, with a view to proving or disproving a causal link between a factor and an observed outcome (Oates 2005, 125). In this experimental study, the observed outcome we are interested in are, in fact, the *dependent variables* (DV) assessed in the aforementioned research questions:

- **DV1.** The translation of alt texts for images in the localised web product
- **DV2.** The appropriateness of alt texts for images in the localised web product

In contrast, a factor, or *independent variable* (IV), is a "circumstance or characteristic that is manipulated or controlled to elicit a change in a human response" (MacKenzie 2013, 161) or, in other words, a change in the dependent variable(s). For the purposes of this research, we decided to manipulate the following independent variables to elicit a change in the localisers' work:

- **IV1.** The knowledge of web accessibility (WA)
- **IV2.** The use of accessibility-oriented quality assurance (QA) tools, where two software tools were proposed: *Acrolinx* and *aDesigner*.

An experiment where more than one independent variable is investigated is often referred to as a factorial study (Lazar et al. 2010). The number of conditions under which participants will perform a given task is an important aspect to be considered during the design of a factorial study, and is determined by the total number of independent variables defined, as well as by the levels of each independent variable (Lazar et al. 2010). While *IV1. Knowledge of web accessibility* only has two levels (WA knowledge and no WA knowledge), *IV2. Use of QA tools* has three levels: (i) the performance of the localisation task without any QA tool support, the result of which will be hereinafter referred to as translation version T1 (control condition); (ii) the performance of the localisation task with the help of one QA tool (from now onward called translation version T2); and (iii) the performance of the localisation task with the support of two QA tools (hereinafter named as translation version T3). Therefore, the total number of conditions in the experiment was six (2 in **IV1** x 3 in **IV2** = 6).

The administration of the test conditions in an experiment is either within-subjects or between-subjects. Given the nature of our two independent variables, we followed a split-plot or hybrid design (Lazar et al. 2010, 54), where **IV1** was investigated through a between-group approach and **IV2** through a within-group approach. Participants were thus divided in two groups in order to facilitate the manipulation of the first independent variable: whereas the control group had no previous background on web accessibility, the treatment group was aware of basic web accessibility best practices prior to the experiment (see section 5.2.2.3 for group distribution criteria).

When different levels of a factor are assigned within-subjects, interference between the test conditions may result due to the order of the testing (MacKenzie 2013, 177). To compensate for the so-called sequence effect, which could lead to an undesired improved or degraded performance, the order in which the QA tools chosen were used was counterbalanced. As a result, after asking all participants to perform the localisation task without any QA tool, 14 out of 28 participants (seven per group, randomly chosen) worked on the second translation version (T2) using *aDesigner*, and the other 14 did so using *Acrolinx*. This logic was also applied when requesting participants to generate translation version T3 (see Table 5.1).

Table 5.1. Split-plot design of the controlled web localisation experiment

	Translation Version		
	T1	*T2*	*T3*
	Translation Condition		
	Control	*Experimental*	
Control Group (without WA knowledge)	None	aDesigner	Acrolinx
		Acrolinx	aDesigner
Treatment Group (with WA knowledge)		aDesigner	Acrolinx
		Acrolinx	aDesigner

It should be noted that translation versions were cumulative, that is T1 served as a starting point to produce T2. Similarly, translation version T2, checked with the first tool used and presumably amended according to the testing results it yielded, was the basis to generate the final translation version (T3). The fundamental rationale behind this decision was based on the belief that the use of several evaluation tools is not mutually exclusive but rather the opposite: combining different software capabilities has been already put forward as a possible solution for tools' low effectiveness (Vigo et al. 2013) and could thus lead to better accessibility results.

Similarly, it is worth mentioning that the ultimate goal of counterbalancing the ordering of test conditions was to avoid any potential bias due to confounding variables, such as the practice effect or learning effect. There is the belief that participants may perform better as they progress from one test condition to the next because they get familiar with the task and they do it more effectively (MacKenzie 2013, 177).

For the purposes of this experiment, we did not inform participants they were taking part in an image accessibility-related study until being requested to use the first QA tool, nor did they receive any introduction to the topic until then (see section 5.2.4.2 for a more detailed explanation of the experiment procedure). Trying to assimilate new concepts, such as the notion of web accessibility as well as the utility of text alternatives for images in the web, could involve a higher cognitive load and time investment in the case of the control group. This could have resulted in a less efficient use of the tool proposed in the first place, motivated as well by a potential increased frustration or nervousness. In the same vein, should the order have remained unaltered, the effectiveness of the second tool used could not have been assessed objectively. The main reason is that, when running a check with a second tool, one may expect that a considerable proportion of the errors present in the document would have been already detected and solved thanks to the tool used earlier in the process.

5.2.2 Participants

5.2.2.1 Call for participation

As explained in the Introduction and argued throughout the theoretical framework of this thesis, the population group being studied in this research are web localisers. By web localiser, we understand any person involved in the linguistic, cultural and technical adaptation process of a web page or website, from internationalisation and localisation engineering (Esselink 2002) to translation and web testing. It would be no exaggeration to state that applying probabilistic sampling techniques to recruit a representative sample of this population group seems almost totally unfeasible, since this would imply knowing how many web localisers exist worldwide or at least at a national or regional level. In addition, the wide set of skills needed to perform a web localisation job —completely or partially—, together with the heterogeneous nature of the task itself, renders it inappropriate to assume that all group members will share the same unique characteristics.

A non-probabilistic approach was thus adopted to recruit potential participants by combining snowball and purposive sampling techniques. A call for participation (CFP) was published on 27th October 2014 on the Cod.eX Research Group website[1] and it remained open for four weeks, until 24th November 2014 (see Appendix C). The CFP, available in English and in French, was advertised through social networks, such as Facebook and Twitter; relevant LinkedIn groups targeting translation and web localisation practitioners; mailing lists and newsletters of translators associations and NGOs, including the Rosetta Foundation and the Swiss Association of Translators, Terminologists and Interpreters (ASTTI); Proz.com, the largest online portal for posting and responding to translation job offers; and direct communication with personal contacts from the translation and localisation industry and academia. When distributing the CFP, recipients were asked to suggest people who might be willing to participate or to forward the CFP to other relevant interested groups with the intention of gaining greater access to the target group, creating a snowball effect. The CFP included the necessary criteria to be met in order to participate in the web localisation experiment, the estimated time schedule, the honorarium and other non-monetary benefits associated with the study, namely, a free one-hour webinar on HTML best practices for web localisers (see Appendix C). It was expected that these incentives would encourage a high

[1] http://diarium.usal.es/codex/en/research/projects/web-localisation-study/ Last access: 18th July 2015.

enough response rate. On the other hand, the required conditions contributed to increase the chances of having participants who were likely to produce valuable data to meet the research purpose; a technique known as purposive sampling (Oates 2005, 98). Prerequisites were as follows:

- participants should have English into French as one of their working language combinations;
- they should have been assigned a professional job as translators or web localisers prior to the experiment;
- they should have basic knowledge of HTML; and
- they should have a high French proficiency level, preferably native or bilingual proficiency.

Over a four-week period, interested parties signed up for the study by indicating their availability in Doodle.com, a freeware online scheduling tool. When the Doodle poll was closed, a total of 49 people had showed their willingness to participate in the web localisation experiment. Another nine potential participants expressed their interest to take part in the study later on via e-mail, and were also accepted. All 58 translators were given a participant number (e.g., P58) that would serve to identify themselves and guarantee their anonymity throughout the different stages of the experiment. This ID number shall be used henceforth in this thesis when individual reference to participants will be needed.

5.2.2.2 Selection process

Once the call for participation was closed, all potential participants were sent a personal message with an invitation to complete an online questionnaire; this had to be filled in before confirming that they could participate in the study and start scheduling their sessions. The screening questionnaire was deployed via SurveyMonkey, an accessible online survey development and administration platform. Screening questionnaires are used as a means for qualifying and selecting participants to take part in research (Rubin and Chisnell 2008, 126), and are especially useful when funding restrictions are imposed. While the monetary compensation that participants would receive upon study completion only amounted to CHF 50 (lower than the regular industry rates), the screening questionnaire allowed for a more efficient use of resources.

Although a series of pre-defined parameters had already been established in the CFP, implementing this complementary screening method was deemed necessary for two main reasons. On one hand, it served to verify that participants satisfied all the requirements set forth in the CFP. To this end, they were asked to report about their professional career record, language combinations and mother tongue(s). In addition, they were requested to self-rate their knowledge level on HTML and complete a brief competence test, where they had to highlight all translatable and localisable elements present in an HTML file excerpt. All the above contributed to eliminate any possible threat to the experiment's external validity associated with self-selection sampling, a technique where data is collected from anyone who responds to a CFP (Oates 2005, 98).

On the other hand, such a screening procedure was needed to facilitate the distribution of participants into two groups (control group and treatment group, see section 5.2.2.3), and thus increase the experiment's internal validity (Oates 2005, 131). In this sense, the screening questionnaire also served as a data collection method to gather historical information about participants, which is normally requested under the form of a background questionnaire prior to the experiment. Given the length and relatively complex procedure of the study (see Appendix H), we considered that asking participants to take both a screening and a background questionnaire would have discouraged them from completing all the experiment's tasks.

Concretely, participants were first asked to provide demographic personal information (nationality, age, gender, place of residence, mother tongue). This was followed by a set of questions related to their professional experience in translation and web localisation, as well as training. Before moving to the last section, aimed at gathering information about their competence in web accessibility, HTML and other web technologies, participants were also requested to indicate the type of file formats they regularly worked with and how often they used CAT and other localisation-oriented tools. This last information set was collected to (i) assist the participant selection and group distribution process, and (ii) explore potential correlations between the results of the experiment and the participants' background and preferred working practices. The questionnaire in full can be found in Appendix D.2.

A total of 50 out of 58 potential participants replied to the questionnaire, from which all reported to have a high proficiency level in French, and English into French as one of their working language combinations. Collected data, derived from the other questions, was sorted in the following order of importance with a view to fine-tuning the selection process:

1) Knowledge of HTML
2) Results from the HTML localisation test
3) Web localisation job record (over the last two years 2013/2014 and prior to this period)
4) Manipulation of HTML files as part of translation-related assignments
5) Translation professional experience (in years)

Four respondents were disqualified because they did not return positive results in any of the abovementioned criteria. From the remaining 46, three estimated that they had no knowledge of HTML but obtained a high mark (4 or 5) in the localisation test, so they were accepted to participate in the study. For the question “Which elements would you translate and/or modify from the following HTML file excerpt?” (see Figure 5.1), grade points were distributed as follows: 0 - No response; 1 - Participant would not make any changes in the file, 2 - Participant would translate text in black (lines 8, 13, 14, 19); 3 - Participant would translate text in black plus meta elements keywords and description (lines 5, 6); 4 - Participant would adapt all translatable text and the Content-Language value (line 3);[2] 5 - Participant would do all the above and adapt the `alt` attribute value (line 16).

```
1  <html>
2  <head>
3  <meta http-equiv="Content-Language" content="en">
4  <meta http-equiv="Content-Type" content="text/html; charset=windows-1252">
5  <meta name="keywords" content="restaurant food online takeaway take away order Indian Italian Tapas
   Kebab">
6  <meta name="description" content="Online restaurant serving food from all over the word to take away">
7  <link rel="stylesheet" type="text/css" href="css/styles.css">
8  <title>The Tapas House</title>
9  </head>
10 <body>
11 <table border="0" width="72%" id="table1" bgcolor="#FFFFFF">
12     <tr>
13         <td><h3>What is it?</h3>
14 <h5>Welcome to The Tapas House. An easier way to find and order a takeaway</h5></td>
15         <td width="218">
16 <img border="0" src="images/food_6c.jpg" width="218" height="76" alt="image/food_6c.jpg"></td>
17     </tr>
18 </table>
19 <h6>Welcome | <a href="order/index.htm">Ordering</a> | <a href="help/index.htm">Help</a></h6>
20 <div><img src="images/print.png" width="30" height="30"></div>
21 </body>
22 </html
23
```

Figure 5.1. HTML file excerpt for HTML competence test

While 46 participants were initially selected for the study, 13 people retired before the experiment sessions were scheduled, pointing at professional agenda constraints as the main reason for withdrawal. Similarly, four of the localisers that had already confirmed their participation decided not to continue, two of whom abandoned the experiment upon receipt of the first task's instructions. From the 29 participants that completed all the tasks, one did not provide us with all the deliverables requested (participant P02) so this subject's experiment's output was not considered during data analysis. Despite the fact that the experiment's mortality rate in terms of loss of participants since the beginning of the recruitment process until the end of the research cycle (Saldanha and O’Brien 2014, 32) was considerably high (48%), the final sample of 28 localisers is similar in size to that

[2] Although its use is no longer recommended in HTML 5, most HTML authoring tools still automatically inserted an `http-equiv` attribute for content language definition by the time of the preparation of the test. http://www.w3.org/International/questions/qa-http-and-lang

of other small-scale translation-related and image accessibility-oriented research projects (Guerberof Arenas 2012; Tang 2012), and matches the general size recommendations for statistical analysis to be reliable — it is advised to have from 10 to 12 participants per test condition (Hinderer Sova and Nielsen 2003, 27). In addition, the experiment yielded more than 10,000 observations, which contributed to enhance the validity and the robustness of the research (see section 5.3.3).

5.2.2.3 Participants distribution between groups

In order to measure the first independent variable (*IV1. Knowledge of web accessibility*), special care was taken to divide participants into two different groups. This distribution was carried out on completion of the screening questionnaire by all interested participants. Since web accessibility-related skills have not been traditionally associated with translation and web localisation competence, and they are not yet established as such, a separate call for participation to recruit web localisers with and without this background was not deemed appropriate. In the same vein, no explicit mention was made as to the ultimate goal of the experiment in the call for participation because, to our understanding, it could have had a negative impact on the response rate to our call. Having WA expertise, a not so common demand among translation and web localisation job offers, potential participants might have refrained from signing in for fear of failure or exposure to unknown professional working practices. Similarly, the opposite could have occurred. Bias might have been introduced if participants had prepared themselves —for instance through self-learning techniques— to be eligible, a phenomenon similar to the Hawthorne effect, where people alter their normal behaviour because they know they are being observed (Saldanha and O'Brien 2014, 31).

To begin the classification process, the only criterion used was the knowledge level about web accessibility that participants had self-rated in the screening questionnaire. To the multiple-choice question "Are you familiar with the concept of web accessibility?", five responses were suggested: I have never heard about it; I have heard about it, but I do not know what it is; I have basic knowledge on the subject; I have intermediate knowledge on the subject; I have advanced knowledge on the subject (see question 29, Appendix D.2). The question was strategically positioned in the questionnaire to avoid its identification as one of the experiment's key criteria. While the initial plan was to automatically assign people who had chosen any of the first two answers to the control group, we realised upon data collection that the latter would lead to an unbalanced distribution of participants: from 50 respondents, 58% (N=29) had no previous background on web accessibility and 42% (N=21) reported to have at least basic knowledge on the matter. This disproportion was solved by randomly selecting eight participants from the first group (control) and adding them to the second group (treatment). This procedure was repeated for group counterbalancing purposes after knowing the total number of participants that had withdrawn from the study before scheduling the experiment sessions.

Table 5.2 shows the group distribution of the final sample of 28 localisers. Although we did not intend to reach an equal number of participants per group, the experiment's mortality reported in the previous section resulted in two groups of 14 people each.[3] From the control group, six participants did not know what WA was at the time of recruitment and eight people had heard about it but had not shown any interest on the subject. With regard to the treatment group, five participants have no knowledge of WA. From those who had, six considered it basic, two intermediate and one advanced.

Table 5.2. Distribution of participants per group (control and treatment)

Group	Participants
Control (without WA knowledge)	P04, P16, P18, P19, P23, P25, P31, P34, P36, P40, P41, P42, P52, P53
Treatment (with WA knowledge)	P01, P06, P21, P22, P26, P33, P44, P46, P49, P51, P55, P56, P58, P59[4]

Upon group distribution and before scheduling participants for the study, a special measure was implemented to assure that variation in the outcomes between the two groups, if any, would be attributable to the manipulation of the independent variable **IV1**. It was expected that such a measure would compensate for the differences observed among participants concerning their background in web accessibility. The solution was to offer a webinar to all members of the treatment group prior to the first experiment session which would cover the basics of the subject. While nine participants had reported to already have that background, it was necessary to provide the other five with a foundation in the field. Similarly, we were uncertain about the former nine participants' understanding on what WA is and which best practices should be followed to implement it. The webinar was therefore established as a convenient complementary means to manipulate **IV1.** Details about its content and how it was administered are reported later in this chapter (see section 5.2.4.2).

[3] Participant P02, whose data was not complete and thus not valid for data analysis, belonged to the control group, which initially comprised 15 participants.

[4] Number 13 was not assigned to anybody, thus the existence of P59.

5.2.2.4 Participants demographics and background

In this section, we provide a profile summary of participants belonging to both the control and the treatment group. Some of this information will be referred to in the next chapters (6 and 7) when discussing the results of this experimental study.

Control group

All participants from the control group but one (P19), whose mother tongue was Spanish and reported to have almost a bilingual proficiency in French, were French native speakers (N=14, aged between 24 and 57, $\bar{x}$=32, sd=2.5; four male, ten female), with an average of six years of experience in the translation domain (sd=1.8, min=0.5, max=26) and no background knowledge in web accessibility. Although nationalities included Belgian (N=1), French (N=10), Spanish (N=1) and Swiss (N=2), the location of these participants at the time of the experiment was more varied: France, Germany, Thailand, Spain, Switzerland and the United Kingdom.

Three participants (P19, P36 and P41) had not previously worked in the web localisation field, but reported to have dealt with HTML files in their professional career. During the period November 2013 - November 2014, eight participants worked on web localisation projects and, for five of them, it represented more than 50% of their annual professional activity. The remaining three participants stated that had received web localisation assignments before the aforementioned period and that it accounted for 25% of their work. Only four participants had received specific localisation training (P19, P23, P25 and P41) prior to the experiment. However, 13 people indicated they had basic knowledge of HTML, with three of them reporting intermediate skills. Interestingly, the only participant who considered them self not to have any background on the subject obtained the highest mark in the HTML competence test. This led us to believe that the person might have entered the wrong answer unintentionally. In addition, some participants indicated to also know other mark-up languages such as XML and CSS, as well as programming languages like Java.

Thirteen out of the 14 participants were freelancers working full time (N=11) or part-time (N=2). All but one person (P40) acknowledged using CAT tools for at least 50% of their translation jobs. Tools mentioned were SDL Trados 2007, Studio 2011 and Studio 2014; memoQ; Wordfast; Déjà-Vu; OmegaT and MetaTexis. Three people were also using other tools for certain localisation assignments, such as Dreamweaver, InDesign, PageMaker, Xpress and PhotoFiltre.

Overall, the most commonly manipulated file formats during translation and localisation jobs were doc/docx (79%), html (57%), xls/xlsx (57%), CAT tool-specific bilingual files (43%), xml (28%) and xliff (21%). Three participants

(P04, P31 and P53) stated that they were often requested also to work directly in a Content Management System (CMS).

In general, the control group appeared to have a satisfactory level of technological skills, which play a key role within web localisation competence (Jiménez Crespo 2013, 176). The former is reflected on the variety of tools used and files handled, according to participants' responses to the screening questionnaire. Their background seemed therefore suitable for the purposes of the study.

Treatment group

Within the treatment group, there were two people who mentioned Spanish as their mother tongue but declared to live in a French speaking country and have a high proficiency level in that language. The rest of the participants (N=12) were French native speakers. Participants' age ranged between 23 and 53 ($\bar{x}$=36, sd=2.6; three male, 11 female) and they had an average of nine years of experience in the translation field (sd=2.5, min=0.5, max=27). Nationalities included Belgian (N=2), Canadian (N=1), French (N=8), Moroccan (N=1), Spanish (N=1) and Swiss (N=1). They all carried out the experiment in their country of origin.

Four participants (P06, P22, P49 and P55) had never received a web localisation assignment prior to their participation in our study. However, they had the requested background in HTML and obtained a high score (3-5) in the competence test. From the remaining ten participants, seven had worked on web localisation projects between November 2013 and November 2014 (for five of them, it accounted for at least 50% of their annual activity; for two people, it represented a 25% of their overall annual translation volume).

Five participants indicated that they had followed web localisation courses (P22, P26, P46, P56 and P59). One participant (P21) perceived his HTML skills as advanced, while four rated themselves as intermediate and nine as basic. As noted in the case of the control group, some participants (N=9) also had knowledge of XML and CSS mark-up languages. Java, Perl and Python were mentioned by two participants when being asked about their competence regarding other web technologies and programming languages.

Twelve out of the 14 participants were freelancers working full time (N=10) or part-time (N=2) as professional translators. All participants acknowledged using CAT tools for at least 75% of their translation jobs, except for one (P49), who indicated that clients did not request her to work with this type of technology. No other CAT tools were mentioned apart from the ones used by the control group and listed in the previous section. Seven people referred to the use of complementary tools for localisation-oriented tasks, including image processing and HTML editing environments, such as Photoshop, Illustrator, Dreamweaver and

Xmetal. One participant also mentioned the use of Solid Converter and ABBYY FineReader for file conversion purposes.

Overall, the most commonly manipulated file formats during translation and localisation jobs were doc/docx (50%), CAT tool-specific bilingual files (50%), html (43%), xls/xlsx (43%), xliff (21%) and xml (21%). Three participants (P33, P44 and P56) stated that they were used to working directly into the client's Content Management System (CMS).

It can be observed from the data presented so far that the profile of both groups (control and treatment) seemed fairly balanced and comparable. The main difference between groups that had been strictly controlled was the web accessibility background of their members, as indicated in section 5.2.2.3. While it was not possible to hold all the other factors constant, such as age, gender or previous relevant experience, we believe that the slight variations reported in this regard did not interfere with the results of the study.

5.2.3 Experiment materials

The twofold goal of the localisation experiment, mentioned in the introductory paragraph of section 5.2, inevitably imposed the adoption of a pragmatic approach during the experiment material acquisition process. In order to facilitate the assessment of the rules selected for evaluation (see Chapter 4, section 4.6.), a representative sample of text alternatives violating these rules had to be chosen. In addition, we also aimed at observing how image accessibility was addressed during the web localisation process, so it was not considered appropriate to simply provide participants with a list of images and their corresponding text equivalents for translation (for instance, in an MS Word or Excel file), totally isolated from their web context. Since participants reported a satisfactory level of technological skills and HTML files ranked #2 in the list of most commonly manipulated file formats during translation and localisation jobs, an HTML-based website was chosen as the main source product to localise. However, it proved impossible to find an already existing site comprising a significant number of images and text alternatives which could meet the requirements of the study. Therefore, an experimental website was designed from scratch.

In the following sections, we report how the text alternatives and images that were to be included in the experimental website were selected (5.2.3.1). Then, we describe the web development process and the characteristics of the final site that served as the source document of the web localisation experiment (5.2.3.2). Finally, before moving to the experimental environment section, we present the QA tools evaluated (5.2.3.3).

5.2.3.1 Selection of text alternatives and images

For the purposes of our study, we defined the value of the HTML `<img> alt` attribute as the main unit of investigation, that is, text alternatives were the target population studied. The sampling method followed in selecting the image text equivalents that web localisers would need to translate during the experiment was similar to the one reported in Chapter 4 (section 4.3) for rule development.

The first step was to build an English web corpus (hereinafter referred to as 'web corpus IV') similar in nature to the previous web corpora already presented in Chapter 4. Website retrieval was carried out on 3rd November 2014 using Scrapbook (version 1.5), the same website downloader application used to build web corpora I, II and III. All pages within a distance of one link from the homepage (also included) were retrieved from a total of 44 websites from two different countries: Ireland and South Africa. More than 5,000 English web pages were then processed with Rainbow (version 6.0.26) to extract all text alternatives contained therein (see Chapter 4, section 4.3 for a comprehensive explanation of the alt text extraction and clean-up process). Table 5.3 shows a descriptive summary of the web corpus content, including the total number of pages and images, as well as the amount of images without an `alt` attribute, an empty alt and non-empty alt text. A detailed list of all the websites retrieved per sub-corpus can be found in Appendix A.4.

Table 5.3. Descriptive summary of web corpus IV built for images and text alternatives selection

Sub-corpus	Pages	<img>	No alt	Empty alt	Non-empty alt
Ireland	2,236	10,214	537	2,320	7,357
South Africa	3,141	11,857	1,874	3,323	6,660
Total	5,377	22,071	2,411	5,643	14,017

After data collection, non-appropriate text alternatives were manually coded according to the rule they contravened from the rule sub-set selected for evaluation (see Chapter 4, section 4.6). While it was a time consuming task, it then allowed us to randomly select 20 instances of each rule in order to get a preliminary sample of 200 unique text alternatives. The images they were describing were also stored for later usage during the web development process. One of the advantages of working with such a multimodal corpus was the possibility of checking the images selected in context. The latter served to verify whether the annotations made with regard to the appropriateness of the text alternatives gathered were pertinent. When this was not the case, the corpus was explored again to retrieve new representative instances of the rule in question.

As highlighted by Saldanha and O'Brien (2014, 71), it is recommended to ensure that the sample includes the full range of variability in that population and that there is a certain balance among the different values in the range. In order to guarantee that these criteria were met, 20 instances of empty `alt` attributes and 20 cases of appropriate text alternatives were also added to the sample. Similarly, since the existence of images with no `alt` attribute within the `<img>` HTML element was also observed in all the web corpora built for this research, the final sample included instances of this image category as well.

Upon this preliminary selection of 260 images, their content was carefully examined with a view to defining possible themes around which the website could be designed. However, for the website to be aesthetically similar to the average site users are used to browsing, the image sample had to be reduced. This belief was reinforced by the heterogeneity of the subjects represented through the images selected, which would unavoidably lead to a chaotic website that could have an impact on localisers' overall performance. As a first step towards limiting the scope of the website and obtaining a coherent textual and visual content, the sample was limited to a total of 130 images whose associated text alternatives are shown in Appendix B.1.

5.2.3.2 Experimental website

Design and development

The aforementioned corpus-based approach enabled the selection of real examples of appropriate and non-appropriate text alternatives. However, as pointed out earlier, their distribution across the different web pages had to be done in a meaningful way to preserve the website representativeness. To this end, we first created a website mock-up where images were predominant, but without compromising its overall visual design. As can be readily seen in Figure 5.2, the website structure was simple enough to accommodate a high proportion of visual content, which could then be easily compensated by harmoniously integrating small pieces of text.

This template, created from scratch for the purposes of this experimental study, together with the themes identified through data observation, inspired the development of a website about a fake development campaign, *Together against poverty*, whose main partners were the following: (i) the Mail & Guardian South African journal, (ii) the University of Johannesburg, and (iii) the United Nations Volunteers (UNV) Programme. The selection of these three existing institutions is innocently based on the arbitrary presence of their logos in the image sample built for the study.

Once the theme was chosen, the web development process continued as follows: first, a webpage for each partner was created, corresponding to files index.html, index2.html and index3.html respectively. All web documents shared the same graphic and content layout (see Appendix B.3), which will be described in more detail in section 5.3.4. Some images from the final corpus sample were replaced with icons serving the same purpose in order to match the style sheet of the website created, but no modifications were made to the text alternatives. Likewise, some structural and decorative images, as well as the social media icons, were used more than once but the alt text always differed. Images and textual content were then distributed across the three pages, always trying to simulate a fully functional website. For instance, wherever a 'Print' option was available, the use of the JavaScript print() method would enable the user to have access to the Print Dialog Box. In the same vein, the navigation menu allowed the user to browse through the different pages, and sample files could be downloaded from the documents section (see Figure 5.2).

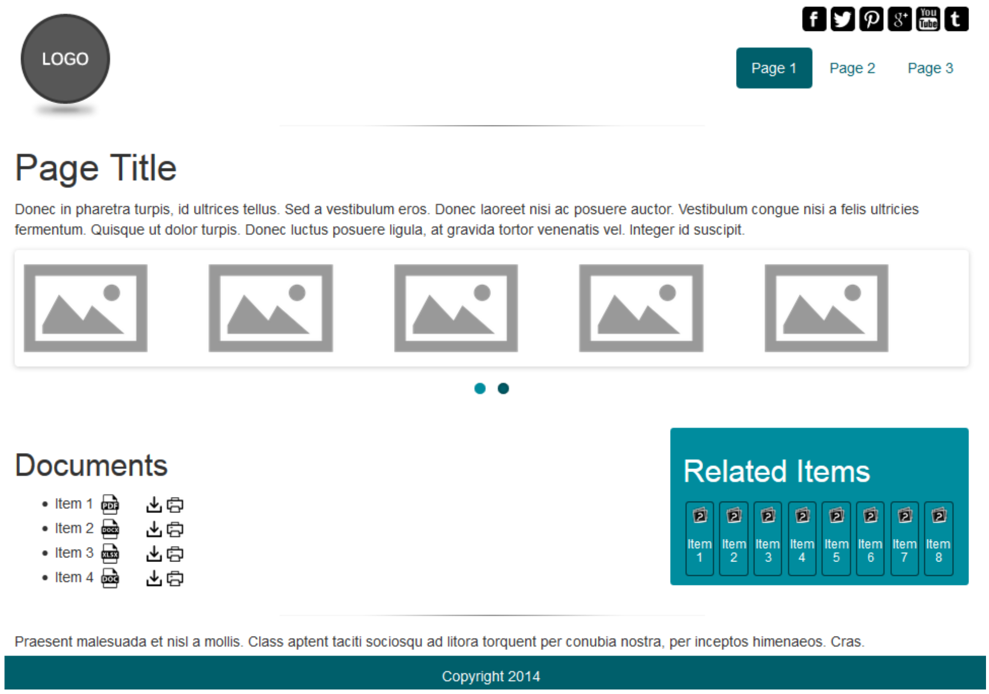


Figure 5.2. Mockup of the experimental website depicting its macrostructure

All the above was of utmost importance in the particular case of images whose alt text appropriateness is strongly dependent on image functionality (e.g., the 'Print' or the 'Download' buttons). In this sense, special mention should be made of the social media section, located at the top of the page, on the right-hand side. Since the aforementioned institutions have social media presence, it was decided not to link our experimental website to their real profiles. Instead, a hyperlink placeholder was introduced (`href="#"`), following web development standard practices. This technique was also used to indicate the existence of child pages, which we intentionally did not include in the website macrostructure to avoid any potential confusion among participants during the web localisation task.

Content

As mentioned before, text alternatives for images were our main object of study. Therefore, observing and analysing how web localisers translated and interacted with other textual elements of the website was out of the scope of this thesis. Still, their presence was necessary not only to preserve the integrity of the website, but also to avoid drawing the participants' attention to the high proportion of images contained in the website and thus raising suspicion about the ultimate goal of the study.

The main textual content was less than 100 words and consisted of three full sentences (73 words) describing the campaign that were repeated on each web page. The remaining body text included section, document and element titles of maximum three words length. In addition, we added a page title and keywords for each file, and a disclaimer which was introduced as an HTML comment on each web page. Table 5.4 summarises the website word count per text location and web document (the full source text can be consulted in Appendix B.2). As illustrated in the table, the source website contained a total of 767 translatable words, 229 of which were repetitions (campaign description –within the main content– and disclaimer). It was expected that most participants would either identify this repeated text when skimming through the website or thanks to the use of a CAT tool, so the final unique segment-based word count was estimated at 538. The value of the `alt` attributes accounted for sixty-four per cent (64%) of the overall translation word count.

Table 5.4. Source text word count per text location and HTML document

Text	index.html	index2.html	index3.thml	Total website
`<head>` content*	5	6	4	15
`<img>` alt content	95	132	121	348
Main content (body)	94	100	93	287
Disclaimer†	39	39	39	117
Total words	233	277	257	767
Total 100% match	0	113	116	229
Total unique words	**233**	**164**	**141**	**538**

*`<title>` and `<meta>` content / †HTML comment

It is worth highlighting that the number of alt text words per web page does not proportionally correlate to the number of images available per HTML file. While there is a difference of 37 words between alt texts in index.html and index2.html, both documents contained the same number of images (45). The third web page (index3.thml) had 40 images (see Table 5.5). This image ratio is in agreement with the average number of images per page observed in Jiméne -Crespo's (2008, 273) web-comparable corpus ($\bar{x}$=35.40 images, original sub-corpus; $\bar{x}$=46.98 images, localised sub-corpus). Similarly, it should be noted that, although the distribution of images across web pages was balanced (45-45-40), this was not the case regarding the rules violated by their associated alt texts. As shown in Table 5.5, the distribution of the ten text alternatives per rule between the different pages was arbitrary and depended on the content flow within each web page. Overall, we believe that the resulting experimental website adequately fit the purposes of the study without compromising its validity.

Table 5.5. Distribution of images per web page, according to alt text rule violated and attribute value type (appropriate, non-existent, empty)

Rule/file	FAA3-social	U4-decor	FAC1-accueil	DDR3-logo	DDR2-image	DDA1-credits	FAA4-print	U2-placeholder	FAA1-files	FDA3-structure	app-alt	no-alt	alt-empty	**Total**
index.html	5	5	2	3	3	3	4	4	3	4	2	4	3	**45**
index2.html	3	4	4	4	6	2	4	4	3	4	3	1	3	**45**
index3.html	2	1	4	3	1	5	2	2	4	2	5	5	4	**40**

5.2.3.3 Accessibility-oriented QA tools

As stated in previous sections, web localisers participating in the experiment had to check the outcome of their work for image accessibility. Since one of the goals of the study was to observe the impact of applying the controlled language (CL) rules formalised through Acrolinx on the appropriateness of text alternatives, the use of this software during the experiment seemed unavoidable. Given that its client-server architecture and functionalities were presented in Chapter 4 (see section 4.4), we shall only describe here how Acrolinx was configured for the study.

A special rule package was first built by choosing the ten rules selected for evaluation and then stored in the local server which had been installed in all the computers participants would have access to. The latter served to ensure that they could only receive feedback concerning these ten rules, and not the full 40-rule set. In order to avoid any confusion, ours was the only rule set available on the server. During the experiment, participants had access to these rules through the Acrolinx Batch Checker (version 2.7.0), one of Acrolinx's clients for website verification. Before each session, the researcher defined the document and checked the options which had to be enabled during the task[5], so that participants did not have to invest any time configuring the tool. Similarly, we made sure that the rule help

[5] A screenshot of Acrolinx Batch Checker 2.7.0 interface can be found in Figure 4.2 (Chapter 4).

could be consulted, if needed. Once the session started, no further intervention was needed from the researcher.

While the effect of implementing *Acrolinx's* suggestions on the final image accessibility level achieved could be measured by comparing the *Acrolinx*-mediated translation with the outcome of the control condition, where no QA tool was used, we deemed it appropriate to introduce a second tool for comparison. The latter would not only add value to the study, but also enable us to know whether our proposed solution could be more suitable for alt text verification than other existing software. We also considered relevant the possibility of suggesting *Acrolinx* as a complementary solution to general web accessibility evaluation tools which, as discussed in Chapter 3, do not always offer sufficient coverage in terms of language-related accessibility verification. Taking into consideration all the above, it seemed reasonable to choose a tool that, while being designed for full WCAG 2.0 conformance checking, also included specific features for image accessibility assessment. A systematic review of existing software with such characteristics revealed that *aDesigner* was the most convenient tool for the purposes of our study.

Eclipse *ACTF aDesigner* is a freeware visual impairment simulator which, at the same time serves the user as an evaluation tool to check websites against different sets of HTML accessibility guidelines, such as the US Section 508 of the Rehabilitation Act and the popular WCAG 2.0. In 2007, it switched from being a proprietary IBM product to an open source project. This might explain the inclusion of a third set of guidelines: the Japanese Industrial Standard (JIS X 8341-3). *aDesigner* also helps users to check accessibility of ODF documents and Flash content. At the time of the experiment, the most up-to-date version (1.0.0) was dated from June 2012[6].

The reasons that motivated the selection of *aDesigner* are the following: on one hand, it is a desktop application, like *Acrolinx's* client for web page verification, whose user interface is a priori intuitive and relatively simple to use; in addition, the results reporting format is very similar to *Acrolinx's*, since errors detected are organised per rule violated and a description of the problem is also provided. On the other hand, as far as image accessibility checking is concerned, *aDesigner* provides some clues about the alt text appropriateness of the images found, instead of just detecting if an `<img> alt` attribute is present or not, which is a popular feature among other tools such as WAVE or TAW[7]. Apart from highlighting the absence of `alt` attributes as clear errors, aDesigner always offers a general reminder to users regarding success criteria 1.1.1 of the WCAG 2.0: “Check that all non-text content that is presented to the users has a text alternative that serves the

[6] In March 2015, version 1.2.0 was released: http://www.eclipse.org/actf/downloads/tools/aDesigner/index.php Last access: 30th July 2015.

[7] See Chapter 4 for a comprehensive explanation of the functionalities of general web accessibility evaluation tools.

equivalent purpose”. Similarly, *aDesigner* marks one-character and certain one-word alt texts as “probably errors”. For instance, it would identify 'banner', 'spacer' or 'image' as words within the text alternative and indicate that they possibly lead to inaccessible images (see Figure 5.3). Since *Acrolinx* CL rules also cover these assumptions, *aDesigner* lent itself particularly well to the tool comparison scenario planned.

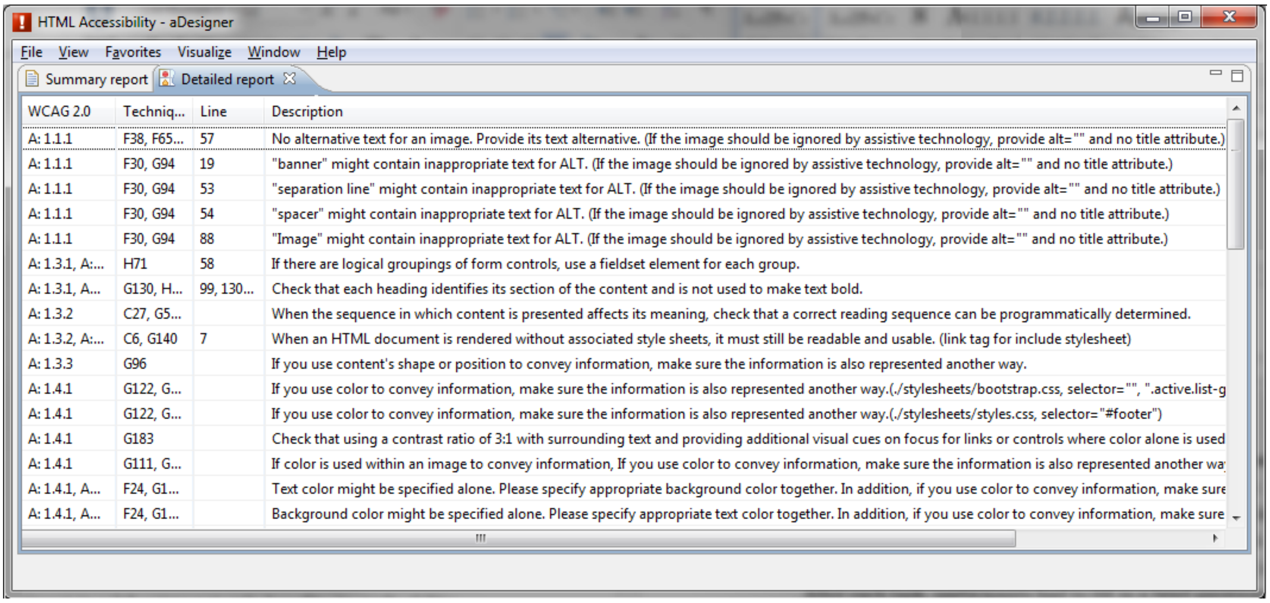


Figure 5.3. aDesigner's feedback after WCAG 2.0 conformance check

aDesigner is a Rich Client Platform (RCP) application, so it does not require the user to go through a multi-step installation process. Instead, the tool can be quickly launched directly from the executable file whenever needed. In this sense, the only measures taken prior to the experiment were (i) storing the executable file in the computers that the participants would have access to, and (ii) configure the accessibility evaluation preferences so that check results were only based on WCAG 2.0.

5.2.4 Experimental environment

Conducting an experiment in a lab setup not only benefits its replicability but also facilitates a higher control of environmental factors and other variables which are not necessarily under investigation and might be a source of bias. When human subjects are involved, however, this might imply observing them carrying out tasks they are used to performing, but outside their usual environment. This improves the internal validity of the research but, at the same time, compromises it external validity (MacKenzie 2013, 140–143). Since the scope of this experimental study was limited to the *translation product*, and not the *process*, conducting the experiment in a lab setting was not considered decisive for general research validity. It was decided, therefore, that the study would be carried out remotely, that is, participants would be able to perform the tasks requested in the place of their choice. One advantage of this approach was the possibility of recruiting people other than Geneva residents, where the research was based and which we believe led to a higher response rate to our call for participation. It also enhanced the ecological validity of the experiment, since participants could use the resources they were familiar with (hardware, software) to perform the tasks requested.

Still, certain conditions were established to ensure that instructions were correctly followed, as well as guaranteeing the validity of the data collected. One measure adopted to achieve the latter was asking participants to use screen recording software during all the experiment's sessions. The installation and proper use of such a programme required participants to perform all tasks in a computer running Windows 2000 or a newer version of this operating system. When access to a computer with such system requirements was not possible, participants were offered the possibility to remotely access a computer equipped with such software, located in the computer lab of the Faculty of Translation and Interpreting (FTI), at the University of Geneva (UNIGE). Specific time frames were also allocated for each session, with a view to avoiding task fragmentation and participants' distraction.

5.2.4.1 Tasks and associated task-specific questionnaires

The experiment was conducted on the basis of a web localisation assignment which comprised two main requests: localising the website described in section 5.2.3.2 and checking the localised version for image accessibility with the help of two accessibility-oriented QA tools: *aDesigner and Acrolinx* (see section 5.2.3.3). After each task, participants had to fill out a brief questionnaire. This research instrument was administered with the primary goal of gaining first-hand information about how each task was carried out, participants' opinion about the tools tested and their perceived role as contributors to image and web accessibility.

Localisation task and associated questionnaire

In the first task, participants were asked to localise the aforementioned website from English into French. They were informed about the existence of link placeholders (see section 5.2.3.2) and requested to perform the task as if, when clicking on them, the user would be taken to real web pages. No restrictions were indicated with regard to the methodology followed to carry out the task: participants were free to use the CAT tool of their choice or to localise the web pages without the help of any translation technology. They were given access to a locally stored website, which contained three HTML files and their associated folders. Nevertheless, given the scope of the study, no localisation engineering was required. In the task briefing sheet, participants were asked **not to** (i) add a language selection page or any other language selector mechanism to the source and target web pages; (ii) make any changes to the Cascade Style Sheet (CSS) file to adapt font type, font size or website colours; (iii) modify file names, folders content or the structure of the website (see Appendix G.1). In short, they were requested to only manipulate the three HTML files and deliver a fully functional website in the target language. This deliverable was considered as translation version T1. As previously mentioned, no reference was made to image accessibility requirements in the localisation briefing sheet, since this localisation task would serve as the control condition to observe whether text alternatives are considered as translatable text without explicitly mentioning it, as well as to measure the effect of using accessibility-oriented QA tools on the achievement of a higher level of alt text visibility and appropriateness.

In the post-task questionnaire, participants belonging to the control group were requested to indicate whether they had used a CAT tool for the task. If they had, they were asked to report why and if any changes were made outside of the tool. In an attempt to understand the localisation approach adopted, participants were also invited to comment on what elements they had translated or modified, and the audience for whom they thought their work was intended. The last questions aimed at knowing whether participants would have taken the same steps to perform a similar task in a real life scenario. The same questionnaire was administered to the treatment group members, for whom a special section was devoted to gather knowledge about how the webinar was perceived and whether they believed it had influenced their performance. The full questionnaire, covering questions addressed to both groups, can be found in Appendix D.3.

Quality assurance task and associated questionnaire

In the second task, participants were requested to use two accessibility-oriented QA tools in order to assess the localised French website for image accessibility. Participants were recommended to run at least one check per web page, but could perform as many checks as they wanted per tool. Both tools would generate a report per check, which also had to be stored for later delivery. However, participants could not randomly move from one tool to another. Instead, the checking software had to be used in the order indicated on the task briefing sheet. Localisers were therefore encouraged to first make all the changes they deemed appropriate to translation version T1 according to the first tool's feedback, if any. Before moving to the second tool, participants were asked to hand in what would be later considered as translation version T2. At a later stage, after having used the second tool proposed, participants would deliver the final localised website (translation version T3).

As explained in section 5.2.1, the order in which tools were used was counterbalanced in order to avoid any potential bias. This motivated the design of two different QA scenarios. In **Scenario A**, T2 was generated after having used *aDesigner* and T3 was saved once all the necessary checks were run with *Acrolinx*. In contrast, instructions in **Scenario B** were to use first *Acrolinx* before handing in T2 and then move to *aDesigner* to prepare the final version to be delivered (T3). Regardless of the scenario suggested, the task briefing sheet for participants from both groups started with a short introduction to the topic of web accessibility, where emphasis was placed on how visually-impaired users access images in the web. Detailed guidance on how to use each tool was then offered in the same document (see Appendices G.2 and G.3). Table 5.6 shows the distribution of participants per scenario.

Table 5.6. Distribution of participants per quality assurance task scenario (A or B)

Group	Scenario	Participants						
Control (without WA knowledge)	Scenario A	P18	P25	P34	P40	P41	P42	P52
	Scenario B	P04	P16	P19	P23	P31	P36	P53
Treatment (with WA knowledge)	Scenario A	P01	P33	P44	P46	P49	P55	P58
	Scenario B	P06	P21	P22	P26	P51	P56	P59

On task completion, participants were asked to respond to a post-task questionnaire touching upon different aspects related to the second assignment they had received. It included questions about *aDesigner* and *Acrolinx* usage, utility and performance during the QA task. This core section was complemented with another two; one about participants' QA practices, and another one about the role of web localisers in the achievement of web and image accessibility. While the complete questionnaire can be found in Appendix D.4, in Chapter 7 we will address each question individually, and the type of data gathered will be presented in more detail (see section 7.3).

5.2.4.2 Procedure

After distribution of participants between groups (see section 5.2.2.3), two data collection sessions (one per task) were scheduled on an individual basis. From 30th November to 8th December, selected participants were contacted personally via e-mail or Skype to schedule both sessions according to their availability. While they were encouraged to perform both tasks in the same day, each session timeframe was adjusted to accommodate each participant's needs. The experiment sessions were finally held between 8th December 2014 and 15th January 2015 (see Appendix H).

Each session was conceived to be of 90 minutes' length, after which the post-task questionnaire should not take participants longer than 15 extra minutes.

Nevertheless, participants could take more time if needed or available. Task briefing sheets were sent via e-mail ten minutes before the agreed starting time. For both tasks, as discussed above, there was no need for participants to come and use the facilities at the University of Geneva, but they were requested instead to record their screen activity. Post-task questionnaires could be completed without being recorded.

In addition, it is worth noting that participants had neither used nor received any training on *Acrolinx* or *aDesigner* prior to the experiment. In order not to introduce any bias, access to both tools was only given when needed for the quality assurance task. Similarly, with a view to avoiding the distribution of *Acrolinx* licenses and its installation on each participant's personal computer, localisers were asked during the second session to remotely access a computer located at the UNIGE, where both *Acrolinx* and *aDesigner* had been previously installed. This released them from the burden of having to install two additional pieces of software, which can be particularly complex in the case of *Acrolinx*.

All the above considered, we shall now describe the overall experiment procedure. Upon completion of the localisation task and its associated questionnaire, participants had to send two deliverables to the researcher: the screen recording video file and the localised website (T1). The next step was to transfer the T1 files to the computer that would be accessed for the quality assurance task. During this task, participants were given the possibility of only working on the remote desktop or using it jointly with their personal laptop (for instance, to be able to implement the QA tools' suggestions directly in their CAT tool environment). In this sense, the participant could easily drag and drop files from his personal desktop to the remote desktop and viceversa. Once the task was finished, deliverables had to be stored on the remote desktop for easy retrieval by the researcher. They included the video file of the second recorded session, the reports of all the checks performed per tool, and the last two versions of the localised website: T2 and T3. Participants could then complete the second questionnaire or do it later in the day. Figure 5.4 illustrates the procedure that has just been described.

Except for participant P21, who performed both tasks using the remote desktop because he deliberately chose not to install the screen recording software requested, all participants followed the procedure above. For the treatment group, the webinar was held before the first experiment session, on the same day. Irrespective of which group they belonged to, participants had to provide the researcher with a written informed consent form before taking part in the study. One week prior to the first scheduled session, participants were sent the following materials together in the following format:

1) *An introduction to the study.* This document explained how the study would be organised, the duration of the sessions and the software that needed to be

installed. Participants from the treatment group also received information about the webinar (see Appendix E.1).

2) *Screen recording software user guide.* Instructions were given about how to download and install the software, as well as on how to set it up for the experiment (see Appendix E.2).

3) *Computer remote control software user guide.* In this document, the participant would find, as in the previous one, indications on which software to download, install and use in the second session (see Appendix E.3).

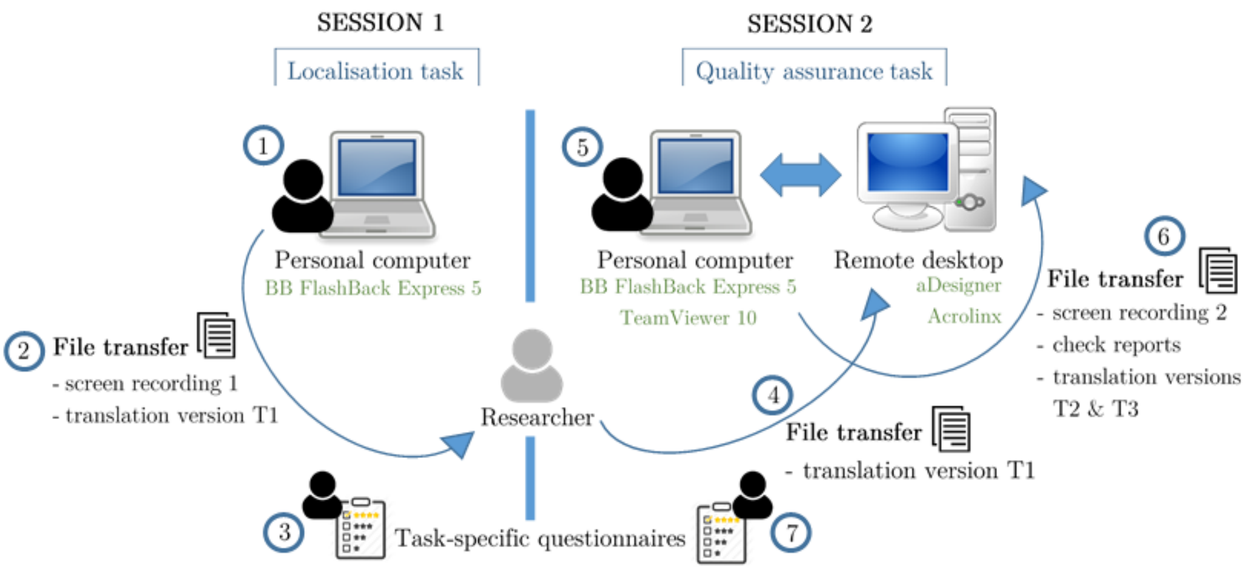


Figure 5.4. Overview of experiment's step-by-step procedure

Finally, it is worth mentioning that, throughout the sessions, an online platform was available to remotely assist participants. If a technical problem or any doubt related to the task briefs would arise, participants could access the platform anonymously and consult the researcher. In most cases, they used this resource to inform the researcher about task completion. Occasionally, it was used to receive clarification about certain functionalities of the QA tools proposed or confirmation that instructions were being correctly followed. While participants' screens were being monitored during the second session, direct intervention was only needed once to set Mozilla Firefox as the default browser for participant P18 to correctly visualise *Acrolinx* rule help files. No other major incident was registered within the six week period during which the experiment was held.

Apparatus

A computer running Windows 2000 or a newer version of this operating system was required to participate in the study. A screen of at least 17 inches was recommended, especially for the second task, when access to a remote desktop was required. Apart from the two main accessibility-oriented QA tools described in section 5.2.3.3, the software applications and services used in the experiment were the following:

1) *Adobe Connect.* This web conferencing software allows the host or presenter to remotely share visual and audio content across devices. This platform served as a virtual classroom where participants from the treatment group followed the webinar in real time prior to the first experiment session (see section 5.2.4.2). In addition, Adobe Connect chat functionality was enabled for all participants irrespective of the group during each session, mainly as a troubleshooting platform. The latter facilitated a smoother and more efficient communication between the researcher and participants. Access to this proprietary software was provided by the UNIGE.
2) *BB FlashBack Express 5.* This freeware screen recording software enabled participants to record their full screen activity. Although it also captured keystrokes and mouse activity, they were not considered during data analysis. Participants could test the software before the first scheduled session. The recording output was exported as an .fbr file which could then be visualised in BB FlashBack Express Player.
3) *TeamViewer 10.* The software used to remotely control the computer where *Acrolinx* and *aDesigner* were installed was TeamViewer 10. It not only provided participants with the possibility to access the UNIGE desktop, but also facilitated a quick and easy file transfer to and from the remote computer.
4) *SurveyMonkey.* The post-task questionnaires were developed and administered through SurveyMonkey. JavaScript and cookies had to be enabled in the participant's browser in order for the site to work properly. This web application was specifically chosen because of its accessibility features. Although this characteristic was not relevant for the post-task questionnaires' administration during the localisation experiment, it was necessary for the subsequent user evaluation (see section 5.3.4). Acquiring the *Select* pricing plan was deemed appropriate since the free version of the software only allowed ten questions per survey.
5) *UNIGE File Transfer Protocol (FTP) Service.* The website to be localised during the first session of the experiment was made available to participants via the FTP of the Faculty of Translation and Interpreting at the UNIGE. Although they were initially expected to deliver the experiment's output also using this service, a test conducted prior to the main study (see section 5.2.4.2) showed that other file retrieval measures were needed.
6) *WeTransfer.* As depicted in Figure 5.3, upon completion of the quality assurance task, participants were requested to transfer all the deliverables, including the screen recording video file, to the remote desktop. However,

this action could not always be carried out successfully due to participants' limited internet connectivity or the large size of the media file generated by BB FlashBack. When this was the case, WeTransfer enabled a secure and easy transfer of up to 2GB. Files could be downloaded from WeTransfer's server over a period of one week before they were automatically deleted.

Research ethics

When an experimental research project involves the participation of human subjects, it is of utmost importance to provide people with relevant information about the nature of the study and its procedure, so that they can make an informed decision as to whether or not they wish to take part in it. Potential participants should also be assured that their privacy will be protected and that they might be informed of any potential risks the study may entail (Lazar et al. 2010, 376–386). In this sense, as Saldanha & O'Brien highlight (2014, 43), “informed consent is one of the core principles in ethically-designed research.”

In the localisation experiment, the above was sought through different documents. In the call for participation, reference was made to the fact that participating in the study would not imply any known risk and that withdrawing was possible at any time (see Appendix C). After their expression of interest, participants received a consent form which should be duly signed prior to the experiment. By signing this document, they were acknowledging to have understood all the conditions associated with the study, such as the need to install and use a screen recording tool. They were also informed that no monetary compensation would be assigned if these conditions were not met. In addition, through this form, the researcher was granted permission to observe and analyse the files delivered by the participant upon completion of all the tasks requested (see Appendix F).

Participants' anonymity was guaranteed when interaction between them was unavoidable ‒such as in some webinar sessions‒ as well as throughout the whole experimental process thanks to the use of personal ID numbers. Privacy protection was also offered by indicating in the informed consent form that the recording software which they needed to use would not record any sound or their working environment (for instance, through the webcam). When this promise was not enough and people would still refuse to install BB FlashBack Express 5 on their personal computer, an alternative to participation was also offered (Lazar et al. 2010, 382), namely, performing all the tasks requested during the experiment remotely, on a computer located at the University of Geneva.

As Hinderer Sova and Nielsen point out (2003, 119–120), people feel more satisfied when their contribution is appreciated. In this sense, gratitude was not only expressed through the incentives offered (CHF 50 and a free one-hour webinar). Each time participants completed a new step in the process, they were sent a personal e-mail with a thank-you note, acknowledging receipt of the deliverables.

Finally, it is also worth noting that an electronic copy of the material presented in the webinar was also sent to those who requested it once the experiment concluded.

Webinar on web accessibility

A webinar on HTML best practices for web localisers had been announced in the call for participation as a non-monetary compensation for participants that would agree to take part in the study. While it was imperative for members of the treatment group to attend the webinar before the localisation task, as introduced in section 5.2.2.3, participants belonging to the control group were free to decide if they wanted to receive it or not. At the time of this thesis writing, only five out of the 14 participants from this group requested and attended the webinar advertised.

Instructions on how to gain access to the webinar platform were embedded in an e-mail message sent to participants ten minutes before the starting time agreed. Only then were they informed that the HTML best practices they would learn about were oriented to assure a higher web accessibility level in their localisation work. A total of nine webinars were provided, three of which were attended by only one participant. When more than two people were invited to join the webinar, they would log in using their participant ID to preserve their anonymity. Access to Adobe Connect was password protected (see Figure 5.5). The webinar had a duration of approximately 50 minutes, with a ten-minute session allocated at the end for questions and answers. Participants could interact with the presenter via the chat pod or audio input, once the functionality "raise hand" was activated. This would alert the host of the session that someone wanted to intervene. Adobe Connect allowed the presenter to share her screen with the attendants, as illustrated in Figure 5.6.

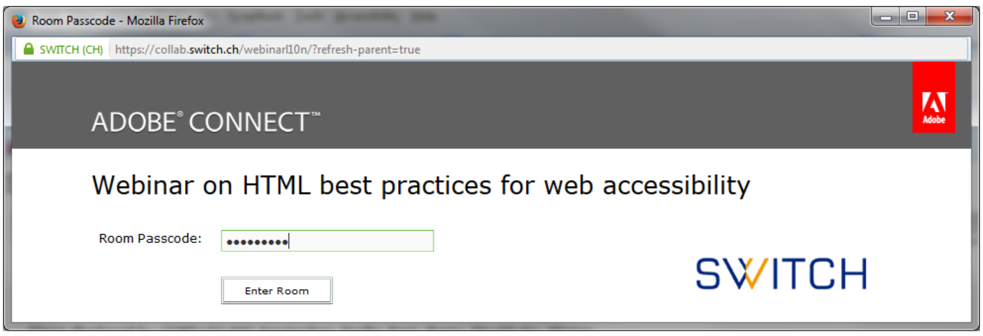


Figure 5.5. Password-protected access to webinar platform

The webinar was administered by the researcher in the form of a PowerPoint presentation slide show. To begin with, an introductory section about the meaning of accessibility in the broader sense preceded the explanation of how web accessibility is being addressed by the W3C, and which assistive technologies are often used by people with disabilities to interact with the web. This introduction was followed by the core content of the webinar: a walk through the Web Content Accessibility Guidelines (WCAG) 2.0 (see Table 2.1, Chapter 2, for an overview). The more relevant guidelines and success criteria (SC) for web localisers were pointed out and illustrated with examples. The SC that received particular attention were the following: 1.1.1 *Non-text Content*, 1.3.1 *Info and Relationships*, 2.4.1 *Bypass Blocks*, 2.4.2 *Page Titled*, 2.4.4 *link Purpose (In Context)*, 2.4.9 *Link Purpose (Link Only)*, 3.1.1 *Language of Page*, 3.1.2 *Language of Parts*, 3.1.4 *Abbreviations*, 3.1.5 *Reading Level*, and 3.3.2 *Labels or Instructions*. The webinar was concluded with a slide about how web accessibility implementation is integrated within the web development cycle, after which we invited participants to think about how this should be done in the case of the multilingual web. To better understand what had just been presented, participants were encouraged at the end of the session to watch a six minute highly instructive video[8] about how blind users perceive the web by means of a screen reader.

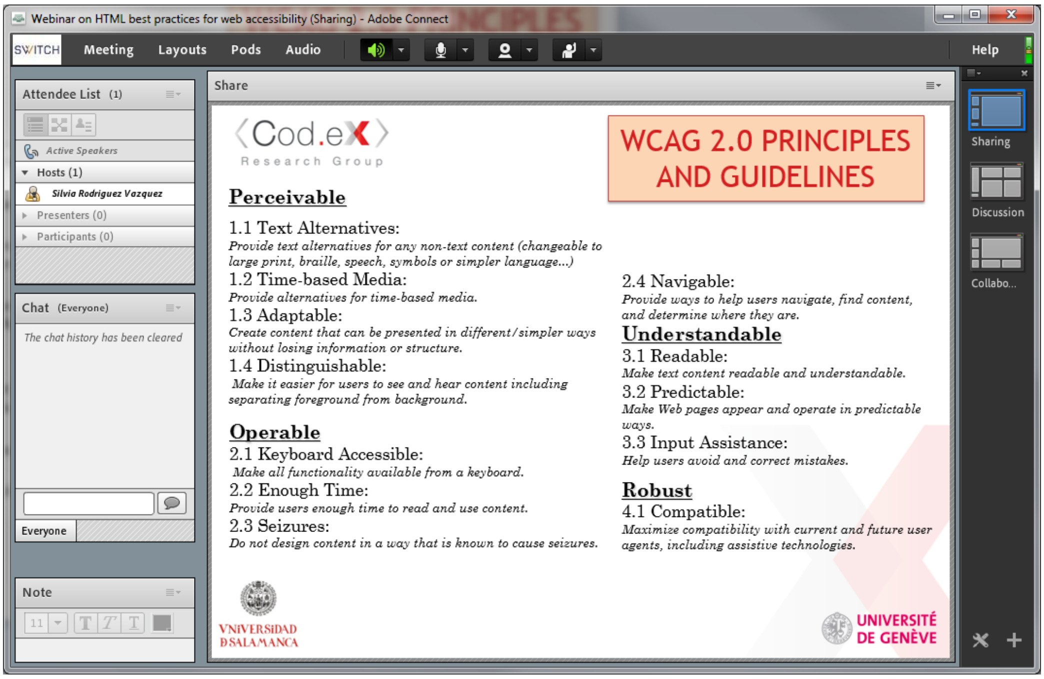


Figure 5.6. Webinar administration environment

While no special emphasis was placed on any of the SC and associated techniques reviewed, detailed explanations were given with regards to image accessibility. Concretely, the difference between decorative, descriptive and functional images was highlighted to participants, who could see on the screen which value should be ideally assigned to the `alt` attribute in each case thanks to

[8] 'Introduction to the Screen Reader', University of Wisconsin: https://www.youtube.com/watch?v=VvWCnFjAGgo Last access: 22nd July 2015.

meaningful examples. From the moment that attention was drawn to the next relevant SC, image accessibility was not further referred to in the webinar. We are therefore confident that no bias was introduced as a result of the researcher unintentional expectancy effect (Saldanha and O'Brien 2014, 30).

The material used for the webinar was leveraged from the web accessibility seminars held at the USAL and the UNIGE in the autumn and spring semester respectively from November 2012 to May 2015. The observational study reported in Rodríguez Vázquez (2014) showed that, after following the aforementioned seminars, localisation students took web accessibility best practices into account for the course final projects, which demonstrated not only a high degree of acceptability of the topic, but also the seminars' content validity and immediate applicability. Although it is generally understood that students are more receptive and open to profession-related challenges and new technologies, it was expected that the webinar would have a similar impact on our participants' performance.

Methodology test

Prior to providing participants with information about the experiment's logistics, it was important to check whether the preparation material for the study and the experiment documents were usable and easy to understand. Similarly, validating the procedure developed (see Figure 5.3) was considered necessary to ensure the experiment's feasibility and potentially improve its reliability, should any flaw in the experimental design be identified.

To this end, two people were recruited to complete the localisation and QA tasks, as well as their related questionnaires. They were two female graduate students, employed as research and teaching assistants at the Faculty of Translation and Interpreting of the UNIGE. Both were French native speakers with an MA in Translation and Interpreting by the aforementioned university, one of whom had followed the web accessibility seminar held in May 2013 as part of the Localisation and Project Management course. None of them had previously used any of the tools proposed and were not familiar with the ultimate goal of the research.

The test was conducted simultaneously by both participants on 27th November 2014. One week before, they had been sent all the documents needed to install and configure the screen recording and the computer remote control software, which they did without experiencing any particular problem, acknowledging that instructions were clear and that both programmes were easy to install. While for the first session, they were asked to follow the same instructions, task briefings were different for the second session, where one tested scenario A and the other one used the QA tools in the order established for scenario B. Both sessions were completed successfully by the two participants within the time initially scheduled, since none of the tasks requested took them more than 80 minutes. Overall, only two major technical issues were identified.

Firstly, participants were told to upload the first session's deliverables to the FTP server from where the source website to be localised could be downloaded. However, when trying to do so, the size of the videos (between 400Mb and 600Mb) generated by BB FlashBack seemed to cause a never ending 'uploading' loop which eventually resulted in web browser failure. Likewise, after the QA task, participants were requested to drag and drop the video file of the second session from their personal computer to the remote desktop. Although no apparent problem was detected, TeamViewer indicated that the transfer would take several hours, which was regarded as inconvenient by both participants. They could work on other tasks or leave the computer unattended during that period of time, but either option would entail a higher time investment and a risk of file loss. It was therefore decided to propose WeTransfer as an alternative solution for file transfer upon task completion, as shown in Appendices G.1, G.2 and G.3.

Secondly, one participant reported that conducting the second task was particularly challenging due to the small size of the remote desktop screen. She performed the entire test on a 13-inches laptop, estimating that this might as well have hindered the usability of TeamViewer 10. In order to prevent this from jeopardising the ecological validity of the study, we requested participants to carry out both tasks using a screen of at least 17 inches. Likewise, instructions on how to adjust the visual settings in TeamViewer 10 were included in the customised user guide sent to all participants in the main study (see Appendix E.3).

No other changes were implemented in the experiment procedure, which was otherwise found appropriate by the two volunteers who took part in the feasibility study. The use of Adobe Connect was introduced later in the process, so the software was not tested in such a context. Similarly, the webinar planned for the main experiment was not offered to the participant who already had some web accessibility background because she had followed a similar training module one year before. Data collected during the methodology test will not be considered in the research findings reported in Chapters 6 and 7.

5.3 Stage 2. User evaluation

In Chapter 2, by adopting a functional-pragmatic approach, we framed web accessibility as a relevant component of web localisation quality. This necessarily implies that a greater emphasis is placed on the end user as the targeted community when applying specific quality measurements. For the purposes of this thesis, we have defined screen reader users as the main subgroup in the target language community and image accessibility as the primary quality measurement. Therefore, in order to assess the outcome of the web localisation experiment described in the previous section, we have conducted a user evaluation in which the appropriateness of text alternatives was judged by seven screen reader users. We have thus assumed that, the more appropriate alt texts are, the higher the level of image accessibility achieved.

In this section, we will first discuss which considerations were taken into account when designing the evaluation methodology (5.3.1). We will then provide information about the end users' recruitment and profile (5.3.2), as well as the evaluation data set (5.3.3). Finally, we will conclude this section with a comprehensive description of the evaluation environment (5.3.4).

5.3.1 Evaluation design

As stated by Lawton Henry (2007, 96), when focus is only placed on the technical aspects of accessibility, the human interaction aspect can be lost. In this sense, we did not want to aim only at verifying whether participants involved in the localisation experiment had made sure that every `<img>` element contained an `alt` attribute. Instead, we sought to investigate whether the value introduced for each `alt` attribute was pertinent with respect to the image function and context within the web page. This distinction goes in agreement with algorithmic and judgement testing for image accessibility compliance, as understood by Thatcher et al. (2006, 408). While algorithmic testing serves to verify the presence of an `alt` attribute, judgement comes in when there is a need to evaluate the quality of the image text equivalent.

Drawing upon Nord's concept of instrumental translation (1997) −or the notion of covert translation, as defined by House (2013, 535)−, which denotes a translation that enjoys the status of an original source text in the target culture and whose audience is equally directly addressed as the source's, a monolingual evaluation approach was adopted. A linguistic-driven comparison of both source and target image text equivalents was thus out of the scope of our research. Similarly, our interest did not lie in performing a thorough linguistic analysis of the target alt texts, but in observing (i) whether text alternatives had been identified as translatable content and, as such, adapted in the localised web product; and (ii) how end users perceived them.

In the sections that follow, we report the lessons learned from an exploratory user evaluation carried out prior to the web localisation experiment and the assessment of its outcome (5.3.1.1) and we describe how text alternatives' appropriateness was finally operationalised (5.3.1.2).

5.3.1.1 Exploratory study

According to Rubin & Chisnell (2008, 29), an exploratory study is conducted quite early in the development cycle, when a product is still in the preliminary stages of being defined and designed. During the rule development process presented in Chapter 4 (see section 4.3), we deemed it appropriate to conduct a preliminary user evaluation with a view to testing some of the rules defined so far for alt text checking. The main objective of this exploratory study was to verify our initial assumptions about the rules' effectiveness regarding alt text quality improvement. However, it also served us to test whether a third-party evaluation methodology that had already been designed to assess the impact of *Acrolinx's* controlled language (CL) rules within the framework of a European research project coordinated by the UNIGE was also appropriate for the purposes of our research. The exploratory study took place in mid-March 2014 and the results have been published in Rodríguez Vázquez et al. (2014). Since the twelve rules tested were preliminary and suffered further modifications until reaching a refined definitive version, we will only discuss here the most relevant aspects of the evaluation methodology followed.

Evaluation data set

A total of 901 web pages belonging to a small web corpus of twelve websites (see Appendix A.2) were verified against the rules selected using *Acrolinx Batch Checker*. The web document collection, retrieved with the help of Scrapbook during the last week of February 2014, consisted of seven Swiss French websites not present in web corpus I (see Appendix A.1) and five websites that were part of it, but whose content or structure had changed since 2011. It contained 3,207 images, 1,731 of which had a unique non-empty alt text. After the batch check, 110 text alternatives flagged as inappropriate by the tool were randomly selected and changes were then manually applied in order to satisfy the rule described.[9] This led to a sample of 110 raw text alternatives and 110 controlled text alternatives (N=220).

Participants

Four participants were recruited for the user evaluation following a purposive sampling technique. Because we implicitly wanted to validate whether our partial formalisation of alt texts' appropriateness was pertinent, we deliberately selected a biased sample of evaluators who were web accessibility experts. Two were sighted users and two were blind users (aged between 28 and 39, $\bar{x}$=35, sd=4.99, all males). All participants had been working on web accessibility projects for more than four years ($\bar{x}$=5.5, sd=3.1) and self-rated their knowledge about the appropriateness of text alternatives for images in the web as advanced.

Evaluation environment

The goal of the evaluation was to know whether improvements in the alt texts' appropriateness were achieved thanks to the application of the CL rules. Following

[9] At the time of the exploratory study, no improvement suggestions were offered by *Acrolinx*. Rule help files did not include a comprehensive description of the problem detected either. Modifications made were thus based on a brief two-sentence description of the rule that had been contravened, as well as on the background knowledge of the researcher with regards to image accessibility.

the method adopted by Gerlach (2015, 110–114), we requested users to judge whether a controlled text alternative was more appropriate than the corresponding raw text alternative. For each pair, they were asked to give their judgement according to a five-point scale: first clearly better, first slightly better, about the same, second slightly better, second clearly better. Alt text pairs were presented in a randomised order to prevent evaluators from knowing which was the text alternative that had been modified with a view to improving its quality.
Both sighted and blind users carried out this comparative assessment in SurveyMonkey,[10] the survey administration tool previously presented in this chapter. In order to offer users the opportunity to make an informed judgement, the following mechanisms were established. Each text alternative proposal had an embedded link to the page in which its corresponding image appeared. At the top of each page, we introduced a "Skip to image for evaluation" link that jumped the evaluator down to an anchor located before the image whose alt text needed to be assessed. This technique was used to facilitate the quick and easy location of the image, especially for blind users. Sighted users were visually presented with the image as well, directly in the evaluation environment.

Lessons learned

While the exploratory study was conducted successfully and returned significantly positive results for half of the rules tested (Rodríguez Vázquez et al. 2014), the following limitations were observed and taken into account during the design of the main user evaluation:

- *Assessment methodology*: The comparison-based evaluation method implemented was convenient because it enabled users to perform the task relatively faster than if alt texts had to be assessed individually. However, the major drawback was that evaluators did not receive any specific indications about the criteria that define a "good" alt text. The question asked for each alt text pair was the following: Which image text alternative is better in terms of accessibility? (in French: 'Quel équivalent textuel convient le mieux en termes d'accessibilité ?'). Then, users were free to decide what they would take into consideration before making their judgement, rendering results highly subjective in nature. Since this was considered as a threat to measurement validity (Saldanha and O'Brien 2014, 33), it was decided that, prior to the main user evaluation, the concept of appropriateness when referring to image text alternatives had to be carefully operationalised. Similarly, when the intermediate value judgement "about the same" was made, no mechanism was

[10] In this case, it was crucial for the feasibility of the alt text assessment task by screen reader users to create and accessible questionnaire. According to SurveyMonkey accessibility disclaimer, surveys created and administered using this software are compliant with the US Section 508 of the Rehabilitation Act, which relies on the WCAG 2.0.

available to know whether evaluators had thought that the alt texts proposed were equally “good” or equally “bad”. An individual assessment was therefore considered more suitable to accurately measure the impact of the rules and maximise the validity of the results.

- *Evaluation setup*: By providing users with the possibility of assessing image accessibility in context, the evaluation environment set up was very similar to that of user testing, one of the most efficient accessibility evaluation techniques, reviewed in Chapter 2. While this contributed to boost the ecological validity of the study, we believe that blind and sighted evaluators did not perform the task under comparable conditions. Sighted users could easily understand the context of the image by quickly examining the surrounding content, even unconsciously. Nevertheless, for blind users, exploring the web page involved a higher effort. In fact, they openly acknowledged that they had checked the web site provided only in 25% of the cases before making their judgement. A different mechanism was thus later defined to allow evaluators to assess the alt text taking into account the context of its corresponding image.
- *Participants*: As web accessibility experts, users pointed at their own professional experience as the main foundation for their judgements. This might have had an influence on the results, especially when the image context was not checked. For the final user evaluation, we planned to include users with different backgrounds. In addition, although input from sighted users was important for validating our initial assumptions about the utility of the rules, we limited the pool of participants to blind evaluators, who are the main end user group in our research.
- *Sample size*: Finally, we believe that the small sample of text alternatives subjected to evaluation (both per rule violated and as a whole) prevented statistical significance from appearing when analysing the impact of certain rules. Similarly, we considered that the number of evaluators should be ideally higher in order to draw more solid conclusions.

5.3.1.2 Measurement of DV2. Appropriateness of alt texts

As indicated at the beginning of this chapter, the main concept addressed in two of our three research questions is the appropriateness of text alternatives for images. Translation quality will therefore be measured upon this criterion. It is generally understood that the more appropriate the alt text is, the higher the level of image accessibility achieved. In turn, appropriateness depends on the pertinence and accuracy of the description provided for a given image, according to its function and context within a web page. Since a comprehensive discussion about the matter has been presented in previous chapters of this thesis, we shall only describe here the tailor-made evaluation metric designed for the purposes of the experimental study.

The exploratory study carried out revealed that the abstract notion of text alternatives' appropriateness had to be defined in advance and presented to the evaluators in order to be properly measured. For the operationalisation of this dependent variable (Saldanha and O'Brien 2014, 24), we developed a grading scale on four levels: 1. *Not appropriate*, 2. *Acceptable*, 3. *Pertinent*, 4. *Very pertinent* (see Table 5.7). The definition of each value in this ordinal scale was drawn upon the literature and the personal experience of two screen reader users who checked the criteria for content validity (Oates 2005, 227).

Table 5.7. Graded Likert-type scale with four rating levels to assess alt text appropriateness

Score	Value	Criteria
1	Not appropriate (*Pas acceptable*)	The text alternative is not appropriate for the image, according to the location and context described.
2	Acceptable (*Acceptable*)	The text alternative is acceptable for the image, according to the location and context described, but not all the information provided is necessarily pertinent.
3	Pertinent (*Pertinent*)	The text alternative provides minimal but sufficient and correct information about the image, according to the location described.
4	Very pertinent (*Très pertinent*)	The text alternative provides complete and precise information about the image, according to the location and context described.

Relying on accessibility conformance-based metrics, which are founded on whether success criteria (SC) of given guidelines are met, was not deemed appropriate because we wanted to assess only one common web accessibility failure (SC 1.1.1). Furthermore, text alternatives' quality goes beyond merely inspecting the code for accessibility conformance, rendering unsuitable binary scoring scales, such as the one used in the failure-rate metric (Sullivan and Matson 2000). On the other hand, assessing the appropriateness of alt texts by measuring their length or syntactic complexity was dismissed since previous work has highlighted that these are subjective parameters and that it is difficult to establish a clear baseline (Petrie et al. 2005). The rating approach adopted in our research is based on the proposal of Fischer and Wyatt (2011), who advocate for a higher level of precision in accessibility test procedures. Since their graded rating system is intended to assess all WCAG SC, it includes five rating levels: pass, marginally acceptable, partly acceptable, marginally unacceptable and fail. Nevertheless, given the nature of the SC we wanted to evaluate, we decided to provide just one negative value in our scale (1. Not appropriate), thus leaving room for nuances only in the case of positive annotations. Similarly, our rating scale is in agreement with other metrics designed for measuring text comprehensibility (Roturier 2006, 85–86), acceptability and readability (O'Brien 2010), which are also four-point scales defined for human evaluation of translation output.

Another example of end user-focused translation quality measurement that can be found in the literature is usability. In Doherty & O'Brien (2012), users rated their satisfaction regarding the usability of raw machine translated text on a five-point Likert scale. Likewise, in Doherty (2012), participants' opinion about the readability and comprehensibility of the text they had read was measured using the same standard scale. In this study, other methods were triangulated with user judgement to investigate the impact of CL rules on the aforementioned dependent variables, including eye-tracking, automatic evaluation metrics and retrospective interviews. While debriefings with screen reader users could have also proved useful in our study, it was decided that the vast amount of quantitative data collected (see section 5.3.3) could be sufficient to infer the annotation trend of each evaluator. Moreover, we estimated that the task duration would be between five and eight hours long (see section 5.3.4.1) and considered that users would not be comfortable with devoting a higher time investment to the study.

As House asserts (2013, 545), any evaluation depends on a large variety of factors that necessarily enter into a social evaluative judgement. However, the intrinsic subjectivity of human judgements can be compensated by trying to define a priori the criteria on which they will be based. We believe that the scale presented in Table 5.7 conveniently denotes our attempt to turn alt text appropriateness into a measurable entity. In addition, the assignment of scores from 1 to 4 to each level of the scale seemed advantageous not only to apply quantitative data analysis to what originally would be treated as qualitative data, but also for end users to understand the hierarchical distribution of the values, where 1 should be understood as synonym with a "wrong" alt text and 4 with an "excellent" text alternative proposal.

5.3.2 Participants

5.3.2.1 Recruitment process

A snowball technique (Oates 2005, 98) was chosen to recruit screen-reader users, with a view to maximising the possibilities of having evaluators with diverse backgrounds, as concluded from the exploratory study (see section 5.3.1.1). The call for participation (CFP) was advertised on 4th April 2015 through social networks, including Facebook and Twitter; the Round Table mailing list, a discussion forum for visually-impaired language professionals; and direct communication with personal contacts who either matched the criteria we were looking for or had access to members of the population under investigation.

Requirements were (i) having advanced proficiency in French and good knowledge of English, and (ii) being experienced users of assistive technology and the Web. The first requirement made the recruitment process particularly challenging and, together with the inherent difficulty of reaching out to this population sub-group, it might explain the final low response rate. In this sense, applying crowdsourcing techniques (based on the principle of microtasking) might have proved helpful, but splitting the evaluation task was not deemed appropriate. The reason is that, in order to get reliable results, each participant would have needed to annotate all alt texts corresponding to one of the rules violated. Still, following this procedure, alt texts produced by the same translator would have been annotated by different evaluators, thus not allowing us to correctly assess their overall performance.

Along with the requirements, the CFP clearly stated that the evaluation was framed within a research project on web localisation and image accessibility. Likewise, it announced that assessing text alternatives was the primary scope of the study. Nevertheless, further details such as the duration and the task procedure were only provided upon expression of interest. Communication with potential participants prior to the user study was mainly written, via e-mail. However, some participants preferred to interact with the researcher through the exchange of voice recordings or Skype voice calls. For instance, participant B2 declared: “I prefer receiving audio messages, as they give me a break from the metallic voice of my screen reader”. From the nine people who initially signed up for the evaluation, only seven completed the task. Although it might be considered a small sample size, five users is the standard number often required in HCI studies (Lazar et al. 2010, 263). Likewise, this number is referred to when the sample of evaluators needed to perform a quality assessment task is discussed in Saldanha and O'Brien (2014, 107).

5.3.2.2 Participants profile

All seven participants were blind (aged between 26 and 41, $\bar{x}=34$, sd=5.77, three female, four male). At the time of the evaluation, they were all located in a French speaking country or in a country where one of the official languages was French: Canada (N=1), France (N=2) and Switzerland (N=4). However, only five declared to have French as their mother tongue. The remaining two were German native speakers with a high proficiency in French. English was the most common second language spoken by all evaluators. All declared to have advanced skills with regards to the use of their assistive technology. One blind user indicated that VoiceOver for Mac OS was his preferred screen-reader. The rest were regular JAWS users, occasionally choosing NVDA when the former was not available.

Despite our efforts to include a wide variety of profiles, we observed two main groups: linguists and web experts. On one hand, female participants had more than three years of experience in language-related jobs ($\bar{x}=4$, sd=1): two were translators and the third participant worked as a document transcription provider. All acknowledged using braille displays for better work performance, and two of them reported to have some basic knowledge of web accessibility. On the other hand, male participants reported to have a rather technical background, with three of them working as web accessibility consultants. The fourth evaluator was an experienced web developer who had some knowledge of the matter but never worked in accessibility-related projects. All had more than four years of professional experience ($\bar{x}$=7.5, sd=3.51).

5.3.3 Evaluation data set

The primary data[11] retrieved from the experiment described in section 5.2 consisted of 84 versions of the same website (three per localiser, corresponding to T1, T2 and T3), accounting for 252 web pages, with a total of 10,920 images. These large numbers had to be reduced for a manual human evaluation to be feasible. Randomly selecting a sub-set of web pages was not an option, since we wanted to examine the improvements, if any, made by individual localisers throughout the different translation versions. Basing the sampling process on a selection of participants had to be dismissed if significance tests were to be applied for data analysis. Another possibility was to measure alt text appropriateness taking only into account one image type (for instance, functional images). However, this would not have allowed us to study the effectiveness of all the CL rules selected for evaluation.

Therefore, we finally collected all `alt` attribute values produced during the controlled localisation experiment (28 translators × 130 images × 3 translation versions, N=10,920 text alternatives). These data were isolated and coded following the methodology described in Chapter 4 (section 4.3), used to prepare web corpora I, II and IV for exploitation. A sample of 2,189 unique text alternatives was obtained after data clean-up. This reduction was achieved by simply eliminating duplicated text alternatives across participants and translation versions. The final evaluation data set was then organised per image. The number of different localised text alternatives to be assessed varied across images and ranged between three (N=3) and 32 (see Appendix J to check the exact number of text alternatives assessed per image).

[11] As explained in section 5.2.4, screen recordings were only collected for ensuring that tasks were correctly followed and will thus not be considered for analysis in this thesis. The same applies to tool check reports, which serve us later to verify whether the checks were run correctly.

5.3.4 Evaluation environment

Upon satisfactory administration of the exploratory alt text evaluation as a questionnaire-like exercise implemented through SurveyMonkey (see section 5.3.1.1), the same data generation method was maintained for the main user evaluation study.[12] Nonetheless, as mentioned earlier, substantial changes were made regarding the task and the procedure that followed with respect to the preliminary study conducted.

5.3.4.1 Task

Screen-reader users were requested to assess the appropriateness of 2,189 different text alternatives, corresponding to the 130 images of the experimental website. To do so, they had to base their judgement on the scale presented in section 5.3.1.2. Simulating the standard conditions of a real accessibility user testing setup, as was done in the exploratory study, would have implied asking evaluators to browse 2,189 web pages. Since this seemed an unreasonable demand, users received instead detailed information about the website from which all images had been extracted: what the campaign was about and who its initiators were. The structure of the website was also described, indicating that each page corresponded to one of the campaign partners. Similarly, evaluators were provided with exhaustive information about the macrostructure (header, body, footer) of each web page and an overview of its content (see Figure 5.7). These comprehensive explanations aimed at helping users to better picture the website from where the images had been retrieved, and contributed to increase the ecological validity of the study.

Each image was presented to the evaluators on a separate page of SurveyMonkey for ease of navigation purposes. At the top of each page, the evaluator would find the following data: (i) the website's page where a given image appeared, (ii) its relative location within that page (as per the macrostructure shown in Figure 5.7), and (iii) a neutral description of the image's context, referring to the elements or sections illustrated in Figure 5.7 when necessary. Immediately after, blind users could read how many text alternatives they would need to assess, as well as the list of text alternatives, introduced in a random order (see Figure 5.8).

[12] Questionnaires are one of the standard user evaluation paradigms, and they may include open-eded questions to allow respondents to answer in a free narrative form, and/or closed questions with a predetermined set of responses deemed appropriate by the evaluator (Jay et al. 2008).

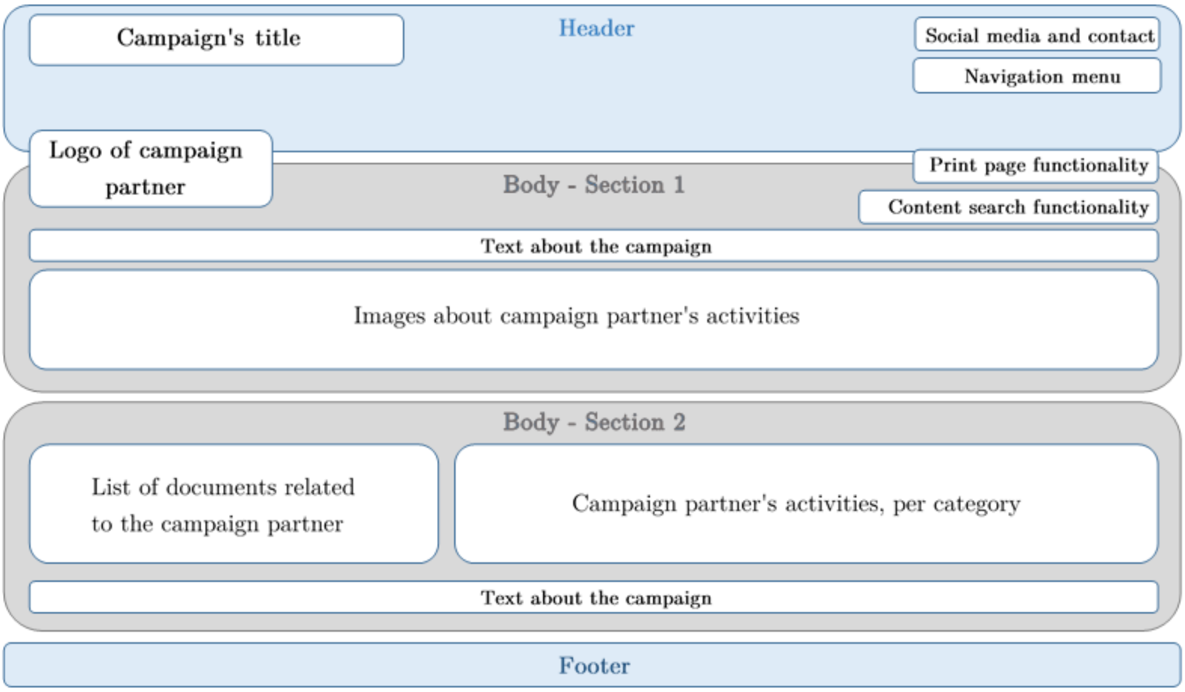
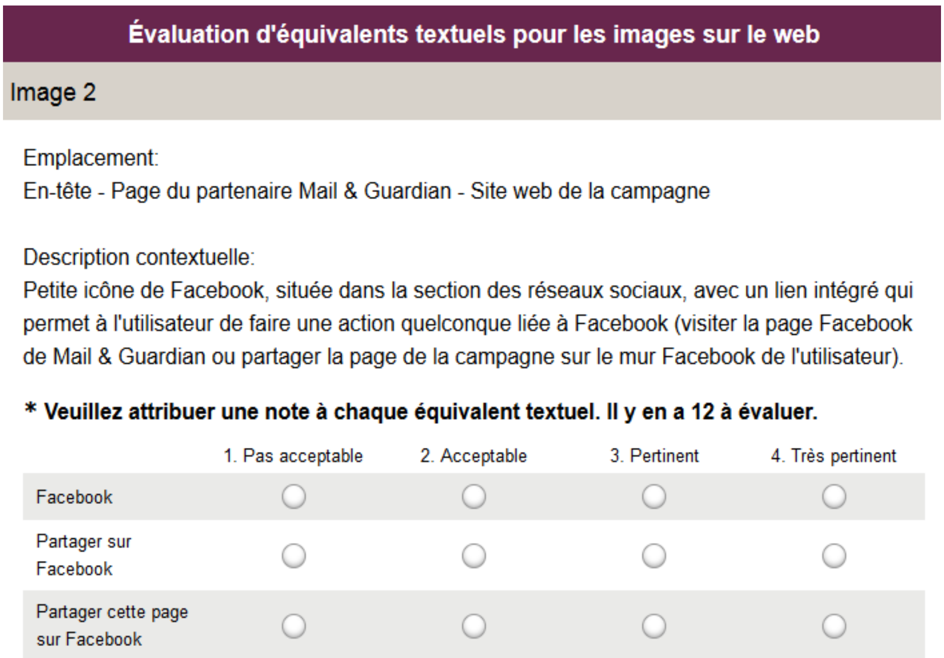


Figure 5.7. Standard web page macrostructure (header, body, footer) and content, as explained to evaluators



Évaluation d'équivalents textuels pour les images sur le web

Image 2

Emplacement:
En-tête - Page du partenaire Mail & Guardian - Site web de la campagne

Description contextuelle:
Petite icône de Facebook, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Facebook (visiter la page Facebook de Mail & Guardian ou partager la page de la campagne sur le mur Facebook de l'utilisateur).

*** Veuillez attribuer une note à chaque équivalent textuel. Il y en a 12 à évaluer.**

	1. Pas acceptable	2. Acceptable	3. Pertinent	4. Très pertinent
Facebook	○	○	○	○
Partager sur Facebook	○	○	○	○
Partager cette page sur Facebook	○	○	○	○

Figure 5.8. Text alternatives evaluation environment

The contextual description provided for each image can be found in Appendix J. As can be seen from the list of 130 descriptions, some are repeated. The reason is that, as explained previously in this chapter, certain images were used more than once in the experimental website. Merging their associated text alternatives in a unique list to avoid these duplicates would have compromised the usability of the evaluation environment. It should be remembered that, although the platform was accessible, navigating through data presented as a table (see Figure 5.8) is still cumbersome, since the screen reader announces the column and/or row heading each time the user enters a new cell. Therefore, the 130 images were presented individually.

5.3.4.2 Procedure

The user evaluation took place between 12th and 22nd April 2015. Upon acceptance of the task, evaluators were sent via e-mail an accessible MS Word file with all the instructions needed to perform the evaluation in French (see Appendix I). The document included the following contents: an introduction to the study; a detailed description of the website's purpose and each page's macrostructure, information about how the questionnaire was organised, and an explanation of the score values in our rating scale. A brief statement guaranteeing confidentiality of the data gathered was also added for the users' comfort. Through the first page of the questionnaire we collected demographic data about the evaluators. The second page contained questions aimed at gathering users' personal general opinions about the accessibility of multilingual websites and of images on the web. From the third page on, each one of the 130 images was presented together with a list of text alternatives to evaluate.

Participants were told how instances of no `alt` attribute or null alt text were coded, in order to facilitate its interpretation during the evaluation task. We also informed them that the image text alternatives they were about to assess were extracted from a website translated by multiple translators, hence the presence of alt text not only in French, but also in English. Still, they were requested to assign scores on the premise that, in a real life scenario, they would be browsing a French website. We deliberately decided not to send them the source English website for reference to avoid any potential bias. We believe that, if given access to this website, evaluators would have been tempted to assess the text alternatives based on their textual translation accuracy with respect to the source, and not on their appropriateness in terms of accessibility.

We had estimated that the assessment exercise could take between five and twelve hours, so we also provided evaluators with tips about how to enable the browser's cookies before starting the SurveyMonkey questionnaire. This would allow them to take breaks during the task or work on it over several days without losing their responses to the questions already completed. No other technical requirements were specified, since SurveyMonkey supports most browsers, including Chrome, Firefox, Safari and Internet Explorer, which is the most common browser used by screen reader users. The link to the questionnaire was both included in the e-mail sent to evaluators and at the end of the instructions file. Each screen-reader user received a monetary compensation of CHF 100 as an acknowledgement for the work done.

5.4 Hypotheses

In this section we recall the three main research questions of the experimental study, already announced in Chapter 1 (see section 1.4) and we set forth their associated hypotheses and sub-hypotheses. While **R1** can be answered only by analysing the data collected from the web localisation experiment, drawing conclusions about **R2** and **R3** requires the analysis of the data gathered through the user evaluation.

R1. Are image text alternatives considered as translatable elements by localisers during the web localisation process?

- **H0 (null hypothesis).** Image text alternatives are not considered by localisers as translatable elements during the web localisation process.
- **H1 (alternative hypothesis).** Image text alternatives are considered by localisers as translatable elements during the web localisation process.
    - **H1.1.** Having basic knowledge of web accessibility helps localisers to identify text alternatives as translatable content during the web localisation process.
    - **H1.2.** The use of accessibility-oriented QA tools helps localisers to identify text alternatives as translatable content during the web localisation process.

R2. Does the use of accessibility-oriented QA tools during the web localisation process result in more appropriate text alternatives for images?

- **H0 (null hypothesis).** There is no difference in terms of appropriateness between image text alternatives that have been validated by means of accessibility-oriented QA tools and those which have not been validated.
- **H1 (alternative hypothesis).** Using accessibility-oriented QA tools during the web localisation process has a positive impact on the appropriateness of translated text alternatives for images.
    - **H1.1.** When only one evaluation tool is used, a controlled language tool with style-oriented rules for image text alternatives' checking like *Acrolinx* helps localisers to achieve more appropriate text alternatives than a general web accessibility evaluation tool like *aDesigner*.

- **H1.2.** When two evaluation tools are used, a controlled language tool with style-oriented rules for image text alternatives' checking like *Acrolinx* leads to more improvements than a general web accessibility evaluation tool like *aDesigner*, irrespective of the order in which they are used.
- **H1.3.** Using two different tools trigger more improvements in terms text alternatives' appropriateness than using only one.

R3. Does knowledge of web accessibility help localisers to produce more appropriate text alternatives for images?

- **H0 (null hypothesis).** There is no difference in terms of appropriateness between image text alternatives that are translated by localisers with web accessibility knowledge and those translated by localisers without that previous knowledge.
- **H1 (alternative hypothesis).** Having web accessibility knowledge has a positive impact on the appropriateness level achieved in the translated text alternatives for images.
  - **H1.1.** Localisers with basic knowledge of web accessibility that use accessibility evaluation tools during the localisation process achieve more appropriate text alternatives than localisers who do not have previous knowledge of web accessibility and have used the same tools.
  - **H1.2.** Localisers with basic knowledge of web accessibility that use a controlled-language tool with style-oriented rules for image text alternatives' checking during the localisation process achieve more appropriate text alternatives than localisers who do not have previous knowledge of web accessibility and use the same controlled-language tool.

5.5 Summary

The core experimental study of this research consisted of two stages: a controlled web localisation experiment (Stage 1) whose output was then subjected to human evaluation using a remote user testing technique (Stage 2). The ultimate goal of the overall study was to investigate whether localisers identified text alternatives as translatable elements during the web localisation process, and whether the resulting localised alt texts provided a pertinent and accurate description for the image in question, according to its function and context within the web page. For the purposes of the study, we manipulated two independent variables (IVs): having web accessibility (WA) knowledge (**IV1**) and using two different accessibility-oriented QA tools (**IV2**), with a view to observing their effect on the image text alternatives' visibility and appropriateness (see Figure 5.9).

At the beginning of this chapter, we stated the three research questions we sought to address. We then discussed the methodological aspects related to the first stage of this empirical study, including its design, participants, material and setup (section 5.2). A similar structure was followed in section 5.3 to describe how the user evaluation was conducted. Finally, in the previous section (5.4), we formulated the hypotheses and sub-hypotheses that will be tested in the next chapter (6).

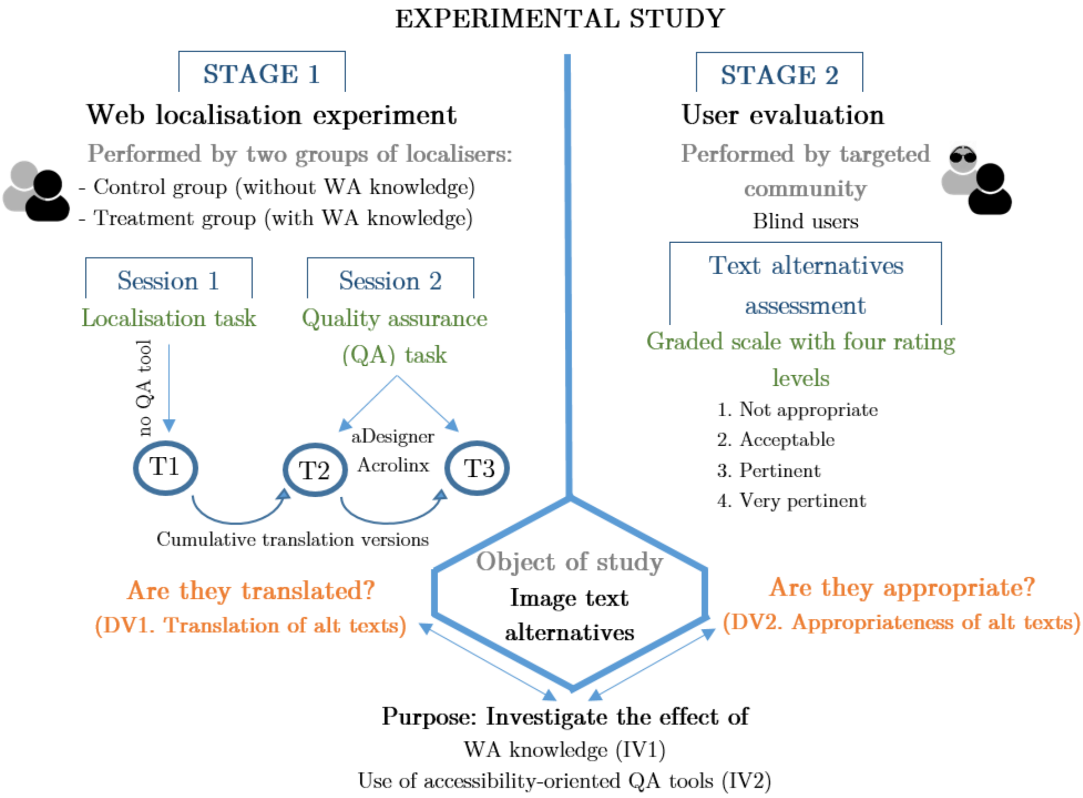


Figure 5.9. Overview of the experimental study

Chapter 6

Experimental study: main results and discussion

This chapter aims at examining the research questions associated with the central experimental study of this thesis by testing the hypotheses and sub-hypotheses set forth in Chapter 5. To this end, we present how data collected throughout the experimental study was processed and analysed, and we report and discuss the main results obtained.

6.1 Overview

The empirical results of this thesis, which represent the output of the research strategy adopted to attain Goal 3 (see Chapter 1, section 1.3.), are described and interpreted in Chapters 6 and 7. The present chapter (6) focuses on the effect of the two primary independent variables —*IV1. Knowledge of web accessibility (WA)* and *IV2. Use of accessibility-oriented quality assurance (QA) tools*— on the two main dependent variables of our study —*DV1. Translation of alt texts* and *DV2. Appropriateness of alt texts*. Additionally, we also consider the crossed effect of the aforementioned IVs and the QA scenario (A or B) followed by localisers during the QA task on our two DVs.

The study of **DV1** is covered in section 6.2, while the analysis of **DV2** is addressed in section 6.3. Apart from the results and their interpretation, each section provides, when appropriate, a description of how the data was processed and coded, as well as the statistical methods used. At the end of each section, we come to a conclusion as to whether our hypotheses and sub-hypotheses can be supported or need to be rejected. Before starting with the description and discussion of our findings, we deem it relevant, however, to first (i) summarise the type of data gathered throughout Stages 1 and 2 of our study and how these were organised (section 6.1.1), and then to (ii) identify the experimental variables investigated in both this and the next chapter (section 6.1.2).

6.1.1 Experimental data

Our experimental study yielded two different sets of data. In Stage 1, we collected information about both the localisation and quality assurance (QA) process in the form of screen recording video files and check reports from the accessibility-oriented QA tools used.[1] This data set, as we announced in the previous chapter, was primarily used to ensure the validity of the research (i.e., to check that instructions were correctly followed and to verify how product-related data was generated should problems occur during data processing and coding). Therefore, it will not be considered for analysis. In addition, throughout Stages 1 and 2, we retrieved data from web documents and three questionnaires: two post-task questionnaires associated with the web localisation task and the quality assurance task respectively, and the text alternatives evaluation questionnaire completed by screen reader users. These data represent the main source of empirical evidence that will be analysed and reported both in the present chapter and in Chapter 7.

To facilitate the observation of the data and its statistical analysis, we built three core databases:

a) *Text alternatives localisation database* (DB1): This consisted of a comparative summary of all the instances of text alternatives gathered during the web localisation experiment, with a total of 21,840 records. It comprised 28 individual sub-databases (one per localiser), including all possible alt text comparisons between the source text (ST) and among translation versions (T1, T2 and T3, see Table 5.1, Chapter 5): ST-T1, ST-T2, ST-T3, T1-T2, T1-T3 and T2-T3. Data in DB1 was explored with a view to investigating the identification of text alternatives as translatable elements by localisers (see section 6.2).

b) *User evaluation database* (DB2): This included 76,440 records accounting for all judgments made by the seven blind users (hereinafter also referred to as 'judges') regarding the appropriateness of the text alternatives produced by web localisers (28 localisers × 130 images × 3 translation versions × 7 judges). The outcome of the evaluation consisted initially of 15,323 records (2,189 text alternatives annotated × 7 judges). Duplicated text alternatives (N=8,731 out of 10,920 produced by the localisers), which had been previously removed from the evaluation data set to lighten the burden of the evaluation task, were then added to the database with their corresponding scores, hence the high number of records indicated above. The database contained ten fields: localiser ID, group, image ID, image source, image alt text, alt text type, judge ID, score, translation version (T1, T2, T3), and tool used (hereinafter also named 'translation condition': none, Acrolinx, aDesigner, both). To allow for a more refined analysis of the data, the last two variables were combined to create an 11th field named 'tool specific', which accounted for both the tools and the order in which they were used, leading to five

[1] See section 5.2.4.1 and 5.2.4.2 (Chapter 5) for more details about the experimental procedure and the files delivered by the participants.

different levels: none (T1), Acrolinx (T2), aDesigner (T2), Both-last-Acrolinx (T3) and Both-last-aDesigner (T3). Database DB2 was exploited to draw conclusions regarding the appropriateness of localised text alternatives (see section 6.3).

c) *Questionnaires' responses database* (DB3): This consisted of three individual databases of 28 registers each, corresponding to the screening questionnaire, the localisation post-task questionnaire and the quality assurance post-task questionnaire. While almost all data collected is quantitative in nature, some open-ended questions included in the post-task questionnaires also returned qualitative data. We use the latter to supplement the observation and statistical analysis of the former. Data compiled in DB3 will be discussed in Chapter 7.

6.1.2 Experimental variables

In order to investigate the three research questions set forth in Chapter 5 (see section 5.1), we took into consideration not only the main IVs and DVs stated at the beginning of this chapter, but also other secondary independent (or explanatory) variables associated to the experimental design and the profile of the localisers. Table 6.1 provides an overview of all the variables we studied (a total of 11), the nature of the data they are associated with, and how they were operationalised (measurements).[2]

In what follows we briefly explain which variables were analysed (individually or in interaction with others) depending on the research question we were seeking to answer:

R1: *Are image text alternatives considered as translatable elements by localisers during the web localisation process?*

To investigate **R1**, we focused on the study of *DV1. Translation of alt texts* by measuring how it was influenced by the two main independent variables manipulated (*IV1. WA knowledge* and *IV2. Use of accessibility-oriented QA tools).* After observing data distribution, we considered it relevant to also measure the impact of two additional explanatory variables, related to our study's experimental design, namely the QA scenario followed and the alt text type. While the analysis of **DV1** was done without making any distinction with regard to the nature of the text alternatives produced by localisers, special attention was paid, in a subsequent step, to two of the four alt text categories: ‘empty `alt` attributes’ and ‘no `alt` attributes’ (see Table 6.1).

R2: *Does the use of accessibility-oriented QA tools during the web localisation process result in more appropriate text alternatives for images?*

[2] The explanation of the different measurements or levels defined will be provided when the corresponding variables will be analysed throughout the two chapters (6 and 7).

Our second research question was investigated by analysing the impact of **IV2** on *DV2. Appropriateness of alt texts.* We complemented our general results by observing the crossed effect of **IV2** and the order in which tools were used (i.e. the QA scenario followed). Similarly, we explored (i) the level of appropriateness reached by localisers in each alt type, depending on the tools used, and (ii) the impact of consulting the tools' help support on the overall quality of the alt texts. The analysis of these three secondary independent variables ('alt text type', 'check of Acrolinx help', and 'check of aDesigner error descriptions', see Table 6.1) will be addressed in Chapter 7.

R3: *Does knowledge of web accessibility help localisers to produce more appropriate text alternatives for images?*

Finally, to answer our third research question, we observed the effect of *IV1. WA knowledge* on **DV2.** The study **IV1** was then combined with **IV2** to examine the crossed effect of both independent variables on **DV2** (as we did for **DV1**). A more complex analysis was then carried out to explore the combined impact of these two IVs with a third variable: the QA scenario followed. While these data were deemed sufficient to answer **R3**, we considered it pertinent to explore the effect of other variables, not necessarily related to the accessibility-oriented QA tools used, on **DV2**: the use of CAT tools, the localisers' HTML self-rated knowledge, and the scores they obtained in the HTML test, during the selection process. These three secondary explanatory variables will be studied in Chapter 7.

As we have already announced in the introductory paragraphs of section 6.1, we have decided, for the sake of clarity, to organise the presentation of the results in both chapters (6 and 7) around the study of the different variables, instead of doing it according to the research questions investigated. Yet, a reminder about the latter will be included at the beginning of each corresponding section.

Table 6.1. Overview of the variables considered during data analysis

Variable			Name	Data type	Measurement / Levels
Dependent (response)			translation of alt texts (DV1)	ratio	total alt texts edited per translation version (T1, T2, T3)
			appropriateness of alt texts (DV2)	ordinal	1 = Not appropriate 2 = Acceptable 3 = Pertinent 4 = Very pertinent
Independent (explanatory)	**Primary**		WA knowledge (IV1)	categorical	control group (without WA) treatment group (with WA)
			Use of accessibility-oriented QA tools (IV2)	categorical	*4 translation conditions:* aDesigner / Acrolinx/ none / both
	Secondary	**Experimental design**	QA Scenario	categorical	*2 scenarios (based on tool order):* – aDesigner-Acrolinx (QA Scenario A) – Acrolinx-aDesigner (QA Scenario B)
			Alt text type	categorical	*4 alt text categories (based on the nature of the text alternatives assigned to the 130 images included in the experimental website, see Table 5.5):* – inappropriate alt texts, selected for the evaluation of 10 CL rules (see Table 4.6) – empty `alt` attributes – no `alt` attributes – appropriate alt texts
			Use of CAT tools	categorical	use of CAT / no use of CAT
			Check of Acrolinx rule help	ordinal	1 = Not checked 2 = Checked for 25% of errors 3 = Checked for 50% of errors 4 = Checked for 75% of errors 5 = Help always checked
			Check of aDesigner error descriptions	ordinal	1 = Not checked 2 = Checked for 25% of errors 3 = Checked for 50% of errors 4 = Checked for 75% of errors 5 = Help always checked
		Participants	HTML self-rated knowledge	ordinal	1 = No knowledge 2 = Basic knowledge 3 = Intermediate knowledge 4 = Advanced knowledge
			HTML test	ordinal	1 = No changes 2 = Translation of text in black 3 = Translation of 2 + keywords & description 4 = Translation of 3 + adaptation of Content-Language 5 = 4 + Translation of alt text

6.2 Translation of text alternatives

The results reported in this section will serve to test the hypotheses and sub-hypotheses related to research question **R1:** *Are image text alternatives considered as translatable elements by localisers during the web localisation process?*

6.2.1 Data processing and coding

Upon completion of the web localisation experiment (see Chapter 5, section 5.2), we collected three different versions of the localised website per participant, which contained three HTML-based web pages each: version T1, which was not validated through any accessibility-oriented QA tool; version T2, which included changes made, if any, after having used the first tool assigned; and version T3, which was the definitive localised website, based on T2 and covering modifications made, if any, according to the second tool's test results. These amounted to a total of 252 HTML files. In order to extract all text alternatives contained therein, we processed all the files following the same methodology described in previous chapters for alt text sampling: by processing the files through Rainbow and a Python script, we got 252 clean MS Excel files with three columns: alt text, image source, and source HTML file (see Chapter 4, section 4.3 for a comprehensive explanation of the alt text extraction process). When problems occurred during the file processing procedure (e.g., extraction of less than 130 text alternatives per file; inaccurate separation of text content from markup), two measures were adopted in parallel with a view to identifying the source of the error and amend it accordingly: (i) web documents were attentively analysed both manually and automatically[3], and (ii) screen recording videos were visualised. The vast majority of errors spotted were minor issues caused by localisers' negligence when manually editing the target HTML files (missing start or end tags, and missing open or closed inverted commas). Since our main object of study were the image text alternatives and not the difficulties encountered by web localisers during the web localisation process, all mark-up errors found were corrected before data analysis and the output of all 28 participants was considered as usable data.

The resulting clean MS Excel files were then grouped and merged into one single spreadsheet per participant. At the end, each file contained the same image metadata as before (image source and source HTML file) and four alt text columns (the source alt text plus one alt text column for each translation version). Once we obtained the 28 spreadsheets, we converted them into tab-separated values (.tsv) files so that they could be processed with the help of the C# Windows Forms application developed by Gerlach (2015), already presented in Chapter 4. The tool

[3] The tool used for automatically checking the robustness of problematic web documents was the W3C Markup Validation Service: https://validator.w3.org/ Last access: September 5th, 2015.

contains, among others, a utility to compare and mark differences between text strings located in two different columns, which we used to run all possible comparisons between the source text and the different translation versions (T1, T2 and T3). The comparison output is conveniently stored in a .tsv file and an HTML file, where differences are highlighted in a different colour and the number of edits is reported. Unlike other advanced measures that yield an automatic score which needs then to be interpreted, such as the Translation Edit Rate (TER) (Snover et al. 2006), Gerlach's tool simply computes the Levenshtein distance at a word level, that is, it calculates the difference between two text strings by counting the total amount of insertions, deletions and substitutions of single words. For instance, within the following alt text pair, two edits were registered: *Main* page - Page *principale* = 2, that is, 1 insertion + 1 substitution.

For the purposes of this research, we coded all alt text pairs with an edit distance of 0 as “non-translated” and those with an edit distance of 1 or higher as “translated” in the comparison ST-T1. This decision was made on the premise that any text alternative that has been subjected to the slightest modification in the target document must have been previously identified −consciously or unknowingly− by the localiser as a translatable element.[4] When comparing translation versions T1-T2 and T2-T3, generated by localisers during the QA task, we used the same binary coding. However, we did not consider alt texts edited as translated, but simply as modified; that is, we did not make the distinction between alt texts edited for the first time in T2 or T3 and those that had already undergone modifications in T1 and were edited again in subsequent translation versions. Through this data coding approach, we have also been able to annotate localisers' reactions regarding the absence of `alt` attributes and the presence of empty `alt` attributes.[5]

6.2.2 Statistical methods

In order to answer **R1**, we examined the number of text alternatives translated by localisers during the web localisation task and the number of text alternatives edited by localisers during the quality assurance task. Thus we primarily worked with numerical data compiled under database DB1 (see section 6.1.1). To this end, we used descriptive statistics (Hardy and Bryman 2009; Lazar et al. 2010) such as measures of central tendency (mean, median), and measures of spread or data distribution (minimum, fractiles, maximum, standard deviation), which also

[4] Although it is true that minor changes to alt texts (e.g. deletion of one letter, addition of an extra space) could have been made by localisers unintentionally while editing other parts of the code, we are assuming that the odds of this happening are relatively low. A thorough study of the localisation process by visualising the videos would be needed to confirm this assumption.

[5] See the brief discussion in section 6.2.3.1 about the possibility of these alt text types being a case of ‘non-translation’.

enabled us to understand the characteristics of our data sample. These summary statistics were then used to compare results across groups, translation versions, tools used and proposed QA scenarios.

In addition, independent-samples t-tests[6] were used to investigate whether the differences observed across groups by translation version and tool used were statistically significant. This parametric statistical test was chosen because, after performing a Shapiro-Wilk normality test, we observed that the data related to DV1 was normally distributed. We have used an alpha level of .05 for all statistical tests (i.e. if $p<0.05$, it was assumed that the difference between the two measured values was significant). Normality and significance tests were performed with the R statistical software.

Finally, it should be mentioned that, as previously announced in this thesis (see Chapter 5, section 5.3.1), our research did not aim at studying the localisation of text alternatives from the traditional perspective of translation as equivalence. Therefore, no further qualitative analysis was performed regarding alt texts' linguistic adequacy or fidelity to the source.

6.2.3 Data analysis and interpretation

In this section, we first present the overall results concerning DV1 without taking into consideration the WA knowledge of the participants (section 6.2.3.1). We then analyse the translation data according to the group who produce it (control, N=14 participants; and treatment, N=14 participants). We first do it by translation version (section 6.2.3.2) and then by QA tool used (section 6.2.3.3). Finally, we compare the results obtained in both groups taking into account the QA scenario followed −where the order in which tools were used differed− (section 6.2.3.4), and we analyse actions taken by localisers with regard to images with no `alt` attribute and images with a null alt or empty `alt` attribute (section 6.2.3.5).

6.2.3.1 Overall results (all participants)

In what follows, we will look at DV1 data per (i) translation version (T1, T2, T3) and (ii) QA tool used (none, Acrolinx, aDesigner)

Translation version

Summary statistics per translation version regarding the number of alt texts edited by localisers are shown in Table 6.2.[7] In general terms, we have observed

[6] T-tests are the most widely adopted statistical procdure for mean comparison (Lazar et al. 2010, 76). In this case, we have used the independent-samples type because the means studied were contributed by different groups of participants.

[7] The following abbreviations are used in Tables 6.2, 6.3, 6.4 and 6.5: **#P** (number of participants); **#Alt** (total number of alt texts edited); **%** (proportion of alt texts edited); **SD** (standard deviation); **Min** (Minimum); **Max** (Maximum); **Q_1** (first quartile); **Q_3** (third quartile).

that the amount of alt texts modified decreases gradually from T1 to T3 (see Figure 6.1, box plot to the left). In the first translation version (T1), which served as the control translation condition since no accessibility-oriented QA tool was used by any localiser, participants translated 1,745 out of 3,640 text alternatives (47.94%), with an average of 62.32 alt texts edited per localiser (sd=40.72). These numbers get slightly smaller in T2, where participants made changes to 1,051 (28.87%) text alternatives ($\bar{x}$=37.53, sd=36.14); and T3, where localisers modified less than one quarter of the alt texts available (N=776, 21.32%, $\bar{x}$=27.71, sd=28.98). As it can be readily seen from the relatively large standard deviations reported,[8] the amount of alt texts edited differs considerably across participants. This variability, which probably derives from the fact that the distribution of localisers between groups is not yet being yet taken into account, is illustrated in Figure 6.1 (box plot to the left).[9] Yet, irrespective of the data variability observed, we consider that the gradual decrease in the number of alt texts edited throughout the three translation versions is understandable, since it is generally expected that the higher the number of QA checks performed, the lower the number of errors left in the localised website and, therefore, of alt texts to be refined.

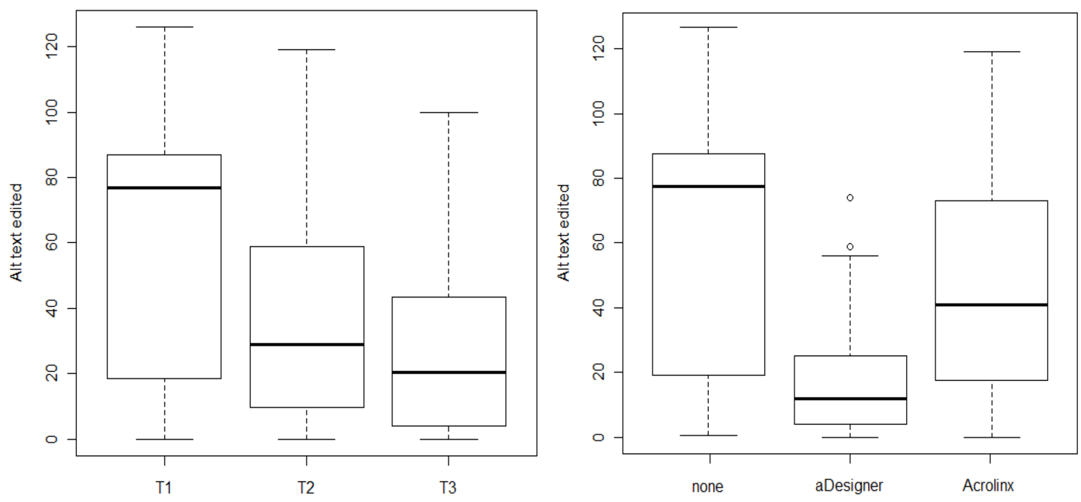


Figure 6.1. Box plots showing the overall dispersion of alt texts edited across translation versions (left) and tools used (right)

[8] The standard deviation of the sample mean tells us how far the typical estimate is away from the actual population mean. This means that, the larger the standard deviation is, the larger the average distance each data point is from the arithmetic mean of the distribution (Hardy and Bryman 2009, 40–43).

[9] The horizontal dark line that splits the box into two parts denotes the median. While the lower and upper whiskers represent the minimum and the maximum values in the data subset respectively, the length of the box is delimited by the 1st (Q_1) and 3rd (Q_3) quartiles (values below which 25% and 75% of the data fall respectively), i.e. the range of the middle half of each data set (Diez et al. 2012, 28).

Table 6.2. Summary statistics per translation version and tool regarding the number of alt texts edited by localisers

#P	Version	#Alt	%	Mean	Median	SD	Min	Max	Q_1	Q_3
28	T1	1745	47.94	62.32	77	40.72	0	126	24.75	87
28	T2	1051	28.87	37.53	29	36.14	0	119	9.75	59
28	T3	776	21.32	27.71	20.5	28.98	0	100	4	40.75
28	aDesigner	512	14.07	18.28	12	19.34	0	74	4	25
28	Acrolinx	1315	36.13	46.96	41	37.36	0	119	18.25	72.5

One may question why more than half of the text alternatives (52.06%) remained unaltered in T1. In this sense, it must be noted that if localisers were not to be held accountable for ensuring the appropriateness of the text alternatives in the target website (i.e. if this was a real-case scenario where the client would not have asked localisers to look into image accessibility issues or expected them to do so), certain text alternatives would not need, a priori, to be edited in order to be linguistically adapted to the target language, as the French translation was formally equivalent to the English source text. In the literature, this phenomenon is often referred to as 'direct translation' (Munday 2009, 182). Within our original experimental website, examples included images with `alt` attribute values containing just proper names (e.g., I04 'Pinterest', I34 'Acrobat Reader'); placeholders (e.g., I41 '#', I112 'alt'); or isolated words whose form is the same in both source and target languages (e.g., I24 'image', I125 'chat logo 3'). Another text string that could have been considered by participants as a 'non-translation' case is the title of the campaign (I07 'Together against poverty'). In addition, images without `alt` attributes (N=10) or with empty `alt` attributes (N=10) might not have been regarded as elements that require action from a translation perspective (see section 6.2.3.5). We have estimated that all these cases may account for approximately 27% of the total amount of text alternatives contained in the website, so there is still a degree of uncertainty regarding the remaining 25% of alt texts that were not modified. We hypothesise that well-constructed full sentences were easier to recognise by translators, rather than isolated nouns or verbs, which might have gone unnoticed among the surrounding HTML code when no CAT tool was used. Similarly, it is likely that not all participants delivered a fully localised website, presumably due to lack of time; hence the relatively low translation rate.

Use of QA tools

When overall results are analysed considering the use of QA tools, differences with regard to the number of alt texts edited become more noticeable (see again Table 6.2 and Figure 6.1, box plot to the right). Whereas, on average, participants made changes to 18.28 text alternatives after using aDesigner (sd=19.34, N=512, 14.07%), Acrolinx triggered modifications in 36.13% of alt texts (N=1315, x̄=46.96, sd=37.36). As depicted in the corresponding box plot, the maximum value observed for Acrolinx (Max=119) doubles the maximum value observed for aDesigner (Max=56) if we do not consider the outliers (74 and 59, labelled with two dots), which correspond to the number of alt texts edited by participants P40 and P42 respectively. A breakdown of the results per participant is presented in Appendix K.1 (see Table K.1).

This first overall comparison of the data gathered for both tools reveals that, irrespective of the order in which they were used, Acrolinx proved more efficient when it comes to drawing participants' attention to the existence of text alternatives. Yet, while we are certain that Acrolinx triggered a considerably higher number of alt text modifications, further analysis is needed to understand whether those changes led, or not, to an improvement of text alternatives' quality. With respect to aDesigner's performance in terms of boosting alt texts visibility, the tool seems to be less powerful. This might be explained by a low level of specificity as regards the number of different possible issues related to image accessibility that can be detected by the tool (Brajnik 2004), or the complexity of the explanations provided in the test reports concerning the problems flagged. These assumptions will be clarified when analysing the data related to DV2 in section 6.3.

6.2.3.2 Results per translation version by group

After obtaining a first glimpse of the results as a whole, we performed a second analysis to take into account the independent variable *IV1. Knowledge of web accessibility (WA)* when observing the differences between the data collected for DV1 across translation versions (T1, T2, T3). In other words, we considered the two groups into which participants were divided (control versus treatment) for the analysis, with a view to examining the effect of the independent variable **IV1**. Table 6.3 provides complete summary statistics for the data under analysis. In Figure 6.2, results per group concerning the number of alt texts edited across translation versions are shown side-by-side.

Table 6.3. Summary statistics per group regarding the number of alt texts edited according to the translation version

Group	#P	Version	#Alt	%	Mean	Median	SD	Min	Max	Q_1	Q_3
Control	14	T1	517	28.41	36.93	18.5	41.99	0	111	0	79
Treatment	14	T1	1228	67.47	87.71	85	17.11	66	126	77	90
Control	14	T2	706	38.79	50.43	45.5	43.03	0	119	10.5	87.5
Treatment	14	T2	345	18.96	24.64	21	22.44	0	74	10.5	35.75
Control	14	T3	446	24.51	31.86	25	25.99	0	91	13.25	46.25
Treatment	14	T3	330	18.13	23.57	10	32.14	0	100	0	28.5

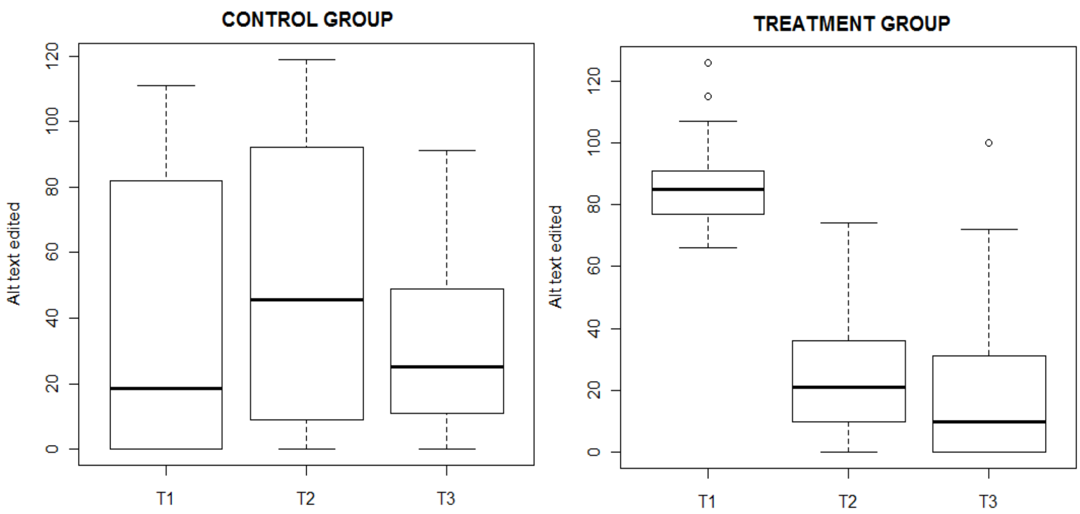


Figure 6.2. Box plots showing the overall dispersion of alt texts edited across translation versions within the control group (left) and the treatment group (right)

When comparing both box plots, a noteworthy contrast can be particularly observed in translation versions T1 and T2. While participants in the control group translated, on average, 36.93 text alternatives in T1 (sd=41.99, N=517, 28.41%), localisers from the treatment group translated, on average, 87.71 alt texts out of 130 (sd=17.11, N=1,228, 67.47%). These results suggest that the participants who had attended the webinar on accessibility-oriented HTML best practices for localisers were more confident about the need —and importance— of translating text alternatives than those who had not. In fact, it seems that localisers with a basic background on WA were able to identify them as such much more easily than participants from the control group. These assumptions are supported by the differences observed between groups in terms of data variability. On one hand, the standard deviation reported for the treatment group is appreciably lower than the control group's. On the other hand, as depicted in Figure 6.2, the length of the boxes representing the number of alt texts edited in T1 differs substantially. Whereas in the control group six participants did not translate any text alternative (Min=0, Q_1=0), the minimum value observed within the treatment group was 66 (Q_1=77). This difference between groups with regard to T1 is confirmed by the independent-samples t-test, which supports that members of the treatment group translated a significantly higher number of alt texts than the control group, $t(17)=4.19$, $p<0.001$.

Variability within the data related to the second translation version (T2) follows a similar pattern. However, the highest number of edits is registered this time by the control group, whose members made changes to 706 text alternatives (38.79%, $\bar{x}$=50.43, sd=43.03), as opposed to the treatment group, that performed modifications on just half of them (N=345, 18.96%, $\bar{x}$=24.64, sd=22.44). Data gathered indicate, therefore, that the introduction of an accessibility-oriented QA tool triggered different reactions in the two groups. In the case of the control group, it seems that it served localisers to identify image text alternatives as translatable elements and propose an equivalent in the target language. In contrast, one would hypothesise that, for the treatment group, the tool proved beneficial for quality refinement. In other words, the distribution of the data suggests that, while the control group showed a high degree of hesitation regarding the translation of image text alternatives during both the web localisation task and the first part of the quality assurance task, the treatment group seems to have followed a more classical localisation procedure, in which the use of a QA tool simply helped to fine-tune and consolidate the work already done. The independent-samples t-test reveals, however, that the differences across groups with regard to the number of alt texts edited is not statistically significant: $t(19)=1.99$, $p=0.06$.

As far as translation version T3 is concerned, results in terms of alt texts edited are closer between groups (control: N=446, 24.51%, $\bar{x}=31.86$, sd=25.99; treatment: N=330, 18.13%, $\bar{x}=23.57$, sd=32.14), which explains the outcome of the t-test: $t(25)=0.75$, $p=0.46$. The degree of variability observed within each data subset is similar as well. These data might indicate that the effect of having basic WA knowledge of WA on alt text editing rates is more blurred when two QA tools are used. However, at this stage of the data analysis, no stronger inferences can yet be drawn yet. In the following sections, we further examine our first research question (**R1**) by looking at group results according to the tool used (6.2.3.3) and the QA scenario followed (6.2.3.4).

6.2.3.3 Results per accessibility-oriented QA tool by group

In this section, we provide insight into the general interaction effect of having WA knowledge (**IV1**) and using accessibility-oriented QA tools (**IV2**) —regardless of the order followed— on the identification of alt texts as translatable elements by localisers. Firstly, we offer a general overview of the findings by combining results from T2 and T3 (use of both tools). Secondly, we present the results by tool.

In Figures 6.3 and 6.4, the bar charts (to the left) depict the total amount of alt texts translated in T1 as opposed to the sum of text alternatives modified in T2 and T3 per group. Data suggest that the combined use of aDesigner and Acrolinx led localisers from the control group to edit almost twice the number of image text alternatives (N=1,152) when compared to the treatment group (N=675). As illustrated in both figures, these numbers contrast with the amount of alt texts translated by each group in T1 (N=517 and N=1,228 respectively). These observations confirm that basic awareness on web accessibility issues indeed helped localisers to identify and translate more text alternatives without any QA automated support. In turn, tools appear to have been a key instrument for the control group to find and edit a higher number of alt texts. Interestingly, if we look at the proportion of text alternatives modified when using each one of the tools (see Figures 6.3 and 6.4, pie charts to the right), we see that Acrolinx triggered changes in more than 70% of the total number of alt texts edited in both groups (control and treatment).

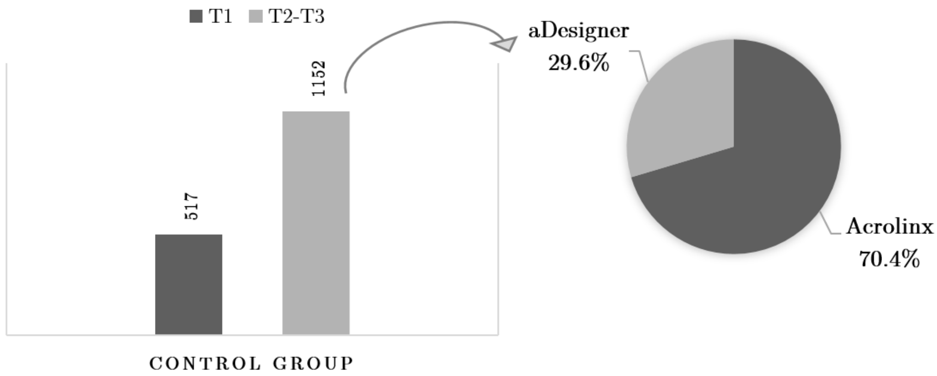


Figure 6.3. Summary of alt texts edited by the control group

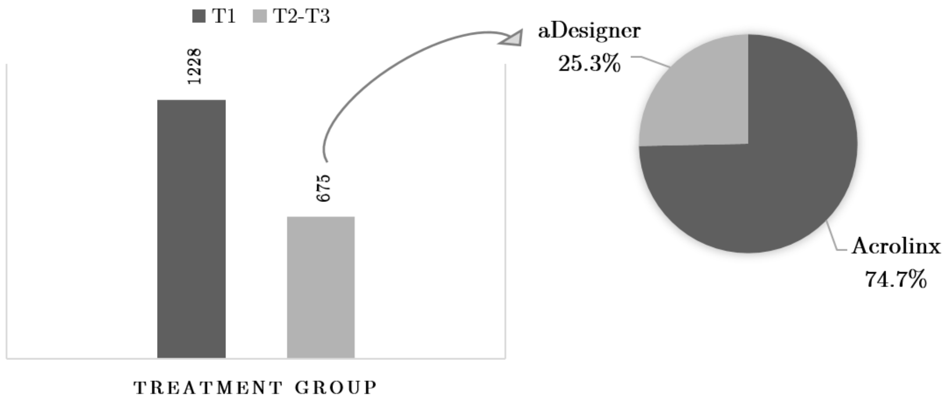


Figure 6.4. Summary of alt texts edited by the treatment group

In Table 6.4, we provide a more detailed breakdown of the data under analysis, which is illustrated as well in the box plots shown in Figure 6.5. As can be seen, aDesigner appears to have been slightly more useful for participants from the control group. It triggered changes in 341 alt texts ($\bar{x}$=24.36, sd=23.57), while within the treatment group, it led to modifications in 171 text alternatives ($\bar{x}$=12.21, sd=11.92). We believe that this might be due to the fact that aDesigner detects the presence of `<img>` elements with no `alt` attribute and, by default, presents one-character or one-word alt texts as potential errors. Although the implications of both cases for image accessibility were presumably known by localisers from the treatment group, they might not have been so obvious for participants in the control group, who most probably ignored these issues in T1 and only learned about them when checking aDesigner's test reports. No solid conclusions can be drawn, however, until examining the data taking into account the order in which the tool was used (see section 6.2.3.4).

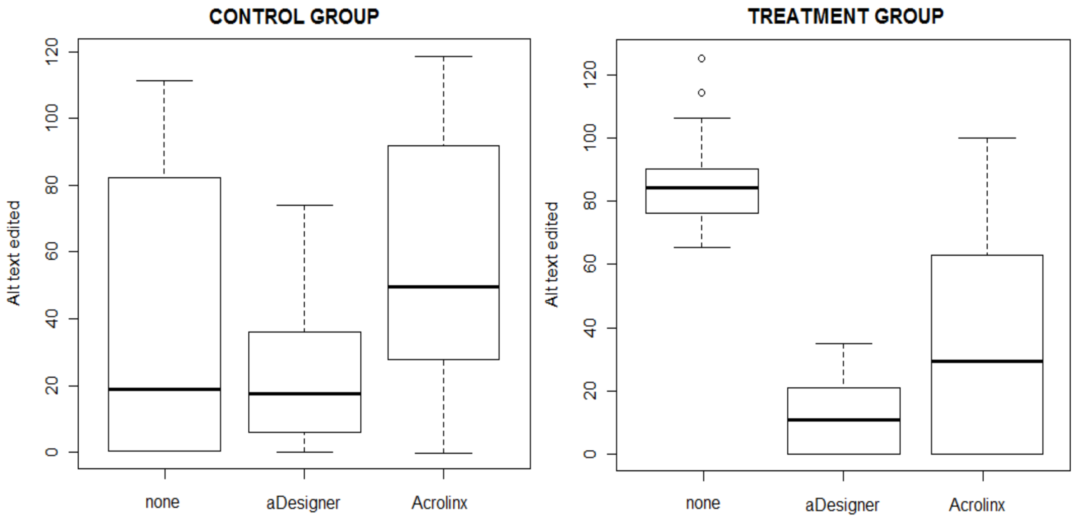


Figure 6.5. Box plots showing the overall dispersion of alt texts edited depending on the tool used within the control group (left) and the treatment group (right)

Table 6.4. Summary statistics per group regarding the number of alt texts edited according to the tool used

Group	#P	Version	#Alt	%	Mean	Median	SD	Min	Max	Q_1	Q_3
Control	14	none	517	28.41	36.93	18.5	41.99	0	111	0	79
Treatment	14	none	1228	67.47	87.71	85	17.11	66	126	77	90
Control	14	aDesigner	341	18.74	24.36	17.5	23.57	0	74	6.75	20.5
Treatment	14	aDesigner	171	9.40	12.21	11	11.92	0	35	1	20.5
Control	14	Acrolinx	811	44.56	57.93	50	39.37	0	119	30.5	91.75
Treatment	14	Acrolinx	504	27.69	36	29.5	33.02	0	100	4	62

Although with this always in mind, we could state that Acrolinx's test results and documentation seem to have been equally enlightening to both the control and the treatment group during the QA task. Data suggest that, thanks to the overall use of this tool, the former made changes to a total of 811 text alternatives ($\bar{x}$=57.93, sd=39.37) and the latter edited 504 alt texts ($\bar{x}$=36, sd=33.02). There are two interesting observations that emerge from the comparison of these data. On one hand, one would expect that the treatment group would not have needed to make such a large amount of changes after T1. It seems, however, that localisers from this group at first performed a word-by-word translation of the alt texts they found, and then realised during the QA task —presumably thanks to the use of Acrolinx— that half of these were not appropriate enough. On the other hand, we could infer that Acrolinx's suggestions and explanations of the errors flagged were comprehensible enough for the control group, who despite not having any background on web accessibility, tried to follow Acrolinx's guidance and managed to edit an amount of text alternatives similar to the number of alt texts modified in T1. The data presented in the next section will help us to understand whether the order in which tools were used had an impact on the alt text edit rates of both groups.

6.2.3.4 Results per QA scenario by group

As explained in Chapter 5 (section 5.2.1), we initially decided to counterbalance the order in which tools were used during the QA task simply to avoid any potential bias due the practice effect. However, when looking closely at how data was distributed across the different translation versions (T1, T2, T3), we realised that results regarding the number of alt texts edited could vary considerably depending on the QA scenario proposed to participants. Therefore, we decided to analyse the data associated with **DV1** taking into account this secondary independent variable. In contrast with what we have done in previous sections of this chapter, overall group results are complemented here with results per participant. The combined discussion of these data (general and individual results) will allow us to better observe the differences between Scenario A, where aDesigner was used to produce T2 and Acrolinx served to fine-tune translation version T3; and Scenario B, where the aforementioned tools were used in reverse order. Table 6.5 provides summary statistics per tool by group, taking into account the order of use (i.e. indicating whether they were used to produce version T2 or T3).

Table 6.5. Summary statistics per group regarding the number of alt texts edited taking into account the order in which tools were used

Group	#P	Version	Tool	#Alt	%	Mean	Median	SD	Min	Max	Q_1	Q_3
Control	7	T2	aDesigner	193	21.21	27.57	15	29.54	0	74	4.5	47.5
Control	7	T2	Acrolinx	513	56.37	73.28	92	43.81	0	119	45.5	105.5
Treatment	7	T2	aDesigner	92	10.11	13.14	12	12.47	0	35	5	17.5
Treatment	7	T2	Acrolinx	253	27.80	36.14	36	25.04	0	74	21	50.5
Control	7	T3	aDesigner	148	16.26	21.14	20	17.51	5	56	8.5	25
Control	7	T3	Acrolinx	298	32.75	42.57	38	29.83	0	91	26.5	58
Treatment	7	T3	aDesigner	79	8.68	11.28	4	12.27	0	31	2	20
Treatment	7	T3	Acrolinx	251	27.58	35.86	16	41.65	0	100	0	67.5

Control group

Data concerning the performance of the control group is depicted in Figure 6.6. As can be clearly seen in the box plots, Acrolinx triggered changes in a higher number of text alternatives than aDesigner, irrespective of the order in which it was used by localisers who have no WA knowledge. When introduced after aDesigner, Acrolinx led to modifications in 60.69% (N=298) of the alt texts edited (N=491) during the QA task. Likewise, when proposed in the first place, it motivated changes in 77.61% (N=513) of the text alternatives modified in the last two translation versions (T2 and T3, N=661) (see Table 6.6, where the highest value observed after comparing results from translation versions T2 and T3 per participant is shown in italics).

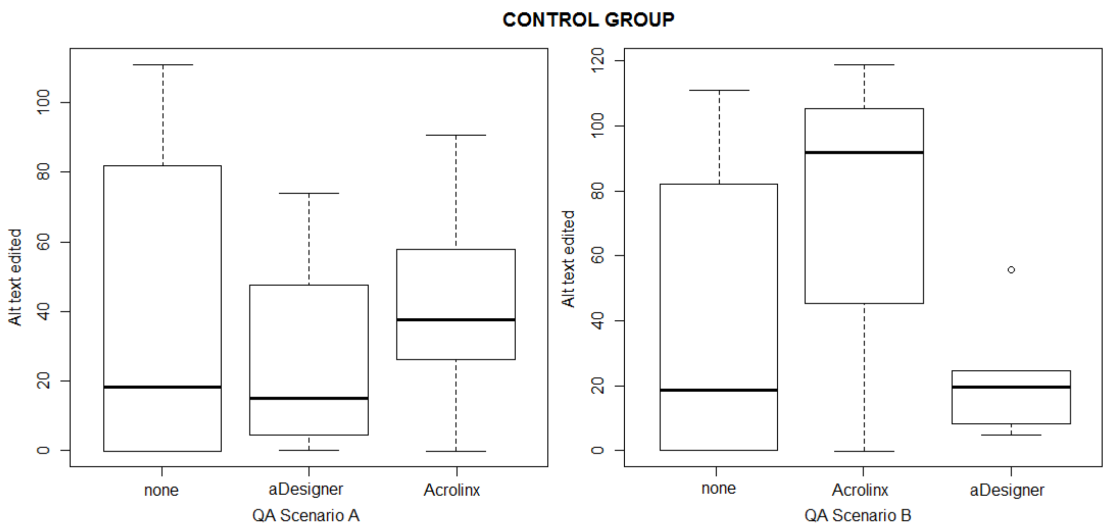


Figure 6.6. Box plots showing the overall dispersion of alt texts edited by the control group per QA scenario

With a view to observing whether the differences in the number of alt text edits made using different tools within the same translation version were statistically significant, we run two independent-samples t-tests.[10] The comparison of the means between the results of the localisers who used Acrolinx in T2 and those from the localisers who used aDesigner in T2 revealed that, in fact, there is a statistically significant difference across tools: $t(11)=2.29$, $p<0.05$. This means that, at T2, Acrolinx led to a significantly higher number of alt text edits than aDesigner. When the same t-test was applied for the comparison T3-Acrolinx – T3-aDesigner, we

[10] It should be noted, however, that in this case, the outcome of these tests should be only regarded as orientative, as the different translation versions are dependent on one another, i.e. results from T2 are influenced by what was done in T1, and those from T3 are directly related to what was achieved in T2.

found, however, that the statistically significant differences disappear: t(10)=1.64, p=0.13.

Table 6.6. Alt texts edited per participant within the control group (no WA knowledge), by QA scenario

QA Scenario A: *aDesigner-Acrolinx*					QA Scenario B: *Acrolinx-aDesigner*				
Participant	**T1**	**T2**	**T3**	**Total tools**	**Participant**	**T1**	**T2**	**T3**	**Total tools**
P18	82	0	0	**0**	P04	111	*51*	25	**76**
P25	6	15	*28*	**43**	P16	70	*92*	25	**117**
P34	87	0	*67*	**67**	P19	0	*40*	20	**60**
P40	0	*74*	49	**123**	P23	88	*102*	56	**158**
P41	0	*36*	25	**61**	P31	0	*109*	6	**115**
P42	42	59	*91*	**150**	P36	0	*119*	11	**130**
P52	31	9	*38*	**47**	P53	0	0	*5*	**5**
Total	**248**	**193**	***298***	**491**	**Total**	**269**	***513***	**148**	**661**

The data presented so far seems to suggest that Acrolinx was more effective when used before aDesigner for the control group. The contrast between results from participants P40 and P41 (Scenario A) and those from participants P31 and P36 (Scenario B) can serve to illustrate this statement (see Table 6.6). Prior to the QA task, none had translated any image text alternative. When P40 and P41 were presented with aDesigner in the first place, they managed to translate 74 and 36 text alternatives respectively. Conversely, P31 and P36 translated 109 and 119 out of 130 alt texts respectively when using Acrolinx as the first QA tool.

Within the control group, two other participants (P19 and P53, QA Scenario B) noticed the existence of text alternatives for the first time during the QA task. Nevertheless, Acrolinx does not seem to have been as useful for them. While P53 did inform the researcher that the complexity of both tools (Acrolinx and aDesigner) had prevented her from editing alt texts without fear of making a mistake, an in-depth analysis of the localisation process data collected should be carried out to better explain why results from P19 appear to be unusually distant from the rest of the data. Lower edit rates in the case of participants P25, P41 and P52 (QA Scenario A) might be due to the fact that all three reported to have worked, due to time limitations, only on one of three pages of the website.

Treatment group

Although not as compelling as in the case of the control group, data from the treatment group points equally at Acrolinx as the tool that triggered the highest amount of alt text edits, regardless of the order in which it was used (see Figure 6.7). While aDesigner led to modifications in 26.82% (N=92) of all the alt texts edited (N=343) during the QA task in Scenario A, localisers following

Scenario B applied changes to 76.20% (N=253) of all text alternatives edited in T2 and T3 (N=332), presumably following Acrolinx's feedback. Similarly, 73.18% (N=251) of the alt texts edited during the QA task when instructions showed Scenario A were triggered by Acrolinx (see Table 6.7, where the highest values are shown in italics).

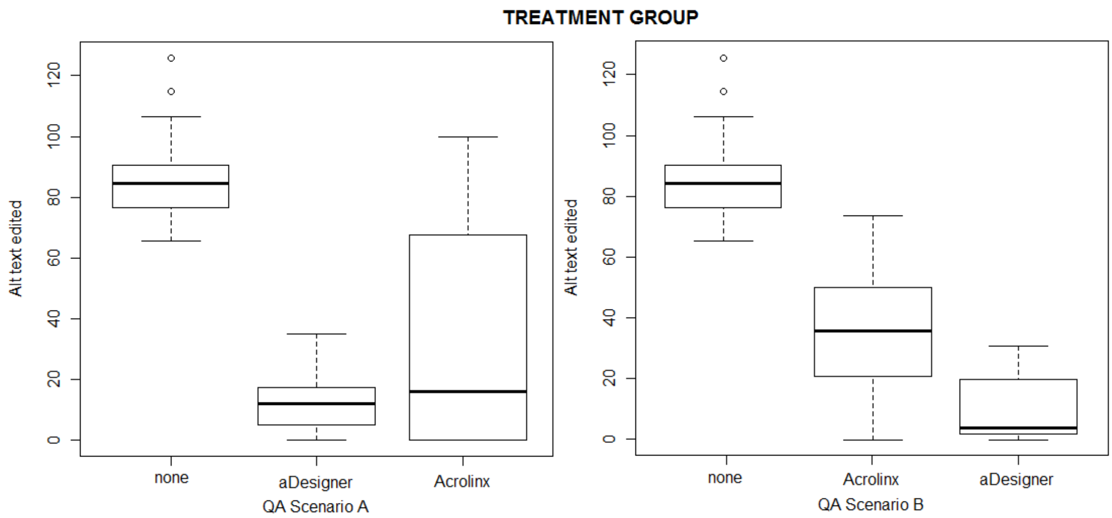


Figure 6.7. Box plots showing the overall dispersion of alt texts edited by the treatment group per QA scenario

Table 6.7. Alt texts edited per participant within the treatment group (with WA knowledge) by QA scenario

QA Scenario A: *aDesigner-Acrolinx*					QA Scenario B: *Acrolinx-aDesigner*				
Participant	**T1**	**T2**	**T3**	**Total tools**	**Participant**	**T1**	**T2**	**T3**	**Total tools**
P01	84	*23*	0	**23**	P06	74	*74*	0	**74**
P33	91	0	0	**0**	P21	107	*19*	4	**23**
P44	72	10	*100*	**110**	P22	66	*36*	19	**55**
P46	80	12	*16*	**28**	P26	126	*23*	4	**27**
P49	86	0	0	**0**	P51	77	*59*	21	**80**
P55	87	12	*72*	**84**	P56	77	0	0	**0**
P58	86	35	*63*	**98**	P59	115	*42*	31	**73**
Total	**586**	**92**	***251***	**343**	**Total**	**642**	***253***	**79**	**332**

Following the same analysis procedure as with data from the control group, we used independent-sample t-tests to observe whether the differences across the number of alt texts edits triggered by each tool within the same translation version were significant. The mean comparison at T2 revealed that, although Acrolinx led to a considerable higher number of edits than aDesigner, this difference was not statistically significant: t(9)=2.18, p=0.058. A similar non-significant outcome was found when tool results were compared at T3: $t(7)=1.50$, $p=0.18$. This indicates that, in the case of the treatment group, the order in which tools were used did not have strong implications in terms of alt text editing. For instance, in contrast to the control group, the amount of text alternatives edited by the treatment group when Acrolinx was used after aDesigner (N=251) is almost the same as the amount observed when Acrolinx was used in the first place (N=253). We notice that the same trend occurred with aDesigner (see Table 6.7).

It is worth highlighting that, during the QA assurance task, four participants did not perform any further changes (P18, from the control group, see Table 6.6; and P33, P49 and P56 from the treatment group, see Table 6.7). Once they completed the experimental study, P18, P49 and P56 acknowledged that they did not amend their alt text translations as per the tools' suggestions because they considered that this job should be carried out by a web accessibility expert. As localisers, they believed that they should maintain the same level of accessibility achieved in the original web product, even if it was low. In turn, participant P33 said that the topic was new to her and more training would be needed to correctly implement the changes suggested in the tools' check reports. The visualisation of the screen recordings and the verification of the check reports from these five participants demonstrated that they had used both tools during the QA task. Therefore, although translation versions T2 and T3 were exactly the same as T1, we still included them as valid data during the analysis.

All in all, the findings reported in this section reveal that, while the distribution of participants between different QA scenarios did not have a significant impact on the overall results of the treatment group, the order in which tools were used proved to have a strong effect on those from the control group, who edited a significantly higher number of alt texts at T2 when Acrolinx was used instead of aDesigner. We hypothesise that, being less technical and more language-oriented, Acrolinx's feedback concerning image accessibility was easier to understand than aDesigner's input for the control group. In this sense, it is likely that after having spent some time trying to follow aDesigner's suggestions, the localisers from the control group who followed QA Scenario A felt tired, confused or frustrated, thus losing focus on the task when we proposed the use of Acrolinx. Interestingly, when the independent variable **IV1**. *WA knowledge* is not taken into account (i.e., when no distinction is made between groups) the observation that emerges from the data is the same as in the case of the control group: when two QA tools are proposed, Acrolinx appears to be more effective when it comes to alt text editing if used in the first place (T2, N=766 out of 993, 77.14%) rather than after aDesigner (T3, N=549 out of 834, 65.83%). Results per participant, regardless of the group they belonged to, can be found in Table K.2 of Appendix K.1.

6.2.3.5 Images with empty `alt` attributes and no `alt` attributes

In previous sections, we have provided insight into the translation and adaptation of text alternatives for images during the localisation process. So far, when reporting our findings, we have treated images on an equal basis, that is, no distinction has been made between images with text alternatives of one or more words and images with empty (or null) `alt` attributes. Similarly, each new insertion of an `alt` attribute within an `<img>` element for which this attribute was missing in the source website has been counted as an edited text alternative.

In this section, focus is shifted to empty `alt` attributes and no `alt` attributes as units of investigation, with a view to examining how they were treated by localisers. As we have discussed in Chapter 3 (section 3.3.1.3), empty `alt` attributes are accepted and recommended for images with a decorative or aesthetic purpose. Nonetheless, the absence of an `alt` attribute itself is considered as a violation of one of the most well-known best practices for web accessibility (Caldwell et al. 2008). Unlike images with a full text equivalent, these two particular cases might require localisers to go further than a mere textual translation. As pointed out by Jiménez Crespo (2008, 430), addressing image accessibility during the web localisation process may entail web technical adjustments, such as creating new text alternatives or adapting existing ones.

At this point, it should be remembered that from the 130 images included in the original website to be localised during the experiment, ten images had an empty `alt` attribute and ten images did not have any `alt` attribute. The results that follow are based on two main assumptions. Firstly, we understand that a decrease in the number of images with no `alt` attributes throughout the three different translation versions (T1, T2, T3) is equivalent to an increase in the number of localised text alternatives. Secondly, we consider that the higher the number of images with new empty `alt` attributes, the higher the number of localised alt texts. In both cases, we can infer that localisers must have edited the HTML source code of the localised website to add the `alt` attribute or to delete the text value it contained.

Empty `alt` attributes

Tables 6.8 and 6.9 present the results regarding the presence of empty `alt` attributes within the localisation output gathered from the control group and the treatment group respectively, according to the QA scenario followed. The first difference observed within the control group (see Table 6.8) is related to translation version T1, in which only one participant (P04) added a textual value to four originally empty image `alt` attributes. Regarding T2 and T3, localisers following QA Scenario B left a higher number of images with empty `alt` attributes (T2, N=132; T3, N=234) than those using the tools in the reverse order (T2, N=59; T3, N=113). However, in contrast to the overall trend observed in the previous section concerning the use of tools (Acrolinx seemed to trigger changes in a higher number of text alternatives), the increase in the amount of empty `alt` attributes appears to be related to the combined use of two tools instead of one, rather than to the use of a specific tool.

Table 6.8. Total number of images with an empty `alt` attribute per participant within the control group (no WA knowledge), by QA scenario

Scenario A: *aDesigner-Acrolinx*				Scenario B: *Acrolinx-aDesigner*			
Participant	**T1**	**T2**	**T3**	**Participant**	**T1**	**T2**	**T3**
P18	10	10	10	P04	6	5	24
P25	10	12	24	P16	10	13	19
P34	10	10	18	P19	10	10	29
P40	10	0	0	P23	10	2	55
P41	10	11	12	P31	10	63	69
P42	10	6	39	P36	10	29	25
P52	10	10	10	P53	10	10	13
Total	**70**	**59**	**113**	**Total**	**66**	**132**	**234**

Table 6.9. Total number of images with an empty `alt` attribute per participant within the treatment group (with WA knowledge) by QA scenario

Scenario A: *aDesigner-Acrolinx*				Scenario B: *Acrolinx-aDesigner*			
Participant	**T1**	**T2**	**T3**	**Participant**	**T1**	**T2**	**T3**
P01	10	4	4	P06	10	29	29
P33	10	10	10	P21	10	11	13
P44	10	16	38	P22	10	10	17
P46	10	5	3	P26	39	39	39
P49	10	10	10	P51	10	20	14
P55	10	20	45	P56	10	10	10
P58	10	8	12	P59	45	45	24
Total	**70**	**73**	**122**	**Total**	**134**	**164**	**146**

At a participant level, it is worth highlighting the case of P42 and P23, who reduced the number of empty `alt` attributes in T2 and increased it drastically in T3. Still, since each one followed a different QA scenario, these changes cannot be attributed to the use of any tool in particular. Interestingly, P53, who only edited five text alternatives in the entire QA task, changed three alt texts to transform them into null `alt` attributes. Data from P40 shows unexpected results: while in T1 the number of empty `alt` attributes remains unaltered, it goes down to zero in T2 and does not change in T3.

As far as the treatment group is concerned, we notice that there is a clear increase in the total number of empty `alt` attributes when we compare overall results from

T3 (N=122) against those from T1 (N=70) in QA Scenario A (see Table 6.9). Nevertheless, the difference between these translation versions in QA Scenario B is considerably lower (T3, N=146; T1, N=134). This contrasts with the results that have been presented above in the case of the control group. We hypothesise that this might be due to the results of two outliers (P26 and P59), who included 39 and 45 null `alt` attributes respectively in T1 (see Table 6.9). An additional difference arises when comparing the results from both groups by tool used. Within the treatment group, Acrolinx seems to have triggered a higher amount of empty `alt` attributes, irrespective of the order in which it was used. In QA Scenario B, aDesigner's suggestions appear to have been particularly meaningful only for P59, who reduced the number of null `alt` attributes from T2 to T3 by almost half.

Overall, data collected suggest that localisers were keen to include more empty `alt` attributes in the localised website than the ones that already existed in the source website. Although it emerges from a different research approach (experimental versus corpus-based), this finding is in agreement with the observations made by Tercedor Sánchez and Jiménez Crespo (2007, 138) when studying a comparable Web corpus. The authors found that the localised sub-corpus contained 386,299 empty `alt` attributes, as opposed to the original sub-corpus, that had 103,229. In that localisation study, the difference observed might have been motivated by differences in terms of total amount of images per corpus. However, in our study, the number of images did not change across translation versions.

We think that localisers might have considered that a higher number of images on the website should not be announced to blind users because their content was not of added-value to understand the surrounding information, hence the unexpected increment of null `alt` attributes. It is likely that this strategy was also considered safer than introducing a non-appropriate full alt text, or simply easier than coming up with a high quality text alternative, which is, in fact, the most common reason for using an empty `alt` attribute among web developers (Tang 2012, 53–54). The increase in the amount of null `alt` attributes might also be caused by a good performance of both tools when explaining the contexts in which images should be labelled with an empty `alt` attribute. Finally, although one would expect that localisers from the treatment group would feel more confident in reconsidering the value (decorative, functional, referential) of the website's images already in T1, no conclusive differences have been noticed between groups. In order to draw further inferences as regards the effectiveness of each tool proposed, the level of appropriateness achieved in the edited text alternatives will need to be taken into consideration (see section 6.3).

Absence of `alt` attributes

Results concerning the presence of images without an `alt` attribute are shown in Tables 6.10 (control group) and 6.11 (treatment group). Just as in the case of empty `alt` attributes, data is presented according to the scenario followed by participants during the QA task.

The most striking observation within the control group emerges from the comparison of the results per QA scenario. Whereas in Scenario A, the number of images missing an `alt` attribute increases from T1 (N=70) to T3 (N=77), we can see that localisers following Scenario B managed to surprisingly reduce the amount of non-accessible images from 75 (T1) to 17 (T3). Interestingly, within this scenario, Acrolinx seems to have guided three participants (P04, P19 and P23) towards the elimination of the `alt` attribute of three, seven, and 41 images respectively in T2. Conversely, aDesigner led to much better results, particularly helping the aforementioned participants to reduce the amount of images without an `alt` attribute in T3. From all the data gathered in Scenario A, results from participant P40 are again counterintuitive. We hypothesise that this participant might have misunderstood how a null `alt` attribute is implemented, so she deleted the attribute itself whenever she thought the image did not contain relevant information instead of just introducing an empty value after it. Finally, it is worth highlighting that only two other localisers (P25 and P42) achieved a lower proportion of images without the said attribute after completing the QA task according to QA scenario A.

Table 6.10. Total number of images without an `alt` attribute per participant within the control group (no WA knowledge), by QA scenario

Scenario A: *aDesigner-Acrolinx*				Scenario B: *Acrolinx-aDesigner*			
Participant	**T1**	**T2**	**T3**	**Participant**	**T1**	**T2**	**T3**
P18	10	10	10	P04	15	18	3
P25	10	1	1	P16	10	9	2
P34	10	10	10	P19	10	17	1
P40	10	27	27	P23	10	52	0
P41	10	10	10	P31	10	8	2
P42	10	10	9	P36	10	4	0
P52	10	10	10	P53	10	10	9
Total	**70**	**78**	**77**	**Total**	**75**	**118**	**17**

Table 6.11. Total number of images without an alt attribute per participant within the treatment group (with WA knowledge) by QA scenario

Scenario A: *aDesigner-Acrolinx*				Scenario B: *Acrolinx-aDesigner*			
Participant	**T1**	**T2**	**T3**	**Participant**	**T1**	**T2**	**T3**
P01	10	9	9	P06	10	10	10
P33	10	10	10	P21	10	8	6
P44	10	1	6	P22	10	10	1
P46	10	9	9	P26	1	1	0
P49	10	10	10	P51	10	7	2
P55	10	5	5	P56	10	10	10
P58	10	2	2	P59	6	6	1
Total	**70**	**46**	**51**	**Total**	**57**	**52**	**30**

Contrary to expectations, participants from the treatment group did not systematically identify images with no alt attribute as problematic during the web localisation task (see columns 'T1', Table 6.11). Additionally, while there has been an overall decrease in the amount of images without an alt attribute from T1 to T3 in both scenarios, only six out of the 14 participants managed to reduce it at least by a half, i.e. to get five or less images with no alt attribute in T3. Results from Scenario B appear to be slightly more satisfactory when compared to those from Scenario A, but not as compelling as in the case of the control group. As far as the use of tools is concerned, Acrolinx has proved counterproductive only in the case of P44 (Scenario A), who in T2, presumably thanks to aDesigner, added an alt attribute to nine out of the ten images that did not have one in T1, but then deleted five in T3, leaving a total of six images without said attribute. For the other 13 participants, the joint use of both tools seemed beneficial.

From a general perspective, the localised versions of the website contained fewer images with no alt attribute than the source website. Once again, this finding mirrors those of the single previous web localisation study that has examined the presence or absence of image text alternatives (Tercedor Sánchez and Jiménez Crespo 2007; Jiménez Crespo 2008). Within the comparable web corpus, the alt attribute was found in 89.21% of all the pages compiled under the localised sub-corpus, as opposed to the original sub-corpus, where the attribute appeared in only 61.71% of the pages (Jiménez Crespo 2008, 426).

In terms of tools, aDesigner appears to have triggered the highest number of insertions of alt attributes in both groups, especially if used after Acrolinx (QA Scenario B). This is understandable, given that, unlike the latter, the former does identify <img> elements with a missing text alternative.[11] Surprisingly enough,

[11] See Chapter 4 (section 4.5) for an explanation of Acrolinx's capabilities regarding the detection of missing and empty alt attributes.

Acrolinx seems to have motivated some minor improvements in this respect, but more emphasis should be placed on the analysis of localisation process data to verify if these few changes were truly due to the alt text guidance provided by Acrolinx— for instance, in the rule help (see Chapter 4, section 4.5). In addition, we had anticipated that the treatment group would dare to make more changes with regard to HTML editing. Nevertheless, it seems that the use of tools has allowed participants from the control group to bridge the knowledge gap and obtain similar results concerning the reduction of images with `alt` attributes.

Conclusions

In Section 6.2, we addressed research question **R1**: *Are image text alternatives considered as translatable elements by localisers during the web localisation process?* The evidence presented thus far has allowed us to observe and discuss the influence of different independent (or explanatory) variables, including the two primary ones contained in research questions **R2** and **R3**, on our first dependent (or response) variable: *DV1. Translation of alt texts.*

Upon data analysis and interpretation, we can only partially support our main hypothesis, which stated that *image text alternatives are considered by localisers as translatable elements during the web localisation process.* As shown earlier in this chapter, not every participant from our experimental study translated the alt texts during the web localisation task. In fact, there were six localisers who did not translate any text alternative at all. Our findings seem to be in agreement with the observations made by Fernández Costales (2010) when studying the translatedness of image text equivalents in university websites, who found that there is still a low level of awareness about the need to translate these elements for accessibility purposes.

When we looked at the data from localisers who actually edited alt texts, results demonstrated a considerable degree of variability as regards the awareness about the existence of (and the need to translate) these elements, especially across groups. In fact, localisers from the treatment group translated, on average, twice the number of alt texts than participants from the control group, and it was proved that the differences between groups at T1, when no automated support was provided, were statistically significant. We can therefore accept our first sub-hypothesis: **H1.1.** *Having basic knowledge of web accessibility helps localisers to identify text alternatives as translatable content during the web localisation process.* We believe that this finding denotes an important flaw with regard to current web localisation practices. As we noted in Chapter 3 (section 3.5.1.1), alt texts are repeatedly referred to as translatable elements of an HTML document in the localisation-related literature (Mata Pastor 2009b, 552; Roturier 2015, 88). Yet, they are still not easily identified as such. Given their technical profile, localisers are expected to recognise all translatable text within a web document, and the added-value of having some WA background should not be decisive for such an endeavour. In future work, when examining the data collected on the localisation process, it would be interesting to see whether the use of CAT tools had an impact on the visibility of image text alternatives. As explained in Chapter 3 (section 3.5.2.2), text extraction rules need to be carefully set up before starting the localisation of mark-up documents, such as HTML. In the next chapter (7), we will consider whether CAT tools influenced the achievement of more appropriate text alternatives, which will enable us to draw clearer conclusions about their effectiveness regarding the rendering of alt texts to localisers.

As far as the use of accessibility-oriented QA tools is concerned, findings reported indicate that when participants used Acrolinx, they managed to edit a higher number of text alternatives than with aDesigner alone. Similarly, data have shown that, for localisers without basic knowledge of web accessibility, the combined use of both tools triggered modifications in twice the number of text alternatives if compared to the total amount of alt texts edited without any QA tool support. Additionally, the use of tools proved particularly beneficial for the six localisers who had not identified text alternatives as translatable elements during the web localisation task. We are thus confident to confirm our second sub-hypothesis **H1.2.** *The use of accessibility-oriented QA tools helps localisers to identify text alternatives as translatable content during the web localisation process*. This finding is of particular importance, given the results reported above with regard to H1.1. They imply that the use of these types of tools could prove beneficial to bridge the knowledge gap and somehow cover any potential limitations in this regard observed in localisation-oriented tools.

Finally, we deem it important to add that, contrary to what we expected, the order in which tools were used had an effect on the number of alt texts edited by localisers in the control group. Those using Acrolinx in the first place, before aDesigner, edited a significantly higher number of alt texts than those who followed the sequence aDesigner-Acrolinx. In this sense, it is worth noting that QA scenario B seems to have been not only beneficial to spot and edit full image text equivalents, but also to identify images with no `alt` attribute and amend the code accordingly. These preliminary results in terms of tool usage seem to already confirm the benefits of complementing general WAE tools with CL software for which we have advocated throughout the thesis. In addition, evidence appears to suggest that such a combination can be of particular assistance to those who have never received WA training.

All the above considered, it would be legitimate to already state that WA awareness and the use of accessibility-oriented QA tools can contribute to improve not only the alt text translation rate in multilingual websites, but also the image accessibility level of the localised website with respect to the source. This last assumption concerning the appropriateness of the changes made is based on the results observed with regard to the higher presence of `alt` attributes in the localised product, as well as on the premise that an image with this attribute is more accessible than an image without it. In the next section, we shall contrast this belief with the data obtained from the user evaluation, and explore if this statement can be extended to the rest of the text alternatives that were edited by localisers.

6.3 Appropriateness of localised text alternatives

In section 6.2, we have examined research question **R1** by investigating the role that targeted tools and knowledge can play in the identification of image text alternatives as translatable elements during the web localisation process. Indeed, we have shown that more alt texts are edited when localisers have a basic understanding of what web accessibility entails or use accessibility-oriented QA tools. In this section, our aim is to explore whether these changes have led or not to a significant improvement in terms of alt text quality, measured on the basis of their degree of appropriateness (see Chapter 5, section 5.3.1.2). The results reported will serve to test the hypotheses and sub-hypotheses related to research question **R2:** *Does the use of accessibility oriented QA tools during the web localisation process result in more appropriate text alternatives for images?* and research question **R3:** *Does knowledge of web accessibility help localisers to produce more appropriate text alternatives for images?* (see section 5.4, Chapter 5, for the complete list of hypotheses).

6.3.1 Data processing and coding

In order to investigate the above mentioned research questions, we jointly analysed the data collected throughout the two stages of our experimental study: the web localisation experiment (Stage 1) and the user evaluation (Stage 2). The results reported throughout section 6.3 are based, therefore, on database DB2, consisting of 76,440 alt text assessments (see section 6.1.1).

Below we offer an explanation of the levels in which the variables that will be treated henceforth in this chapter were divided (see Table 6.1 for a full list). This information should serve to better understand the interpretation of the results that we provide:

a) Appropriateness of alt texts:

As indicated in Chapter 5 (see section 5.3.4), the seven blind users recruited for the evaluation were requested to annotate text alternatives according to their degree of appropriateness, previously operationalised through a grading scale as follows: *1. Not appropriate - 2. Acceptable - 3. Pertinent - 4. Very pertinent.* When reporting the results, we will assume that, if alt texts were assigned scores 2, 3 or 4, a higher degree of alt text quality (and thus of image accessibility) was obtained in the localised website with respect to the source. Similarly, it should be noted that we will often refer to 3 and 4 as positive scores.

b) Translation condition:

When discussing the effect of the use of accessibility-oriented QA tools on the appropriateness of alt texts, we will consider the following four levels within **IV2**:[12]

- *none*, equivalent to translation version T1, when no tool was used, taking into account data from all 28 localisers;
- *Acrolinx*, equivalent to translation version T2, with data from the 14 localisers who use the tool before aDesigner;
- *aDesigner*, equivalent to translation version T2, with data from the 14 localisers who use the tool before Acrolinx; and
- *both*, equivalent to translation version T3, when both tools were used, taking into account data from all 28 localisers.

c) QA scenario:

In order to assess the impact that the QA scenario had on the quality of image text alternatives, we divided the aforementioned translation condition 'both' into two levels. This allowed us to make a distinction between the tools used to generate T3. Hence, when we will study the crossed effect of the QA scenario and other independent variables on DV2, five levels will be referred to:

- *none*, equivalent to translation version T1, when no tool was used, taking into account data from all 28 localisers;
- *Acrolinx*, equivalent to translation version T2, with data from the 14 localisers who use the tool before aDesigner;
- *aDesigner*, equivalent to translation version T2, with data from the 14 localisers who use the tool before Acrolinx;
- *Both-last-Acrolinx,* equivalent to translation version T3, when Acrolinx was used after aDesigner, taking into account data from 14 localisers; and
- *Both-last-aDesigner,* equivalent to translation version T3, when aDesigner was used after Acrolinx, taking into account data from 14 localisers.

6.3.2 Statistical methods

[12] Notice that the number of localisers will be reduced by half when the distribution of participants between groups will be considered during data analysis. The same applies for numbers shown in c) QA scenario.

To analyse *DV2. Appropriateness of alt texts*, we combined traditional descriptive statistics with regression modelling and significance tests. Before conducting this type of analysis, it was crucial to take into account two important characteristics of the data gathered. On one hand, the score per alt text is based on a Likert-type grading scale, i.e. data concerning **DV2** is ordinal, which cannot be assumed to be normally distributed. On the other hand, data comes from a repeated-measures experimental design, i.e. a study in which subjects were measured repeatedly under different conditions. Therefore, it seemed inevitable to assume that (i) there was a correlation between the observations made, and that (ii) the hierarchical structure of the data had to be taken into account when using inferential statistics.

After considering the above, the analysis approach adopted was mainly based on the use of parametric statistical tests, assuming that they would be robust, despite the non-normal distribution of the data, due to the vast amount of observations. Exceptionally, non-parametric tests or additional statistical analyses were run on the same data for comparison purposes in order to check whether similar conclusions were supported. Different tests and analyses were performed depending on the number of independent variables and the variable data type considered. These included analyses of variance (ANOVAs), particularly suitable for comparing the means in studies involving more than two conditions; and regression analyses, often used to investigate the relationship between one dependent variable and a number of independent variables (Lazar et al. 2010, 69–96). We have used an alpha level of .05 for all statistical tests. Further details about the specific statistical tests carried out and the models run are presented in each section before reporting the results they yielded.

The overall inferential statistical analysis of **DV2** was conducted in collaboration with two expert teams: the Servei d'Estadística at the Universitat Autònoma de Barcelona (UAB) and the Research Center for Statistics at the University of Geneva (UNIGE). While to measure the effect of the independent variable 'alt text type' on the quality of image text alternatives (see Chapter 7, section 7.2.1) we used SAS 9.3, the rest of the analyses included in this chapter and the subsequent one (Chapter 7) were performed with the R statistical software (library lme4 and multcomp). The choice of tool depended on the team of statisticians with whom we collaborated for each analysis.

6.3.3 Data analysis and interpretation

In the sections that follow, we analyse and discuss the effect of the two primary independent variables we have manipulated −WA knowledge (**IV1**) and use of accessibility-oriented QA tools (**IV2**)− on our second main dependent variable *DV2. Appropriateness of alt texts.* Similarly, we also observe the interaction effect of these variables (individually and jointly) and the QA scenario proposed to web localisers. To facilitate the flow of the argumentation, we will start with the analysis and discussion of the alt text quality results without taking into account the distribution of localisers between groups (section 6.3.3.1) and then present the evidence gathered considering their WA background (section 6.3.3.2).

6.3.3.1 Use of accessibility-oriented QA tools

Results by translation condition

In what follows we report the general results of the study with regard to the level of appropriateness of the image text alternatives generated by localisers. To this end, we look at the data gathered per translation condition (see section 6.3.1 for an explanation of the levels). Summary statistics about these data are shown in Table 6.12 and illustrated in Figure 6.8.[13]

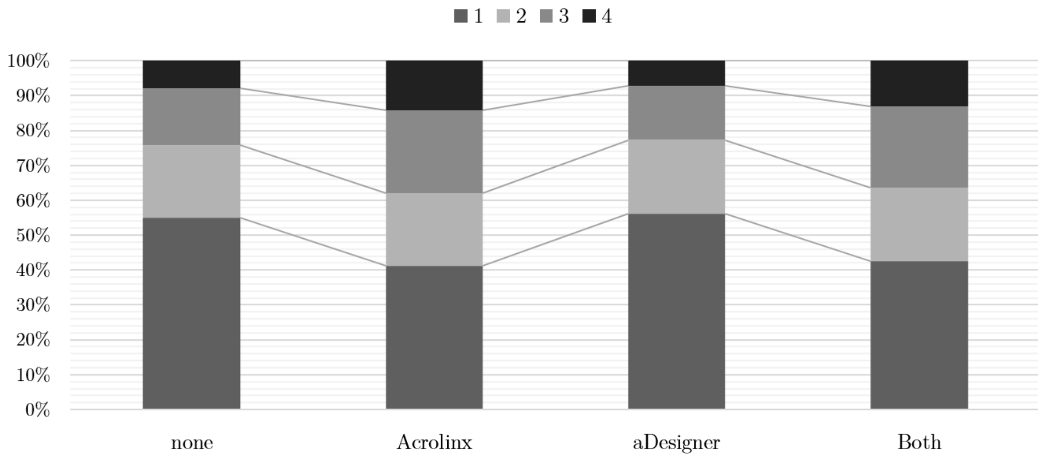


Figure 6.8. Overall distribution of alt text scores by translation condition: none, Acrolinx, aDesigner, both

Table 6.12. Overall proportion (%) of alt text scores by translation condition: none, Acrolinx, aDesigner, both

Score	Value	none	%	Acrolinx	%	aDesigner	%	Both	%
1	Not appropriate	14006	54.97	5252	41.22	7155	56.16	10841	42.55
2	Acceptable	5298	20.79	2642	20.74	2696	21.16	5377	21.10
3	Pertinent	4194	16.46	3028	23.77	1976	15.51	5952	23.36
4	Very Pertinent	1982	7.78	1818	14.27	913	7.17	3310	12.99

Secondly, we can notice reasonable differences between scores obtained for alt texts in T2 depending on the tool used. When localisers run checks with Acrolinx,

[13] To better interpret the gray scale colour coding used both in this figure and in Figure 6.9, notice that score 1 (not appropriate) is always at the bottom of the graphic (i.e. lowest part of the y-axis) and score 4 (very pertinent) is always at the top.

they managed to render accessible more than half of the alt texts available (58.78%). Conversely, despite having used aDesigner, 56.16% of the alt texts they produced in T2 were considered as not appropriate by blind users. As observed when comparing results for the pair none-both, there is a substantial difference between the proportions of positive scores for the pair Acrolinx-aDesigner. While, after using aDesigner, 15.51% of the localised alt texts were seen as pertinent and 7.17% as very pertinent, the proportion of positive scores obtained was higher after using Acrolinx (3=23.77%, 4=14.27%). Interestingly, the highest proportion of very pertinent alt texts and the lowest proportion of non-accessible images were obtained in T2 presumably thanks to the use of Acrolinx alone (see Table 6.12).

To statistically measure the effect of the use of accessibility-oriented QA tools on the alt texts appropriateness, we performed a repeated measure one-way analysis of variance (ANOVA) on the 4-level score scale, with three random factors (judge, alt text type and localiser). To this end, we used a linear mixed model (LMM) fitted with the program R (library lme4), where the fixed factor was the translation condition.[14] The analysis provides strong evidence that there is an effect of using tools on the scores collected, irrespective of the number or type of tools chosen (χ^2=1764, df=3, p<0.001). The model was complemented with a post-hoc analysis, for which we applied a Tukey's HSD (honest significant difference) correction (Tukey 1949) using the R library multcomp. The multiple comparisons of means (see Table 6.13) of the scores gathered per translation condition (none, aDesigner, Acrolinx, both) show that the difference between the scores obtained for text alternatives produced in translation version T3 (both tools) and those of translation version T1 (without any tool) is highly significant (p<0.001).[15] Additionally, we have found that (i) there is a significant difference in the scores of alt texts amended after using one tool (aDesigner or Acrolinx, translation version T2) compared to those obtained for alt texts in T1 (p<0.001); and that (ii) there is a significant difference between scores assigned to text alternatives checked with Acrolinx and those collected for aDesigner (p<0.001), when only one tool was used (T2). Finally, the analysis also indicates that, overall, the impact of using two tools on the scores gathered for alt texts is significantly greater than the effect of using only one (either aDesigner or Acrolinx, p<0.001).

[14] A linear mixed model (LMM) is a parametric linear regression model for clustered, longitudinal, or repeated-measures data that quantifies the relationships between a continuous dependent variable and various predictor or independent variables. An LMM may include both fixed-effect parameters associated with one or more continuous or categorical covariates and random effects associated with one or more random factors, hence its name (mixed model) (West et al. 2007).

[15] P-values computed for Tukey's test estimates have been adjusted for the multiple comparisons.

Table 6.13. Tukey's test results for scores comparison between translation conditions (four levels)

Comparison of translation conditions*	Estimate	Std. Error[16]	*p*-value[†]
both tools (T3) > none (T1)	0.2977	0.0073	*<0.001*
Acrolinx (T2) > none (T1)	0.2235	0.0096	*<0.001*
both tools (T3) > aDesigner (T2)	0.2146	0.0096	*<0.001*
Acrolinx (T2) > aDesigner (T2)	0.1405	0.0126	*<0.001*
aDesigner (T2) > none (T1)	0.0830	0.0096	*<0.001*
both tools (T3) > Acrolinx (T2)	0.0741	0.0096	*<0.001*

*Data organised in descending order, from highest to lowest estimate. / † Significant p-values shown in italics.

In order to check the robustness of the results against the non-normality of the scores, a non-parametric Kruskal Wallis test (K-W) was also performed. The inconvenience of this second approach is that it does not take the correlation structure of the data (repeated measures) into account. The test confirms the global significant effect of the use of tools on the judge's scores ($p<0.001$). The non-parametric post-hoc analysis (Nemenyi test) results also support the strongly significant differences found in scores between the following pairs: Both tools (T3) > no tool (T1), Acrolinx (T2) > no tool (T1), both tools (T3) > aDesigner (T2), and Acrolinx (T2) > aDesigner (T2); $p<0.001$. However, a weaker significance was observed in the remaining two pairwise comparisons: both tools (T3) > Acrolinx (T2) (p=0.008), and aDesigner (T2) > no tool (T1) (p=0.041). This correlates somehow with the low —although significant— improvements reported according to the Tukey's test, estimated at 0.0741 and 0.0830 respectively (see Table 6.13). Overall, the results from the repeated measure one-way ANOVA are confirmed.

The comparison of these findings with those reported in section 6.2.3.3 reveals that while Acrolinx had triggered changes in a higher number of text alternatives than aDesigner in T2, the improvement achieved with regard to T1 in terms of alt text quality thanks to the edits made is equally significant under both translation conditions. Yet, Acrolinx yields significantly better results when both tools are compared at the same level (T2) (see Table 6.13). In the same vein, we saw that localisers had modified, on average, fewer text alternatives in T3 than in T1 or T2. Still, results seem to indicate that these last few changes led to a significant improvement in the appropriateness of alt texts.

[16] The standard error represents the standard deviation associated with the estimate, i.e it describes the typical error or uncertainty associated with the estimate (Diez et al. 2012, 172). A good rule of thumb is that if the standard error is lower than the estimate, this will most probably be significant.

Results by QA scenario

As announced earlier in this chapter (section 6.3.1), alt texts from translation version T3 were coded according to the last tool used to better assess the effect of the secondary independent variable 'QA scenario'. This resulted in two more levels, namely: (i) *Both-last-Acrolinx,* when T3 was created following QA scenario A; and (ii) *Both-last-aDesigner* when T3 was generated following QA scenario B. Figure 6.9 and Table 6.14 provide again a summary of the overall proportion of alt text scores by translation condition, but this time in five levels. Although results for conditions *none*, *Acrolinx* and *aDesigner* have been already reported, we include them here once more just for comparison purposes.

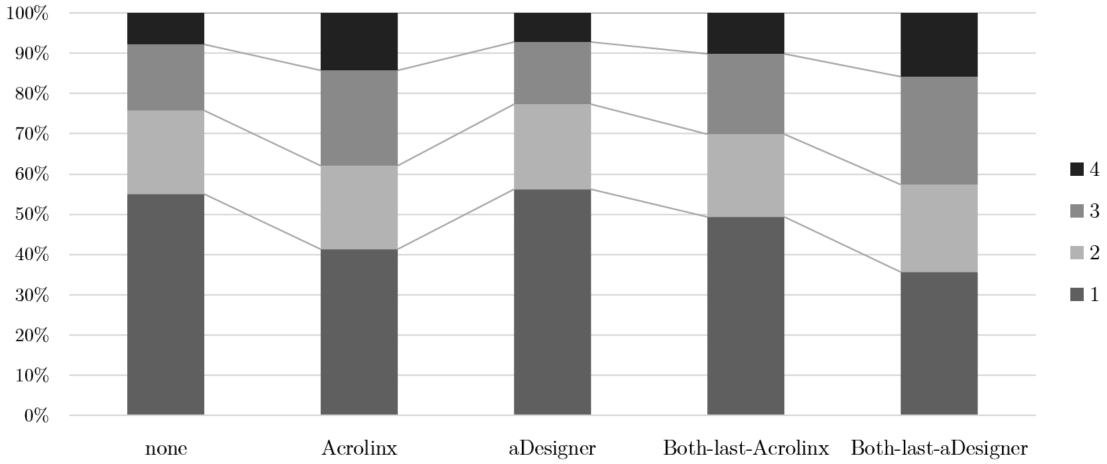


Figure 6.9. Overall distribution of alt text scores by translation condition, where both scores are divided according to the last tool used

Table 6.14. Overall proportion (%) of alt text scores by tool, according to order of use (QA scenario)

Score	Acrolinx	%	aDesigner	%	Both-last-aDesigner	%	Both-last-Acrolinx	%
1	5252	41.22	7155	56.16	4551	35.72	6290	49.37
2	2642	20.74	2696	21.16	2766	21.71	2611	20.49
3	3028	23.77	1976	15.51	3413	26.79	2539	19.93
4	1818	14.27	913	7.17	2010	15.78	1300	10.20

One interesting observation can be made when contrasting the distribution of the alt text scores within the same tool as per order of use. When used after aDesigner, Acrolinx leads to a lower proportion of positive scores than when it is the only tool proposed for alt text QA checking (there is almost a 10% increase in the number of non accessible images (41.22% vs. 49.37%; see Table 6.14, proportion of score 1 for Acrolinx and Both-last-Acrolinx). The most striking difference with regard to the quality of the alt texts produced by localisers is found when looking at aDesigner results. When localisers used aDesigner alone, they managed to render 43.84% of the images accessible. However, when used after Acrolinx, aDesigner helped them to create appropriate text alternatives for 64.28% of the images (these percentages consider all positive scores: 2-3-4). At first sight, it appears therefore that following QA Scenario B contributed to a noticeably higher decrease in the number of not appropriate alt texts than Scenario A if compared to translation condition 'none' (see Figure 6.9).

With a view to understanding whether the differences between T3 and T2 with regard to the scores obtained were equally significant irrespective of the QA scenario proposed, we used the same linear mixed model as for the previous analysis, but this time we considered five levels within the variable *IV2. Use of QA tools.* A non-parametric test was not performed in parallel given the fact that, in the previous analysis, it supported the same conclusions of the ANOVA. The post-hoc analysis determined that the difference between scores for alt texts verified with the help of two tools (T3) and those for alt texts edited after running checks with only one tool (T2) is very significant ($p<0.001$), regardless of the tool order (see Table 6.15). Furthermore, a significant difference was observed when comparing scores within the different conditions (aDesigner vs. Acrolinx) under which translation version T3 was performed ($p<0.001$). Results indicate that when the last automatic checks are run with aDesigner, localisers produce alt texts of significantly better quality than when the last tool used is Acrolinx.

Table 6.15. Tukey's test results for scores comparison between translation conditions (five levels)

Comparison of translation conditions*	**Estimate**	**Std. Error**	***p*-value**†
Both-last-aDesigner (T3) > none (T1)	0.3809	0.0103	*<0.001*
Both-last-aDesigner (T3) > aDesigner (T2)	0.3393	0.0145	*<0.001*
Both-last-Acrolinx (T3) > none (T1)	0.2144	0.0103	*<0.001*
Both-last-Acrolinx (T3) > aDesigner (T2)	0.1728	0.0103	*<0.001*
Both-last-aDesigner (T3) > both-last-Acrolinx (T3)	0.1664	0.0145	*<0.001*
Both-last-aDesigner (T3) > Acrolinx (T2)	0.1159	0.0103	*<0.001*
Acrolinx (T2) > Both-last-Acrolinx (T3)	0.0505	0.0145	*0.002*

*Data organised in descending order, from highest to lowest estimate. / † Significant p-values shown in italics.

As shown in Table 6.15, this more detailed analysis confirms the significant results observed in the first ANOVA as regards the impact of using both tools in comparison with none, no matter which tool is used in the second position. In this table, we have also included the findings resulting from the multiple comparisons of means within tools, depending on the order of use. The analysis reveals that there is a significant difference between the scores obtained when aDesigner was used in T3 and the scores from text alternatives produced during T2 after having run checks with that very same tool ($p<0.001$). Interestingly, in the case of Acrolinx, the significance test points at scores from T2 as significantly better than scores gathered for alt texts in T3 ($p=0.002$), even if the improvement achieved is estimated at only 0.0505 (see Table 6.15). Although we are aware that in these two last comparisons results may be slightly biased, since scores in T3 might be influenced by the impact of the tool used in the first place, we believe they illustrate and support the observations made so far about the interaction effect of the use of tools and the QA scenario followed. Overall, results from quantitative statistics and significance testing suggest that the sequence none-Acrolinx-aDesigner (Scenario B) lends itself particularly well to the goal of improving image accessibility in localised websites. We hypothesise that this might be due to the combined effect of (i) Acrolinx being more specific when flagging errors related to image accessibility, and (ii) aDesigner requiring localisers to be perhaps more aware of WA standards and their implementation. We will closely look at this second assumption when examining DV2 results by group.

6.3.3.2 Knowledge of web accessibility

In this section, we present the findings regarding the appropriateness of localised text alternatives taking also into consideration the effect of the independent variable *IV1. WA knowledge.* Results are therefore presented by group (control and treatment).

Results per translation condition by group

Figure 6.10[17] and Table 6.16 show a comparative summary of the distribution of alt text scores per group, according to the translation condition. A first glance at the data indicates that the most noteworthy differences between the control and the treatment group can be found in conditions 'none' and 'aDesigner'. Almost 60% of the text alternatives produced by the control group during the web localisation task were identified as not appropriate, 20.71% as acceptable, 14.55% as pertinent and 5.38% as very pertinent (see Table 6.16, column 'none'). In the case of the treatment group, the proportion of score 1 is lower (50.58%), and of positive scores higher (3=18.37% and 4=10.18%, see Table 6.16). As it is depicted in Figure 6.10, group differences in score proportions when using aDesigner are similar to those reported for condition ‘none’.

[17] In the case of Figures 6.10 and 6.11, the gray scale coulour coding goes from left to right, being score 1 (not appropriate) to the extreme left and score 4 (very pertinent) to the extreme right.

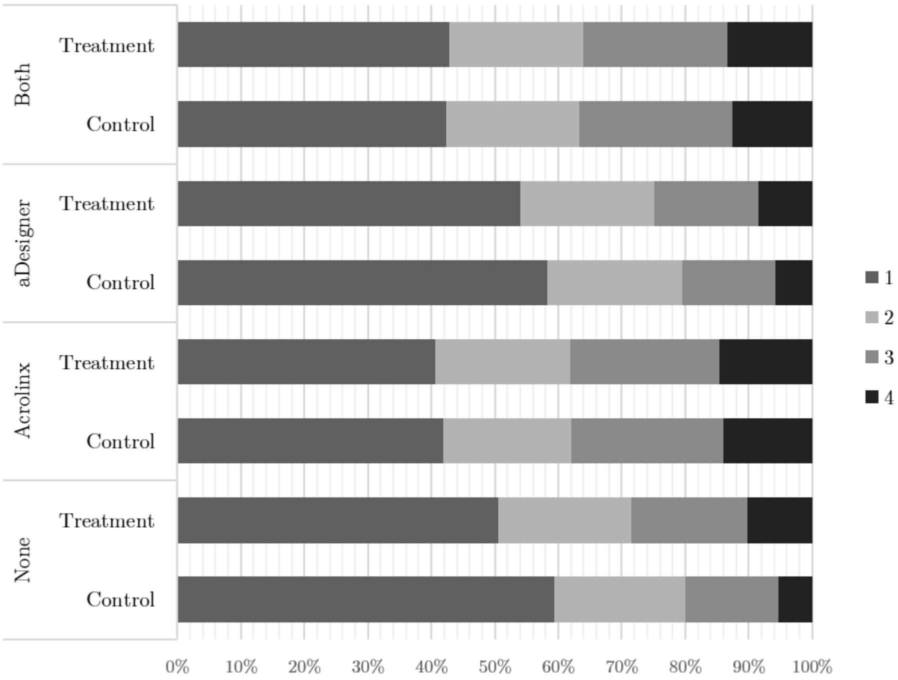


Figure 6.10. Distribution of alt text scores per group, by translation condition: none, Acrolinx, aDesigner, both

Table 6.16. Proportion (%) of alt text scores per group, by translation condition: none, aDesigner, Acrolinx, both

Score	Group	none	%	Acrolinx	%	aDesigner	%	Both	%
1	Control	7562	59.36	2668	41.88	3714	58.30	5391	42.32
	Treatment	6444	50.58	2584	40.57	3441	54.02	5450	42.78
2	Control	2639	20.71	1283	20.14	1348	21.16	2674	20.99
	Treatment	2659	20.87	1359	21.33	1348	21.16	2703	21.22
3	Control	1854	14.55	1528	23.99	936	14.69	3063	24.04
	Treatment	2340	18.37	1500	23.55	1040	16.33	2889	22.68
4	Control	685	5.38	891	13.99	372	5.84	1612	12.65
	Treatment	1297	10.18	927	14.55	541	8.49	1698	13.33

Additionally, it is worth noting that the total number of 3 and 4 scores seems to be higher for the treatment group, irrespective of the translation condition. As an interesting fact, it should also be mentioned that both groups obtained the highest proportion of positive alt text scores (2-3-4) when running checks with Acrolinx alone. Figure 6.10 also clearly illustrates that it was with this tool that both groups obtained the highest proportion of very pertinent text alternatives (control=13.99%, treatment=14.55%; see black boxes of the stacked bar chart). Surprisingly enough, the use of aDesigner seems to be counterproductive for the treatment group, whose data shows that there was a 4% increase in the number of non appropriate alt texts from translation version T1 to T2. Differences between groups appear to be more blurred when localisers used both tools during the QA task (see Figure 6.10, *Both*, top part). Contrary to what was observed for other translation conditions, the proportion of non appropriate text alternatives is almost the same for both groups, with minor differences appearing only in the positive alt texts scores (see Table 6.16). Overall, quantitative results per group seem to be in line with those presented in section 6.3.3.1 (see Figure 6.8) when referring to the same data but without taking into account the *IV1. WA knowledge*.

In an attempt to analyse the effect of this variable on the scores obtained under each translation condition, we used a linear mixed model with three random factors (judge, alt text type and localiser) and two fixed factors (group and translation condition). First, a Wald test (Engle 1984) was carried out to estimate the general impact of having WA knowledge of the alt texts quality. The test showed that the overall effect of **IV1** is not significant (χ^2= 1.0369, df=1, p=0.3085). We believe that this negative result might be influenced by the experimental design itself. Since participants were tested under multiple different conditions, it is likely that the impact of having a basic understanding of what web accessibility entails might differ depending on the tool(s) used. A post-hoc analysis (using the R package multcomp) was thus conducted to measure the interaction effect of both independent variables (**IV1** and **IV2**).

The multiple comparisons of means (see Table 6.17) of the scores gathered per group in each translation condition (none, aDesigner, Acrolinx, both) reveal that there is a significant difference between the scores assigned to alt texts produced by the treatment group and the control group in T1 (p=0.007). Similarly, the difference between groups as regards the scores obtained for alt texts produced after using aDesigner alone (T2) is significant (p=0.017). These data provide convincing evidence about the positive effect of having WA knowledge on the quality of text alternatives when localisers lack any automated QA support or just use aDesigner. Nevertheless, the significance test shows that, when localisers use Acrolinx alone or together with aDesigner, the significant effect of having WA knowledge disappears (see Table 6.17). These results corroborate what we had anticipated when observing the distribution of alt text scores per group and translation condition (see Figure 6.10). Notice that the Tukey's test also indicates that the effect of having WA knowledge is actually negative when the two tools proposed are used by the two groups (see Table 6.17). This means that, although the difference in scores is not significant, the control group did obtain slightly better scores than the treatment group when they jointly used aDesigner and Acrolinx, regardless of the order.

Table 6.17. Tukey's test results for scores comparison between groups: Treatment > Control

Translation condition*	Estimate	Std. Error	*p*-value†
none (T1)	0.2219	0.0918	*0.007*
aDesigner (T2)	0.1122	0.0529	*0.017*
Acrolinx (T2)	0.0200	0.1844	0.457
both tools (T3)	-0.0047	0.1353	0.514

*Data organised in descending order, from highest to lowest estimate. / † Significant p-values shown in italics.

When these data are checked against the total amount of alt texts edited per group by translation version and tool (see sections 6.2.3.2 and 6.2.3.3), the following remarks can be made. We saw that, during T2, the control group had performed changes in twice the number of text alternatives than the treatment group, Acrolinx being the tool that triggered the majority of them. It thus follows that those modifications were appropriate when Acrolinx was used, but not enough when aDesigner was the proposed tool. Similarly, during T3, few changes were performed by both groups. It seems, however, that those made by the control group were very pertinent and helped them to achieve a degree of alt text quality comparable to that of the treatment group. The most interesting inference that can be drawn from these findings is that using an accessibility-oriented QA tool like Acrolinx can help localisers cover the lack of previous background on web accessibility by not only boosting alt texts' visibility, as we have suggested before (section 6.2.4), but also by providing useful guidance on how to render images accessible. Additionally, given the results above, one could believe that tools are not that effective for localisers who received only some basic training on WA because this kind of software may have pointed to more complex accessibility aspects than the ones they had reviewed before, thus making them feel less confident about their knowledge.

Results per QA scenario by group

After the overall analysis per translation condition, we also investigated the differences between groups with regard to the effect on the alt texts' appropriateness of using the two tools in a different order. Figure 6.11 shows a cumulative stacked bar chart depicting the distribution of scores per group by translation condition in five levels (as opposed to the one presented in Figure 6.10). Results from conditions 'none', 'Acrolinx' and 'aDesigner' are kept in the chart for comparison purposes. The corresponding data is shown in Table 6.18.

In the previous section we reported that when participants had run QA checks with both tools, differences between groups with regard to the quality of the alt texts produced were almost non-existent and thus not significant. Nevertheless, if we consider which tool was used to generate T3, differences across groups become more noticeable. When the last tool proposed during the QA task was Acrolinx (Scenario A), the control group reduced the number of non-appropriate text alternatives only by approximately 8% (T1=59.36%; T2=58.30%; T3=51.92%). However, when localisers from this group followed QA Scenario B (Acrolinx-aDesigner), a 27% decrease was observed in the number of non accessible images from T1 to T3 (32.72%). Interestingly, following this QA scenario led participants from the control group to achieve slightly better results than the treatment group, who registered a higher proportion of score 1 at T3 (38.73%) than the former when using aDesigner as the last tool (see Table 6.18). It is also worth noting that, regardless of the group, the highest proportions of scores 3 (Pertinent) and 4 (Very pertinent) were obtained when using aDesigner in the second position (control group: 3=29.23%, 4=16.17%; treatment group: 3=24.35%, 4=15.38%; see Table 6.18). The distribution of alt text scores per participant and QA scenario followed are presented in Appendix K.2 (Tables K.3 to K.6 and Figures K.1 to K.4).

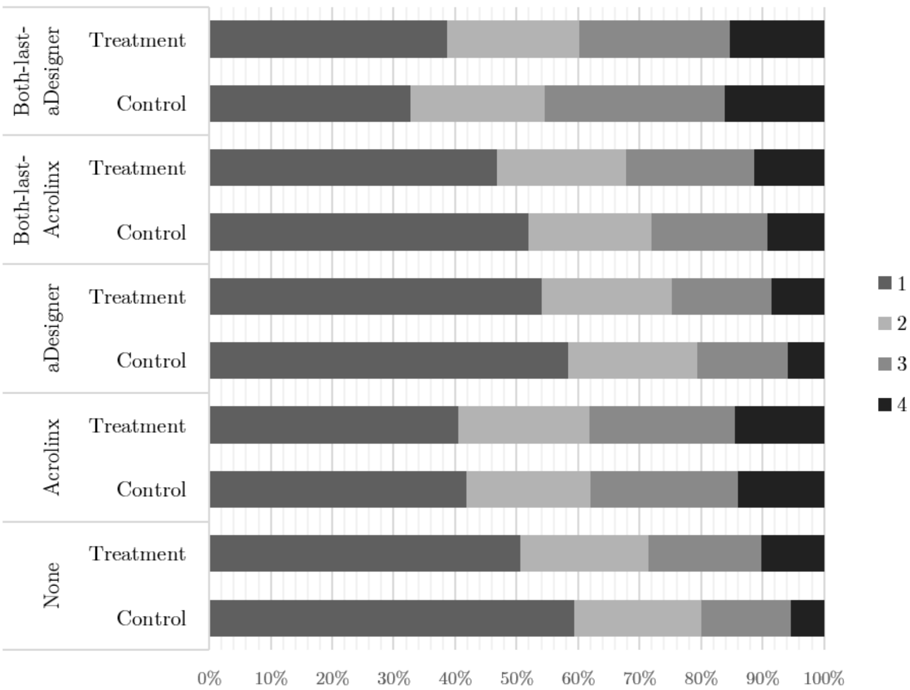


Figure 6.11. Distribution of alt text scores per group, by translation condition, where both scores are divided according to the last tool used

Table 6.18. Proportion (%) of alt text scores per group by tool, according to order of use (QA scenario)

Score	Group	Acrolinx	%	aDesigner	%	Both-last-Acrolinx	%	Both-last-aDesigner	%
1	Control	2668	41.88	3714	58.30	3307	51.92	2084	32.72
	Treatment	2584	40.57	3441	54.02	2983	46.83	2467	38.73
2	Control	1283	20.14	1348	21.16	1280	20.09	1394	21.88
	Treatment	1359	21.33	1348	21.16	1331	20.89	1372	21.54
3	Control	1528	23.99	936	14.69	1201	18.85	1862	29.23
	Treatment	1500	23.55	1040	16.33	1338	21.00	1551	24.35
4	Control	891	13.99	372	5.84	582	9.14	1030	16.17
	Treatment	927	14.55	541	8.49	718	11.27	980	15.38

Overall, the sequence aDesigner-Acrolinx (QA scenario A) triggered fewer improvements for both groups, although the alt texts produced by the treatment group received the highest proportion of positive scores (2-3-4) in all translation conditions. These data suggest that while both scenarios lead to an improvement of the alt texts' appropriateness, QA Scenario B seems to have been more effective, particularly for the control group. This echoes the conclusions drawn upon the analysis of the amount of alt texts edited by each group depending on the QA scenario (see section 6.2.3.4). At a tool level, the general observations made in section 6.3.3.1 appear to be confirmed since, for both groups, Acrolinx triggered higher scores when used during T2 instead of T3, and aDesigner proved more useful in T3 than in T2.

To estimate whether the differences observed between groups were statistically significant or not, we replaced the four-level tool fixed factor (none, Acrolinx, aDesigner, both) with a five-level factor (none, Acrolinx, aDesigner, Both-last-Acrolinx, Both-last-aDesigner) in the linear mixed model (LMM) used for previous analyses. The model was complemented with a post-hoc analysis, for which we applied again a Tukey's HSD correction (Tukey 1949) using the R library multcomp. This allowed us to compare the performance of both groups at T3, depending on the QA tool used. Results shown in Table 6.19 (which complement those presented in Table 6.17) reveal that there is no significant difference between the scores assigned to alt texts produced by the treatment group in T3 using Acrolinx and the scores obtained by the control group under the same translation condition (p=0.205). Similarly, no significant differences between groups are found when aDesigner was the tool used during T3 (p=0.732). In fact, the test shows a negative estimate for the comparison Treatment > Control at T3 (Both-last-aDesigner), which supports the observations made earlier with regard to the high proportions of positive scores (3-4) registered by the control group under this condition.

Table 6.19. Tukey's test results for scores comparison between groups at T3: Treatment > Control

Translation condition	Estimate	Std. Error	*p*-value†
Both-last-Acrolinx (T3)	0.1151	0.1396	0.205
Both-last-aDesigner (T3)	-0.1246	0.2019	0.732

† Significant p-value shown in italics.

Impact of WA knowledge, tool used and QA scenario

The general linear hypotheses tested and reported in Tables 6.17 and 6.19 served us to compare group results at the same level (or translation condition). However, in order to measure the crossed effect of the three independent variables (WA knowledge, tool used and QA scenario) on the alt text scores, we looked directly at the LMM estimates (see Table 6.20).[18] Within the LMM output, the intercept should be interpreted as the mean of the outcome (i.e. the alt text score) when all the predictors (i.e. the three independent variables considered in the analysis) have a value of zero. In our study, the intercept is determined by the results from the control group at T1, when no tool was used, and is estimated at 1.6594. Each predictor estimate (i.e. predicted improvement per additional factor level) shown in Table 6.20 should thus be added to the intercept to calculate the other predicted mean scores. To understand whether these improvement estimates were significant or not, we computed the *p*-value for each one of them (significant *p*-values are shown in italics).

Before discussing the results of the LMM, let us consider the pair T1-T2 (Acrolinx) for illustrative purposes. The LMM indicates that the predicted mean score for the control group at T2 when Acrolinx was used is 2.0836 (μ=1.6594 + β_1=0.4242, where the improvement achieved thanks to the tool is understood as significant: p<0.001, see Table 6.20). In the case of the treatment group, the predicted score is lower, and it is estimated at 1.9871 (μ=1.6594 + β_1=0.4242 + α=0.2219 + α_1= -0.3184), where the significant positive effect of having WA knowledge (p=0.010) is in fact nullified by the use of Acrolinx, which has a significant negative impact (p<0.001) on the score.

[18] For comparison purposes, a generalised linear model (GLM) was also performed, with a view to accounting for the specific distribution of the response variable (ordinal). The output of the cumulative linear mixed model used is reported in Appendix K.2 (Table K.7). While estimates differ from those yielded by the LMM, significance results are uniform across the two regression models. Hence, it was decided to carry out the analysis using only a LMM, since it is technically less complex to run and interpret.

Table 6.20. LMM output predicting the interaction effect of three independent variables: WA knowledge, tool used and QA scenario followed on the alt texts score

Coefficient	Group	Translation condition	Estimate	Std. Error	*p*-value
μ	**control**	**none (T1)**	**1.6594***	0.2219	*<0.001*
β_4	control	Both-last-aDesigner (T3)	0.6119	0.0145	*<0.001*
β_1	control	Acrolinx (T2)	0.4242	0.0145	*<0.001*
α	treatment	none (T1)	0.2219	0.0968	*0.010*
β_3	control	Both-last-Acrolinx (T3)	0.2096	0.0145	*<0.001*
β_2	control	aDesigner (T2)	0.0382	0.0145	*0.004*
α_3	treatment	Both-last-Acrolinx (T3)	0.0096	0.0205	0.319
α_2	treatment	aDesigner (T2)	0.0068	0.0205	0.369
α_1	treatment	Acrolinx (T2)	-0.3184	0.0205	*<0.001*
α_4	treatment	Both-last-aDesigner (T3)	-0.4631	0.0205	*<0.001*

*Intercept (predicted value of the dependent variable when all the independent variables are equal to 0).

The LMM output suggests that the improvements achieved in terms of alt texts appropriateness from T1 to T3 were statistically significant for the control group, regardless of the QA scenario followed. Nevertheless, interesting observations can be made regarding the treatment group. First, it seems that the only significantly positive effect on the score is triggered by their previous WA knowledge. When QA Scenario A (aDesigner-T2 - Both-last-Acrolinx-T3) was followed, there were minor improvements (0.0068 and 0.0096 respectively), but not significant. Surprisingly enough, using the tools in the order established by QA Scenario B appears to have had a significantly negative effect on the alt texts score at T3 (see Table 6.20).

To facilitate the comparison between groups and the interpretation of the LMM output, we have calculated the predicted mean score per group across the five different translation conditions, as we have previously done for the pair T1-T2 (Acrolinx) (see Table 6.21). These predicted scores are illustrated in the double line chart presented in Figure 6.12, where the black line represents the control group and the red line accounts for the treatment group. The distance between the x-axis and the first marker (small dot) indicates the effect of the tool(s). In turn, the effect of WA knowledge is depicted by the distance between the closest marker to the x-axis and the second marker right above it. When predicted values per QA scenario are compared across groups, the most striking observation that emerges is the following: while for the control group, the crossed effect of having used first Acrolinx and then aDesigner is higher than when tools were used in reverse order (predicted score: 2.2713 as opposed to 1.8690), the interaction effect of having knowledge and the use of tools has only proved to be positive for localisers who followed QA Scenario A (aDesigner-Acrolinx, predicted score: 2.1005).

Table 6.21. Predicted average alt text score per group across translation conditions (five levels)

Translation condition	*Control group* Interaction effect	Predicted score ($\bar{x}$)	*Treatment group* Interaction effect	Predicted score ($\bar{x}$)
none (T1)	μ	1.6594	$\mu + \alpha$	1.8816
Acrolinx (T2)	$\mu + \beta_1$	2.0836	$\mu + \beta_1 + \alpha + \alpha_1$	1.9871
aDesigner (T2)	$\mu + \beta_2$	1.6976	$\mu + \beta_2 + \alpha + \alpha_2$	1.9263
Both-last-Acrolinx (T3)	$\mu + \beta_3$	1.8690	$\mu + \beta_3 + \alpha + \alpha_3$	2.1005
Both-last-aDesigner (T3)	$\mu + \beta_4$	2.2713	$\mu + \beta_4 + \alpha + \alpha_4$	2.0301

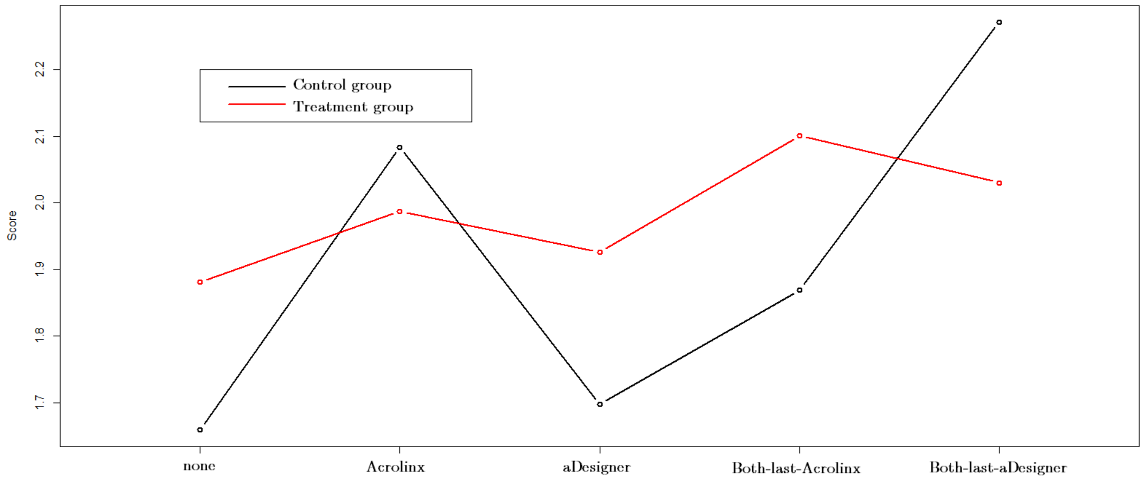


Figure 6.12. Predicted average alt text score per group across translation conditions (five levels)

All in all, the regression analysis and the significance tests performed reveal that using one accessibility-oriented QA tool alone or a combination of two complementary tools is significantly beneficial for localisers who do not have a background in web accessibility to achieve more appropriate text alternatives in the localised website. Regardless of the tool order, improvements achieved by the control group are significant, although the sequence Acrolinx-aDesigner (QA Scenario B) leads to better results. Specifically, the LMM indicates that the positive impact associated with the use of Acrolinx alone is higher than the positive effect of having received basic training on web accessibility (see Table 6.20, Estimates column). This finding is in line with the Tukey's test results presented in Table 6.17. As a general remark, we could thus infer that with regard to image accessibility assurance, using a CL checker like Acrolinx would be more effective than following a short WA training, at least for localisation professionals.

As far as the treatment group is concerned, the order in which tools are used affects localisers' performance in terms of alt text quality improvement. When they followed the sequence aDesigner-Acrolinx, they produced slightly better text alternatives, but the improvements achieved were not significant with respect to those implicitly obtained thanks to having previous knowledge of web accessibility. Interestingly, the alt text editing effort was almost the same for this group regardless of the QA scenario (alt texts edited during QA task: Scenario A=343, Scenario B=332; see section 6.2.3.4). It thus follows that changes performed in T2 and T3 under QA Scenario A were not pertinent enough to trigger a significant improvement. Likewise, data suggests that the use of Acrolinx and aDesigner combined was counterproductive under QA Scenario B.

6.3.4 Conclusions

In Section 6.3, we presented the results related to the appropriateness of localised image text alternatives (dependent variable **DV2**). We are now in a better position to provide an answer to research questions **R2**: *Does the use of accessibility-oriented QA tools during the web localisation process result in more appropriate text alternatives for images?* and **R3**: *Does knowledge of web accessibility help localisers to produce more appropriate text alternatives for images?* In order to do so, we re-examine their underlying main hypotheses and sub-hypotheses and conclude whether they can be supported or need to be rejected.

6.3.4.1 Use of accessibility-oriented QA tools

The independent variable *IV2. Use of accessibility-oriented QA tools* has been operationalised in our study through two specific tools: Acrolinx and aDesigner. The analyses performed have revealed that, when localisers ran checks with one or both tools, results in terms of alt texts' appropriateness were significantly better than those achieved when none of these tools were used. We can thus confirm our main alternative hypothesis associated with **R2**, which stated that *using accessibility-oriented QA tools during the web localisation process has a positive impact on the appropriateness of translated text alternatives for images*.

More specifically, a significant improvement with regard to alt text quality was observed from T1 to T2, regardless of the tool proposed. Nevertheless, when data from T2 were compared, the statistical tests carried out indicated that Acrolinx led localisers to achieve significantly more appropriate alt texts than aDesigner. This finding confirms our first sub-hypothesis: **H1.1.** *When only one evaluation tool is used, a controlled language tool with style-oriented rules for image text alternatives' checking like Acrolinx helps localisers to achieve more appropriate text alternatives than a general web accessibility evaluation tool like aDesigner*. The implications of supporting this hypothesis are of considerable importance for both the localisation and accessibility community. On one hand, our study suggests that linguistic aspects can have indeed a strong impact on the accessibility of the multilingual Web, and that, therefore, it seems reasonable to involve content producers, like localisers, in accessibility assurance processes. On the other hand, we have shown that relying on NLP-based software like Acrolinx, which moves away from the practices currently followed during automated WAE, can lead to significantly higher levels of image accessibility (and potentially of other language-related aspects of WA). By demonstrating that a CL checker could cover the limitations identified in current state-of-the-art automated solutions for image accessibility verification (see Chapter 3, section 3.4), our work somehow confirms Lawton Henry and Abou-Zahra's (2014) belief that research and development in natural language processing (NLP) can bring new insights into the web accessibility field.

The order in which the two accessibility-oriented QA tools proposed were used was also taken into account when we analysed the data collected from the study. During the QA task, 14 participants had been randomly requested to check their localisation proposals for image accessibility using first aDesigner and then Acrolinx (QA Scenario A), while the other 14 participants had been asked to use them in reverse order (QA Scenario B). Although this decision was initially taken to avoid a biased improved or degraded performance of the said tools due to the so-called sequence effect, we have found that the appropriateness of alt texts was affected differently by the use of tools depending on the order of use. The observation of the alt text scores distribution across the five different translation versions (T1-none, T2-aDesigner, T2-Acrolinx, T3-Both-last-aDesigner, T3-Both-last-Acrolinx) and the inferential statistical methods applied showed that Acrolinx was significantly more effective when used in the first place than when used after aDesigner. From this finding we were able to infer that Acrolinx alone leads to significantly better results than aDesigner and Acrolinx together, if this precise order is respected (QA Scenario A). This reinforces the conclusions we have drawn earlier with regard to the benefits of using CL software for accessibility purposes.

In addition, we have found that aDesigner triggered the production of alt texts of significantly higher quality when used after Acrolinx than when used alone. This result was later corroborated when observing that localisers who followed QA Scenario B obtained significantly better results concerning alt texts' appropriateness than those who followed QA Scenario A. We are confident, therefore, in rejecting sub-hypothesis **H1.2**, which predicted that *when two evaluation tools are used, a controlled language tool with style-oriented rules for image text alternatives' checking like Acrolinx leads to more improvements than a general web accessibility evaluation tool like aDesigner, irrespective of the order in which they are used.* As we have highlighted earlier in this chapter, this might have been due to limitations in the experimental design, such as the constraints imposed in terms of time. One could believe that aDesigner, being a more complete tool in terms of WCAG 2.0 coverage, required localisers to spend more time filtering results and interpreting them, thus making them invest less time in Acrolinx. Similarly, the fact that aDesigner has not been originally developed to be used by professionals with a more linguistic background might have caused confusion or frustration when used in the first place, hence discouraging localisers from making efforts in subsequent steps of the experiment.

While the order of use proved to be a decisive factor with regard to the effectiveness of both tools as far as image accessibility assurance is concerned, data analysis has revealed that the quality of the alt texts produced in T3, irrespective of the tool used, was significantly higher than the quality of text alternatives included in T2. Hence, we can accept our last more general sub-hypothesis (**H1.3**), which stated that *using two different tools triggers more improvements in terms of text alternatives' appropriateness than using only one.* This finding is in agreement with the conclusions of Vigo et al. (2013), already highlighted in Chapter 2 (section 2.3.3), who reported that the effectiveness of automated testing techniques during WAE can be improved when two complementary tools are used with a view to boosting coverage, completeness and correctness. In the case of image accessibility assessment, we have demonstrated that the combination of a tool that performs syntactic and heuristic checks like aDesigner and a tool that verifies linguistic aspects can be optimal to assure an acceptable level of image accessibility in localised websites.

Finally, it should be noted that all the conclusions we have drawn so far with regard to the use of QA tools rely on the assumption that the improvements achieved in terms of image accessibility have been primarily triggered by the tools themselves. Yet, we have not examined the check reports yielded by Acrolinx and aDesigner to carefully study the precision and recall of both tools, and thereby ensure that the achievement of higher levels of alt text quality was not a result of chance or influenced by other variables. While we are aware of this limitation, we are confident that the analysis of the impact of several identified variables on the appropriateness of localised text alternatives, such as the localisers' previous knowledge of WA (see next section, 6.3.4.2) or the verification of the tools' help support (see Chapter 7), to name just a few, will serve to reinforce the validity of our findings.

6.3.4.2 Knowledge of web accessibility

To facilitate the manipulation of *IV1. Web accessibility knowledge*, participants were divided into two groups. In contrast to the control group, the treatment group attended a webinar on best accessibility practices for localisers prior to the study. In order to test the hypotheses associated with **R3**, we compared the alt text quality results of both groups. The analyses carried out have indicated that the participants who had followed that brief training produced significantly more appropriate texts alternatives in T1 than those who had not. Although at a different level ‒given that it is the first time that an accessibility study is conducted with localisation professionals‒, this finding is consistent with similar investigations reported in the literature where the effectiveness of short training sessions on web accessibility concerns was examined. For instance, Chevalier and Ivory (2003) found that, after an introductory webinar on accessible-aware web development, novice WA evaluators (with a web design background) were up to 31% more effective than other untrained novice evaluators in detecting accessibility barriers. In addition, it could be stated that the outcome of our experiment echoes that of the preliminary observational study we conducted with localisation students (Rodríguez Vázquez 2014), thus endorsing the idea that increasing accessibility awareness among future localisation professionals (e.g., by introducing WA training in the translation and localisation curricula) could significantly contribute to enhance the accessibility of the multilingual Web.

All in all, however, we believe that the evidence gathered through the experimental study can only partially support our main hypothesis (**H1**), which reads as follows: *Having web accessibility knowledge has a positive impact on the appropriateness level achieved in the translated text alternatives for images.* The reason is that we have found that the effect of having some basic web accessibility background on the final alt text quality varies according to which and how many QA tools are used to check the localised website for image accessibility. This might be related to the phenomenon of the 'evaluator effect' that we have commented on earlier in this thesis (see Chapter 2, section 2.3.3). It could be hypothesised that if localisers had had a higher level of expertise on WA issues, our study would have yielded more conclusive results. This assumption is based on prior work where human-based accessibility evaluations with expert and novice evaluators have revealed that, while the former are more judgemental, the latter tend to underestimate the severity of accessibility barriers and are more likely to produce false positives (Yesilada et al. 2009; Brajnik et al. 2012). Within the particular context of our thesis, it is possible that localisers who attended the WA webinar (novice evaluators themselves) had similar problems when using and interpreting the feedback from the QA tools proposed. In this sense, it would be interesting to complement our study with another one involving expert evaluators with a localisation background, with a view to further investigating the impact of the evaluator effect on the final alt text quality.

In terms of tools usage, we found that aDesigner alone did not prove to help localisers from the control group to achieve comparable alt text quality results to those from the treatment group. Nevertheless, when participants used Acrolinx to generate T2, the positive effect of having web accessibility knowledge disappeared, as it was somehow compensated by the high performance of the CL checker. Moreover, the linear mixed model estimated that the positive impact associated with the use of Acrolinx alone was higher than the positive effect of having received basic training on web accessibility. In addition, when we compared the group results at T3, we noticed that the control group had performed slightly better in terms of alt texts' appropriateness than the treatment group, although the differences found were not significant. The comparison between groups also yielded interesting results with regard to the QA scenario followed. When Acrolinx was the last tool used to run an image accessibility check (QA Scenario A), the treatment group obtained better (yet not significant) results than the control group. Nevertheless, the alt text quality outcome was the opposite when QA Scenario B was followed.

Based on the findings summarised above, we need to reject the following two sub-hypotheses:

H1.1 *Localisers with basic knowledge of web accessibility that use accessibility evaluation tools during the localisation process achieve more appropriate text alternatives than localisers who do not have previous knowledge of web accessibility and have used the same tools;* and

H1.2 *Localisers with basic knowledge of web accessibility that use a controlled-language tool with style-oriented rules for image text alternatives' checking during the localisation process achieve more appropriate text alternatives than localisers who do not have previous knowledge of web accessibility and use the same controlled-language tool.*

In this regard, our research suggests that the use of accessibility-oriented QA tools can help localisers without any WA background reach a similar (and even higher) level of effectiveness during accessibility assessments to that of localisers who have received some basic training on the topic. Although we are aware that this finding might be influenced by the fact that localisers from the treatment group had only received a one-hour introductory session on the matter, our results contrast —at least from an image accessibility point of view— with the claims made in the literature about the negative effects of relying on automated accessibility testing alone (Vigo et al. 2013). Indeed, based on the outcome of our experimental study, we could state that if the proper tool (or the adequate combination of several tools) is integrated in the web product quality assurance workflow, acceptable levels of accessibility could be achieved even if the users of such tools do not have an extensive expertise on the matter. This does not necessarily imply that WA training is not essential for localisers to ensure the transfer and adaptation of accessibility from the source to the target website, which, as we have seen before, can be critical when no automated support is offered. Yet, it brings new insights into the potential of considering the introduction of WAE tools in current localisation QA workflows, including within teams that do not have accessibility experts. The additional findings that will be presented in Chapter 7 will help us to expand on our study's implications.

6.4 Summary

In this chapter, we have reported and discussed the results from the central empirical study of this thesis. While section 6.2 covered the data analysis and interpretation concerning *DV1. Translation of alt texts*, in section 6.3 we presented how *DV2. Appropriateness of localised text alternatives* had been influenced by our two primary independent variables (*IV1. Web accessibility knowledge* and *IV2. Use of accessibility-oriented QA tools*). Similarly, we have examined the interaction effect of both IVs (individually and jointly) with the QA scenario followed. These data allowed us to test the hypotheses and sub-hypotheses associated with the three research questions of the experimental study, summarised in Table 6.22 together with the results highlights.

The evidence gathered has served us to determine the extent to which localisers are capable of assuring that an acceptable level of image accessibility is achieved in the target web product, which was one of the goals of the present thesis (see section 1.3, Chapter 1). Our study has shown that the ability of localisation professionals to, first and foremost, identify text alternatives as translatable elements, and then assess (and improve, when needed) their appropriateness according to the context and purpose of the images they replace is strongly influenced by two main factors: (i) their degree of awareness about what WA entails and how it can be implemented, and (ii) the tools they have at their disposal to assist them during the QA phase of the web localisation process.

We have found that, contrary to expectations, the need to translate image text equivalents during the web localisation process is not taken for granted by localisers. Only those who have some background in web accessibility seem to try to ensure that these elements are localised in the target website. In addition, we have observed that the use of accessibility-oriented QA tools is of particular assistance for localisers who are not aware of accessibility standards and the way in which blind users interact with the Web. More specifically, using a CL checker like Acrolinx, helps them identify a significantly higher number of alt texts than aDesigner, which is a rather general WAE tool. Furthermore, we noticed that for these localisers the joint use of both tools can lead to twice the number of alt text edits when compared to the localisation product they would deliver if these tools were not available.

In terms of alt text quality, our study confirms what we had anticipated. Knowing that web accessibility requirements exists and that these are key for people with disabilities to access the Web contributes to produce a significantly higher number of appropriate image text alternatives in the target website. Nonetheless, surprising results have been observed with regard to the use of QA tools. It appears that following Acrolinx suggestions and recommendations regarding image accessibility leads to better results in terms of alt text quality than having followed a one-hour webinar about web accessibility. Having WA knowledge appears to be significantly more beneficial only when a WAE tool based on syntactic and heuristic checks like aDesigner is used. This means, in fact, that Acrolinx provides relevant guidance for understanding which alt text formulations are more convenient depending on the type of image they are associated with, to the extent that it is comparable to the inferences on the matter that can be drawn by localisers who are accessibility aware. Moreover, we know now that the combination of CL software and general automated WAE support can enhance the quality of image text alternatives in localised websites.

The findings reported in this chapter have implications at several levels. First, they confirm that existing limitations of state-of-the-art web and image accessibility evaluation tools with regard to the detection of inappropriate alt texts can be covered by the application of tailored controlled-language rules by means of a CL checker. Second, by demonstrating that localisers play an important role in the production of multilingual websites, we have strengthened the arguments that we have put forward throughout this thesis towards a higher level of involvement of these professionals in the achievement of a more accessible Web for all. Third, and as result of the above, our study may well have a bearing on current localisation QA approaches, as it has shown that not considering accessibility issues can have a significantly negative impact on the quality of the localised web product in terms of user experience. The discussion about the findings of our experimental study and their implications will be continued in the next chapter (7).

Table 6.22. Overview of hypothesis testing outcome and results highlights

Research Question		Hypothesis	Supported or rejected?	Results highlights
R1. Are alt texts considered as translatable elements during web localisation?	H1	Alt texts are considered as translatable elements	Partially supported	- A considerable degree of variability was observed as regards the awareness about the existence of text alternatives and the need to translate them, especially across groups. - Six (N=6) out of 28 participants did not translate any text alternative at all.
	H1.1	Web accessibility knowledge facilitates the identification of alt texts	Supported	- The treatment group translated, on average, twice the number of alt texts than the control group. - The difference across groups with regard to total number of alt texts edited is significant. - The only two participants who spotted images with no `alt` attributes and added them in T1 were from the treatment group.
	H1.2	The use of QA tools facilitates the identification of alt texts	Supported	- The six localisers who had not translated any alt text in T1 identified them either in T2 or T3 - Acrolinx triggered a higher number of edits than aDesigner. - For the control group, when only one tool was used, Acrolinx led to a significantly higher number of edits than aDesigner - For the control group, the combined used of Acrolinx and aDesigner triggered modifications in twice the number of text alternatives if compared to alt texts edited in T1.

Table 6.22. (continued)

Research Question		Hypothesis	Supported or rejected?	Results highlights
R2. Does the use of accessibility-oriented QA tools result in more appropriate image alt texts?	H1	The use of QA tools has a positive impact on alt texts' appropriateness	Supported	- The combined use Acrolinx and aDesigner (irrespective of the order) led to alt texts of significantly higher quality than those produced when no tool was used. - Using Acrolinx or aDesigner individually triggered the production of alt texts of significantly higher quality than those generated when no tool was used.
	H1.1	If only one tool is used, Acrolinx leads to better results than aDesigner	Supported	- Alt texts from T2, edited after having used Acrolinx, were significantly more appropriate than alt texts from T2, produced with the help of aDesigner.
	H1.2	If both tools are used, Acrolinx leads to better results than aDesigner, regardless of the order of use	Rejected	- When Acrolinx was used to generate T2, text alternatives received significantly higher scores than alt texts from T3, produced with the help of the same tool. - When aDesigner was used after Acrolinx, alt texts were significantly more appropriate than the text alternatives from T2, when aDesigner was the first tool proposed.
	H1.3	Using two QA tools leads to better results than using only one	Supported	- Alt text quality results were significantly better in T3-Both-last-Acrolinx than in T2-aDesigner. - Alt text quality results were significantly better in T3-Both-last-aDesigner than in T2-Acrolinx.

Table 6.22. (continued)

Research Question		Hypothesis	Supported or rejected?	Results highlights
R3. Does web accessibility knowledge help to produce more appropriate image alt texts?	H1	Web accessibility knowledge has a positive impact on alt texts' appropriateness	Partially supported	- The alt texts produced by the treatment group in T1 were of significantly higher quality than the alt texts produced by the control group. - The effect of having a basic web accessibility background varied according to which and how many QA tools were used to check T1 for image accessibility.
	H1.1	Web accessibility knowledge and the use of QA tools leads to better results than only using QA tools	Rejected	- The treatment group produced significantly more appropriate text alternatives than the control group when aDesigner was used individually. - When aDesigner was after Acrolinx (QA Scenario B), the control group obtained better (yet not significant) results in terms of alt texts' appropriateness than the treatment group.
	H1.2	If combined with Acrolinx, web accessibility knowledge leads to better results than only using Acrolinx	Rejected	-The positive impact associated with the use of Acrolinx alone was higher than the positive impact of having received basic training on web accessibility. - The differences between groups as regards alt texts scores from T2-Acrolinx was not significant.

Chapter 7

Experimental study: additional findings

In this chapter, we continue the discussion about the results of the experimental study introduced in Chapter 5. Specifically, we complement the findings reported so far in Chapter 6 with regard to the appropriateness of the text alternatives produced localisers, and we present and comment on the subjective data gathered through the post-task questionnaires.

7.1 Overview

Throughout the previous chapter (6), we have measured the effect of WA knowledge (**IV1**), the use of accessibility-oriented QA tools (**IV2**) and the order in which these were used (QA scenario) —one of the seven secondary IVs— on our two main independent variables: the translation of image text alternatives (**DV1**) and their appropriateness (**DV2**). The analyses conducted have yielded sufficient evidence to reject or accept the hypotheses and sub-hypotheses formulated with regard to the research questions associated with our experimental study, stated in Chapter 5 (section 5.4). Nonetheless, we believe that examination of the other six secondary independent variables that we have presented in Table 6.1 (see Chapter 6, section 6.1.2), as well as of the information collected about the localisation and the quality assurance processes by means of the post-task questionnaires (see Chapter 5, section 5.2.4.1) can be of added-value for our work and contribute to a more refined understanding of our main research findings.

In the present chapter, we particularly seek to complement and expand the conclusions we have already drawn concerning the alt text quality of the text alternatives produced by localisers.[1] To this end, we will quantify the impact that four secondary explanatory variables related to our experimental design and another two secondary variables linked to the profile of the participants had on the appropriateness scores assigned by the seven blind users to image alt texts (section 7.2). The second part of this chapter is dedicated to explore localisers' feedback about the tasks they carried out, the tools they used and their overall opinions about web accessibility implementation and evaluation in the context of

[1] We decided to focus on this particular DV, and not on the translation of alt texts, since we believe (i) it is more relevant for continuing the discussion about image accessibility, (ii) the amount of observations gathered for this variable is considerably higher and (iii) the nature of the data (ordinal) lends itself particularly well to the application of inferential statistics.

multilingual websites (section 7.3). These additional findings will serve us to reach more informed conclusions about our investigation.

7.2 Study of other secondary independent variables

This section examines whether the following variables influenced or not the level of image accessibility achieved in the localised websites delivered by the participants of our study:

a) Alt text type

This variable, which will be explored in section 7.2.1, is based on the nature of the 130 image text alternatives included in the source website. The list of these alt texts can be found in Appendix B.1 and their distribution per web page is depicted in Table 5.5. (Chapter 5). To facilitate the analysis of this variable, we divided it into four categories:

- *Inappropriate alt texts*: Out of the 130 image text alternatives, 100 were not appropriate. These alt texts, of at least one-character length, were introduced in the experimental website in order to assess the impact of the 10 CL rules selected for evaluation, shown in Table 4.6 (Chapter 4).
- *empty* `alt` *attributes:* A total of 10 images in the source website contained null or empty `alt` attributes (i.e. `alt=""`).
- *no* `alt` *attributes:* A total of 10 images in the source website had no `alt` attributes and were thus assumed to be totally inaccessible.
- *appropriate alt texts:* Based on our expertise on the topic, a total of 10 images in the source website contained alt texts of at least one-character length that were appropriate, according to the image communicative value.

b) HTML knowledge

As explained in Chapter 5 (see section 5.2.2.2), participants' HTML skills prior to the experimental study were measured through their self-rated knowledge of this markup language (question 28, Appendix D.2), and a short HTML localisation test administered during the selection process (question 29, Appendix D.2). These two different variables will be jointly studied in section 7.2.2.

c) Use of CAT tools

With a view to increasing the ecological validity of our study, we allowed participants to use the CAT tool of their choice or to localise the website without the help of any translation technology. This variable, which is based on the participants' reported use of such tools, will be examined in section 7.2.3.

d) Check of Acrolinx rule help

As we explained in Chapter 4 (section 4.5.2), Acrolinx does not only provide the user with a test report containing the errors found, but also with an explanation of the issues flagged. Within the context of this thesis, the rule help was designed to offer information about the communicative value of the images that the alt texts flagged were associated with, as well as related examples of appropriate and non-appropriate text alternatives. This variable was analysed taking into account the frequency with which localisers reported to have consulted the rule help: (i) never, (ii) for 25% of the errors, (iii) for 50% of the errors, (iv) for 75% of the errors, or (v) for every error flagged by the tool.

e) Check of aDesigner error descriptions

When a website is checked for accessibility issues with aDesigner, a detailed report is returned to the user. When image accessibility issues are found, a description of the problem is provided (see, for instance, Figure 5.3 in Chapter 5). We have also analysed whether consulting those descriptions was determinant for localisers to achieve high quality alt texts. The variable was operationalised as the previous one, considering how often error descriptions were read by localisers: (i) never, (ii) for 25% of the errors, (iii) for 50% of the errors, (iv) for 75% of the errors, or (v) for every error flagged by the tool.

The results we present in this section are based on the information contained in databases DB2 and DB3 (see Chapter 6, section 6.1.1). During data analysis, both descriptive and inferential statistics were used. The investigation on the impact of the alt text type was carried out applying regression analysis methods, which will be explained in the corresponding section prior to the interpretation of the results. To statistically measure the individual effect of the other secondary variables, we performed a repeated measures one-way ANOVA per variable, with three random factors (judge, alt text type and localiser). To this end, we included seven fixed factors (group, translation condition, and the five secondary variables) in the linear mixed model (LMM) fitted for the study of the two primary IVs and their interaction with the QA scenario followed (Chapter 6, sections 6.3.3.1 and 6.3.3.2), A Wald test (Engle 1984) was carried out to measure the significance of the results yielded by each ANOVA.

7.2.1 Alt text type

This section aims at investigating whether the appropriateness of localised text alternatives was influenced by the alt text type included in the source website. The translation conditions that will be considered during the analysis of this variable vary depending on each alt text category, as per the descriptions offered at the beginning of section 7.2. Additionally, it should be noted that results presented are based on data from all 28 participants.

In what follows, we first look at the level of appropriateness achieved in alt texts which initially were not appropriate for the images they were associated with (section 7.2.1.1). Then, we observe the scores gathered for images which originally had an empty `alt` attribute (section 7.2.1.2) and no `alt` attribute (section 7.2.1.3). The analysis of the variable 'alt text type' is completed with the observation of the scores gathered for images which contained an appropriate alt text in the source website (section 7.2.1.4). Given the high volume of data analysed for this particular secondary independent variable, a separate section (7.2.1.5) will be dedicated to cover the conclusions about its overall impact on alt text quality.

7.2.1.1 Inappropriate alt texts

Measuring the effect of this particular alt text category (*inappropriate alt texts*) on the appropriateness of the text alternatives produced by localisers will serve us to assess the impact of applying 10 of the controlled language (CL) rules we developed within the framework of this thesis for alt text quality improvement in French (see Chapter 4). Among the 10 rules selected, three aimed at identifying inappropriate descriptive content (DDA1-credits, DDR2-image, DDR3-logo), five aimed at detecting inappropriate functional content (FAA1-files, FAA3-social, FAA4-print, FAC1-accueil, FDA3-structure), and two sought to identify uninformative content within the alt text (U2-placeholder, U2-decor). Descriptions and examples of these rules can be found in Tables 4.2, 4.3 and 4.4 respectively (section 4.4, Chapter 4).

While it would have been interesting to explore, for instance, whether the application of certain rules was more effective for the treatment group than for the control group (e.g. in the case of more complex rules where having WA knowledge may have helped with the interpretation of Acrolinx's feedback) or viceversa (e.g. in the case of rules covering rather basic image accessibility errors that the treatment group was already aware of), we focused on the examination of each rule impact as a whole. The rationale behind this decision was that we wanted to be able to draw general conclusions about the usefulness of the alt text writing recommendations we formalised through the CL rules. This would allow us to confirm, with empirical evidence, whether some of the best practices with regard to alt text formulation already highlighted in the literature (see section 3.3.2, Chapter 3) are indeed effective, and whether the new ones that we have put forward can be safely added to that list of recommendations on how to write appropriate text alternatives.

Given the fact that the application of the CL rules was only possible through Acrolinx, we narrowed the scope of our analysis to the following translation condition pairwise comparisons: (i) *Acrolinx (T2) - none (T1)*, and (ii)

Both-last-Acrolinx (T3) - aDesigner (T2). For the purposes of this study, we assumed that the differences observed in the scores obtained across the different translation conditions were triggered by the application of the said rules, as per their description in Acrolinx's check reports and rule help. For instance, if in the comparison Acrolinx (T2) - none (T1), the alt texts of a given image that initially contravened one of the CL rules received a higher score under condition *Acrolinx* than under condition *none*, we inferred that this improvement was motivated by the application of that rule.

Acrolinx (T2) - none (T1)

In Figure 7.1[2] we compare the proportion of scores per rule obtained for text alternatives which were produced without any automated QA support (translation condition *none*) and those from alt texts edited after using Acrolinx in the first place, before aDesigner. Figure 7.2 shows the mean scores (y-axis) per rule (x-axis). The red line corresponds to scores collected for the first translation version (T1) and the blue line represents the mean of the scores given to the controlled image text equivalents. Table K.8 (see Appendix K.2) offers a comprehensive summary of the scores distribution, mean and standard deviation per rule for the translation conditions under analysis.

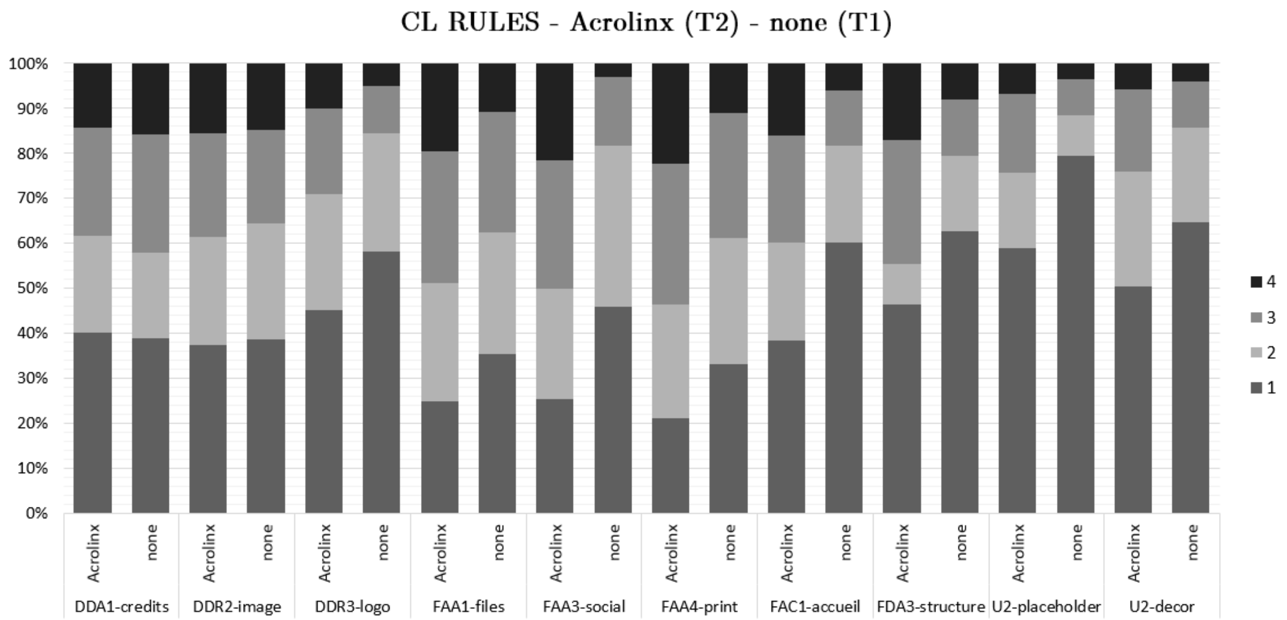


Figure 7.1. Distribution of scores per rule by translation condition: T2 (Acrolinx) - T1 (none)

[2] To better interpret the gray scale colour coding used both in this figure and in Figures 7.4, 7.13, 7.15, 7.16 and 7.17, notice that score 1 (not appropriate) is always at the bottom of the graphic (i.e. lowest part of the y-axis) and score 4 (very pertinent) is always at the top.

As can be readily seen when looking at Figure 7.1, the proportion of score 1 (*Not appropriate*) decreases considerably from T1 to T2 when the following eight rules are applied: DDR3-logo (T1, 58.27%; T2, 45.31%), FAA1-files (T1, 35.41%; T2, 24.90%), FAA3-social (T1, 45.92%; T2, 25.41%), FAA4-print (T1, 33.06%; T2, 21.12%), FAC1-accueil (T1, 60.31%; T2, 38.37%), FDA3-structure (T1, 62.76%; T2, 46.33%), U2-placeholder (T1, 79.59%; T2, 58.98%), and U2-decor (T1, 64.59%; T2, 50.41%). It is also worth noting that the proportion of pertinent and very pertinent alt texts (scores 3 and 4) increases when these rules are followed. Concretely, it seems that the rules aiming at better describing the action triggered by a functional image (FAA1-file, FAA3-social and FAA4-print) led localisers to offer an appropriate text alternative for nearly 75% of all the images.

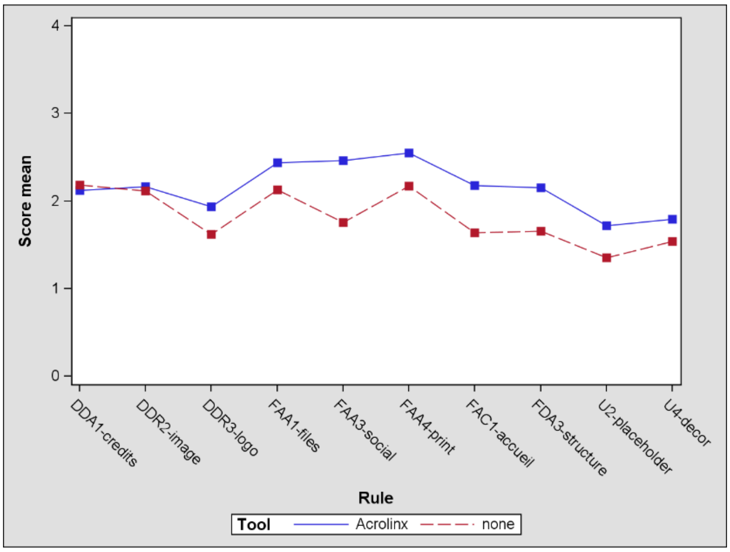


Figure 7.2. Comparison of mean scores per rule by translation condition: Acrolinx (T2) - none (T1)

Conversely, it appears that the rules suggesting the deletion of content within the text alternative for adequacy or redundancy reasons (DDA1-credits and DDR2-image) did not have a noteworthy impact on the scores obtained under translation condition *Acrolinx* with respect to those from *none*. Alt texts describing structural and decorative images (rules FDA3-structure and U2-decor) seem to have been particularly problematic for localisers. On average, and regardless of the translation version, the proportion of non-appropriate text alternatives in these cases was equal to or higher than the proportion of those annotated with positive scores (2-3-4). We hypothesise that this might be due to the subjectivity that regularly implies deciding whether an image has or not a decorative or presentational function, and whether an empty `alt` attribute would be or not more suitable than a full-phrase alt text. Images with text alternatives contravening the rule U2-placeholder, which a priori should have been easily identified by localisers as inappropriate, appear to have been regarded by blind users as the most inaccessible ones. Still, Acrolinx support (either through automatic detection of errors or through informative guidance) helped localisers to reduce the proportion of non appropriate alt texts under this category from 79.59% to 58.98%.

The differences between mean scores per translation condition depicted in Figure 7.2 support these observations. Notice that, with the exception of rules DDR3-logo ($\bar{x}$=1.94), U2-placeholder ($\bar{x}$=1.72), and U2-decor ($\bar{x}$=1.79), the average score per rule in T2 (Acrolinx) is always above 2 (alt text of acceptable quality), as opposed to results from T1 (none), where mean scores are below 2 in six out of the ten rules —with the exception of DDA1-credits ($\bar{x}$=2.19), DDR2-image ($\bar{x}$=2.12), FAAA1-files ($\bar{x}$=2.13), and FAA4-print ($\bar{x}$=2.17). In Figure 7.3, which is composed of ten comparative line charts with the same colour coding used in Figure 7.2, differences between mean scores across rules are presented at an image level. The corresponding data can be found in Tables K.9 to K.18 (Appendix K.2). In this same appendix, bar charts illustrating the distribution of scores per image are also included (Figures K.5 to K.14).

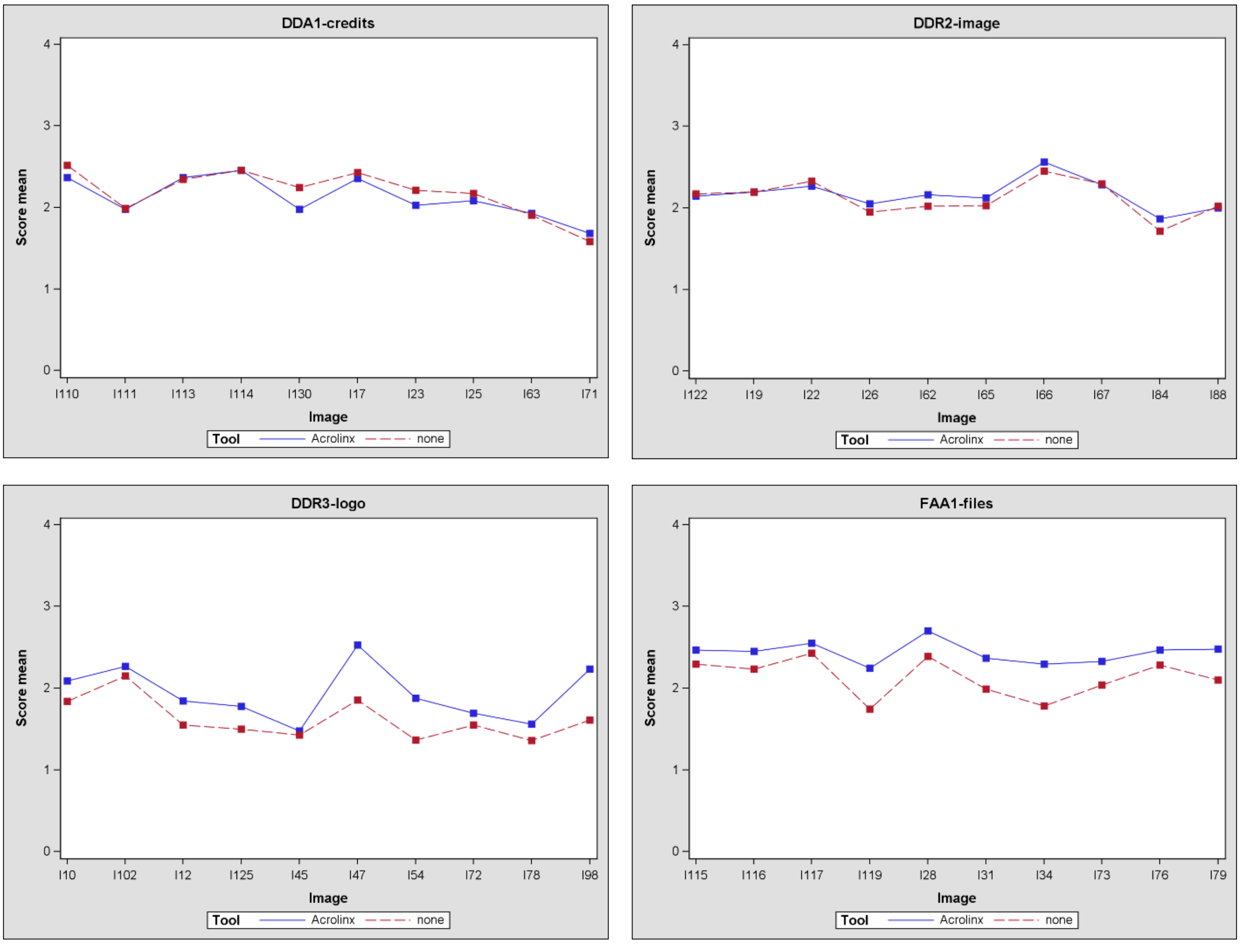


Figure 7.3. Comparison of mean scores per image by rule: Acrolinx (T2) - none (T1)

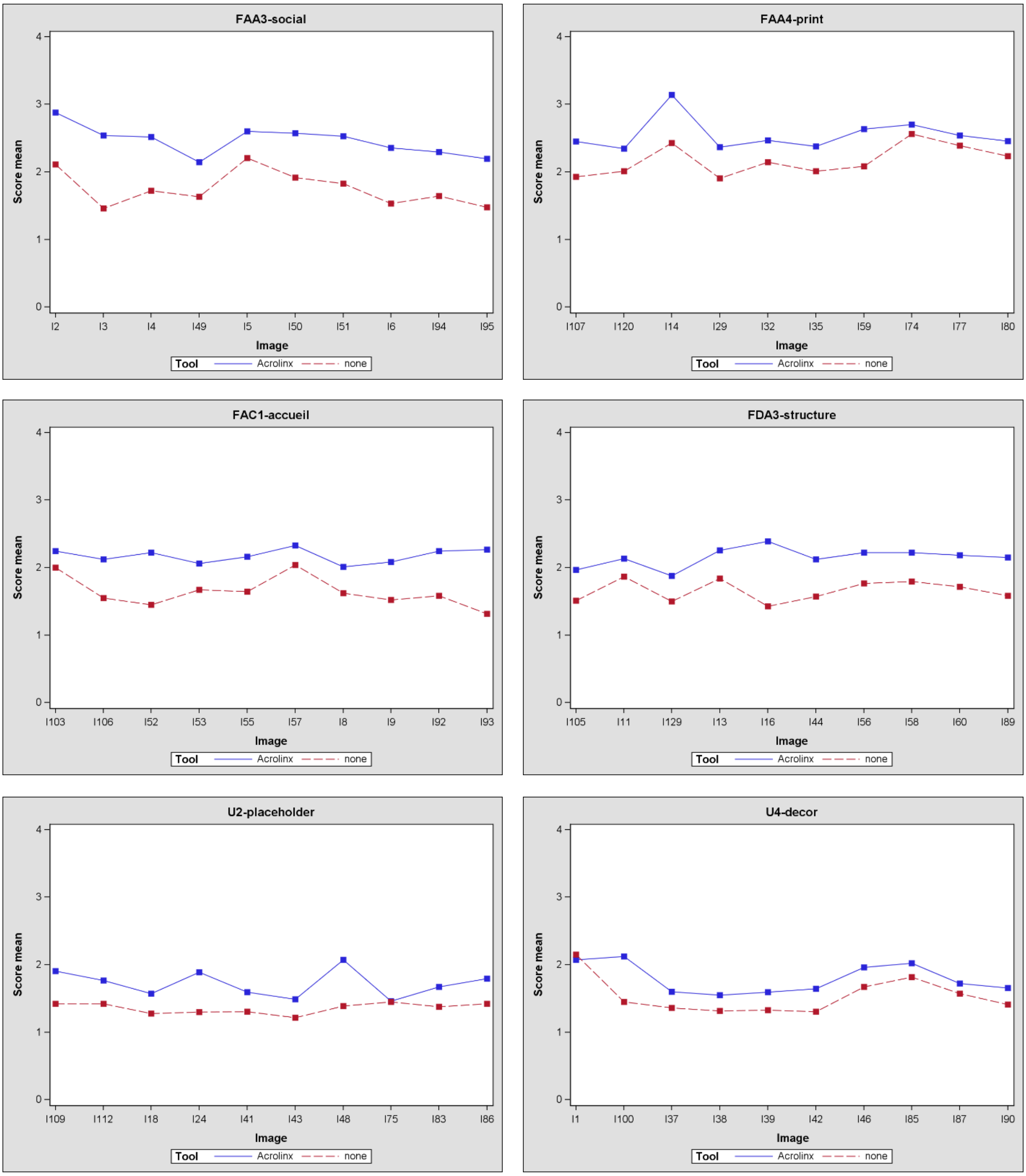


Figure 7.3 (continued)

The highest mean score within a rule image group was always registered for text alternatives produced during T2 (Acrolinx), with two exceptions: rules DDA1-credits and U4-decor, where the highest mean score was obtained for alt texts generated in T1 (none): $\bar{x}$=2.52 (image I110) and $\bar{x}$=2.15 (image I1), respectively. The clearest improvement pattern across the different images per rule is observed in four out of the five rules aimed at improving the alt text quality of functional images (FAA1-files, FAA3-social, FAA4-print and FAC1-accueil), were text alternatives produced after using Acrolinx obtained a mean score higher than 2 for all images. This regular pattern is also noticeable within the image group of rule FDA3-structure, but the average score for alt texts produced in T2 was slightly lower than 2 in images I105 ($\bar{x}$=1.97) and I129 ($\bar{x}$=1.88). It is also interesting to note that at least two images per rule registered an alt text average score lower than 2 in T1, when no QA tool was proposed.

Although Figures 7.1, 7.2 and 7.3 certainly illustrate that improvements were achieved after the application of almost every rule —with the exception of DDR2-image and DDA1-credits—, a second analysis was conducted to measure if the observed score differences between translation conditions were statistically significant. In order to account for the non-normality of the data distribution, given by the ordinal-based evaluation metric defined for the study, a multinomial regression model was used, taking into account three random factors (judge, alt text type and translator). Since the estimates of Generalised Linear Models (GLM) with hierarchical data can be complex to interpret, results are expressed in Table 7.1 in terms of odds ratio (OR)[3] (Hardy and Bryman 2009, 267).

Table 7.1. [Multinomial regression] Odds Ratio (OR) estimates per rule for scores comparison between translation conditions: Acrolinx (T2) > none (T1)

Rule	**Odds Ratio**	**Lower OR**	**Upper OR**
DDA1-credits	0.921	0.800	1.061
DDR2-image	1.068	0.929	1.228
DDR3-logo	*2.719*	2.331	3.171
FAA1-files	*2.150*	1.869	2.473
FAA3-social	-	-	-
FAA4-print	*2.134*	1.855	2.455
FAC1-accueil	*3.921*	3.376	4.553
FDA3-structure	*3.413*	2.932	3.974
U2-placeholder	-	-	-
U4-decor	*2.825*	2.413	3.308

[3] The OR offers a way to quantify the size of significant correlations, and indicates how the likelihood of one variable level varies in response to changes in another variable (Gries and Wulff 2012).

The ORs will allow us to observe if an improvement in the alt text score is more likely to happen when a given rule is applied than when localisers overlook it. It is generally understood that, if the OR estimate is higher than 1 and the confidence interval (delimited by the lower and upper ORs) does not contain the value 1, we can infer that the difference between scores of raw and controlled text alternatives is statistically significant. The analysis provides strong evidence that there is indeed a significant difference between the scores obtained for text alternatives associated with six of the ten rules after using Acrolinx and those of alt texts produced when no tool was used (see rules with OR in italics, Table 7.1). In addition, results indicate that the odds of obtaining a high score when text alternatives are amended according to these rules are, on average, almost three times greater than the odds of obtaining a high score when translators do not use the CL checker. In the particular case of rules FAA3-social and U2-placeholder, we have observed that the model did not return any conclusive results because none of the evaluators had assigned a score of 4 (Very pertinent) to any text alternative of the first translation condition associated with images I3 and I18. In an attempt to adapt the analysis to the characteristics of the data sample, we fit the GLM assuming data distribution to be binomial. Hence, to estimate the effectiveness of the ten rules, we associated score 1 (negative value) to level 0 and scores 2-3-4 (positive value) to level 1 and applied a binomial logistic regression model.

Based on the estimates obtained from this second model, we first predicted the average proportion[4] of positive scores (2-3-4) that could be achieved if rules were or were not to be applied (see Table 7.2). The highest differences across translation versions are observed in the case of rules FAC1-accueil and U2-placehoder. Data indicates that the probability of obtaining a score higher than 1 (not appropriate alt text) when applying these rules is estimated at 62.11% and 40.95% respectively, as opposed to 31.91% and 11.33% when localisers do not follow them. Similarly, it is worth mentioning that the probability of obtaining a positive score is also lower than 50% if rules FDA3-structure, U4-decor and DDR3-logo were not to be applied. Finally, data shown in Table 7.2 suggests that the highest probability of producing appropriate text alternatives is achieved when rules FAA1-files, FAA3-social and FAA4-print are followed, and it is estimated at 75.92%, 76.38% and 79.79% respectively.

In order to understand whether the evaluation results per rule were significantly different in the pairwise comparison *Acrolinx (T2) > none (T1)*, we also applied a Tukey's HSD correction (Tukey 1949), as we had done in previous analyses. The multiple comparisons of means of the scores gathered per rule in each of the aforementioned translation conditions (see Table 7.3) corroborate the results of the

[4] These average proportions (shown under the column *Mean* in Table 7.2) can be interpreted as the estimated probability of producing an appropriate image text alternative when rules are followed.

multinomial logistic regression analysis and indicate that the impact of applying all rules but DDA1-credits and DDR2-image on the alt text quality is highly significant (p<0.001). The Tukey's test estimates can also help us understand whether any of the rules contributed to reducing the appropriateness of the localised alt texts instead of boosting it. In this sense, two interesting observations emerge from the results obtained for rules DDA1-credits and DDR2-image. While the impact of applying the latter is not significant, data suggests that there was a slight improvement from T1 to T2, estimated at 0.0290 (see Table 7.3). However, in the case of the former, the Tukey's test returns a negative estimate (-0.0266), which means that the application of this rule was counterproductive. Significance results can be also expressed in terms of OR (see Table 7.4) and should be interpreted as explained before. The OR estimates confirm that the three rules which demonstrated having the highest impact on the alt text's appropriateness are FAA3-social, FAC1-accueil and U2-placeholder, since the odds of obtaining a high score when text alternatives are amended according to these rules are, on average, almost four times greater than the odds of obtaining a high score when they are not followed (see Table 7.2, highest ODs shown in bold).

Table 7.2. Estimated average proportion of positive scores (2-4) per rule for translation conditions Acrolinx (T2) and none (T1)

Rule	Translation Condition	Mean	Std. Error Mean	Lower Mean	Upper Mean
DDA1-credits	Acrolinx	0.6004	0.0374	0.5253	0.6711
	none	0.6068	0.0355	0.5354	0.6738
DDR2-image	Acrolinx	0.6287	0.0316	0.5650	0.6881
	none	0.6219	0.0298	0.5620	0.6782
DDR3-logo	Acrolinx	0.5567	0.0750	0.4090	0.6950
	none	0.3360	0.0670	0.2193	0.4769
FAA1-files	Acrolinx	0.7592	0.0299	0.6959	0.8129
	none	0.6017	0.0366	0.5283	0.6707
FAA3-social	Acrolinx	0.7638	0.0407	0.6752	0.8342
	none	0.5415	0.0540	0.4354	0.6440
FAA4-print	Acrolinx	0.7979	0.0275	0.7386	0.8465
	none	0.6528	0.0359	0.5796	0.7195
FAC1-accueil	Acrolinx	0.6211	0.0375	0.5453	0.6913
	none	0.3191	0.0332	0.2579	0.3873
FDA3-structure	Acrolinx	0.5365	0.0499	0.4386	0.6317
	none	0.3014	0.0412	0.2271	0.3877
U2-placeholder	Acrolinx	0.4095	0.0235	0.3644	0.4562
	none	0.1133	0.0102	0.0948	0.1347
U4-decor	Acrolinx	0.5012	0.0659	0.3747	0.6275
	none	0.2737	0.0517	0.1845	0.3855

Table 7.3. Tukey's test results per rule for scores comparison between translation conditions: Acrolinx (T2) > none (T1)

Rule	Estimate	Std. Error	*p*-value*
DDA1-credits	-0.0266	0.0815	0.998
DDR2-image	0.0290	0.0819	0.997
DDR3-logo	0.9090	0.0872	*<0.001*
FAA1-files	0.7360	0.0888	*<0.001*
FAA3-social	1.0071	0.0891	*<0.001*
FAA4-print	0.7419	0.0931	*<0.001*
FAC1-accueil	1.2520	0.0834	*<0.001*
FDA3-structure	0.9870	0.0834	*<0.001*
U2-placeholder	1.6919	0.0965	*<0.001*
U4-decor	0.9809	0.0868	*<0.001*

* Significant p-values shown in italics

Table 7.4. [Binomial regression] Odds Ratio (OR) estimates per rule for scores comparison between translation conditions: Acrolinx (T2) > none (T1)

Rule	Odds Ratio	Lower OR	Upper OR
DDA1-credits	0.974	0.780	1.216
DDR2-image	1.029	0.823	1.287
DDR3-logo	*2.482*	1.957	3.148
FAA1-files	*2.088*	1.638	2.660
FAA3-social	***2.738***	2.147	3.491
FAA4-print	*2.100*	1.629	2.707
FAC1-accueil	***3.497***	2.785	4.391
FDA3-structure	*2.683*	2.137	3.369
U2-placeholder	***5.430***	4.173	7.066
U4-decor	*2.667*	2.104	3.379

Overall, the descriptive and inferential statistical analyses performed reveal that, if respected, most of the rules we have defined to be automated through Acrolinx lead to a significant quality improvement of the text alternatives suggested in the target language website. One of the rules which seems to have no significant impact on the alt text appropriateness according to the evaluators' subjective judgement is avoiding the use of redundant expressions such as “Image of” or “Image which shows...” Interestingly, as we have seen Chapter 3 (section 3.3.2.1) this is one of the most recurrent best practices for alt text editing in the literature (Craven 2006; Tercedor Sánchez and Prieto Velasco 2009; Tang 2012). We could hypothesise that this might be derived from the evaluation context itself, where no image was included to accompany the text and, thus, redundancy could not be fully sensed by screen reader users.[5] Still, we believe that the latter should not pose a threat to our study's validity since, being highly experienced users of screen readers and the web, participants were expected to be fully aware of how the alt text would read in a real case scenario.

Additionally, it appears that rules aimed at enhancing the quality of alt texts of functional images are particularly effective. We believe that improvements in these type of text alternatives were seen by blind users as crucial for a better human-computer interaction. Finally, while the analysis of the check reports would be needed to verify whether rules were, in fact, not correctly applied by the tool (e.g. due to a high number of false negatives), or whether localisers simply ignored Acrolinx feedback, the negative (although not significant) impact observed in rule DDA1-credits reflects the continuing uncertainty about where to place image copyright information, especially after the introduction of new HTML 5 elements `<figure>` and `<figcaption>`.[6] Results from the pairwise comparison Acrolinx (T2) - none (T1) seem to indicate that information about the image copyright is not regarded by screen reader users as particularly disturbing or superfluous.

T3 (Both-last-Acrolinx) vs. T2 (aDesigner)

In this section, we present the rule evaluation results taking into account the alt text scores gathered for translation conditions *Both-last-Acrolinx (T3) - aDesigner (T2)*; that is, we measure the impact of applying the ten rules by means of Acrolinx after running checks with aDesigner (QA Scenario A). The comparative bar chart in Figure 7.4 shows the proportion of scores per rule, by translation condition. Figure 7.5 illustrates the alt text mean scores (y-axis) per rule (x-axis). As in previous double line charts, the blue line represents the mean of the scores given to controlled image text alternatives. The red line corresponds to the average scores of alt texts produced in T2 when aDesigner was the first QA tool proposed to localisers. Table K.19 (see Appendix K.2) offers a comprehensive summary of the scores distribution, mean and standard deviation per rule for the translation conditions under analysis.

Data collected suggests that the eight rules that proved to have a considerable impact in the pairwise comparison *Acrolinx (T2) - none (T1)* also showed a noticeable degree of effectiveness when applied after having run QA checks with aDesigner. While improvements in terms of alt text appropriateness were achieved thanks to these rules, score 1 (Not appropriate) proportions are relatively higher

[5] See section 3.3.1.1 (Chapter 3) for an explanation of how images are rendered to visually-impaired users.

[6] See discussion thread in WebAim mailing list between the researcher and Steve Faulkner, co-editor of McCarron et al. (2015) : http://webaim.org/discussion/mail_message?id=26930 Accessed September 29, 2015.

than the ones presented in the previous section, when Acrolinx was used in the first place: DDR3-logo (T2, 65.41%; T3, 58.06%), FAA1-files (T2, 43.27%; T3, 39.29%), FAA3-social (T2, 45.92%; T3, 31.43%), FAA4-print (T2, 33.86%; T3, 25.82%), FAC1-accueil (T2, 71.12%; T3, 61.33%), FDA3-structure (T2, 72.96%; T3, 55.51%), U2-placeholder (T2, 87.65%; T3, 72.65%), and U2-decor (T2, 71.73%; T3, 59.49%).

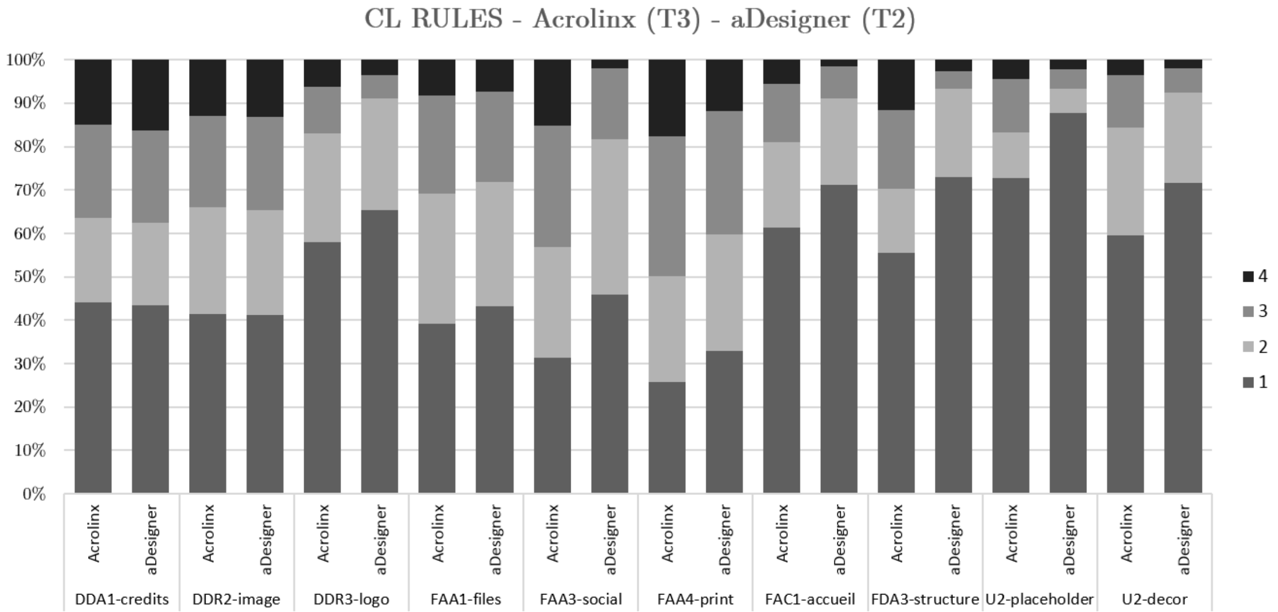


Figure 7.4. Distribution of scores per rule by translation condition: T3 (Both-last-Acrolinx) - T2 (aDesigner)

In the same vein, it should be mentioned that the differences between mean scores per rule was higher in the pairwise comparison Acrolinx (T2) - none (T1) than in the one under analysis (see Figure 7.5). Concretely, attention should be paid to rules FAA1-files (T2, $\bar{x}$=1.92; T3, $\bar{x}$=2.00) and FAC1-accueil (T2, $\bar{x}$=1.29; T3, $\bar{x}$=1.63), which demonstrated having a substantially higher impact when no other tool had been previously used. As far as the proportion of scores 3 and 4 (*Pertinent* and *Very pertinent*) is concerned, data shows that alt texts from T3 were more positively annotated than alt texts from T2 in the aforementioned eight rules. Results regarding the remaining two ─DDA1-credits and DDR2-image─ are also in line with those already reported earlier during the comparison Acrolinx (T2) - none (T1), and score proportions are almost identical across translation conditions. As shown in Table K.19 and illustrated in Figure 7.5, the average score per rule in T3 (Both-last-Acrolinx) is equal to or above 2 (alt text of acceptable quality) in only three out of the eight more productive rules (FFA1-files, $\bar{x}$=2.00; FAA3-social, $\bar{x}$=2.27; and FAA4-print, $\bar{x}$=2.42). In the same vein, it is also worth noting that, regardless of their degree of effectiveness, all rules led to higher average scores in T2 (Acrolinx) than in T3 (Both-last-Acrolinx). This should not be necessarily seen as a degraded performance level of the tool; it is also likely that using aDesigner in T2 had contributed to improve the quality of the alt texts generated in T1, hence the lower differences between scores.

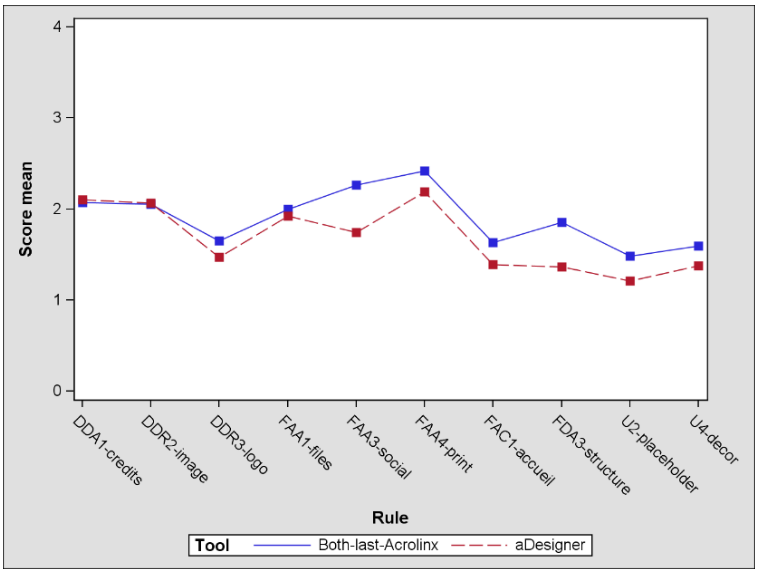


Figure 7.5. Comparison of mean scores per rule by translation condition: Both-last-Acrolinx (T2) - aDesigner (T2)

If we look at the difference between mean scores across images within each rule (see Figure 7.6), the rules that show the clearest and more constant improvement pattern are (i) FAA3-social, where the alt texts per image generated during T3 obtained always an average score higher than 2, with the highest being that of image I14: $\bar{x}$=3.16; and (ii) FDA3-structure, where the highest mean score registered was in image I13: $\bar{x}$=2.18. From the other three rules associated to functional images, FAA1-files appears to have had the lowest impact. While the last two rules (U2-placeholder and U4-decor) triggered improvements, mean scores per image remained under 2.00. A breakdown of score proportions per image, by rule, is provided in Tables K.20 to K.29 and Figures K.15 to K.24 (see Appendix K.2).

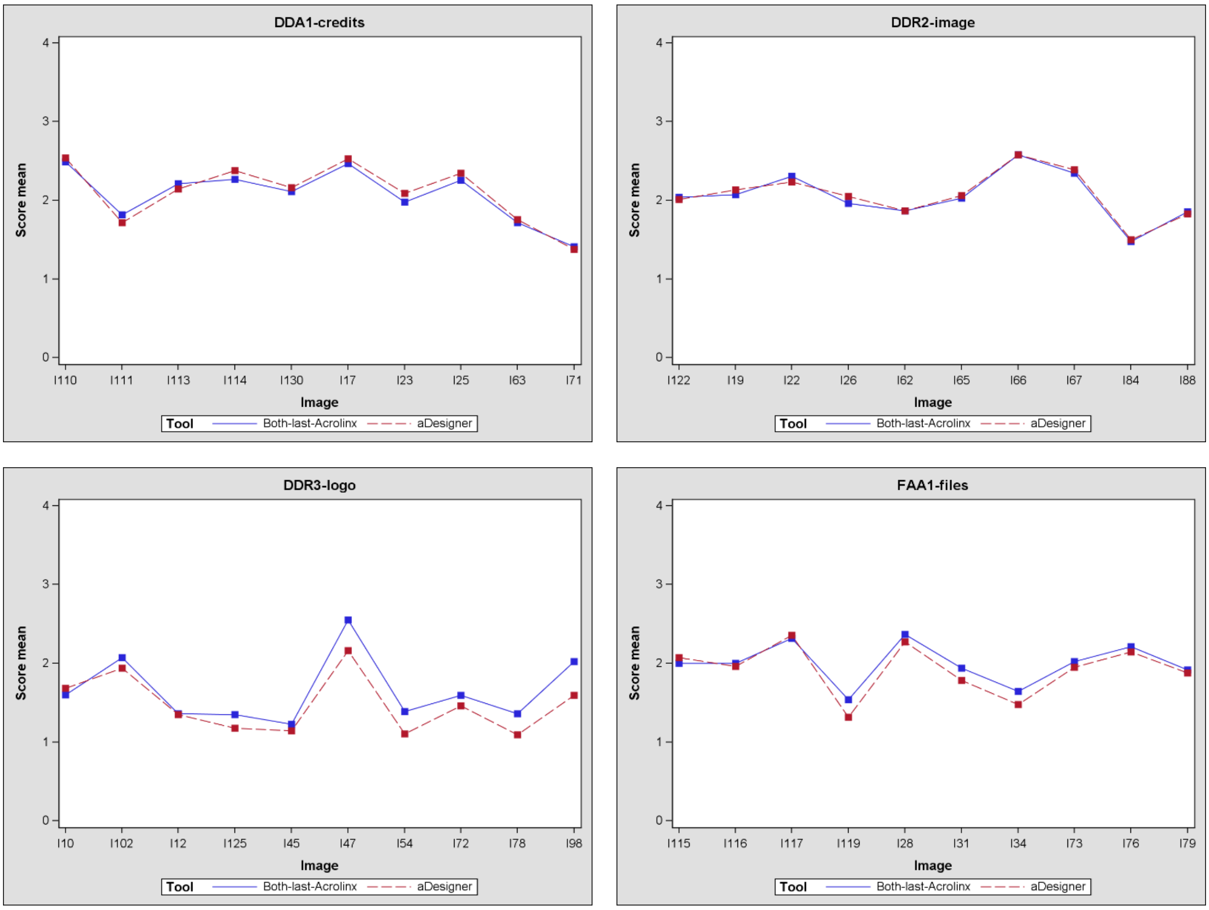


Figure 7.6. Comparison of mean scores per image by rule: Both-last-Acrolinx (T3) - aDesigner (T2)

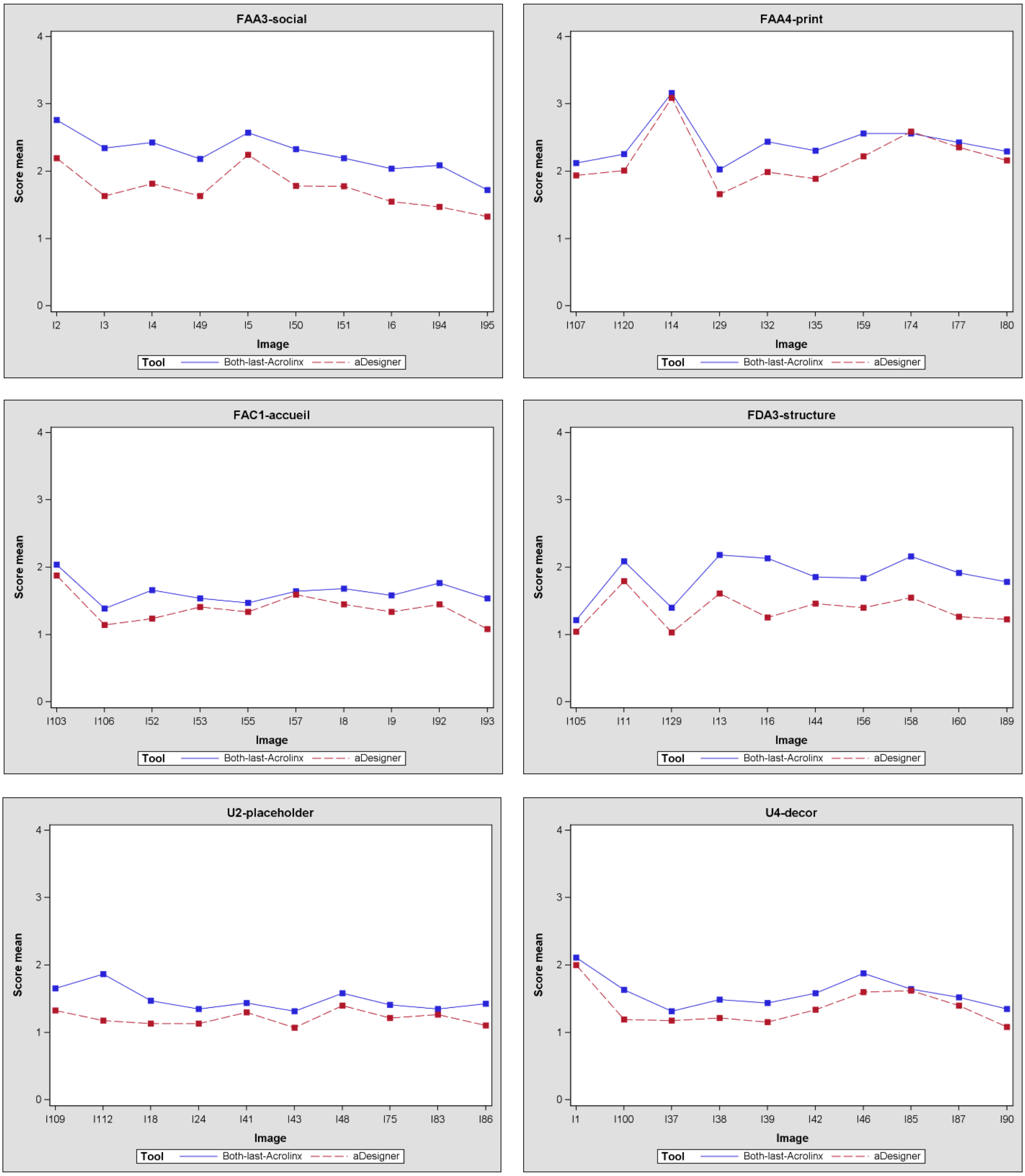


Figure 7.6 (continued)

To estimate whether the impact observed was statistically significant, we followed the same analytical procedure described previously for the pairwise comparison Acrolinx (T2) - none (T1). We modelled the data assuming its distribution to be binomial and performed a binomial logistic regression analysis. Based on the estimates returned by the model, we first computed the estimated average proportion of positive scores that could be achieved if rules were or were not to be applied after using aDesigner (see Table 7.5). The highest differences across translation versions are observed in the case of rules FDA3-structure and FAA3-social. Data indicates that the probability of obtaining a score higher than 1 (not appropriate alt text) when applying these rules after having used aDesigner is estimated at 43.79% and 70.25% respectively, as opposed to 25.61% and 54.87% when localisers do not follow them. It is interesting to notice that, in the case of FDA3-structure, the estimated average proportion of positive scores is below 50% in both translation conditions (when the rule is applied and when the rule is not followed). This can also be also observed in the case of other rules, such as DDR3-logo, U2-placeholder, U4-decor or FAC1-accueil. These results contrast with those presented in Table 7.2, where the probability of producing an appropriate alt text was above 50% if these rules were applied. We understand that this could be due to the nature of the changes made in T2, after receiving the input of aDesigner. It is possible that the patterns of the new alt texts were not covered by the aforementioned CL rules, thus making it less likely to produce higher quality alt texts.

The Tukey's test (Tukey 1949) yielded significant results for seven out of the ten rules under evaluation (see Table 7.5). They indicate that when rules DDR3-logo (p=0.003), FAA3-social (p<0.001), FAA4-print (p=0.003), FAC1-accueil (p<0.001), FDA3-structure (p<0.001), U2-placeholder (p<0.001) and U4-decor (p<0.001) were followed, localisers managed to produce alt texts of significant higher quality than with the help of aDesigner. These results are in agreement with those presented before, when Acrolinx was used in the first place (QA Scenario B), with the exception of the rule FAA1-files, which did not contribute to a significant improvement in the alt texts' appropriateness. Tukey's test estimates also show that the application of rules DDA1-credits and DDR2-image was counterproductive, although their use did not lead to a significant decrease in the alt texts' quality.

To complement the results presented so far, we have computed the OR per rule for the pairwise comparison of translation conditions Both-last-Acrolinx (T3) > aDesigner (T2). The OR estimates provided in Table 7.7 serve to corroborate the significant impact of the aforementioned seven rules. In addition, they indicate that the odds of obtaining a high score when text alternatives are amended according to rules such as FAA3-social, FDA3-structure and U2-placeholder are, on average, almost two times greater than the odds of obtaining a high score when only aDesigner's recommendations are followed (see Table 7.7, highest ODs shown in bold).

Table 7.5. Estimated average proportion of positive scores (2-4) per rule for translation conditions Both-last-Acrolinx (T3) and aDesigner (T2)

Rule	Translation Condition	Mean	Std. Error Mean	Lower Mean	Upper Mean
DDA1-credits	Both-last-Acrolinx	0.5600	0.0384	0.4840	0.6333
	aDesigner	0.5675	0.0382	0.4916	0.6403
DDR2-image	Both-last-Acrolinx	0.5889	0.0326	0.5238	0.6509
	aDesigner	0.5910	0.0325	0.5260	0.6529
DDR3-logo	Both-last-Acrolinx	0.4052	0.0733	0.2729	0.5529
	aDesigner	0.3203	0.0664	0.2059	0.4614
FAA1-files	Both-last-Acrolinx	0.6112	0.0380	0.5348	0.6826
	aDesigner	0.5698	0.0391	0.4921	0.6441
FAA3-social	Both-last-Acrolinx	0.7025	0.0468	0.6035	0.7855
	aDesigner	0.5487	0.0551	0.4401	0.6529
FAA4-print	Both-last-Acrolinx	0.7506	0.0315	0.6840	0.8072
	aDesigner	0.6786	0.0362	0.6040	0.7452
FAC1-accueil	Both-last-Acrolinx	0.3818	0.0376	0.3113	0.4576
	aDesigner	0.2810	0.0326	0.2217	0.3490
FDA3-structure	Both-last-Acrolinx	0.4379	0.0494	0.3446	0.5358
	aDesigner	0.2561	0.0387	0.1877	0.3389
U2-placeholder	Both-last-Acrolinx	0.2718	0.0201	0.2342	0.3130
	aDesigner	0.1219	0.0130	0.09866	0.1496
U4-decor	Both-last-Acrolinx	0.3976	0.0632	0.2823	0.5255
	aDesigner	0.2609	0.0512	0.1733	0.3728

Table 7.6. Tukey's test results per rule for scores comparison between translation conditions: Both-last-Acrolinx (T3) > aDesigner (T2)

Rule	Estimate	Std. Error	*p*-value*
DDA1-credits	-0.0302	0.0929	0.998
DDR2-image	-0.0087	0.0930	1.000
DDR3-logo	0.3684	0.1014	*0.003*
FAA1-files	0.1716	0.0938	0.358
FAA3-social	0.6636	0.0976	*<0.001*
FAA4-print	0.3546	0.1018	*0.004*
FAC1-accueil	0.4574	0.0981	*<0.001*
FDA3-structure	0.8168	0.0992	*<0.001*
U2-placeholder	0.9893	0.1211	*<0.001*
U4-decor	0.6257	0.1030	*<0.001*

* Significant p-values shown in italics

Table 7.7. Odds Ratio (OR) estimates per rule for scores comparison between translation conditions: Both-last-Acrolinx (T3) > aDesigner (T2)

Rule	**Odds Ratio**	**Lower OR**	**Upper OR**
DDA1-credits	0.970	0.753	1.250
DDR2-image	0.991	0.769	1.278
DDR3-logo	1.445	1.096	1.906
FAA1-files	1.187	0.919	1.534
FAA3-social	**1.942**	1.488	2.534
FAA4-print	1.426	1.080	1.882
FAC1-accueil	1.580	1.209	2.065
FDA3-structure	**2.263**	1.726	2.967
U2-placeholder	**2.689**	1.933	3.742
U4-decor	1.870	1.411	2.477

In summary, the rule evaluation results seem to be consistent irrespective of the translation condition under which localisers used Acrolinx (as the first of two QA tools proposed or after having run checks with aDesigner). The analyses performed suggest, on one hand, that the impact of applying the two rules aiming at detecting uninformative alt texts and the five rules covering the identification of inappropriate text alternatives associated with functional images is significant, with the exception of FAA1-files at T3. On the other hand, the evaluation results provide evidence that only one out of the three rules selected for refining text alternatives produced for descriptive images (DDR3-logo) triggered significant improvements in terms of alt text quality. Interestingly enough, the explicitation in the alt text of an image being a logo has been traditionally more controversial than using redundant expressions such as "Image of...", covered in rule DDR2-image (see Chapter 3, section 3.3.2.2). Based on the evidence gathered in our study, it could be stated that, in fact, end users are more disturbed by the former than by the latter. Further conclusions about the rule impact according to the use in which tools were used and the overall impact of these finds for the guidance on how to write appropriate text alternatives for images will be discussed in section 7.2.1.5.

7.2.1.2 Empty `alt` attributes

In this section, we focus on the scores assigned to localised alt texts whose corresponding images originally had originally an empty `alt` attribute. In this sense, it is important to mention that the data presented henceforth does not exclusively cover the annotation of empty alt texts, but rather of all alt texts (of zero-length or more-than-one-character length) proposed by localisers in the different translation versions (T1, T2, T3) for the ten images which, in the English source website, had an empty `alt` attribute (see Appendix B.1, images marked as *alt-empty*). Since both Acrolinx and aDesigner provide indications about when to use empty `alt` attributes, we have considered all five levels of *IV2. Use of accessibility-oriented QA tools* to observe the interaction effect of this independent variable with this particular alt text type on the scores obtained.

Figure 7.7[7] and Table 7.8 illustrate and summarise the distribution of scores per translation condition, taking into account the QA scenario followed (five levels). Although there are not clearly strong differences across translation conditions in terms of score proportions, a closer look at the data reveals that localisers who followed Scenario B (Acrolinx-aDesigner) seem to have taken wiser decisions as regards the use of empty `alt` attributes, since the lowest proportion of score 1 and the highest proportion of score 4 have been achieved in translation condition T3-Both-last-aDesigner (1=29.59%, 4=13.47%; see Table 7.8). Interestingly, the highest proportion of score 1 and the lowest proportion of score 4 were registered for alt texts generated during T2, when aDesigner was used in the first place (1=38.98%, 4=8.37%). We hypothesise that the difference between the high scores obtained in T3-Both-last-aDesigner and T2-aDesigner might be derived from the degree of completeness of the help support provided by the tools. One could believe that aDesigner's explanations of why a null `alt` attribute should be implemented were not clear enough for participants, who may have later better understood the technique when consulting Acrolinx rule help.

In Figure 7.8, we compare the alt text mean scores obtained per image when aDesigner was used at T2 and T3 (see blue line in both charts) against the mean scores achieved per image for text alternatives produced in the corresponding preceding translation versions (T1-none and T2-Acrolinx). A detailed breakdown of score proportions, means and standard deviations per image is presented in Tables K.30 and K.31 and illustrated in Figures K.25 and K.26 (see Appendix K.2). While similar patterns are depicted in both charts, one can observe that alt text mean scores are higher when no tool was used than when localisers had the help of aDesigner (see line chart to the left, Figure 7.8). In contrast, mean scores per image at T3 when aDesigner was used after Acrolinx are almost always greater than those achieved at T2 when Acrolinx was first tool proposed. Mean score variation across images in both pairwise comparisons may be due to the nature of the images themselves.

For instance, images I108 and I61 were transparent images described as such to blind users (see Appendix J); hence the pertinence of associating them to zero-length alt texts. However, the use of an empty `alt` attribute could be seen as debatable in images I27 and I30, which represented PDF icons and, although not

[7] In the case of Figures 7.7, 7.9, 7.11 and 7.14, the gray scale coulour coding goes from left to right, being score 1 (not appropriate) to the extreme left and score 4 (very pertinent) to the extreme right.

having a functional purpose, might have been interpreted by evaluators as images with an informative meaning. A comprehensive qualitative study of these cases would be needed, however, to confirm this hypothesis.

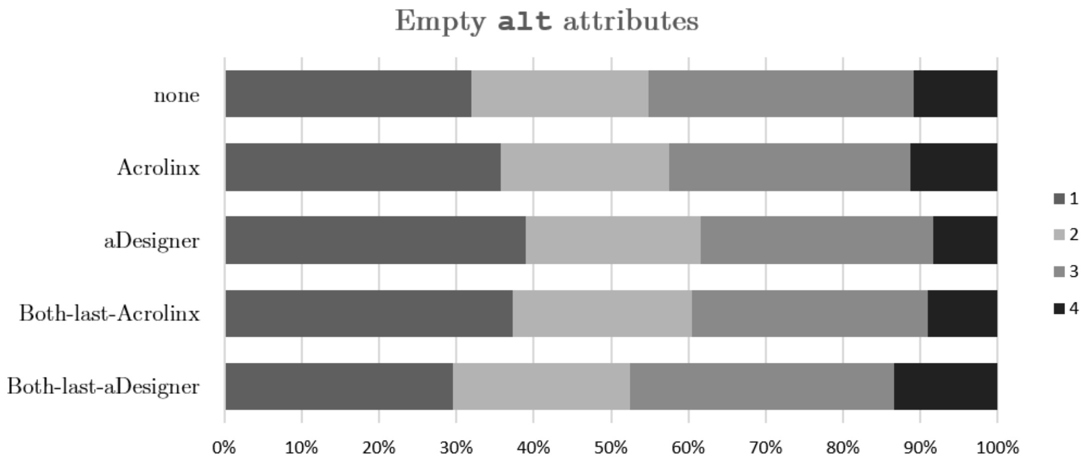


Figure 7.7. Distribution of scores per translation condition (five levels) assigned to images which originally had an empty `alt` attribute

Table 7.8. Score mean, standard deviation and proportions (%) per translation condition (five levels) for images which originally had an empty alt attribute

#P	Version	Tool	Mean	SD	1	%	2	%	3	%	4	%
14	T3	aDesigner	2.31	1.04	290	29.59	224	22.86	334	34.08	132	13.47
14	T3	Acrolinx	2.11	1.01	366	37.35	226	23.06	299	30.51	89	9.08
14	T2	aDesigner	2.08	1.01	382	38.98	221	22.55	295	30.10	82	8.37
14	T2	Acrolinx	2.18	1.04	350	35.71	213	21.73	306	31.22	111	11.33
28	T1	none	2.24	1.02	625	31.89	450	22.96	672	34.29	213	10.87

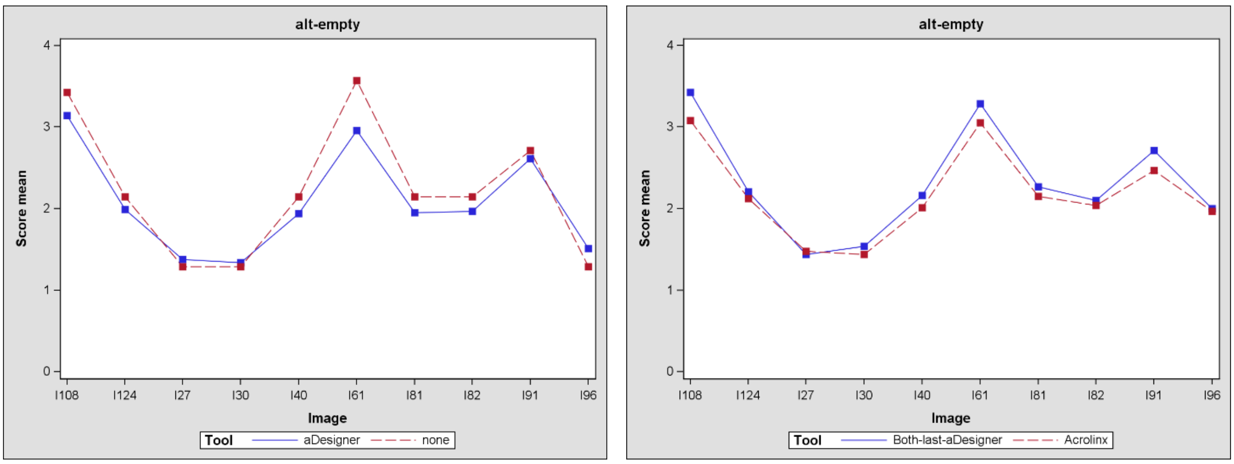


Figure 7.8. Pairwise comparison of mean scores for images which originally had an empty `alt` attribute: aDesigner (T2) vs. none (T1) (left), and Both-last-aDesigner (T3)

With a view to guaranteeing a certain homogeneity as regards the statistical methods chosen to measure the interaction of effect of the alt text type and the use of QA tools on the alt texts scores obtained, we adopted the same inferential analysis approach as for the rule evaluation, presented in the previous section. As shown in Table 7.9, the probability of producing an appropriate alt text in the target website when empty `alt` attributes are present in the source is above 60% in all five translation conditions. Notice that, in comparison to the control condition (T1-none), the estimated average proportion of positive scores (2-4) is only higher if the two QA tools are used in the following order: Acrolinx-aDesigner (77.65%). Nevertheless, the multiple comparisons of means of the scores gathered per translation condition for this alt text type (see Table 7.10) reveal that the difference between the scores obtained at T3-aDesigner (QA scenario B) and those gathered for T1-none is not significant. In fact, the Tukey's test (Tukey 1949) results indicate that the only significant improvement in terms of alt text appropriateness was achieved when using aDesigner to assess T2, which in turn had been previously checked for image accessibility with Acrolinx (p=0.003). This finding is confirmed by the corresponding OR estimate for this translation condition pair (see Table 7.11).

Table 7.9. Estimated average proportion of positive scores (2-4) per translation condition (five levels) for images which originally had an empty `alt` attribute

Translation Condition	Mean	Std. Error Mean	Lower Mean	Upper Mean
Both-last-aDesigner	0.7765	0.0883	0.5617	0.9040
Both-last-Acrolinx	0.6702	0.1123	0.4288	0.8462
aDesigner	0.6461	0.1161	0.4029	0.8317
Acrolinx	0.6938	0.1079	0.4556	0.8598
none	0.7467	0.0955	0.5228	0.8881

Table 7.10. Tukey's test results of scores comparison between translation conditions for images which originally had an empty `alt` attribute

Comparison of translation conditions*	Estimate	Std. Error	*p*-value †
Both-last-aDesigner > Acrolinx	0.4275	0.1197	*0.003*
Acrolinx > aDesigner	0.2159	0.1162	0.341
Both-last-aDesigner > none	0.1642	0.1050	0.520
Both-last-Acrolinx > aDesigner	0.1070	0.1157	0.887
Acrolinx > none	-0.2633	0.1024	0.076
Both-last-Acrolinx > none	-0.3721	0.1019	*0.002*
aDesigner > none	-0.4791	0.1015	*<0.001*
Both-last-Acrolinx > Both-last-aDesigner	-0.5363	0.1194	*<0.001*

*Data organised in descending order, from highest to lowest estimate. / † Significant p-values shown in italics.

The most striking result to emerge from the data under analysis is that the use of either aDesigner or Acrolinx alone leads to a decrease in the alt texts' quality and, in the case of aDesigner, this negative effect is significant (-0.4791, $p<0.001$). Similarly, the significance test provides strong evidence that the participants who followed QA Scenario A obtained significantly lower scores for this alt text type than those who followed QA Scenario B (-0.5363, $p<0.001$). What is more, localisers made significantly wiser translation decisions regarding the use of empty `alt` attributes when no QA automated solution was proposed than when they used aDesigner followed by Acrolinx (-0.3721, $p=0.002$, see Table 7.10).

Table 7.11. Odds Ratio (OR) estimates of scores comparison between translation conditions for images which originally had an empty `alt` attribute

Comparison of translation conditions	Odds Ratio	Lower OR	Upper OR
Both-last-aDesigner > Acrolinx	1.533	1.106	2.126
Acrolinx > aDesigner	1.241	0.988	1.559
Both-last-aDesigner > none	1.178	0.959	1.448
Both-last-Acrolinx > aDesigner	1.113	0.887	1.396
Acrolinx > none	0.769	0.629	0.939
Both-last-Acrolinx > none	0.689	0.564	0.842
aDesigner > none	0.619	0.508	0.756
Both-last-Acrolinx > Both-last-aDesigner	0.585	0.463	0.739

These somehow unexpected results seem to suggest that, overall, the use of tools was not critical to help localisers decide whether or not to transform zero-length alt texts into phrase-based text alternatives. Besides, it appears that this was more challenging for localisers than performing the opposite action. This belief is supported by the observations made in Chapter 6 (see section 6.2.3.5), where we have reported a considerable increase in the amount of empty `alt` attributes across translation versions. In other words, it could be stated that participants seem to have been keener to propose new empty `alt` attributes than reconsidering the appropriateness of the null alt texts already present in the source. This could be attributed to the fact that aDesigner recommends users to consider the use of empty `alt` attributes by default when a potentially inappropriate alt text is found, and some Acrolinx rules (e.g. FDA3-structure, U4-decor) aim at precisely detecting text alternatives associated with images that, in principle, should have a null alt. In addition, it should be noted that the positive findings presented in this section regarding the tool sequence Acrolinx-aDesigner (QA Scenario B) are in accord with those already reported in Chapter 6. However, for this particular alt text type, the use of Acrolinx alone or after aDesigner has demonstrated to have had a negative impact in the final image accessibility level, as opposed to results yielded in previous analyses.

7.2.1.3 Absence of `alt` attributes

The data that will be presented in this section corresponds to the ten images which were originally included in the source website without an `alt` attribute (see Appendix B.1, images marked as *no-alt*). For the purposes of these data analysis and interpretation, we assumed that any increase in the scores assigned to localised text alternatives was due, at least, to the introduction of an `alt` attribute. Figure 7.9 and Table 7.12 provide a general and a detailed overview of the score distribution for the `alt` attribute values under analysis across all five translation conditions. Although at first sight they might seem to show a rather surprising outcome as regards the high proportion of not appropriate text alternatives, we had already foreseen these results (see section 6.2.3.5, Chapter 6) since we observed that only three out of the 28 participants had managed to reduce the number of images without an `alt` attribute down to zero in their last translation version (T3).

Surprisingly enough, the lowest proportion of score 1 has been registered in T2 when Acrolinx was used (1=81.94%), despite the fact that this tool does not detect the absence of the said attribute. This contrasts with the data presented earlier in Chapter 6 (see section 6.2.3.5, Tables 6.10 and 6.11), which suggested that aDesigner had triggered the highest number of insertions of `alt` attributes. We could thus infer that, while the tool prompted the use of `alt` attributes, it did not provide enough guidance on how to produce appropriate alt texts. It seems, however, that Acrolinx covered that gap, since it triggered the highest proportion of pertinent and very pertinent text alternatives (T2-Acrolinx: 3=7.35%, 4=4.80%; see Table 7.12). Finally, it is worth noting that results obtained in the last translation version (T3) seem to be very similar, irrespective of the QA scenario followed.

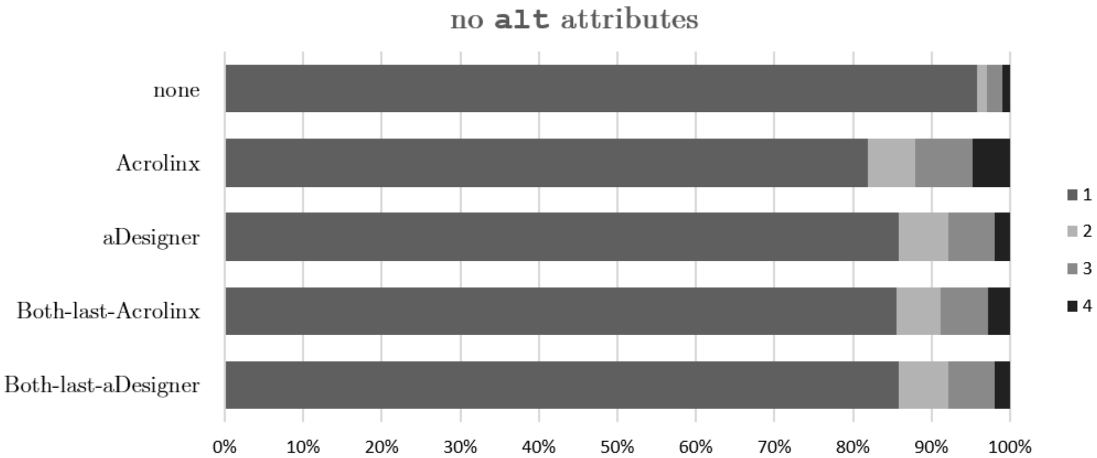
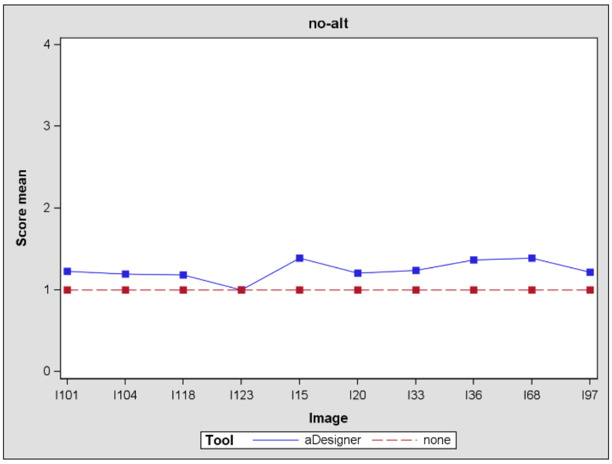


Figure 7.9. Distribution of scores per translation condition (five levels) assigned to images which originally had no `alt` attribute

Table 7.12. Score mean, standard deviation and proportions (%) per translation condition (five levels) for images which originally had no `alt` attribute

#P	Version	Tool	Mean	SD	1	%	2	%	3	%	4	%
14	T3	aDesigner	1.24	0.65	841	85.82	62	6.33	58	5.92	19	1.94
14	T3	Acrolinx	1.26	0.69	838	85.51	55	5.61	60	6.12	27	2.76
14	T2	aDesigner	1.24	0.65	841	85.82	62	6.33	58	5.92	19	1.94
14	T2	Acrolinx	1.35	0.81	803	81.94	58	5.92	72	7.35	47	4.80
28	T1	none	1.08	0.42	1877	95.77	24	1.22	39	1.99	20	1.02

Figure 7.10 shows the average scores obtained per image in translation conditions none, T2-aDesigner, T2-Acrolinx and T3-Both-last-aDesigner. As can be readily seen from both charts, neither the use of aDesigner (see blue line, left chart) nor the use of Acrolinx alone (see red line, right chart) led to a mean score equal to or higher than 2, which we considered as the minimum threshold for the target image to be accessible. Nevertheless, improvements become more noticeable if we look at the average scores obtained when aDesigner was used after Acrolinx (see blue line, right chart). A detailed breakdown of score proportions, means and standard deviations of this alt text type is presented in Tables K.32 and K.33 and illustrated in Figures K.27 and K.28 (see Appendix K.2).



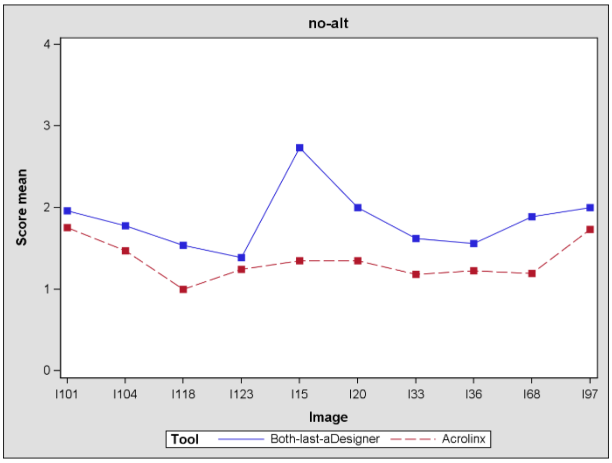


Figure 7.10. Pairwise comparison of mean scores for images which originally had no `alt` attribute: aDesigner (T2) vs. none (T1) (left), and Both-last-aDesigner (T3) vs. Acrolinx (T2) (right)

As we have previously done with data from other alt text types, we used a binomial logistic regression analysis in order to understand the interaction effect of the use of tools and the introduction of images with no `alt` attribute in the source website on the appropriateness of the localised alt texts. Based on the data analysed, the model allowed us to predict that the probability of producing a localised alt text of acceptable quality is lower than 4% when there are images with no `alt` attributes in the source web product and localisers do not count on automated QA support (see Table 7.13). The model also suggests that the estimated average proportion of positive scores increases if localisers use aDesigner (13.35%), aDesigner followed by Acrolinx (13.64%) or Acrolinx alone (17.13%).

Table 7.13. Estimated average proportion of positive scores (2-4) per translation condition (five levels) for images which originally had no alt attribute

Translation Condition	Mean	Std. Error Mean	Lower Mean	Upper Mean
Both-last-aDesigner	0.4409	0.0419	0.3610	0.5240
Both-last-Acrolinx	0.1364	0.0214	0.0995	0.1842
aDesigner	0.1335	0.0211	0.0972	0.1805
Acrolinx	0.1713	0.0253	0.1272	0.2267
none	0.0389	0.0072	0.0270	0.0560

In spite of the relatively low probabilities predicted by the model, the post-hoc analysis provides strong evidence that, in our experimental study, the improvements achieved in terms of alt text quality by localisers when using aDesigner alone, Acrolinx alone, or Acrolinx followed by aDesigner in the case of images which originally had no `alt` attribute were statistically significant (p<0.001, see Table 7.14). In addition, the Tukey's test indicates that, although the difference between scores is not significant, localisers under translation condition T2-Acrolinx obtained higher alt text scores than those under translation condition T2-aDesigner. A significant difference is observed, however, between scores obtained by localisers at T3, depending on the tool used. Those who used aDesigner followed by Acrolinx (QA Scenario A) produced alt texts ─once the alt attribute was introduced─ of significantly lower quality than those who used Acrolinx followed by aDesigner (QA Scenario B) (-1.6078, p<0.001, see Table 7.14). The positive impact of following this last scenario is confirmed by the OR estimates shown in Table 7.15. They indicate that the odds of obtaining a positive score when using first Acrolinx and then aDesigner is almost 20 times greater than when none of them are used. Conversely, the odds are considerably lower if the tools are used in the inverse order (3.895, see Table 7.15, third row).

Taken together, these results suggest that the individual or combined use of accessibility-oriented QA tools can both facilitate the identification of missing `alt` attributes and the production of acceptable text alternatives. Still, better results with regard to image accessibility in the target website can be achieved when a CL checker like Acrolinx is used followed by a general WAE tool as aDesigner. While aDesigner's capability of identifying missing `alt` attributes was initially regarded as its main advantage over Acrolinx, our findings demonstrate that this extra feature did not have a significant impact on the final alt text quality. As we have hypothesised earlier in this chapter, we attribute these results to the guidance offered by the tools to amend the errors flagged. More insights about this issue will be provided in sections 7.2.4 and 7.2.5.

Table 7.14. Tukey's test results of scores comparison between translation conditions for images which originally had no `alt` attribute

Comparison of translation conditions*	**Estimate**	**Std. Error**	***p*-value†**
Both-last-aDesigner > none	2.9675	0.1309	*<.0001*
Acrolinx > none	1.6286	0.1405	*<.0001*
Both-last-Acrolinx > none	1.3597	0.1452	*<.0001*
Both-last-aDesigner > Acrolinx	1.3389	0.1071	*<.0001*
aDesigner > none	1.3343	0.1457	*<.0001*
Acrolinx > aDesigner	0.2943	0.1250	0.018
Both-last-Acrolinx > aDesigner	0.0254	0.1302	0.999
Both-last-Acrolinx > Both-last-aDesigner	-1.6078	0.1132	*<.0001*

*Data organised in descending order, from highest to lowest estimate. / † Significant p-values shown in italics.

Table 7.15. Odds Ratio (OR) estimates of scores comparison between translation conditions for images which originally had no `alt` attribute

Comparison of translation conditions	**Odds Ratio**	**Lower OR**	**Upper OR**
Both-last-aDesigner > none	19.443	13.605	27.787
Acrolinx > none	5.097	3.474	7.478
Both-last-Acrolinx > none	3.895	2.621	5.788
Both-last-aDesigner > Acrolinx	3.815	2.848	5.109
aDesigner > none	3.797	2.552	5.650
Acrolinx > aDesigner	1.342	0.954	1.888
Both-last-Acrolinx > aDesigner	1.026	0.719	1.463
Both-last-Acrolinx > Both-last-aDesigner	0.200	0.147	0.273

7.2.1.4 Appropriate `alt` attributes

Throughout the three previous sections, we have already covered the analysis of localisation data associated to 120 out of the 130 images contained in the experimental website: (i) 100 images which, in the source web product, had alt texts that contravened the ten CL rules evaluated; (ii) ten images which originally had an empty `alt` attribute; and (iii) ten images for which the said attribute was missing. The remaining ten images had been assigned to the English website with what, to best of our knowledge, were appropriate phrase-based alt texts. In this section, we aim at observing whether the use of accessibility-oriented QA tools had a negative impact on the alt text quality achieved in the localised product for this alt text type.

Figure 7.11 depicts the distribution of the alt text scores across the different translation conditions. The corresponding data can be found in Table 7.16. The most distinctive results are observed in translation condition T3-Both-last-aDesigner, which registered the lowest proportion of score 1 (26.84%) and highest proportion of score 4 (21.53%). Interestingly, the mean score obtained in T1, when no tool was proposed, and the mean score obtained in T3, when the last tool used was aDesigner, are the same ($\bar{x}$=2.45, see Table 7.16). The average scores registered in the other translation conditions are slightly lower but never below 2. These almost unnoticeable differences across translation conditions can be also observed at an image level (see Figure 7.12). For reference purposes, exact score proportions, means and standard deviations of this alt text type are presented in Tables K.34 and K.35 and illustrated in Figures K.29 and K.30 (see Appendix K.2). Before looking into the significance of the results, it is worth noting that while we were confident to have included in the source website acceptable image text alternatives, at least 25% of the alt texts in this set were considered by end users as non-appropriate. Although this outcome may be well due to a flaw in our experimental design, we attribute it to the inherent subjectivity of the alt text assessment task, as what might be regarded by some end users as a pertinent text alternative could be seen by others as extremely long, uninformative or unnecessary. Another explanation for this may be associated with the quality of the translations themselves.

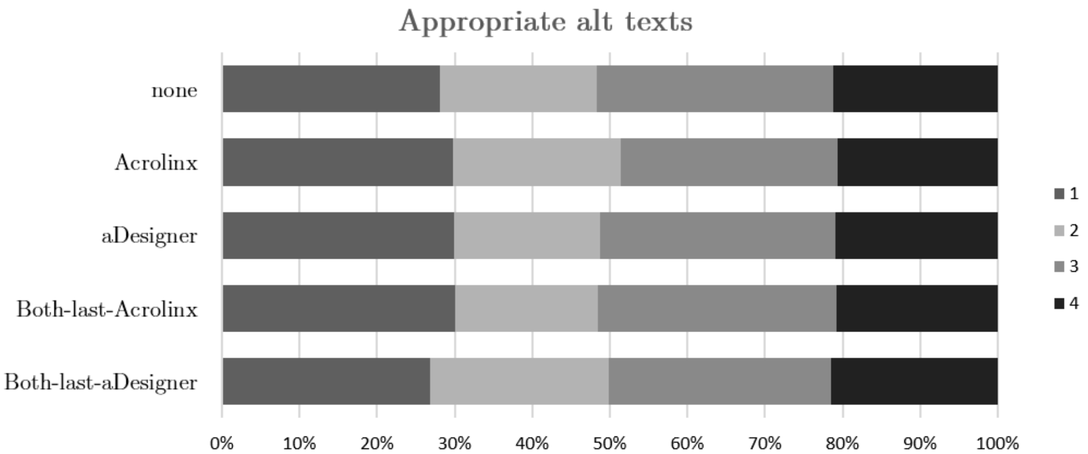


Figure 7.11. Distribution of scores per translation condition (five levels) assigned to images which originally had an appropriate alt text

Table 7.16. Score mean, standard deviation and proportions (%) per translation condition (five levels) for images which originally had an appropriate alt text

#P	Version	Tool	Mean	SD	1	%	2	%	3	%	4	%
14	T3	aDesigner	2.45	1.10	263	26.84	225	22.96	281	28.67	211	21.53
14	T3	Acrolinx	2.42	1.12	294	30.00	181	18.47	302	30.82	203	20.71
14	T2	aDesigner	2.42	1.12	293	29.90	185	18.88	297	30.31	205	20.92
14	T2	Acrolinx	2.39	1.12	292	29.80	212	21.63	274	27.96	202	20.61
28	T1	none	2.45	1.11	550	28.06	397	20.26	596	30.41	417	21.28

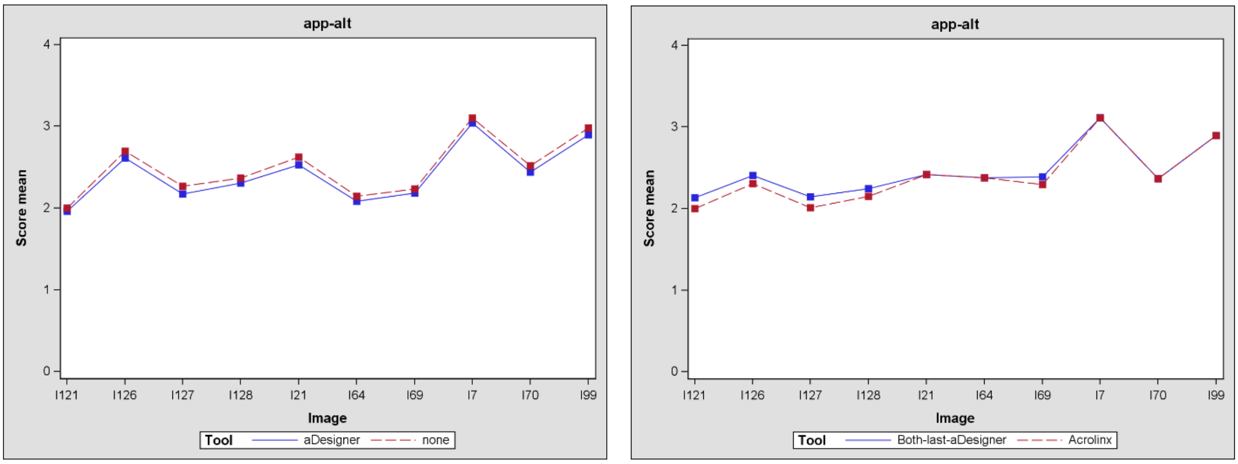


Figure 7.12. Pairwise comparison of mean scores for images which originally had an appropriate alt text: aDesigner (T2) vs. none (T1) (left), and Both-last-aDesigner (T3) vs. Acrolinx (T2) (right)

To verify whether changes in the scores collected across translation versions were or not, we followed the same analytical procedure as in previous alt text types. To this end, we applied a binomial logistic regression model to the data and, based on the estimates it yielded, we predicted the average proportion of positive scores (2-4) to be obtained per translation condition. As shown in Table 7.17, the probability of transferring the alt text appropriateness achieved in the source web product to the target website is higher than 70%, regardless of the translation condition under which localisers are working. The highest probability is estimated at 74.78% for those who follow QA Scenario B (Acrolinx-aDesigner). The data presented in Table 7.17 supports the observations thus far in this thesis when reporting the score distribution across translation conditions.

Table 7.17. Estimated average proportion of positive scores (2-4) per translation condition (five levels) for images which originally had an appropriate alt text

Translation Condition	Mean	Std. Error Mean	Lower Mean	Upper Mean
Both-last-aDesigner	0.7478	0.0411	0.6592	0.8197
Both-last-Acrolinx	0.7162	0.0441	0.6224	0.7944
aDesigner	0.7172	0.0440	0.6236	0.7952
Acrolinx	0.7183	0.0440	0.6248	0.7960
none	0.7356	0.0411	0.6477	0.8082

The Tukey's test (multiple comparisons of means) and the Odds Ratio did not return any significant results, so we have considered it unnecessary to include the corresponding data here. According to these findings, we can thus infer that neither the localisation of presumably appropriate alt texts nor their automated verification by means of accessibility-oriented QA tools like Acrolinx or aDesigner leads to a significant decrease in the target website of the alt text quality level already achieved in the source text. In addition, results reveal that, although minor improvements can be obtained through the combined use of Acrolinx and aDesigner, localisers can produce text alternatives of acceptable quality without the help of any automated QA support if the source website features a sufficient image accessibility level. In future work, it would be interesting to try to support this assertion by considering data from a larger data sample.

7.2.1.5 Conclusions on the alt text type

The study of how the nature of the text alternatives originally introduced in the source website influenced the final level of image accessibility achieved in the localised website has yielded interesting results, especially when this variable was considered in interaction with *IV2. Use of accessibility-oriented QA tools.*

By focusing on the set of **inappropriate text alternatives** that we purposely presented to participants, we have been able to assess the impact of 10 out of the 40 CL rules developed in the context of this thesis to detect problematic alt texts in French. Overall, we have observed that the application of eight out of the ten CL rules contributed to significantly increase the quality of localised alt texts from T1, when no tool was used, to T2, when Acrolinx served as the main NLP-based technology to implement those rules. The highest impact was achieved with the application of *FAA3-social* and *FAC1-accueil*, key for an enhanced interaction of blind users with functional images, and *U2-placeholder*, which aimed at detecting uninformative alt texts. While the latter is usually covered by general WAE tools which rely on heuristic methods, the former are part of the new recommendations we have put forward based on the Swiss web corpus we exploited during rule development (see Appendix A.1).

When Acrolinx was used after aDesigner (QA Scenario A), we found the likelihood of obtaining better results thanks to the CL rules in terms of alt text quality was lower than when Acrolinx was used alone, thus supporting the main conclusions drawn so far with regard to the effectiveness of the sequence Acrolinx-aDesigner. This could be due to the variability of the modifications made by localisers based on aDesigner's feedback, which our rules may have not covered, or to the translation proposals made by localisers, who might have moved away from the original idea expressed in the source text. This is, in fact, one of the limitations of automated CL approaches, as it is never possible to predict all possible language formulations that can be made by a person. Within the particular context of this thesis, this is exacerbated by the fact that text formulations are strongly linked to the interpretation of visual content in a web page, which is not easily automated, as discussed in Chapter 3 (see section 3.4.2). Yet, we have been able to prove the significant effectiveness of seven out of the 10 rules with respect to the appropriateness of the text alternatives produced with the help of aDesigner at T2.

Some of the most striking results were observed concerning the rules DDR3-logo and DDR2-image. These aimed at removing redundant information from alt texts for the sake of succinctness. In the case of the former, we have demonstrated that users tend to be against an abusive used of the word ‘logo’ to describe an image that shows a logotype, thus contributing to the ever-evolving discussion in the literature about its presence in alt texts. With regard to the latter (DDR2-image), we have observed that redundancy (i.e. using expressions such as ‘Picture of...’, ‘Image representing...’) was not considered as particularly problematic by end-users. These findings contrast with the results of our exploratory pilot study, where this rule led to significantly more appropriate text alternatives (Rodríguez Vázquez et al. 2014). A possible explanation for this difference may be that, in the pilot study, the rules were applied by the researcher, who had presumably a higher level of image accessibility expertise than the localisers who took part in our experiment. In this sense, as indicated earlier in this chapter, a more finegrained analysis of the tools' test reports and the application of rules by members of the two groups (control and treatment) may prove useful in the future to expand on this conclusions with regard to each rule's degree of effectiveness.

All in all, we believe that our rule evaluation can have practical applications for practitioners who seek guidance about how to write appropriateness text alternatives for images. First, based on empirical evidence, we have confirmed some of the recommendations already found in the literature, which the following rules rely on: *U2-placeholder*, which states that isolated undescriptive words (e.g. page, image) and word sets serving as placeholders (e.g. Insert content here) should be avoided; *U4-decor*, which states that decorative images should not be described; and *FDA3-structure*, which states that images that provide structure to the web document should not include a denotative description (e.g. vertical line).

In addition, we have demonstrated the effectiveness of some of the recommendations that emerged from our corpus analysis: *FAA1-files*, which states that when the image enables the user to perform an action related to a file, an imperative verbal form indicating the action, instead of the name of the file or its format, should be used (e.g. Download Geneva's map); *FAA3-social*, which suggests the same guideline as the previous rule but applied to social media actions, with specific indications about the information to be given about them (e.g. ‘Share this page on my Facebook wall’ would be more appropriate than ‘Share on Facebook’); or *FAA4-print*, which also suggest the use of imperative verbal forms for print-related actions, in combination with the specification of the content of the file that will be printed. Of course, these guidelines are based on the most general recommendation that alt texts for functional images should describe the purpose of the image, not its content. Yet, this had never been tested before with regard to the particular functions described above, nor had practitioners given a sufficient number of examples about how to implement such recommendation. Finally, it should be mentioned that, while we are aware that the recommendations we have formalised have been only tested for the French language, we consider that they could at least provide guidance about the type of issues (verb tenses, information to be given to the user, etc.) that should be considered when trying to create high quality alt texts in other languages.

As far as the category of **empty `alt` attributes** is concerned, results have again confirmed the trend we have been observing thus far. More appropriate text alternatives are achieved for images who initially had a null alt attribute when localisers use first Acrolinx and then aDesigner. Still, the only significant difference in terms of scores was noticed from T2-Acrolinx to T3-aDesigner. This could be attributed to what is known as the learning or practice effect (MacKenzie 2013, 177), i.e. by running several automated tests in a row with the two tools, localisers may have more clearly understood (thanks to the guidance provided) when an empty `alt` attribute should be inserted. If this was the case, when tools were used in reverse order, similar results should have been obtained. However, we have demonstrated that following QA Scenario A can lead to significantly lower alt text quality results than QA Scenario B. The fact that no significant improvements were observed when only one tool was used denotes that localisers are keener to propose new empty `alt` attributes than reconsidering the appropriateness of the null alt texts already present in the source. This implies somehow that a higher level of awareness is needed among localisation professionals for them to show more confident attitudes towards accessibility implementation.

The high proportions of score 1 (*non-appropriate*) observed in the category of images which originally had **no `alt` attributes** correlates with the results we presented in Chapter 6 concerning the low edit rates observed for this category. They also strengthen the severity of not providing images with a text equivalent and the significant negative impact that this practice can have on the overall accessibility level of a website; two remarks that the literature, and especially the WCAG 2.0, so much insist on. In the particular context of multilingual websites, our study has shown that the probability of achieving an acceptable degree of image accessibility when the source contains `<img>` elements that lack an `alt` attribute is lower than 4% when no automated support to detect this problem is available. This means that (i) either localisers do not feel empowered enough to make changes to the source code (even if this has positive effects on the final accessibility of the localised website), (ii) or more emphasis needs to be placed in accessibility-oriented issues in localisation training.

When accessibility-oriented QA tools are facilitated, gains in image accessibility are strongly significant for this specific alt text category. Concretely, the combined use of Acrolinx followed by aDesigner (QA Scenario B) can lead to image accessibility levels in the target website that are 20 times higher than in the source. It should be noted, however, that the improvements were are reporting are probably based on the insertion of an `alt` attribute in the target document when there was not one in the source document. A qualitative analysis of the values introduced by localisers would be needed to assess the degree of appropriateness achieved in each new text alternative. All in all, although the rate of images without alt attributes on the Web has decreased over the last years, as we have noted earlier in this thesis (see section), our study suggests that not performing WA assessments or the lack of accessibility awareness among the localisation community could lead to spread wrong practices in the multilingual Web.

Finally, our findings concerning the quality of localised text alternatives for images which initially had **appropriate alt texts** have been, at least, comforting. While we expected that, thanks to the use of targeted tools, quality could be enhanced, no significant differences have been observed across translation versions. This should not, however, be seen as negative result, as it also shows the tools' good performance in terms of correctness (i.e. their ability to reduce false positives), a parameter that is often looked at when discussing the effectiveness of automated testing (see Chapter 2, section 2.3.2). Tools have not contributed to downgrade the quality of the alt texts either. All in all, the most noteworthy conclusion that can be drawn from the analysis of this alt text category is that the localisation process has not resulted in a lower image accessibility level. We could thus infer that if acceptable levels of image accessibility are achieved in the source website, it is likely that the target version will at least maintain such degree of accessibility. This statement is only based on the translation and adaptation of text alternatives originally present in the source. Nevertheless, further studies would be needed to investigate whether this would be also the case if higher levels of localisation are required (website redesign, replacement of images to fit target culture conventions, etc.).

7.2.2 HTML knowledge

Apart from the nature of the alt texts introduced in the website requested for translation, another factor that could have had an impact on the alt text quality results reported so far is the participants' HTML knowledge. One may believe that a localiser with more basic background on the topic might not be fully aware of the function of the `alt` attribute, while those with more advanced skills in HTML are more likely to identify and adapt alt texts, as well as to be keen to render images more accessible.

Table 7.18 summarises the responses given by the 28 participants (overall and per group) to question 28 of the screening questionnaire (Appendix D.2). Although these were already reported when providing information about the participants' background in Chapter 5 (see section 5.2.2.4), we include them again here for contextualisation purposes. As indicated in the aforementioned section of Chapter 5, the participant from the control group (P42) who considered not to have any background on the subject obtained the highest mark in the HTML test and was thus accepted.

When observing the proportional distribution of the alt text scores according to the participants' self-rated HTML knowledge (see Table 7.19 and Figure 7.13), it is possible to notice that the highest proportion of very appropriate alt texts was achieved by the participants who acknowledged to have intermediate and advanced HTML skills (11.33% and 17.69% respectively). Similarly, it is worth noting that they registered a lower proportion of not appropriate text alternatives (44.56% and 30.62% respectively) than those who reported a basic HTML background (51.25%). Nevertheless, the Wald test (Engle 1984) indicates that these differences are not significant (χ^2=1.638, df=1, p=0.20). It thus follows that the differences in terms of HTML skills across participants did not have a significant impact on the quality of the alt texts they produced and, therefore, they do not pose a threat to the research's validity.

Table 7.18. Response count and percent (%) for question 28 (D.2): HTML knowledge

How would rate your knowledge of HTML?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
None	1	7%	-	-	1	4%
Basic	10	71%	9	64%	19	68%
Intermediate	3	21%	4	29%	7	25%
Advanced	-	-	1	7%	1	4%

Table 7.19. Overall alt text score mean, standard deviation and proportions (%) according to participants' self-rated HTML knowledge

HTML knowledge	**Mean**	**SD**	**1**	**%**	**2**	**%**	**3**	**%**	**4**	**%**
None	1.91	1.04	1338	49.01	571	20.92	545	19.96	276	10.11
Basic	1.87	1.04	26584	51.25	10545	20.33	9642	18.59	5099	9.83
Intermediate	2.01	1.06	8496	44.46	4151	21.72	4298	22.49	2165	11.33
Advanced	2.29	1.08	836	30.62	746	27.33	665	24.36	483	17.69

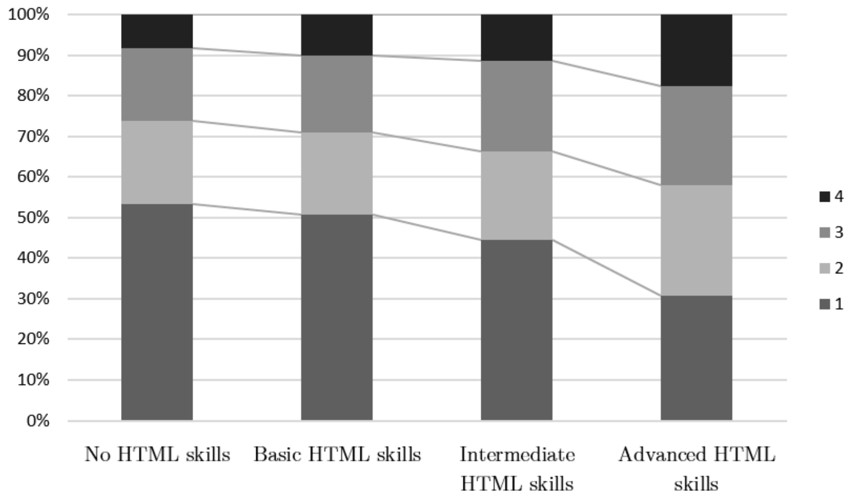


Figure 7.13. Overall distribution of alt text scores according to participants' self-rated HTML knowledge

The HTML background information participants provided us with was complemented with an HTML test, presented in the form of a semi-open question (see Appendix D.2, question 29). Participants were invited to list the localisable elements of an HTML file excerpt (see Figure 5.1, Chapter 5) and their responses were then coded on a graded scale, from *1. The participant would not make any changes* to *5. The participant identified all localisable elements, including `alt` attribute values.* Responses to the HTML test are summarised in Table 8.20. From those who said they would make some changes, twelve (N=12 out of 26) indicated that they would prefer to localise the HTML file excerpt using a CAT tool or an HTML editor instead of a regular advanced text editor with a view to avoiding corrupting the code. Interestingly, only two participants recognised the unique alt text contained in the excerpt. This could be explained by the fact that it was not represented in the same colour as the other translatable text and, in addition, the `alt` attribute value was identical to the `<img> src` value.

Table 7.20. Response count and percent (%) for question 29 (D.2): HTML test

Which elements would you translate/modify from the HTML file excerpt proposed?

Response	Control	%	Treatment	%	Overall	%
1. No changes	1	7%	1	7%	2	7%
2. Page title & body content	5	36%	4	29%	9	32%
3. 2 + page keywords & description	6	43%	1	7%	7	25%
4. 3 + Content-Language value	1	7%	7	50%	8	29%
5. 4 + alt attribute values	1	7%	1	7%	2	7%

The Spearman's rank correlation coefficient, a non-parametric test to measure the strength of relation between two variables when the data are not normally distributed, was used to determine the relationship between the participants' self-rated HTML skills and the results of the HTML test. Surprisingly, the test showed a very low correlation ($r_s = 0.067$). However, before drawing any conclusions, we first observed the distribution of the alt text scores according to the participants' HTML test results. Data gathered suggest that the higher the number of localisable elements identified in the HTML excerpt, the higher the quality of the text alternatives produced (see Figure 7.14 and Table 7.21).

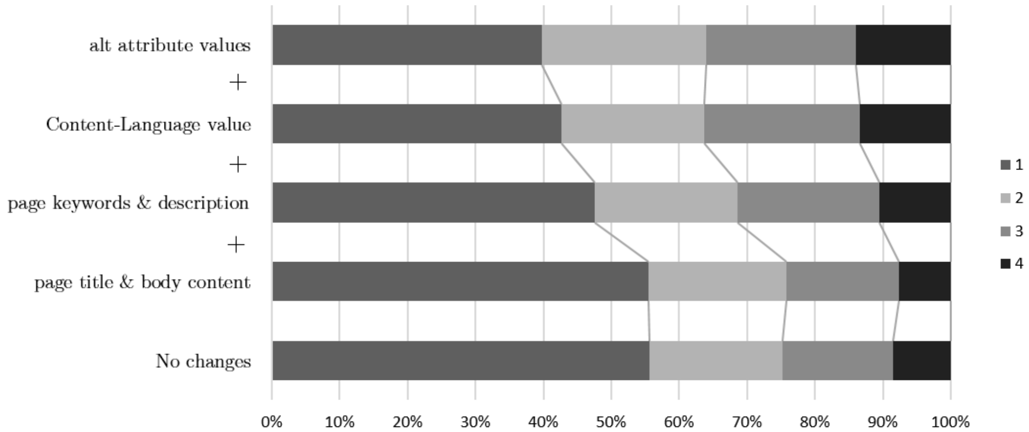


Figure 7.14. Overall distribution of alt text scores according to participants' HTML test results

Table 7.21. Overall alt text score mean, standard deviation and proportions (%) according to participants' HTML test results

HTML test	Mean	SD	1	%	2	%	3	%	4	%
No changes	1.78	1.00	3034	55.57	1074	19.67	887	16.25	465	8.52
Page title & body content	1.76	0.98	13639	55.51	4987	20.30	4070	16.56	1874	7.63
page keywords & description	1.94	1.05	9091	47.57	4027	21.07	3992	20.89	2000	10.47
Content-Language value	2.07	1.09	9316	42.66	4608	21.10	4991	22.85	2925	13.39
alt attribute values	2.10	1.08	2174	39.82	1317	24.12	1210	22.16	759	13.90

A repeated measures one-way ANOVA revealed that the differences that exist between the alt text scores obtained when we take into account the participants' HTML test results are significant (χ^2=6.31, df=1, p=0.012). If we consider this finding alone, we could conclude that having good HTML skills can lead to the production of significantly more appropriate text alternatives. A possible explanation for the low correlation observed between the participants' self-rated knowledge and the test results might be that they underestimated their skills. Another possible explanation may be that the selection process caused a bias in localisers' responses. Since in the call for participation we had listed having basic HTML skills as one of the main requirements, choosing this option in the questionnaire might have been regarded as the safest way to participate in the study.

7.2.3 Use of CAT tools

As explained in Chapter 5 (see section 5.2.4), during the experimental study, participants were free to use the CAT tool of their choice or to localise the web pages without the help of any translation technology. We know from on our experience that certain tools do not recognise by default image text alternatives as translatable elements. Differences regarding text extraction and tag recognition might also occur across different versions of the same tool (see discussion in Chapter 3, section 3.5.2.2). Hence, it would be no exaggeration to think that this independent variable, given by the experimental design itself, could have had an effect on the experiment's outcome. In fact, although not directly related to accessibility issues, previous research has already demonstrated that the use of CAT tools can have a positive impact on the final translation quality (Morado Vázquez 2012).

In the web localisation post-task questionnaire, 39% (N=11) of the participants said that they had not used a CAT tool to produce translation version T1 (see Table 7.22). Those who had were asked to give reasons for their decision. Approximately half of them (N=9, 53%; control=2, 33%; treatment=7, 64%) indicated that they had used a CAT tool to reduce the time effort. Eight participants (N=8, 47%; control=4, 67%; treatment=4, 36%) acknowledged that they had used it for consistency and quality purposes, as well as to avoid the translation of repeated text thanks to the use of a translation memory (TM).[8] The most alleged reason was to avoid corrupting the code (format protection) and to identify easier the text to translate (N=13, 76%; control=5, 83%; treatment=8, 73%). For instance, P56 stated: “Pour détecter rapidement le texte à traduire (sans balises) et dans un souci de cohérence et de rapidité.” Finally, it is worth mentioning that four people from the treatment group (N=4, 36%) indicated that they had used a CAT tool because it was easier than using any other tool and because they were used to working with them.

[8] We did not provide participants with a TM. Some reported that they had created one for the specific purposes of our study, while others indicated that they had used TMs from their past translation projects.

Table 7.22. Response count and percent (%) for question 1 (D.3): Use of CAT tools

Have you used a CAT tool during session 1 (T1)?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I have	6	43%	11	79%	17	61%
No, I have not	8	57%	3	21%	11	39%

Of the 17 participants who used a CAT tool, only five (N=5, 29%) claimed to have made some changes in T1 outside the CAT tool translation environment (see Table 7.23). All five participants used Notepad++. Two of them said that they had just made minor changes while revising the translated text. Concretely, P59 stated: "Tous les éléments à traduire n'ont pas été extraits par l'outil de TAO. Modifications nécessaires en texte brut à des fins de vérification et d'harmonisation." Overall, the main reason reported was to edit the code where needed (adaptation of language code, insertion of HTML character entities; for instance, to add non-breaking spaces). Interestingly, none mentioned the fact of having added `alt` attributes when they were missing, whereas as reported in Chapter 6 (section 6.2.3.5), P59 had indeed added the said attribute to four `<img>` elements, and P26 to nine images.

Table 7.23. Response count and percent (%) for question 3 (D.3): Changes made outside the CAT tool environment

Have you made any changes to your initial translation directly on target HTML files, outside the CAT tool environment?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I have	1	17%	4	36%	5	29%
No, I have not	5	83%	7	64%	12	71%

Figure 7.15 illustrates the data presented in Table 7.24, which compares the mean scores, standard deviations and score proportions obtained for alt texts from T1 by participants who used a CAT tool and those who did not. Although this last group seems to have achieved better results, differences are almost non-existent. This observation is confirmed by the repeated measures one-way ANOVA, which reveals that the use of CAT tools did not have a significant impact on the alt text scores (χ^2=0.133, df=1, p=0.715). The freedom to choose how to perform the task was given to localisers in attempt to boost the ecological validity of the study, even if we were aware that it could pose a threat to its internal validity. However, this finding serves to rule out this possibility.

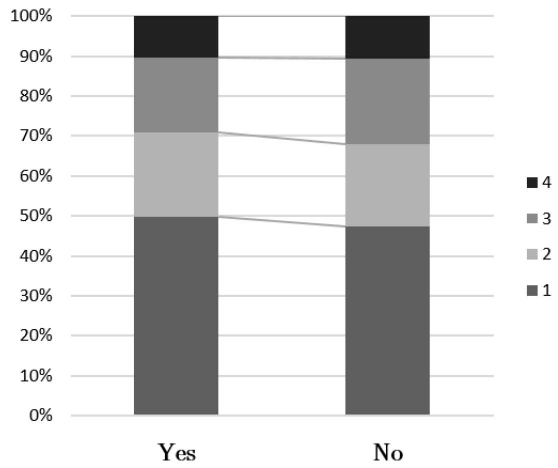


Figure 7.15. Overall distribution of alt text scores with regard to the use of CAT tools during T1

Table 7.24. Overall alt text score mean, standard deviation and proportions (%) with regard to the use of CAT tools during T1

Use of CAT	Mean	SD	1	%	2	%	3	%	4	%
Yes, I have	1.90	1.04	23038	49.64	9817	21.15	8708	18.76	4847	10.44
No, I have not	1.95	1.05	14216	47.34	6196	20.63	6442	21.45	3176	10.58

7.2.4 Check of Acrolinx rule help

Ten items on the quality assurance post-task questionnaire (Appendix D.4) aimed at gathering data about the participants' opinion and perceived performance of Acrolinx, five of which focused specifically on Acrolinx's rule help. As previously discussed in this thesis (see Chapter 2, section 2.3.3 and Chapter 3, section 3.4.3), web accessibility evaluation (WAE) tools do not generally provide a comprehensive explanation about the reasoning behind the errors they flag, thereby reducing their effectiveness. By further developing Acrolinx rule documentation (see section 4.5.2, Chapter 4), we sought to cover this major drawback. Therefore, we hypothesise that localisers who checked the rule documentation, available at a one-click distance from Acrolinx's check report, obtained a clearer overview of the issues reported and thus produced alt texts of higher quality.

In an attempt to validate this hypothesis, we first asked participants how often they had checked the rule help. Three people from the control group and three people from the treatment group (N=6, 21%) replied that they had only checked Acrolinx's general report and thus did not consult the rule-specific documentation (see Table 7.25). From the 22 participants who did consult it, 19 (86%) indicated that the explanations given were useful (see Table 7.26). When explicitly asked about the utility of the examples provided in the rule help, 20 participants (91%)

claimed that they had found them useful too (see Table 7.27). Although one would have expected the control group to check the documentation more often, data shows that the treatment group felt almost the same urge to do it. Similarly, given their basic background on web accessibility, one would have believed that the treatment group would have found the explanations and the examples provided not so advantageous. However, both groups shared the same impression in this regard.

Table 7.25. Response count and percent (%) for question 6 (D.4): Check of Acrolinx rule help (frequency)

Have you checked the rule help associated to each one of the errors flagged by Acrolinx?

Response	Control	%	Treatment	%	Overall	%
Never	3	21%	3	21%	6	21%
25% of errors flagged	5	36%	3	21%	8	29%
50% of errors flagged	2	14%	4	29%	6	21%
75% of errors flagged	3	21%	1	7%	4	14%
Always	1	7%	3	21%	4	14%

Table 7.26. Response count and percent (%) for question 7 (D.4): Acrolinx rule help usefulness

Overall, have you found the rule help useful?

Response	Control	%	Treatment	%	Overall	%
Yes, I have	9	82%	10	91%	19	86%
No, I have not	2	18%	1	9%	3	14%
I do not know	-	-	-	-	-	-

Table 7.27. Response count and percent (%) for question 8 (D.4): Acrolinx rule help examples

More specifically, have you found the examples available in the rule help useful?

Response	Control	%	Treatment	%	Overall	%
Yes, I have	10	91%	10	91%	20	91%
No, I have not	1	9%	1	9%	2	9%
I do not know	-	-	-	-	-	-

Two further questions were asked relating to the rule help content. They required participants to express their opinion about the possibility of including images illustrating the examples provided, as well as audio clips reproducing how a screen reader user would perceive a given image. As expected, almost two thirds of the participants declared that they would have liked to have that type of complementary information when consulting the rule help (control: N=9, 82%; treatment: N=8, 73%). We believe that this feedback may also reflect the opinion of other content authors who are not web accessibility experts and that it could be helpful for web accessibility evaluation tool developers in general.

Figure 7.16 depicts the overall distribution of alt text scores according to how often the Acrolinx rule help was checked. Data suggests that localisers who consulted the rule documentation for at least 75% of the errors flagged managed to render more than 60% of the images accessible. Those who claimed to have always checked the rule help obtained better results than those who never consulted it (mean scores: $\bar{x}$=2.02 and $\bar{x}$=1.89 respectively). Nonetheless, they did not register the lowest proportion of not appropriate alt texts. Table 7.28 provides more detailed results.

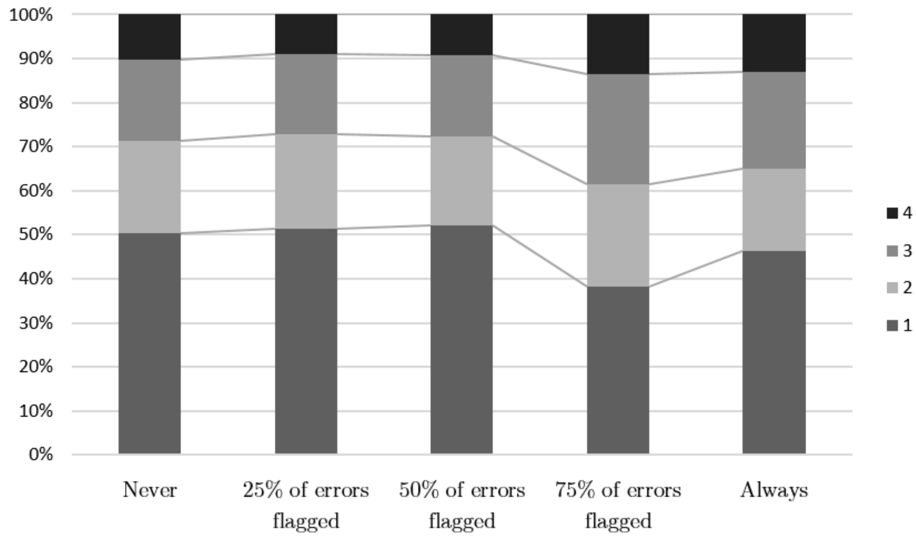


Figure 7.16. Overall distribution of alt text scores according to Acrolinx rule help check frequency

Table 7.28. Overall alt text score mean, standard deviation and proportions (%) according to Acrolinx rule help check frequency

Check of Acrolinx rule help	Mean	SD	1	%	2	%	3	%	4	%
Never	1.89	1.04	8266	50.46	3402	20.77	3016	18.41	1696	10.35
25% of errors flagged	1.85	1.01	11209	51.32	4719	21.61	3963	18.15	1949	8.92
50% of errors flagged	1.85	1.02	8548	52.19	3315	20.24	3019	18.43	1498	9.15
75% of errors flagged	2.14	1.07	4167	38.16	2536	23.22	2754	25.22	1463	13.40
Always	2.02	1.10	5064	46.37	2041	18.69	2398	21.96	1417	12.98

In order to examine whether these differences were significant or not, we used a repeated measures one-way ANOVA. The Wald test concluded that having checked the Acrolinx rule help had a significant and positive impact on the alt texts scores obtained (χ^2=7.2005, df=1, p=0.007). This means that not only was Acrolinx's documentation consulted and found useful, but also decisive to achieve more appropriate text alternatives. This finding highlights the importance of properly documenting the automated checking process and corroborates the assumptions made by Petrie et al. (2011). It also relates with the conclusions that Tang (2012) drew in her doctoral thesis, where she demonstrated that different user groups (accessibility specialists, website developers and content providers) were able to select more pertinent information to be included in an image alt text when they received the appropriate guidance.

7.2.5 Check of aDesigner error descriptions

In the quality assurance post-task questionnaire (Appendix D.4), participants were also requested to give their feedback about aDesigner. Three questions (17, 18 and 19) focused on the descriptions of the issues flagged by the tool. We first asked participants whether they had checked the explanations related to the image accessibility-related issues detected by aDesigner. It should be remembered that, contrary to Acrolinx, error descriptions provided by aDesigner are included directly in the check report (see, for instance, Figure 5.3 in Chapter 5). Nine participants (N=9 out of 28, 32%; control: N=4, 29%; treatment: N=5, 36%) replied that they had never read them (see Table 7.29). We need to assume, therefore, that either (i) they did not apply any changes to their corresponding translation versions when using aDesigner, or (ii) they only relied on the 'Visualise blind usability' function, which allows the user to see some of the issues flagged in context.[9]

As shown in Table 7.30, explanations given to describe image-accessibility problems were found equally useful by the control group and the treatment group. Nevertheless, the latter did not appreciate the examples provided as much as the former (see Table 7.31). A possible explanation for this might be that, most of the time, examples in aDesigner are somehow hidden, often in the form of suggestions. For instance, when no `alt` attribute is found for an `<img>` element, the description of the error is as follows: “No alternative text for an image. Provide its text alternative. If the image should be ignored by assistive technology, provide `alt=""` and not `title` attribute.” While this might have seemed illustrative for the control group, it might well have been regarded as elementary information by the treatment group members, who presumably were already aware on how to intervene when an image should not be presented to the blind user and were expecting, perhaps, more revealing and specific examples.

[9] This feature of aDesigner allows users to simulate how a blind person would perceive a given web page by means of her screen reader. After running a check, only the accessibility issues classified as 'Error' (as opposed to 'Probably error', 'Need confirmation' or 'Human check') are visually identified in the web page with an exclamation mark. When the user hovers over it, the description of the issue flagged is shown. See Chapter 3 (section 3.4) for a more comprehensive discussion of the general functionalities of web accessibility evaluation tools and aDesigner in particular.

Table 7.29. Response count and percent (%) for question 17 (D.4): Check of aDesigner error descriptions

Have you checked the problem description of each one of the image accessibility-related errors flagged by aDesigner?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Never	4	29%	5	36%	9	32%
25% of errors flagged	2	14%	1	7%	3	11%
50 %of errors flagged	4	29%	5	36%	9	32%
75% of errors flagged	3	21%	-	-	3	11%
Always	1	7%	3	21%	4	14%

Table 7.30. Response count and percent (%) for question 18 (D.4): usefulness of aDesigner error descriptions

Overall, have you found aDesigner's explanation of the image accessibility-related issues useful?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I have	8	80%	8	89%	16	84%
No, I have not	2	20%	1	11%	3	16%
I do not know	-	-	-	-	-	-

Table 7.31. Response count and percent (%) for question 19 (D.4): aDesigner help examples

More specifically, have you found the examples included in aDesigner help useful?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I have	8	80%	3	33%	11	58%
No, I have not	2	20%	6	67%	8	42%
I do not know	-	-	-	-	-	-

As previously done in the case of Acrolinx, we combined the data gathered from the quality assurance post-task questionnaire regarding the frequency with which participants had checked aDesigner's help with the alt text scores they obtained when they used the tool. The bar chart presented in Figure 7.17 depicts that consulting aDesigner's explanations did not necessarily lead to a clear improvement pattern in terms of alt text quality. The highest proportion of not appropriate text alternatives was registered by those who claimed to have checked aDesigner's guidance for at least 75% of the errors flagged (score 1=60.94%). This contrasts with the observations made with respect to Acrolinx, where those who checked the rule help with the same frequency obtained, in fact, the best results. One of the most surprising aspects of the data is that the highest alt mean score was achieved when participants did not check the error descriptions shown in the report ($\bar{x}$=2.04). As reported in Table 7.32, the average score remained slightly below 2 when localisers always consulted aDesigner's explanations, as well as when they only did so in 25% of the cases.

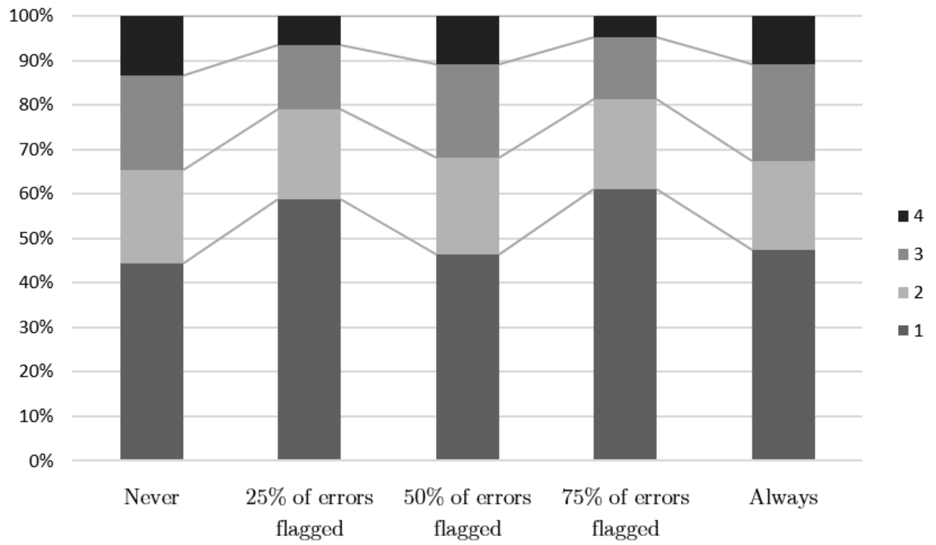


Figure 7.17. Overall distribution of alt text scores according to according to aDesigner help check frequency

Table 7.32. Overall alt text score mean, standard deviation and proportions (%) according to aDesigner help check frequency

Check of aDesigner help	Mean	SD	1	%	2	%	3	%	4	%
Never	2.04	1.09	10900	44.36	5142	20.93	5264	21.42	3264	13.28
25% of errors flagged	1.69	0.95	4806	58.68	1670	20.39	1173	14.32	541	6.61
50% of errors flagged	1.96	1.05	11387	46.35	5352	21.78	5172	21.05	2659	10.82
75% of errors flagged	1.62	0.89	4991	60.94	1672	20.42	1143	13.96	384	4.69
Always	1.96	1.06	5170	47.34	2177	19.94	2398	21.96	1175	10.76

To statistically measure the impact of having checked aDesigner's help concerning image accessibility-related issues on the alt text scores obtained, we used a repeated measures one-way ANOVA. The analysis revealed that consulting aDesigner's explanations led to the production of text alternatives of significantly lower quality (χ^2=7.219, df=1, p=0.007). The negative impact is inferred from the estimate value returned by the mixed linear model (-0.04024, *se*=0.01410). This result could be explained by the low level of specificity of the tool with regard to image accessibility. It has been claimed that the more specific a tool is (that is, the higher the number of contextual and helpful suggestions proposed), the more effective it is (Brajnik 2004). Since aDesigner only detects the absence of `alt` attributes and of isolated words within the `alt` attributes that could be the source of inappropriate text alternatives, error descriptions remain quite vague, probably in an attempt to avoid offering misleading guidance. The negative impact reported could also be attributed to the short length of the descriptions and the lack of variety in the examples offered. It is worth noting, however, that this finding should be interpreted within the context of the results reported for aDesigner in Chapter 6 (sections 6.3.3.1 and 6.3.3.2), which showed that, overall, the tool was efficient.

7.2.6 Overall conclusions

In section 7.2, we have provided a comprehensive analysis of how the final level of appropriateness achieved in the target websites delivered by localisers was influenced by a total of six secondary independent variables. Since we have already discussed the findings and the impact concerning one of them ('alt text type', see section 7.2.1.5), we will focus here on the other five.

One of the variables that we considered to have possibly affected localisers' performance in terms of image accessibility was their **self-rated HTML knowledge**. While, when analysing the data, we observed that the highest proportion of very appropriate alt texts was achieved by the participants who had acknowledged to have intermediate and advanced HTML skills, we have found that the differences in terms of alt text appropriateness across localisers were not significantly related to their reported background. However, after analysing the results of the **HTML localisation test** participants carried out during the selection process, we noticed that the higher the number of localisable elements identified in the test, the higher the quality of the text alternatives produced. Surprisingly, these differences were found to be statistically significant, but not correlated to their self-reported knowledge. As we noted earlier, this could be derived from (i) localisers underestimating their skills, or (ii) some degree of bias caused by the selection process (we had simply asked for basic knowledge on the matter, so it is likely that we influenced their response to the question). Given the data gathered, we are inclined to believe that having good HTML skills can lead to the production of significantly more appropriate text alternatives.

The use of **CAT tools**, however, seems to be less influential. Although 11 out of 28 localisers indicated that they had not used any translation software during the experiment, statistical tests indicated that this did not have an effect on the quality of alt texts produced. We consider, nonetheless, that further analyses are needed to reinforce this assertion. For instance, looking into the web localisation process data (e.g. the video recordings) would allow us to observe whether those who used CAT tools (N=17) checked the text extraction settings of the tool prior to the localisation task or even during the subsequent QA task, after noticing that the accessibility-oriented QA tools were flagging text that looked unfamiliar for them.

The complementary analyses performed concerning the appropriateness of the alt texts produced by localisers across the different translation conditions have revealed that the effectiveness of the QA tools was affected by the quality of the error descriptions they provided. We have observed that consulting the **Acrolinx rule help** contributed significantly to increase the appropriateness of the alt texts that localisers produced. This confirms the assumptions we have been making throughout this thesis that, if automated testing is complemented by relevant guidance on the errors flagged, the evaluation process can lead to considerable improvements in terms of accessibility. In addition, this finding suggests that the explanations, examples and references we introduced in each rule help file (see Figure 4.4, Chapter 4) were pertinent and useful. In the future, this material could prove particularly beneficial for the development of accessibility-oriented web localisation courses. Further research might explore whether the guidance we developed can also be useful for other web professionals without a strong linguistic background.

In contrast to Acrolinx-related findings, checking **aDesigner's error descriptions** has resulted in text alternatives of significantly lower quality. While we knew from the beginning that guidance offered by aDesigner was limited compared to Acrolinx's, this strongly negative result was not expected. One could hypothesise that it could be related to the relatively small coverage aDesigner provides on image accessibility issues −despite the fact that, when compared to other WAE tools, as we indicated in Chapter 5 (section 5.2.3.3), aDesigner seems to be the one offering the highest level of specificity−. Overall, this evidence strengthens the arguments set forth before regarding the need to provide comprehensive explanations not only about the errors presented to the user, but also about how this could be corrected. In the case of image accessibility, offering information about how blind users interact with the Web and the different communicative values that images may convey can be of added value.

7.3 Post-task questionnaires

As announced at the beginning of the present chapter, this section provides a summary of the qualitative data collected through the two post-task questionnaires administered to the web localisers who took part in our experimental study. These data serve to complement the quantitative results reported in the previous section (7.2) as well as in Chapter 6.

7.3.1 Web localisation post-task questionnaire

The web localisation post-task questionnaire (see Appendix D.3) aimed at gathering information on (i) the use of CAT tools, (ii) the participants' impressions about the task, and (iii) the treatment group's feedback about the webinar they followed prior to the study. Since the first block of questions was already covered earlier in this chapter, in this section the focus is shifted to the last two.

7.3.1.1 Participants' impressions about the task

In an attempt to know whether localisers had identified image text alternatives as such during the web localisation task, we asked them to list the web elements that they believed they had translated or modified in T1 (question 5). Although we had given them a sample answer ("I have translated the title of each web page") to illustrate which kind of information we were expecting to receive, some localisers responded more broadly, either indicating that they had simply translated the text proposed by the CAT tool they were using (N=4, 14%), or saying that they had translated the textual content which was visually represented in the browser (N=6, 21%).

Concrete answers to question 5 are summarised in Table 7.33. Sixteen (N=16, 57%) out of the 28 participants explicitly mentioned image text alternatives when listing the web elements they had translated. Although none of them used the proper technical name in French ("équivalents textuels"), we assumed they were making reference to alt texts when they included in the list elements such as "descriptions des images" (P19), "texte de survol des divers éléments graphiques" (P23), "titres d'images" (P25), or "légendes des photos" (P42), to name just a few. As reported in Chapter 6 (see section 6.2.3.1), only six participants had not translated any alt text in T1 (see Table K.1, Appendix K). It thus follows that six of the 22 participants who did translate them did not necessarily recognise them as image text alternatives or they simply did not mention it in the post-task questionnaire.

Table 7.33. Response count and percent (%) for question 5 (D.3): web elements localised

Could you please list the web elements that you believe have translated and/or modified?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Main body content	5	36%	11	79%	16	57%
Alt attribute values	6	43%	10	71%	16	57%
Page titles	3	21%	9	64%	12	43%
Page keywords	5	36%	2	14%	7	25%
Language code	2	14%	2	14%	4	14%

Similarly, we asked localisers whether they would have made any further changes to T1 if they had been given more time. The six participants from the control group who replied positively to this question (see Table 7.34) gave different explanations, including the willingness, if they had the time, to (i) translate the text embedded in images, such as the title of the campaign (I7, see Appendix J); (ii) translate the title and keywords contained in the web page `<head>`; and (iii) proofread the translation before delivering it. Interestingly, four out of the five localisers from the treatment group who acknowledged they would have made more changes made reference to image text alternatives, indicating that they would have (i) included more comprehensive descriptions in certain cases, (ii) added `alt` attributes where they were missing, and (iii) harmonise the use of empty `alt` attributes. P21 also mentioned that he would have checked the scripts and the CSS file, in case there was text that needed to be translated. The level of awareness with regard to alt texts shown here by the treatment group is in agreement with the quantitative results presented as regards the translation of text alternatives (DV1).

Table 7.34. Response count and percent (%) for question 6 (D.3): time constraints

Should you have had more time, would you have made any further changes?						
Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I would have	6	43%	5	36%	11	39%
No, I would have not	8	57%	9	64%	17	61%

Irrespective of the time constraints, 61% of the participants (N=17) indicated that they would have proceeded differently if the task proposed would have been part of a real localisation assignment (see Table 7.35). Concretely, four participants (N=4, 36%) acknowledged that they would have used different tools; three (N=3, 27%) said that they would have requested more instructions from the client on what to localise and what not, two (N=2, 18%) said that they would have pre-processed the files prior to the localisation task, and another two (N=2, 18%) recognised that they would have checked the files for HTML best practices compliance before submitting the job. The remaining participants made reference again to the time pressure by indicating that they would have taken more time to perform revision tasks. We believe that most of the reasons alleged are understandable and that they reflect the limitations of any translation experimental environment.

Table 7.35. Response count and percent (%) for question 8 (D.3): real localisation assignment

If this was a real localisation assignment, would you have proceeded in the same way?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I would have	7	50%	10	36%	17	61%
No, I would have not	7	50%	4	64%	11	39%

Finally, participants were asked about the type of audience they had thought about when making translation decisions. Our aim was to know whether they had considered the possibility of visually-impaired users having access to the web content they had localised. From the 28 participants, only seven (N=7, 25%) mentioned the blind community, six of which were from the treatment group. Given the theme featured in the experimental website, the rest of the answers pointed at the general public or at people interested in humanitarian causes as the main target audience (see Table 7.36).

Table 7.36. Response count and percent (%) for question 10 (D.3): target audience

Which type of audience have you thought about when making translation decisions?

Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
General public	12	86%	6	43%	18	64%
Humanitarian causes	9	64%	5	36%	14	50%
Blind web users	1	7%	6	43%	7	25%

7.3.1.2 Participants' feedback on the webinar

At the end of the web localisation post-task questionnaire, the treatment group was presented with specific questions about the webinar (questions 11 to 15, see Appendix D.3). Overall, the feedback received was highly positive. Seventy-one per cent of the participants (N=10 out of 14, 71%) found it pertinent, 64% (N=9) interesting, 43% (N=7) useful, and 21% (N=3) inspiring. One person mentioned that the subject was technical and thus difficult to follow, and two people highlighted that although the presentation was clear, there was a lot of content to assimilate in such a short period of time.

To the question *Do you believe that the knowledge acquired during the webinar helped you to render your localised website more accessible?* , 57% of the localisers (N=8) replied positively and 14% (N=2) said 'I don't know'. The remaining four participants chose the option 'Other' and placed emphasis on the fact that the webinar was very instructive, but more time would have been needed during the web localisation task to implement some of the best practices learned. Similarly, they mentioned that having a copy of the presentation would have been of great help.

Localisers were also asked whether they believed that the knowledge acquired during the webinar had made the localisation task more difficult. From the 14 participants, three (N=3, 21%) said yes. The reasoning behind their answers was as follows: P59 indicated that he felt somehow obliged to pay extra attention to accessibility issues, which was not a common practice in his regular assignments. The other two (P6 and P56) made reference to the time constraints and the need to further develop their accessibility-related skills in order to feel more confident about their localisation decisions. From the remaining 11 participants, 57% (N=8) reported that they had not perceived the task as more difficult, one said 'I don't know', and two indicated that it was not a matter of difficulty but of time, since being aware of (and implementing) basic accessibility best practices required an additional effort which inevitably rendered the localisation task longer.

7.3.2 QA post-task questionnaire

Upon completion of the quality assurance task, participants were requested to report on (i) their perceived performance of the two accessibility-oriented QA tools used, and (ii) their personal opinion about localisation, web accessibility, and QA practices (see Appendix D.4).

7.3.2.1 Participants' feedback on the tools used

Apart from commenting on the usefulness of Acrolinx's and aDesigner's help as regards the description of the errors flagged (see sections 7.2.4 and 7.2.5), participants were also given the opportunity to provide their feedback with regard to the tool's documentation clarity. As illustrated in Figure 7.18,[10] Acrolinx's error explanations were regarded as clear or very clear by more than two thirds of the participants, as opposed to aDesigner's, which were only perceived as clear or very clear by half of the participants (see Table 7.37). These results are in line with the observations made earlier in this chapter about the positive impact of having checked Acrolinx rule help and the negative effect of consulting aDesigner's help on the quality of the localised image text alternatives.

[10] To better interpret the gray scale colour coding used both in this figure and in Figures 7.19 and 7.20, notice that the highest percentage interval (75%-100%) (not appropriate) is always at the bottom of the graphic (i.e. lowest part of the y-axis) and the lowest one (0%) is always at the top.

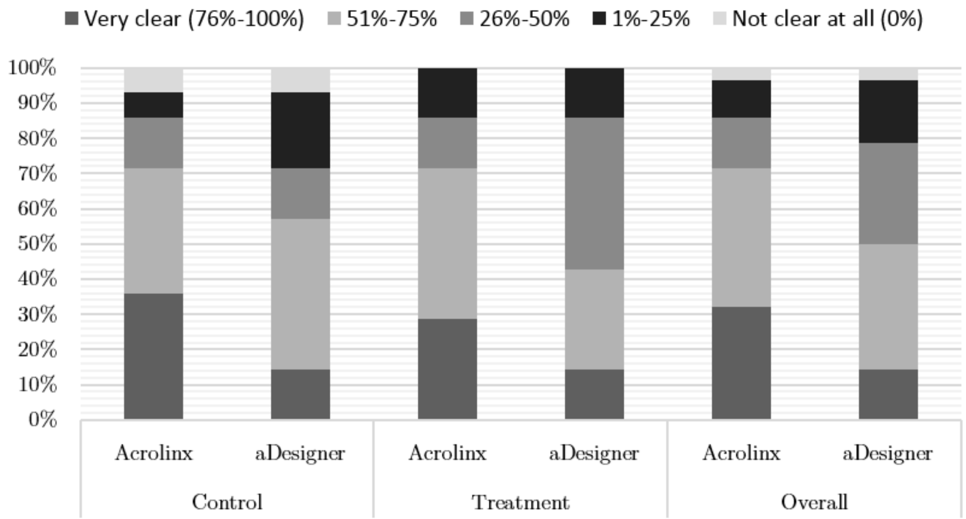


Figure 7.18. Perceived tools' documentation clarity (overall and per group)

Table 7.37. Response count and percent (%) per tool for questions 5 and 16 (D.4): tools' documentation clarity

How would you rate the tool's documentation clarity as regards image accessibility?

	Control		**Treatment**		**Overall**	
Response	**Acrolinx**	**aDesigner**	**Acrolinx**	**aDesigner**	**Acrolinx**	**aDesigner**
Very clear (76%-100%)	5 (36%)	2 (14%)	4 (29%)	2 (14%)	9 (32%)	4 (14%)
51%-75%	5 (36%)	6 (43%)	6 (43%)	4 (29%)	11 (39%)	10 (36%)
26%-50%	2 (14%)	2 (14%)	2 (14%)	6 (43%)	4 (14%)	8 (29%)
1%-25%	1 (7%)	3 (21%)	2 (14%)	2 (14%)	3 (11%)	5 (18%)
Not clear at all (0%)	1 (7%)	1 (7%)	-	-	1 (4%)	1 (4%)

Similarly, participants were asked to judge the tools' specificity and correctness, as understood by Brajnik (2004), with regard to image accessibility. As far as the tools' specificity is concerned, data gathered have revealed that, while the control group estimated that aDesigner was slightly more specific than Acrolinx, the treatment group showed a clearer preference for the latter over the former (see Figure 7.19 and Table 7.38). This might be explained by the fact that the treatment group had been exposed to multiple alt text sample scenarios during the web accessibility webinar and thus were able to better appreciate Acrolinx's specificity.

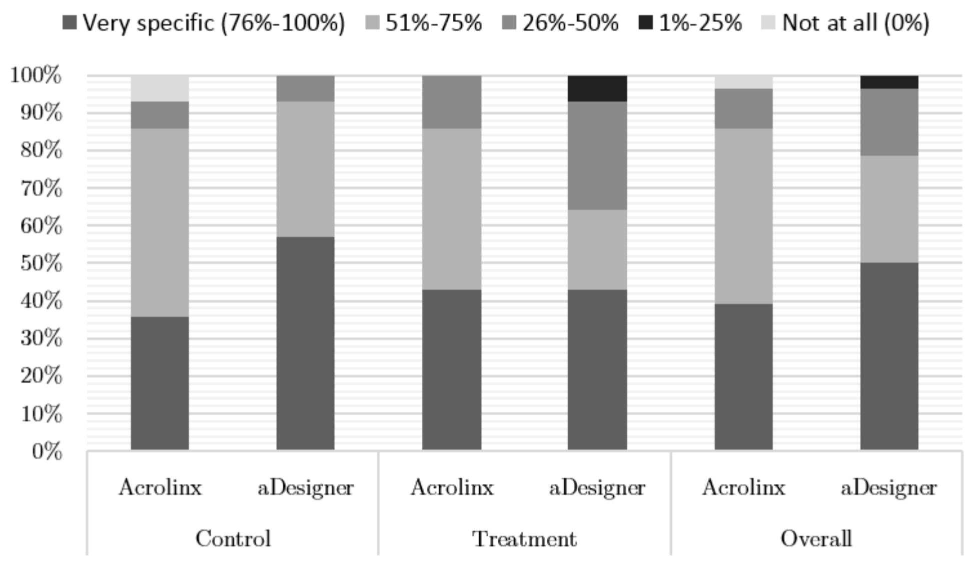


Figure 7.19. Perceived tools' specificity (overall and per group)

Table 7.38. Response count and percent (%) per tool for questions 3 and 14 (D.4): tools' specificity

How would you rate the tool's specificity as regards image accessibility?

	Control		**Treatment**		**Overall**	
Response	**Acrolinx**	**aDesigner**	**Acrolinx**	**aDesigner**	**Acrolinx**	**aDesigner**
Very specific (76%-100%)	5 (36%)	8 (57%)	6 (43%)	6 (43%)	11 (39%)	14 (50%)
51%-75%	7 (50%)	5 (36%)	6 (43%)	3 (21%)	13 (46%)	8 (29%)
26%-50%	1 (7%)	1 (7%)	2 (14%)	4 (29%)	3 (11%)	5 (18%)
1%-25%	-	-	-	1 (7%)	-	1 (4%)
Not specific at all (0%)	1 (7%)	-	-	-	1 (4%)	-

Both groups' opinions seemed to be closer in terms of the tools' correctness, pointing at Acrolinx as the tool which, according to their judgement, reported more image accessibility issues that were indeed true problems (see Figure 7.20 and Table 7.39). One interesting observation that emerged from the data comparison is that, above all, the significant differences with regard to the tools' effectiveness found during the quantitative data analysis are not so clearly reflected in the participants' subjective assessment.

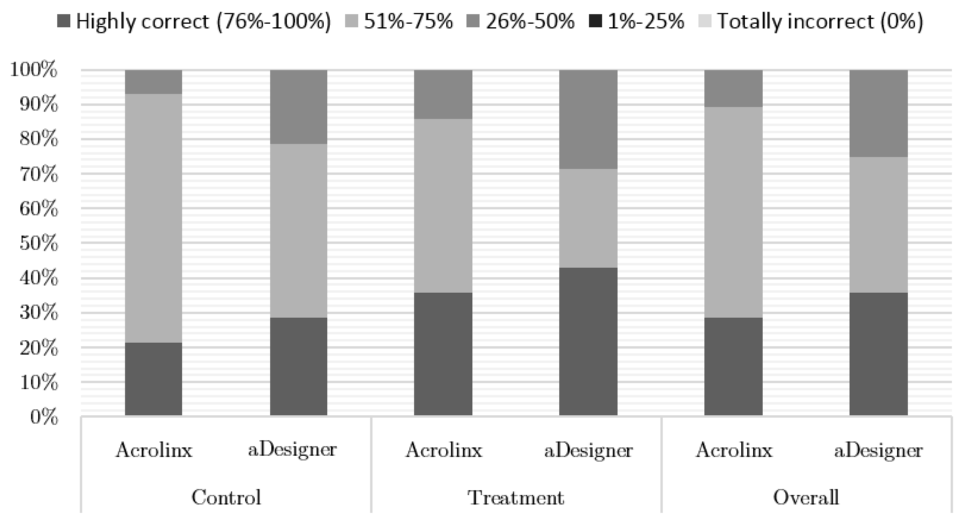


Figure 7.20. Perceived tools' correctness (overall and per group)

Table 7.39. Response count and percent (%) per tool for questions 4 and 15 (D.4): tools' correctness

How would you rate the tool's correctness as regards image accessibility?

	Control		Treatment		Overall	
Response	**Acrolinx**	**aDesigner**	**Acrolinx**	**aDesigner**	**Acrolinx**	**aDesigner**
Highly correct (76%-100%)	3 (21%)	4 (29%)	5 (36%)	6 (43%)	8 (29%)	10 (36%)
51%-75%	10 (71%)	7 (50%)	7 (50%)	4 (29%)	17 (61%)	11 (39%)
26%-50%	1 (7%)	3 (21%)	2 (14%)	4 (29%)	3 (11%)	7 (25%)
1%-25%	-	-	-	-	-	-
Totally incorrect (0%)	-	-	-	-	-	-

In an attempt to offer participants the possibility of freely expressing themselves about the tools, we included a couple of open questions where they could indicate what they had appreciated and disliked the most from each one of them. Acrolinx's most welcomed feature was the rule help, which was mentioned by almost all participants. Ten of them (N=10, 26%) pointed specifically at the specificity of the tool and the clarity of the check reports, and six (N=6, 21%) made reference to the relevance of the examples given in the rule help. Five localisers (N=5, 18%) highlighted that Acrolinx was easy to use and that suggestions offered after running a check were useful. Interestingly, four people (N=4, 14%) placed emphasis on its pertinence for translators or any other content editors with no background on image accessibility. For instance, participant P34 said:

P34: “L'outil explique et illustre clairement les problèmes d'accessibilité concernant les images, ce qui permet au localisateur de visualiser le contenu sous un autre angle

> (ce qui n'est pas négligeable lorsqu'il s'agit de traiter du texte codé manquant souvent de contexte).”

Only one localiser acknowledged that he had not liked the tool at all. Otherwise, according to the participants' subjective assessment, the most negative aspect of Acrolinx was the lack of a ‘WYSIWYG’ mode in which errors could be visualised in context (N=9 out of 28, 32%). In addition, five participants (N=5, 18%) did not like the order in which errors flagged appeared in the report, and one participant (P4) mentioned that he would have preferred to have the rule help in French, instead of English. All in all, participants claimed that they had found the tool interesting, useful and intuitive.

When localisers were asked about what they had appreciated the most from aDesigner, the majority of responses pointed at the possibility of seeing the error flagged in context (N=15, 54%) and highlighted that editing the HTML files within aDesigner's environment would have been ideal. Other positive aspects included the completeness of the tool (N=3, 11%) and the visual disability simulation functionality (N=3, 11%). Isolated comments were made with regard to the reporting of missing `alt` attributes (e.g. P4 indicated: “Cet outil signale les images qui n'ont pas d'attribut `alt`, alors que Acrolinx ne le fait pas”) and the concision of error descriptions (e.g. P25 said: “J'ai aimé la facilité d'utilisation et la manière dont chaque erreur est brièvement décrite”). Nevertheless, this was regarded by half of the participants (N=14, 50%) as the major drawback of the tool.

> P01: “Je soulignerais le manque de clarté des énoncés des problèmes dans les rapports.”
> P52: “Les phrases explicatives étaient très hermétiques.”

Nine participants (N=9, 32%) also reported that they had not liked the tools' interface and that aDesigner was more difficult to use than Acrolinx (e.g. P48 said: “Je n'ai pas trouvé cet outil convivial, il est beaucoup moins intuitif que Acrolinx”). This belief might have been motivated by the fact that, compared to Acrolinx, aDesigner covers a wider range of accessibility issues and one needs to be familiar with the WCAG 2.0 structure to find image-related warnings. Overall, participants considered that using this tool alone was not sufficient to create appropriate text alternatives for images.

While eight localisers (N=8, 29%) explicitly indicated that they preferred Acrolinx over aDesigner because of the clarity of the error descriptions and its more user-friendly interface, more than half of the participants (N=16%, 57%) believed that the joint use of both tools would be more beneficial for image accessibility verification (see Table 7.40). The main reason given was their complementarity. Concretely, five localisers (all of which had followed QA Scenario B) made the same comment about the order in which both tools should be used. For illustrative purposes, we introduce here the answer of one of them:

P21: "Si je m'en tiens à l'expérience des séances 1 et 2 et à la manière dont je conçois le rôle du localisateur par rapport à l'accessibilité, l'apport différencié de l'un et l'autre peut être utile. Leur utilisation serait ordonnée dans l'ordre : vérification avec Acrolinx, puis contrevérification avec aDesigner."

Table 7.40. Response count and percent (%) per tool for question 27 (D.4): tool preference

In your opinion, which tool would be more useful for the localiser to create appropriate text alternatives for images?

Response	Control	%	Treatment	%	Overall	%
Acrolinx	3	21%	5	36%	8	29%
aDesigner	3	21%	1	7%	4	14%
Both tools	8	57%	8	57%	16	57%
None	-	-	-	-	-	-

These data go in line with some of the main conclusions drawn previously in Chapter 6, where we supported that (i) the use of two tools could lead to the production of alt texts of significantly better quality, and that (ii) the aforementioned QA scenario proved to be the most efficient one.

7.3.2.2 Participants' QA practices and impressions about localisation and accessibility

Almost half per cent of the localisers who took part in the experimental study indicated that they were using quality assurance tools on a regular basis (see Table 7.41). Six out of the 12 participants who said they were not using them claimed that they had never had the urge or the need to do so (N=6, 43%), and four people (N=4, 29%) replied that they were working within a more complex localisation workflow and other actors were in charge of implementing quality assurance measures, often mentioning project managers. The remaining participants gave different reasons, including that these tools were only useful for handling high volume translation assignments (N=2, 14%); that their clients did not often request them to use QA tools (N=2, 14%); and that they always relied on their experience to perform revision and proofreading tasks (N=2, 14%). Conversely, irrespective of their current QA practices, almost every participant (N=24, 86%; control: N=12; treatment: N=12) believed that accessibility-oriented QA tools should be used together with other regular QA automated solutions implemented during the web localisation process.

Table 7.41. Response count and percent (%) for question 23 (D.4): use of QA tools

Do you regularly use any quality assurance (QA) tool?						
Response	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Yes, I do	6	43%	6	43%	12	43%
No, I do not	8	57%	6	43%	14	50%
Other	-	-	2	14%	2	7%

The quality assurance post-task questionnaire also lent itself particularly well to gathering participants' opinions about the role of localisers as contributors not only to image accessibility, but also to multilingual web accessibility in general. In order to do so, we first asked them if they agreed with the task that they had been requested to carry out during the study, that is, adapt the image text alternatives initially proposed in the source to improve the accessibility level of the target website.

As shown in Table 7.42, more than two thirds of the participants answered affirmatively to the question (N=22, 79%), and some of them even illustrated their response with examples. For instance, P53 indicated that "[...] s'il manque une description d'image, les localisateurs devraient être autorisés à en ajouter une". Six participants (21%), however, showed a more sceptical attitude. While most considered that rendering images accessible was important, some believed that this task should be explicitly requested by the client. Others argued that, should the localiser have the necessary knowledge to identify accessibility problems, he could inform the client or the web developer about any potential threats to accessibility before taking any action. For example, participants P23 and P59 stated:

> P23: "Je n'avais encore jamais eu à aborder les équivalents textuels sur un plan purement d'accessibilité. Il serait souhaitable, sauf souhait contraire du donneur d'ordre, d'adapter les équivalents textuels. Néanmoins, cela devrait être clairement inclus dans l'étendue des tâches demandées par le donneur d'ordre et ne peut pas être considéré comme tâche standard."

> P59: "Si le client n'a pas fait ces efforts, je ne pense pas que le traducteur doive faire ce travail à sa place. [...] La vérification et la mise en conformité en termes d'accessibilité représentent un travail supplémentaire qui ne doit pas être inclus dans les prestations de traduction de base. Bien sûr, [...] rien n'empêche le traducteur de pointer les manquements aux règles d'accessibilité et de proposer à son client des améliorations qui bénéficieront alors à la fois à la version source et à la version cible d'un site Web."

Table 7.42. Response count and percent (%) for question 2 (D.4): improvement of image accessibility

Do you think that localisers should be entitled to adapt the image alt text initially proposed in the source website in order to improve the accessibility level of the target website?

Response	Control	%	Treatment	%	Overall	%
Yes, I do	10	71%	12	86%	22	79%
No, I do not	-	-	-	-	-	-
Other	4	29%	2	14%	6	21%

In the same perspective, we asked participants whether each language version of a multilingual website should be assessed for accessibility separately. Sixty-four per cent (64%, N=18 out of 28) of the localisers replied affirmatively, acknowledging that each language version should be rendered accessible individually, since content adaptations may have occurred during the localisation process (see Table 7.43). Conversely, five localisers (N=5, 18%) considered that accessibility was universal and, therefore, it could be automatically transferred from the source to the target website. These results contrast with the responses given to the same question by web accessibility experts (see Chapter 2, section 2.5), of whom only 37% believed that localised versions of a website should be checked separately for accessibility. The remaining five localisers (N=5, 18%) relied on the principle of universal design as the key to assure multilingual web accessibility. For instance, P1 stated:

> P01: “Les deux : l'accessibilité devrait être universelle et donc une caractéristique par défaut du site web original (penser à la première version linguistique d'un site web avec l'esprit de l'accessibilité pour tous), ce qui ne dispensera pas d'apporter des changements incontournables dans la version localisée en raison de différences culturels, linguistiques, stylistiques etc.”

Table 7.43. Response count and percent (%) for question 29 (D.4): multilingual web accessibility

Do you think that each language version of a multilingual web site should be rendered accessible separately?

Response*	Control	%	Treatment	%	Overall	%
Yes, I do	9	64%	9	64%	18	64%
No, I do not	2	14%	3	21%	5	18%
Other	3	21%	2	14%	5	18%

*Simplified (see Appendix D.4, question 29, to check options given in full)

The overall positive attitude towards the involvement of web localisers in the achievement of more accessible multilingual websites observed in the participants' responses so far was also reflected in the last three questions of the QA post-task questionnaire. To the question *Which approach should localisers adopt as regards web accessibility?*, only one participant from the treatment group replied that web localisers do not need to address web accessibility issues (option a, see Table 7.44). The highest proportion of responses was registered for option c): “Localisers should keep the same accessibility level, even if it is low, but inform the client or the web developer about the accessibility issues identified” (N=12, 43%). Communication with the client or the web developer was also seen as essential by another nine participants (N=9, 32%), who additionally believed that, instead of maintaining a passive approach when finding web accessibility barriers, they should achieve an acceptable accessibility level in the target website.

Table 7.44. Response count and percent (%) for question 31 (D.4): accessibility approach during web localisation

In your opinion, which approach should localisers adopt as regards web accessibility?

Response*	**Control**	**%**	**Treatment**	**%**	**Overall**	**%**
Option a)	-	-	1	7%	1	4%
Option b)	1	7%	1	7%	2	7%
Option c)	5	36%	7	50%	12	43%
Option d)	2	14%	1	7%	3	11%
Option e)	1	7%	-	-	1	4%
Option f)	5	36%	4	29%	9	32%

*Simplified (see Appendix D.4, question 31, to check options given in full)

Lastly, participants were requested to express their opinion with regard to web accessibility training for web localisation professionals. Over half of them (option c, N=15, 54%) considered that, in order to carry out their job more professionally, localisers should have at least a basic background on what web accessibility entails and how it can be achieved. Others believed that a rather intermediate (option d, N=5, 18%) or advanced (option e, N=4, 14%) knowledge level was needed instead (see Table 7.45).

Although, when replying to question 30, two participants argued that it is web accessibility experts and not localisers who should be ultimately responsible for rendering multilingual websites accessible (option b), and another two believed that it should be a joint effort between the former and the latter (option f), all of them acknowledged that it would be relevant to integrate web accessibility training within the web localisation curriculum. This belief was emphasised by eight participants in their final remarks, who claimed that hands-on labs and more than a one-hour training session would be needed.

Table 7.45. Response count and percent (%) for question 30 (D.4): accessibility-related skills

Do you think that the localiser should have accessibility-related skills to carry out his job in a more professional manner?

Response*	Control	%	Treatment	%	Overall	%
Option a)	-	-	-	-	-	-
Option b)	1	7%	1	7%	2	7%
Option c)	7	50%	8	57%	15	54%
Option d)	2	14%	3	21%	5	18%
Option e)	2	14%	2	14%	4	14%
Option f)	2	14%	-	-	2	7%

*Simplified (see Appendix D.4, question 30, to check options given in full)

7.3.3 Conclusions

The data presented in sections 7.3.1 and 7.3.2 have not only provided an insight into how the 28 participants who took part in our experimental study perceived the localisation and the quality assurance tasks, but have also served to complement the quantitative results reported in Chapter 6, as well as in section 7.2 of the present chapter. The information gathered through both questionnaires has allowed us to confirm that (i) not every participant had identified image text alternatives as translatable elements during the production of translation version T1, especially within the control group, and that (ii) the use of accessibility-oriented QA tools not only boosted their visibility but also contributed to increase awareness with regard to their functionality and purpose. For instance, P23 concluded:

> P23: “Cela a été un exercice intéressant qui m'a permis d'aborder un point de vue différent de l'accessibilité. J'ai eu affaire à bien des situations où le texte alternatif était négligé (ce qui est malheureusement encore une situation au semblant majoritaire), ou bien eu à localiser des infobulles [...] permettant une indexation plus efficace des images par les moteurs de recherche, mais jamais dans l'optique de rendre le site accessible à tous et notamment aux personnes souffrant d'un handicap. J'ai notamment appris l'utilité de l'attribut alt vide pour l'accessibilité par rapport à son absence (qu'il m'arrivait de supprimer si je n'en voyais pas l'utilité et que j'avais la permission de modifier le code, ce que je ne pense plus faire dorénavant).”

Similarly, we have been able to observe that, although localisers showed a clear preference for Acrolinx over aDesigner because of its comprehensive error descriptions and its user-friendliness, the joint effectiveness of Acrolinx and aDesigner was acknowledged by the majority of the participants, irrespective of the group they belonged to. In addition, both tools were regarded as a suitable complement to other QA automated solutions which may be implemented during the web localisation process.

Localisers also seemed to agree with regard to their role as contributors to multilingual web accessibility. Most advocated for the achievement of web accessibility by both web developers and localisation professionals, and believed that informing the client about any accessibility barriers identified during the localisation process was equally crucial for that collaborative effort to be more productive. Participants did not disregard the possibility of implementing accessibility best practices in the target product themselves, although the task was seen by some of them as potentially time-consuming. In fact, certain localisers claimed that time constraints imposed during the experimental study prevent them from looking more carefully at the use of null alts or introducing missing `alt` attributes, which could have led to better alt text quality results.

The webinar attended by the treatment group was considered useful, instructive and pertinent for web localisers. The belief that web accessibility training could be beneficial for web localisers to perform their job more professionally was also acknowledged by the participants who had not followed the webinar prior to the study. Overall, all appeared to embrace the proposal of introducing web accessibility to localisation trainees and the need for localisation professionals to have at least basic knowledge on the subject, regardless of the complementary use of accessibility-oriented QA tools. This findings complement and strengthen those from the first survey presented in this thesis (see section 2.5, Chapter 5), where WA experts considered that localisers should be held accountable for the level of accessibility achieved in multilingual websites.

7.4 Summary

This chapter aimed to present the additional findings of the central experimental study of this thesis. More specifically, we sought to expand on the main conclusions we had drawn in Chapter 6 with regard to the appropriateness of localised alt texts, which were based on the analysis of how this property was influenced by the manipulation of three independent variables during our web localisation experiment: (i) the WA knowledge of the participants, (ii) the use of accessibility-oriented QA tools, and (iii) the QA scenario followed, which was, in turn, dependent on the tool order proposed. To this end, in section 7.2, we looked into the effect of the other six secondary independent variables shown in Table 6.1 (see Chapter 6) on the scores assigned by blind users to the alt texts produced by the localisers (*1. Not appropriate, 2. Acceptable, 3. Pertinent, 4. Very pertinent*). All the evidence collected has allowed us to provide further insight into research questions **R2** and **R3**.

Research question **R2** asks whether the use of accessibility-oriented QA tools during the web localisation process results in more appropriate text alternatives for images. Based on the hypothesis testing, in Chapter 6 we concluded that such tools are indeed crucial for these professionals to achieve an acceptable degree of image accessibility in the target website, but that different levels of appropriateness are reached depending on the tool used (Acrolinx or aDesigner) and on the order in which they are introduced during the QA localisation phase. In this chapter (section 7.2.1), we have observed that the effect of using these tools on the quality of localised alt texts also varies depending on the nature of the alt texts contained in the source website.

When the original alt texts are inappropriate, it appears that Acrolinx helps localisers to achieve better appropriateness results, both when it is used alone and after aDesigner. The effectiveness that Acrolinx showed in identifying problematic text alternatives relies on ten out of the 40 CL rules developed within the framework of this thesis to automate the alt text verification process. Our study has demonstrated that the individual application of 8 out of these 10 rules (especially those aimed at improving alt texts for functional images) leads to higher quality text alternatives in French. When it comes to images which had an empty `alt` attribute in the source website, the individual use of accessibility-oriented QA tools proved to be insufficient, and only the combination of Acrolinx followed by aDesigner seemed to result in more appropriate alt texts, although the improvements observed were not significant. The opposite occurred in the case of images which lacked an `alt` attribute in the source web pages: using Acrolinx, aDesigner or both tools contributed to enhance the image accessibility level of the localised websites. Finally, we showed that, when localisers receive a website with appropriate text alternatives, tools do not help to significantly improve their quality, nor do they lead to worse alt texts.

Another factor that appeared to influence the performance of the two tools tested with regard to the achievement of appropriate text alternatives is the error repair guidance they provide (see sections 7.2.4 and 7.2.5). While localisers who consistently checked Acrolinx rule help to try to solve the issues flagged by the tool reached significantly better results in terms of alt text quality, those who relied on aDesigner error descriptions saw a significant downgrade in the image accessibility level of the website they were working on. These findings reinforce the need for giving importance not only to tools' completeness, correctness and specificity, but also to the comprehensiveness and usefulness of the guidance they provide to help evaluators understand the meaning of the problems detected and how they can be solved.

Research question **R3** aimed to investigate whether having some background on WA helps localisers to produce more appropriate text alternatives for images. One of the key results reported in Chapter 6 was that, when no accessibility-oriented QA tools are used, WA knowledge matters. In section 7.2.2, we showed that HTML skills are equally important in order to achieve acceptable levels of image accessibility in localised websites. In addition, the analysis of the data we gathered through the localisation and QA post-task questionnaires has revealed that, although using accessibility-oriented QA tools did have an impact on the appropriateness of alt texts, the use of CAT tools did not affect the overall alt text quality.

In this chapter, we also reviewed localisers' opinions on Acrolinx and aDesigner, their overall performance during the two tasks requested, and their impressions about the introduction of WA best practices during the web localisation process. Although subjective in nature, the information collected supports both our theoretical arguments towards the consideration of the needs of users with disabilities during web localisation, and the evidence gathered through the survey about multilingual web assessment practices for accessibility (see Chapter 2, section 2.5) and the localisation experiment itself. Their views can be summarised as follows: While Acrolinx was thought to provide more comprehensive guidance about how to ensure image accessibility, the combined use of Acrolinx and aDesigner was regarded as more efficient to achieve such an endeavour. Additionally, the one-hour webinar offered on WA best practices for localisers was found to be pertinent, interesting, useful and inspiring, although only a half of the participants considered that it had helped them to render their websites more accessible, at least within the particular context of the study, during which restrictions in terms of time were imposed. Above all, we consider it relevant to highlight that almost 80% of the localisers thought that they should be entitled to improve the image accessibility level of the source, and that accessibility considerations should be part of the basic training in web localisation.

Chapter 8

Conclusions and future research directions

This last chapter is dedicated to offering a summary of the investigation presented in this thesis, highlighting its main results and framing the relevance of our research for the fields in which it resides. In addition, we outline the limitations of our work and present the future research paths that can be followed to complement and expand our findings.

8.1 Overview of the research

With this thesis, our intent has been to underscore the scarce attention paid to the multilingualism of the Web in prior web accessibility research and, additionally, the low level of importance attributed so far by both web localisation practitioners and scholars to the needs of people with disabilities —one of the population groups within the target audience. More specifically, by adopting an interdisciplinary perspective and a multimethod research approach, our work set out to investigate the role of localisers in the achievement of more accessible multilingual websites for blind users, placing special emphasis on ways to assure access to one type of digital content that particularly poses more challenges for this community: images on the Web.

In this investigation, our departure point —or general research hypothesis— has been that localisers do indeed have a crucial role to play in enhancing the multilingual web browsing experience of blind individuals. By understanding localisation as a process whose ultimate goal is to ensure that a website that was originally available in only one language is transformed into a properly functional website (linguistically, culturally and technically) with at least two different language versions, we have sought to demonstrate that localisers should be held accountable for the accessible rendering of the localised web content, together with the other actors involved in the multilingual web development cycle. In order to do so, we have relied both on theoretical arguments (grounded on the related literature and software review) and on evidence-based foundations. In what follows, we summarise the former —mostly covered in Chapters 2 and 3— as well as the methodological framework to obtain the latter —outlined in Chapter 1 and elaborated primarily in Chapters 4 and 5. This will later bring us on to the presentation of this thesis' achievements in section 8.2, which covers the main findings presented and the conclusions drawn in Chapters 6 and 7.

Summary of supporting arguments

Our investigation has been grounded on the belief that localisation professionals should contribute to the efforts being devoted worldwide to endorse the W3C Web for All principle because embracing the accessibility recommendations covered in the Web Content Accessibility Guidelines (WCAG) 2.0 is not only a question of social responsibility, but also a beneficial strategy at multiple levels: technical, political, legal and financial. In addition, as we have pointed out on several occasions throughout the thesis, accessibility is often seen by both scholars and practitioners as a joint effort. If further accessibility-targeted actions were taken by all people participating in the production of web content, overcoming the major obstacles still faced today during the implementation of the aforementioned guidelines and their evaluation would be a less complicated and more sustainable endeavour.

In this work, we looked at the nature of these obstacles by considering (i) the accessibility of *websites* as a whole and (ii) the more specific case of *images*, with a view to suggesting how the involvement of localisers in the accessibilisation of multilingual websites could prove helpful. Apart from the typical limitations related to the web development process, such as lack of time, funding or resources, we realised that current barriers to the achievement of both accessible websites and images are, overall, closely linked to two main dimensions: (i) the experience (in broad terms and regarding web accessibility) of those who create web content and eventually check it for accessibility issues and (ii) the support offered by technologies to render content accessible.

As far as *web accessibility* assurance is concerned, we have seen that understanding WCAG 2.0 guidelines and success criteria (SC) requires not only instrumental and technical skills, but also language-related expertise, which is not common among web developers and designers. Furthermore, prior work had revealed that web accessibility evaluation (WAE) tools do not currently feature enough WCAG 2.0 coverage, nor do they offer comprehensive guidance for those who lack the necessary training on how to implement each SC. Obstacles in accessibility automation often emerge from those SC which bear some degree of subjectivity or are related to linguistic aspects. In this context, we argued that provided localisers have some web accessibility (WA) knowledge, their multifaceted profile could be of added value to address those particular aspects of the WCAG 2.0.

When the scope was narrowed down to *image accessibility* assurance, the literature confirmed that most difficulties were associated with the formulation of image text equivalents. Writing appropriate alt texts taking into account images' communicative value (descriptive, functional or decorative) is essential to achieve higher levels of image accessibility. However, these are not easily attained since multiple web content producers (e.g. web developers, web designers) do not always have the analytical and language skills needed or the time required. By reviewing the literature, we found that this problem is often exacerbated by the lack of support from which they suffer, both in terms of a complete (and well documented) set of guidelines about which preferred lexicon and syntactic structures to be used during alt text formulation, and of automated software capable of identifying inappropriate text alternatives. In this sense, we argued that image accessibility problems in source websites may be transferred to their localised versions (i) if localisers are not aware of the functionality of alt texts, which would probably result in overlooking the fact that they are translatable elements, or (ii) if they do not have the necessary knowledge and/or tools to assure that the text alternatives they produce are indeed appropriate according to the purpose of the images they are associated with. Given the flaws of current WAE tools and state-of-the-art image accessibility-oriented technology, we suggested that localisers could leverage the capabilities of controlled language checking software, which is already used in multiple localisation contexts for varied purposes, including the enhancement of source text translatability and the facilitation of the machine translation post-editing process.

Summary of the methodological framework

Taking into account all of the above, at the beginning of our investigation we established three concrete goals that would ultimately allow us to better understand whether our arguments could be sustained:

- **Goal 1.** *To investigate whether established procedures exist to ensure the accessibility of multilingual websites in general*, with a view to providing insight into how they could be more smoothly integrated into the localisation workflow.
- **Goal 2.** *To propose an automated testing solution based on controlled language (CL) technology for assuring the generation of appropriate text alternatives for images on the Web during the localisation process,* with a view to covering some of the limitations of existing guidance and tools designed for image accessibility evaluation.
- **Goal 3.** *To determine the extent to which localisers are capable of assuring that an acceptable level of image accessibility is achieved in the target web product*, with a view to identifying which factors help them to achieve such an endeavour or, on the contrary, hamper the task.

In order to reach these goals, we adopted an empirical investigation approach, for which we triangulated different research strategies (a survey, the design and creation of an IT component and a controlled experiment) as well as data generation methods (questionnaires and web documents), as depicted in Figure 1.6 (see Chapter 1). By following this kind of approach, we have gathered evidence from multiple sources of data, thus contributing to strengthen the soundness of our findings. Specifically:

a) We assembled and exploited four different web corpora (see Appendix A) comprising 152 websites with over 12,000 pages, with the number of images contained therein amounting to more than 55,000;
b) we surveyed 67 web professionals with at least two years of experience in web accessibility from almost every continent;
c) we recruited 28 localisers working with the language pair English-French, asked them to perform a localisation task followed by a QA task under different conditions, and analysed the resulting 252 localised web pages (with a total of 10,920 images) for image accessibility;
d) we surveyed these localisers on a varied number of factors related to the whole localisation process they followed; and
e) we recruited 7 blind users and analysed their judgments on the appropriateness of the alt texts produced by localisers for the aforementioned amount of images (i.e. a final evaluation data set of 76,440 records).

8.2 Achievements

In broad terms, the exploitation of the multiple sets of data indicated above and the statistical analysis undertaken have contributed to extend our knowledge of why and how accessibility should be considered during the web localisation process. This is, in fact, one of the key achievements of our investigation, since it represents the first empirical research work that has studied more in depth the convergences between localisation and accessibility illustrated in Jiménez Crespo's map of Localisation Studies (see Figure 1.5, Chapter 1).

We believe that, despite the limitations of our investigation (see section 8.3), overall we have attained the goals initially set forth. In what follows, we will present our key findings (section 8.2.1), outline their impact and point out what their added value is for the research areas which have delimited the framework of our thesis (section 8.2.2).

8.2.1 Summary of main findings

Goal 1

Results related to Goal 1 are derived from the study of the literature and the survey described in Chapter 2, section 2.5.

- *The W3C does not offer any methodology or recommendations about how to achieve accessibility in multilingual websites.*

Existing guidance on how to implement accessibility best practices or check web content to see whether these have been followed does not indicate how to proceed in the case of localised websites in this regard. The W3C only acknowledges that if the different versions of a multilingual website can be used independently from one another (e.g. when they are available under different URLs), then each version should be assessed for accessibility individually.

- *WA experts do not follow a standardised procedure when checking multilingual websites for accessibility issues.*

Web professionals with WA expertise believe that the ideal solution to guarantee that multilingual websites are accessible is to ensure that every language version satisfies the W3C accessibility standards. Yet they acknowledge that, in real life situations, this almost never occurs due to lack of time, human resources and lack of additional language skills. As a result, it seems that although it is seen as undesirable, the common practice is to focus only on one language version and assume that the others have a similar level of accessibility. In addition, it appears that culture-embedded elements and textual content do not normally receive too much attention during WA audits, despite the fact that these are particularly important in the case of localised websites.

- *WA experts agree that accessibility should become a primary goal for localisation practitioners.*

Professionals involved in the web development cycle share our view that localisers are to be held accountable for the accessibility level of the content they produce, to the same extent as designers and developers and even more than webmasters. Similarly, the accessibility community believes that localisers' complete background and skills can be of particular benefit during the WA assessments carried out during the last QA phase before a multilingual website becomes operational.

Goal 2

Results related to Goal 2 include the outcome of the design and creation research strategy adopted (see Chapter 4) and the findings from its evaluation.

- *A 40 CL rule set aimed at identifying inappropriate alt texts in French was developed using Acrolinx technology.*

Based on real examples of high-quality and low-quality text alternatives, we defined three sub-sets of guidelines to ensure the appropriateness of alt texts in French for images with a descriptive value (sub-set A, 14 rules, see section 4.4.1, Chapter 4) and a functional value (sub-set B, 18 rules, see section 4.4.2, Chapter 4). Additionally, we developed 8 rules to identify image text equivalents containing uninformative text strings (sub-set C, 8 rules, see section 4.4.3, Chapter 4). These rules proscribe the use of certain terms and expressions and suggest which syntactic structures are more suitable for each type of alt text, depending on the communicative function of the image. These guidelines were formalised using Acrolinx's state-of-the-art technology, which also offers a client to automate their application in HTML files: Acrolinx IQ Batch Checker (2.7.0).

- *For each rule developed, comprehensive guidance was created to help users understand why a given alt text is not appropriate and how it should be improved.*

When the 40 CL rules are applied using Acrolinx and errors are flagged, users can consult the rule help to better interpret the problem identified. Specifically, each rule help file we developed includes a brief introduction to image accessibility, an explanation of why a certain alt text is not appropriate, examples that support that explanation and recommendations on how to amend the text flagged. Rule help files also feature links to official documentation provided by the W3C. An example is shown in Figure 4.4, Chapter 4).

- *From the 10 CL rules evaluated, we confirmed that at least 8 lead to a significantly higher level of appropriateness in localised alt texts.*

When the source website contains inappropriate text alternatives, the likelihood of producing appropriate alt texts in the localised version is between twice and three times higher when the rules we developed are applied. Particularly, rules aimed at improving the appropriateness of alt texts for functional images, which should describe the action they trigger upon interaction with the user, lead to significantly better results in terms of alt text quality. The significant effect of most of the rules tested is stronger when Acrolinx is the only tool used to try to boost the appropriateness of localised text alternatives, but this is generally maintained when the application of the rules is done after performing checks with another tool as well. The two out of the ten rules that did not prove to be as efficient were related to the insertion of image copyright information in the alt text and the use of redundant expressions such as “image of...” or “An image that shows...”

- *Overall, a CL-based checking solution can offer better support in the achievement of image accessibility in localised websites than a general WAE tool.*

The proposal of adopting a controlled-language based approach to assure high levels of image accessibility was motivated by the limitations of current tools used for that purpose, including their inability to distinguish appropriate from inappropriate text alternatives or to validate text in languages other than English. Our work has confirmed that following such an approach during the web localisation process can result in significantly more appropriate French alt texts in the target website than when a general WAE tool is used. Furthermore, our proposal proved to be significantly beneficial not only in terms of detection of inappropriate alt texts, but also with regard to the error repair guidance offered. When localisers consulted the rule help files alluded to above, the quality of their alt texts increased significantly.

Goal 3

The last goal of this thesis was attained by giving the answers to three specific research questions. Here we sum up the main findings of the hypothesis testing process, outlined in Table 6.22 (Chapter 6), together with the results of the additional analyses presented in Chapter 7:

R1. *Are image text alternatives considered by localisers as translatable elements during the web localisation process?*

Contrary to expectations, alt texts are not systematically translated by localisers. We observed a noteworthy degree of variability with regard to the awareness of the existence of these text elements and the need to translate them, both in the experiment data and in the subjective data gathered through the questionnaires we administered to localisers. It was interesting to note that those who had some WA background were able to identify and translate a significantly higher number of alt texts than localisers who had not received any previous training on accessibility concerns. In addition, the experimental study we conducted revealed that localisers do not feel comfortable enough to add `alt` attributes to `<img>` elements which did not have any in the source text. When it comes to the treatment of empty `alt` attributes during the localisation process, localisers appear to be keener to propose new null alt texts (i.e. transform full text alternatives found in the source website into zero-length alt texts) than reconsidering the appropriateness of the empty `alt` attributes already present in the source.

Our investigation has also shown that the use of accessibility-oriented QA tools facilitates the identification of alt texts as translatable elements. We found that those localisers who had not touched any alt text in the first translation version they delivered were then able to spot them during the QA process. More specifically, we observed that Acrolinx triggered, overall, a higher number of alt text edits during the QA phase than aDesigner, the general WAE tool. In fact, for localisers without any WA knowledge, the former proved to be significantly more effective than the latter in this regard. Similarly, we noticed that for this group of participants, the combined use of both tools led to modifications in twice the number of text alternatives if compared to the translation version they had produced before the QA task.

R2. *Does the use of accessibility-oriented QA tools during the web localisation process result in more appropriate text alternatives for images?*

One of the key findings to emerge from our investigation is that the use of automated support can help localisers to produce a significantly higher level of image accessibility in the target website than if no QA technology is used. We have been able to corroborate that this occurs not only when two tools with different capabilities are combined, but also when localisers only have one at their disposal (Acrolinx or aDesigner). A more refined analysis of the data revealed that NLP-based software like Acrolinx performs significantly better when used alone than WAE tools like aDesigner, which do not offer a high level of specificity with regard to image accessibility assessment. Only when aDesigner is used after Acrolinx are significant results obtained. From these facts we inferred that the best solution to assure the appropriateness of localised alt texts is to rely first on Acrolinx, which focuses more on linguistic correctness, and then use aDesigner, which helps to quickly identify missing alt texts and the most common uninformative ones.

In addition, we observed that the use of accessibility-oriented QA tools proved effective in improving alt text quality when the source text alternatives were inappropriate, but it did not boost nor degrade the level of appropriateness of those alt texts which were originally conformant with accessibility standards in the source website. Similarly, we deem it important to note that, while automated support during the QA task proved to have a positive impact on the final level of image accessibility achieved, the use of translation software (notably of CAT tools) seemed to have no significant effect.

R3. *Does knowledge of web accessibility help localisers to produce more appropriate text alternatives for images?*

As in the case of alt text translatedness, we found that having some background in WA led to significantly better results, this time in terms of alt text quality. Our data indicates that following a one-hour training session on accessibility best practices has a strongly significant impact on the appropriateness of the text alternatives produced by localisers. However, one of the more striking findings of our investigation is that this advantage can be leveraged by localisation professionals with presumably no WA background when the adequate tools are used. Indeed, our study indicates that when Acrolinx recommendations are followed by both localisers with and without WA knowledge, differences across groups with regard to alt text appropriateness are not significant. In addition, when the QA scenario Acrolinx-aDesigner was applied, we showed that localisers who were not aware of accessibility issues achieved slightly better results. Apart from the localisers' WA background, we also took into account their HTML skills when analysing the level of image accessibility they managed to reach in the target website. When contrasting data from participants, we were able to observe that, in fact, good knowledge of this markup language helps localisers obtain significantly more appropriate text alternatives for images.

8.2.2 Research impact and contributions to related fields

Taking the above findings into consideration, as well as the other additional conclusions progressively drawn throughout the thesis, it could be stated that the need for localisers to embrace accessibility that we had initially claimed has been endorsed not only by the subjective opinions of the web professionals (web designers, web developers, accessibility consultants and localisers) we surveyed, but also by the objective data we collected from our experimental study. At a more specific level, our investigation has made several noteworthy contributions to the fields of Accessibility, Controlled Language and Web Localisation.

Accessibility

Our research work adds to the growing body of literature investigating effective and sustainable ways of assuring a more inclusive Web for All. The novelty of our investigation resides in the fact that (i) we have studied the behaviour of a group of web professionals who had not been considered, to date, as potential contributors to the aforementioned universal access goal, and that (ii) we have demonstrated that their actions can indeed have an impact on the overall accessibility of the Web. In addition, we have opened a new line of discussion within Web Accessibility Studies by bringing the distinction between monolingual and multilingual websites to the forefront. From an accessibility perspective, we believe that the latter are considerably more challenging than the former, not only because content is available in different languages and web elements may vary across versions, but also because their production involves a higher number of actors, with different skills and perceptions about the needs of people with disabilities. The insights gained from our research with regard to the lack of standardised procedures to assure multilingual web accessibility have shed light on the urgent need to develop new WA evaluation methodologies for this particular type of websites if the evaluator effect we have often referred to in the thesis is to be reduced.

At an image accessibility level, our work may have different practical applications. Firstly, we have demonstrated that the ineffectiveness of current automated solutions can be covered by looking into the capabilities of software from other areas of expertise, such as NLP. Based on the benefits we have highlighted thus far with regard to the use of Acrolinx (and its combination with tools based on markup syntax checks), we could suggest the integration of controlled language checkers within accessibility evaluation workflows. Secondly, the corpus-based study that has led to the development of CL rules has also helped us to understand that, in fact, appropriate alt texts follow similar patterns and that these can be transformed into more language-focused guidelines for writing high quality alt texts. While we have identified those more frequently followed in French —which may be certainly useful for web professionals who need to create text alternatives in that language—, they may also serve as a guide for content producers working with other languages, as they can help practitioners understand the kind of formulations mostly preferred by blind users depending on the image communicative value.

Controlled Language

Our investigation could also be regarded as innovative from the perspective of Controlled Language research. It represents the first study that has investigated the applicability of CL software for accessibility purposes and our findings at least look promising. In the particular case of French, we have shown that certain accessibility-related aspects can indeed be formalised under relatively simple CL rules. This may contribute to draw the attention of CL researchers into the accessibility domain, who may seek to examine whether other text-based accessibility issues can be addressed following a similar approach. Furthermore, CL applications in localisation studies have normally focused on different language-related quality dimensions of the target product, such as its readability or its translatability, which overall could only have an effect on the accessibility of textual content. Through our study, we have shown that controlling how text is produced can also facilitate the access to visual information in a web document.

Web Localisation

Last but not least, we believe that our research work can have direct implications on future localisation best practices. By providing the first scientific account of how accessibility is addressed during the localisation process, we have been able to observe that awareness about the needs of people with disabilities is still relatively low within the localisation community. Accessibility implementation is not yet regarded as a ‘must do’ —as we have illustrated with the particular case of adaptation of text alternatives— which can bring negative consequences to the overall accessibility level of the multilingual web. In contrast to this last statement, we have also shown that, with little action on the side of localisers, the opposite can occur as well. Concretely, our findings point to the need for further empowering the localiser to take part in the web development cycle. It seems that they do not feel confident enough to make changes to the original source code, even if it is for the sake of facilitating the access to localised content by a wider variety of users.

Similarly, our research has contributed to the field of Localisation Studies by reporting the first empirical study where the source product requested for localisation was composed of HTML documents, and by bringing in end users with disabilities as reviewers or experts to assess the localisation output quality. Moreover, the urge for localisers to pay more attention to accessibility concerns, which was recognised by experts from the web accessibility community and reflected in the results of our empirical study, has provided interesting insights into the consideration of accessibility as a localisation quality indicator. This could have an impact on three specific dimensions: (i) *the localisation QA process:* our findings lead us to believe that an accessibility evaluation phase should be introduced as part of the localisation QA process, by relying both on automated and manual testing techniques; (ii) *components of the general localisation competence:* apart from the traditional instrumental, technical and linguistic skills, our research has pointed at other competences that could prove advantageous for the achievement of fully functional localised websites, such as understanding how accessibility standards are implemented and how people with disabilities use assistive technologies to perceive and interact with the Web; and (iii) *localisation training:* our investigation has shown the positive effects of accessibility awareness and the use of accessibility-oriented tools, thus suggesting that accessibility-related aspects of the web development cycle could be of added value if introduced to the localisation curriculum.

8.3 Limitations and future work

As the first large study about the accessibility of multilingual websites that has attempted to examine the links between the three aforementioned areas of research, this thesis has seen its scope reduced in terms of the type of web content investigated and the languages covered. In this sense, we have looked more in depth into image accessibility issues, providing only occasionally examples of other shared concerns between accessibility and localisation. The choice of images as the central object of study inevitably led to restricting our main interest to the blind community. As far as language coverage is concerned, although in our more theoretical chapters most of our arguments have been language-independent, our empirical work —both the development of CL rules and the localisation experimental study— has mainly focused on the English-French language pair. These two broad limitations, together with the relatively small samples of the populations we studied (both animate –WA experts, localisers, end users– and inanimate –images, alt texts, web pages), could be seen as arguable weaknesses of our research.

Throughout the thesis, we have tried to describe the measures taken to compensate for these general limitations, particularly with a view to boosting the internal and ecological validity of our studies, as well as to guaranteeing their reproducibility. Although, while doing so, we have also regularly pointed to the limitations associated with such measures, what follows is a review of the major shortcomings of the approaches taken to attain our three concrete goals. This last section of the chapter will also serve to put forward possible future research directions.

Goal 1 was addressed from a rather general perspective. The fact that we have only relied on the study of the literature and the subjective input of web accessibility experts could make our findings less generalisable. Further studies based on direct observations could be carried out in order to more closely examine the accessibility assessment practices followed by these professionals when dealing with websites with at least two different language versions. In addition, it would be interesting to compare the policies and legislations covering web accessibility issues in countries with more than one official language against those of monolingual countries, with a view to understanding whether internal methods or guidelines exist to ensure that information of public interest is accessible for speakers of all languages used within national territories. Our findings have also raised important questions regarding the current state of multilingual websites in terms of WCAG 2.0 conformance. In this respect, studies should be undertaken to find out whether different levels of compliance are achieved per language version, as well as to estimate the potential impact that these differences could have, for instance, in the browsing experience of bilingual end users. Notwithstanding these limitations, we see our general report about multilingual web accessibility assurance as a well-motivated starting point for further investigations by both the accessibility and the localisation research communities.

In order to reach **Goal 2**, certain compromises had to be made. The first one was linked to the nature of the web corpus chosen to feed the CL rule development process. We are aware that the type and amount of images used in a website may vary considerably across genres. In this sense, the formalisation of language-based guidelines on how to write appropriate alt texts could have proved to be more challenging if, for instance, tourism or commercial websites were taken as a baseline, instead of governmental and public sector websites –particularly because the latter usually bear more functional images, while the former are regularly expected to contain a high amount of images with a descriptive value. To ensure the portability of the rules developed, further studies could assess how well our guidelines assist evaluators in image accessibility audits within these other website types. Precisely related to the rule evaluation approach adopted, it is worth noting that another limitation of our study was that the measurement of the rules' effectiveness was solely based on the quality of the final localised product. We believe that a more refined analysis of the data could lead to interesting findings regarding the usability of each rule depending on the WA background of the person that applies it. In the same vein, the study of Acrolinx check reports could shed light on new structures or linguistic patterns that we may not have covered in our rules and that might have been spontaneously proposed by localisers. Last but not least, it would be pertinent to explore whether Acrolinx and its rules register the same level of performance when used by other web professionals with a more technical-oriented profile, such as web developers or designers.

Finally, a number of limitations should be considered when interpreting the results associated with **Goal 3**. Some of these are related to the experimental design itself. For example, guided by the need to produce simple HTML documents to facilitate the evaluation of as many image accessibility issues as possible, the nature of our experimental website was mainly static (or semi-static if we take into account the introduction of some basic JavaScript code). In the future, it would be important to also explore whether the transfer and achievement of image accessibility is affected by different factors when localisers need to deal with dynamic websites. Similarly, given the large amount of data collected, it was decided to mainly adopt a quantitative approach to operationalise the translatedness and appropriateness of text alternatives. While this is, at the same time, one of the strengths of our investigation —as it has allowed us to apply advanced inferential statistics and report on the significance of our results— a greater focus on qualitative aspects of the localisation output would be desired to expand on our conclusions about the level of image accessibility achieved by localisers in the target product (e.g. to obtain a more fine-grained definition of what is understood as alt text quality).

The profile of the localisers recruited might also have been a source of bias, since the majority of them were working as freelancers. One could hypothesise that in-house localisation teams within large corporations may be better trained and pay more attention to accessibility issues (maybe as a result of their companies' interest in the benefits of accessibility, legally and financially). In this sense, it is possible that if expert evaluators with a localisation background had participated in the study, our results with regard to the importance of having WA knowledge may have turned out stronger, or the positive effects of using accessibility-oriented QA tools less conclusive. Observing the behaviour of such a group could be seen to be of added value in the future to further explore the phenomenon of the evaluator effect in the context of multilingual websites. Nonetheless, we believe that our study and its findings are still relevant, particularly if we take into account current increasing trends in the localisation industry towards outsourcing. In addition, our experimental design has also allowed us to discover the didactic potential of the aforementioned types of tools which, as we noted earlier, could prove beneficial in raising accessibility awareness during the localisation QA process.

The above is directly linked to another limitation of our investigation which, at the same time, holds the highest potential for future work: little attention was paid, in general, to the localisation process-oriented data we collected. In our opinion, the screen recordings in particular will be a valuable source of evidence for continuing our research on localisation and accessibility. Based on the insights from these videos (e.g. regarding the amount of time dedicated to inspecting the visual rendering of the website while translating or the strategies followed to address problems related to the source code), further research work could attempt to determine how accessibility best practices could be more smoothly integrated within the localisation workflow, thus avoiding disruptive redesign efforts at later stages in the multilingual web development cycle. The inferences drawn from the analysis of these recordings could be then contrasted and expanded, for instance, with the output of other methods typical of process-oriented research, such as eye-tracking studies.

We expect that these additional analyses will also assist web accessibility training for future localisation professionals. While the relevance of acquiring WA knowledge and know-how was generally welcomed by the students who took part thus far in our seminars on the topic and our pilot teaching materials proved to be successful (Rodríguez Vázquez 2014), the selection of the content taught was not based on scientific evidence. Furthermore, discussing WA as a separate topic within the localisation course may be regarded as an ineffective pedagogical technique. To encourage 'accessibility thinking' among localisation trainees, we argue that WA should be integrated throughout the full curriculum, as it has already been done in computer and information science education (Waller et al. 2009; Wang 2012). In this sense, our future work could be grounded on the pedagogical framework we are currently developing, which advocates for a communicative, object and social approach to localisation teaching (Torres del Rey et al. 2014):

> "Localisation students must learn to play a three-fold role, vis-a-vis the digital product and the production process, as: **mediators** of the communicative value generated by the digital product as a cultural object and as a technical extension of human physical and intellectual abilities; **stakeholders** in the distribution of the localisable object's functional and informative values; and **negotiators** of their own (as professionals) and of their localisable technological product's social position and responsibility."

For, you see, so many out-of-the-way things had happened lately, that Alice had begun to think that very few things indeed were really impossible.

In Lewis Carrol, *Alice's Adventures in Wonderland*

References

Note: Due to the large number of links provided, the last access date (30th March 2016) applies to all links.

Abou-Zahra, Shadi, ed. 2005. 'Selecting Web Accessibility Evaluation Tools'. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/eval/selectingtools.html.

———. 2008. 'Web Accessibility Evaluation'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 79–106. Human-Computer Interaction Series. London: Springer-Verlag.

Access for All. 2011. 'Étude 2011 Sur L'accessibilité Des Sites Web Suisses'. Zurich: Access for all. http://www.access-for-all.ch/ch/publikationen.html.

Acrolinx GmbH. 2011. 'Acrolinx Linguistic IDE'. *Acrolinx tool documentation.*

———. 2012. 'Acrolinx IQ Batch Checker User Guide (Version 2.7)'. *Acrolinx tool documentation.*

Aikawa, Takako, Lee Schwartz, Ronit King, Monica Corston-Oliver, and Carmen Lozano. 2007. 'Impact of Controlled Language on Translation Quality and Post-Editing in a Statistical Machine Translation Environment'. In *Proceedings of Machine Translation Summit XI*, 1–7. Copenhagen (Denmark): European Association for Machine Translation. http://research.microsoft.com/apps/pubs/default.aspx?id=69483.

Akpınar, Elgin, and Yeliz Yesilada. 2015. '"Old Habits Die Hard!": Eyetracking Based Experiential Transcoding: A Study with Mobile Users'. In *Proceedings of the 12th Web for All Conference*, 12:1–12:5. W4A '15. New York, USA: ACM Press. doi:10.1145/2745555.2746646.

Älli, Sami. 2012. 'Some Challenges for Developing an Easy-to-Read Website'. In *Proceedings of the W3C Easy-to-Read on the Web Syposium*, Paper 5. http://www.w3.org/WAI/RD/2012/easy-to-read/paper5/.

Alonso, Fernando, José Luis Fuertes, Ángel Lucas González, and Loïc Martínez. 2010. 'On the Testability of WCAG 2.0 for Beginners'. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 9:1–9:9. W4A '10. New York, USA: ACM Press. doi:10.1145/1805986.1806000.

Aluísio, Sandra Maria, and Caroline Gasperin. 2010. 'Fostering Digital Inclusion and Accessibility: The PorSimples Project for Simplification of Portuguese Texts'. In *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, 46–53. YIWCALA '10.

Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1868701.1868708.

Anastasiou, Dimitra, and Reinhard Schäler. 2010. 'Translating Vital Information: Localisation, Internationalisation, and Globalisation'. *Journal Syn-Theses*, 13–27.

Andersen, Aaron. 2010. 'Future Web Accessibility: SVG'. *WebAIM - Web Accessibility in Mind*. http://webaim.org/blog/future-web-accessibility-svg/.

Andreu-Vall, Mar, and Mari-Carmen Marcos. 2012. 'Evaluación de Sitios Web Multilingües: Metodología Y Herramienta Heurística'. *El Profesional de La Información* 21 (3): 254–60.

Andronikos, Nikos, Rossen Atanassov, Tavmjong Bah, Amelia Bellamy-Royds, Brian Birtles, Bogdan Brinza, Cyril Concolato, et al., eds. 2015. 'Scalable Vector Graphics (SVG) 2'. W3C Editor's Draft. https://svgwg.org/svg2-draft/Overview.html.

Asakawa, Chieko. 2005. 'What's the Web Like If You Can'T See It?' In *Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)*, 1–8. W4A '05. New York, USA: ACM Press. doi:10.1145/1061811.1061813.

———. 2014. 'Can a Blind Person Understand Your World?' In *Proceedings of the 11th Web for All Conference*. W4A '14. Seoul, Republic of Korea. New York, USA: ACM Press.

Asakawa, Chieko, and Hironobu Takagi. 2008. 'Transcoding'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 231–60. Human-Computer Interaction Series. London: Springer-Verlag.

Barreto, Armando. 2008. 'Visual Impairments'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 3–14. Human-Computer Interaction Series. London: Springer-Verlag.

Barthe, Kathy, Claire Juaneda, Dominique Leseigneur, Jean-Claud Loquet, Claude Morin, Jean Escande, and Annick Vayrette. 1999. 'GIFAS Rationalized French: A Controlled Language for Aerospace Documentation in French'. *Technical Communication* 46 (2): 220–29.

Bedford, Aurora. 2014. 'Icon Usability'. Blog. *Nielsen Norman Group: Evidence-Based User Experience Research, Training, and Consulting.* July 27. http://www.nngroup.com/articles/icon-usability/.

Berners-Lee, Tim. 1992. 'Word-Wide Web: The Information Universe'. *Electronic Networking* 2 (1): 52–58.

———. 2007. *The Future of the World Wide Web*. http://www.w3.org/People/Berners-Lee/Overview.html#Talks.

Berners-Lee, Tim, and Daniel Connolly, eds. 1995. 'Hypertext Markup Language - 2.0'. World Wide Web Consortium (W3C).

Bigham, Jeffrey P. 2007. 'Increasing Web Accessibility by Automatically Judging Alternative Text Quality'. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*, 349–52. IUI '07. Honolulu, Hawaii, USA: ACM Press. doi:10.1145/1216295.1216364.

———. 2009. 'Intelligent Interfaces Enabling Blind Web Users to Build Accessibility Into the Web'. Seattle, Washington, USA: University of Washington. http://www.sigaccess.org/2009/10/intelligent-interfaces-enabling-blind-web-users-to-build-accessibility-into-the-web/.

Bigham, Jeffrey P., Anna C. Cavender, Jeremy T. Brudvik, Jacob O. Wobbrock, and Richard E. Lander. 2007. 'WebinSitu: A Comparative Analysis of Blind and Sighted Browsing Behavior'. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, 51–58. Assets '07. New York, USA: ACM Press. doi:10.1145/1296843.1296854.

Bigham, Jeffrey P., Ryan S. Kaminsky, Richard E. Ladner, Oscar M. Danielsson, and Gordon L. Hempton. 2006. 'WebInSight: Making Web Images Accessible'. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, 181–88. Assets '06. Portland, Oregon. New York, USA: ACM Press. doi:10.1145/1168987.1169018.

Bigham, Jeffrey P., and Richard E. Ladner. 2007. 'Accessmonkey: A Collaborative Scripting Framework for Web Users and Developers'. In *Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility (W4A)*, 25–34. W4A '07. Banff, Canada. New York, USA: ACM Press. doi:10.1145/1243441.1243452.

Bolaños Medina, Alicia, María Jesús Rodríguez Medina, Lydia Bolaños Medina, and Luis Losada García. 2005. 'Analysing Digital Genres: Function and Functionality in Corporate Websites of Computer Hardware'. *Ibérica*, no. 9: 123–47.

Borodin, Yevgen, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. 2010. 'More Than Meets the Eye: A Survey of Screen-Reader Browsing Strategies'. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 13:1–13:10. W4A '10. Raleigh, USA. New York, USA: ACM Press. doi:10.1145/1805986.1806005.

Bowker, Lynne. 2015. 'Translatability and User eXperience: Compatible or in Conflict?' *Localisation Focus. The International Journal of Localisation.* 14 (2).

Brajnik, Giorgio. 2004. 'Comparing Accessibility Evaluation Tools: A Method for Tool Effectiveness'. *Universal Access in the Information Society* 3 (3): 252–63. doi:10.1007/s10209-004-0105-y.

———. 2008a. 'A Comparative Test of Web Accessibility Evaluation Methods'. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, 113–20. Assets '08. New York, USA: ACM Press. doi:10.1145/1414471.1414494.

———. 2008b. 'Beyond Conformance: The Role of Accessibility Evaluation Methods'. In *Web Information Systems Engineering – WISE 2008 Workshops*, edited by Sven Hartmann, Xiaofang Zhou, and Markus Kirchberg, 63–80. Lecture Notes in Computer Science 5176. Springer Berlin Heidelberg.

———. 2009. 'Validity and Reliability of Web Accessibility Guidelines'. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 131–38. Assets '09. New York, USA: ACM Press. doi:10.1145/1639642.1639666.

Brajnik, Giorgio, Andrea Mulas, and Claudia Pitton. 2007. 'Effects of Sampling Methods on Web Accessibility Evaluations'. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, 59–66. Assets '07. Tempe, Arizona, USA. New York, USA: ACM Press. doi:10.1145/1296843.1296855.

Brajnik, Giorgio, Yeliz Yesilada, and Simon Harper. 2010. 'Testability and Validity of WCAG 2.0: The Expertise Effect'. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, 43–50. ASSETS '10. Orlando, Florida. New York, USA: ACM Press. doi:10.1145/1878803.1878813.

———. 2011. 'The Expertise Effect on Web Accessibility Evaluation Methods'. *Human-Computer Interaction*, Springer, 26 (3): 246–83. doi:10.1080/07370024.2011.601670.

———. 2012. 'Is Accessibility Conformance an Elusive Property? A Study of Validity and Reliability of WCAG 2.0'. *ACM Transactions on Accessible Computing* 4 (2): 8:1–8:28. doi:10.1145/2141943.2141946.

Bredenkamp, Andrew, Berthold Crysmann, and Mirela Petrea. 2000. 'Looking for Errors: A Declarative Formalism for Resource-Adaptive Language Checking'. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*.

Bühler, Christian, Helmut Heck, Annika Nietzio, MortenGoodwin Olsen, and Mikael Snaprud. 2008. 'Monitoring Accessibility of Governmental Web Sites in Europe'. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, 5105:410–17. Lecture Notes in Computer Science. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-70540-6_59.

Cadieux, Pierre, and Bert Esselink. 2002. 'GILT: Globalization, Internationalization, Localization, Translation'. *Globalization Insider* 11 (1): 1–5.

Cadwell, Patrick. 2008. 'Readability and Controlled Language: Does the Study of Readability Have Merit in the Field of Controlled Language, and Is Readability Increased by Applying Controlled-Language Rules to Texts?' MA thesis. Dublin: Dublin City University (DCU).

Caldwell, Ben, Michael Cooper, Loretta Guarino Reid, and Gregg Vanderheiden, eds. 2008. 'Web Content Accessibility Guidelines (WCAG) 2.0'. W3C Recommendation. http://www.w3.org/TR/2008/REC-WCAG20-20081211/.

Carreras Montoto, Olga. 2014. 'Textos Alternativos, Imágenes Accesibles. Herramientas de Ayuda: Mapa de Decisión Y Wizard Online'. Blog. *Usable Y Accesible*. July 18. http://olgacarreras.blogspot.com.es/2014/07/textos-alternativos-imagenes-accesibles.html.

Center for Universal Design. 1997. 'The Principles of Universal Design, Version 2.0.' Raleigh, NC: North Carolina State University.

Chen, Alex, and Simon Harper. 2008. 'Web Evolution: Method and Materials'. Kilburn Building, Oxford Road, Manchester, M13 9PL, UK: School of Computer Science, The University of Manchester. http://wel-eprints.cs.man.ac.uk/74/.

Chevalier, Aline, and Melody Ivory. 2003. 'Web Site Designs: Influences of Designer's Experience and Design Constraints'. *International Journal of Human-Computer Studies* 58 (1): 57–87.

Chisholm, Wendy, and Shawn Lawton Henry. 2005. 'Interdependent Components of Web Accessibility'. In *Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)*, 31–37. W4A '05. New York, USA: ACM Press. doi:10.1145/1061811.1061818.

Chisholm, Wendy, Gregg Vanderheiden, and Ian Jacobs, eds. 1999. 'Web Content Accessibility Guidelines 1.0'. W3C Recommendation. http://www.w3.org/TR/WCAG10/.

Clark, Joe. 2002. *Building Accessible Websites*. Indianapolis, Ind: New Riders Press. http://joeclark.org/book/.

Colina, Sonia. 2008. 'Translation Quality Evaluation: Empirical Evidence for a Functionalist Approach'. *The Translator* 14 (1): 97–134.

Connor, Joshue O. 2012. *Pro HTML5 Accessibility*. 1st edition. Berkeley, CA : New York: Apress.

Cooper, Martyn, David Sloan, Brian Kelly, and Sarah Lewthwaite. 2012. 'A Challenge to Web Accessibility Metrics and Guidelines: Putting People and Processes First'. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 20:1–20:4. W4A '12. Lyon, France. New York, USA: ACM Press. doi:10.1145/2207016.2207028.

Cooper, Michael. 2005. 'Requirements for WCAG 2.0 Checklists, Techniques, and Test Files'. W3C Working Draft. https://www.w3.org/WAI/GL/WCAG20/WD-wcag2-tech-req-20050125.html.

Cooper, Michael, Andrew Kirkpatrick, and Joshue O Connor, eds. 2015a. 'Techniques for WCAG 2.0. Techniques and Failures for Web Content Accessibility Guidelines 2.0'. W3C Working Group Note. http://www.w3.org/TR/2015/NOTE-WCAG20-TECHS-20150226/.

———. , eds. 2015b. ‘Understanding WCAG 2.0. A Guide to Understanding and Implementing Web Content Accessibility Guidelines 2.0’. W3C Working Group Note. http://www.w3.org/TR/UNDERSTANDING-WCAG20/.

Craven, Timothy C. 2006. ‘Some Features of Alt Texts Associated with Images in Web Pages’. *Information Research* 11 (2). http://www.informationr.net/ir/11-2/paper250.html.

Cunningham, Katie. 2012. *Accessibility Handbook*. 1st edition. Beijing: O’Reilly Media.

Dahlström, Erik, Patrick Dengler, Anthony Grasso, Chris Lilley, Cameron McCormack, Doug Schepers, and Jonathan Watt, eds. 2011. ‘Scalable Vector Graphics (SVG) 1.1 (Second Edition)’. W3C Recommendation. https://www.w3.org/TR/SVG11/.

Debove, Antonia, Sabrina Furlan, and Ilse Depraetere. 2011. ‘A Contrastive Analysis of Five Automated QA Tools (QA Distiller 6.5.8, Xbench 2.8, ErrorSpy 5.0, SDLTrados 2007 QA Checker 2.0 and SDLX 2007 SP2 QA Check)’. In *Perspectives on Translation Quality*, edited by Ilse Depraetere, 161–92. Text, Translation, Computational Processing (TTCP) 9. Germany: De Gruyter Mouton.

DePalma, Donald A., and Benjamin B. Sargent. 2014. ‘Why Localization Matters for Corporate Buyers’. Common Sense Advisory (CSA). https://www.commonsenseadvisory.com/AbstractView.aspx?ArticleID=21553.

Depraetere, Ilse, and Thomas Vackier. 2011. ‘Comparing Formal Translation Evaluation and Meaning-Oriented Translation Evaluation: Or How QA Tools Can(not) Help’. In *Perspectives on Translation Quality*, edited by Ilse Depraetere, 25–50. Text, Translation, Computational Processing (TTCP) 9. Germany: De Gruyter Mouton.

Diez, David M., Christopher D. Barr, and Mine Çetinkaya-Rundel. 2012. *OpenIntro Statistics*. Second Edition. Creative Commons. http://www.openintro.org/stat/textbook.php.

Diggs, Joanmarie, James Craig, Shane McCarron, and Michael Cooper, eds. 2015. ‘Accessible Rich Internet Applications (WAI-ARIA) 1.1’. W3C Working Draft. http://www.w3.org/TR/wai-aria-1.1/.

Doherty, Stephen. 2012. ‘Investigating the Effects of Controlled Language on the Reading and Comprehension of Machine Translated Texts: A Mixed-Methods Approach’. PhD thesis. Dublin: Dublin City University (DCU). http://doras.dcu.ie/16805/.

Doherty, Stephen, and Sharon O’Brien. 2012. ‘A User-Based Usability Assessment of Raw Machine Translated Technical Instructions’. In *Proceedings of the 10th Biennial Conference of the Association for Machine Translation in the Americas (AMTA 2012)*. San Diego, CA.

Donney, Jerry, Steve Murphy, Chris Sacre, Alexander Scholz, and David Walters. 2008. ‘Globalization of Graphics: Delineating a Research into Using the Scalable Vector Graphics File Format to Improve the Translation of Graphics’. In *Proceedings of* *the 26th Annual ACM International Conference on Design of Communication*, 87–92. SIGDOC '08. New York, USA: ACM Press. doi:10.1145/1456536.1456554.

DRC. 2004. 'The Web: Access and Inclusion for Disabled People'. London: Disability Rights Commission.

Drndarevic, Biljana, Sanja Štajner, and Horacio Saggion. 2012. 'Reporting Simply: A Lexical Simplification Strategy for Enhancing Text Accessibility'. In *Proceedings of the W3C Easy-to-Read on the Web Syposium*, Paper 7. http://www.w3.org/WAI/RD/2012/easy-to-read/paper7/.

Dunne, Keiran J. 2006. 'Putting the Cart behind the Horse: Rethinking Localization Quality Management'. In *Perspectives on Localization*, edited by Keiran J. Dunne, 95–117. Amsterdam: John Benjamins.

———. 2011. 'From Vicious to Virtuous Cycle. Customer-Focused Translation Quality Management Using ISO9001 Principles and Agile Methodologies'. In *Translation and Localization Project Management: The Art of the Possible*, edited by Keiran J. Dunne and Elena S. Dunne, 153–87. Amsterdam: John Benjamins.

———. 2015. 'Localization'. In *The Routledge Encyclopedia of Translation Technology*, edited by Chan Sin-wai, 550–62. Oxon, New York: Routledge.

Edwards, Allistair D.N. 2008. 'Assistive Technologies'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 141–62. Human-Computer Interaction Series. London: Springer-Verlag.

Eggert, Eric, and Shadi Abou-Zahra, eds. 2014a. 'Images - W3C Web Accessibility Tutorials Series'. W3C/WAI. http://www.w3.org/WAI/tutorials/images/.

———. , eds. 2014b. 'Web Accessibility Evaluation Tools List'. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/ER/tools/.

———. 2014c. 'Web Accessibility Tutorials: Images'. WAI, W3C. http://www.w3.org/WAI/tutorials/images/.

Engle, Robert F. 1984. 'Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics'. In *Handbook of Econometrics*, edited by Zvi Griliches and Michael Intriligator. Vol. II. Elsevier.

Ernst & Young. 2012. 'Informe Sobre La Ceguera En España'. Report by Ernst & Young in collaboration with Retinaplus+ and ONCE.

Esselink, Bert. 2000. *A Practical Guide to Localization*. Amsterdam: John Benjamins. https://benjamins.com/#catalog/books/liwd.4/main.

———. 2002. 'Localization Engineering: The Dream Job?' *Tradumática*, no. 1. http://www.fti.uab.es/tradumatica/revista/articles/besselink/art.htm.

———. 2003. 'Localisation and Translation'. In *Computers and Translation. A Translator's Guide.*, edited by Harlold Somers, 67–86. John Benjamins. https://benjamins.com/#catalog/books/btl.35.17nyb/details.

———. 2006. ‘The Evolution of Localization’. In *Translation Technology and Its Teaching*, edited by Anthony Pym, Alexander Perekrestenko, and Bram Starink, 21–29. Tarragona, Spain: Servei de publicacions, Universitat Rovira i Virgili. http://isg.urv.es/seminars/2003_localization_online/esselink.html.

Faulkner, Steve. 2011. ‘HTML5 Accessibility Chops: The Alt Decision’. *The Paciello Group*. https://www.paciellogroup.com/blog/2011/04/html5-accessibility-chops-the-alt-decision/.

———. 2012. ‘HTML5 Accessibility Chops: Title Attribute Use and Abuse’. *The Paciello Group*. https://www.paciellogroup.com/blog/2012/01/html5-accessibility-chops-title-attribute-use-and-abuse/.

———. 2015. ‘Using the HTML Title Attribute’. *The Paciello Group*. https://www.paciellogroup.com/blog/2010/11/using-the-html-title-attribute/.

Faulkner, Steve, Hans Hillen, and David MacDonald, eds. 2015. ‘Notes on Using ARIA in HTML’. W3C Working Draft. http://www.w3.org/TR/aria-in-html/.

Fernandes, Nádia, Nikolaos Kaklanis, Konstantinos Votis, Dimitrios Tzovaras, and Luís Carriço. 2014. ‘An Analysis of Personalized Web Accessibility’. In *Proceedings of the 11th Web for All Conference*, 19:1–19:10. W4A ’14. Seoul, Republic of Korea. USA: ACM Press. doi:10.1145/2596695.2596698.

Fernández Costales, Alberto. 2009. ‘Translation 2.0. The Localization of Institutional Websites under the Scope of Functionalist Approaches’. In *Translation and the (Trans)formation of Identities*, edited by Dries De Crom.

———. 2010. ‘Traducción, localización e internacionalización: El caso de las páginas web universitarias’. PhD thesis. Oviedo: Universidad de Oviedo.

Fischer, Detlev, and Tiffany Wyatt. 2011. ‘The Case for a WCAG-Based Evaluation Scheme with a Graded Rating Scale’. In *W3C Symposium on Website Accessibility Metrics*, edited by Joshue O Connor Markel Vigo Giorgio Brajnik, article 7. http://www.w3.org/WAI/RD/2011/metrics/paper7/.

Folaron, Deborah. 2006. ‘A Discipline Coming of Age in the Digital Age’. In *Perspectives on Localization*, edited by Keiran J. Dunne, 195–219. Amsterdam; Philadelphia: John Benjamins.

———. 2010. ‘Web and Translation’. In *Handbook of Translation Studies*, 1:446–51. Amsterdam; Philadelphia: John Benjamins.

———. 2012. ‘Digitalizing Translation’. *Translation Spaces*, John Benjamins, 1: 5–31.

Freire, Andre P., Cibele M. Russo, and Renata P. M. Fortes. 2008. ‘A Survey on the Accessibility Awareness of People Involved in Web Development Projects in Brazil’. In *Proceedings of the 2008 International Cross-Disciplinary Conference on Web Accessibility (W4A)*, 87–96. W4A ’08. New York, USA: ACM Press. doi:10.1145/1368044.1368064.

Freitas, Diamantino. 2010. 'Accessibility and Design for All Solutions Through Speech Technology'. In *Speech Technology. Theory and Applications*, edited by Fang Chen and Kristina Jokinen, 271–99. Springer. http://dx.doi.org/10.1007/978-0-387-73819-2_14.

Fuertes Castro, José Luis, and Loïc Martínez Normand. 2007. 'Accesibilidad Web'. *TRANS. Revista de Traductología*, no. 11: 135–54.

Gambino, Orazio, Roberto Pirrone, and Fabrizio DiGiorgio. 2014. 'Accessibility of the Italian Institutional Web Pages: A Survey on the Compliance of the Italian Public Administration Web Pages to the Stanca Act and Its 22 Technical Requirements for Web Accessibility'. *Universal Access in the Information Society*, 1–8. doi:10.1007/s10209-014-0381-0.

Gerlach, Johanna. 2015. 'Improving Statistical Machine Translation of Informal Language: A Rule-Based Pre-Editing Approach for French Forums'. PhD thesis. Geneva: University of Geneva.

Gerlach, Johanna, Victoria Porro Rodriguez, Pierrette Bouillon, and Sabine Lehmann. 2013. 'Combining Pre-Editing and Post-Editing to Improve SMT of User-Generated Content'. In *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice*, edited by Sharon O'Brien, Michel Simard, and Lucia Specia, 45–53. The European Association for Machine Translation. http://archive-ouverte.unige.ch/unige:30952.

Gibb, Forbes, and Ioannis Matthaiakis. 2007. 'A Framework for Assessing Web Site Localisation'. *The Electronic Library* 25 (6): 664–78.

Gómez, David, Alberto Heredia, and Sonia Martín. 2007. 'Usabilidad Y Accesibilidad Para Mejorar El Posicionamiento Web'. In *II Congreso AMADIS (Accesibilidad a Los Medios Audiovisuales Para Personas Con Discapacidad)*, 243–49. Madrid: Real Patronato sobre Discapacidad.

Gouadec, Daniel. 2007. *Translation as a Profession*. Amsterdam; Philadelphia: John Benjamins.

———. 2010. 'Quality in Translation'. In *Handbook of Translation Studies*, edited by Yves Gambier and Luc van Doorslaer, 1:270–75. Amsterdam: John Benjamins.

Gries, Stephan Th., and Stefanie Wulff. 2012. 'Regression Analysis in Translation Studies'. In *Quantitative Methods in Corpus-Based Translation Studies*, edited by Michael P. Oakes and Meng Ji, 35–52. Amsterdam; Philadelphia: John Benjamins.

Guerberof Arenas, Ana. 2012. 'Productivity and Quality in the Post-Editing of Outputs from Translation Memories and Machine Translation'. PhD thesis. Tarragona: Universitat Rovira i Virgili. http://www.tdx.cat/handle/10803/90247.

Gutiérrez y Restrepo, Emmanuelle, and Loïc Martínez Normand. 2010. 'Localization and Web Accessibility'. *Tradumàtica*, no. 8 (December). http://www.fti.uab.cat/tradumatica/revista/num8/articles/10/10art.htm.

Haddad, Ghassan. 2003. 'Globalization Image Assistant (GIA)'. *Multilingual Computing and Technology*. https://multilingual.com/article-detail/?art_id=1093.

Hanson, Vicki L., and John T. Richards. 2013. 'Progress on Website Accessibility?' *ACM Transactions on the Web* 7 (1): 2:1–2:30. doi:10.1145/2435215.2435217.

Hardy, Melissa, and Alan Bryman, eds. 2009. *The Handbook of Data Analysis*. Paperback Edition. Los Angeles, CA, USA: SAGE Publications Ltd.

Harper, Simon, and Alex Chen. 2012. 'Web Accessibility Guidelines: A Lesson from the Evolving Web'. *World Wide Web* 15 (1): 61–88.

Harper, Simon, Eleni Michailidou, and Robert Stevens. 2009. 'Toward a Definition of Visual Complexity As an Implicit Measure of Cognitive Load'. *ACM Trans. Appl. Percept.* 6 (2): 10:1–10:18. doi:10.1145/1498700.1498704.

Harper, Simon, Anwar Ahmad Moon, Markel Vigo, Giorgio Brajnik, and Yeliz Yesilada. 2015. 'DOM Block Clustering for Enhanced Sampling and Evaluation'. In *Proceedings of the 12th Web for All Conference*, 15:1–15:10. W4A '15. New York, USA: ACM Press. doi:10.1145/2745555.2746649.

Harper, Simon, and Yeliz Yesilada, eds. 2008a. *Web Accessibility: A Foundation for Research*. Human-Computer Interaction Series. London: Springer-Verlag.

———. 2008b. 'Web Accessibility and Guidelines'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 61–78. London: Springer-Verlag.

Hartley, Anthony, and Cécile Paris. 2001. 'Translation, Controlled Languages, Generation'. In *Exploring Translation and Multilingual Text Production: Beyond Content*, edited by Erich Steiner and Colin Yallop, 307–26. Text, Translation, Computational Processing (TTCP) 3. Germany: De Gruyter Mouton.

Hertzum, Morten, and Niels Ebbe Jacobsen. 2001. 'The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods'. *International Journal of Human-Computer Interaction* 13 (4): 421–43.

Hickson, Ian, Robin Berjon, Steve Faulkner, Travis Leithead, Erika D. Navara, Edward O'Connor, and Silvia Pfeiffer, eds. 2014. 'HTML 5. A Vocabulary and Associated APIs for HTML and XHTML'. W3C Recommendation. http://www.w3.org/TR/html5/.

Hinderer Sova, Deborah, and Jakob Nielsen. 2003. 'How to Recruit Participants for Usability Studies'. Fremont: Nielsen Norman Group. http://www.nngroup.com/reports/how-to-recruit-participants-usability-studies/.

Hong, Soongoo, Pairin Katerattanakul, and Seok Jeong Joo. 2008. 'Evaluating Government Website Accessibility: A Comparative Study'. *International Journal of Information Technology & Decision Making* 07 (03): 491–515. doi:10.1142/S0219622008003058.

Hornbæk, Kasper, and Erik Frøkjær. 2008. ‘A Study of the Evaluator Effect in Usability Testing’. *Human–Computer Interaction* 23 (3): 251–77. doi:10.1080/07370020802278205.

Horton, Sarah, and Laura Leventhal. 2008. ‘Universal Usability’. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 345–56. Human-Computer Interaction Series. London: Springer-Verlag.

House, Juliane. 2013. ‘Quality in Translation Studies’. In *The Routledge Handbook of Translation Studies*, edited by Carmen Millán and Francesca Bartrina, 534–47. New York: Routledge.

Huijsen, Willem-Olaf. 1998. ‘Controlled Language - An Introduction’. In *Proceedings of the Second International Workshop on Controlled Language Applications – CLAW 98*, edited by Teruko Mitamura, 1–15. Pittsburgh, Pennsylvania: Language Technologies Institute, Carnegie Mellon University. http://books.google.es/books/about/Claw_98.html?id=rU48ywAACAAJ&redir_esc=y.

Hu, Jianying, and Amit Bagga. 2003. ‘Identifying Story and Preview Images in News Web Pages’. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, 640 – . ICDAR ’03. Washington, DC, USA: IEEE Computer Society. http://dl.acm.org/citation.cfm?id=938980.939588.

Inclusion Europe. 2009. ‘Information for All. European Standards for Making Information Easy to Read and Understand’. European Commission. http://www.inclusion-europe.org/etr/en/european-easy-to-read-standards.

Insieme. 2015. ‘Facile à Surfer’. http://insieme.ch/fr/produit/facile-a-surfer/

International Organization for Standardization. 2012a. ‘ISO/IEC 40500:2012 Information Technology -- W3C Web Content Accessibility Guidelines (WCAG) 2.0’. International Organization for Standardization (ISO). http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=58625.

———. 2012b. ‘ISO/IEC 20071-11:2012 Information Technology — User Interface Component Accessibility — Part 11: Guidance for Alternative Text for Images’. International Organization for Standardization (ISO). https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:20071:-11:ed-1:v1:en.

International Telecommunication Union. 2015. ‘ICT - Facts and Figures’. ITU - ICT Data and Statistics Division. http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

Ishida, Richard. 2015. ‘Internationalization Quick Tips for the Web’. http://www.w3.org/International/quicktips/.

Ivory, Melody. 2003. *Automated Web Site Evaluation.* Human-Computer Interaction Series. Dordrecht: Springer.

Ivory, Melody, and Aline Chevalier. 2002. 'A Study of Automated Web Site Evaluation Tools'. University of Washington, Department of Computer Science.

Jacobsen, Niels Ebbe, Morten Hertzum, and Bonnie E. John. 1998. 'The Evaluator Effect in Usability Tests'. In *CHI 98 Cconference Summary on Human Factors in Computing Systems*, 255–56. CHI '98. New York, USA: ACM Press. doi:10.1145/286498.286737.

Jay, Caroline, Darren Lunn, and Eleni Michailidou. 2008. 'End User Evaluations'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 107–26. Human-Computer Interaction Series. London: Springer-Verlag.

Jiménez Crespo, Miguel Ángel. 2008. 'El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo'. PhD thesis. Granada: Universidad de Granada. http://digibug.ugr.es/handle/10481/1908#.U9Yqv7HiNuQ.

———. 2009a. 'Navegación Accesible'. In *Materiales Multimedia Para Todos. Inclusión Y Accessibilidad En Educación.*, 185–204. Traducción Accesible. Granada: Tragacanto.

———. 2009b. 'The Evaluation of Pragmatic and Functionalist Aspects in Localisation: Towards a Holistic Approach to Quality Assurance'. *The Journal of Internationalization and Localisation* 1: 60–93.

———. 2010. 'Localization and Writing for a New Medium: A Review of Digital Style Guides'. *Tradumàtica*, Web localisation, , no. 8 (December). http://www.fti.uab.es/tradumatica/revista/num8/articles/08/08central.htm.

———. 2013. *Translation and Web Localization.* London: Routledge.

Kahle, Brewster. 2002. 'Editors' Interview. The Internet Archive'. *RLG DigiNews* 6 (3). http://www.rlg.org/preserv/diginews/diginews6-3.html.

Keysers, Daniel, Marius Renn, and Thomas M. Breuel. 2007. 'Improving Accessibility of Html Documents Bygenerating Image-Tags in a Proxy'. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, 249–50. Assets '07. New York, USA: ACM Press. doi:10.1145/1296843.1296896.

Kilgarriff, Adam, and Gregory Grefenstette. 2003. 'Web as a Corpus'. *Computational Linguistics* 29 (3): 333–47.

Klein, Eduard, Anton Bolfing, and Markus Riesch. 2014. 'Checking Web Accessibility with the Content Accessibility Checker (CAC)'. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler, 109–12. Lecture Notes in Computer Science 8547. Springer International Publishing. http://link.springer.com/chapter/10.1007/978-3-319-08596-8_16.

Klopman, Exequiel. 2016. 'Preparing Image Text for CAT Tools'. *Multilingual Computing and Technology.*

Korpela, Jukka. 2012. 'Guidelines on Alt Texts in Img Elements (revised Version)'. https://www.cs.tut.fi/~jkorpela/html/alt.html.

Kuhn, Tobias. 2014. 'A Survey and Classification of Controlled Natural Languages'. *Computational Linguistics* 40 (1): 121–70. doi:10.1162/COLI_a_00168.

Lakó, Cristian. 2014. 'Localizing Websites: Shifting Focus onto the End-User'. Iaşi (Romania): Universitatea „Alexandru Ioan Cuza" din Iaşi.

Lawton Henry, Shawn, ed. 2005a. 'Essential Components of Web Accessibility'. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/intro/components.php.

———. , ed. 2005b. 'Introduction to Web Accessibility'. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/intro/accessibility.php.

———. 2007. *Just Ask: Integrating Accessibility Throughout Design*. Lulu.com. http://uiaccess.com/accessucd/print.html.

———. , ed. 2008. 'The WCAG 2.0 Documents'. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/intro/wcag20.

———. 2009. 'How WCAG 2.0 Differs from WCAG 1.0'. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/WCAG20/from10/diff.php.

Lawton Henry, Shawn, and Shadi Abou-Zahra. 2014. 'The Role of Accessibility in a Universal Web'. In *Proceedings of the 11th Web for All Conference*. W4A '14. Seoul, Republic of Korea. New York, USA: ACM Press. doi:10.1145/2596695.2596718.

Lazar, Jonathan, Aaron Allen, Jason Kleinman, and Chris Malarkey. 2007. 'What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users'. *International Journal of Human-Computer Interaction* 22 (3): 247–69. doi:10.1080/10447310709336964.

Lazar, Jonathan, Alfreda Dudley-Sponaugle, and Kisha-Dawn Greenidge. 2004. 'Improving Web Accessibility: A Study of Webmaster Perceptions'. *The Compass of Human-Computer Interaction* 20 (2): 269–88. doi:10.1016/j.chb.2003.10.018.

Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. John Wiley & Sons.

Leitner, Marie-Luise, and Christine Strauss. 2010. 'Organizational Motivations for Web Accessibility Implementation – A Case Study'. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, 392–99. Lecture Notes in Computer Science 6179. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-14097-6_62.

Leitner, Marie-Luise, Christine Strauss, and Christian Stummer. 2014. 'Web Accessibility Implementation in Private Sector Organizations: Motivations and Business Impact'. *Universal Access in the Information Society*, 1–12. doi:10.1007/s10209-014-0380-1.

Lionbridge. 2015. '2015 State of Website Localization Report'. USA. https://ww1.lionbridge.com/website-localization-report/

Lopes, Rui, Daniel Gomes, and Luís Carriço. 2010. 'Web Not for All: A Large Scale Study of Web Accessibility'. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 10:1–10:4. W4A '10. Raleigh, USA. New York, USA: ACM Press. doi:10.1145/1805986.1806001.

MacKenzie, I. Scott. 2013. *Human-Computer Interaction: An Empirical Research Perspective*. USA: Elsevier. http://www.yorku.ca/mack/HCIbook/.

Madugalla, Anuradha. 2015. 'Accessible On-Line Graphics'. In *Proceedings of the 12th Web for All Conference*, 41:1–41:2. W4A '15. New York, USA: ACM Press. doi:10.1145/2578726.2746672.

Makoushina, Julia. 2007. 'Translation Quality Assurance Tools: Current State and Future Approaches'. In *Proceedings of Translating and the Computer 29*. London, UK.

Mangiron, Carmen, Pilar Orero, and Minako O'Hagan, eds. 2014. *Fun for All. Translation and Accessibility Practices in Video Games*. Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien: Peter Lang.

Mankoff, Jennifer, Holly Fait, and Tu Tran. 2005. 'Is Your Web Page Accessible? A Comparative Study of Methods for Assessing Web Page Accessibility for the Blind'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, 41–50. ACM Press.

Marcus, Aaron. 1996. 'Icon and Symbol Design Issues for Graphical User Interfaces'. In *International User Interfaces*, edited by Elisa del Galdo and Jakob Nielsen, 257-70. New York, NY, USA: John Wiley & Sons.

Mata Pastor, Manuel. 2005. 'Localización Y Traducción de Contenido Web'. In *Traducción Y Localización - Mercado, Gestión Y Tecnologías*, edited by Detlef Reineke, 187–252. Las Palmas de Gran Canaria: Anroat Ediciones.

———. 2009a. 'Algunas Pautas Para El Tratamiento de Imágenes Y Contenido Gráfico En Proyectos de Localización (I)'. *Entreculturas: Revista de Traducción Y Comunicación Intercultural*, no. 1: 513–32.

———. 2009b. 'Algunas Pautas Para El Tratamiento de Imágenes Y Contenido Gráfico En Proyectos de Localización (II)'. *Entreculturas: Revista de Traducción Y Comunicación Intercultural*, no. 1: 533–69.

Matausch, Kerstin, and Annika Nietzio. 2012. 'Easy-to-Read and Plain Language: Defining Criteria and Refining Rules'. In *Proceedings of the W3C Easy-to-Read on the Web Syposium*, Paper 11. http://www.w3.org/WAI/RD/2012/easy-to-read/paper11/.

Matis, Nancy. 2011. 'Quality Assurance in the Translation Workflow – A Professional's Testimony'. In *Perspectives on Translation Quality*, edited by Ilse Depraetere, 147–60. Text, Translation, Computational Processing (TTCP) 9. Germany: De Gruyter Mouton.

McCarron, Shane, Liam R. E. Quin, and Steve Faulkner, eds. 2015. 'HTML5: Techniques for Providing Useful Text Alternatives'. W3C Working Group Note. http://www.w3.org/TR/html-alt-techniques/.

McDonough, Julie. 2006. 'Hiding Difference: On the Localization of Websites'. *The Translator* 12 (1): 85–103.

McEwan, Tom, and Ben Weerts. 2007. 'ALT Text and Basic Accessibility'. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 2*, 71–74. BCS-HCI '07. Swinton, UK, UK: British Computer Society. http://dl.acm.org/citation.cfm?id=1531407.1531426.

Michailidou, Eleni, Simon Harper, and Sean Bechhofer. 2008. 'Investigating Sighted Users' Browsing Behaviour to Assist Web Accessibility'. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, 121–28. Assets '08. New York, NY, USA: ACM. doi:10.1145/1414471.1414495.

Miesenberger, Klaus, and Andrea Petz. 2014. '"Easy-to-Read on the Web": State of the Art and Needed Research'. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler, 161–68. Lecture Notes in Computer Science 8547. Springer International Publishing. http://link.springer.com/chapter/10.1007/978-3-319-08596-8_25.

Montalt i Resurrecció, Vicent. 2003. 'La Traducción de Géneros Electrónicos: El Caso de La Localización'. In *Panorama Actual de La Investigación En Traducción E Interpretación*, edited by Miguel Ángel García Peinado and Emilio Ortega Arjonilla, 3rd ed., 313–28. Traducción En El Atrio 1. Granada: Editorial Atrio.

Morado Vázquez, Lucia. 2012. 'An Empirical Study on the Influence of Translation Suggestions' Provenance Metadata'. PhD thesis. Limerick (Ireland): University of Limerick.

Mosquera, Alejandro, Elena Lloret, and Paloma Moreda. 2012. 'Towards Facilitating the Accessibility of Web 2.0 Texts through Text Normalisation'. In *Proceedings of the Natural Language Processing for Improving Textual Accessibility (NLP4ITA) Workshop*. Istambul, Turkey.

Munday, Jeremy, ed. 2009. *The Routledge Companion to Translation Studies*. Revised Edition. Oxon: Routledge.

Muñoz Sánchez, Pablo. 2008. *Image Localization Manager*. Guía rápida. http://algomasquetraducir.com/image-localization-manager/

Nauert, Sandra. 2007. 'Translating Websites'. In *Proceedings of the Marie Curie Euroconferences MuTra: LSP Translation Scenarios*. EU High Level Scientific Conference Series. Vienna. http://www.euroconferences.info/proceedings/2007_Proceedings/2007_proceedings.html.

Nevile, Charles, and Mark Sadecki, eds. 2015. 'HTML5 Image Description Extension (longdesc)'. W3C Recommendation. http://www.w3.org/TR/html-longdesc/.

Nielsen, Jakob. 2014. 'Icon Classification: Resemblance, Reference, and Arbitrary Icons'. Blog. *Nielsen Norman Group: Evidence-Based User Experience Research, Training, and Consulting*. August 17. http://www.nngroup.com/articles/classifying-icons/.

Nielsen, Jakob, and Hoa Loranger. 2006. *Prioritizing Web Usability*. 1st edition. Berkeley, California: New Riders.

Nord, Christiane. 1997. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome.

———. 2010. 'Functionalist Approaches'. In *Handbook of Translation Studies*, edited by Yves Gambier and Luc van Doorslaer, 1:120–28. Amsterdam: John Benjamins.

Nyberg, Eric, Teruko Mitamura, and Willem-Olaf Huijsen. 2003. 'Controlled Language for Authoring and Translation'. In *Computers and Translation. A Translator's Guide.*, edited by Harlold Somers, 245–81. John Benjamins. https://benjamins.com/#catalog/books/btl.35.17nyb/details.

Nyong, Ndon. 2009. 'Guidelines for Creating Meaningful and Effective Alternative Text'. University of Saskatchewan.

Oates, Briony J. 2005. *Researching Information Systems and Computing*. London: SAGE Publications Ltd.

O'Brien, Sharon. 2003. 'Controlling Controlled English. An Analysis of Several Controlled Language Rule Sets'. *Proceedings of EAMT-CLAW*, St. Jerome Publishing, 3: 105–14.

O'Brien, Sharon. 2006. 'Machine-Translatability and Post-Editing Effort: An Empirical Study Using Translog and Choice Network Analysis'. PhD thesis. Dublin: Dublin City University (DCU).

———. 2010. 'Controlled Language and Readability'. In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 143–65. American Translators Association Scholarly Monograph Series, XV. Amsterdam; Philadelphia: John Benjamins.

———. 2012a. 'Towards a Dynamic Quality Evaluation Model for Translation'. *JoSTrans. The Journal of Specialised Translation.*, no. 17: 55–77.

———. 2012b. 'Translation as Human–computer Interaction'. *Translation Spaces*, John Benjamins, 1 (1): 101–22.

O'Brien, Sharon, and Johann Roturier. 2007. 'How Portable Are Controlled Language Rules? A Comparison of Two Empirical MT Studies'. In *Proceedings of MT Summit XI*, 345–52. Copenhagen (Denmark).

Ó Broin, Ultan. 2003. 'Image Localization and New Technology'. *Multilingual Computing and Technology*. http://www.multilingual.com/articleDetail.php?id=1047.

———. 2004. ‘Accessibility Is Just Another Language’. *Multilingual Computing and Technology*. http://www.multilingual.com/articleDetail.php?id=675.

Ogden, Charles Kay. 1930. *Basic English: A General Introduction with Rules and Grammar*. London: Paul Treber and Co., Ltd.

O'Hagan, Minako, and David Ashworth. 2002. *Translation-Mediated Communication in a Digital World: Facing the Challenges of Globalization and Localization*. Clevedon: Multilingual Matters Limited.

Olsen, Morten Goodwin, Mikael Snaprud, and Annika Nietzio. 2010. ‘Automatic Checking of Alternative Texts on Web Pages’. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, 425–32. Lecture Notes in Computer Science 6179. Berlin: Springer. http://link.springer.com/chapter/10.1007/978-3-642-14097-6_68.

Paciello, Mike. 2000. *Web Accessibility for People with Disabilities*. 1st edition. Lawrence, Kan.: Berkeley, CA: CRC Press.

Paek, Seungyup, and John R. Smith. 1998. ‘Detecting Image Purpose in World-Wide Web Documents’. In *Proc. IS&T/SPIE Symp. on Electronic Imaging: Science and Technology - Document Recognition*. San Jose, CA.

Pascual, Afra, Mireia Ribera, Toni Granollers, and Jordi L. Coiduras. 2014. ‘Impact of Accessibility Barriers on the Mood of Blind, Low-Vision and Sighted Users’. *Procedia Computer Science* 27 (0): 431–40. doi:http://dx.doi.org/10.1016/j.procs.2014.02.047.

Pernice, Kara, and Jakob Nielsen. 2001. ‘Usability Guidelines for Accessible Web Design’. Fremont: Nielsen Norman Group.

Petrie, Helen, Fraser Hamilton, and Neil King. 2004. ‘Tension, What Tension?: Website Accessibility and Visual Design’. In *Proceedings of the 2004 International Cross-Disciplinary Workshop on Web Accessibility (W4A)*, 13–18. W4A '04. New York, USA: ACM Press. doi:10.1145/990657.990660.

Petrie, Helen, Fraser Hamilton, Neil King, and Pete Pavan. 2006. ‘Remote Usability Evaluations With Disabled People’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1133–41. CHI '06. New York, USA: ACM Press. doi:10.1145/1124772.1124942.

Petrie, Helen, Chandra Harrison, and Sundeep Dev. 2005. ‘Describing Images on the Web: A Survey of Current Practice and Prospects for the Future’. In *Proceedings of the 3rd International Conference on Universal Access in Human Computer Interaction*.

Petrie, Helen, and Omar Kheir. 2007. ‘The Relationship Between Accessibility and Usability of Websites’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 397–406. CHI '07. New York, USA: ACM Press. doi:10.1145/1240624.1240688.

Petrie, Helen, Christopher Power, David Swallow, Carlos A. Velasco, Blaithin Gallagher, Mark Magennis, Emma Murphy, Sam Collin, and Keren Down. 2011. 'The Value Chain for Web Accessibility: Challenges and Opportunities'. In *Proceedings of the Accessible Design in the Digital World (ADDW) Workshop*. Lisbon, Portugal.

Petrie, Helen, Andreas Savva, and Christopher Power. 2015. 'Towards a Unified Definition of Web Accessibility'. In *Proceedings of the 12th Web for All Conference*, 35:1–35:13. W4A '15. New York, USA: ACM Press. doi:10.1145/2745555.2746653.

Pieters, Simon, Anne van Kesteren, Philip Jägenstedt, Domenic Denicola, Ian Hickson, Steve Faulkner, Travis Leithead, Erika Doyle Navara, Edward O'Connor, and Robin Berjon, eds. 2015. 'HTML 5.1'. W3C Working Draft. http://www.w3.org/TR/html51/.

PLAIN, Plain Language Action and Information Network. 2011. 'Federal Plain Language Guidelines'. http://www.plainlanguage.gov/.

Pool, Jonathan. 2006. 'Can Controlled Languages Scale to the Web?' In *Proceedings of the 5th International Workshop on Controlled Language Applications (CLAW)*. Cambridge, Massachusetts. http://www.mt-archive.info/05/CLAW-2006-Pool.pdf.

Porro Rodríguez, Victoria, Johanna Gerlach, Pierrette Bouillon, and Violeta Seretan. 2014. 'Rule-Based Automatic Post-Processing of SMT Output to Reduce Human Post-Editing Effort'. In *Proceedings of Translating and the Computer 36*.

Power, Christopher, André Freire, Helen Petrie, and David Swallow. 2012. 'Guidelines Are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 433–42. CHI '12. New York, USA: ACM Press. doi:10.1145/2207676.2207736.

Prieto Velasco, Juan Antonio. 2009. *Traducción E Imagen: La Información Visual En Textos Especializados*. Traducción Accesible. Granada: Tragacanto.

Putnam, Cynthia, Kathryn Wozniak, Mary Jo Zefeldt, Jinghui Cheng, Morgan Caputo, and Carl Duffield. 2012. 'How Do Professionals Who Create Computing Technologies Consider Accessibility?' In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, 87–94. ASSETS '12. Boulder, Colorado, USA: ACM. doi:10.1145/2384916.2384932.

Pym, Anthony. 2003. 'Localization and the Training of Linguistic Mediators for the Third Millennium'. In *The Challenges of Translation and Interpretation in the Third Millennium*, edited by Naji Oueijan and Boulos Sarru, 23–30. Beirut: Notre Name University, Louaize.

———. 2004. *The Moving Text: Localization, Translation, and Distribution*. Amsterdam; Philadelphia: John Benjamins.

———. 2011. ‘Website Localization’. In *The Oxford Handbook of Translation Studies*, edited by Kirsten Malmkjær and Kevin Windle, 410–24. Oxford, England: Oxford University Press.

———. 2014. *Exploring Translation Theories*. Second edition. Oxon, New York: Routledge.

Quah, Chiew Kin. 2006. *Translation and Technology*. Houndmills England ; New York: Palgrave Macmillan.

Quesenbery, Whitney. 2008. ‘The Five Dimensions of Usability’. In *Content and Complexity: Information Design in Technical Communication*, edited by Michael J. Albers and Beth Mazur. Mahwah: Lawrence Erlbaum Associates, Inc.

Raggett, Dave, Arnaud Le Hors, and Ian Jacobs, eds. 1999. ‘HTML 4.01 Specification’. W3C Recommendation. http://www.w3.org/TR/html401/.

Regan, Bob, and Andrew Kirkpatrick. 2008. ‘Multimedia and Graphics’. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 287–99. London: Springer-Verlag.

Rello, Luz, Gaurang Kanvinde, and Ricardo Baeza-Yates. 2012. ‘Layout Guidelines for Web Text and a Web Service to Improve Accessibility for Dyslexics’. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 36:1–36:9. W4A ’12. New York, NY, USA: ACM. doi:10.1145/2207016.2207048.

Reuther, Ursula, and Antje Schmidt-Wigger. 2000. ‘Designing a Multi-Purpose CL Application’. In *Proceedings of the Third International Workshop on Controlled Language Applications (CLAW 2000)*. Seattle, Washington, USA.

Richards, John T., Kyle Montague, and Vicki L. Hanson. 2012. ‘Web Accessibility As a Side Effect’. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, 79–86. ASSETS ’12. Boulder, Colorado, USA: ACM. doi:10.1145/2384916.2384931.

Rodríguez Vázquez, Silvia. 2013. ‘Towards Defining the Role of Localisation Professionals in the Achievement of Multilingual Web Accessibility’. *Revista Tradumàtica* 11 (1): 383–88.

———. 2014. ‘Introducing Web Accessibility to Localization Students: Implications for a Universal Web’. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, 333–34. ASSETS ’14. New York, NY, USA: ACM. doi:10.1145/2661334.2661414.

———. 2015a. ‘A Controlled Language-Based Evaluation Approach to Ensure Image Accessibility during Web Localisation’. *Translation Spaces* 4 (2): 187–215.

———. 2015b. ‘Exploring Current Accessibility Challenges in the Multilingual Web for Visually-Impaired Users’. In *The 24th International World Wide Web Conference (WWW) 2015 Companion Volume*. New York, NY, USA: ACM. doi:10.1145/2740908.2743010.

———. 2015c. 'Unlocking the Potential of Web Localizers as Contributors to Image Accessibility: What Do Evaluation Tools Have to Offer?' In *Proceedings of the 12th Web for All (W4A) Conference*. New York, NY, USA: ACM. doi:doi.org/10.1145/2745555.2746662.

———. 2016. 'Measuring the Impact of Automated Evaluation Tools on Alternative Text Quality: A Web Translation Study'. In *Proceedings of the 13th Web for All Conference*. W4A '16. New York: ACM.

Rodríguez Vázquez, Silvia, and Anton Bolfing. 2013. 'Multilingual Website Assessment for Accessibility: A Survey on Current Practices'. In *Proceedings of the 15th ACM SIGACCESS International Conference on Computers and Accessibility*. Bellevue, Washington (USA). http://dl.acm.org/citation.cfm?id=2513428.

Rodríguez Vázquez, Silvia, Pierrette Bouillon, and Anton Bolfing. 2014. 'Applying Accessibility-Oriented Controlled Language (CL) Rules to Improve Appropriateness of Text Alternatives for Images: An Exploratory Study'. In *Proceedings of the 9th Language Resources and Evaluation Conference*. Reykjavik (Iceland). http://www.lrec-conf.org/proceedings/lrec2014/summaries/640.html.

Rodríguez Vázquez, Silvia, and Jesús Torres del Rey. 2012. 'A Communicative Approach to Evaluate Web Accessibility Localisation Using a Controlled Language Checker: The Case of Text Alternatives for Images'. *Localisation Focus. The International Journal of Localisation.* 11 (1): 27–39.

Rösener, Christoph. 2010. 'Computational Linguistics in the Translator's Workflow: Combining Authoring Tools and Translation Memory Systems'. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids*, 1–6. CL&W '10. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1860657.1860658.

Roturier, Johann. 2006. 'An Investigation into the Impact of Controlled English Rules on the Comprehensibility, Usefulness and Acceptability of Machine-Translated Technical Documentation for French and German Users'. Dublin City University (DCU). http://doras.dcu.ie/18190/.

———. 2015. *Localizing Apps*. New York: Routledge.

Rubin, Jeffrey, and Dana Chisnell. 2008. *Handbook of Usability Testing*. Second Edition. Indianapolis, Ind: Wiley Publishing, Inc.

Saggion, Horacio, Elena Gómez Martínez, Esteban Etayo, Alberto Anula, and Lorena Bourg. 2011. 'Text Simplification in Simplext: Making Texts More Accessible'. *Procesamiento Del Lenguaje Natural*, no. 47 (September): 341–42.

Saldanha, Gabriela, and Sharon O'Brien. 2014. *Research Methodologies in Translation Studies*. Manchester, UK: Routledge.

Sandrini, Peter. 2005. 'Website Localization and Translation'. In *Challenges of Multidimensional Trnaslation (MuTra)*. EU-High-Level Scientific Conference Series. Saarbrücken, Germany. www.euroconferences.info/proceedings/2005_Proceedings/2005_Sandrini_Peter.pdf.

———. 2008. 'Localization and Translation'. Edited by Heidrun Gerzymisch-Arbogast, Gerhard Budin, and Gertrud Hofer. *MuTra Journal*, LSP Translation Scenarios. Selected Contributions to the EU Marie Curie Conference, Vienna 2007, 2: 167–91.

Savage, Terry Michael, and Karla E. Vogel. 2014. *An Introduction to Digital Multimedia.* Burlington: Jones & Bartlett Learning.

Savourel, Yves. 2001. *XML Internationalization and Localization.* 1 edition. Indianapolis, Ind: Sams Publishing.

Schade, Amy, and Jakob Nielsen. 2013. 'E-Commerce User Experience. Vol. 10: International Users'. Fremont, USA: Nielsen Norman Group.

Schäler, Reinhard. 2003. 'The Cultural Dimension in Software Localisation'. *Localisation Focus. The International Journal of Localisation.* 1 (2): 5–9.

———. 2007. 'Reverse Localisation'. *Localisation Focus. The International Journal of Localisation.* 6 (1): 39–48.

———. 2008. 'Linguistic Resources and Localisation'. In *Topics in Language Resources for Translation and Localisation*, edited by Elia Yuste Rodrigo, 195–214. Amsterdam: John Benjamins.

———. 2010. 'Localization and Translation'. In *Handbook of Translation Studies*, edited by Yves Gambier and Luc van Doorslaer, 1:209–14. Amsterdam: John Benjamins Publishing Co.

Schiller, Annette. 2006. 'Website Translation: Issues of Cohesion'. In *Proceedings of the 5th International AELFE Conference*, 535–41. Zaragoza: Prensas Universitarias de Zaragoza. http://dialnet.unirioja.es/servlet/libro?codigo=12483.

Schwitter, Rolf. 2002. 'English as a Formal Specification Language'. In *Proceedings of 13th International Conference on Database and Expert Systems Applications (DEXA 2002)*, 228–32. Aix-en-Provence, France.

———. 2015. 'Controlled Language'. In *The Routledge Encyclopedia of Translation Technolgoy*, edited by Chan Sin-wai, 450–64. Abingdon, Oxon: Routledge.

Schwitter, Rolf, and Marc Tilbrook. 2006. 'Writing RSS Feeds in a Machine-Processable Controlled Natural Language'. In *Proceedings of the 5th International Workshop on Controlled Language Applications (CLAW).* Cambridge, Massachusetts. www.mt-archive.info/05/CLAW-2006-Schwitter.pdf.

Sharif, Ather, Babak Forouraghi, and Shengqi Gong. 2015. 'EvoGraphs - A jQuery Plugin to Create Web Accessible Graphs'. presented at the The Graphical Web 2015, Pittsburgh, Pennsylvania, USA. http://athersharif.com/presentations/graphicalweb2015/assets/player/KeynoteDHTMLPlayer.html#0.

Shneiderman, Ben. 2000. 'Universal Usability'. *Commun. ACM* 43 (5): 84–91. doi:10.1145/332833.332843.

Singh, Nitish, and Arun Pereira. 2005. *The Culturally Customized Web Site: Customizing Web Sites for the Global Marketplace*. Elsevier Butterworth-Heinemann.

Sin-wai, Chan. 2013. 'Approaching Localization'. In *The Routledge Handbook of Translation Studies*, edited by Carmen Millán and Francesca Bartrina. New York: Routledge.

Slatin, John M. 2001. 'The Art of ALT: Toward a More Accessible Web'. *Computers and Composition* 18 (1): 73–81.

Sloan, David. 2006. 'The Effectiveness of the Web Accessibility Audit as a Motivational and Educational Tool in Inclusive Web Design'. PhD thesis. Dundee, Scotland, UK: University of Dundee.

Sloan, David, Andy Heath, Fraser Hamilton, Brian Kelly, Helen Petrie, and Lawrie Phipps. 2006. 'Contextual Web Accessibility - Maximizing the Benefit of Accessibility Guidelines'. In *Proceedings of the 2006 International Cross-Disciplinary Workshop on Web Accessibility (W4A)*, 121–31. W4A '06. Edinburgh, UK: ACM. doi:10.1145/1133219.1133242.

Smith, Jared. 2011. 'Web Accessibility and SEO'. *WebAIM - Web Accessibility in Mind*. http://webaim.org/blog/web-accessibility-and-seo/.

———. 2013. 'WebAIM's Hierarchy for Motivating Accessibility Change'. *WebAIM - Web Accessibility in Mind*. http://webaim.org/blog/motivating-accessibility-change/.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. 'A Study of Translation Edit Rate with Targeted Human Annotation'. In *In Proceedings of Association for Machine Translation in the Americas*, 223–31.

Sorge, Volker, Mark Lee, and Sandy Wilkinson. 2015. 'End-to-End Solution for Accessible Chemical Diagrams'. In *Proceedings of the 12th Web for All Conference*, 6:1–6:10. W4A '15. New York, NY, USA: ACM. doi:10.1145/2745555.2746667.

Štajner, Sanja, Richard Evans, Constantin Orăsan, and Ruslan Mitkov. 2012. 'What Can Readability Measures Really Tell Us About Text Complexity?' In *Proceedings of the Natural Language Processing for Improving Textual Accessibility (NLP4ITA) Workshop*, 14–21. Istambul, Turkey.

Sullivan, Terry, and Rebecca Matson. 2000. 'Barriers to Use: Usability and Content Accessibility on the Web's Most Popular Sites'. In *Proceedings on the 2000 Conference on Universal Usability*, 139–44. CUU '00. Arlington, USA: ACM. doi:10.1145/355460.355549.

Suojanen, Tytti, Kaisa Koskinen, and Tiina Tuominen. 2015. *User-Centered Translation.* Translation Practices Explained. New York: Routledge.

Takagi, Hironobu, and Chieko Asakawa. 2000. 'Transcoding Proxy for Nonvisual Web Access'. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies*, 164–71. Assets '00. New York, NY, USA: ACM. doi:10.1145/354324.354371.

Tang, Lisa. 2012. 'Producing Informative Text Alternatives for Images'. Saskatoon, Canada: University of Saskatchewan.

Tang, Lisa, and Jim Carter. 2011. 'Verbalizing Images'. In *HCI International 2011 – Posters' Extended Abstracts*, edited by Constantine Stephanidis, 173:394–98. Communications in Computer and Information Science. London ; New York: Springer.

Tercedor Sánchez, Maribel. 2005. 'Aspectos Culturales En La Localización de Productos Multimedia'. *Quaderns. Revista de Traducció.* 12: 151–60.

———. 2010. 'Translating Web Multimodalities: Towards Inclusive Web Localization'. *Tradumàtica*, Web localisation, , no. 8 (December). http://www.fti.uab.cat/tradumatica/revista/num8/articles/09/9art.htm.

Tercedor Sánchez, Maribel, Esperanza Alarcón Navío, Juan Antonio Prieto Velasco, and Clara Inés López Rodríguez. 2009. 'Images as Part of Technical Translation Courses: Implications and Applications'. *JoSTrans. The Journal of Specialised Translation.*, no. 11 (January): 143–68.

Tercedor Sánchez, Maribel, and Miguel Ángel Jiménez Crespo. 2007. 'Accesibilidad, Imágenes Y Traducción Técnica'. In *II Congreso AMADIS (Accesibilidad a Los Medios Audiovisuales Para Personas Con Discapacidad)*, 129–40. Madrid: Real Patronato sobre Discapacidad.

Tercedor Sánchez, Maribel, Clara Inés López Rodríguez, and Juan Antonio Prieto Velasco. 2006. 'Accesibilidad Web a Través de La Descripción de Imágenes'. In *I Congreso AMADIS (Accesibilidad a Los Medios Audiovisuales Para Personas Con Discapacidad)*, 73–82. Madrid: Real Patronato sobre Discapacidad.

Tercedor Sánchez, Maribel, and Juan Antonio Prieto Velasco. 2009. 'Aprender Con Imágenes'. In *Materiales Multimedia Para Todos. Inclusión Y Accessibilidad En Educación.*, 67–86. Traducción Accesible. Granada: Tragacanto.

Thatcher, Jim, Michael R. Burks, Christian Heilmann, Shawn Lawton Henry, Andrew Kirkpatrick, Patrick H. Lauke, Bruce Lawson, et al. 2006. *Web Accessibility: Web Standards and Regulatory Compliance.* 1 edition. Berkeley, CA : New York: FriendsofED.

Theofanos, Mary Frances, and Janice Redish. 2006. 'Guidelines for Accessible and Usable Web Sites: Observing Users Who Work With Screen Readers'. *Authors' Version of Theofanos and Redish. Reprinted and Expanded from Interactions, 10(6), 34-51*, 1–24. doi:10.1145/947226.947227.

The Paciello Group. 2015. 'HTML5 Feature Accessibility Support in Major Browsers'. *HTML5 Accessibility*. http://www.html5accessibility.com/.

Topac, Vasile, and Vasile Stoicu-Tivadar. 2012. 'Evaluation of Terminology Labeling Impact over Readability'. In *Proceedings of the W3C Easy-to-Read on the Web Syposium*, Paper 1. http://www.w3.org/WAI/RD/2012/easy-to-read/paper1/.

Torres del Rey, Jesús, and Lucia Morado Vázquez. 2015. 'XLIFF and the Translator: Why Does It Matter?' *Tradumàtica*, no. 13: 584–607.

Torres del Rey, Jesús, Lucia Morado Vázquez, Silvia Rodríguez Vázquez, and Emilio Rodríguez Vázquez de Aldana. 2014. 'Localisation Training in Translation Studies: A Communicative, Social and Object-Oriented Approach'. presented at the Second International Conference on Research into the Didactics of Translation, Barcelona, Spain, July 8.

Torres del Rey, Jesús, and Silvia Rodríguez Vázquez. 2013. 'Traducción Y Accesibilización de Discurso Público En Formato Web'. In *Traducción, Política(s), Conflicto: Legados Y Retos Para La Era Del Multiculturalismo*, edited by África Vidal Claramonte and Rosario Martín Ruano, 121–42. Interlingua. Granada: Comares.

Torres del Rey, Jesús, and Emilio Rodríguez V. de Aldana. 2013. 'Localisation Standards for Joomla! Translator-Oriented Localisation of CMS-Based Websites'. *Localisation Focus. The International Journal of Localisation.* 12 (1): 4–14.

———. 2014. 'La Localización de Webs Dinámicas: Objetos, Métodos, Presente Y Futuro'. *JoSTrans. The Journal of Specialised Translation.*, no. 21 (January): 153–76.

Trewin, Shari, Brian Cragun, Cal Swart, Jonathan Brezin, and John Richards. 2010. 'Accessibility Challenges and Tool Features: An IBM Web Developer Perspective'. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 32:1–32:10. W4A '10. Raleigh, USA: ACM. doi:10.1145/1805986.1806029.

Tukey, John W. 1949. 'Comparing Individual Means in the Analysis of Variance'. *Biometrics* 5 (2): 99–114.

UCBA. 2012. 'Handicap Visuel et Cécité : évolution En Suisse'. Union centrale suisse pour le bien des aveugles.

USERLab. 2012. *Text Alternatives Tool for Images (TATI)*. University of Saskatchewan. http://userlab.usask.ca/TATI/Instructions.php.

Velasco, Carlos A., and Shadi Abou-Zahra, eds. 2014. 'Developers' Guide to Features of Web Accessibility Evaluation Tools'. W3C Editor's Draft. https://w3c.github.io/w3c-waet/WAET.html.

Velleman, Eric, and Shadi Abou-Zahra, eds. 2014. 'Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0'. W3C Working Group Note. http://www.w3.org/TR/WCAG-EM/.

Velleman, Eric, and Thea van der Geest. 2013. 'Page Sample Size in Web Accessibility Testing: How Many Pages Is Enough?' In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 61:1–61:2. ASSETS '13. Bellevue, Washington (USA): ACM. doi:10.1145/2513383.2513408.

Vieritz, Helmut, Daniel Schilberg, and Sabina Jeschke. 2010. 'Merging Web Accessibility and Usability by Patterns'. In *Computers Helping People with Special Needs*, edited by Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, 336–42. Lecture Notes in Computer Science 6179. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-14097-6_54.

Vigo, Markel. 2009. 'Automatic Assessment of Contextual Web Accessibility from an Evaluation, Measurement and Adaptation Perspectives'. San Sebastián, Spain: University of the Basque Country.

Vigo, Markel, and Giorgio Brajnik. 2011. 'Automatic Web Accessibility Metrics: Where We Are and Where We Can Go'. *Interact. Comput.* 23 (2): 137–55. doi:10.1016/j.intcom.2011.01.001.

Vigo, Markel, Justin Brown, and Vivienne Conway. 2013. 'Benchmarking Web Accessibility Evaluation Tools: Measuring the Harm of Sole Reliance on Automated Tests'. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, 1:1–1:10. W4A '13. Rio de Janeiro, Brazil: ACM. doi:10.1145/2461121.2461124.

Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2014. 'A Picture Is Worth a Thousand (coherent) Words: Building a Natural Description of Images'. November. http://googleresearch.blogspot.co.uk/2014/11/a-picture-is-worth-thousand-coherent.html.

von Ahn, Luis, and Laura Dabbish. 2004. 'Labeling Images with a Computer Game'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 319–26. CHI '04. New York, NY, USA: ACM. doi:10.1145/985692.985733.

von Ahn, Luis, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. 2006. 'Improving Accessibility of the Web with a Computer Game'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 79–82. CHI '06. New York, NY, USA: ACM. doi:10.1145/1124772.1124785.

Vorburger, Michael. 1999. 'ALTifier - A Web Accessibility Enhancement Tool'. École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. http://www.vorburger.ch/projects/alt/.

W3C-RDWG, Research and Development Working Group. 2014. 'Research Report on Easy to Read on the Web'. W3C Editor's Draft. http://www.w3.org/WAI/RD/2012/easy-to-read/note/ED-E2R.

Waller, Annalu, Vicki L. Hanson, and David Sloan. 2009. 'Including Accessibility Within and Beyond Undergraduate Computing Courses'. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 155–62. Assets '09. Pittsburgh, Pennsylvania, USA: ACM. doi:10.1145/1639642.1639670.

Wang, Ye Diana. 2012. 'A Holistic and Pragmatic Approach to Teaching Web Accessibility in an Undergraduate Web Design Course'. In *Proceedings of the 13th Annual Conference on Information Technology Education*, 55–60. SIGITE '12. New York, NY, USA: ACM. doi:10.1145/2380552.2380568.

Wassmer, Thomas. 2003. 'Tools for Localizing Multimedia Applications: Considerations for Working with Images, Flash Movies and Digital Video in Internet Presentations'. *Multilingual Computing and Technology*. http://www.multilingual.com/articleDetail.php?id=658.

Watanabe, Willian M., Arnaldo Candido Jr., Marcelo A. Amâncio, Matheus de Oliveira, Thiago A. S. Pardo, Renata P. M. Fortes, and Sandra M. Aluísio. 2010. 'Adapting Web Content for Low-Literacy Readers by Using Lexical Elaboration and Named Entities Labeling'. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 8:1–8:9. W4A '10. New York, NY, USA: ACM. doi:10.1145/1805986.1805998.

Watson, Léonie. 2013. 'Using ARIA to Enhance SVG Accessibility'. *The Paciello Group*. https://www.paciellogroup.com/blog/2013/12/using-aria-enhance-svg-accessibility/.

WebAIM. 2010. 'Writing Clearly and Simply'. http://webaim.org/techniques/writing/.

———. 2013. 'Training Others - Why Accessibility? Motivating Learners to Bring About Change'. *WebAIM - Web Accessibility in Mind*. http://webaim.org/articles/training/motivate.

———. 2015a. 'Longdesc Test Cases'. http://webaim.org/techniques/alttext/longdesctestcases.htm.

———. 2015b. 'Alternative Text'. September 3. http://webaim.org/techniques/alttext/.

West, Brady T., Kathleen B. Welch, Andrzej Galecki, and Brenda W. Gillespie. 2007. *Linear Mixed Models. A Practical Guide Using Statistical Software*. Chapman & Hall/CRC.

WHO. 2002. 'Towards a Common Language for Functioning, Disability and Health'. Geneva: World Health Organisation. http://www.who.int/classifications/icf/en/.

———. 2011. 'World Report on Disability'. Geneva: World Health Organisation. http://www.who.int/disabilities/world_report/2011/en/.

———. 2014. 'Visual Impairment and Blindness'. http://www.who.int/mediacentre/factsheets/fs282/en/.

———. 2016. 'International Statistical Classification of Diseases and Related Health Problems 10th Revision'. http://apps.who.int/classifications/icd10/browse/2016/en.

Winograd, Terry, and Fernando Flores. 1987. *Understanding Computers and Cognition: A New Foundation for Design.* 1st ed. Norwood, NJ: Ablex Publishing.

World Institute on Disability. 1999. 'Telecommunications Problems and Design Strategies for People with Cognitive Disabilitites'.

Wyner, Adam, Krasimir Angelov, Guntis Barzdins, Danica Damljanovic, Brian Davis, Norbert E. Fuchs, Stefan Hoefler, et al. 2010. 'On Controlled Natural Languages: Properties and Prospects'. In *Controlled Natural Language*, edited by Norbert E. Fuchs, 281–90. Lecture Notes in Artificial Intelligence 5972. Berlin, Heidelberg: Springer-Verlag.

Yesilada, Yeliz, Giorgio Brajnik, and Simon Harper. 2009. 'How Much Does Expertise Matter?: A Barrier Walkthrough Study with Experts and Non-Experts'. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 203–10. Pittsburgh, Pennsylvania, USA: ACM. doi:10.1145/1639642.1639678.

Yesilada, Yeliz, Giorgio Brajnik, Markel Vigo, and Simon Harper. 2012. 'Understanding Web Accessibility and Its Drivers'. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 19:1–19:9. W4A '12. Lyon, France: ACM. doi:10.1145/2207016.2207027.

Yesilada, Yeliz, Robert Stevens, Simon Harper, and Carole Goble. 2007. 'Evaluating DANTE: Semantic Transcoding for Visually Disabled Users'. *ACM Trans. Comput.-Hum. Interact.* 14 (3). doi:10.1145/1279700.1279704.

Yunker, John. 2003. *Beyond Borders: Web Globalization Strategies.* USA: New Riders.

Zamanian, Mostafa, and Pooneh Heydari. 2012. 'Readability of Texts: State of the Art'. *Theory and Practice in Language Studies* 2 (1): 43–53.

Appendices

Appendix A. Research web corpora

A.1 Web corpus I - Rule development

Table A.1. Descriptive summary of web corpus I - List of websites and total number of pages, images and alt texts (absent, empty, non-empty) per website

ID	Title (fr)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Ch-a1	Journal 20 minutes	www.20min.ch	126	1544	340	9	1195
Ch-a2	Journal 24 heures	www.24heures.ch	54	540	46	315	179
Ch-a3	Adecco - Services en Ressources Humaines	www.adecco.ch	11	25	2	6	17
Ch-a4	Portail du Gouvernement suisse	www.admin.ch	42	93	1	37	55
Ch-a5	Office fédéral de la santé publique	www.bag.admin.ch	35	106	11	44	51
Ch-a6	Ville de Bâle[1]	www.basel.ch	23	53	0	18	35
Ch-a7	Canton de Berne	www.be.ch	34	49	0	7	42
Ch-a8	Ville de Berne	www.berne.ch	15	135	24	53	58
Ch-a9	Office fédéral de la statistique	www.bfs.admin.ch	71	272	3	61	208
Ch-a10	Tribunal fédéral	www.bger.ch	33	150	111	9	30
Ch-a11	Ville de Bienne	www.biel-bienne.ch	46	164	54	78	32
Ch-a12	Chancellerie fédérale	www.bk.admin.ch	41	188	4	30	154
Ch-a13	Canton de Bâle-Ville	www.bs.ch	23	53	0	18	35
Ch-a14	Office fédéral des assurances sociales	www.bsv.admin.ch	27	26	0	14	12
Ch-a15	Transport en commun de Bâle	www.bvb.ch	42	175	49	125	1
Ch-a16	Tribunal administratif fédéral	www.bvger.ch	18	18	0	8	10
Ch-a17	Portail des autorités suisses	www.ch.ch	50	194	7	81	106
Ch-a18	Coop - Supermarché en ligne	www.coopathome.ch	21	278	32	91	155
Ch-a19	Département fédéral des affaires étrangères	www.eda.admin.ch	52	405	28	137	240
Ch-a20	Département fédéral de l'interieur	www.edi.adminc.h	29	59	0	18	41
Ch-a21	Département fédéral des finances	www.efd.admin.ch	33	114	0	76	38

[1] At the time of the web corpus retrieval, Ch-a6 and Ch-a13 were two different websites. Their content has been later merged into www.bs.ch (Last access: 27th July 2015).

Table A.1. Descriptive summary of web corpus I (continued)

ID	Title (fr)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Ch-a22	Département fédéral de justice et police	www.ejpd.admin.ch	30	57	1	25	31
Ch-a23	École polytechnique fédérale de Lausanne	www.epfl.ch	43	200	34	65	101
Ch-a24	Département fédéral de l'économie[2]	www.evd.admin.ch	30	85	0	37	48
Ch-a25	Canton de Fribourg	www.fr.ch	1	7	0	0	7
Ch-a26	Communauté tarifaire de Fribourg	www.frimobil.ch	8	84	16	36	32
Ch-a27	Canton de Genève	www.ge.ch	54	201	10	80	111
Ch-a28	jobs.ch - Site d'emploi en Suisse	www.jobs.ch	38	1535	45	1	1489
Ch-a29	jobwinner.ch - Site d'emploi en Suisse	www.jobwinner.ch	6	18	9	1	8
Ch-a30	Canton du Jura	www.ju.ch	32	217	1	1	215
Ch-a31	Ville de Lausanne	www.lausanne.ch	15	190	34	102	54
Ch-a32	Journal Le Matin	www.lematin.ch	22	649	44	28	577
Ch-a33	LeShop.ch - Supermarché en ligne	www.leshop.ch	3	27	8	3	16
Ch-a34	Journal Le Temps	www.letemps.ch	29	476	34	283	159
Ch-a35	Ville de Lugano	www.lugano.ch	32	174	13	100	61
Ch-a36	Office fédéral de météorologie et de climatologie MétéoSuisse	www.meteosuisse.admin.ch	67	710	26	52	632
Ch-a37	MySwitzerland.com - Portail de tourisme	www.myswitzerland.com	78	1334	82	830	422
Ch-a38	Canton de Neuchâtel	www.ne.ch	7	9	5	0	4
Ch-a39	Parlement suisse	www.parlament.ch	29	90	35	3	52
Ch-a40	La Poste - Société de service postal	www.post.ch	39	66	3	45	18
Ch-a41	Chemins de fer fédéraux suisses	www.sbb.ch	70	303	28	132	143
Ch-a42	Swisscom - Télécommunications	www.swisscom.ch	64	309	49	131	129

[2] The department has changed since and, as of July 2015, is known as the Federal Department of Economic Affairs, Education and Research. The URL has changed as well: www.wbf.admin.ch (Last access: 27th July 2015).

Table A.1. Descriptive summary of web corpus I (continued)

ID	Title (fr)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Ch-a43	Portail officiel d'informations sur la Suisse[3]	www.swissworld.org	52	127	0	60	67
Ch-a44	Transports publics lausannois	www.t-l.ch	35	36	15	7	14
Ch-a45	Transports publics genevois	www.tpg.ch	60	196	20	53	123
Ch-a46	Département fédéral de l'environnement, des transports, de l'énergie et de la communication	www.uvek.admin.ch	37	84	5	29	50
Ch-a47	Département fédéral de la défense, de la protection de la population et des sports	www.vbs.admin.ch	56	299	0	22	277
Ch-a48	Canton de Vaud	www.vd.ch	48	141	0	98	43
Ch-a49	Ville de Genève	www.ville-geneve.ch	83	227	1	42	184
Ch-a50	Ville de Winterthur	www.ville.winterthur.ch	19	22	0	0	22
Ch-a51	Canton du Valais	www.vs.ch	18	258	37	63	158
Ch-a52	Canton de Zoug	www.zg.ch	7	6	0	2	4
		Total	**1,938**	**12,778**	**1,267**	**3,566**	**7,945**

[3] At the time of this thesis writing, the content of this website had been integrated into Ch-a19: https://www.eda.admin.ch/aboutswitzerland/fr/home.html (Last access: 27th July 2015).

A.2 Web corpus II - Exploratory pilot study

Table A.2. Descriptive summary of web corpus II - List of websites and total number of pages, images and alt texts (absent, empty, non-empty) per website

ID	Title (fr)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Ch-b1	Université de Genève	www.unige.ch	76	270	82	11	177
Ch-b2	Université de Lausanne	www.unil.ch	75	329	17	178	134
Ch-b3	Université de Fribourg	www.unifr.ch	55	251	56	136	59
Ch-b4	Université de Neuchâtel	www.unine.ch	211	703	4	551	148
Ch-b5	Faculté de traduction et interprétation de l'Université de Genève	www.unige.ch/traduction-interpretation/faculte.html	52	68	16	1	51
Ch-b6	Département fédéral de l'économie, de la formation et de la recherche	www.wbf.admin.ch	32	36	0	11	25
Ch-b7	Radio télévision suisse	www.rts.ch	89	1062	5	260	797
Ch-b8	Transports publics lausannois	www.t-l-ch	28	32	15	5	12
Ch-b9	Canton de Neuchâtel	www.ne.ch	119	147	40	77	30
Ch-b10	Canton du Jura	www.jura.ch	43	132	1	2	129
Ch-b11	Ville de Bienne	www.biel-bienne.ch	70	163	5	3	155
Ch-b12	Tribunal fédéral	www.bger.ch	51	14	0	0	14
		Total	**901**	**3,207**	**241**	**1,235**	**1,731**

A.3 Web corpus III - Rule selection for evaluation

Table A.3. Descriptive summary of the Beligum sub-corpus from web corpus III. List of websites and total number of pages, images and alt texts (absent, empty, non-empty) per website

ID	Title (fr)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Be1	Service public fédéral Affaires étrangères	www.diplomatie.belgium.be	122	148	0	26	122
Be2	Portail de l'Agriculture wallonne	www.agriculture.wallonie.be	46	531	94	112	325
Be3	Portail des services publics belges	www.belgium.be	96	126	0	54	72
Be4	Brussels airlines	www.brusselsairlines.com	203	220	46	13	161
Be5	Fédération Royale Belge des transporteurs et des prestataires de services logistiques	www.febetra.be	172	25	2	14	9
Be6	Société régionale wallonne du transport	www.infotec.be	49	107	18	56	33
Be7	Chambre des Représentants de Belgique	www.lachambre.be	100	256	126	96	34
Be8	Portail de la Wallonie	www.wallonie.be	69	417	0	279	138
Be9	Portail belge pour la recherche et l'innovation	www.research.be	28	109	69	26	14
Be10	Institut royal des Sciences naturelles de Belgique	www.naturalsciences.be	71	260	3	20	237
Be11	Société nationale des chemins de fer belges	www.belgianrail.be	166	543	0	16	527
Be12	Société des transports intercommunaux de Bruxelles	www.stib-mivb.be	88	440	0	0	440
Be13	Université de Liège	www.ulg.ac.be	89	270	14	11	245
Be14	Université de Mons	www.portail.umons.ac.be	140	1219	154	413	652
Be15	Université libre de Bruxelles	www.ulb.ac.be	103	265	119	82	64
Be16	Ville de Louvain	www.leuven.be	139	800	162	546	92
Be17	Ville de Liège	www.liege.be	27	52	7	1	44
Be18	Ville de Mons	www.mons.be	193	1,002	16	78	908
Be19	Ville de Bruxelles	www.bruxelles.be	93	249	23	8	218

Table A.3. Descriptive summary of the Beligum sub-corpus from web corpus III (continued)

ID	**Title (fr)**	**URL**	**Pages**	**<img>**	**No alt**	**Empty alt**	**Non-empty alt**
Be20	Ville de Charleroi	www.charleroi.be	98	204	6	38	160
Be21	Journal L'Echo	www.lecho.be	174	530	81	230	219
Be22	Journal Le Soir	www.lesoir.be	156	926	64	65	797
		Total	**2,422**	**8,699**	**1,004**	**2,184**	**5,511**

Table A.4. Descriptive summary of the Canada sub-corpus from web corpus III. List of websites and total number of pages, images and alt texts (absent, empty, non-empty) per website

ID	**Title (fr)**	**URL**	**Pages**	**<img>**	**No alt**	**Empty alt**	**Non-empty alt**
Ca1	Assemblée Nationale du Québec	www.assnat.qc.ca	88	261	6	114	141
Ca2	Assemblée législative de l'Ontario	www.ontla.on.ca	32	22	0	6	16
Ca3	Universités Canada. Porte-parole des universités canadiennes	www.aucc.ca	41	115	5	54	56
Ca4	Gouvernement du Canada	www.canada.ca	71	52	0	48	4
Ca5	Gouvernement du Québec	www.gouv.qc.ca	33	69	1	2	66
Ca6	Gouvernement du Nouveau-Brunswick	www2.gnb.ca	106	454	52	39	363
Ca7	Gouvernement de l'Ontario	www.ontario.ca	27	100	0	48	52
Ca8	Ville de Montréal	www.ville.montreal.qc.ca	71	317	4	87	226
Ca9	OC Transpo. Transports urbains d'Ottawa	www.octranspo1.com	55	217	10	38	169
Ca10	Ville d'Ottawa	www.ottawa.ca	124	216	69	3	144
Ca11	Parlement du Canada	www.parl.gc.ca	31	133	17	5	111
Ca12	Gouvernement de Manitoba	www.gov.mb.ca	50	187	14	14	159
Ca13	Santé et Services sociaux des Territoires du Nord-Ouest	www.hss.gov.nt.ca	82	69	1	19	49

Table A.4. Descriptive summary of the Canada sub-corpus from web corpus III (continued)

ID	Title (fr)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Ca14	Université Saint-Paul d'Ottawa	www.ustpaul.ca	242	277	10	77	190
Ca15	RTC - Réseau du transport de la Capitale	www.rtcquebec.ca	64	235	54	154	27
Ca16	Université de Montréal	www.umontreal.ca	22	200	6	43	151
Ca17	Université du Québec	www.uquebec.ca	61	414	212	95	107
Ca18	Ville du Québec	www.ville.quebec.qc.ca	65	263	52	21	190
Ca19	Air Canada	www.aircanada.com	65	192	99	25	68
Ca20	Via Rail Canada - Société ferroviaire	www.viarail.ca	69	245	2	41	202
Ca21	Journal Le Devoir	www.ledevoir.com	189	2,123	58	240	1,825
Ca22	Le Journal de Montréal (24hrs Montréal)	www.journaldemontreal.com[1]	115	2,315	226	911	1,178
		Total	**1,703**	**8,476**	**898**	**2,084**	**5,494**

[1] General landing page: www.24hrs.ca

A.4 Web corpus IV - Image and text alternatives selection

Table A.5. Descriptive summary of the Ireland sub-corpus from web corpus IV. List of websites and total number of pages, images and alt texts (absent, empty, non-empty) per website

ID	Title (en)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Ie1	Aer Lingus	www.aerlingus.com	76	140	70	1	69
Ie2	Environmental Protection Agency	www.epa.ie	111	153	11	18	124
Ie3	The Consumers' Association of Ireland	www.thecai.ie	28	99	2	2	95
Ie4	Irish Aviation Authority	www.iaa.ie	27	35	0	0	35
Ie5	Food Safety Authority of Ireland	www.fsai.ie	47	52	1	8	43
Ie6	Dublin Bus - Public Transport Service	www.dublinbus.ie	95	674	98	36	540
Ie7	Cork City Council	www.corkcity.ie	88	157	0	25	132
Ie8	Dublin City Council	www.dublincity.ie	115	154	1	18	135
Ie9	Official city portal for Dublin	www.dublin.ie	46	205	16	18	171
Ie10	Limerick City and County Council	www.limerick.ie	294	1,255	19	235	1,001
Ie11	Department of Foreign Affairs and Trade	www.dfa.ie	107	306	2	61	243
Ie12	Department of Health	www.health.gov.ie	68	30	6	0	24
Ie13	Institute of Publich Health (IPH)	www.publichealth.ie	47	35	2	8	25
Ie14	Houses of Oireachtas	www.oireachtas.ie	129	105	2	24	79
Ie15	Irish Rail - Rrailway system operator	www.irishrail.ie	71	67	32	10	25
Ie16	University College Dublin	www.ucd.ie	41	208	17	51	140
Ie17	Cork Institute of Technology (CIT)	www.cit.ie	156	1,437	4	129	1,304
Ie18	National University of Ireland Galway	www.nuigalway.ie	51	231	0	35	196
Ie19	University of Limerick	www.ul.ie	125	454	81	181	192
Ie20	Trinity College Dublin	www.tcd.ie	34	121	20	48	53
Ie21	The Irish Times	www.irishtimes.com	245	2,351	69	274	2,008
Ie22	Irish Mirror	www.irishmirror.ie	235	1,945	84	1,138	723
		Total	**2,236**	**10,214**	**537**	**2,320**	**7,357**

Table A.6. Descriptive summary of the South Africa sub-corpus from web corpus IV. List of websites and total number of pages, images and alt texts (absent, empty, non-empty) per website

ID	Title (en)	URL	Pages	<img>	No alt	Empty alt	Non-empty alt
Za1	Government of South Africa	www.gov.za	107	82	0	0	82
Za2	Parliament of the Republic of South Africa	www.parliament.gov.za	281	1,787	334	300	1,153
Za3	City of Johannesburg	www.joburg.org.za	130	386	57	13	316
Za4	City of Cape Town	www.capetown.gov.za	72	246	10	35	201
Za5	National Department of Health	www.health.gov.za	49	21	20	0	1
Za6	MyCiTi - Cape Town Integrated Rapid Transit (IRT) system	www.myciti.org.za	68	205	86	59	60
Za7	Brand South Africa's Portal	www.southafrica.info	59	582	191	62	329
Za8	City of Tshwane	www.tshwane.gov.za	98	247	0	0	247
Za9	University of Cape Town	www.uct.ac.za	368	463	34	39	390
Za10	Cape Peninsula University of Technology	www.cput.ac.za	265	179	23	35	121
Za11	University of South Africa	www.unisa.ac.za	31	106	23	4	79
Za12	South African Railways	www.southafricanrailways.co.za	13	246	0	52	194
Za13	Portal of the eThekwini Municipality	www.durban.gov.za	59	1,474	102	26	1,346
Za14	University of Johannesburg	www.uj.ac.za	442	1,930	30	619	1,281
Za15	University of Western Cape	www.uwc.ac.za	115	591	309	145	137
Za16	Western Cape Government Portal	www.westerncape.gov.za	86	195	0	0	195
Za17	Journal Mail & Guardian	www.mg.co.za	529	2,412	306	1,792	314
Za18	Times LIVE news portal	www.timeslive.co.za	88	148	31	34	83
Za19	Stellenbosch University	www.sun.ac.za	106	227	114	78	35
Za20	Tshwane University of Technology	www.tut.ac.za	103	112	76	19	17
Za21	Municial Directory of South Africa	www.localgovernment.co.za	28	171	128	0	43
Za22	South African Department of Labour	www.labour.gov.za	44	47	0	11	36
		Total	**3,141**	**11,857**	**1,874**	**3,323**	**6,660**

Appendix B. Experimental website

B.1 Source text alternatives

Table B.1. List of 130 text alternatives included in each web file and associated alt type or rule violated

Image ID	Text alternative (en)	Rule/alt type	HTML file
I1	banner	U4-decor	index.html
I2	Facebook	FAA3-social	index.html
I3	Share	FAA3-social	index.html
I4	Pinterest	FAA3-social	index.html
I5	Newsletter	FAA3-social	index.html
I6	Video	FAA3-social	index.html
I7	Together against poverty	app-alt	index.html
I8	Presentation Page	FAC1-accueil	index.html
I9	Home	FAC1-accueil	index.html
I10	UNV Logo	DDR3-logo	index.html
I11	separation line	FDA3-structure	index.html
I12	Company Logo	DDR3-logo	index.html
I13	spacer	FDA3-structure	index.html
I14	Print page	FAA4-print	index.html
I15	[Not alt]	no-alt	index.html
I16	trans	FDA3-structure	index.html
I17	Fifa president Sepp Blatter. Photo: Adam Davy/PA	DDA1-credits	index.html
I18	Untitled	U2-placeholder	index.html
I19	Pic shows collect picture of Jothini Sriskandapalan	DDR2-image	index.html
I20	[Not alt]	no-alt	index.html
I21	Students representing Zimbabwe dancing on stage	app-alt	index.html
I22	Pictured: Provost Dr Patrick Prendergast with the Board members and staff	DDR2-image	index.html
I23	Bray Head (Photo: Robert Pritchard)	DDA1-credits	index.html
I24	Image	U2-placeholder	index.html
I25	Graduate student from UJ. Photographer: Jason Alden/Bloomberg	DDA1-credits	index.html
I26	Picture of a DT Bus	DDR2-image	index.html
I27	[alt EMPTY]	alt-empty	index.html
I28	Report of Ad Hoc comittee (.PDF)	FAA1-files	index.html

Table B.1. List of 130 source text alternatives (continued)

Image ID	Text alternative (en)	Rule/alt type	HTML file
I29	Printer	FAA4-print	index.html
I30	[alt EMPTY]	alt-empty	index.html
I31	Download	FAA1-files	index.html
I32	printable format	FAA4-print	index.html
I33	[Not alt]	no-alt	index.html
I34	Acrobat Reader	FAA1-files	index.html
I35	For print	FAA4-print	index.html
I36	[Not alt]	no-alt	index.html
I37	decoration	U4-decor	index.html
I38	thumb	U4-decor	index.html
I39	pictogram	U4-decor	index.html
I40	[alt EMPTY]	alt-empty	index.html
I41	#	U2-placeholder	index.html
I42	Photo Thumb	U4-decor	index.html
I43	Title	U2-placeholder	index.html
I44	Split	FDA3-structure	index.html
I45	logo	DDR3-logo	index.html
I46	cover photo	U4-decor	index2.html
I47	Find us on Facebook Logo	DDR3-logo	index2.html
I48	Insert content here	U2-placeholder	index2.html
I49	Tumblr	FAA3-social	index2.html
I50	Email	FAA3-social	index2.html
I51	RSS	FAA3-social	index2.html
I52	Page Title	FAC1-accueil	index2.html
I53	Main Page	FAC1-accueil	index2.html
I54	Logo 34x33	DDR3-logo	index2.html
I55	homepage	FAC1-accueil	index2.html
I56	Break	FDA3-structure	index2.html
I57	main page of the website	FAC1-accueil	index2.html
I58	transparent	FDA3-structure	index2.html
I59	Print	FAA4-print	index2.html
I60	space	FDA3-structure	index2.html
I61	[alt EMPTY]	alt-empty	index2.html
I62	Academic Honour for UL Professor image	DDR2-image	index2.html
I63	Photograph: Daragh Mc Sweeney/Provision	DDA1-credits	index2.html
I64	A woman speaking	app-alt	index2.html
I65	Graduation Picture	DDR2-image	index2.html
I66	Pictured is President of Ireland Michael D. Higgins and his wife Sabina	DDR2-image	index2.html
I67	Image of a refugee	DDR2-image	index2.html
I68	[Not alt]	no-alt	index2.html
I69	Diego Costa runs with the ball	app-alt	index2.html

Table B.1. List of 130 source text alternatives (continued)

Image ID	Text alternative (en)	Rule/alt type	HTML file
I70	A student poses for a photo in front of the university	app-alt	index2.html
I71	Iveagh House © Michael Foley photography	DDA1-credits	index2.html
I72	Document Logo	DDR3-logo	index2.html
I73	scanned PDF file	FAA1-files	index2.html
I74	PDF version for print	FAA4-print	index2.html
I75	Embedded Image	U2-placeholder	index2.html
I76	Plan (in pdf format)	FAA1-files	index2.html
I77	(printer-friendly version)	FAA4-print	index2.html
I78	Logos	DDR3-logo	index2.html
I79	Order Paper pdf Format	FAA1-files	index2.html
I80	PDF (printable)	FAA4-print	index2.html
I81	[alt EMPTY]	alt-empty	index2.html
I82	[alt EMPTY]	alt-empty	index2.html
I83	blank	U2-placeholder	index2.html
I84	News article image	DDR2-image	index2.html
I85	pictogram of a bus	U4-decor	index2.html
I86	test 1	U2-placeholder	index2.html
I87	bookcover 1	U4-decor	index2.html
I88	This is an image of the air quality index for health	DDR2-image	index2.html
I89	Line	FDA3-structure	index2.html
I90	collage	U4-decor	index2.html
I91	[alt EMPTY]	alt-empty	index3.html
I92	Welcome page	FAC1-accueil	index3.html
I93	mobile site	FAC1-accueil	index3.html
I94	Share this site	FAA3-social	index3.html
I95	Follow us	FAA3-social	index3.html
I96	[alt EMPTY]	alt-empty	index3.html
I97	[Not alt]	no-alt	index3.html
I98	Logo Instagram	DDR3-logo	index3.html
I99	Send this page by email	app-alt	index3.html
I100	edition cover	U4-decor	index3.html
I101	[Not alt]	no-alt	index3.html
I102	Mail & Guardian Logo	DDR3-logo	index3.html
I103	UJ Home	FAC1-accueil	index3.html
I104	[Not alt]	no-alt	index3.html
I105	slide left	FDA3-structure	index3.html
I106	Landing Page	FAC1-accueil	index3.html
I107	Printer friendly page	FAA4-print	index3.html
I108	[alt EMPTY]	alt-empty	index3.html

Table B.1. List of 130 source text alternatives (continued)

Image ID	Text alternative (en)	Rule/alt type	HTML file
I109	*	U2-placeholder	index3.html
I110	The Central Bank of Ireland's former headquarters on Dame Street. Photograph: Matt Kavanagh	DDA1-credits	index3.html
I111	Photo: Scott Halleran/Getty Images	DDA1-credits	index3.html
I112	alt	U2-placeholder	index3.html
I113	Rachel O'Byrne as Mary Queen of Scots (courtesy of Keith Dixon)	DDA1-credits	index3.html
I114	Social entrepreneur and businesswoman Wendy Luhabe during the campaign. Image: Paul Botes	DDA1-credits	index3.html
I115	Adobe PDF format document	FAA1-files	index3.html
I116	Large PDF file	FAA1-files	index3.html
I117	PDF version	FAA1-files	index3.html
I118	[Not alt]	no-alt	index3.html
I119	Adobe Acrobat	FAA1-files	index3.html
I120	Printing of registration form.pdf	FAA4-print	index3.html
I121	Medical doctors	app-alt	index3.html
I122	an image of the landscape of Ireland	DDR2-image	index3.html
I123	[Not alt]	no-alt	index3.html
I124	[alt EMPTY]	alt-empty	index3.html
I125	chat logo 3	DDR3-logo	index3.html
I126	Useful Information for Refugees	app-alt	index3.html
I127	UNV efforts in cultural field	app-alt	index3.html
I128	More about our peace involvement	app-alt	index3.html
I129	Slide right	FDA3-structure	index3.html
I130	Several Nines (c) Studio Design	DDA1-credits	index3.html

B.2 Website source text

Table B.2. Translatable text included in the Mail & Guardian partner's web page (index.html)

Location	Text
`<head>` content	campaign, Mail&Guardian, Together Against Poverty
`<img> alt` content	[See Table B.1. Images I1-I45]
Main content (body)	The 2014 Together Against Poverty Campaign is a joint partnership between two leading African institutions, Mail&Gardian and the University of Johannesbourg, together with the United Nations Volunteers (UNV) programme. Get to know each partner's main activities and key actors in this knowledge sharing portal! Through youth mobilization, UNV and its African partners aim at fostering community-based projects for sustainable development, with the ultimate goal of poverty eradication. Designed by: Print this page Search Related Documents Press coverage reports Affiliated members Sustainable development projects Other topics Justice Agriculture Gender Elections Civil Rights People Industry Others
Disclaimer	Disclaimer: This website is built for academic research purposes (test pages) and must not be published online by the participants at this study. The author declines all responsibility regarding copyright infringements due to unauthorized online publishing of these pages.

Table B.3. Translatable text included in the University of Johannesburg partner's web page (index2.html)

Location	Text
`<head>` content	campaign, University of Johannesburg, Untitled Document
`<img> alt` content	[See Table B.1. Images I46-I90]
Main content (body)	The 2014 Together Against Poverty Campaign is a joint partnership between two leading African institutions, Mail&Gardian and the University of Johannesbourg, together with the United Nations Volunteers (UNV) programme. Get to know each partner's main activities and key actors in this knowledge sharing portal! Through youth mobilization, UNV and its African partners aim at fostering community-based projects for sustainable development, with the ultimate goal of poverty eradication. Designed by: Print this page Search How do we fight against poverty? Public information sessions Countinuos education Careers center and jobs fair Our University Board Dean Vice dean Campus Transport International Register Statistics
Disclaimer	Disclaimer: This website is built for academic research purposes (test pages) and must not be published online by the participants at this study. The author declines all responsibility regarding copyright infringements due to unauthorized online publishing of these pages.

Table B.4. Translatable text included in the United Nations Volunteers partner's web page (index3.html)

Location	Text
<head> content	campaign, UNV, Untitled Document
<img> alt content	[See Table B.1. Images I91-I130]
Main content (body)	The 2014 Together Against Poverty Campaign is a joint partnership between two leading African institutions, Mail&Gardian and the University of Johannesbourg, together with the United Nations Volunteers (UNV) programme. Get to know each partner's main activities and key actors in this knowledge sharing portal! Through youth mobilization, UNV and its African partners aim at fostering community-based projects for sustainable development, with the ultimate goal of poverty eradication. Designed by: Print this page Search Our involvement Partnership agreements Community networks Fundraising efforts Registration forms Volunteering Areas Health Land Gender Science Dialogues Refugees Culture Peace
Disclaimer	Disclaimer: This website is built for academic research purposes (test pages) and must not be published online by the participants at this study. The author declines all responsibility regarding copyright infringements due to unauthorized online publishing of these pages.

B.3 Experimental website design

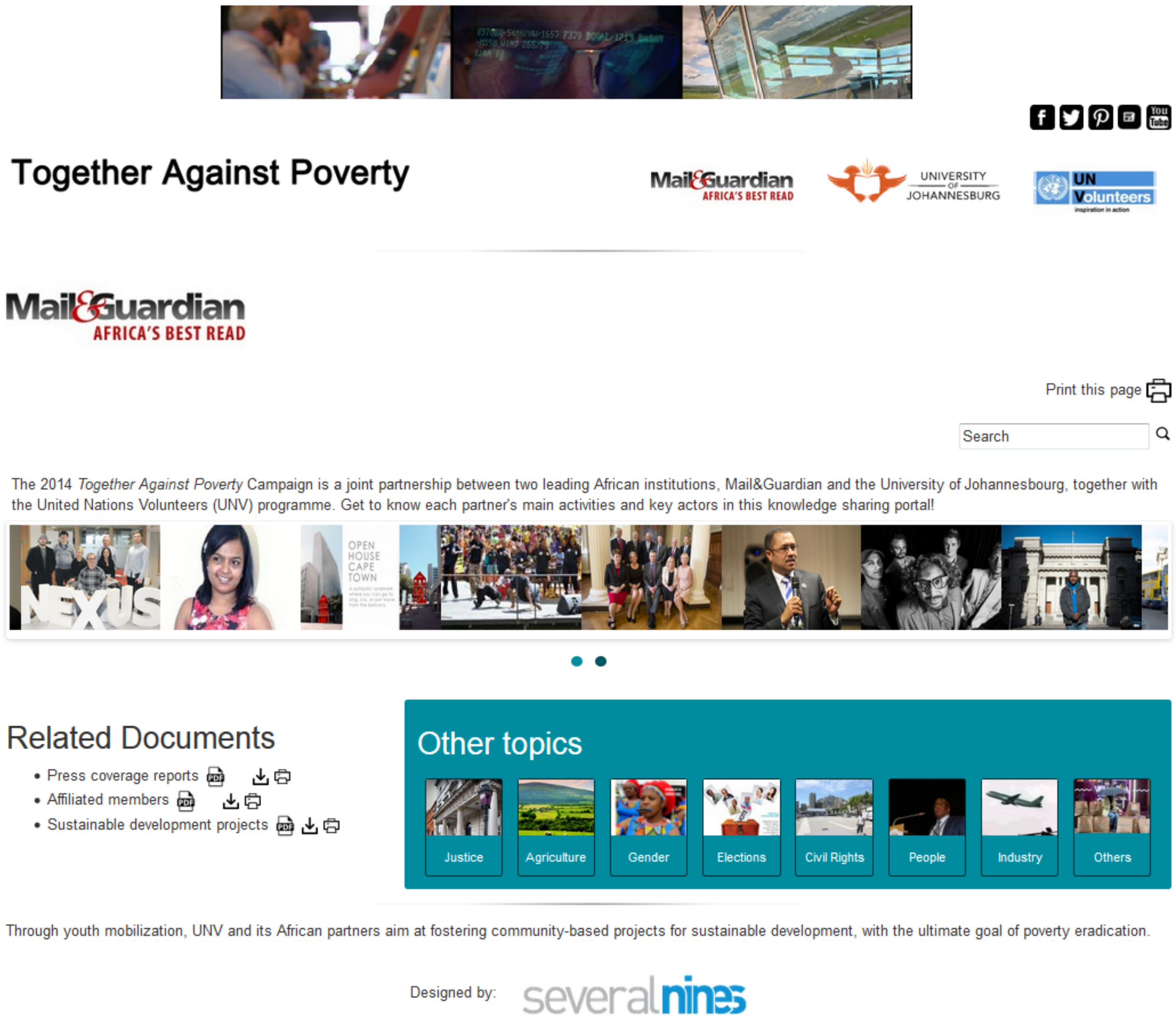


Figure B.1. Screenshot of Mail & Guardian partner's web page (index.html)

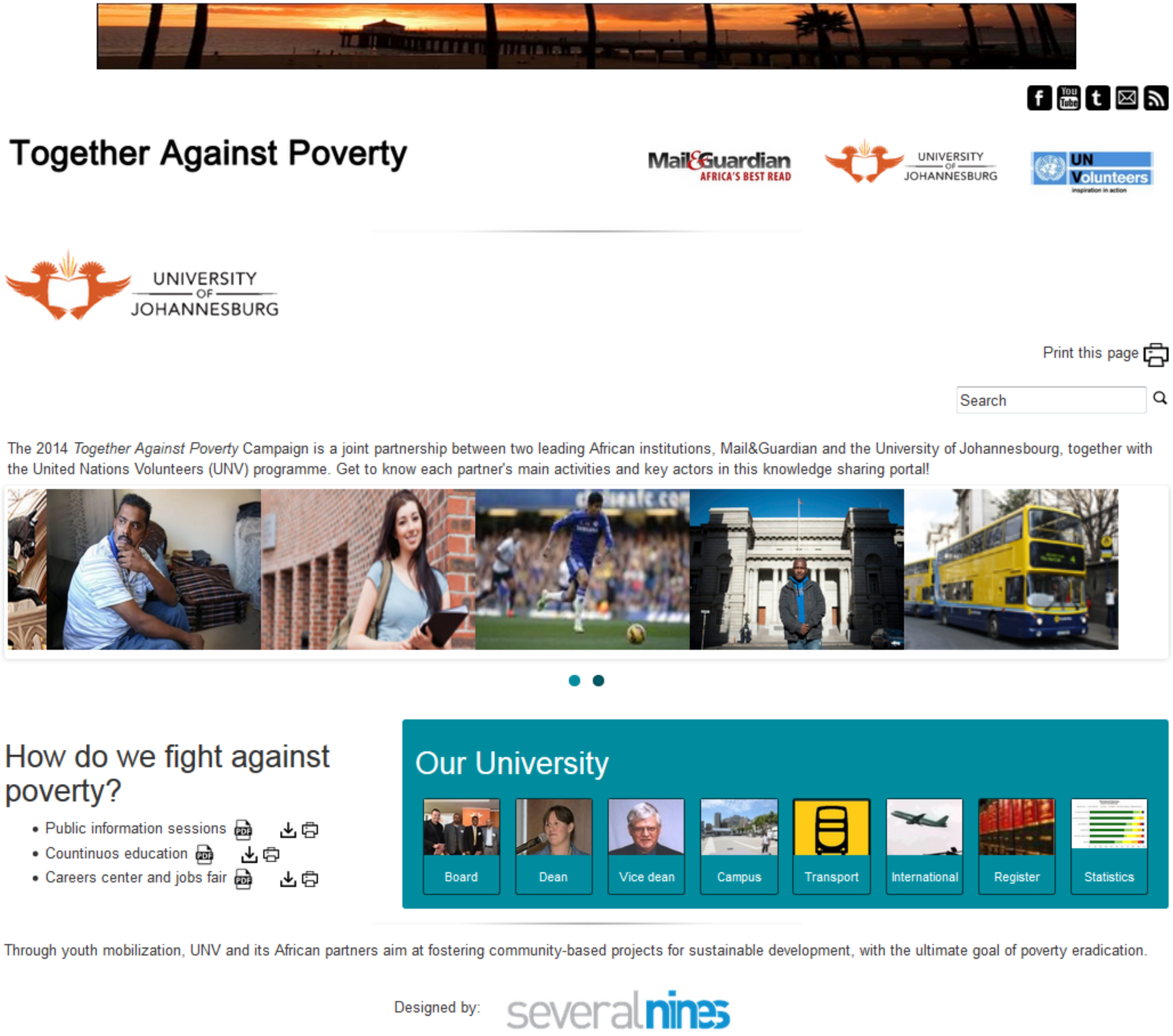


Figure B.2. Screenshot of the University of Johannesburg partner's web page (index2.html)

Figure B.3. Screenshot of the United Nations Volunteers (UNV) partner's web page (index3.html)

Appendix C. Stage 1: Call for participation

Note:

The call for participation (CFP) was launched in English and French. The English version is presented below, while the French version can be consulted in: http://diarium.usal.es/codex/fr/recherche/projects/ Last access: 28th March 2016.

CALL FOR PARTICIPATION

Research study on web localisation

At the Faculty of Translation of Interpreting (FTI) of the University of Geneva (Switzerland), we are looking for translation professionals who would be interested in taking part in a web translation-related research study. In the sections below you will find information about the profile required to participate in the study, the estimated time schedule, as well as the honorarium and other non-financial benefits associated with the study.

PARTICIPANT PROFILE

We are looking for web localisers or translation professionals with at least some basic knowledge of HTML. Participants should have **English > French** as one of their working language combinations, preferably with French as mother tongue.

No other particular background, knowledge or skills are needed.

HONORARIUM AND OTHER BENEFITS

This exercise is part of a PhD research project, for which unfortunately we have limited funds. Upon completion of the study, each participant will be economically compensated in the amount of **CHF 50** as an acknowledgement for the work done. Once the experiment is finished, you will receive a payment form in which we will request your bank account details to make the transfer.

Besides, a **free one-hour webinar** about advanced HTML best practices for translators and localisers will be offered to all participants taking part in the study.

RISKS

There are no known risks associated with the study. Taking part in it is voluntary. Should the participant feel upset or uncomfortable, he or she may quit at any time.

EXPERIMENT WORKING ENVIRONMENT

There is no travel involved. Participants will be able to remotely participate in the study. The software needed to access the experiment working environment only runs on **Windows**. For participants who do not use this operating system, we can offer remote access to a virtual Windows environment for free.

TIME SCHEDULE

The experiment will consist of two sessions of approximately 1h30 maximum each. They will be held between December 8th and December 18th 2014, depending on participants' availability.

HOW TO SIGN IN

We have opened a Doodle poll for participants to sign in. We request each participant to indicate his or her name and valid email address. However, please note that these data will not be visible to other participants and we will only use it for communication purposes prior to the study. The call for participation is **open for 4 weeks, until November 24th, 2014**. Please do not hesitate in sharing the link with other colleagues.

Doodle poll to sign in for the research study: https://doodle.com/x9v7yv4ykzqphvx5

Detailed information about the research study and the next steps to come will be sent via email to participants in due time.

If you have any questions, please do not hesitate to contact Silvia Rodríguez Vázquez at silvia.rodriguez@unige.ch.

Thank you very much in advance for your collaboration!

Silvia

Silvia Rodríguez Vázquez
Cod.eX Research Group
Department of Translation Technology (TIM)
Faculty of Translation and Interpreting, University of Geneva (UNIGE)

Appendix D. Research questionnaires

D.1 Web accessibility experts

Note:

The content of this questionnaire is presented in English, since it is the original language in which it was distributed.

Introduction

Dear respondents,

Thank you for taking the time to complete this questionnaire, which should not take you more than 20-30 minutes. It contains between 20 and 30 questions, depending on the respondent's profile.

This study is part of larger research project being carried out at the University of Geneva (Switzerland) and the University of Salamanca (Spain) that aims at defining how web localisation professionals could contribute to achieve a higher degree of web accessibility in multilingual websites, in particular regarding textual alternatives to visual content. Localisation is understood as the translation and cultural adaptation of digital products, such as websites and software, from a given source language and culture to one or several languages and cultures.

One step of this project requires participation of people that have worked on web accessibility assessment and who are not localisers. Through this questionnaire, the researchers try to understand current accessibility assessment practices of multilingual websites (websites available in two or more languages).

All questions contained in this questionnaire are strictly confidential and will only be used for scientific purposes. Personal data will not be revealed under any circumstances.

Thanks again for your collaboration!

Screening questions

1. Location

2. Age

3. Gender

    - ◻ Male
    - ◻ Female
    - ◻ N/A (not applicable)

4. Which is your first language?

5. Do you understand and fluently speak other languages different than your mother tongue?

    - ◻ Yes
    - ◻ No

6. [If yes, question 5] Which one(s)?

7. How do you rate your knowledge of web accessibility?

    a) I do not know anything about the subject.
    b) I am interested in the subject, but I do not have too much experience.
    c) I have performed some web accessibility-related jobs.
    d) I mainly perform web accessibility-related jobs.
    e) N/A (not applicable)

8. How many websites have you assessed in terms of accessibility in the past year? (it does not matter if they were monolingual or multilingual)

9. Are you currently enrolled as a student?

    a) Yes, at a high school or equivalent
    b) Yes, at a college/university
    c) Yes, in graduate school

d) Yes, as a part-time student
e) No, I am not currently enrolled as a student

Please specify in which year are you (e.g. I am in my 3rd year, Computer Science)

10. Current position (you can select more than one option)

- ☐ Researcher in the web accessibility field
- ☐ Web accessibility consultant
- ☐ Webmaster
- ☐ Web developer
- ☐ Web designer
- ☐ Other (please specify)

11. How long have you been working professionally in the web accessibility field? (e.g. Three years)

12. When you are not performing a web assessment task, do you use any assistive technology as your primary means of accessing the web?

- ☐ Yes
- ☐ No

13. [If yes, question 12] Which one(s)? (e.g. I use a screen reader to browse the web: NVDA)

Web Accessibility Evaluation Experience

This section aims to obtain information about the participants' web accessibility evaluation experience.

14. Which method do you use for web accessibility assessment? (you can select more than one option)

    a) Inspection method (i.e. the evaluator checks the website for accessibility against a given set of accessibility guidelines)
    b) Automated testing (i.e. the evaluator checks the website for accessibility using an automated accessibility tool)
    c) Screening techniques (i.e. the evaluator checks the website for accessibility using it in a way that some sensory, motor or cognitive capabilities of the user are (artificially) reduced)
    d) Subjective assessment (i.e. the evaluator hires a panel of users who are asked to explore the website autonomously and send back their opinions)
    e) User testing (i.e. real end users are individually asked to perform goal-free or goal-oriented navigation on a website while being observed by evaluators)
    f) Other (please specify)

15. Do you combine them?

    - ☐ Yes
    - ☐ No

16. [If yes, question 15] Which ones?

17. [If yes, question 15] How often?

    - ☐ Never
    - ☐ Sometimes
    - ☐ Often
    - ☐ Always
    - ☐ N/A (not applicable)

18. When performing a web accessibility assessment task, how much time do you spend on textual accessibility-related issues?

    a) I do not assess textual accessibility.
    b) I spend up to 25% of the time assessing textual accessibility.

c) I spend up to 50% of the time assessing textual accessibility.
d) I spend up to 75% of the time assessing textual accessibility.
e) I only check the website for textual accessibility because I am specialised on that area.
f) Other (please specify)

19. When performing a web accessibility assessment task, do you also take into consideration culture-related elements that should be adapted (e.g., symbols, shapes, colours, signs)?

- □ Never
- □ Sometimes
- □ Often
- □ Always
- □ N/A (not applicable)

Experience in Multilingual Web Assessment for Accessibility

This section aims to obtain information about the participants' multilingual web accessibility evaluation experience and current practices.

20. How often do you assess websites that are multilingual (i.e. that are available in two or more languages)?

a) None of the websites I evaluate are multilingual
b) A few, up to 25% of the websites I evaluate are multilingual
c) 2650% of the websites I evaluate are multilingual
d) 5175% of the websites I evaluate are multilingual
e) Almost all the websites I evaluate are multilingual (more than 75%)
f) All the websites I evaluate are multilingual

21. [If not a), question 20] When performing a web accessibility assessment task of a multilingual website, do you test just one language version of the website?

- □ Yes
- □ No

22. [If yes, question 21] Why?

- □ I only speak one language
- □ Although I speak other languages, I do not feel comfortable assessing websites that are not in my mother tongue

- ◻ I do not have time to check all available languages
- ◻ It is not necessary to check all language versions of a website
- ◻ Other (please specify)

23. [If not a), question 20] Please estimate how often the following statements describe your web accessibility assessment procedure of multilingual websites (you need to select an option in each statement)

	1. Never	2. Sometimes	3. Often	4. Always	N/A
As I check a web element or functionality, I recheck it in the other language versions.	◻	◻	◻	◻	◻
I first check one language version (all pages available), and then I check the other language versions.	◻	◻	◻	◻	◻
I just check my native language version, and if I find a major accessibility problem, I check if it is present in the other language versions too.	◻	◻	◻	◻	◻
I just check my native language version, and I assume that the other language versions have the same degree of accessibility as the one I checked.	◻	◻	◻	◻	◻

Other (please specify)

24. If you would be assessing a multilingual website for accessibility, would you appreciate receiving also the feedback of the localisation professional(s) who adapted the web to the other language(s) available?

- ◻ Yes
- ◻ No
- ◻ N/A (not applicable)

- ◻ It depends on the web element or functionality I am assessing (e.g. text, images, multimedia content…). Please kindly specify which ones.

25. Do you consider that multilingual websites and monolingual websites should be tested for accessibility differently?
    - ◻ Yes
    - ◻ No
    - ◻ N/A (not applicable)

26. Which elements of the web would be, in your opinion, worth looking at separately? (you can select more than one option)
    - ◻ Semantic structure (sections, headings…)
    - ◻ Textual content (e.g. titles, alternative text for images, body text of web pages…)
    - ◻ Navigation and hyperlinks
    - ◻ Graphical content (e.g. images)
    - ◻ Multimedia content (e.g., videos, sounds…)
    - ◻ Presentation layout (colours, style sheets…)
    - ◻ Other (please specify)

27. In the particular case of images, why would be important, in your opinion, to ask for feedback to the localisation professional who worked on the multilingual website? (you can select more than one option)
    - ◻ Because the images in the language version I am assessing might contain culture related information that should have been taken into account in the localised versions.
    - ◻ Because images might vary from one version to the other, and I cannot assess the quality of their alternative text in other languages different than my native language.
    - ◻ Because some images might have been removed or changed in certain language versions and I do not understand why.
    - ◻ I do not think that asking the localisation professional about the accessibility of images is necessary.
    - ◻ Other (please specify)

28. Who do you think should be responsible for making a multilingual website accessible? (you can select more than one option)

	Yes	No	N/A
Web developers	☐	☐	☐
Web editors	☐	☐	☐
Webmasters	☐	☐	☐
Web localisation professionals (when the website is adapted and translated into another language)	☐	☐	☐

Other (please specify)

29. What would be, in your opinion, the ideal accessibility assessment procedure to follow in the case of multilingual websites?

30. Is there anything else you would like to add that is relevant to this topic? Please feel free to share any thoughts with us.

Thank you

We will share with you the results of this survey at the end of the study. For further information, please contact:

Silvia Rodríguez Vázquez
Multilingual Information Processing Department
Faculty of Translation and Interpreting
University of Geneva

Email: Silvia.Rodriguez@unige.ch

Please, leave us your email address if you want us to send you the results at the end of the study.

D.2 Participants screening - Web localisation experiment

Note:

The content of the screening questionnaire launched to recruit participants for Stage 1 of the experimental study (see Chapter 5) is presented in French, since it is the original language in which it was distributed.

Introduction

Chers participants,

Merci de prendre le temps de remplir ce questionnaire avant de participer à l'étude. Cela ne devrait pas vous prendre plus de 15 minutes.

Toutes les questions contenues dans ce questionnaire sont strictement confidentielles et ne seront utilisées qu'à des fins scientifiques. Les données personnelles ne seront révélées en aucune circonstance.

En outre, veuillez noter que le but de cette étude n'est pas d'évaluer les participants individuellement. Nous savons que vous êtes de bons professionnels de la traduction! Nous devons collecter ces données uniquement pour des raisons liées à la validité de la recherche.

Merci de votre collaboration !

Silvia

Profil personnel

Numéro d'identification du participant (veuillez utiliser le numéro d'identification qui vous a été attribué)

1. Lieu (où habitez-vous ?)
2. Nationalité (par exemple, française)
3. Âge

4. Sexe

   - ☐ Femme
   - ☐ Homme
   - ☐ S/O (sans opinion)

5. Quelle est votre langue maternelle ?

Profil professionnel

6. Quelles sont vos combinaisons linguistiques ? (par exemple, anglais > français, espagnol > français)

7. Profession actuelle (si nécessaire, vous pouvez choisir plus d'une option)

   - ☐ Traducteur
   - ☐ Réviseur
   - ☐ Ingénieur en localisation (localisation engineer)
   - ☐ Testeur (tester)
   - ☐ Terminologue
   - ☐ Webmaster
   - ☐ Éditeur de contenus Web
   - ☐ Gestionnaire de projets de traduction (project manager)
   - ☐ Autre (veuillez préciser)

8. Parallèlement, suivez-vous une formation ? (par exemple, une formation continue, une maîtrise...)

   - ☐ Oui
   - ☐ Non

9. [Si oui, question 8] Quel type de formation suivez-vous ? (par exemple, « je suis une formation continue sur la traduction automatique (6 semaines de cours) »; « je suis dans ma 2e année d'une maîtrise en traduction »...)

10. Depuis combien de temps travaillez-vous professionnellement dans le domaine la traduction ? (par exemple, trois ans)

11. Avez-vous travaillé au cours de l'année écoulée (novembre 2013novembre 2014) sur des mandats de localisation web ?

    - ◻ Oui
    - ◻ Non

12. [Si oui, question 11] Veuillez préciser si ces mandats de localisation web font partie de votre activité professionnelle principale ou s'il s'agit de mandats ponctuels.

    a. Les mandats de localisation web constituent mon activité professionnelle principale (plus de 75% de mes mandats sont liés à la localisation web)
    b. Les mandats de localisation web représentent un pourcentage élevé de mon activité professionnelle (entre 51% et 75% de mes mandats sont liés à la localisation web)
    c. Les mandats de localisation web représentent un pourcentage considérable de mon activité professionnelle (entre 26% et 50% de mes mandats sont liés à la localisation web)
    d. Les mandats de localisation web représentent un pourcentage faible de mon activité professionnelle (entre 1% et 25% de mes mandats sont liés à la localisation web)
    e. Les mandats de localisation web sur lesquels je travaille actuellement sont exceptionnels Je ne travaille pas habituellement sur des mandats de localisation web.
    f. Autre (veuillez préciser)

13. Avez-vous travaillé sur des mandats de localisation web dans le passé (avant novembre 2013) ?
    a. Oui, c'était mon activité professionnelle principale (plus de 75% de mes mandats étaient liés à la localisation web)
    b. Oui, les mandats de localisation web représentaient un pourcentage élevé de mon activité professionnelle (entre 51% et 75% de mes mandats étaient liés à la localisation web)
    c. Oui, les mandats de localisation web représentaient un pourcentage considérable de mon activité professionnelle (entre 26% et 50% de mes mandats étaient liés à la localisation web)
    d. Oui, les mandats de localisation web représentaient un pourcentage faible de mon activité professionnelle (25% de mes mandats étaient liés à la localisation web)
    e. Oui, mais les mandats de localisation web ne représentaient pas mon activité professionnelle principale (entre 1% et 25% de mes mandats étaient liés à la localisation web)

f. Oui, mais j'ai accepté des mandats de localisation de façon temporaire. Je ne travaille pas habituellement sur des mandats de localisation web.
g. Non, je n'ai jamais travaillé sur des mandats de localisation web.
h. Autre (veuillez préciser)

14. Travaillez-vous à titre d'indépendant (freelance) ?

    a. Oui
    b. Non
    c. Ponctuellement

15. [Si a) ou c), question 14] Veuillez indiquer le(s) type(s) de client(s) pour lequel/lesquels vous travaillez à titre d'indépendant (vous pouvez choisir plus d'une option)

    - ◻ Client privé
    - ◻ Prestataire de services linguistiques (LSP)
    - ◻ Agence de traduction (multinationale)
    - ◻ Département de traduction dans une agence/entreprise/société qui ne se dédie pas uniquement aux services de traduction
    - ◻ ONG
    - ◻ Organisation internationale
    - ◻ Autre (veuillez préciser)

16. [Si b), question 14] Veuillez indiquer qui est votre employeur :

    - ◻ Prestataire de services linguistiques (LSP)
    - ◻ Agence de traduction (multinationale)
    - ◻ Département de traduction dans une agence/entreprise/société qui ne se dédie pas uniquement aux services de traduction
    - ◻ ONG
    - ◻ Organisation internationale
    - ◻ Autre (veuillez préciser)

17. Avez-vous reçu une formation spécifique sur la localisation web ?

    - ◻ Oui
    - ◻ Non

18. [Si oui, question 17] Pourriez-vous nous donner plus d'informations sur cette formation ? (par exemple, « J'ai pris un cours spécifique sur la localisation web

dans le cadre de mes études de traduction / ma maîtrise » ; « J'ai suivi des webinaires en ligne » ; « J'ai participé à des séminaires sur la localisation web au travail » ...)

19. Travaillez-vous avec des outils de traduction assistée par ordinateur (TAO) ?

- ◻ Oui
- ◻ Non

20. [Si oui, question 19] Quel(s) outil(s) de TAO utilisez-vous ? (par exemple, « J'utilise SDL Trados Studio 2011 et MemoQ »)

21. Avec quelle fréquence ?

   a. Je les utilise pour un nombre très limité de mes mandats de localisation web (entre 1% et 25% de mes mandats)
   b. Je les utilise pour un nombre limité de mes mandats de localisation web (entre 26% et 50% de mes mandats)
   c. Je les utilise pour un nombre important de mes mandats de localisation web (entre 51% et 75% de mes mandats)
   d. Je les utilise pour presque tous mes mandats de localisation web (entre 76% et 99% de mes mandats)
   e. Je les utilise pour tous mes mandats de localisation web
   f. Je ne les utilise pas pour mes mandats de localisation web, mais je les utilise pour d'autres mandats non liés à la localisation web
   g. Autre (veuillez préciser)

22. Pourquoi ? (vous pouvez choisir plus d'une option)

   a. Je n'ai pas reçu de formation sur l'utilisation d'outils de TAO
   b. Je crois qu'ils ne sont pas utiles pour les mandats de localisation web
   c. Je ne possède pas de licence pour aucun outil de TAO
   d. Mon client ne me demande pas d'utiliser des outils de TAO
   e. Mon employeur ne me demande pas d'utiliser des outils de TAO
   f. Autre (veuillez préciser)

23. Utilisez-vous un autre logiciel à des fins de localisation dans votre travail quotidien ? (par exemple, Dreamweaver comme éditeur de langage HTML, Photoshop pour le traitement des images...)

- ◻ Oui
- ◻ Non

24. [Si oui, question 23] Le(s)quel(s) ?

25. [Si oui, question 23] Avec quelle fréquence ?

a. Je les utilise pour un nombre très limité de mes mandats de localisation web (entre 1% et 25% de mes mandats)
b. Je les utilise pour un nombre limité de mes mandats de localisation web (entre 26% et 50% de mes mandats)
c. Je les utilise pour un nombre important de mes mandats de localisation web (entre 51% et 75% de mes mandats)
d. Je les utilise pour presque tous mes mandats de localisation web (entre 76% et 99% de mes mandats)
e. Je les utilise pour tous mes mandats de localisation web
f. Autre (veuillez préciser)

26. Lorsque vous recevez des mandats de localisation web, dans quel(s) format(s) travaillez-vous pour la traduction du contenu textuel ? (Vous pouvez choisir plus d'une option)

- ◻ DOC / DOCX
- ◻ XLS / XLSX
- ◻ XML
- ◻ HTML
- ◻ XHTML
- ◻ Gettext PO
- ◻ XLIFF (Veuillez préciser la version de XLIFF si vous la connaissez) :
- ◻ Fichiers bilingues exclusifs des outils de TAO (par exemple, SDL .ttx) Veuillez préciser de quel outil :
- ◻ Je travaille directement dans le système de gestion de contenu du client (CMS) Veuillez préciser quel CMS :
- ◻ Je travaille directement sur une plateforme en ligne fournie par mon employeur
- ◻ Autre (veuillez préciser)

27. Êtes-vous familier avec le concept d'accessibilité du Web ?

a. Non, je n'ai jamais entendu parler de l'accessibilité du Web
b. Oui, mais je n'ai pas de connaissances sur le sujet
c. Oui, j'ai quelques connaissances de base sur le sujet
d. Oui, j'ai quelques connaissances intermédiaires sur le sujet
e. Oui, j'ai des connaissances avancées sur le sujet

f. Autre (veuillez préciser)

28. Comment évalueriez-vous vos connaissances sur le langage HTML ?

   a. Je n’ai pas de connaissances sur le langage HTML
   b. J’ai quelques connaissances de base sur le langage HTML
   c. J’ai quelques connaissances intermédiaires sur le langage HTML
   d. J’ai des connaissances avancées sur le langage HTML
   e. Autre (veuillez préciser)

29. Dans l'extrait suivant tiré d'une page HTML, quel(s) élément(s) traduiriez/modifieriez-vous ? S’il vous plaît, veuillez indiquer non seulement les balises, mais aussi le texte qui devrait être traduit et/ou modifié.

```html
<html>
<head>
<meta http-equiv="Content-Language" content="en">
<meta http-equiv="Content-Type" content="text/html; charset=windows-1252">
<meta name="keywords" content="restaurant food online takeaway take away order Indian Italian Tapas
Kebab">
<meta name="description" content="Online restaurant serving food from all over the word to take away">
<link rel="stylesheet" type="text/css" href="css/styles.css">
<title>The Tapas House</title>
</head>
<body>
<table border="0" width="72%" id="table1" bgcolor="#FFFFFF">
    <tr>
        <td><h3>What is it?</h3>
<h5>Welcome to The Tapas House. An easier way to find and order a takeaway</h5></td>
        <td width="218">
<img border="0" src="images/food_6c.jpg" width="218" height="76" alt="image/food_6c.jpg"></td>
    </tr>
</table>
<h6>Welcome | <a href="order/index.htm">Ordering</a> | <a href="help/index.htm">Help</a></h6>
<div><img src="images/print.png" width="30" height="30"></div>
</body>
</html
```

   a. Je ne voudrais pas faire de changements
   b. Je voudrais faire les changements suivants (veuillez préciser quels changements)
   c. Je voudrais faire les changements suivants, mais je préfèrerais utiliser un outil de TAO ou un éditeur de langage HTML pour être sûr que le code ne soit pas corrompu (veuillez préciser quels changements)

30. Connaissez-vous d'autres technologies Web ? (par exemple, CSS, XML, JAVA ...)

   - ◻ Oui
   - ◻ Non
   - ◻ Autre (veuillez préciser)

31. [Si oui, question 30] Laquelle/lesquelles ?

32. Y a-t-il d'autres informations ou remarques que vous aimeriez ajouter en relation avec le flux de travail que vous suivez pour vos mandats de localisation web ? Par exemple, le nombre de membres de l'équipe et leurs fonctions, des protocoles spécifiques que vous suivez... Nous vous remercions d'avance de partager avec nous vos idées et d'autres remarques que vous considérez importantes.

Comment avez-vous eu connaissance de cette étude ? (par exemple, par la liste de diffusion de l'ASTTI)

Adresse email (veuillez introduire la même adresse email que vous avez indiquée lors de l'inscription à l'étude)

Remerciement

Nous vous remercions d'avoir rempli ce questionnaire.

Dans les prochains jours, nous prendrons à nouveau contact avec vous pour vous donner plus d'informations sur le déroulement de l'étude.

Très cordialement,

Silvia

Silvia.Rodriguez@unige.ch

D.3 Web localisation task

Notes:

The content of the web localisation post-task questionnaire (Stage 1 of the experimental study, see Chapter 5) is presented in French, since it is the original language in which it was distributed. Questions 11-15 were only shown to and answered by participants from the treatment group.

Introduction

Cher participant,

Merci de prendre le temps de remplir ce questionnaire lié à l'exercice que vous avez fait pendant cette première séance. Cela ne devrait pas vous prendre plus de 1015 minutes.

Toutes les questions contenues dans ce questionnaire sont strictement confidentielles et ne seront utilisées qu'à des fins scientifiques. Les données personnelles ne seront révélées en aucune circonstance.

En outre, veuillez noter que le but de cette étude n'est pas d'évaluer les participants individuellement. Nous savons que vous êtes de bons professionnels de la traduction ! Nous devons collecter ces données uniquement pour des raisons liées à la validité de la recherche.

Merci de votre collaboration !

Silvia

Remarques générales

Numéro d'identification du participant (veuillez utiliser le numéro d'identification qui vous a été attribué)

1. Avez-vous utilisé un outil de traduction assistée par ordinateur (TAO) pendant cette première séance?
   - ◻ Oui
   - ◻ Non

2. [Si oui, question 1] Pourriez-vous nous expliquer les raisons pour lesquelles vous avez utilisé un outil de traduction assistée par ordinateur (TAO) ?

3. Avez-vous fait des changements directement sur les fichiers HTML sans passer par un outil de traduction assistée par ordinateur (TAO) ?

   - ◻ Oui
   - ◻ Non

4. [Si oui, question 3] Pourriez-vous nous expliquer les raisons pour lesquelles vous avez fait ces changements directement sur les fichiers HTML et dans quel outil vous les avez faits?

5. Pourriez-vous énumérer les éléments du site web que vous avez traduits ou modifiés ? (par exemple : « J'ai traduit le titre de la page »)

6. Si vous aviez eu le temps, auriez-vous traduit ou modifié d'autres éléments ?
   - ◻ Oui
   - ◻ Non

7. [Si oui, question 6] Le(s)quel(s) ?

8. Si ce mandat de localisation web avait été un mandat réel, auriez-vous effectué la tâche de la même manière, c'est-à-dire en suivant la même procédure de localisation ?
   - ◻ Oui
   - ◻ Non

9. [Si non, question 8] Pourquoi ?

10. À quel type de destinataires avez-vous pensé lors de vos choix de traduction ?

11. Comment définiriez-vous le webinaire que vous avez suivi avant d'effectuer cet exercice ? (vous pouvez choisir plus d'une option)
   - ◻ Pertinent
   - ◻ Non pertinent
   - ◻ Intéressant
   - ◻ Ennuyeux
   - ◻ Inspirant
   - ◻ Difficile
   - ◻ Utile
   - ◻ Superflu
   - ◻ Autre (veuillez préciser)
12. Croyez-vous que les connaissances que vous avez acquises pendant le webinaire vous ont aidé à rendre votre site localisé plus accessible ?
   - ◻ Oui
   - ◻ Non
   - ◻ S/O (sans opinion)
   - ◻ Autre (veuillez préciser)
13. [Si non, question 12] Pourquoi ?

14. Croyez-vous que les connaissances que vous avez acquises pendant le webinaire que vous avez suivi avant d'effectuer cet exercice ont rendu la tâche de localisation plus difficile?
   - ◻ Oui
   - ◻ Non
   - ◻ S/O (sans opinion)
   - ◻ Autre (veuillez préciser)
15. [Si oui, question 14] Pourquoi ?

16. Y a-t-il d’autres informations ou remarques que vous aimeriez ajouter en relation avec cette première séance de l'étude ?

Adresse email (veuillez utiliser la même adresse email que vous avez indiquée lors de l'inscription à l'étude)

Remerciement

Nous vous remercions d'avoir rempli ce questionnaire !

Silvia

Silvia.Rodriguez@unige.ch

D.4 Quality assurance task

Notes:

The content of the quality assurance post-task questionnaire (Stage 1 of the experimental study, see Chapter 5) is presented in French, since it is the original language in which it was distributed. Questions were the same both for participants from the control group and the treatment group.

Introduction

Cher participant,

Merci de prendre le temps de remplir ce questionnaire lié à l'exercice que vous avez fait pendant cette deuxième séance. Cela ne devrait pas vous prendre plus de 1520 minutes. Nous vous conseillons de le remplir juste après avoir fini la séance, mais vous pouvez aussi le remplir plus tard, si nécessaire.

Toutes les questions contenues dans ce questionnaire sont strictement confidentielles et ne seront utilisées qu'à des fins scientifiques. Les données personnelles ne seront révélées en aucune circonstance.

En outre, veuillez noter que le but de cette étude n'est pas d'évaluer les participants individuellement. Nous savons que vous êtes de bons professionnels de la traduction! Nous devons collecter ces données uniquement pour des raisons liées à la validité de la recherche.

Merci de votre collaboration !

Silvia

Remarques générales

Numéro d'identification du participant (veuillez utiliser le numéro d'identification qui vous a été attribué)

1. Avant cette deuxième séance, aviez-vous identifié les équivalents textuels pour les images web en tant que tels ? Merci d'illustrer, si possible, votre réponse par des exemples.

2. Pensez-vous que les localisateurs devraient pouvoir adapter les équivalents textuels pour les images web initialement proposés dans la source, afin d'améliorer l'accessibilité du site ? Veuillez préciser les raisons de votre réponse.

Acrolinx IQ Batch Checker

3. Comment évalueriez-vous la précision de l'outil Acrolinx IQ Batch Checker en ce qui concerne la pertinence des équivalents textuels pour les images web ?

   (Précision ou “specificity” en anglais : nombre des différents problèmes qui peuvent être détectés et décris par l'outil à propos de la pertinence des équivalents textuels pour les images web)

   a. L'outil est très précis (76%100%)
   b. L'outil est assez précis (51%75%)
   c. L'outil est plus ou moins précis (26%50%)
   d. L'outil n'est pas très précis (1%25%)
   e. L'outil n'est pas du tout précis (0%)

4. Comment évalueriez-vous la justesse de l'outil Acrolinx IQ Batch Checker en ce qui concerne la pertinence des équivalents textuels pour les images web ?

   (Justesse ou “correctness” en anglais : proportion des problèmes détectés par l'outil qui sont réellement de vrais problèmes)

   a. L'outil est très juste (76%100%)
   b. L'outil est assez juste (51%75%)
   c. L'outil est plus ou moins juste (26%50%)
   d. L'outil n'est pas très juste (1%25%)
   e. L'outil n'est pas du tout juste (0%)

5. Comment évalueriez-vous la clarté des rapports générés par l'outil Acrolinx IQ Batch Checker en ce qui concerne la pertinence des équivalents textuels pour les images du Web ?

   a. Les rapports présentent l'information d'une manière très claire
   b. Les rapports présentent l'information d'une manière assez claire
   c. Les rapports présentent l'information d'une manière plus ou moins claire
   d. Les rapports ne présentent pas l'information d'une manière très claire
   e. Les rapports ne présentent pas du tout l'information de manière claire

6. Avez-vous vérifié la documentation apportée pour chaque règle identifiée par Acrolinx IQ Batch Checker dans le rapport de vérification ?
   a. Je n'ai vérifié la documentation d'aucune des règles identifiées dans les rapports
   b. J'ai vérifié la documentation des règles dans environ 25% des cas
   c. J'ai vérifié la documentation des règles dans environ 50% des cas
   d. J'ai vérifié la documentation des règles dans environ 75% des cas
   e. J'ai vérifié la documentation de toutes les règles identifiées dans les rapports
7. [Si pas a), question 6] De manière générale, avez-vous trouvé la documentation des règles utile ?
   - ☐ Oui
   - ☐ Non
   - ☐ Autre (veuillez préciser)
8. [Si pas a), question 6] De manière générale, avez-vous trouvé utiles les exemples donnés dans la documentation des règles ?
   - ☐ Oui
   - ☐ Non
   - ☐ Autre (veuillez préciser)
9. [Si pas a), question 6] Auriez-vous trouvé utile que la documentation inclue des images, et pas seulement les exemples d'équivalents textuels pertinents et non pertinents ?
   - ☐ Oui
   - ☐ Non
   - ☐ Autre (veuillez préciser)
10. [Si pas a), question 6] Auriez-vous trouvé utile que la documentation inclue des fichiers son, recréant la voix d'un lecteur d'écran, pour chacun des exemples présentés pour chaque règle ?
    - ☐ Oui
    - ☐ Non
    - ☐ Autre (veuillez préciser)
11. Qu'avez-vous aimé le plus de l'outil Acrolinx IQ Batch Checker ?

12. Qu'avez-vous aimé le moins de l'outil Acrolinx IQ Batch Checker ?

13. Y a-t-il d'autres informations ou remarques que vous aimeriez ajouter en relation avec Acrolinx IQ Batch Checker ?

aDesigner

14. Comment évalueriez-vous la précision de l'outil aDesigner en ce qui concerne la pertinence des équivalents textuels pour les images web ?

    (Précision ou "specificity" en anglais : nombre des différents problèmes qui peuvent être détectés et décris par l'outil à propos de la pertinence des équivalents textuels pour les images web)

    a. L'outil est très précis (76%100%)
    b. L'outil est assez précis (51%75%)
    c. L'outil est plus ou moins précis (26%50%)
    d. L'outil n'est pas très précis (1%25%)
    e. L'outil n'est pas du tout précis (0%)

15. Comment évalueriez-vous la justesse de l'outil aDesigner en ce qui concerne la pertinence des équivalents textuels pour les images web ?

    (Justesse ou "correctness" en anglais : proportion des problèmes détectés par l'outil qui sont réellement de vrais problèmes)

    a. L'outil est très juste (76%100%)
    b. L'outil est assez juste (51%75%)
    c. L'outil est plus ou moins juste (26%50%)
    d. L'outil n'est pas très juste (1%25%)
    e. L'outil n'est pas du tout juste (0%)

16. Comment évalueriez-vous la clarté des rapports générés par l'outil aDesigner en ce qui concerne la pertinence des équivalents textuels pour les images du Web ?

    a. Les rapports présentent l'information d'une manière très claire
    b. Les rapports présentent l'information d'une manière assez claire
    c. Les rapports présentent l'information d'une manière plus ou moins claire
    d. Les rapports ne présentent pas l'information d'une manière très claire
    e. Les rapports ne présentent pas du tout l'information de manière claire

17. Avez-vous vérifié la documentation apportée pour chaque problème identifié par aDesigner en relation avec la pertinence des équivalents textuels pour les images du Web ?

    a. Je n'ai vérifié la documentation d'aucun des problèmes identifiés en relation avec les images
    b. J'ai vérifié la documentation dans environ 25% des cas
    c. J'ai vérifié la documentation dans environ 50% des cas
    d. J'ai vérifié la documentation dans environ 75% des cas
    e. J'ai vérifié la documentation de tous les problèmes identifiés en relation avec les images

18. [Si pas a), question 17] De manière générale, avez-vous trouvé la documentation des problèmes relatifs aux images utile ?

    - ☐ Oui
    - ☐ Non
    - ☐ Autre (veuillez préciser)

19. [Si pas a), question 17] De manière générale, avez-vous trouvé utiles les exemples montrés dans la documentation ?

    - ☐ Oui
    - ☐ Non
    - ☐ Autre (veuillez préciser)

20. Qu'avez-vous aimé le plus de l'outil aDesigner?

21. Qu'avez-vous aimé le moins de l'outil aDesigner ?

22. Y a-t-il d'autres informations ou remarques que vous aimeriez ajouter en relation avec aDesigner ?

Assurance qualité (QA) dans le processus de localisation - Remarques générales

23. Travaillez-vous avec des outils d'assurance qualité (QA tools) ? (par exemple, Xbench, QA Distiller, fonctionnalité d'assurance qualité des outils de TAO...)

    - ☐ Oui
    - ☐ Non

- ◻ Autre (veuillez préciser)

24. [Si non, question 23] Y a-t-il une raison particulière pour laquelle vous n'utilisez pas d'outils d'assurance qualité ? (par exemple, « Il y a une personne qui s'en occupe dans mon agence de traduction », « Mon client n'est pas si exigeant » ...)

Assurance qualité (QA) dans le processus de localisation - Accessibilité

25. Pensez-vous que les outils de vérification d'accessibilité web pourraient compléter les outils d'assurance qualité utilisés habituellement dans le processus de localisation web ?
    - ◻ Oui
    - ◻ Non
    - ◻ S/O (sans opinion)
    - ◻ Autre (veuillez préciser)

26. [Si non, question 25] Pourquoi ?

27. D'après vous, quel outil, parmi ceux que vous avez utilisés lors de la dernière séance de l'étude, serait le plus utile au localisateur pour proposer des équivalents textuels pertinents pour les images web ?
    a. Acrolinx IQ Batch Checker
    b. aDesigner
    c. Les deux
    d. Aucun des deux

28. Pourriez-vous préciser les raisons de votre réponse ?

29. Considérez-vous que l'accessibilisation de chaque version linguistique d'un site web multilingue doive se faire séparément ?
    a. Non, je crois que l'accessibilité est universelle. Si le site web original était accessible, la version localisée sera elle aussi accessible.
    b. Oui, je crois que, lors du processus de localisation, il peut y avoir des changements par rapport au contenu initial du site et, par conséquent, l'accessibilisation de chaque version linguistique devrait se faire séparément.
    c. Autre (veuillez préciser)

30. Croyez-vous que le localisateur devrait avoir des connaissances sur l'accessibilité web pour effectuer son travail de façon plus professionnelle ?

    a. Non, je crois que rendre les sites accessibles n'est pas si important.
    b. Non, je crois que c'est aux experts en accessibilité web de rendre les sites accessibles, et que ce n'est en tout cas pas aux localisateurs de le faire.
    c. Oui, je pense que les localisateurs devraient avoir au moins des connaissances de base sur le sujet.
    d. Oui, je pense que les localisateurs devraient avoir des connaissances intermédiaires sur le sujet.
    e. Oui, je pense que les localisateurs devraient avoir des connaissances avancées sur le sujet.
    f. Autre (veuillez préciser)

31. D'après vous, quelle approche devrait suivre le localisateur en ce qui concerne l'accessibilité web ?

    a. Le localisateur ne doit pas faire attention à des aspects liés à l'accessibilité web
    b. Le localisateur devrait conserver le même niveau d'accessibilité obtenu dans la source, même si la source n'est pas accessible
    c. Le localisateur devrait conserver le même niveau d'accessibilité obtenu dans la source, même si la source n'est pas accessible, mais aussi informer le client (ou développeur web) à propos des problèmes d'accessibilité identifiés
    d. Le localisateur devrait conserver le même niveau d'accessibilité obtenu dans la source, mais seulement si la source est accessible
    e. Si la source n'est pas accessible, le localisateur devrait obtenir un niveau d'accessibilité acceptable dans la cible
    f. Si la source n'est pas accessible, le localisateur devrait obtenir un niveau d'accessibilité acceptable dans la cible, ainsi qu'informer le client (ou développeur web) à propos des problèmes d'accessibilité identifiés

32. Croyez-vous qu'il serait pertinent d'inclure un module sur l'accessibilité web dans les cours de formation en localisation web ?

    - ☐ Oui
    - ☐ Non
    - ☐ S/O (sans opinion)
    - ☐ Autre (veuillez préciser)

33. Y a-t-il d'autres informations ou remarques que vous aimeriez ajouter en relation avec cette deuxième séance de l'étude ?

Adresse email (veuillez utiliser la même adresse email que vous avez indiquée lors de l'inscription à l'étude)

Remerciement

Nous vous remercions d'avoir complété ce questionnaire !

Si vous n'avez pas encore suivi le webinaire offert avec cette étude, veuillez indiquer vos disponibilités sur ce sondage Doodle.

Silvia

Silvia.Rodriguez@unige.ch

D.5 User evaluation

Notes:

The content of the user evaluation questionnaire (Stage 2 of the experimental study, see Chapter 5) is presented in French, since it is the original language in which it was distributed. For economy of space, only one out of the 130 image evaluation exercises is presented here under the section “Évaluation des équivalents textuels pour les images : Exercice”. The complete evaluation data set can be found in Appendix J.

Introduction

Cher évaluateur,

Nous vous remercions d'avoir accepté de participer à cet exercice d'évaluation. Votre contribution est extrêmement précieuse pour notre projet de recherche et nous apprécions votre volonté de collaborer avec nous.

Pour toute information concernant le but de l'étude, la structure du questionnaire, ainsi que d'autres renseignements importantes sur cet exercice, veuillez consulter le document Word que vous avez reçu par courrier électronique.

Si vous êtes prêt à commencer, veuillez cliquer sur le bouton «Suiv.»

Questions démographiques

Adresse email

1. Nationalité

2. Pays de résidence

3. Âge

4. Quelle est votre langue maternelle ?

5. Parlez-vous d'autres langues couramment ?

- ◻ Oui
- ◻ Non

6. [Si oui, question 5] Laquelle/lesquelles ?

7. Profession

8. Années d'expérience professionnelle

9. Comment évalueriez-vous vos connaissances sur l'accessibilité web ?
   a. Je n’ai pas de connaissances sur le sujet.
   b. J'ai quelques connaissances, mais je n'ai pas eu l'occasion de travailler sur des projets liés à l'accessibilité web.
   c. J'ai travaillé sur quelques projets liés à l'accessibilité web.
   d. Je travaille principalement sur des projets liés à l'accessibilité web.
   e. Autre (veuillez préciser)

10. [Si c. ou d., question 9] Combien de temps avez-vous travaillé professionnellement dans le domaine de l'accessibilité web ?

11. Utilisez-vous des technologies d'assistance comme votre principal moyen d'accès à l’internet ?
    - ◻ Oui
    - ◻ Non

12. [Si oui, question 11] De quelle technologie s'agit-il ? (Exemple : J'utilise un lecteur d'écran pour naviguer sur le web : NVDA)

Accessibilité des sites multilingues et des images sur le web

13. Avec quelle fréquence naviguez-vous sur des sites multilingues (sites avec du contenu parallèle en deux langues ou plus) ?
    a. 100% des sites que je consulte sont multilingues
    b. Entre 76% et 99% des sites que je consulte sont multilingues
    c. Entre 51% et 75% des sites que je consulte sont multilingues
    d. Entre 26% et 50% des sites que je consulte sont multilingues
    e. Entre 1% et 25% des sites que je consulte sont multilingues
    f. Je consulte seulement des sites monolingues
    g. Autre (veuillez préciser)

14. Comment accédez-vous à des sites multilingues ?
    a. J'y accède toujours avec la voix que j'ai par défaut sur mon lecteur d'écran, indépendamment de la langue du site web
    b. Je change manuellement la voix de mon lecteur d'écran en fonction des langues disponibles sur le site web
    c. La voix de mon lecteur d'écran change automatiquement quand je passe d'un page web dans une langue à une autre page web dans une langue différente
    d. Autre (veuillez préciser)

15. D'après vous, l'accès à des sites multilingues entraîne-t-il davantage de défis que l'accès à des sites monolingues ? Veuillez illustrer votre réponse avec des exemples concrets, si possible.

16. Dans votre vie quotidienne, avec quelle fréquence trouvez-vous des équivalents textuels pour des images dans une langue différente de celle de la page web que vous êtes en train de consulter ?
    a. Toujours
    b. Souvent
    c. Parfois
    d. Jamais

17. D'après vous, sur un site internet multilingue, les équivalents textuels pour les images qui n'ont pas été traduits... (vous pouvez choisir plus d'une option)
    a. entravent totalement mon expérience de navigation sur le web

b. sont gênants, mais cela n'a pas de conséquences graves pour mon expérience de navigation sur le web
c. ne me dérangent pas du tout, mais seulement si je comprends la langue
d. ne me dérangent pas du tout, même si je ne comprends pas la langue
e. Autre (veuillez préciser)

18. En général, dans votre vie quotidienne, avec quelle fréquence trouvez-vous des images inaccessibles sur le web?

a. Toujours
b. Souvent
c. Parfois
d. Jamais

19. Quel est votre point de vue par rapport au cas particulier des images décoratives (dont la raison d'être est purement esthétique) ?

a. J'aime bien qu'elles soient décrites, même si cela peut ralentir ma navigation sur le web
b. J'aime bien qu'elles soient décrites, mais je n'aime pas la perte de temps que cela implique
c. Je crois qu'elles ne devraient pas du tout être décrites
d. Autre (veuillez préciser)

20. Comment souhaiteriez-vous qu'une image avec un lien intégré soit décrite ?

Évaluation des équivalents textuels pour les images : Introduction

La tâche d'évaluation elle-même commence maintenant. À partir de ce moment, chaque page de SurveyMonkey sera consacrée à une seule image.

Veuillez vous rappeler que vous y trouverez d'abord la description du contexte de l'image: sur quelle page et dans quelle partie de la page l'image apparaît, et quel est son contenu et/ou fonction. Une liste d'équivalents textuels potentiels vous sera ensuite proposée pour cette image. Après la description contextuelle de chaque image, nous annoncerons le nombre total d'équivalents textuels proposés.

Vous devrez attribuer une note à chaque équivalent textuel de la liste selon l'acceptabilité et la pertinence de l'équivalent textuel par rapport au contexte décrit.

Au bas de chaque page, un espace « Commentaire » optionnel vous permettra d'ajouter vos éventuelles remarques. Veuillez ouvrir le document Word que vous avez reçu par courrier électronique pour consulter, entre autres, les informations concernant les aspects suivants :

a) Structure des pages web dont sont tirées les images
b) Valeurs des notes sur l'échelle d'évaluation
c) Précisions par rapport au codage des équivalents textuels

Bonne continuation et merci d'avance de votre patience et bonne volonté !

Évaluation des équivalents textuels pour les images : Exercice

Image 1

Emplacement :

Entête Page du partenaire Mail & Guardian Site web de la campagne

Description contextuelle : Image située en haut de la page, avant la section des réseaux sociaux et le menu de navigation, et qui est presque aussi large que la page web. Il s'agit d'un collage comprenant trois images, montrant différents scénarios. La première image (à gauche) est floue. Deux hommes semblent parler au téléphone. Dans la deuxième image (au centre du collage), le portrait d'un homme portant des lunettes est affiché. Seuls les sourcils, les lunettes et le nez sont visibles. La troisième image (à droite) montre une tour de contrôle du trafic aérien.

1. Veuillez attribuer une note à chaque équivalent textuel. Il y en a 10 à évaluer.

	1. Pas acceptable	2. Acceptable	3. Pertinent	4. Très pertinent
bannière	☐	☐	☐	☐
alt vide	☐	☐	☐	☐
banner	☐	☐	☐	☐
bandeau d'en-tête (mise en page)	☐	☐	☐	☐
no alt	☐	☐	☐	☐
bannière du site	☐	☐	☐	☐
Image d'en-tête	☐	☐	☐	☐
alt "" no title attribute	☐	☐	☐	☐
banière	☐	☐	☐	☐
image du bandeau de la page Web de l'UVJ	☐	☐	☐	☐

Commentaire (optionnel)

[*The evaluation exercise about Image 1 has been included here for illustrative purposes. The structure of the exercise was exactly the same for Images 2-130*].

Remerciement

Vous avez terminé l'exercice d'évaluation.

Nous vous remercions énormément de votre collaboration !

Pour plus d'informations sur la rétribution, veuillez prendre contact avec Mme Silvia Rodríguez Vázquez.

Courriel : Silvia.Rodriguez@unige.ch

Merci !

Appendix E. Preparation kit prior to Stage 1

E.1 Introduction to the study

Note:

This document corresponds to the version sent to participants of the treatment group (with WA knowledge). In the version received by the control group, participants were invited to take part in the WA webinar after the study.

INTRODUCTION À L'ÉTUDE

Merci d'avoir accepté de participer à notre étude sur la traduction de contenus web et d'avoir rempli le questionnaire en ligne. Dans les sections suivantes, vous trouverez des informations pertinentes sur l'étude. Veuillez les lire attentivement.

ORGANISATION ET DURÉE DE L'ÉTUDE
L'étude sera divisée en deux séances.

Première séance : Avant la première séance, vous recevrez un webinaire gratuit d'environ 45 minutes sur les bonnes pratiques en langage HTML pour les traducteurs et les localisateurs. Nous vous enverrons toutes les informations nécessaires pour participer à ce webinaire un jour avant la date convenue entre le participant et le chercheur.

Après une pause de 30 minutes, vous serez invité à effectuer un exercice de localisation web (traduction de fichiers HTML), puis à remplir un bref questionnaire lié à la tâche. Nous vous ferons parvenir les instructions détaillées par e-mail 15 minutes avant le début de l'exercice. La séance ne devrait pas durer plus d'une heure et demie (1 heure et 30 minutes). Le chercheur sera disponible pendant toute la séance via chat, sur la plateforme Adobe Web Meeting. Vous recevrez les informations nécessaires pour accéder au chat avec le reste des instructions.

Pour faire l'exercice, les participants seront libres d'utiliser, s'ils le souhaitent, l'outil (ou les outils) de traduction assistée par ordinateur (TAO) de leur choix. Il n'y a toutefois aucune obligation dans ce sens; c'est-à-dire qu'il n'y a pas d'instructions précises sur la procédure à suivre pour effectuer l'exercice de localisation web.

Par contre, afin de garantir la validité de la recherche, nous aurions besoin d'enregistrer cette séance. Nous vous prions par conséquent de bien vouloir installer et utiliser un logiciel qui permet d'enregistrer votre écran.

BBFlashBack Express 5 est un logiciel gratuit que vous pouvez installer pour cette première séance; vous pourrez le désinstaller quand vous n'aurez plus besoin. Veuillez noter que ce logiciel n'enregistrera que l'écran de votre ordinateur. Il n'enregistrera pas les sons ou votre espace de travail. Vous trouverez des instructions détaillées sur la manière de l'installer et de l'utiliser dans le document **BBFlashBackExpress_ GuideUtilisation.pdf**. Nous vous invitons à lire ce document, à installer le logiciel et à le tester dès à présent afin de pouvoir vous aider si vous avez des questions.

Après l'exercice de localisation web, nous vous demanderons d'enregistrer les fichiers résultant de la première séance (traductions et fichier vidéo) et de nous les envoyer, selon les instructions que vous recevrez au début de la séance.

Deuxième séance : Au cours de la deuxième séance, nous vous demanderons d'utiliser deux outils d'assurance qualité (AQ) différents et de remplir un deuxième questionnaire lié à cette dernière tâche. Afin de pouvoir avoir accès à ces outils, vous recevrez des informations d'identification uniques pour vous connecter à distance aux ordinateurs disponibles dans notre salle informatique, à l'aide d'un logiciel de contrôle à distance des ordinateurs via Internet. Nous vous ferons également parvenir des instructions précises 15 minutes avant le début de la deuxième séance. La séance ne devrait pas durer plus d'une heure et demie (1 heure et 30 minutes).

Pour cette deuxième séance, nous vous prions de bien vouloir installer et utiliser un logiciel de contrôle à distance des ordinateurs via Internet : *Real VNC Viewer.* Vous recevrez des instructions détaillées sur la manière de l'installer et l'utiliser au cours de cette semaine, avant vendredi 5 décembre 2014.

Après le deuxième exercice, nous vous demanderons d'enregistrer les fichiers résultant de la deuxième séance et de nous les envoyer, selon les instructions que vous recevrez au début de la séance.

ALTERNATIVE POUR LA PREMIÈRE SÉANCE

Cette étude cherche idéalement à recréer les conditions de travail habituelles des participants. C'est pour cette raison que nous vous serions reconnaissante de suivre la procédure décrite ci-dessus pour la première séance. Cependant, si vous le souhaitez, vous pouvez aussi faire le premier exercice en vous connectant à distance à un ordinateur disponible dans notre salle informatique. Si vous choisissez cette solution, vous n'aurez pas besoin d'installer le logiciel d'enregistrement de l'écran sur votre ordinateur. Si vous préférez cette option, veuillez informer Mme. Rodríguez par courriel.

RISQUES

Il n'y a pas de risques connus associés à l'étude. La participation est complètement volontaire. Si le participant se sent mal à l'aise, il peut se retirer de l'étude à tout moment.

Après avoir lu ce document, veuillez vous référer au ***Formulaire de consentement éclairé*** que vous avez reçu aussi par courriel.

Si vous avez des questions, n'hésitez pas à prendre contact avec Silvia Rodríguez Vázquez : silvia.rodriguez@unige.ch.

E.2 BB FlashBack Express 5 - Customised user guide for participants

---

BB FlashBack Express 5 - Installation et utilisation

BB FlashBack Express 5 est un logiciel gratuit qui permet d'enregistrer l'écran de l'ordinateur, et dont l'interface est facile à utiliser. Veuillez lire attentivement les instructions ci-dessous.

TÉLÉCHARGEMENT ET INSTALLATION DE BB FLASHBACK EXPRESS 5

1. Pour télécharger le logiciel, veuillez suivre le lien suivant (la dernière version du logiciel est seulement disponible en anglais, donc veuillez vous assurer que vous êtes sur la version anglaise du site) :
   http://www.bbsoftware.co.uk/bbflashbackexpress/download.aspx
2. Même si le logiciel est gratuit, il est nécessaire d'avoir une licence. Pour la recevoir, veuillez introduire votre adresse e-mail, puis cliquez sur ***Download*** (figure 1).

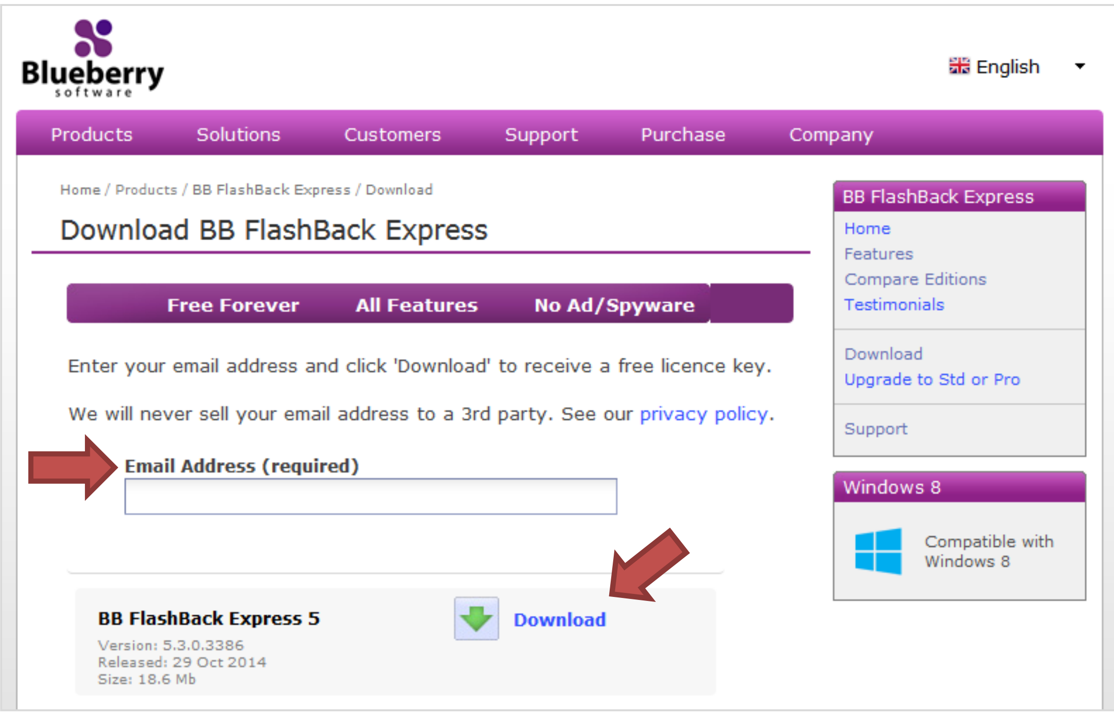


Figure 1. Page de saisie de l'adresse e-mail

3. Veuillez ensuite cliquer sur le bouton ***Download here*** pour obtenir le fichier exécutable *bbfbex5.exe* (figure 2).

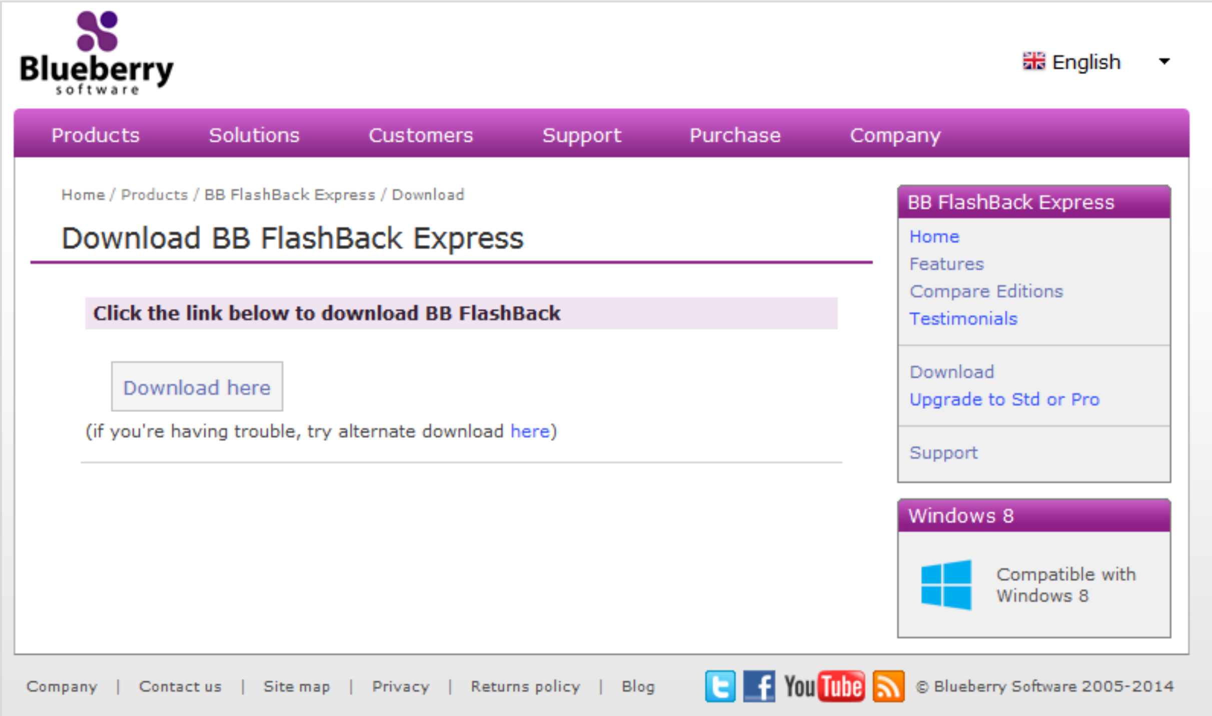


Figure 2. Bouton de téléchargement du logiciel

4. Lancez le fichier exécutable *bbfbex5.exe* et suivez les instructions de l'assistant d'installation jusqu'à la fin (Figure 3). Avant de fermer l'assistant d'installation, vous aurez l'option de démarrer le logiciel directement après l'installation.

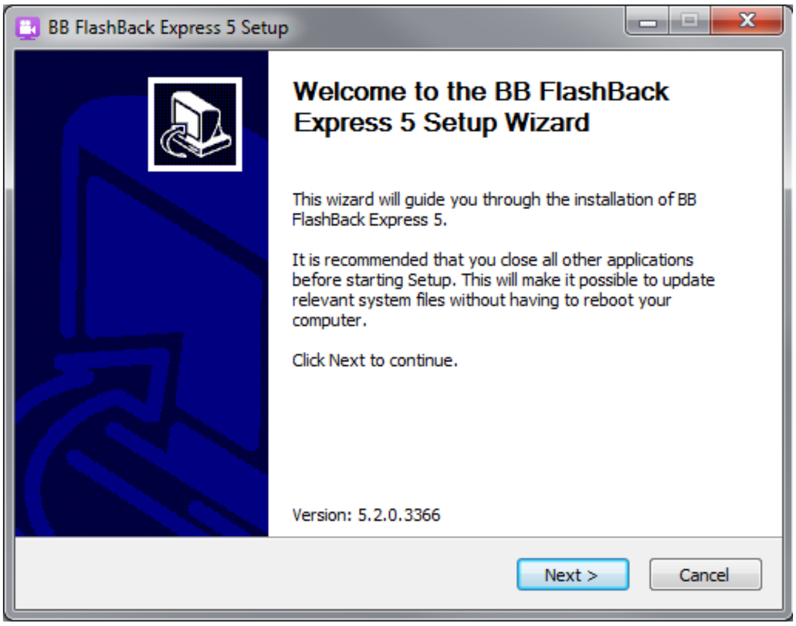


Figure 3. Assistant d'installation

5. À chaque fois que vous démarrerez le logiciel, vous verrez une annonce qui vous demandera si vous voulez la version payante de BBFlashBack. Veuillez ignorer cette annonce et cliquer sur ***Continue*** (figure 4).

Figure 4. Annonce de la version payante de BB FlashBack

6. Vous devrez ensuite introduire le numéro de licence que vous avez reçu par courriel, puis cliquez sur ***Continue*** pour pouvoir commencer à utiliser le logiciel (figure 5).

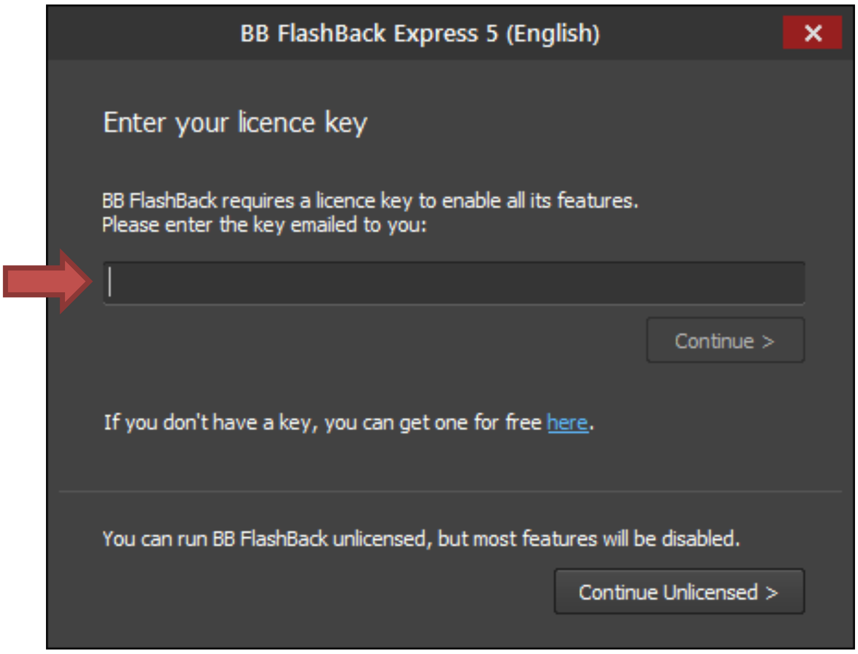


Figure 5. Fenêtre de saisie du numéro de licence

DÉFINITION DES PARAMÈTRES DE BB FLASHBACK EXPRESS 5

1. Après avoir introduit le numéro de licence, vous arriverez sur la fenêtre d'accueil du logiciel (figure 6). Veuillez cliquer sur ***Record Your Screen*** pour définir les options d'enregistrement.

Figure 6. Fenêtre d'accueil de BB FlashBack Express 5

2. Lors de l'utilisation du logiciel pendant l'étude, il n'est pas nécessaire de cocher les options ***Record Sound*** ou ***Record Webcam***. Veuillez vous assurez de <u>ne pas cocher</u> les cases correspondantes (figure 7).

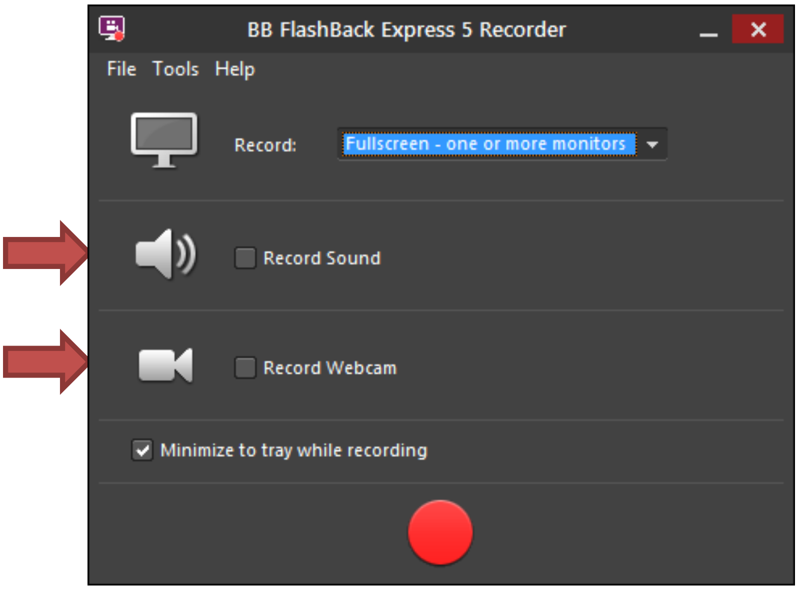


Figure 7. Options d'enregistrement

3. Par contre, nous vous prions de vous assurer que vous avez bien coché les options suivantes :

Enregistrement des frappes clavier :

- Pour accéder aux options avancées, veuillez aller sur le menu ***Tools > Options***.

- Sous la section ***Keystroke Recording***, veuillez vous assurer que les deux cases sont cochées (figure 8).

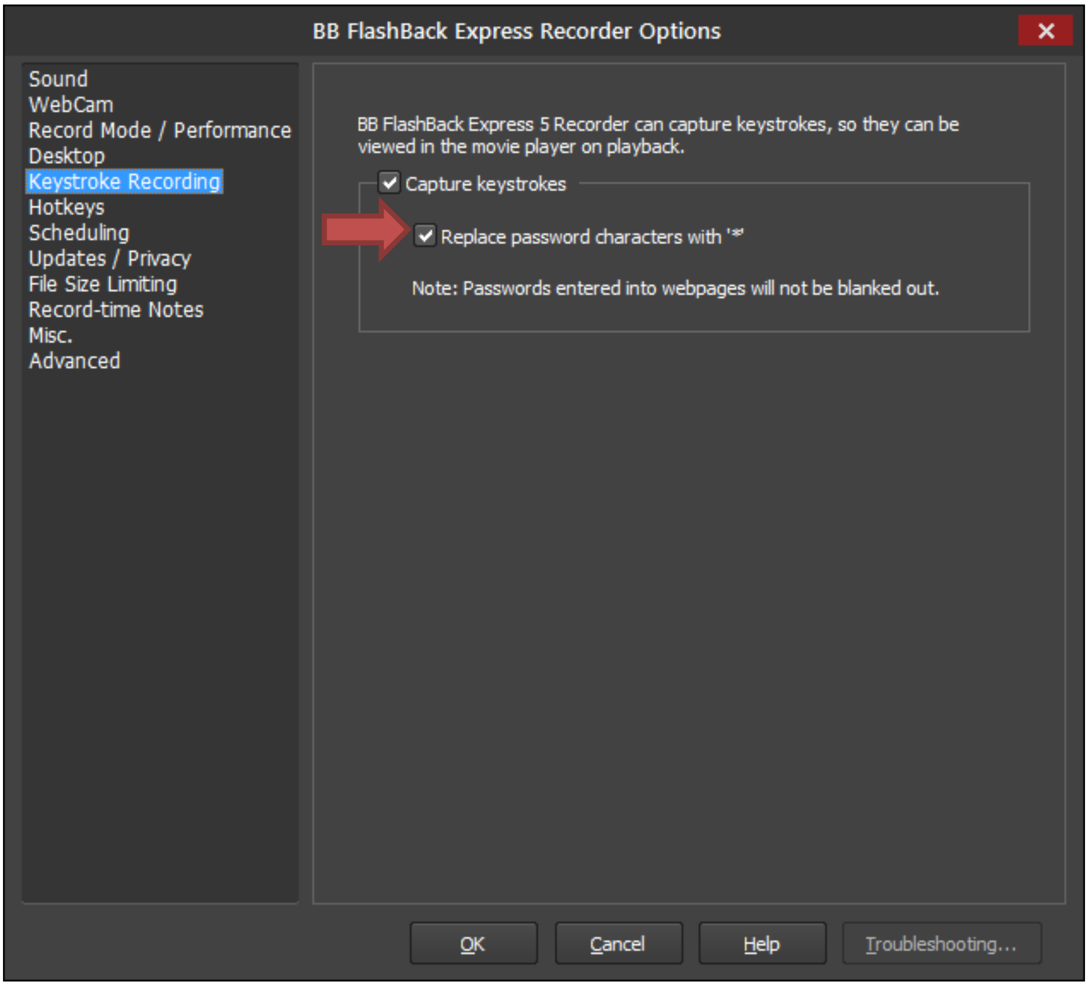


Figure 8. Options d'enregistrement des frappes clavier

Autres options importantes :

- Sous la section ***Misc.*** (figure 9), veuillez vous assurer que les cases suivantes sont cochées :

a) Keep PC awake and disable screensavers when recording
b) Show 3-2-1 countdown when recording starts

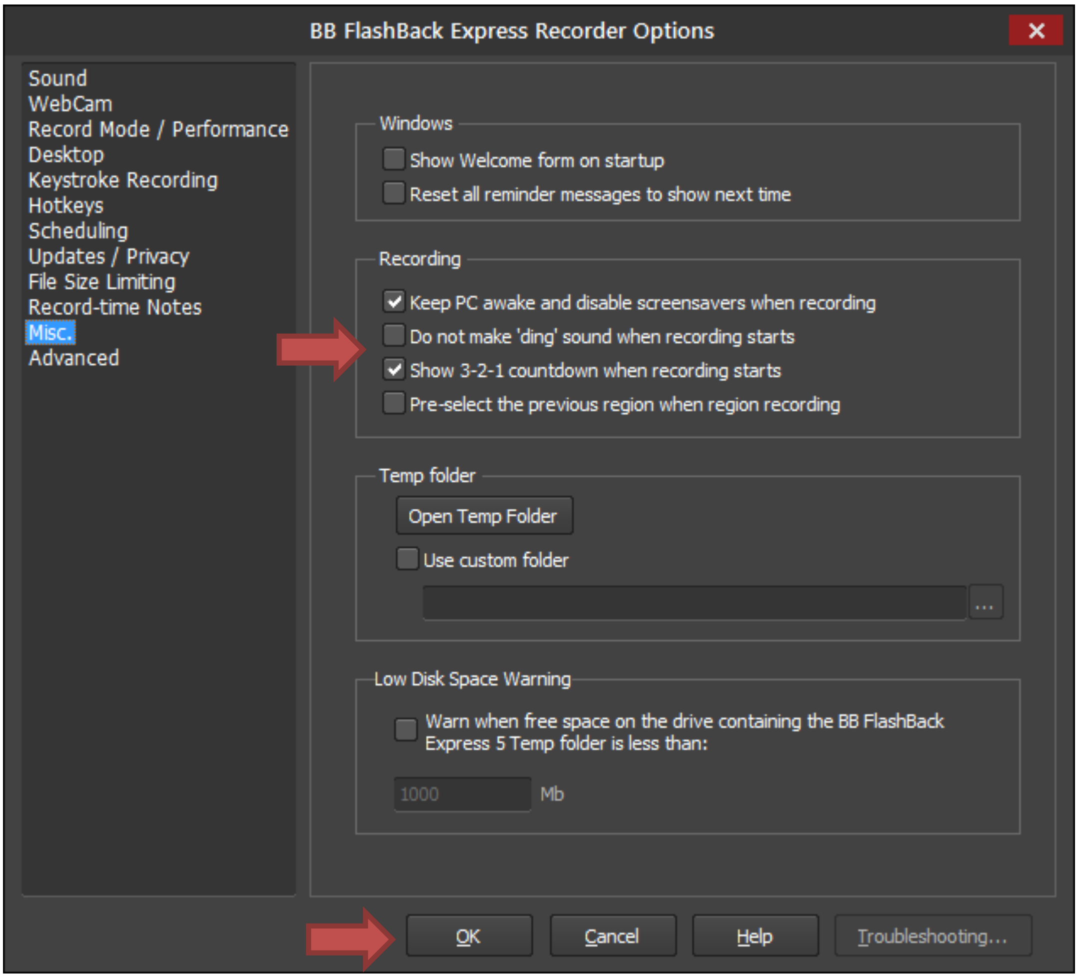


Figure 9. Options d'enregistrement avancées

- Enfin, cliquez sur ***OK*** pour valider les changements effectués.

UTILISATION DE BB FLASHBACK EXPRESS 5

1. Après avoir appuyé sur ***OK***, vous vous trouverez à nouveau sur fenêtre montrée à la figure 7. Pour commencer à enregistrer l'écran, veuillez appuyer sur le bouton rond rouge ***Start recording***.

2. Pour arrêter l'enregistrement, veuillez appuyer sur le bouton carré rouge ***Stop recording*** (figure 10).

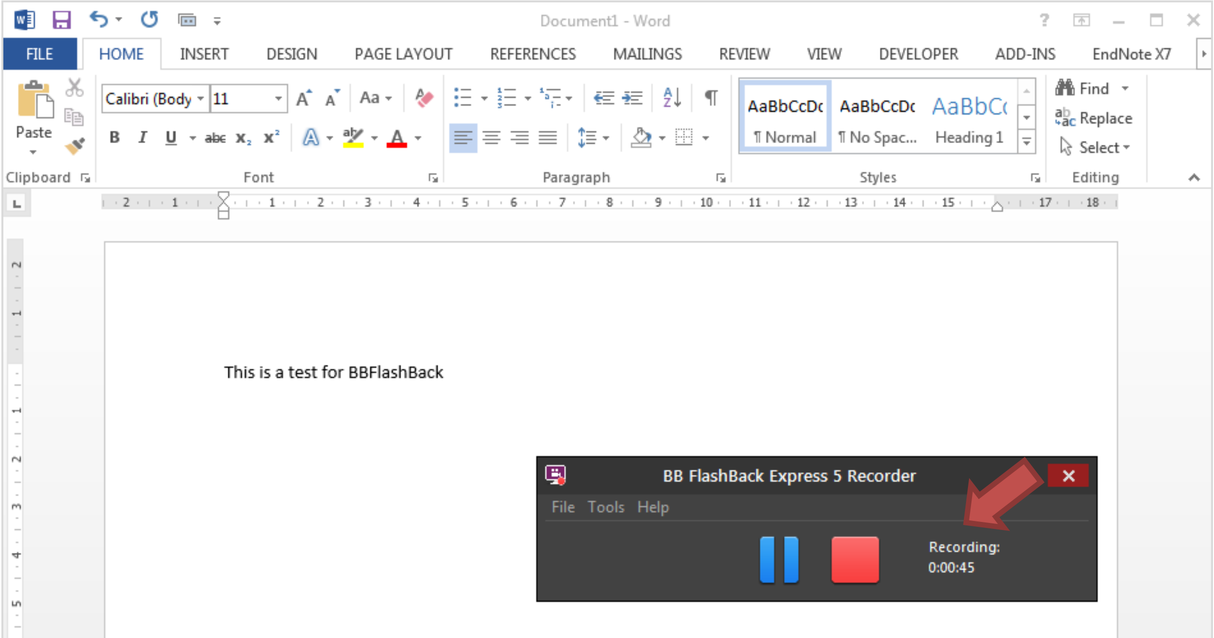


Figure 10. Bouton carré rouge pour arrêter l'enregistrement

3. Quand vous arrêterez l'enregistrement, BB FlashBack vous donnera l'option d'enregistrer la vidéo. Veuillez cliquer sur ***Save*** (figure 11).

Figure 11. Enregistrer la vidéo

4. Après chaque séance, nous vous demanderons de nous envoyer le fichier vidéo (en format .fbr), avec vos traductions.

N'hésitez pas à tester le logiciel avant la première séance de l'étude et à nous poser des questions, si nécessaire.

Silvia Rodríguez Vázquez
silvia.rodriguez@unige.ch

E.3 TeamViewer 10 - Customised user guide for participants

TeamViewer - Installation et utilisation

TeamViewer 10 est un logiciel freeware de contrôle de PC à distance dont l'interface est facile à utiliser. Vous en aurez besoin seulement pour la deuxième séance de l'étude. Les participants qui ont choisi de faire la première séance sur l'un de nos ordinateurs à la salle informatique de la Faculté de traduction et interprétation de l'UNIGE, se serviront aussi de TeamViewer 10 pour y avoir accès. Nous vous prions de lire attentivement les instructions ci-dessous.

NOTE IMPORTANTE : Si vous allez effectuer la tâche sur un portable ou un ordinateur avec un écran de 17 pouces ou inférieur, il est possible que la taille des éléments du bureau à distance soit légèrement inférieure aux dimensions habituelles. Essayez, tout d'abord, d'aller sur ***View > Scale > Original***, tout en haut de la fenêtre qui montre le bureau à distance, pour améliorer la qualité de visualisation. Si, lors du test de connexion qui aura lieu 10 minutes avant le début de la tâche, vous ne vous sentez pas à l'aise avec l'environnement de travail, nous vous proposerons un autre outil, également facile et rapide à installer.

TÉLÉCHARGEMENT ET INSTALLATION de TeamViewer

1. TeamViewer 10 est disponible pour Windows, Mac OS X et Linux. Veuillez suivre le lien suivant pour le télécharger : http://www.teamviewer.com/fr/index.aspx. Sur la page principale, cliquez sur ***Version complète gratuite*** (flèche rouge, figure 1).

Figure 1. Téléchargement de TeamViewer 10

2. Lancez le fichier exécutable *TeamViewer_Setup_fr.exe* en double-cliquant. Dans la première fenêtre, veuillez choisir les options ***Installer*** et ***personnellement / non commercialement***, ainsi que cocher la casse ***Afficher les paramètres avancés*** (figure 2).

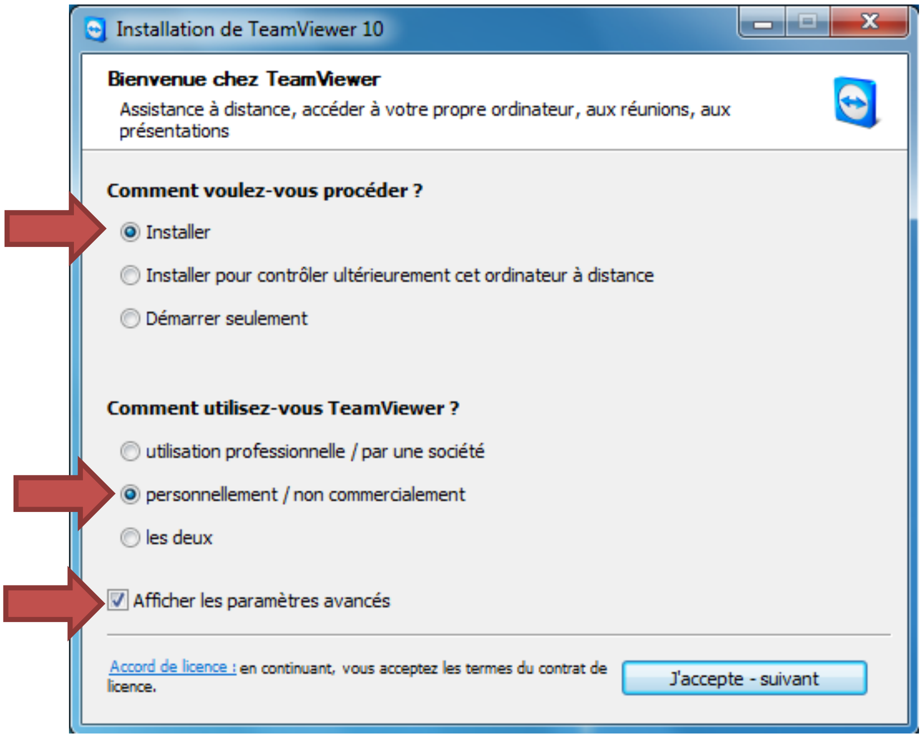


Figure 2. Assistant d'installation de TeamViewer 10

3. Dans la fenêtre des paramètres avancés, veuillez vous assurer que vous ne cochez aucune des trois options disponibles (figure 3). Ensuite, cliquez sur ***Fermer***.

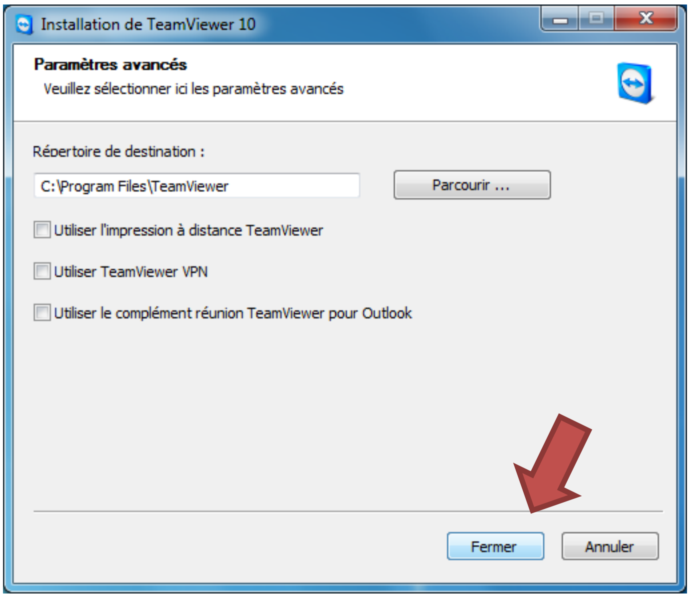


Figure 3. Paramètres avancés de TeamViewer 10

UTILISATION DE TeamViewer 10

1. La figure 4 montre l'interface principale du logiciel. Avec les instructions de la séance, vous recevrez le numéro correspondant à l'**ID du partenaire.** Veuillez rentrer ce numéro (avec le format 000 000 000, avec des espaces), sélectionner l'option ***Contrôle à distance*** et puis cliquer sur le bouton ***Connexion à un partenaire***.

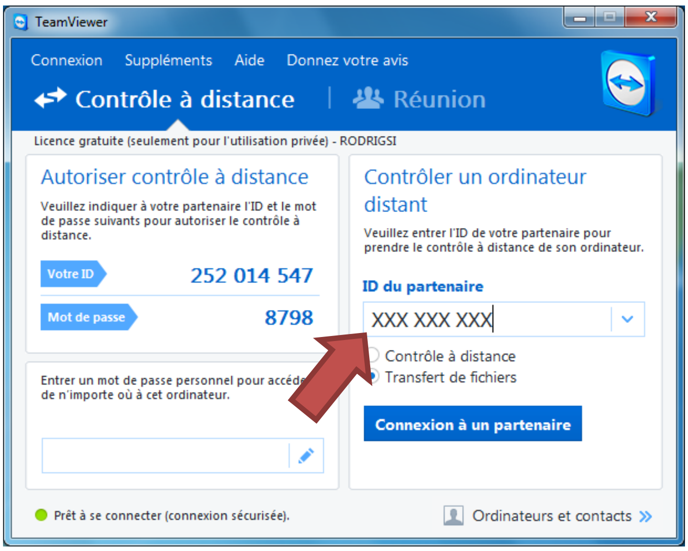


Figure 4. Interface principale de TeamViewer 10

2. Ensuite, un mot de passe vous sera demandé. Veuillez saisir le mot de passe **etude2014** et appuyer sur ***Connexion*** (figure 5).

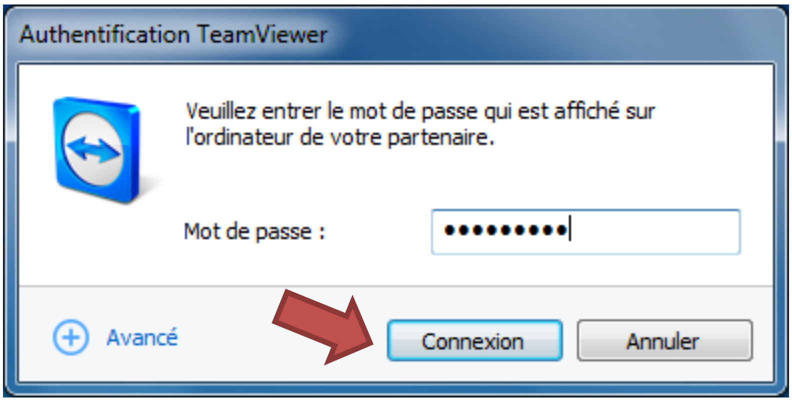


Figure 5. Fenêtre de saisie du mot de passe pour se connecter à l'ordinateur distant

À partir de ce moment-là, vous prendrez le contrôle de l'ordinateur distant et, par conséquent, vous pourrez utiliser tous les logiciels et fonctionnalités disponibles sur cette machine.

ASTUCES POUR L'UTILISATION DE TeamViewer 10

Si, pendant la séance, vous avez besoin de transférer des fichiers ou des dossiers de votre ordinateur à l'ordinateur distant et vice-versa, vous pouvez simplement les sélectionner et les glisser d'un bureau à l'autre (figure 6). **IMPORTANTE:** Veuillez noter qu'il est seulement possible de copier/coller ou glisser des fichiers sur le ***Bureau*** ou le dossier ***Téléchargements***.

Quand le transfert sera effectué, vous verrez une petite fenêtre de confirmation. Veuillez la fermer en cliquant sur ***Fermer*** et ensuite continuer avec l'exercice.

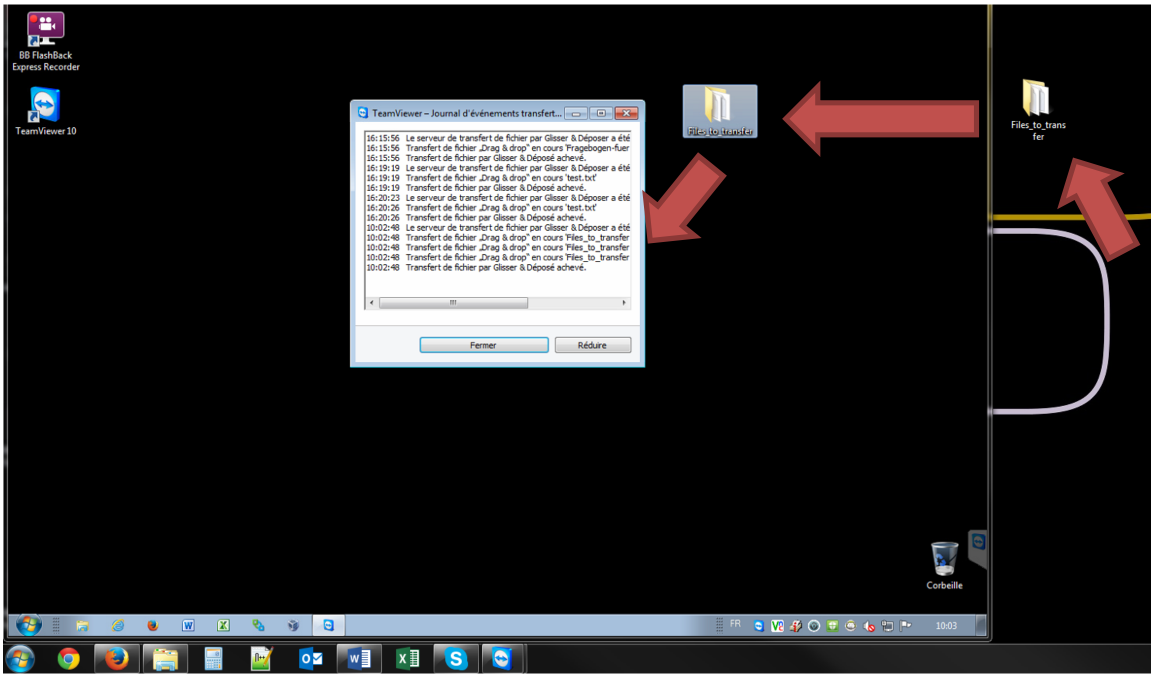


Figure 6. Sélection et glisse fichiers entre les deux bureaux actifs

Pour arrêter la connexion quand vous auriez fini, fermez tout simplement la fenêtre. Si vous avez besoin de vous connecter de nouveau à l'ordinateur distant, vous pouvez retourner à l'interface principale de TeamViewer 10 : le dernier numéro **ID du partenaire** que vous avez introduit reste toujours visible, même après la déconnexion.

Appendix F. Stage 1: Informed consent form

FORMULAIRE DE CONSENTEMENT ÉCLAIRÉ

Merci à nouveau de votre intérêt pour notre étude. Nous vous prions de bien vouloir lire les déclarations ci-dessous, de cocher les cases correspondantes et de nous renvoyer le présent formulaire de consentement éclairé dûment daté et signé. Il n'est pas nécessaire d'imprimer le document, d'ajouter votre signature à la main puis de le scanner. Vous pouvez simplement écrire la date et votre nom complet directement sur ce PDF, au format électronique.

INVITATION À PARTICIPER À UNE ÉTUDE

- [ ] Je confirme que j'ai accepté de participer volontairement à une étude sur la traduction de contenus web.

- [ ] Je confirme que j'ai reçu et lu le document Introduction à l'étude, qui contient des informations pertinentes sur les deux séances de l'étude, telles que la durée estimée et les logiciels nécessaires pour effectuer les exercices.

DROIT DE RETRAIT

- [ ] Je comprends que je peux me retirer de l'étude à tout moment et que, si je décide de le faire, je ne recevrai pas la rétribution indiquée sur l'appel à participation comme compensation pour le travail accompli.

ACCÈS À DISTANCE À L'ENVIRONNMENT INFORMATIQUE

- [ ] J'accepte d'installer le logiciel mentionné sur le document Introduction à l'étude, qui me permettra d'accéder à distance via Internet à l'environnement informatique de l'étude, depuis mon ordinateur personnel, selon les instructions du chercheur (applicable pour la deuxième séance).

- [ ] Je comprends que je vais recevoir des informations d'identification uniques temporaires pour accéder à distance à l'environnement informatique de l'étude et que cet accès sera coupé à la fin de l'étude (applicable pour la deuxième séance).

ENREGISTREMENT DE L'ÉCRAN

- ☐ J'accepte d'utiliser un logiciel permettant l'enregistrement de l'écran pendant chacune des deux séances, selon les instructions du chercheur (applicable pour les deux séances).

- ☐ Je comprends que ce logiciel d'enregistrement de l'écran n'enregistrera que l'écran de l'ordinateur que j'utiliserai pour faire les exercices. Il n'enregistrera pas les sons ou mon espace de travail.

- ☐ Je comprends que je dois m'assurer que toutes les options d'enregistrement de l'écran indiquées dans le document BBFlashBackExpress_GuideUtilisation.pdf sont activées au moment d'enregistrer chaque séance.

- ☐ Je comprends aussi que, si ces options ne sont pas activées, la vidéo résultante ne sera pas valide et, par conséquent, je n'aurai pas droit à la rétribution indiquée sur l'appel à participation.

- ☐ Je permets au chercheur principal de l'étude, Mme Silvia Rodríguez Vázquez, d'observer et d'analyser les fichiers résultant de ma participation à l'étude.

ANONYMAT

- ☐ Je comprends que mon nom et mon adresse e-mail ne seront utilisés qu'à des fins de communication, avant et pendant l'étude, et que, en aucun cas, ces données seront publiées ou révélées.

- ☐ Je comprends que les données recueillies par le chercheur au cours des différentes séances de l'étude peuvent être publiées dans des communications académiques, mais qu'aucune donnée personnelle permettant de m'identifier ne sera pas divulguée.

Numéro d'identification du participant : ______

Date: ______

Signature (nom complet): ______

Appendix G. Stage 1: Task briefing sheets

G.1 Localisation task brief

Note:
The localisation task brief was the same for participants from the control group and the treatment group. The only difference was the link to the post-task questionnaire, since two different ones had to be created to include questions 11-15 (se Appendix D.3) in the case of the treatment group.

ÉTUDE SUR LA LOCALISATION WEB

Première séance (environ 90 minutes)

1. INSTRUCTIONS GÉNÉRALES

Dans cette première partie de l'étude, vous devrez **localiser de l'anglais vers le français** le site internet d'une campagne de lutte contre la pauvreté fictif, dont les pages ont été développées spécialement pour cette étude. Ce site web est composé de **trois fichiers HTML** (*index.html, index2.html* et *index3.html*) ainsi que de cinq dossiers associés (figure 1).

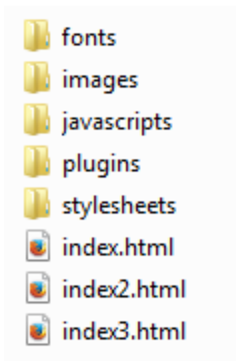


Figure 1. Fichiers du site internet à localiser

Le **temps prévu** pour effectuer cette tâche est d'une heure et demie. Toutefois, vous pouvez prendre tout le temps dont vous avez besoin. Comme nous vous l'avions déjà annoncé dans le document d'introduction à l'étude, il n'y a pas de restrictions par rapport à la méthode de traduction des contenus de ces trois fichiers : c'est-à-dire, vous pouvez utiliser les outils d'aide à la traduction et à la localisation de votre choix, ainsi que des ressources linguistiques monolingues ou multilingues, physiques ou électroniques. Vous pouvez également faire des recherches sur internet, si vous en avez besoin.

Par ailleurs, veuillez noter que :

- **Il ne faut pas** ajouter un sélecteur de langues ou créer un site bilingue qui permettrait aux visiteurs de changer d'une version à l'autre en utilisant des liens hypertexte : veuillez nous retourner uniquement **les fichiers dans la langue cible.**

- **Il ne faut pas** faire des changements structurels ou esthétiques pour lesquels il serait nécessaire de modifier les feuilles de style (couleurs, police, taille du texte, etc.), même si vous trouvez que ces changements seraient idéalement nécessaires pour une localisation complète du site pour la culture cible. Veuillez travailler seulement sur les fichiers HTML.

- **Il ne faut pas** renommer les fichiers HTML. Il n'est donc pas nécessaire d'adapter les liens hypertexte entre les différentes pages du site.

2. PRÉCISIONS SUR LES CONTENUS DU SITE

Étant donné qu'il s'agit d'un site web fictif, nous avons enlevé les pages enfant (ou sous-pages) de ces trois fichiers principaux (*index.html, index2.html* et *index3.html*). Veuillez donc prendre en compte les informations suivantes :

- ➔ Nous sommes conscients que les différents éléments des sections "Other topics" (*index.html*), "Our University" (*index2.html*) et "Volunteering Areas" (*index3.html*) contiennent des liens qui ne marchent pas. Veuillez ignorer ce problème et, le cas échéant, faire comme s'ils contenaient des liens vers des pages avec des informations à propos de chaque sujet.
- ➔ Nous sommes conscients que les liens vers les différents réseaux sociaux et média ne marchent pas. Veuillez ignorer ce problème et, le cas échéant, faire comme si chaque institution avait une participation active sur ces réseaux.

3. TÉLÉCHARGEMENT DES FICHIERS NÉCESSAIRES POUR EFFECTUER LA TÂCHE

Les fichiers source se trouvent dans un fichier .zip (*SiteWeb.zip*) que vous pourrez télécharger en suivant le lien
http://ftp.eti.unige.ch/repository/list/FilesEtude/SiteWeb.zip.

Pour pouvoir y accéder, vous devrez donner nom d'utilisateur et un mot de passe. Veuillez utiliser les données suivantes :

Nom utilisateur : etude2014
Mot de passe : 2014etude

4. RAPPEL SUR BB FLASHBACK EXPRESS 5

IMPORTANT : Avant de commencer l'exercice, n'oubliez pas de lancer l'outil qui vous permettra d'enregistrer votre écran : BB FlashBack Express 5. Vérifiez également que les options indiquées sur le document que vous avez reçu il y a quelques jours soient activées. Consultez-le une dernière fois avant l'exercice, si nécessaire (*BBFlashBackExpress_GuideUtilisation_DerniereVersion.pdf).*

➔ Pour commencer à enregistrer la séance, cliquez sur le bouton rond rouge (figure 2).

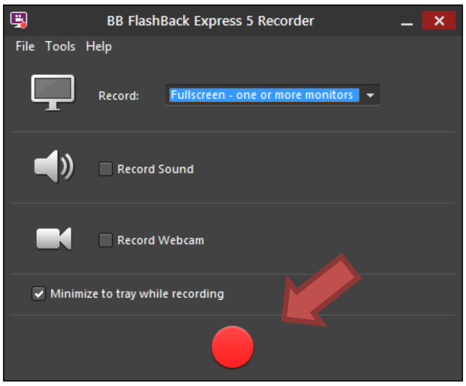


Figure 2. Fenêtre principale de BB FlashBack Express 5

5. FICHIERS À RENDRE

Quand vous aurez terminé l'exercice, veuillez arrêter l'enregistrement en faisant un clic droit sur la petite icône de BBFlashBack, puis faites un clic gauche sur Stop (figure 3).

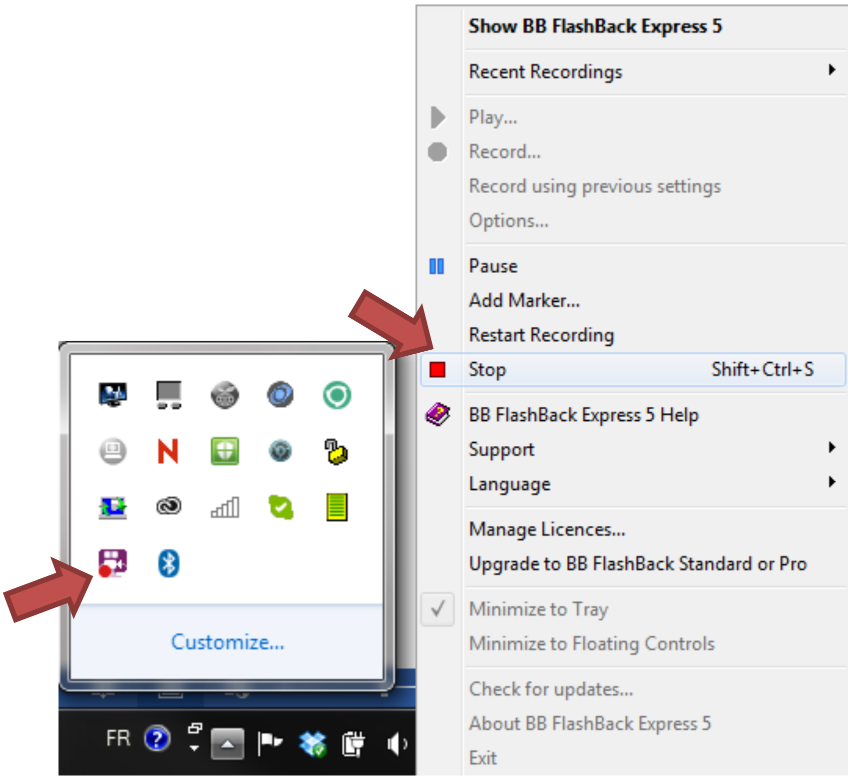


Figure 3. Arrêter l'enregistrement de la séance

Veuillez enregistrer la vidéo résultante en la nommant **seance1.fbr**.

Veuillez ensuite comprimer le fichier vidéo et le dossier **SiteWeb** (avec les trois fichiers HTML en langue cible) en créant un seul fichier final **seance1_participant00.zip** (merci de remplacer les chiffres 00 par le numéro d'identification qui vous a été attribué pour cette étude).

Enfin, veuillez nous envoyer le fichier .zip à l'adresse e-mail **silvia.rodriguez@unige.ch** à l'aide de WeTransfer : https://www.wetransfer.com/

IMPORTANT : Si la deuxième séance se déroulera le même jour, nous vous prions de nous envoyer aussi par e-mail (à l'adresse silvia.rodriguez@unige.ch) **le site web localisé dans un fichier .zip, sans le fichier vidéo**, afin que nous puissions déjà mettre vos fichiers sur l'ordinateur à distance pour la deuxième séance.

6. QUESTIONNAIRE LIÉ À LA TÂCHE EFFECTUÉE

Après l'exercice de localisation, nous vous prions de bien vouloir remplir un bref questionnaire lié à la tâche effectuée. Le remplir ne devrait pas vous prendre plus de 10 minutes.

https://fr.surveymonkey.com/s/premiere_seance_etude2014

MERCI BEAUCOUP DE VOTRE PARTICIPATION À CETTE PREMIÈRE SÉANCE !

Si vous avez un problème technique avec BBFlashBack Express 5, le FTP ou WeTransfer (pour télécharger les fichiers), je serai disponible sur **Skype (silvia_rodriguez_vazquez)** et Adobe Web Meeting, ainsi que joignable par **courriel (silvia.rodriguez@unige.ch).**

G.2 Quality assurance task brief - Scenario A (aDesigner-Acrolinx)

ÉTUDE SUR LA LOCALISATION WEB

Deuxième séance (environ 90 minutes)

Dans cette deuxième partie de l'étude, vous allez utiliser deux logiciels d'assurance qualité pour l'accessibilité web du site que vous avez localisé.

Le World Wide Web Consortium (W3C) a publié en 2008 les **Règles pour l'accessibilité des contenus Web (WCAG) 2.0,** qui définissent la façon de rendre les contenus web plus accessibles aux personnes en situation de handicap. Selon le W3C, suivre ces règles rendra les contenus accessibles à une plus grande variété de personnes en situation de handicap, dont les personnes aveugles et malvoyantes, les personnes sourdes et malentendantes, les personnes ayant des troubles d'apprentissage, des limitations cognitives ou des limitations motrices, entre autres. Souvent, suivre ces règles rendra aussi les contenus Web souvent plus faciles à utiliser pour les utilisateurs en général.

Dans cette deuxième séance, nous vous proposons de **vérifier l'accessibilité des images** du site que vous avez localisé pendant la première partie de cette étude, à l'aide de deux outils de vérification différents. Les images représentent une barrière d'accès aux contenus web spécialement pour les personnes aveugles et malvoyantes. Contrairement aux utilisateurs voyants, les personnes avec des limitations visuelles accèdent au web à l'aide de diverses technologies d'assistance, comme les logiciels d'écran loupe ou les lecteurs d'écran. Les lecteurs d'écran font une lecture du contenu à haute voix, c'est-à-dire qu'ils permettent aux personnes non-voyantes et malvoyantes d'utiliser la synthèse vocale pour naviguer sur le web. Pour cette raison, les équivalents textuels pour les images jouent un rôle important sur chaque page web.

1. SE CONNECTER À UN ORDINATEUR À DISTANCE

Les deux logiciels nécessaires pour effectuer cette dernière tâche sont installés sur les ordinateurs de la salle informatique de la Faculté de traduction et interprétation (FTI) de l'UNIGE. Pour accéder à l'un des ordinateurs :

- ➔ Lancez TeamViewer 10 (que vous avez déjà installé sur votre ordinateur local, en suivant les instructions du document *TeamViewer_GuideUtilisation.pdf*).

➔ Connectez-vous au bureau à distance en utilisant le numéro **ID du partenaire** que vous avez reçu par courriel avec les instructions pour cette deuxième séance (figure 1). Veuillez vous rappeler que le mot de passe est le suivant : **etude2014**

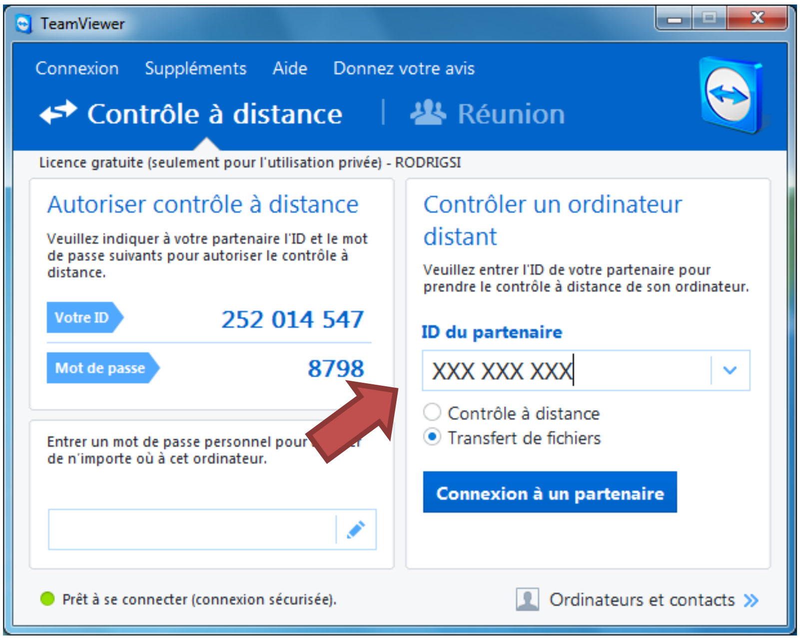


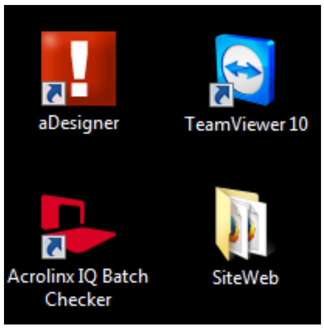
Figure 1. Interface principale de TeamViewer 10

Sur le bureau à distance, vous trouverez :

a) Votre **site web** en langue cible (français)

b) Un raccourci vers **aDesigner**

c) Un raccourci vers **Acrolinx IQ Batch Checker**



2. RAPPEL SUR BB FLASHBACK EXPRESS 5

Comme pour la première séance, veuillez lancer **BB FlashBack Express 5** sur votre ordinateur local pour enregistrer cette dernière tâche de l'étude.

Il est important de lancer BB FlashBack Express 5 depuis votre ordinateur, **et pas depuis l'ordinateur auquel vous êtes connecté.** Vérifiez que les options indiquées sur le document que vous avez reçu il y a quelques jours soient activées. Consultez-le une dernière fois avant l'exercice, si nécessaire (*BBFlashBackExpress_GuideUtilisation_DerniereVersion.pdf).*

➔ Pour commencer à enregistrer la séance, cliquez sur le bouton rond rouge (figure 2).

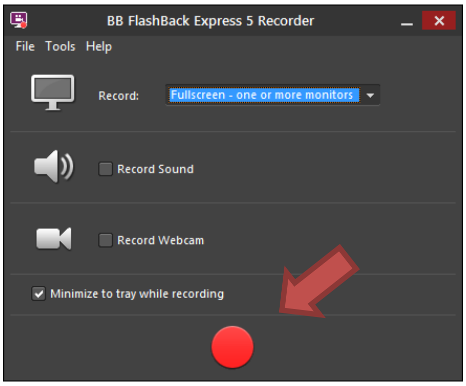


Figure 2. Fenêtre principale de BB FlashBack Express 5

3. VALIDATION DES FICHIERS HTML AVEC ADESIGNER

aDesigner est à la fois un simulateur de limitations visuelles et un outil de vérification d'accessibilité web qui permet aux professionnels du web de s'assurer que les contenus web sont accessibles pour les personnes aveugles et malvoyantes.

➔ Veuillez lancer l'application en double-cliquant sur le raccourci de **aDesigner,** qui se trouve sur le bureau distant. Sélectionnez ensuite l'option ***HTML Accessibility***, puis cliquez sur ***OK*** (figure 3).

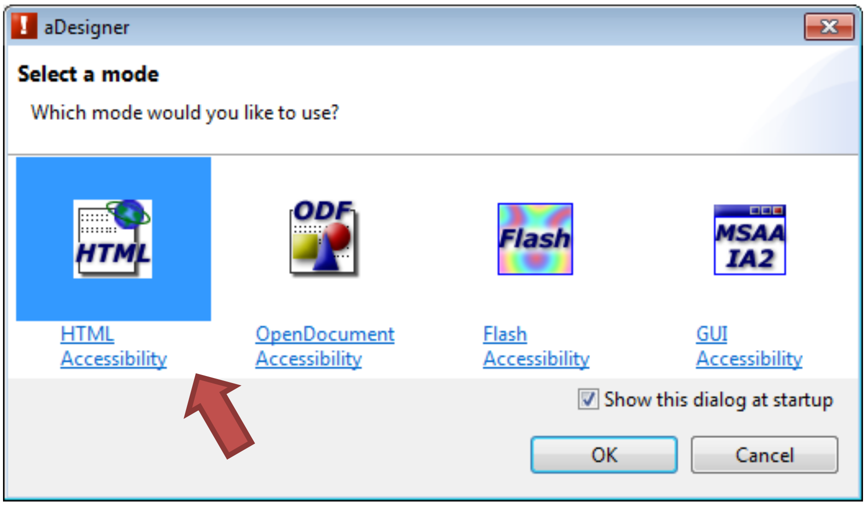


Figure 3. Fenêtre pour sélectionner le type de vérification souhaité

➔ Ouvrez l'un des fichiers HTML de votre site : ***File > Open > Browse > sélection du fichier > Ouvrir > OK.***

➔ À gauche, vous pourrez visualiser la page web sélectionnée. Cliquez ensuite sur le deuxième bouton à gauche, **"Visualize Blind Usability"**, à côté de l'icône des jumelles, pour vérifier le fichier (flèche rouge, figure 4). Veuillez noter que, pour ces instructions, c'est le site web en anglais qui apparaît sur les images d'exemple. Cependant, vous utiliserez la version française.

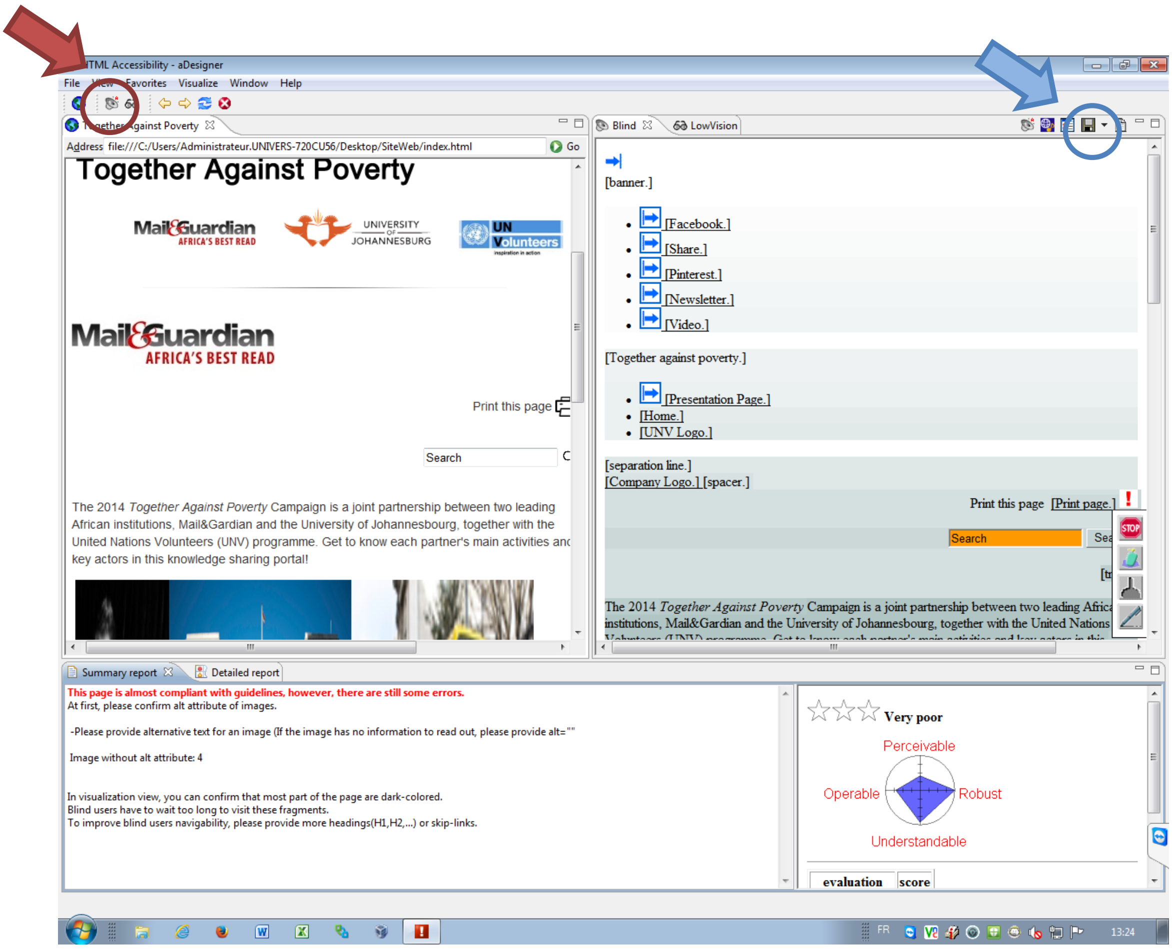


Figure 4. Interface principale de aDesigner

Vous verrez ensuite que, à droite, la même page est affichée sans images, avec les contenus affichés dans l'ordre dans lequel une personne non voyante ou malvoyante les entendrait lors de l'utilisation d'un lecteur d'écran.

En bas, vous trouverez deux onglets de vérification : ***a) Summary report*** et ***b) Detailed report***. Veuillez prendre quelques instants pour jeter un coup d'œil sur les résultats de la vérification.

Nous vous invitons ensuite à suivre les instructions suivantes :

➔ Créez un dossier sur le bureau à distance et nommez-le "aDesigner_reports". Retournez sur aDesigner et enregistrez les résultats au format .csv en cliquant sur ***Save results*** (flèche bleue, figure 4). Il n'est pas nécessaire d'ouvrir les fichiers de rapport que vous générerez et enregistrerez.

- ➔ Ensuite, sous l'onglet ***Detail report***, veuillez trier les résultats par type de problème trouvé en cliquant sur le titre de la colonne ***WCAG 2.0.***
- ➔ Les problèmes liés à l'accessibilité des images sont indiquées sous la catégorie **A: 1.1.1.** S'il y a des problèmes affichés sous cette catégorie, veuillez lire les remarques dans la colonne ***Description*** et, le cas échéant, faites les modifications que vous considérez pertinentes sur votre fichier.

IMPORTANT : Si vous estimez devoir faire des changements sur vos fichiers, vous avez deux possibilités :

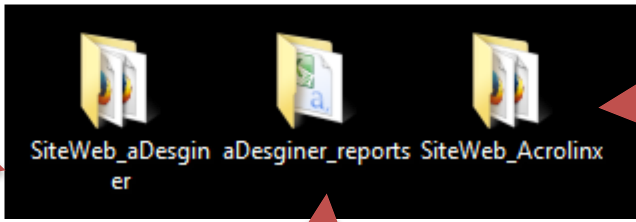
a) Faire les changements sur l'ordinateur à distance (vous avez à votre disposition Notepad++, ainsi que des outils de TAO comme DéjàVu ou SDL Trados).

b) Faire les changements dans le même environnement dans lequel vous avez travaillé pour localiser les fichiers, sur votre ordinateur local. Si vous choisissez cette option, veuillez vous référer à la section ***ASTUCES POUR L'UTILISATION DE TEAMVIEWER 10*** du document *TeamViewer_ GuideUtilisation.pdf,* sur la manière de transférer des fichiers de l'un ordinateur à l'autre.

- ➔ Veuillez suivre cette procédure pour les <u>trois fichiers de votre site.</u> Veuillez enregistrer également les différents rapports générés pour un même fichier (le nombre de rapports à enregistrer par fichier dépendra du nombre de vérifications faites sur ce même fichier, après avoir fait des modifications). Vous pouvez décider quand est-ce que vous êtes satisfait(e) des résultats et que vous êtes prêt(e) à passer à l'étape suivante.
- ➔ Avant de passer à l'étape 4, veuillez changer le nom du dossier de votre site et le remplacer par **SiteWeb_aDesginer**.
- ➔ Faites ensuite une copie du dossier, en l'appelant **SiteWeb_Acrolinx.**

Vous devriez désormais avoir sur le bureau à distance les dossiers suivants :

Un dossier contenant la dernière version de votre site web en français



Une copie exacte du dossier contenant la dernière version de votre site web en français, sur laquelle vous allez travailler à l'-étape suivante

Un dossier contenant les rapports des vérifications faites avec aDesigner

4. <u>VALIDATION DES FICHIERS HTML AVEC ACROLINX IQ BATCH CHECKER</u>

Acrolinx IQ Batch Checker est un outil qui permet de vérifier le contenu textuel des fichiers texte, HTML, XML ou PDF par rapport à un ensemble de règles linguistiques qui touchent à des aspects stylistiques, terminologiques, grammaticaux et orthographiques.

➔ Veuillez lancer l'application en double-cliquant sur le raccourci de **Acrolinx IQ Batch Checker,** qui se trouve sur le bureau distant.

➔ Vous devez maintenant définir les paramètres de vérification de la manière suivante (figure 5) :

1) Choisissez ***HTML*** comme "File type"

2) Choisissez ***UTF-8*** comme "Encoding"

3) Choisissez ***HTML_Image_Accessibility*** comme "Context Segment Definition (CSD)"

4) À l'aide de l'icône représentant un dossier jaune, sélectionnez le dossier contenant les fichiers à vérifier qui se trouve sur le bureau de l'ordinateur à distance "SiteWeb_Acrolinx".

5) Cliquez sur ***Find Files***. Les trois fichiers HTML à vérifier devraient s'afficher dans la colonne à droite. Il ne faut pas les sélectionner. Acrolinx IQ Batch Checker prendra les trois automatiquement lors de la vérification.

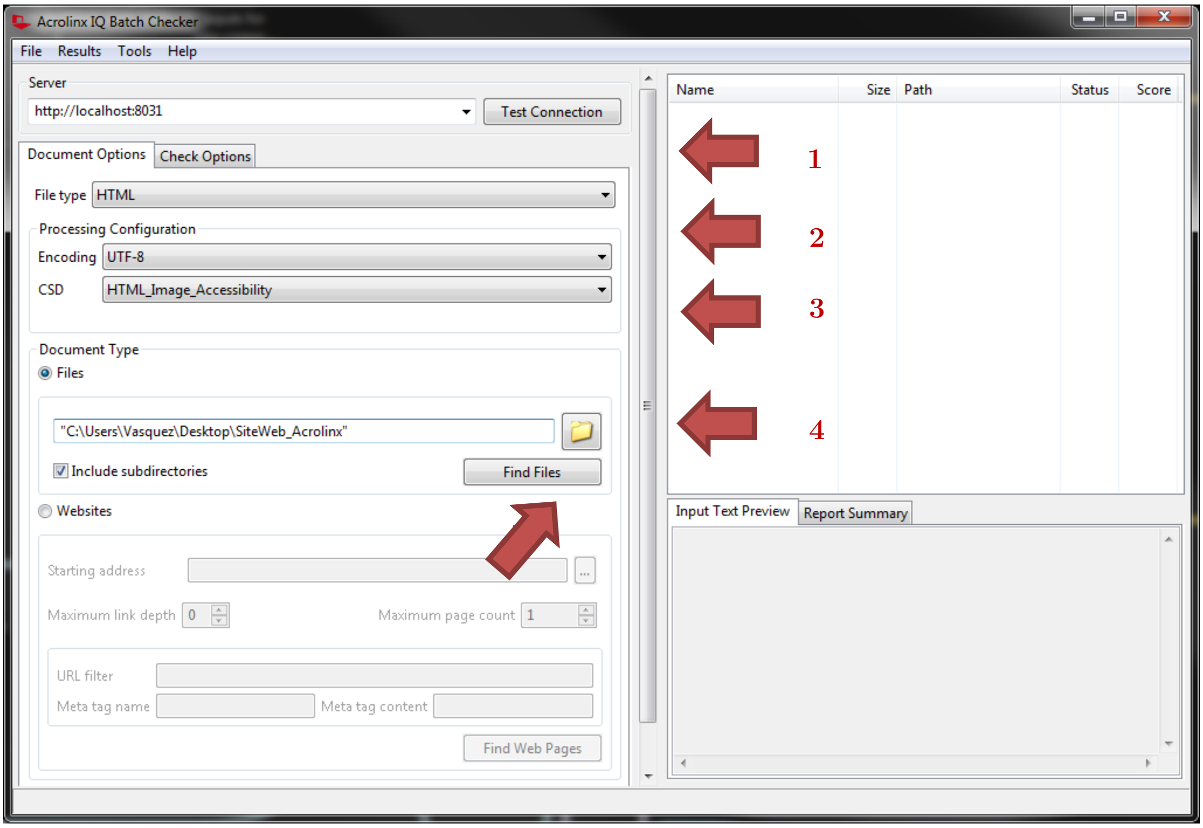


Figure 5. Options liées aux documents à vérifier avec Acrolinx IQ Batch Checker

➔ Assurez-vous ensuite sous l'onglet ***Check Options*** que l'ensemble de règles ***Image_Accessibility*** est sélectionné, que la langue est la bonne (fr) et que seule la case ***Style*** est cochée (figure 6).

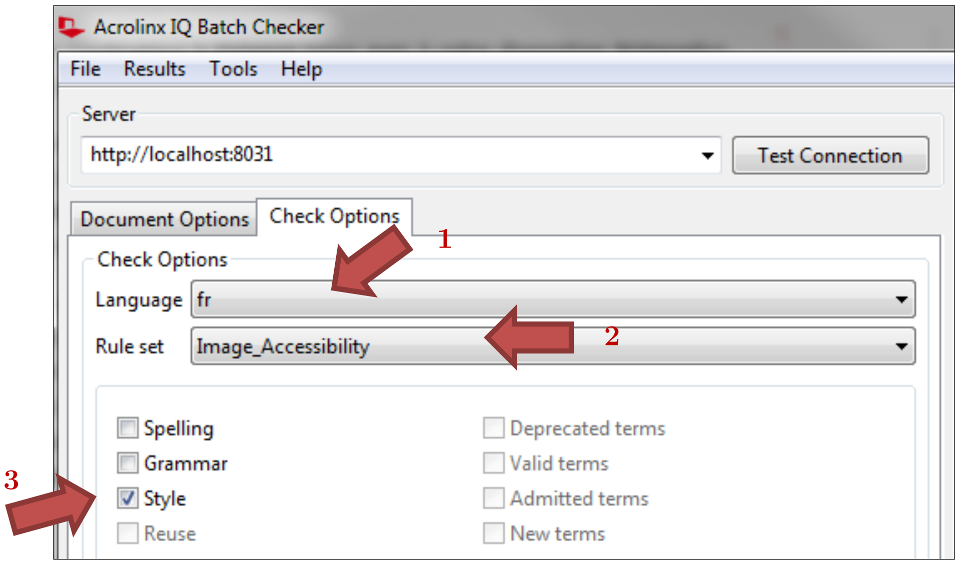


Figure 6. Options de vérification sur Acrolinx IQ Batch Checker

➔ Enfin, retournez sur l'onglet ***Document Options*** et appuyez sur le bouton 

pour générer le rapport de résultats (défilez à l'aide de la barre de défilement à droite, si nécessaire).

Pour visualiser les résultats, veuillez faire un clic droit sur chacun des fichiers évalués et choisissez **Show report.**

Le rapport s'ouvrira dans votre navigateur par défaut. Dans la dernière section du rapport, **Style**, vous trouverez, si l'outil a produit des résultats, les informations suivantes :
1) Un résumé des règles qui portent sur l'accessibilité des images et qui n'ont pas été respectées, et le nombre d'erreurs trouvées.
2) Une section plus détaillée avec les erreurs exactes trouvées et des suggestions d'amélioration, le cas échéant.

En cliquant sur le nom de la règle, vous trouverez une explication plus détaillée de l'erreur identifiée.

La procédure, après avoir fait cette deuxième vérification, est similaire à celle suivie pour aDesigner :

- ➔ Examinez le rapport de chaque fichier et, le cas échéant, faites les modifications que vous considérez pertinentes. Vous pouvez faire plusieures vérifications, si nécessaire, comme dans le cas de aDesigner.
- ➔ Veuillez noter que, dans le cas d'Acrolinx IQ Batch Checker, il n'est pas nécessaire d'enregistrer les fichiers de rapport (la sauvegarde est automatique).

Vous pouvez décider quand est-ce que vous êtes satisfait(e) des résultats et vous êtes prêt(e) à terminer la tâche.

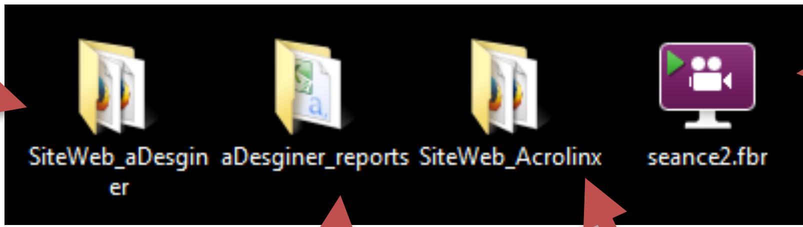
5. FICHIERS À RENDRE

Après avoir terminé l'exercice, veuillez arrêter l'enregistrement de l'écran et sauvegarder la vidéo résultante en la nommant **seance2.fbr**. Transférez ensuite le fichier vidéo vers le bureau de l'ordinateur à distance (sélectionnez et glissez).

Si, pour des raisons liées au type de connexion internet que vous avez ou la taille de la vidéo, le fichier vidéo ne se transfère pas, veuillez créer un fichier .zip avec la vidéo seulement et nous l'envoyer à l'adresse e-mail **silvia.rodriguez@unige.ch** à l'aide de WeTransfer : https://www.wetransfer.com/

Après la séance, nous allons récupérer vos fichiers directement sur l'ordinateur à distance sur lequel vous avez travaillé. Veuillez donc vous assurer que, sur le bureau, vous avez les dossiers et fichiers suivants :

Un dossier contenant la première version de votre site web en français



La vidéo de la deuxième séance

Un dossier contenant les rapports des vérifications faites avec aDesigner

Un dossier contenant la dernière version de votre site web en français

6. QUESTIONNAIRE LIÉ À LA TÂCHE EFFECTUÉE

Après l'exercice d'assurance qualité pour l'accessibilité des images que vous avez fait pendant cette deuxième séance, nous vous prions de bien vouloir remplir un bref questionnaire lié à la tâche effectuée. Le remplir ne devrait pas vous prendre plus de 15 minutes.

https://fr.surveymonkey.com/s/deuxieme_seance_etude2014

NOUS VOUS REMERCIONS INFINIMENT DE VOTRE PARTICIPATION À CETTE ÉTUDE !

Si vous avez un problème technique avec BBFlashBack Express 5, TeamViewer 10 ou les logiciels installés sur l'ordinateur à distance, je serai disponible sur **Skype (silvia_rodriguez_vazquez)** et Adobe Web Meeting, ainsi que joignable par **courriel (silvia.rodriguez@unige.ch)**

G.3 Quality assurance task brief - Scenario B (Acrolinx - aDesigner)

ÉTUDE SUR LA LOCALISATION WEB

Deuxième séance (environ 90 minutes)

Dans cette deuxième partie de l'étude, vous allez utiliser deux logiciels d'assurance qualité pour l'accessibilité web du site que vous avez localisé.

Le World Wide Web Consortium (W3C) a publié en 2008 les **Règles pour l'accessibilité des contenus Web (WCAG) 2.0,** qui définissent la façon de rendre les contenus web plus accessibles aux personnes en situation de handicap. Selon le W3C, suivre ces règles rendra les contenus accessibles à une plus grande variété de personnes en situation de handicap, dont les personnes aveugles et malvoyantes, les personnes sourdes et malentendantes, les personnes ayant des troubles d'apprentissage, des limitations cognitives ou des limitations motrices, entre autres. Souvent, suivre ces règles rendra aussi les contenus Web souvent plus faciles à utiliser pour les utilisateurs en général.

Dans cette deuxième séance, nous vous proposons de **vérifier l'accessibilité des images** du site que vous avez localisé pendant la première partie de cette étude, à l'aide de deux outils de vérification différents. Les images représentent une barrière d'accès aux contenus web spécialement pour les personnes aveugles et malvoyantes. Contrairement aux utilisateurs voyants, les personnes avec des limitations visuelles accèdent au web à l'aide de diverses technologies d'assistance, comme les logiciels d'écran loupe ou les lecteurs d'écran. Les lecteurs d'écran font une lecture du contenu à haute voix, c'est-à-dire qu'ils permettent aux personnes non-voyantes et malvoyantes d'utiliser la synthèse vocale pour naviguer sur le web. Pour cette raison, les équivalents textuels pour les images jouent un rôle important sur chaque page web.

1. SE CONNECTER À UN ORDINATEUR À DISTANCE

Les deux logiciels nécessaires pour effectuer cette dernière tâche sont installés sur les ordinateurs de la salle informatique de la Faculté de traduction et interprétation (FTI) de l'UNIGE. Pour accéder à l'un des ordinateurs :

➔ Lancez TeamViewer 10 (que vous avez déjà installé sur votre ordinateur local, en suivant les instructions du document *TeamViewer_GuideUtilisation.pdf*).

➔ Connectez-vous au bureau à distance en utilisant le numéro **ID du partenaire** que vous avez reçu par courriel avec les instructions pour cette deuxième séance (figure 1). Veuillez vous rappeler que le mot de passe est le suivant : **etude2014**

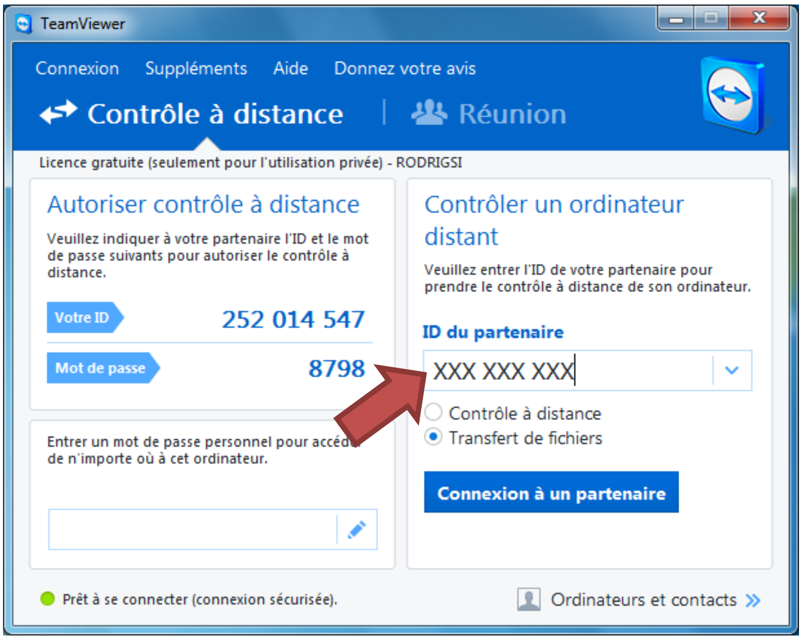


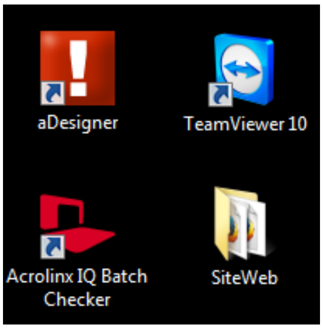
Figure 1. Interface principale de TeamViewer 10

Sur le bureau à distance, vous trouverez :

a) Votre **site web** en langue cible (français)

b) Un raccourci vers **aDesigner**

c) Un raccourci vers **Acrolinx IQ Batch Checker**



2. RAPPEL SUR BB FLASHBACK EXPRESS 5

Comme pour la première séance, veuillez lancer **BB FlashBack Express 5** sur votre ordinateur local pour enregistrer cette dernière tâche de l'étude.

Il est important de lancer BB FlashBack Express 5 depuis votre ordinateur, **et pas depuis l'ordinateur auquel vous êtes connecté.** Vérifiez que les options indiquées sur le document que vous avez reçu il y a quelques jours soient activées. Consultez-le une dernière fois avant l'exercice, si nécessaire (*BBFlashBackExpress_GuideUtilisation_ DerniereVersion.pdf).*

➔ Pour commencer à enregistrer la séance, cliquez sur le bouton rond rouge (figure 2).

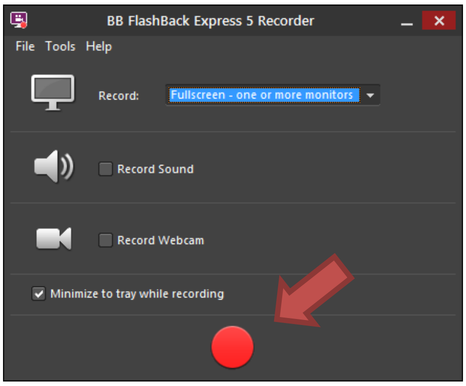


Figure 2. Fenêtre principale de BB FlashBack Express 5

3. VALIDATION DES FICHIERS HTML AVEC ACROLINX IQ BATCH CHECKER

Acrolinx IQ Batch Checker est un outil qui permet de vérifier le contenu textuel des fichiers texte, HTML, XML ou PDF par rapport à un ensemble de règles linguistiques qui touchent à des aspects stylistiques, terminologiques, grammaticaux et orthographiques.

Veuillez lancer l'application en double-cliquant sur le raccourci de **Acrolinx IQ Batch Checker,** qui se trouve sur le bureau distant.

➔ Vous devez maintenant définir les paramètres de vérification de la manière suivante (figure 3) :
1) Choisissez ***HTML*** comme "File type"
2) Choisissez ***UTF-8*** comme "Encoding"
3) Choisissez ***HTML_Image_Accessibility*** comme "Context Segment Definition (CSD)"
4) À l'aide de l'icône représentant un dossier jaune, sélectionnez le dossier contenant les fichiers à vérifier qui se trouve sur le bureau de l'ordinateur à distance "SiteWeb".
5) Cliquez sur ***Find Files***. Les trois fichiers HTML à vérifier devraient s'afficher dans la colonne à droite. Il ne faut pas les sélectionner. Acrolinx IQ Batch Checker prendra les trois automatiquement lors de la vérification.

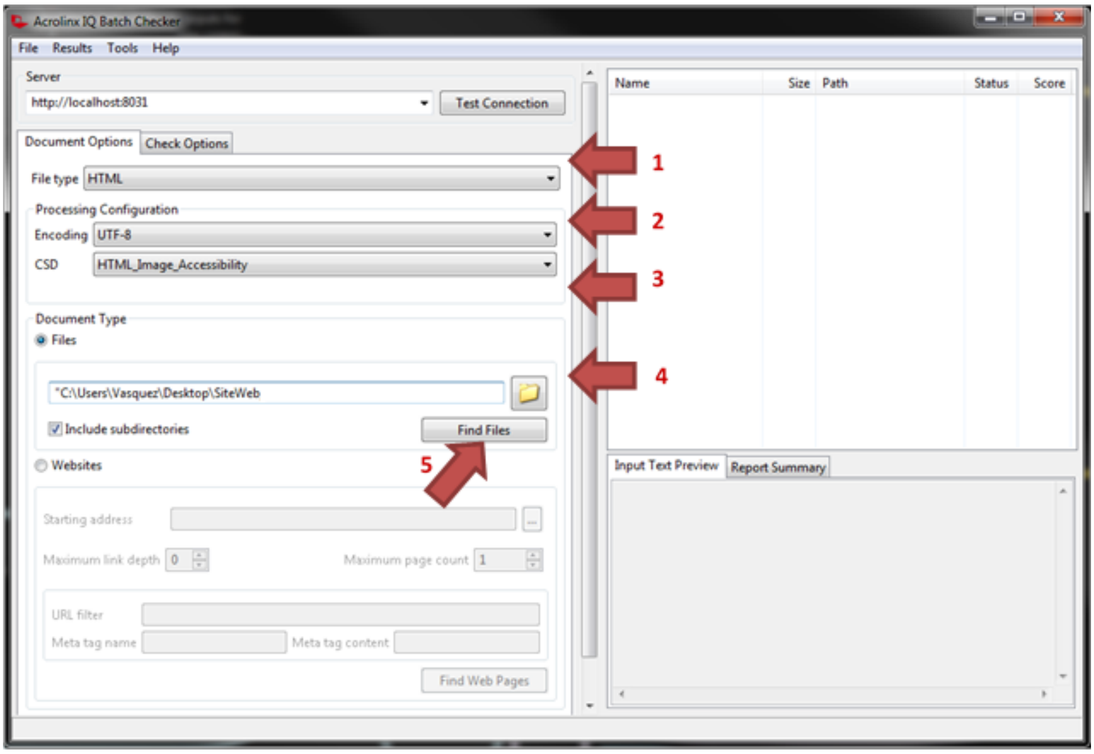


Figure 3. Options liées aux documents à vérifier avec Acrolinx IQ Batch Checker

➔ Assurez-vous ensuite sous l'onglet ***Check Options*** que l'ensemble de règles ***Image_Accessibility*** est sélectionné, que la langue est la bonne (fr) et que seule la case ***Style*** est cochée (figure 4).

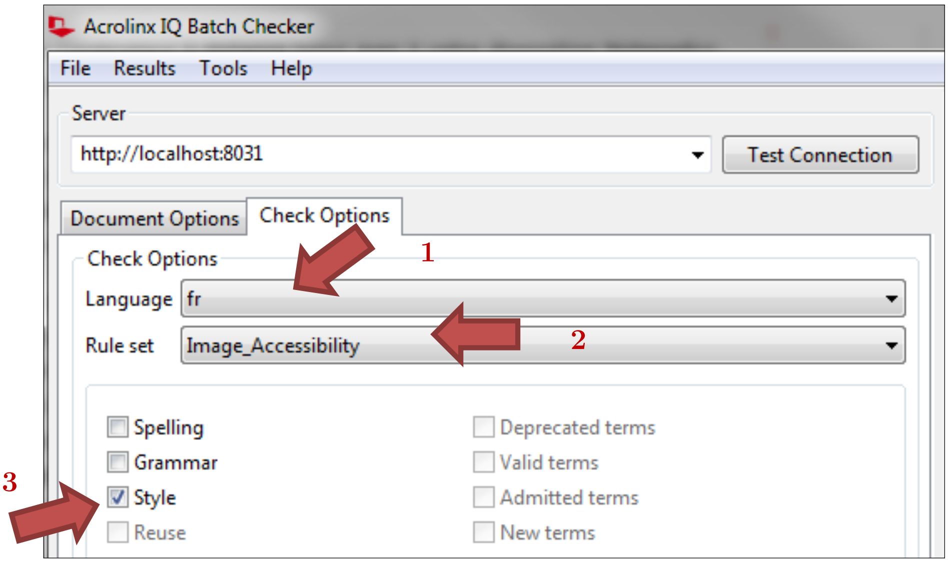


Figure 4. Options de vérification sur Acrolinx IQ Batch Checker

- ➔ Enfin, retournez sur l'onglet ***Document Options*** et appuyez sur le bouton 

pour générer le rapport de résultats (défilez à l'aide de la barre de défilement à droite, si nécessaire).

Pour visualiser les résultats, veuillez faire un clic droit sur chacun des fichiers évalués et choisissez **Show report.**

Le rapport s'ouvrira dans votre navigateur par défaut. Dans la dernière section du rapport, **Style**, vous trouverez, si l'outil a produit des résultats, les informations suivantes :
1) Un résumé des règles qui portent sur l'accessibilité des images et qui n'ont pas été respectées, et le nombre d'erreurs trouvées.
2) Une section plus détaillée avec les erreurs exactes trouvées et des suggestions d'amélioration, le cas échéant.

En cliquant sur le nom de la règle, vous trouverez une explication plus détaillée de l'erreur identifiée.

Nous vous invitons ensuite à suivre les instructions suivantes :

- ➔ Veuillez prendre quelques instants pour examiner le rapport de chaque fichier, lire les remarques sur chacune des règles qui n'a pas été respectée et, le cas échéant, faites les modifications que vous considérez pertinentes sur vos fichiers. Vous pouvez faire plusieures vérifications par fichier, si nécessaire.

IMPORTANT : Si vous estimez devoir faire des changements sur vos fichiers, vous avez deux possibilités :

a) Faire les changements sur l'ordinateur à distance (vous avez à votre disposition Notepad++, ainsi que des outils de TAO comme DéjàVu ou SDL Trados).

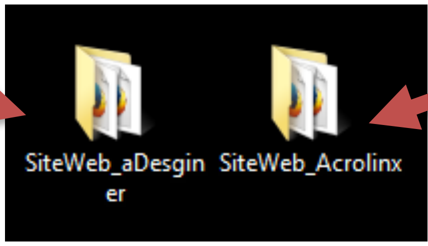
b) Faire les changements dans le même environnement dans lequel vous avez travaillé pour localiser les fichiers, sur votre ordinateur local. Si vous choisissez cette option, veuillez vous référer à la section ***ASTUCES POUR L'UTILISATION DE TEAMVIEWER 10*** du document *TeamViewer_ GuideUtilisation.pdf,* sur la manière de transférer des fichiers de l'un ordinateur à l'autre.

- ➔ Vous pouvez décider quand est-ce que vous êtes satisfait(e) des résultats et vous êtes prêt(e) à passer à l'étape suivante.
- ➔ Avant de passer à l'étape 4, veuillez changer le nom du dossier de votre site (dont les fichiers vous avez modifiés) et le remplacer par **SiteWeb_Acrolinx**.

➔ Faites ensuite une copie du dossier, en l'appelant **SiteWeb_aDesigner.**

Vous devriez désormais avoir sur le bureau à distance les dossiers suivants :

Une copie exacte du dossier contenant la dernière version de votre site web en français, sur laquelle vous allez travailler à l'étape suivante



Un dossier contenant la dernière version de votre site web en français

4. VALIDATION DES FICHIERS HTML AVEC ADESIGNER

aDesigner est à la fois un simulateur de limitations visuelles et un outil de vérification d'accessibilité web qui permet aux professionnels du web de s'assurer que les contenus web sont accessibles pour les personnes aveugles et malvoyantes.

➔ Veuillez lancer l'application en double-cliquant sur le raccourci de **aDesigner,** qui se trouve sur le bureau distant. Sélectionnez ensuite l'option ***HTML Accessibility***, puis cliquez sur ***OK*** (figure 5).

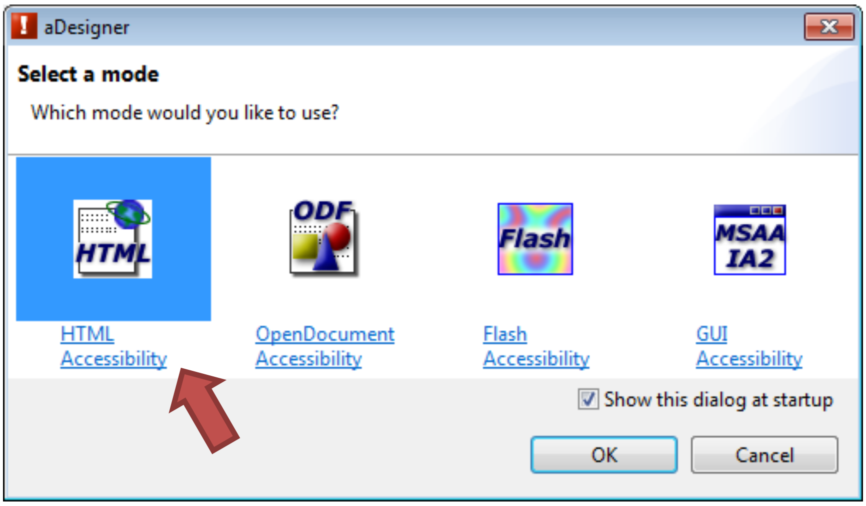


Figure 5. Fenêtre pour sélectionner le type de vérification souhaité

➔ Ouvrez l'un des fichiers HTML de votre dossier "SiteWeb_aDesigner" : ***File > Open > Browse > sélection du fichier > Ouvrir > OK.***

➔ À gauche, vous pourrez visualiser la page web sélectionnée. Cliquez ensuite sur le deuxième bouton à gauche, **"Visualize Blind Usability"**, à côté de l'icône des jumelles, pour vérifier le fichier (flèche rouge, figure 6). Veuillez noter que, pour ces instructions, c'est le site web en anglais qui apparaît sur les images d'exemple. Cependant, vous utiliserez la version française.

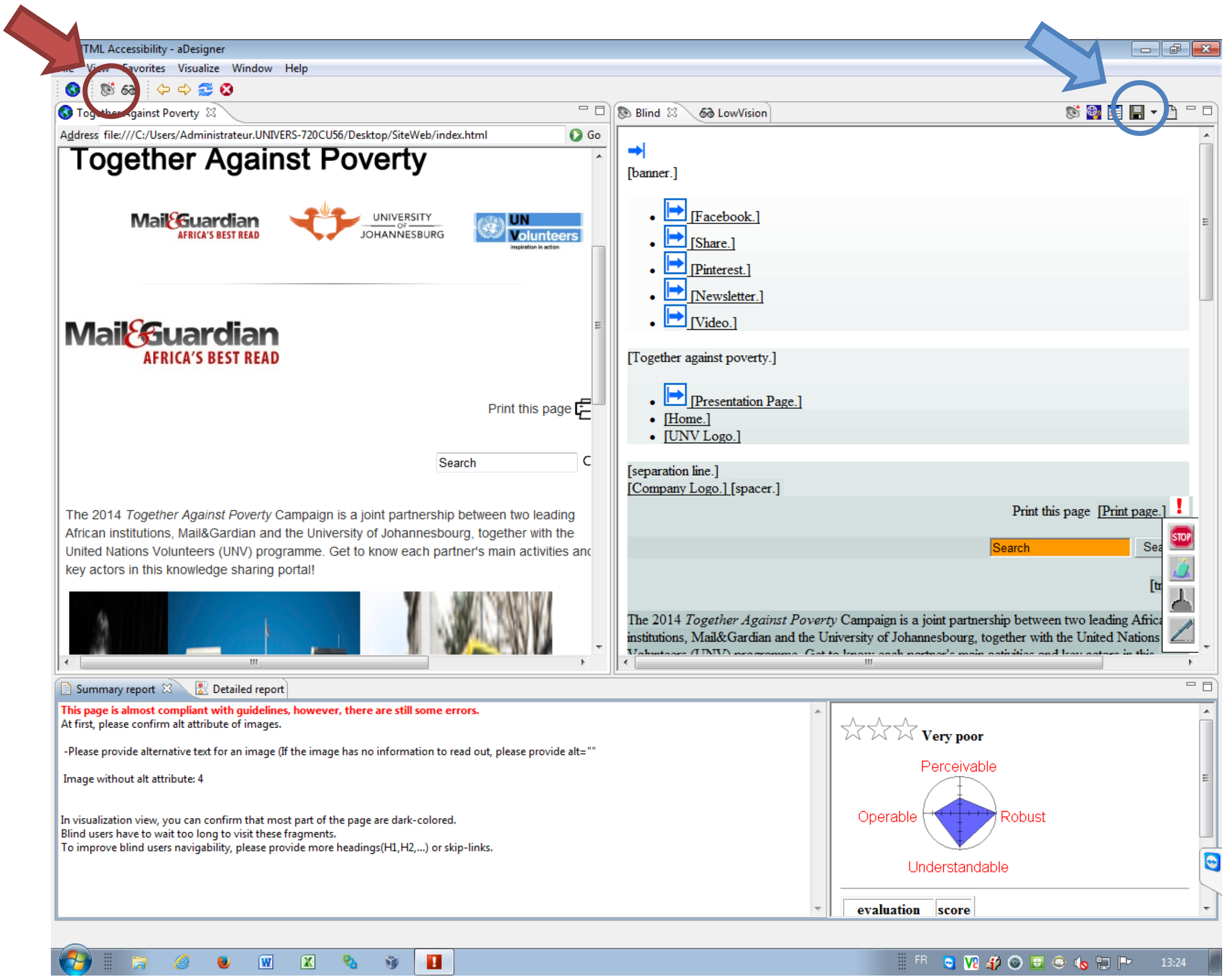


Figure 6. Interface principale de aDesigner

Vous verrez ensuite que, à droite, la même page est affichée sans images, avec les contenus affichés dans l'ordre dans lequel une personne non voyante ou malvoyante les entendrait lors de l'utilisation d'un lecteur d'écran.

En bas, vous trouverez deux onglets de vérification : ***a) Summary report*** et ***b) Detailed report***. Veuillez prendre quelques instants pour jeter un coup d'œil sur les résultats de la vérification.

Nous vous invitons ensuite à suivre les instructions suivantes :

➔ Créez un dossier sur le bureau à distance et nommez-le "aDesigner_reports". Retournez sur aDesigner et enregistrez les résultats au format .csv en cliquant sur ***Save results*** (flèche bleue, figure 6). Il n'est pas nécessaire d'ouvrir les fichiers de rapport que vous générerez et enregistrerez.

➔ Ensuite, sous l'onglet ***Detail report***, veuillez trier les résultats par type de problème trouvé en cliquant sur le titre de la colonne ***WCAG 2.0.***

- ➔ Les problèmes liés à l'accessibilité des images sont indiquées sous la catégorie **A: 1.1.1.** S'il y a des problèmes affichés sous cette catégorie, veuillez lire les remarques dans la colonne ***Description*** et, le cas échéant, faites les modifications que vous considérez pertinentes sur votre fichier.

La procédure, après avoir fait cette vérification, est similaire à celle suivie pour Acrolinx IQ Batch Checker :

- ➔ Veuillez vérifier les trois fichiers de votre site en suivant la même procédure. Veuillez enregistrer également les différents rapports générés pour un même fichier (le nombre de rapports à enregistrer par fichier dépendra du nombre de vérifications faites sur ce même fichier, après avoir fait des modifications).
- ➔ Vous pouvez décider quand est-ce que vous êtes satisfait(e) des résultats et vous êtes prêt(e) à terminer la tâche.

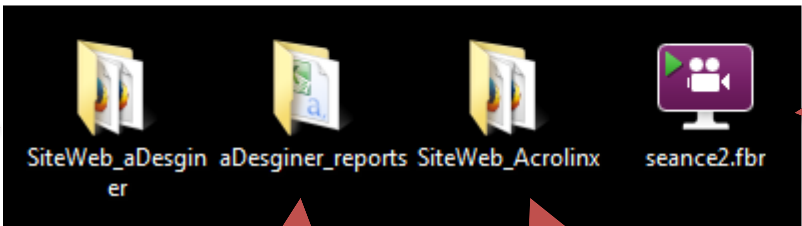
5. FICHIERS À RENDRE

Après avoir terminé l'exercice, veuillez arrêter l'enregistrement de l'écran et sauvegarder la vidéo résultante en la nommant **seance2.fbr**. Transférez ensuite le fichier vidéo vers le bureau de l'ordinateur à distance (sélectionnez et glissez).

Si, pour des raisons liées au type de connexion internet que vous avez ou la taille de la vidéo, le fichier vidéo ne se transfère pas, veuillez créer un fichier .zip avec la vidéo seulement et nous l'envoyer à l'adresse e-mail **silvia.rodriguez@unige.ch** à l'aide de WeTransfer : https://www.wetransfer.com/

Après la séance, nous allons récupérer vos fichiers directement sur l'ordinateur à distance sur lequel vous avez travaillé. Veuillez donc vous assurer que, sur le bureau, vous avez les dossiers et fichiers suivants :

Un dossier contenant la dernière version de votre site web en français



La vidéo de la deuxième séance

Un dossier contenant les rapports des vérifications faites avec aDesigner

Un dossier contenant la première version de votre site web en français

6. QUESTIONNAIRE LIÉ À LA TÂCHE EFFECTUÉE

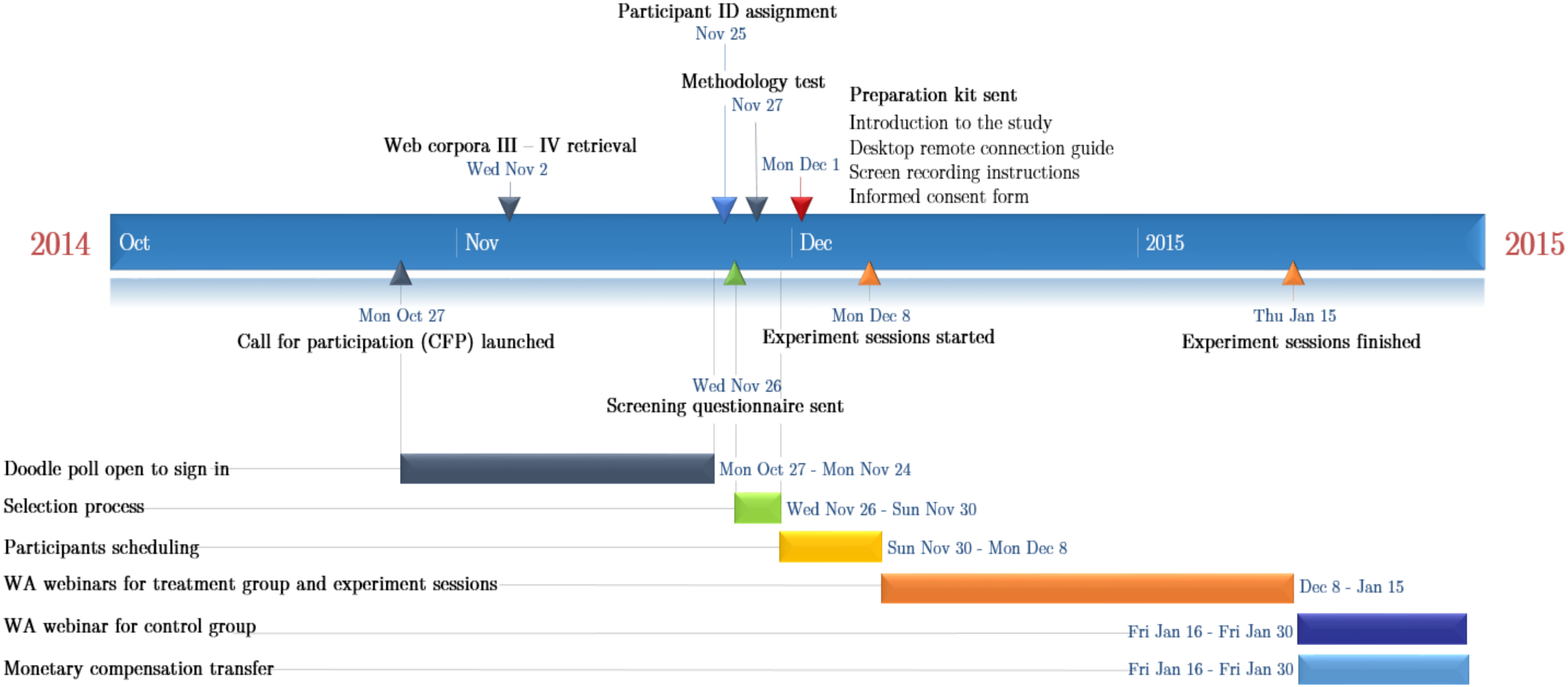
Après l'exercice d'assurance qualité pour l'accessibilité des images que vous avez fait pendant cette deuxième séance, nous vous prions de bien vouloir remplir un bref questionnaire lié à la tâche effectuée. Le remplir ne devrait pas vous prendre plus de 15 minutes.

https://fr.surveymonkey.com/s/deuxieme_seance_etude2014

NOUS VOUS REMERCIONS INFINIMENT DE VOTRE PARTICIPATION À CETTE ÉTUDE !

Si vous avez un problème technique avec BBFlashBack Express 5, TeamViewer 10 ou les logiciels installés sur l'ordinateur à distance, je serai disponible sur **Skype (silvia_rodriguez_vazquez)** et Adobe Web Meeting, ainsi que joignable par **courriel (silvia.rodriguez@unige.ch)**

Appendix H. Experimental study's timeline: Stage 1


Participant ID assignment
Nov 25
Methodology test
Nov 27
Preparation kit sent
Introduction to the study
Desktop remote connection guide
Screen recording instructions
Informed consent form
Mon Dec 1
Web corpora III – IV retrieval
Wed Nov 2
2014
Oct
Nov
Dec
2015
2015
Mon Oct 27
Call for participation (CFP) launched
Mon Dec 8
Experiment sessions started
Thu Jan 15
Experiment sessions finished
Wed Nov 26
Screening questionnaire sent
Doodle poll open to sign in
Mon Oct 27 - Mon Nov 24
Selection process
Wed Nov 26 - Sun Nov 30
Participants scheduling
Sun Nov 30 - Mon Dec 8
WA webinars for treatment group and experiment sessions
Dec 8 - Jan 15
WA webinar for control group
Fri Jan 16 - Fri Jan 30
Monetary compensation transfer
Fri Jan 16 - Fri Jan 30

Appendix I. Stage 2: User evaluation briefing sheet

Évaluation d'équivalents textuels pour les images sur le web
Instructions

Introduction

Nous vous remercions d'avoir accepté de participer à cet exercice d'évaluation. Votre contribution est extrêmement précieuse pour notre projet de recherche et nous apprécions votre volonté de collaborer avec nous.

Le but de cette étude est d'évaluer la pertinence, du point de vue de l'accessibilité web, d'une sélection d'équivalents textuels pour des images sur le web en français, produit par des traducteurs francophones. Les équivalents textuels sont lus par les lecteurs d'écran, quand ceux-ci trouvent des images, permettant rendre accessible le contenu ou la fonction des images aux personnes malvoyantes et non-voyantes.

Les 130 images sélectionnées font partie d'un site internet, créé à des fins de recherche, d'une campagne fictive pour le développement intitulé en anglais « Together Against Poverty », récemment lancé par trois partenaires internationaux: le journal sud-africain Mail & Guardian, l'Université de Johannesburg et le programme des Volontaires des Nations Unies (VNU).

Structure des pages web dont sont tirées les images
Le site contient trois pages web; une page par partenaire. Chaque page présente une structure rigoureusement identique (en-tête, corps de page et pied de page).

Dans l'en-tête de chaque page, à droite, il y a une section avec des liens vers différents réseaux sociaux. Immédiatement en dessous, se trouve le menu de navigation principal, qui permet à l'utilisateur de naviguer d'une page à l'autre. À gauche, le titre de la campagne « Together Against Poverty » est affiché.

Le corps de la page est divisé en deux sections. La première contient les éléments suivants :

a) Logo du partenaire
b) Options d'impression de la page
c) Formulaire de recherche
d) Information sur la campagne
e) Images liées aux activités menées par le partenaire

La deuxième section contient des renseignements plus précis sur le partenaire en question :

a) À gauche, se trouve une liste de documents relatifs au partenaire qui peuvent être téléchargés ou imprimés directement

b) À droite, les activités du partenaire peuvent être consultées, par catégorie.

Enfin, dans le pied de page, sont affichées des informations sur l'entreprise qui a conçu le portail web de la campagne.

Structure du questionnaire sur SurveyMonkey

Sur la première page du questionnaire sur SurveyMonkey, la plateforme d'évaluation choisie pour cet exercice, nous allons vous poser quelques questions démographiques. Sur la deuxième page, vous trouverez des questions par rapport à votre expérience et opinion en matière d'accessibilité des sites multilingues et des images sur le web. Veuillez noter que toutes les réponses données dans le questionnaire sont strictement confidentielles et ne seront utilisées qu'à des fins scientifiques. Les données personnelles ne seront révélées en aucune circonstance.

La tâche d'évaluation elle-même commencera à la troisième page. À partir de ce moment-là, chaque page de SurveyMonkey sera consacrée à une seule image. Vous y trouverez d'abord la description du contexte de l'image: sur quelle page et dans quelle partie de la page l'image apparaît, et quel est son contenu et/ou fonction. Une liste d'équivalents textuels potentiels vous sera ensuite proposée pour cette image. Veuillez noter que le nombre d'équivalents textuels sur chaque liste peut varier selon les images (de 3 à 35 solutions possibles). Après la description contextuelle de chaque image, nous annoncerons le nombre total d'équivalents textuels proposés. Au bas de chaque page, un espace « Commentaire » optionnel vous permettra d'ajouter vos éventuelles remarques.

Pour chaque équivalent textuel, nous aimerions que vous attribuiez une note, sur une échelle de 1 à 4, selon l'acceptabilité et la pertinence de l'équivalent textuel par rapport au contexte décrit. Au moment de la notation, vous pouvez réfléchir à la question suivante: si je naviguais sur cette page web en situation réelle et que je tombais sur cette image, l'équivalent textuel serait-il approprié ?

Valeurs des notes sur l'échelle d'évaluation

Les valeurs associées à chaque note de l'échelle sont les suivantes:

1: Pas acceptable - l'équivalent textuel n'est pas approprié pour le contexte décrit
2: Acceptable - l'équivalent textuel est acceptable par rapport au contexte décrit, mais pas toute l'information offerte est pertinente
3: Pertinent - l'équivalent textuel donne des informations minimales mais suffisantes par rapport au contexte décrit
4: Très pertinent - l'équivalent textuel donne des informations ciblées par rapport au contexte décrit

Codage des équivalents textuels

Avant de commencer l'exercice, veuillez noter que :
a) Si l'une des solutions proposées dans la liste d'équivalents textuels potentiels est un alt vide, vous lirez « alt vide ». Dans le code HTML, l'attribut `alt` permet d'introduire l'équivalent textuel pour l'image. Ainsi, un alt vide serait exprimé dans le code HTML de la page sous la forme `alt=""` (sans aucun texte entre les guillemets). Comme vous le savez déjà, si une image contient un alt vide, les lecteurs d'écran ignoreront le plus souvent l'image et n'annonceront pas sa présence à l'utilisateur.
b) Si l'une des solutions proposées dans la liste d'équivalents textuels potentiels est de ne pas introduire un attribut alt pour l'image dans le code HTML, vous lirez « no alt ». Contrairement aux images avec des alt vides, qui sont souvent ignorées par les lecteurs d'écran, les images sans attribut alt sont signalées par les lecteurs d'écran, mais aucun équivalent textuel n'est fourni. Parfois, les lecteurs d'écran lisent, à la place, le nom du fichier.

Informations supplémentaires

Texte dans deux langues

Le site web a été traduit de l'anglais vers le français par des traducteurs différents. C'est pourquoi vous trouverez parfois des équivalents textuels dans les deux langues. Toutefois, pendant l'évaluation, veuillez toujours attribuer votre note partant du principe que, en situation réelle, vous navigueriez sur un site en français.

Dans les descriptions contextuelles des images, on fera parfois référence à des titres ou rubriques des pages web. Nous les avons laissés en anglais, afin que vous puisez avoir le même contexte de départ que celui des traducteurs. N'hésitez pas à prendre contact avec Mme Rodriguez Vazquez si vous avez besoin des précisions supplémentaires.

Durée

Selon nos estimations, cet exercice d'évaluation peut prendre quatre ou cinq heures au minimum et une journée et demie au maximum. Cependant, il n'est absolument pas nécessaire d'avoir toute une journée libre pour la tâche. Si vous voulez commencer le sondage un jour, vous arrêter après une heure et continuer dans l'après-midi ou même le lendemain, vous pouvez le faire. SurveyMonkey se souviendra de vos réponses (en stockant un cookie dans votre navigateur), ainsi que de la page sur laquelle vous vous trouviez avant de quitter le sondage. Vous pouvez éteindre et rallumer votre ordinateur, mais vous devez toujours utiliser le même navigateur. Veuillez vous assurer aussi que les cookies sont activés.

Cookies

Les cookies sont des fichiers stockés sur votre ordinateur par les sites web que vous visitez et contiennent des informations telles que vos préférences de navigation ou votre statut de connexion. Afin que SurveyMonkey puisse se souvenir de vos réponses, veuillez vous assurer que les cookies sont activés. Vous pouvez nous contacter si vous avez des questions par rapport à l'activation des cookies. Nous essayerons de vous guider en fonction du navigateur que vous utilisez.

Rétribution

Comme vous le savez probablement déjà, cet exercice d'évaluation fait partie d'un projet de recherche doctorale, pour lequel nous n'avons malheureusement que des fonds limités. Chaque participant recevra 100 francs suisses en guise de remerciement pour le précieux travail accompli. À la fin de l'évaluation, nous vous demanderons vos coordonnées bancaires pour effectuer le versement.

Merci encore de votre collaboration et de votre temps, si précieux !

Si vous êtes prêt à commencer, veuillez cliquer sur le lien suivant :

https://fr.surveymonkey.com/r/evaluation-equivalents-textuels

Appendix J. Stage 2: Evaluation data set

Table J.1. Location and contextual description in French, as received by evaluators, of images included in the Mail & Guardian campaign's partner web page (index.html)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I1	10	En-tête	Image située en haut de la page, avant la section des réseaux sociaux et le menu de navigation, et qui est presque aussi large que la page web. Il s'agit d'un collage comprenant trois images, montrant différents scénarios. La première image (à gauche) est floue. Deux hommes semblent parler au téléphone. Dans la deuxième image (au centre du collage), le portrait d'un homme portant des lunettes est affiché. Seuls les sourcils, les lunettes et le nez sont visibles. La troisième image (à droite) montre une tour de contrôle du trafic aérien.
I2	12	En-tête	Petite icône de Facebook, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Facebook (visiter la page Facebook de Mail & Guardian ou partager la page de la campagne sur le mur Facebook de l'utilisateur).
I3	17	En-tête	Petite icône de Twitter, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Twitter (visiter le compte Twitter de Mail & Guardian ou écrire un tweet sur le compte Twitter de l'utilisateur).
I4	13	En-tête	Petite icône de Pinterest, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Pinterest (visiter le compte Pinterest de Mail & Guardian ou partager la page de la campagne sur le compte Pinterest de l'utilisateur).
I5	18	En-tête	Petite icône d'un journal qui représente un bulletin d'information en ligne, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à ce bulletin (s'abonner au bulletin du journal Mail & Guardian ou lire le bulletin).

Table J.1. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I6	24	En-tête	Petite icône de YouTube, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à YouTube (visiter le canal YouTube du journal Mail & Guardian ou visionner une vidéo).
I7	6	En-tête	Image montrant le nom de la campagne en anglais : Together Against Poverty.
I8	23	En-tête	Logo du journal Mail & Guardian, situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I9	23	En-tête	Logo de l'Université de Johannesburg (UJ), situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I10	24	En-tête	Logo du programme des Volontaires des Nations Unies (VNU), situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I11	12	En-tête	Ligne noire et mince qui sépare l'en-tête du corps de page.
I12	28	Corps de page	Logo du journal Mail & Guardian, d'une taille plus grande que celle du logo utilisé dans le menu de navigation, situé au début du contenu principal de la page. En cliquant sur l'image, l'utilisateur est mené vers le site officiel du Mail & Guardian (en dehors du site internet de la campagne).
I13	7	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I14	12	Corps de page	Petite icône d'une imprimante, située à droite du texte « Print this page », au début du corps de page. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur.

Table J.1. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I15	3	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I16	8	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I17	29	Corps de page	Section : Images liées aux activités menées par le partenaire. Monsieur Sepp Blatter, dans une attitude de réflexion.
I18	9	Corps de page	Section : Images liées aux activités menées par le partenaire. Membres du groupe Nexus.
I19	26	Corps de page	Section : Images liées aux activités menées par le partenaire. Portrait d'une dame d'origine indienne.
I20	13	Corps de page	Section : Images liées aux activités menées par le partenaire. Œuvre d'art urbaine intitulée Open House dans une rue du Cap.
I21	24	Corps de page	Section : Images liées aux activités menées par le partenaire. Jeunes qui dansent dans la rue.
I22	31	Corps de page	Section : Images liées aux activités menées par le partenaire. Groupe de personnes qui posent dans une salle intérieure avec des colonnes de style gréco-romain.
I23	13	Corps de page	Section : Images liées aux activités menées par le partenaire. Orateur avec un micro dans la main.
I24	11	Corps de page	Section : Images liées aux activités menées par le partenaire. Vue aérienne en noir et blanc d'un groupe de jeunes hommes.
I25	35	Corps de page	Section : Images liées aux activités menées par le partenaire. Jeune homme devant un bâtiment historique.
I26	23	Corps de page	Section : Images liées aux activités menées par le partenaire. Bus jaune à deux étages.
I27	7	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Press coverage reports.

Table J.1. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I28	25	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Press coverage reports.
I29	16	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant : Press coverage reports.
I30	6	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Affiliated members.
I31	19	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Affiliated members.
I32	20	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant : Affiliated members.
I33	7	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Sustainable development projects.
I34	19	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Sustainable development projects.
I35	16	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant : Sustainable development projects.

Table J.1. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I36	8	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Justice », qui montre un bâtiment ancien, de style gréco-romain. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I37	10	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Agriculture », qui montre une prairie avec une colline en arrière plan. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I38	15	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Gender », qui montre une femme africaine avec des vêtements colorés. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I39	9	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Elections », qui montre une urne et des bulletins de vote avec des portraits de personnes différentes. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I40	10	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Civil Rights », qui montre le siège d'une organisation internationale au bout d'une avenue. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I41	12	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « People », qui montre un orateur devant un micro. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I42	15	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Industry », qui montre un avion qui décolle. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.

Table J.1. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I43	10	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Other topics ». Image de taille moyenne, au-dessus de la rubrique « Others », qui montre un marchand ambulant avec un parapluie entouré de sacs en papier. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I44	10	Pied de page	Ligne noire et mince qui sépare le corps de page du pied de page.
I45	11	Pied de page	Logo de l'entreprise qui a conçu le site internet de la campagne.

Table J.2. Location and contextual description in French, as received by evaluators, of images included in the University of Johannesburg campaign's partner web page (index2.html)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I46	14	En-tête	Image située en haut de la page, avant la section des réseaux sociaux et le menu de navigation, et qui est presque aussi large que la page web. Il s'agit d'un coucher de soleil au bord de la mer avec un ciel nuageux. Il y a quelques palmiers au premier plan et une petite maison dans le fond.
I47	28	En-tête	Petite icône de Facebook, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Facebook (visiter la page Facebook de l'Université de Johannesburg ou partager la page de la campagne sur le mur Facebook de l'utilisateur).
I48	32	En-tête	Petite icône de YouTube, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à YouTube (visiter le canal YouTube de l'Université de Johannesburg ou visionner une vidéo).
I49	14	En-tête	Petite icône de Tumblr, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Tumblr (visiter le blog Tumblr de l'Université de Johannesburg ou partager la page de la campagne sur le blog Tumblr de l'utilisateur).
I50	19	En-tête	Petite icône d'une enveloppe qui représente un courrier électronique, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de partager la page web par courrier électronique.
I51	19	En-tête	Petite icône RSS, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée aux flux RSS (s'abonner au flux RSS de l'Université de Johannesburg ou lire ses actualités sur les flux RSS).
I52	14	En-tête	Image montrant le nom de la campagne en anglais : Together Against Poverty.
I53	21	En-tête	Logo du journal Mail & Guardian, situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.

Table J.2. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I54	16	En-tête	Logo de l'Université de Johannesburg (UJ), situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I55	22	En-tête	Logo du programme des Volontaires des Nations Unies (VNU), situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I56	14	En-tête	Ligne noire et mince qui sépare l'en-tête du corps de page.
I57	21	Corps de page	Logo de l'Université de Johannesburg, d'une taille plus grande que celle du logo utilisé dans le menu de navigation, situé au début du contenu principal de la page. En cliquant sur l'image, l'utilisateur est mené vers le site officiel de l'Université de Johannesburg (en dehors du site internet de la campagne).
I58	5	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I59	10	Corps de page	Petite icône d'une imprimante, située à droite du texte « Print this page », au début du corps de page. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur.
I60	7	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I61	6	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I62	30	Corps de page	Section : Images liées aux activités menées par le partenaire. Homme tenant dans la main un document dans un cadre.
I63	21	Corps de page	Section : Images liées aux activités menées par le partenaire. Train en marche.

Table J.2. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I64	18	Corps de page	Section : Images liées aux activités menées par le partenaire. Oratrice devant un micro.
I65	23	Corps de page	Section : Images liées aux activités menées par le partenaire. Jeunes lors de la remise de diplômes.
I66	25	Corps de page	Section : Images liées aux activités menées par le partenaire. Vue aérienne de personnes à l'intérieur d'un bâtiment historique.
I67	11	Corps de page	Section : Images liées aux activités menées par le partenaire. Homme avec le regard perdu, avec une valise.
I68	10	Corps de page	Section : Images liées aux activités menées par le partenaire. Jeune fille souriante avec un livre dans les mains.
I69	21	Corps de page	Section : Images liées aux activités menées par le partenaire. Joueur de football sur le terrain.
I70	19	Corps de page	Section : Images liées aux activités menées par le partenaire. Jeune homme devant un bâtiment historique.
I71	19	Corps de page	Section : Images liées aux activités menées par le partenaire. Bus jaune à deux étages.
I72	16	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Public information sessions.
I73	22	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Public information sessions.
I74	16	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant : Public information sessions.
I75	19	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Continuous education.

Table J.2. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I76	21	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Continuous education.
I77	18	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant : Continuous education.
I78	10	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Careers center and jobs fair.
I79	28	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Careers center and jobs fair.
I80	18	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant : Careers center and jobs fair.
I81	10	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Board », qui montre quatre hommes en costume devant un poster académique. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I82	10	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Dean », qui montre une oratrice devant un micro. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I83	15	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Vice Dean », qui montre le portrait d'un homme avec des lunettes. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.

Table J.2. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I84	27	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Campus », qui montre un grand espace ouvert, avec des bâtiments en arrière plan. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I85	15	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Transport », qui montre le dessin simplifié d'un bus. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I86	9	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « International », qui montre un avion qui décolle. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I87	15	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Register », qui montre une série de livres avec la même reliure. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I88	30	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Our University ». Image de taille moyenne, au-dessus de la rubrique « Statistics », qui montre un histogramme dont les données ne sont pas lisibles sur l'écran. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I89	8	Pied de page	Ligne noire et mince qui sépare le corps de page du pied de page.
I90	10	Pied de page	Logo de l'entreprise qui a conçu le site internet de la campagne.

Table J.3. Location and contextual description in French, as received by evaluators, of images included in the United Nations Volunteers campaign's partner web page (index3.html)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I91	4	En-tête	Image située en haut de la page, avant la section des réseaux sociaux et le menu de navigation, et qui est presque aussi large que la page web. Il s'agit d'un coucher de soleil au bord de la mer avec un ciel nuageux. Il y a quelques palmiers au premier plan et une petite maison dans le fond.
I92	22	En-tête	Petite icône de Facebook, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Facebook (visiter la page Facebook du programme VNU ou partager la page de la campagne sur le mur Facebook de l'utilisateur).
I93	20	En-tête	Petite icône de Twitter, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Twitter (visiter le compte Twitter du programme VNU ou écrire un tweet sur le compte Twitter de l'utilisateur).
I94	18	En-tête	Petite icône de Pinterest, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Pinterest (visiter le compte Pinterest du programme VNU ou partager la page de la campagne sur le compte Pinterest de l'utilisateur).
I95	22	En-tête	Petite icône de Google+, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Google+ (visiter la page Google+ du programme VNU ou partager la page de la campagne sur la page Google+ de l'utilisateur).
I96	16	En-tête	Petite icône de YouTube, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à YouTube (visiter le canal YouTube du programme VNU ou visionner une vidéo).
I97	16	En-tête	Petite icône de Tumblr, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à Tumblr (visiter le blog Tumblr du programme VNU ou partager la page de la campagne sur le blog Tumblr de l'utilisateur).

Table J.3. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I98	13	En-tête	Petite icône d'Instagram, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de visiter le compte Instagram du programme VNU.
I99	17	En-tête	Petite icône d'une enveloppe qui représente un courrier électronique, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de partager la page web par courrier électronique.
I100	24	En-tête	Petite icône d'un journal qui représente un bulletin d'information en ligne, située dans la section des réseaux sociaux, avec un lien intégré qui permet à l'utilisateur de faire une action quelconque liée à ce bulletin (s'abonner au bulletin du programme VNU ou lire le bulletin).
I101	8	En-tête	Image montrant le nom de la campagne en anglais : Together Against Poverty.
I102	21	En-tête	Logo du journal Mail & Guardian, situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I103	22	En-tête	Logo de l'Université de Johannesburg (UJ), situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I104	16	En-tête	Logo du programme des Volontaires des Nations Unies (VNU), situé dans le menu de navigation qui permet à l'utilisateur de naviguer entre les trois pages du site. En cliquant sur l'image, l'utilisateur est mené vers la page web de ce partenaire sur le site de la campagne.
I105	16	En-tête	Ligne noire et mince qui sépare l'en-tête du corps de page.
I106	28	Corps de page	Logo du programme des Volontaires des Nations Unies, d'une taille plus grande que celle du logo utilisé dans le menu de navigation, situé au début du contenu principal de la page. En cliquant sur l'image, l'utilisateur est mené vers le site officiel du programme VNU (en dehors du site internet de la campagne).

Table J.3. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I107	17	Corps de page	Petite icône d'une imprimante, située à droite du texte « Print this page », au début du corps de page. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur.
I108	3	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I109	5	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I110	35	Corps de page	Section : Images liées aux activités menées par le partenaire. Grand bâtiment de plusieurs étages.
I111	17	Corps de page	Section : Images liées aux activités menées par le partenaire. Tiger Woods se prépare à frapper la balle, le club de golf dans l'air.
I112	4	Corps de page	Image transparente qui rend l'effet visuel d'un espace entre deux éléments quelconques (titres, images, paragraphes) sur l'écran.
I113	32	Corps de page	Section : Images liées aux activités menées par le partenaire. Actrice de théâtre, habillée comme une reine.
I114	33	Corps de page	Section : Images liées aux activités menées par le partenaire. Portrait d'une dame africaine.
I115	23	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Partnership agreements.
I116	32	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Community networks.

Table J.3. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I117	17	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Public information sessions.
I118	9	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône PDF qui se trouve au même niveau que le texte suivant : Fundraising efforts.
I119	17	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une flèche pointée vers le bas. En cliquant sur l'image, l'utilisateur peut télécharger un document PDF. L'icône se trouve au même niveau que le texte suivant : Fundraising efforts.
I120	23	Corps de page	Section : Liste de documents relatifs au partenaire. Petite icône d'une imprimante. En cliquant sur l'image, l'utilisateur fait apparaître la fenêtre des options d'impression associées à l'imprimante connectée à l'ordinateur. L'icône se trouve au même niveau que le texte suivant dans la page source en anglais : Fundraising efforts.
I121	13	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Health », qui montre deux médecins chirurgiens avec des masques. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I122	23	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Land », qui montre un paysage vert, avec un ciel bleu mais nuageux. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I123	6	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Gender », qui montre une femme africaine avec des vêtements colorés. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.

Table J.3. (continued)

Image ID	Alt texts assessed	Location (fr)	Contextual description (fr)
I124	13	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Science », qui montre deux pièces de puzzle qui semblent s'emboîter. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I125	22	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Dialogue », qui montre deux bulles de dialogue. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I126	16	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Refugees », qui montre un dessin en 3D d'une maison avec des balcons. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I127	24	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Culture », qui montre un paysage de montagne. Au premier plan, on voit un monument en pierre couronné par une croix. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I128	21	Corps de page	Section : Activités du partenaire, par catégorie, sous le titre « Volunteer Areas ». Image de taille moyenne, au-dessus de la rubrique « Peace », qui montre le siège d'une organisation internationale avec des drapeaux de pays différents au premier plan. En cliquant sur l'image, l'utilisateur est mené à une page du site avec des renseignements sur ce sujet-là.
I129	16	Pied de page	Ligne noire et mince qui sépare le corps de page du pied de page.
I130	10	Pied de page	Logo de l'entreprise qui a conçu le site internet de la campagne.

Appendix K. Supplementary data

K.1 Translation of text alternatives

Table K.1. Alt texts edited per localiser, by translation version and tool used (1)

Participant	T1	T2	T3	aDesigner	Acrolinx
P01	84	23	0	23	0
P04	111	51	25	25	51
P06	74	74	0	0	74
P16	70	92	25	25	92
P18	82	0	0	0	0
P19	0	40	20	20	40
P21	107	19	4	4	19
P22	66	36	19	19	36
P23	88	102	56	56	102
P25	6	15	28	15	28
P26	126	23	4	4	23
P31	0	109	6	6	109
P33	91	0	0	0	0
P34	87	0	67	0	67
P36	0	119	11	11	119
P40	0	74	49	74	49
P41	0	36	25	36	25
P42	42	59	91	59	91
P44	72	10	100	10	100
P46	80	12	16	12	16
P49	86	0	0	0	0
P51	77	59	21	21	59
P52	31	9	38	9	38
P53	0	0	5	5	0
P55	87	12	72	12	72
P56	77	0	0	0	0
P58	86	35	63	35	63
P59	115	42	31	31	42
Total	**1745**	**1051**	**776**	**512**	**1315**

Table K.2. Alt texts edited per localiser, by QA scenario [(2)]

Scenario A: *aDesigner-Acrolinx*				Scenario B: *Acrolinx-aDesigner*			
Participant	**T1**	**T2**	**T3**	**Participant**	**T1**	**T2**	**T3**
P01	84	*23*	0	P04	111	*51*	25
P18	82	0	0	P06	74	*74*	0
P25	6	15	*28*	P16	70	*92*	25
P33	91	0	0	P19	0	*40*	20
P34	87	0	*67*	P21	107	*19*	4
P40	0	*74*	49	P22	66	*36*	19
P41	0	*36*	25	P23	88	*102*	56
P42	42	*59*	91	P26	126	*23*	4
P44	72	10	*100*	P31	0	*109*	6
P46	80	12	*16*	P36	0	*119*	11
P49	86	0	0	P51	77	*59*	21
P52	31	9	*38*	P53	0	0	*5*
P55	87	12	*72*	P56	77	0	0
P58	86	35	*63*	P59	115	*42*	31
Total	**834**	**285**	***549***	**Total**	**911**	***766***	**227**

Notes:

[(1)] Translation versions T1 and T3 correspond to translation conditions 'none' and 'both' when we referred to the use of accessibility-oriented QA tools in Chapter 8. Hence, results for these conditions are not repeated in Table K.1.

[(2)] The highest value observed after comparing results from translation versions T2 and T3 per participant is shown in italics.

K.2 Appropriateness of text alternatives

Notes:

To better interpret the gray scale colour coding used in the figures included in this appendix, notice that score 1 (not appropriate) is always at the bottom of the graphic (i.e. lowest part of the y-axis) and score 4 (very pertinent) is always at the top.

(3) The following abbreviations are used in Tables K.3 to K.33: **P. ID** (participant ID); **SD** (standard deviation); **%** (proportion of score).

(4) Due to space restrictions, we used the abbreviated form 'Acrolinx' for translation condition 'Both-last-Acrolinx' (T3) in Tables K.19 to K.29, and in Figures K.1, K.3, and K.15 to K.24.

(5) Due to space restrictions, we used the abbreviated form 'aDesigner' for translation condition 'Both-last-aDesigner' (T3) in Tables K.31, K.33 and K.25; and in Figures K.2, K.4, K.26, K.28 and K.30.

Table K.3. Control group, QA Scenario A (T2-aDesigner - T3-Acrolinx). Score mean, standard deviation and proportions (%) per participant (3)

P. ID	Tool	Mean	SD	1	%	2	%	3	%	4	%
P18	none	1.77	0.98	496	54.51	198	21.76	145	15.93	71	7.80
	aDesigner	1.77	0.98	496	54.51	198	21.76	145	15.93	71	7.80
	Both-last-Acrolinx	1.77	0.98	496	54.51	198	21.76	145	15.93	71	7.80
P25	none	1.56	0.84	576	63.30	189	20.77	115	12.64	30	3.30
	aDesigner	1.59	0.84	552	60.66	206	22.64	122	13.41	30	3.30
	Both-last-Acrolinx	1.78	0.96	478	52.53	210	23.08	162	17.80	60	6.59
P34	none	1.75	0.99	512	56.26	181	19.89	146	16.04	71	7.80
	aDesigner	1.75	0.99	512	56.26	181	19.89	146	16.04	71	7.80
	Both-last-Acrolinx	2.05	1.12	420	46.15	157	17.25	205	22.53	128	14.07
P40	none	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	aDesigner	1.56	0.84	581	63.85	181	19.89	119	13.08	29	3.19
	Both-last-Acrolinx	1.65	0.93	562	61.76	156	17.14	143	15.71	49	5.38
P41	none	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	aDesigner	1.71	0.94	515	56.59	203	22.31	135	14.84	57	6.26
	Both-last-Acrolinx	1.75	0.97	506	55.60	187	20.55	152	16.70	65	7.14
P42	none	1.67	0.94	542	59.56	179	19.67	133	14.62	56	6.15
	aDesigner	1.78	0.99	494	54.29	193	21.21	150	16.48	73	8.02
	Both-last-Acrolinx	2.28	1.09	302	33.19	199	21.87	262	28.79	147	16.15
P52	none	1.64	0.92	553	60.77	187	20.55	115	12.64	55	6.04
	aDesigner	1.60	0.88	564	61.98	186	20.44	119	13.08	41	4.51
	Both-last-Acrolinx	1.68	0.96	543	59.67	173	19.01	132	14.51	62	6.81

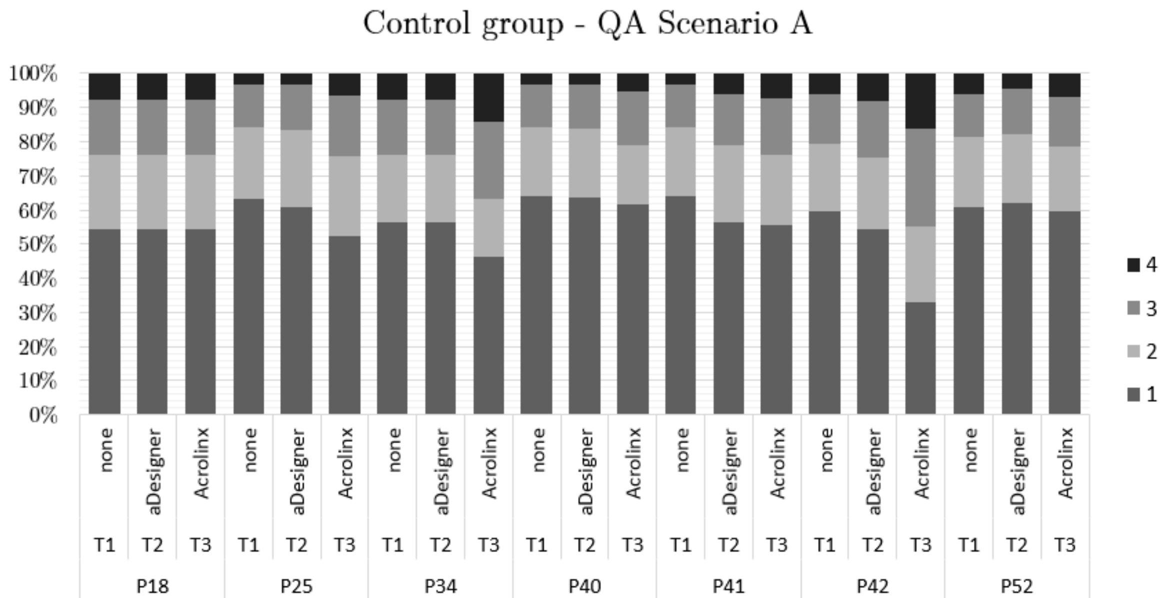


Figure K.1. Distribution of scores per participant by translation version: Control group (QA Scenario A) (4)

Table K.4. Control group, QA Scenario B (T2-Acrolinx - T3-aDesigner). Score mean, standard deviation and proportions (%) per participant

P. ID	Tool	Mean	SD	1	%	2	%	3	%	4	%
P04	none	2.00	1.01	386	42.42	222	24.40	221	24.29	81	8.90
	Acrolinx	2.25	1.07	309	33.96	191	20.99	283	31.10	127	13.96
	Both-last-aDesigner	2.52	1.05	214	23.52	178	19.56	348	38.24	170	18.68
P16	none	1.76	0.99	510	56.04	184	20.22	144	15.82	72	7.91
	Acrolinx	2.25	1.12	325	35.71	195	21.43	232	25.49	158	17.36
	Both-last-aDesigner	2.46	1.07	230	25.27	213	23.41	285	31.32	182	20.00
P19	none	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	Acrolinx	1.74	1.01	533	58.57	163	17.91	133	14.62	81	8.90
	Both-last-aDesigner	1.93	1.07	449	49.34	176	19.34	184	20.22	101	11.10
P23	none	1.79	1.00	495	54.40	195	21.43	139	15.27	81	8.90
	Acrolinx	2.11	1.21	440	48.35	107	11.76	187	20.55	176	19.34
	Both-last-aDesigner	2.69	1.01	147	16.15	209	22.97	333	36.59	221	24.29
P31	none	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	Acrolinx	2.45	1.07	229	25.16	213	23.41	294	32.31	174	19.12
	Both-last-aDesigner	2.49	1.05	210	23.08	221	24.29	303	33.30	176	19.34
P36	none	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	Acrolinx	2.36	1.05	250	27.47	230	25.27	283	31.10	147	16.15
	Both-last-aDesigner	2.38	1.06	252	27.69	214	23.52	293	32.20	151	16.59
P53	none	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	Acrolinx	1.55	0.83	582	63.96	184	20.22	116	12.75	28	3.08
	Both-last-aDesigner	1.55	0.83	582	63.96	183	20.11	116	12.75	29	3.19

Table K.5. Treatment group, QA Scenario A (T2-aDesigner - T3-Acrolinx). Score mean, standard deviation and proportions (%) per participant

P. ID	Tool	Mean	SD	1	%	2	%	3	%	4	%
P01	none	1.77	0.99	503	55.27	191	20.99	141	15.49	75	8.24
	aDesigner	1.82	1.00	473	51.98	205	22.53	153	16.81	79	8.68
	Both-last-Acrolinx	1.82	1.00	473	51.98	205	22.53	153	16.81	79	8.68
P33	none	1.76	1.00	509	55.93	181	19.89	145	15.93	75	8.24
	aDesigner	1.76	1.00	509	55.93	181	19.89	145	15.93	75	8.24
	Both-last-Acrolinx	1.76	1.00	509	55.93	181	19.89	145	15.93	75	8.24
P44	none	1.77	1.00	508	55.82	182	20.00	141	15.49	79	8.68
	aDesigner	1.84	1.02	469	51.54	198	21.76	159	17.47	84	9.23
	Both-last-Acrolinx	2.32	1.09	291	31.98	182	20.00	289	31.76	148	16.26
P46	none	1.72	0.97	527	57.91	178	19.56	141	15.49	64	7.03
	aDesigner	1.67	0.95	548	60.22	172	18.90	129	14.18	61	6.70
	Both-last-Acrolinx	1.68	0.94	541	59.45	181	19.89	130	14.29	58	6.37
P49	none	1.70	0.96	530	58.24	185	20.33	130	14.29	65	7.14
	aDesigner	1.70	0.96	530	58.24	185	20.33	130	14.29	65	7.14
	Both-last-Acrolinx	1.70	0.96	530	58.24	185	20.33	130	14.29	65	7.14
P55	none	1.72	0.96	518	56.92	187	20.55	143	15.71	62	6.81
	aDesigner	1.76	0.96	496	54.51	197	21.65	156	17.14	61	6.70
	Both-last-Acrolinx	2.15	1.09	358	39.34	180	19.78	246	27.03	126	13.85
P58	none	1.78	1.01	507	55.71	181	19.89	139	15.27	83	9.12
	aDesigner	1.98	1.07	416	45.71	210	23.08	168	18.46	116	12.75
	Both-last-Acrolinx	2.33	1.10	281	30.88	217	23.85	245	26.92	167	18.35

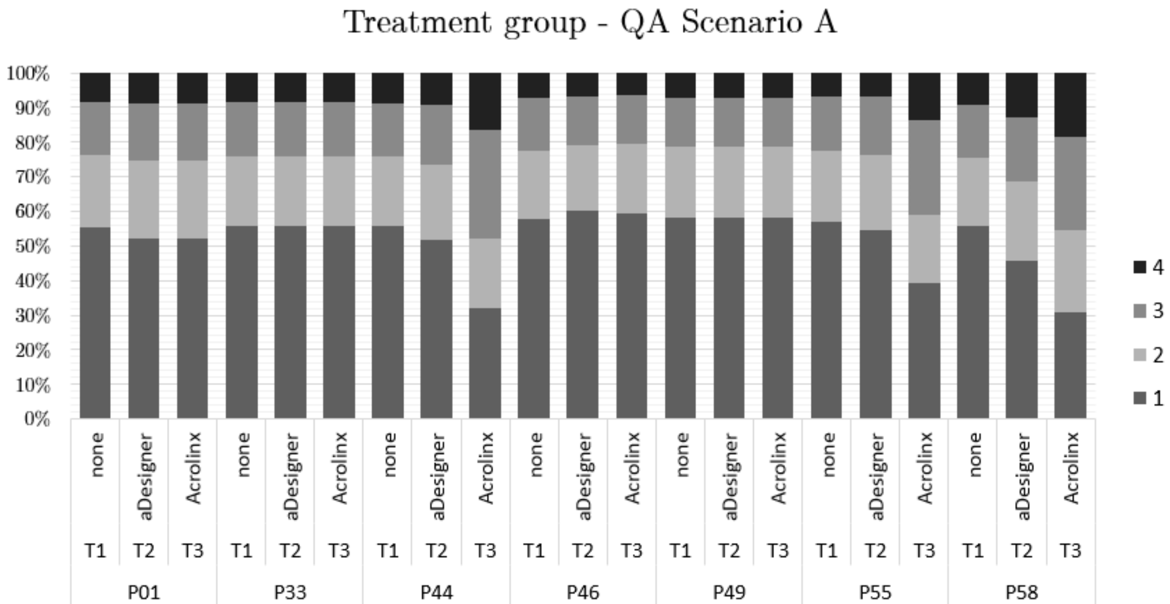


Figure K.2. Distribution of scores per participant by translation version: Treatment group (QA Scenario A)

Table K.6. Treatment group, QA Scenario B (T2-Acrolinx - T3-aDesigner). Score mean, standard deviation and proportions (%) per participant

P. ID	Tool	Mean	SD	1	%	2	%	3	%	4	%
P06	none	1.67	0.94	545	59.89	177	19.45	130	14.29	58	6.37
	Acrolinx	1.98	1.09	431	47.36	175	19.23	191	20.99	113	12.42
	Both-last-aDesigner	1.98	1.09	431	47.36	175	19.23	191	20.99	113	12.42
P21	none	2.24	1.07	293	32.20	256	28.13	214	23.52	147	16.15
	Acrolinx	2.31	1.09	277	30.44	241	26.48	224	24.62	168	18.46
	Both-last-aDesigner	2.33	1.08	266	29.23	249	27.36	227	24.95	168	18.46
P22	none	1.74	0.97	518	56.92	181	19.89	145	15.93	66	7.25
	Acrolinx	1.76	1.01	519	57.03	163	17.91	153	16.81	75	8.24
	Both-last-aDesigner	1.86	1.02	519	57.03	163	17.91	153	16.81	75	8.24
P26	none	2.76	1.03	153	16.81	160	17.58	349	38.35	248	27.25
	Acrolinx	2.76	1.04	151	16.59	170	18.68	332	36.48	257	28.24
	Both-last-aDesigner	2.78	1.03	146	16.04	168	18.46	338	37.14	258	28.35
P51	none	1.71	0.98	544	59.78	158	17.36	140	15.38	68	7.47
	Acrolinx	2.04	1.11	419	46.04	159	17.47	211	23.19	121	13.30
	Both-last-aDesigner	2.09	1.12	399	43.85	162	17.80	215	23.63	134	14.73
P56	none	1.75	1.01	524	57.58	175	19.23	130	14.29	81	8.90
	Acrolinx	1.75	1.01	524	57.58	175	19.23	130	14.29	81	8.90
	Both-last-aDesigner	1.75	1.01	524	57.58	175	19.23	130	14.29	81	8.90
P59	none	2.26	1.03	265	29.12	267	29.34	252	27.69	126	13.85
	Acrolinx	2.24	1.00	263	28.90	276	30.33	259	28.46	112	12.31
	Both-last-aDesigner	2.36	1.03	233	25.60	259	28.46	274	30.11	144	15.82

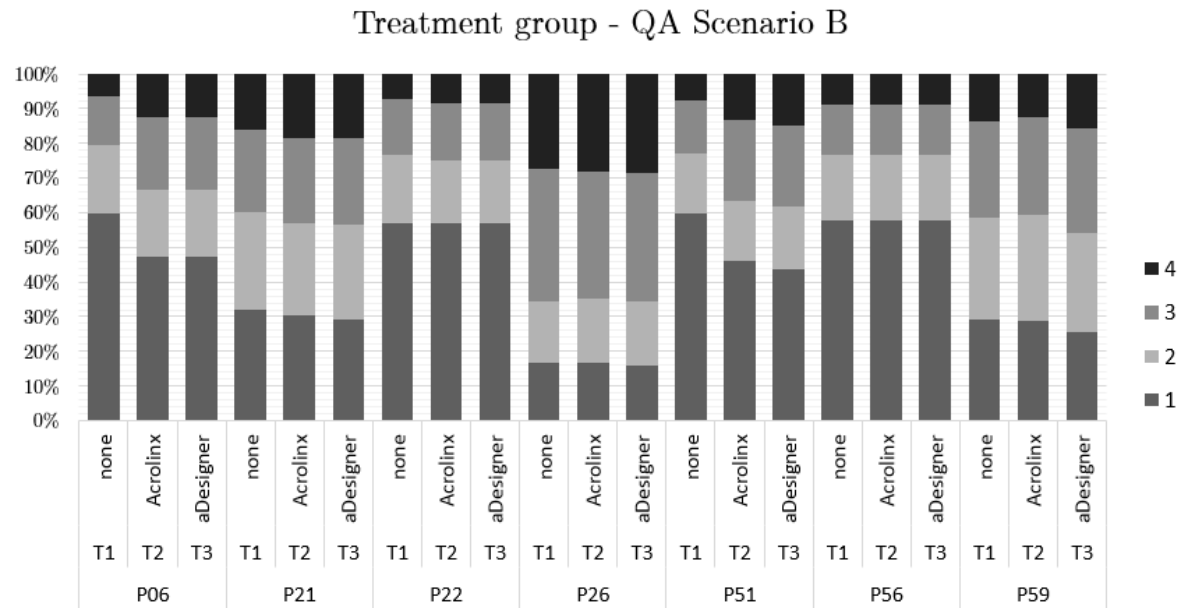


Figure K.3. Distribution of scores per participant by translation version: Treatment group (QA Scenario B)

Table K.7. GLM output predicting the interaction effect of three independent variables (WA knowledge, tool used and QA scenario followed) on the alt texts score

Group	Translation condition	Estimate*	Std. Error	*p*-value†
control	Both-last-aDesigner (T3)	1.505	0.037	*<0.001*
treatment	Both-last-aDesigner (T3)	1.14	0.051	*<0.001*
control	Acrolinx (T2)	1.101	0.037	*<0.001*
treatment	Acrolinx (T2)	0.839	0.052	*<0.001*
treatment	none (T1)	0.589	0.236	*0.012*
control	Both-last-Acrolinx (T3)	0.545	0.037	*<0.001*
control	aDesigner (T2)	0.116	0.037	*0.002*
treatment	aDesigner (T2)	0.005	0.053	0.921
treatment	Both-last-Acrolinx (T3)	0.002	0.052	0.968

*Data organised in descending order, from highest to lowest estimate. / † Significant p-values shown in italics.

Table K.8. Score mean, standard deviation and proportions (%) by rule in the pairwise comparison of translation conditions T2 (Acrolinx) - T1 (none)

Rule	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDA1-credits	Acrolinx	2.12	1.09	394	40.20	210	21.43	237	24.18	139	14.18
	none	2.19	1.12	382	38.98	187	19.08	257	26.22	154	15.71
DDR2-image	Acrolinx	2.17	1.09	367	37.45	235	23.98	227	23.16	151	15.41
	none	2.12	1.08	378	38.57	253	25.82	205	20.92	144	14.69
DDR3-logo	Acrolinx	1.94	1.02	444	45.31	252	25.71	187	19.08	97	9.90
	none	1.62	0.86	571	58.27	257	26.22	104	10.61	48	4.90
FAA1-files	Acrolinx	2.44	1.07	244	24.90	257	26.22	287	29.29	192	19.59
	none	2.13	1.02	347	35.41	264	26.94	264	26.94	105	10.71
FAA3-social	Acrolinx	2.46	1.09	249	25.41	240	24.49	279	28.47	212	21.63
	none	1.75	0.82	450	45.92	351	35.82	150	15.31	29	2.96
FAA4-print	Acrolinx	2.55	1.06	207	21.12	247	25.20	307	31.33	219	22.35
	none	2.17	1.01	324	33.06	275	28.06	272	27.76	109	11.12
FAC1-accueil	Acrolinx	2.17	1.11	376	38.37	215	21.94	231	23.57	158	16.12
	none	1.64	0.92	591	60.31	211	21.53	118	12.04	60	6.12
FDA3-structure	Acrolinx	2.15	1.18	454	46.33	90	9.18	268	27.35	168	17.14
	none	1.66	0.98	615	62.76	165	16.84	121	12.35	79	8.06
U2-placeholder	Acrolinx	1.72	0.98	578	58.98	163	16.63	173	17.65	66	6.73
	none	1.36	0.78	780	79.59	86	8.78	79	8.06	35	3.57
U2-decor	Acrolinx	1.79	0.94	494	50.41	250	25.51	180	18.37	56	5.71
	none	1.54	0.84	633	64.59	207	21.12	100	10.20	40	4.08

Table K.9. Score mean, standard deviation and proportions (%) by image for the rule DDA1-credits in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDA1-credits	**I110**	Acrolinx	2.37	1.21	35	35.71	17	17.35	21	21.43	25	25.51
		none	2.52	1.27	33	33.67	14	14.29	18	18.37	33	33.67
	I111	Acrolinx	1.98	0.92	38	38.78	28	28.57	28	28.57	4	4.08
		none	1.99	0.98	42	42.86	20	20.41	31	31.63	5	5.10
	I113	Acrolinx	2.37	1.19	35	35.71	14	14.29	27	27.55	22	22.45
		none	2.35	1.23	38	38.78	12	12.24	24	24.49	24	24.49
	I114	Acrolinx	2.46	1.23	32	32.65	18	18.37	19	19.39	29	29.59
		none	2.46	1.29	35	35.71	16	16.33	14	14.29	33	33.67
	I130	Acrolinx	1.98	1.06	46	46.94	18	18.37	24	24.49	10	10.20
		none	2.24	1.10	35	35.71	19	19.39	29	29.59	15	15.31
	I17	Acrolinx	2.36	1.12	32	32.65	17	17.35	31	31.63	18	18.37
		none	2.43	1.06	26	26.53	21	21.43	34	34.69	17	17.35
	I23	Acrolinx	2.03	1.02	38	38.78	30	30.61	19	19.39	11	11.22
		none	2.21	0.96	29	29.59	26	26.53	36	36.73	7	7.14
	I25	Acrolinx	2.08	1.08	39	39.80	26	26.53	19	19.39	14	14.29
		none	2.17	1.10	37	37.76	22	22.45	24	24.49	15	15.31
	I63	Acrolinx	1.93	0.92	42	42.86	24	24.49	29	29.59	3	3.06
		none	1.91	0.93	44	44.90	22	22.45	29	29.59	3	3.06
	I71	Acrolinx	1.68	0.90	57	58.16	18	18.37	20	20.41	3	3.06
		none	1.58	0.86	63	64.29	15	15.31	18	18.37	2	2.04

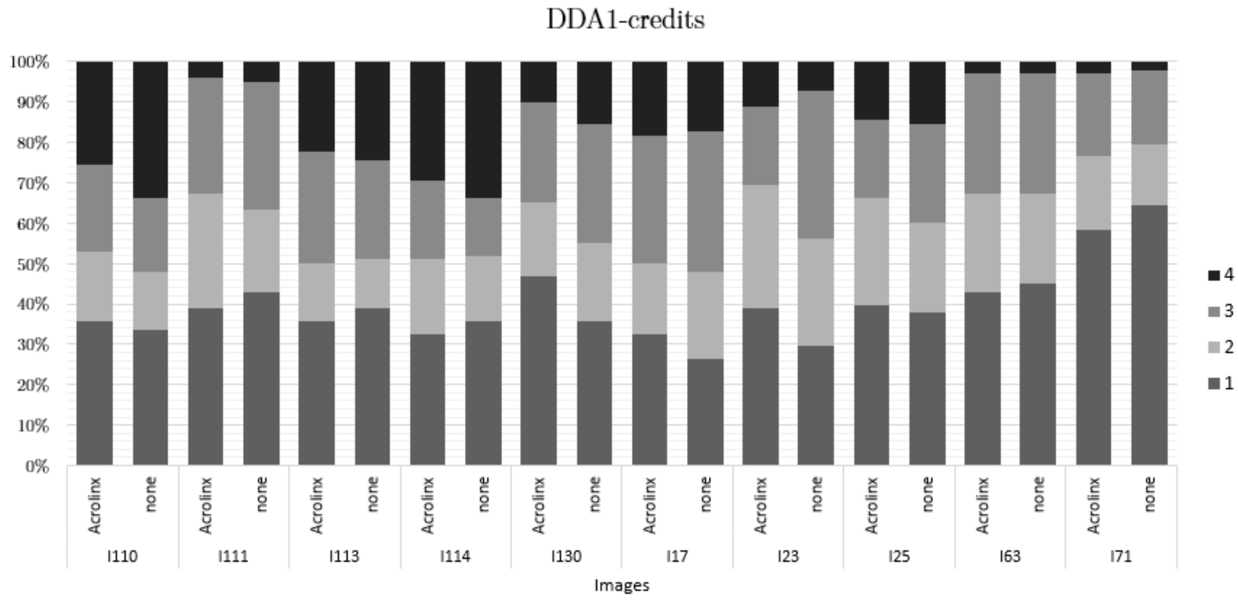


Figure K.4. Distribution of scores by image for the rule DDA1-credits: T2 (Acrolinx) - T1 (none)

Table K.10. Score mean, standard deviation and proportions (%) by image for the rule DDR2-image the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDR2-image	**I122**	Acrolinx	2.14	1.04	33	33.67	31	31.63	21	21.43	13	13.27
		none	2.17	1.04	32	32.65	30	30.61	23	23.47	13	13.27
	I19	Acrolinx	2.19	1.10	37	37.76	19	19.39	28	28.57	14	14.29
		none	2.19	1.09	36	36.73	21	21.43	27	27.55	14	14.29
	I22	Acrolinx	2.27	1.19	39	39.80	14	14.29	25	25.51	20	20.41
		none	2.33	1.21	38	38.78	12	12.24	26	26.53	22	22.45
	I26	Acrolinx	2.05	0.98	36	36.73	29	29.59	25	25.51	8	8.16
		none	1.95	0.88	36	36.73	35	35.71	23	23.47	4	4.08
	I62	Acrolinx	2.16	1.12	36	36.73	28	28.57	16	16.33	18	18.37
		none	2.02	1.05	38	38.78	34	34.69	12	12.24	14	14.29
	I65	Acrolinx	2.12	1.10	38	38.78	25	25.51	20	20.41	15	15.31
		none	2.03	1.06	39	39.80	31	31.63	14	14.29	14	14.29
	I66	Acrolinx	2.56	1.20	28	28.57	17	17.35	23	23.47	30	30.61
		none	2.45	1.21	30	30.61	23	23.47	16	16.33	29	29.59
	I67	Acrolinx	2.29	1.05	27	27.55	32	32.65	23	23.47	16	16.33
		none	2.30	1.08	29	29.59	28	28.57	24	24.49	17	17.35
	I84	Acrolinx	1.87	0.94	46	46.94	23	23.47	25	25.51	4	4.08
		none	1.71	0.85	51	52.04	26	26.53	19	19.39	2	2.04
	I88	Acrolinx	2.00	1.11	47	47.96	17	17.35	21	21.43	13	13.27
		none	2.02	1.16	49	50.00	13	13.27	21	21.43	15	15.31

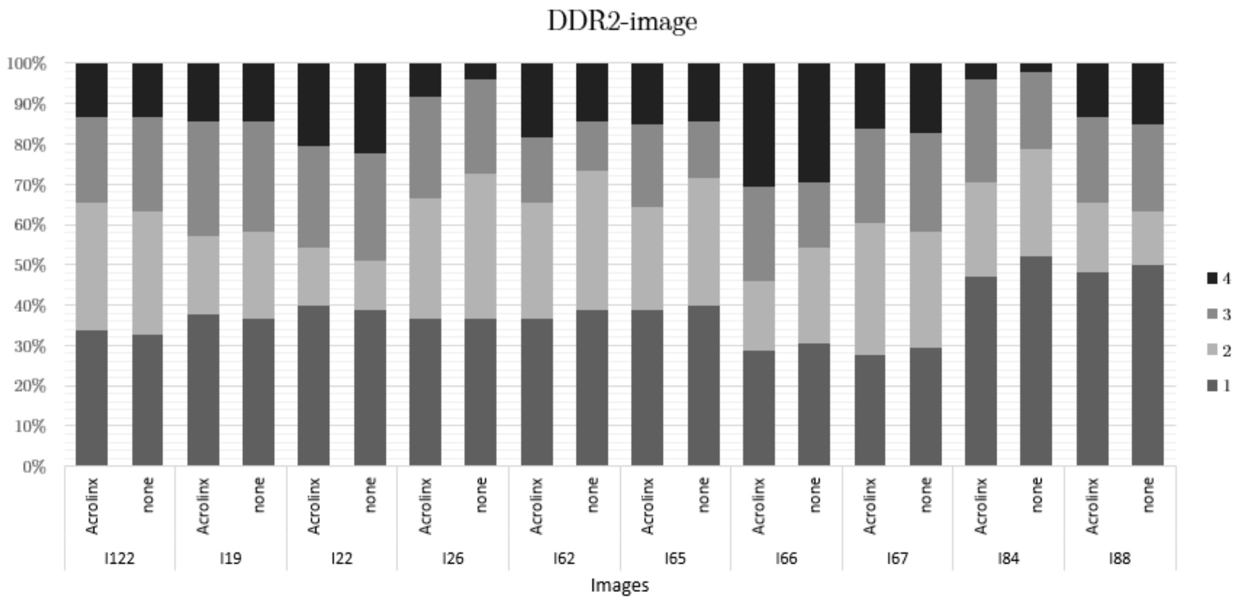


Figure K.5. Distribution of scores by image for the rule DDR2-image: T2 (Acrolinx) - T1 (none)

Table K.11. Score mean, standard deviation and proportions (%) by image for the rule DDR3-logo in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDR3-logo	**I10**	Acrolinx	2.09	0.93	31	31.63	34	34.69	26	26.53	7	7.14
		none	1.84	0.74	33	33.67	51	52.04	11	11.22	3	3.06
	I102	Acrolinx	2.27	1.02	27	27.55	32	32.65	25	25.51	14	14.29
		none	2.15	1.11	37	37.76	25	25.51	20	20.41	16	16.33
	I12	Acrolinx	1.85	1.01	51	52.04	18	18.37	22	22.45	7	7.14
		none	1.55	0.83	62	63.27	21	21.43	12	12.24	3	3.06
	I125	Acrolinx	1.78	0.95	50	51.02	27	27.55	14	14.29	7	7.14
		none	1.50	0.83	65	66.33	22	22.45	6	6.12	5	5.10
	I45	Acrolinx	1.48	0.76	63	64.29	27	27.55	4	4.08	4	4.08
		none	1.43	0.86	74	75.51	12	12.24	6	6.12	6	6.12
	I47	Acrolinx	2.53	1.01	18	18.37	29	29.59	32	32.65	19	19.39
		none	1.86	0.85	38	38.78	41	41.84	14	14.29	5	5.10
	I54	Acrolinx	1.88	1.08	51	52.04	20	20.41	15	15.31	12	12.24
		none	1.37	0.78	77	78.57	9	9.18	9	9.18	3	3.06
	I72	Acrolinx	1.69	0.89	54	55.10	24	24.49	16	16.33	4	4.08
		none	1.55	0.79	60	61.22	24	24.49	12	12.24	2	2.04
	I78	Acrolinx	1.56	0.99	71	72.45	6	6.12	14	14.29	7	7.14
		none	1.36	0.82	80	81.63	5	5.10	9	9.18	4	4.08
	I98	Acrolinx	2.23	1.04	28	28.57	35	35.71	19	19.39	16	16.33
		none	1.61	0.64	45	45.92	47	47.96	5	5.10	1	1.02

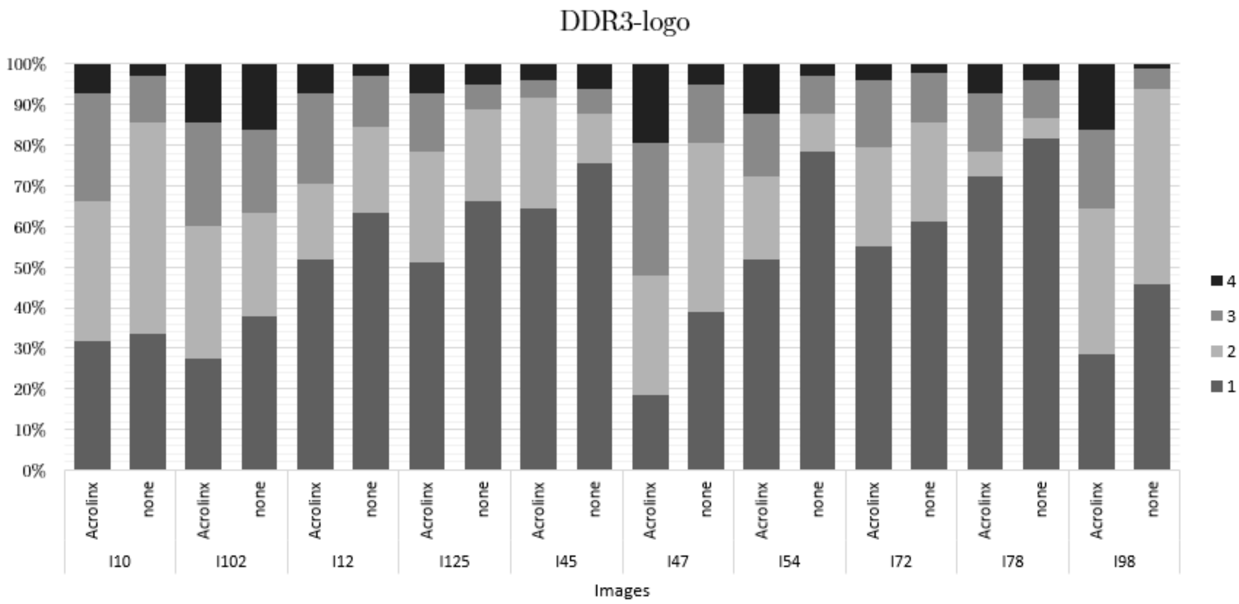


Figure K.6. Distribution of scores by image for the rule DDR3-logo: T2 (Acrolinx) - T1 (none)

Table K.12. Score mean, standard deviation and proportions (%) by image for the rule FAA1-files in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAA1-files	**I115**	Acrolinx	2.47	1.06	23	23.47	25	25.51	31	31.63	19	19.39
		none	2.30	1.03	27	27.55	29	29.59	28	28.57	14	14.29
	I116	Acrolinx	2.45	1.10	27	27.55	20	20.41	31	31.63	20	20.41
		none	2.23	1.10	35	35.71	20	20.41	28	28.57	15	15.31
	I117	Acrolinx	2.55	1.05	20	20.41	25	25.51	32	32.65	21	21.43
		none	2.43	1.10	27	27.55	22	22.45	29	29.59	20	20.41
	I119	Acrolinx	2.24	1.07	32	32.65	24	24.49	28	28.57	14	14.29
		none	1.74	0.92	51	52.04	27	27.55	14	14.29	6	6.12
	I28	Acrolinx	2.70	1.14	22	22.45	16	16.33	29	29.59	31	31.63
		none	2.39	0.99	25	25.51	21	21.43	41	41.84	11	11.22
	I31	Acrolinx	2.37	1.08	26	26.53	29	29.59	24	24.49	19	19.39
		none	1.99	0.91	37	37.76	29	29.59	28	28.57	4	4.08
	I34	Acrolinx	2.30	1.09	30	30.61	26	26.53	25	25.51	17	17.35
		none	1.79	0.96	53	54.08	17	17.35	24	24.49	4	4.08
	I73	Acrolinx	2.33	1.09	27	27.55	32	32.65	19	19.39	20	20.41
		none	2.04	1.01	36	36.73	34	34.69	16	16.33	12	12.24
	I76	Acrolinx	2.47	0.95	15	15.31	39	39.8	27	27.55	17	17.35
		none	2.29	0.94	21	21.43	40	40.82	25	25.51	12	12.24
	I79	Acrolinx	2.48	1.00	22	22.45	21	21.43	41	41.84	14	14.29
		none	2.10	0.98	35	35.71	25	25.51	31	31.63	7	7.14

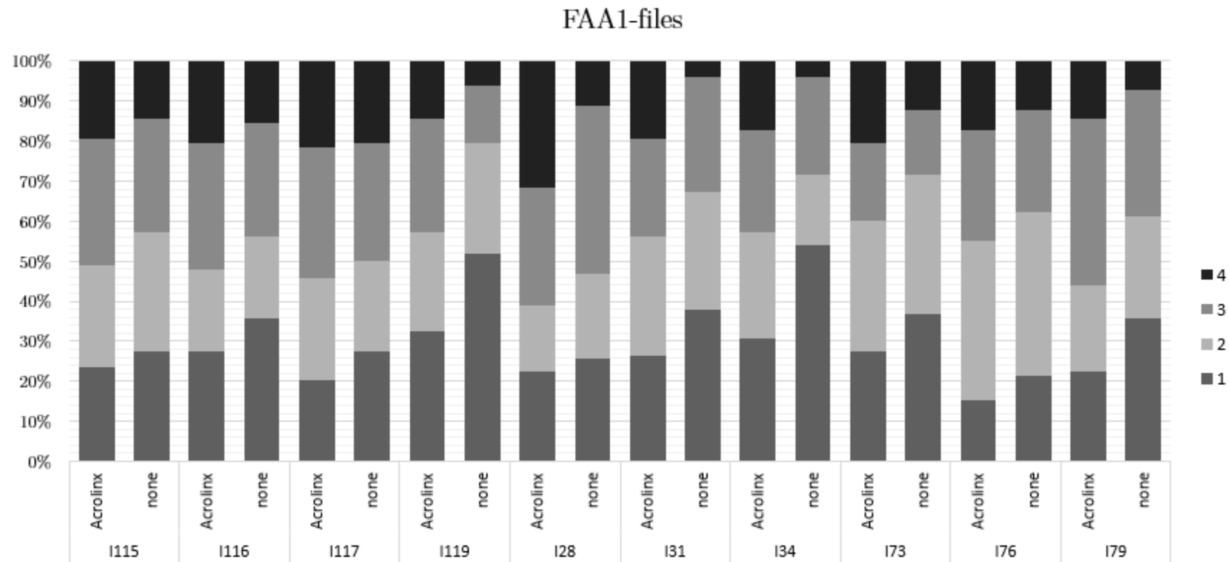


Figure K.7. Distribution of scores by image for the rule FAA1-files: T2 (Acrolinx) - T1 (none)

Table K.13. Score mean, standard deviation and proportions (%) by image for the rule FAA3-social in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAA3-social	**I2**	Acrolinx	2.88	0.96	8	8.16	27	27.55	32	32.65	31	31.63
		none	2.11	0.73	19	19.39	51	52.04	26	26.53	2	2.04
	I3	Acrolinx	2.54	1.19	27	27.55	20	20.41	22	22.45	29	29.59
		none	1.46	0.68	63	64.29	25	25.51	10	10.20	.	.
	I4	Acrolinx	2.52	1.08	23	23.47	22	22.45	32	32.65	21	21.43
		none	1.72	0.73	42	42.86	42	42.86	13	13.27	1	1.02
	I49	Acrolinx	2.14	1.05	33	33.67	32	32.65	19	19.39	14	14.29
		none	1.63	0.79	53	54.08	30	30.61	13	13.27	2	2.04
	I5	Acrolinx	2.60	0.97	17	17.35	22	22.45	42	42.86	17	17.35
		none	2.20	0.79	20	20.41	40	40.82	36	36.73	2	2.04
	I50	Acrolinx	2.57	1.15	24	24.49	22	22.45	24	24.49	28	28.57
		none	1.92	1.02	47	47.96	20	20.41	23	23.47	8	8.16
	I51	Acrolinx	2.53	1.04	22	22.45	20	20.41	38	38.78	18	18.37
		none	1.83	0.83	40	40.82	38	38.78	17	17.35	3	3.06
	I6	Acrolinx	2.36	1.00	23	23.47	31	31.63	30	30.61	14	14.29
		none	1.53	0.63	52	53.06	41	41.84	4	4.08	1	1.02
	I94	Acrolinx	2.30	1.21	37	37.76	19	19.39	18	18.37	24	24.49
		none	1.64	0.85	53	54.08	33	33.67	6	6.12	6	6.12
	I95	Acrolinx	2.19	1.10	35	35.71	25	25.51	22	22.45	16	16.33
		none	1.48	0.74	61	62.24	31	31.63	2	2.04	4	4.08

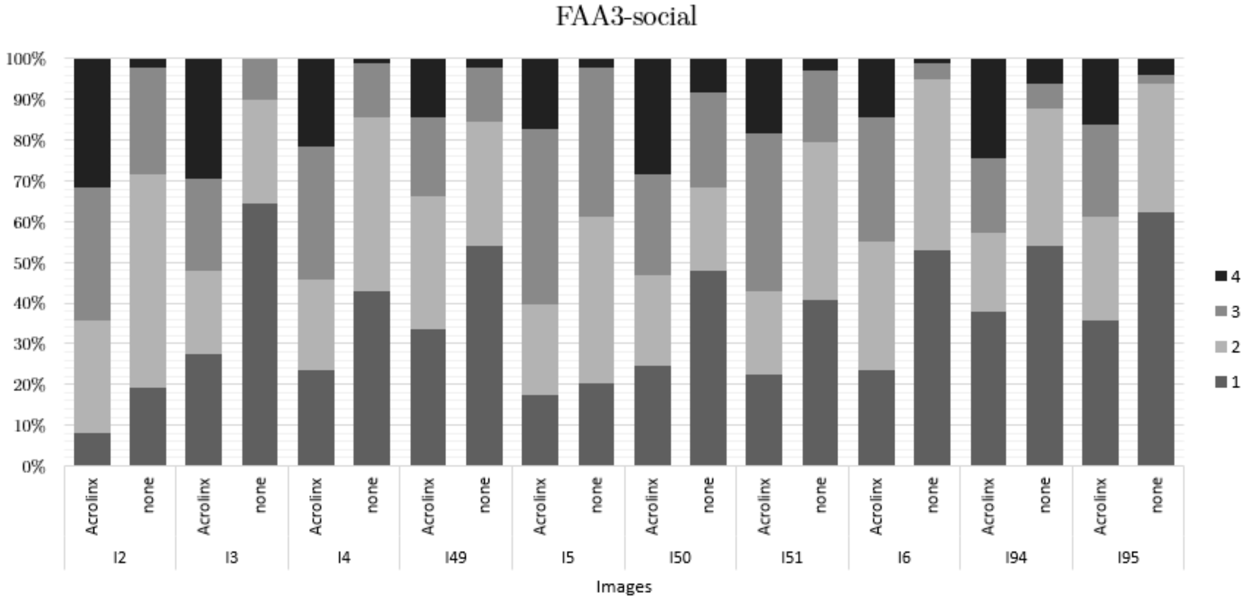


Figure K.8. Distribution of scores by image for the rule FAA3-social: T2 (Acrolinx) - T1 (none)

Table K.14. Score mean, standard deviation and proportions (%) by image for the rule FAA4-print in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAA4-print	**I107**	Acrolinx	2.45	1.08	25	25.51	23	23.47	31	31.63	19	19.39
		none	1.93	0.92	40	40.82	30	30.61	23	23.47	5	5.10
	I120	Acrolinx	2.35	0.92	20	20.41	34	34.69	34	34.69	10	10.20
		none	2.01	0.94	35	35.71	34	34.69	22	22.45	7	7.14
	I14	Acrolinx	3.14	1.04	11	11.22	13	13.27	25	25.51	49	50.00
		none	2.43	1.13	28	28.57	22	22.45	26	26.53	22	22.45
	I29	Acrolinx	2.37	1.13	29	29.59	26	26.53	21	21.43	22	22.45
		none	1.91	1.00	45	45.92	25	25.51	20	20.41	8	8.16
	I32	Acrolinx	2.47	1.07	21	21.43	32	32.65	23	23.47	22	22.45
		none	2.14	0.97	29	29.59	37	37.76	21	21.43	11	11.22
	I35	Acrolinx	2.38	1.09	30	30.61	17	17.35	35	35.71	16	16.33
		none	2.01	0.99	42	42.86	18	18.37	33	33.67	5	5.10
	I59	Acrolinx	2.63	1.07	18	18.37	26	26.53	28	28.57	26	26.53
		none	2.08	0.86	32	32.65	26	26.53	40	40.82	.	.
	I74	Acrolinx	2.70	1.07	19	19.39	17	17.35	36	36.73	26	26.53
		none	2.56	1.17	26	26.53	19	19.39	25	25.51	28	28.57
	I77	Acrolinx	2.54	0.98	17	17.35	28	28.57	36	36.73	17	17.35
		none	2.39	0.98	23	23.47	26	26.53	37	37.76	12	12.24
	I80	Acrolinx	2.46	0.92	17	17.35	31	31.63	38	38.78	12	12.24
		none	2.23	0.95	24	24.49	38	38.78	25	25.51	11	11.22

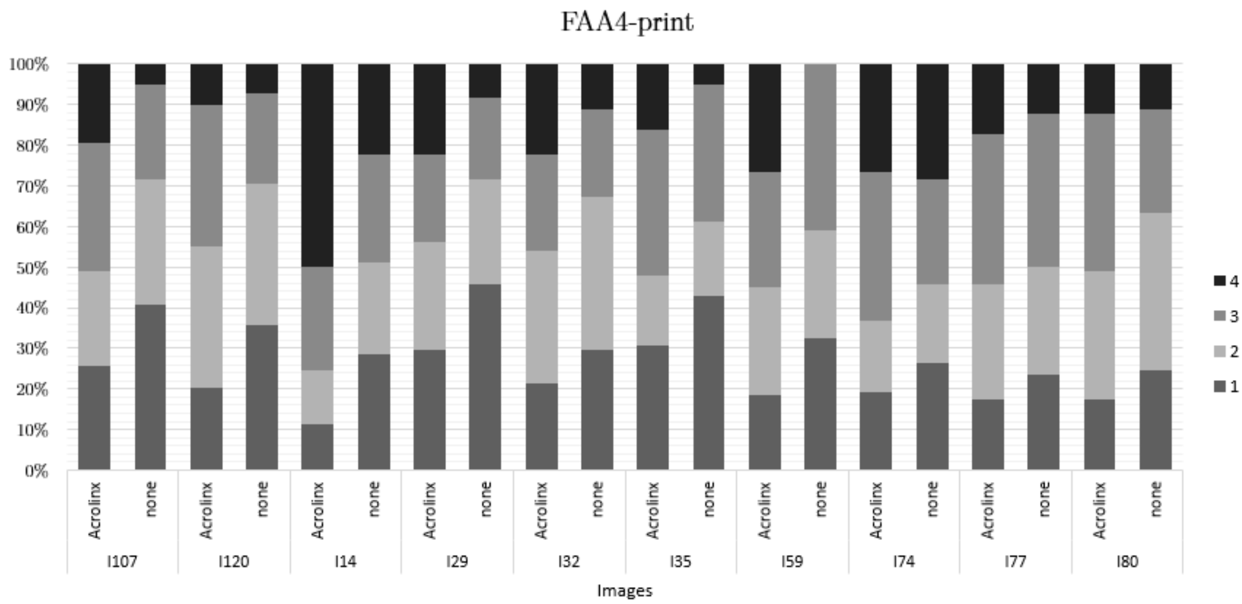


Figure K.9. Distribution of scores by image for the rule FAA4-print: T2 (Acrolinx) - T1 (none)

Table K.15. Score mean, standard deviation and proportions (%) by image for the rule FAC1-accueil in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAC1-accueil	**I103**	Acrolinx	2.24	1.11	36	36.73	17	17.35	30	30.61	15	15.31
		none	2.00	1.03	41	41.84	26	26.53	21	21.43	10	10.20
	I106	Acrolinx	2.12	1.16	44	44.9	14	14.29	24	24.49	16	16.33
		none	1.55	0.98	69	70.41	13	13.27	7	7.14	9	9.18
	I52	Acrolinx	2.22	1.16	37	37.76	22	22.45	19	19.39	20	20.41
		none	1.45	0.90	75	76.53	8	8.16	9	9.18	6	6.12
	I53	Acrolinx	2.06	1.02	39	39.80	23	23.47	27	27.55	9	9.18
		none	1.67	0.94	58	59.18	20	20.41	14	14.29	6	6.12
	I55	Acrolinx	2.16	1.13	38	38.78	23	23.47	20	20.41	17	17.35
		none	1.64	0.96	61	62.24	18	18.37	12	12.24	7	7.14
	I57	Acrolinx	2.33	1.21	37	37.76	15	15.31	23	23.47	23	23.47
		none	2.04	1.07	41	41.84	25	25.51	19	19.39	13	13.27
	I8	Acrolinx	2.01	0.94	34	34.69	37	37.76	19	19.39	8	8.16
		none	1.62	0.71	48	48.98	41	41.84	7	7.14	2	2.04
	I9	Acrolinx	2.08	0.97	33	33.67	33	33.67	23	23.47	9	9.18
		none	1.52	0.75	60	61.22	27	27.55	9	9.18	2	2.04
	I92	Acrolinx	2.24	1.16	37	37.76	19	19.39	23	23.47	19	19.39
		none	1.58	0.84	60	61.22	22	22.45	13	13.27	3	3.06
	I93	Acrolinx	2.27	1.22	41	41.84	12	12.24	23	23.47	22	22.45
		none	1.32	0.70	78	79.59	11	11.22	7	7.14	2	2.04

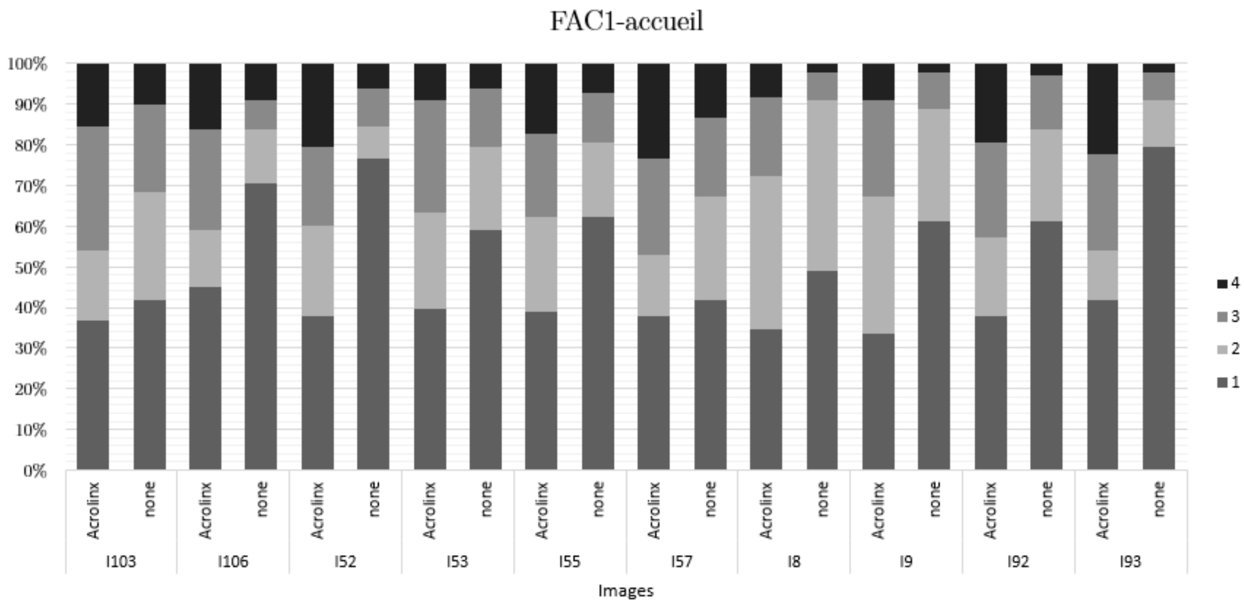


Figure K.10. Distribution of scores by image for the rule FAC1-accueil: T2 (Acrolinx) - T1 (none)

Table K.16. Score mean, standard deviation and proportions (%) by image for the rule FDA3-structure in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FDA3-structure	**I105**	Acrolinx	1.97	1.21	57	58.16	3	3.06	22	22.45	16	16.33
		none	1.51	0.97	74	75.51	5	5.10	12	12.24	7	7.14
	I11	Acrolinx	2.13	1.22	48	48.98	7	7.14	25	25.51	18	18.37
		none	1.87	1.06	51	52.04	20	20.41	16	16.33	11	11.22
	I129	Acrolinx	1.88	1.23	62	63.27	4	4.08	14	14.29	18	18.37
		none	1.50	1.02	77	78.57	3	3.06	8	8.16	10	10.20
	I13	Acrolinx	2.26	1.21	43	43.88	6	6.12	30	30.61	19	19.39
		none	1.84	1.06	54	55.01	16	16.33	18	18.37	10	10.20
	I16	Acrolinx	2.39	1.27	42	42.86	.	.	32	32.65	24	24.49
		none	1.43	0.94	80	81.63	1	1.02	10	10.20	7	7.14
	I44	Acrolinx	2.12	1.16	46	46.94	8	8.16	30	30.61	14	14.29
		none	1.57	0.92	65	66.33	16	16.33	11	11.22	6	6.12
	I56	Acrolinx	2.22	1.12	38	38.78	14	14.29	32	32.65	14	14.29
		none	1.77	0.95	51	52.04	26	26.53	14	14.29	7	7.14
	I58	Acrolinx	2.22	1.18	37	37.76	24	24.49	15	15.31	22	22.45
		none	1.80	0.95	46	46.94	36	36.73	6	6.12	10	10.20
	I60	Acrolinx	2.18	1.22	43	43.88	16	16.33	17	17.35	22	22.45
		none	1.71	0.99	56	57.14	24	24.49	8	8.16	10	10.20
	I89	Acrolinx	2.15	0.97	38	38.78	8	8.16	51	52.04	1	1.020
		none	1.58	0.82	61	62.24	18	18.37	18	18.37	1	1.020

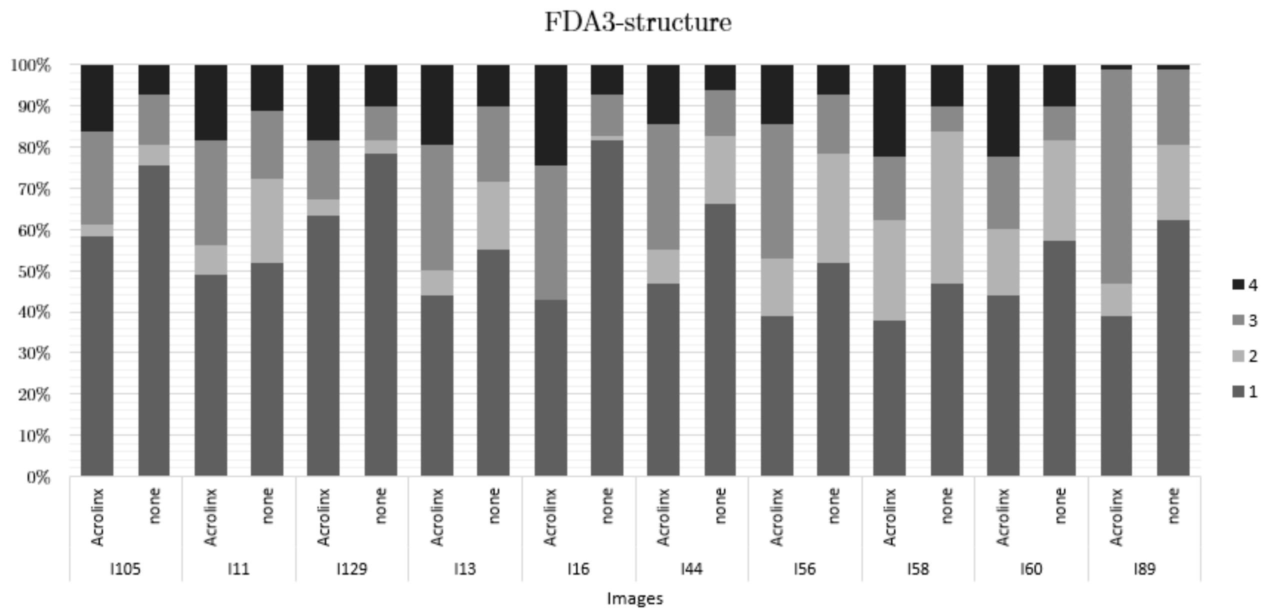


Figure K.11. Distribution of scores by image for the rule FDA3-structure: T2 (Acrolinx) - T1 (none)

Table K.17. Score mean, standard deviation and proportions (%) by image for the rule U2-placeholder in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
U2-placeholder	**I109**	Acrolinx	1.91	1.17	59	60.2	1	1.02	26	26.53	12	12.24
		none	1.42	0.92	80	81.63	1	1.02	11	11.22	6	6.12
	I112	Acrolinx	1.77	1.17	67	68.37	1	1.02	16	16.33	14	14.29
		none	1.42	0.95	81	82.65	1	1.02	8	8.16	8	8.16
	I18	Acrolinx	1.57	0.81	62	63.27	16	16.33	20	20.41	.	.
		none	1.28	0.62	80	81.63	9	9.18	9	9.18	.	.
	I24	Acrolinx	1.89	1.11	55	56.12	10	10.20	22	22.45	11	11.22
		none	1.30	0.75	83	84.69	4	4.08	8	8.16	3	3.06
	I41	Acrolinx	1.59	0.73	53	54.08	33	33.67	11	11.22	1	1.02
		none	1.31	0.72	80	81.63	9	9.18	6	6.12	3	3.06
	I43	Acrolinx	1.49	0.72	62	63.27	25	25.51	10	10.20	1	1.02
		none	1.21	0.56	83	84.69	10	10.20	4	4.08	1	1.02
	I48	Acrolinx	2.07	1.10	43	43.88	17	17.35	26	26.53	12	12.24
		none	1.39	0.74	74	75.51	11	11.22	12	12.24	1	1.02
	I75	Acrolinx	1.46	0.86	72	73.47	12	12.24	9	9.18	5	5.10
		none	1.45	0.83	70	71.43	17	17.35	6	6.12	5	5.10
	I83	Acrolinx	1.67	0.85	53	54.08	27	27.55	15	15.31	3	3.06
		none	1.38	0.77	75	76.53	12	12.24	8	8.16	3	3.06
	I86	Acrolinx	1.80	0.98	52	53.06	21	21.43	18	18.37	7	7.14
		none	1.42	0.84	74	75.51	12	12.24	7	7.14	5	5.10

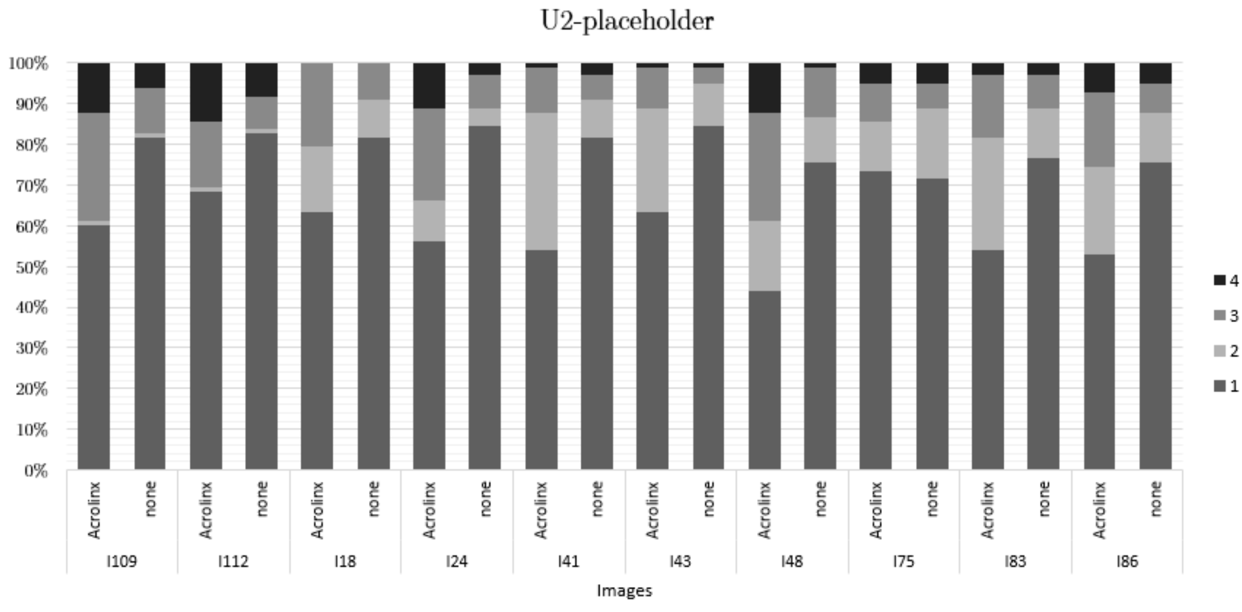


Figure K.12. Distribution of scores by image for the rule U2-placeholder: T2 (Acrolinx) - T1 (none)

Table K.18. Score mean, standard deviation and proportions (%) by image for the rule U4-decor in the pairwise comparison of translation conditions: T2 (Acrolinx) - T1 (none)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
U4 decor	**I1**	Acrolinx	2.07	1.00	38	38.78	22	22.45	31	31.63	7	7.14
		none	2.15	0.91	26	26.53	39	39.80	25	25.51	8	8.16
	I100	Acrolinx	2.12	1.18	46	46.94	10	10.20	26	26.53	16	16.33
		none	1.45	0.85	73	74.49	10	10.20	11	11.22	4	4.08
	I37	Acrolinx	1.60	0.80	57	58.16	24	24.49	16	16.33	1	1.02
		none	1.36	0.79	78	79.59	9	9.18	7	7.14	4	4.08
	I38	Acrolinx	1.55	0.71	55	56.12	33	33.67	9	9.18	1	1.02
		none	1.32	0.71	78	79.59	12	12.24	5	5.10	3	3.06
	I39	Acrolinx	1.59	0.73	54	55.10	30	30.61	14	14.29	.	.
		none	1.33	0.74	79	80.61	9	9.18	7	7.14	3	3.06
	I42	Acrolinx	1.64	0.84	54	55.10	29	29.59	11	11.22	4	4.08
		none	1.31	0.78	82	83.67	7	7.14	4	4.08	5	5.10
	I46	Acrolinx	1.96	1.09	49	50.00	15	15.31	23	23.47	11	11.22
		none	1.67	0.88	55	56.12	24	24.49	15	15.31	4	4.08
	I85	Acrolinx	2.02	0.82	27	27.55	47	47.96	19	19.39	5	5.10
		none	1.82	0.68	31	31.63	56	57.14	9	9.18	2	2.04
	I87	Acrolinx	1.72	0.91	52	53.06	26	26.53	15	15.31	5	5.10
		none	1.57	0.72	54	55.10	33	33.67	10	10.2	1	1.02
	I90	Acrolinx	1.65	0.96	62	63.27	14	14.29	16	16.33	6	6.12
		none	1.41	0.87	77	78.57	8	8.16	7	7.14	6	6.12

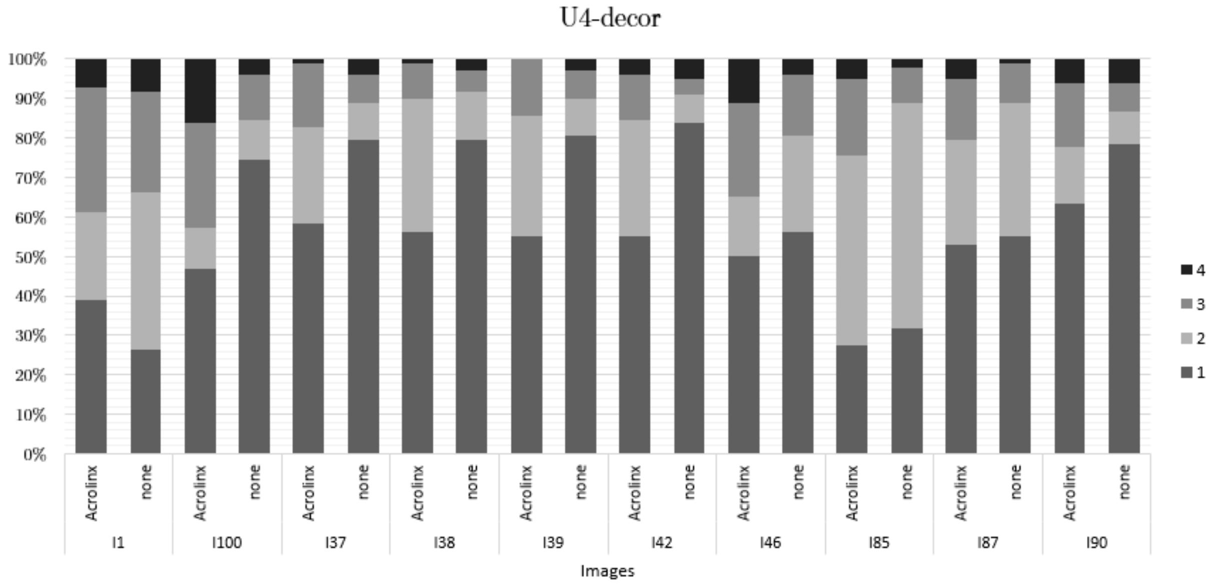


Figure K.13. Distribution of scores by image for the rule U4-decor: T2 (Acrolinx) - T1 (none)

Table K.19. Score mean, standard deviation and proportions (%) by rule in the pairwise comparison of translation conditions T3 (Both-last-Acrolinx) - T2 (aDesigner) [(4)]

Rule	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDA1-credits	aDesigner	2.10	1.14	425	43.37	188	19.18	207	21.12	160	16.33
	Acrolinx	2.07	1.12	432	44.08	192	19.59	209	21.33	147	15.00
DDR2-image	aDesigner	2.07	1.07	403	41.12	238	24.29	211	21.53	128	13.06
	Acrolinx	2.05	1.07	405	41.33	243	24.80	206	21.02	126	12.86
DDR3-logo	aDesigner	1.47	0.75	641	65.41	253	25.82	51	5.20	35	3.57
	Acrolinx	1.65	0.90	569	58.06	245	25.00	105	10.71	61	6.22
FAA1-files	aDesigner	1.92	0.96	424	43.27	280	28.57	205	20.92	71	7.24
	Acrolinx	2.00	0.97	385	39.29	294	30.00	221	22.55	80	8.16
FAA3-social	aDesigner	1.74	0.80	450	45.92	351	35.82	160	16.33	19	1.94
	Acrolinx	2.27	1.06	308	31.43	250	25.51	274	27.96	148	15.10
FAA4-print	aDesigner	2.19	1.02	322	32.86	263	26.84	280	28.57	115	11.73
	Acrolinx	2.42	1.06	253	25.82	239	24.39	315	32.14	173	17.65
FAC1-accueil	aDesigner	1.39	0.69	697	71.12	197	20.10	72	7.35	14	1.43
	Acrolinx	1.63	0.91	601	61.33	193	19.69	132	13.47	54	5.51
FDA3-structure	aDesigner	1.36	0.68	715	72.96	199	20.31	41	4.18	25	2.55
	Acrolinx	1.86	1.09	544	55.51	144	14.69	179	18.27	113	11.53
U2-placeholder	aDesigner	1.21	0.62	859	87.65	56	5.71	44	4.49	21	2.14
	Acrolinx	1.49	0.87	712	72.65	104	10.61	120	12.24	44	4.49
U2-decor	aDesigner	1.38	0.68	703	71.73	203	20.71	55	5.61	19	1.94
	Acrolinx	1.60	0.83	583	59.49	245	25.00	117	11.94	35	3.57

Table K.20. Score mean, standard deviation and proportions (%) by image for the rule DDA1-credits in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDA1-credits	**I110**	Acrolinx	2.49	1.29	33	33.67	20	20.41	9	9.18	36	36.73
		aDesigner	2.54	1.32	33	33.67	18	18.37	8	8.16	39	39.8
	I111	Acrolinx	1.82	0.93	50	51.02	19	19.39	26	26.53	3	3.06
		aDesigner	1.71	0.90	55	56.12	18	18.37	23	23.47	2	2.04
	I113	Acrolinx	2.21	1.18	39	39.80	19	19.39	20	20.41	20	20.41
		aDesigner	2.14	1.18	41	41.84	22	22.45	15	15.31	20	20.41
	I114	Acrolinx	2.27	1.30	45	45.92	9	9.18	17	17.35	27	27.55
		aDesigner	2.38	1.35	43	43.88	9	9.18	12	12.24	34	34.69
	I130	Acrolinx	2.11	1.11	42	42.86	16	16.33	27	27.55	13	13.27
		aDesigner	2.16	1.12	40	40.82	16	16.33	28	28.57	14	14.29
	I17	Acrolinx	2.47	1.05	24	24.49	21	21.43	36	36.73	17	17.35
		aDesigner	2.53	1.07	23	23.47	20	20.41	35	35.71	20	20.41
	I23	Acrolinx	1.98	0.97	39	39.80	30	30.61	21	21.43	8	8.16
		aDesigner	2.09	0.94	34	34.69	26	26.53	33	33.67	5	5.10
	I25	Acrolinx	2.26	1.17	36	36.73	22	22.45	19	19.39	21	21.43
		aDesigner	2.35	1.22	35	35.71	20	20.41	17	17.35	26	26.53
	I63	Acrolinx	1.71	0.87	53	54.08	22	22.45	21	21.43	2	2.04
		aDesigner	1.76	0.83	48	48.98	26	26.53	24	24.49	.	.
	I71	Acrolinx	1.41	0.72	71	72.45	14	14.29	13	13.27	.	.
		aDesigner	1.38	0.70	73	74.49	13	13.27	12	12.24	.	.

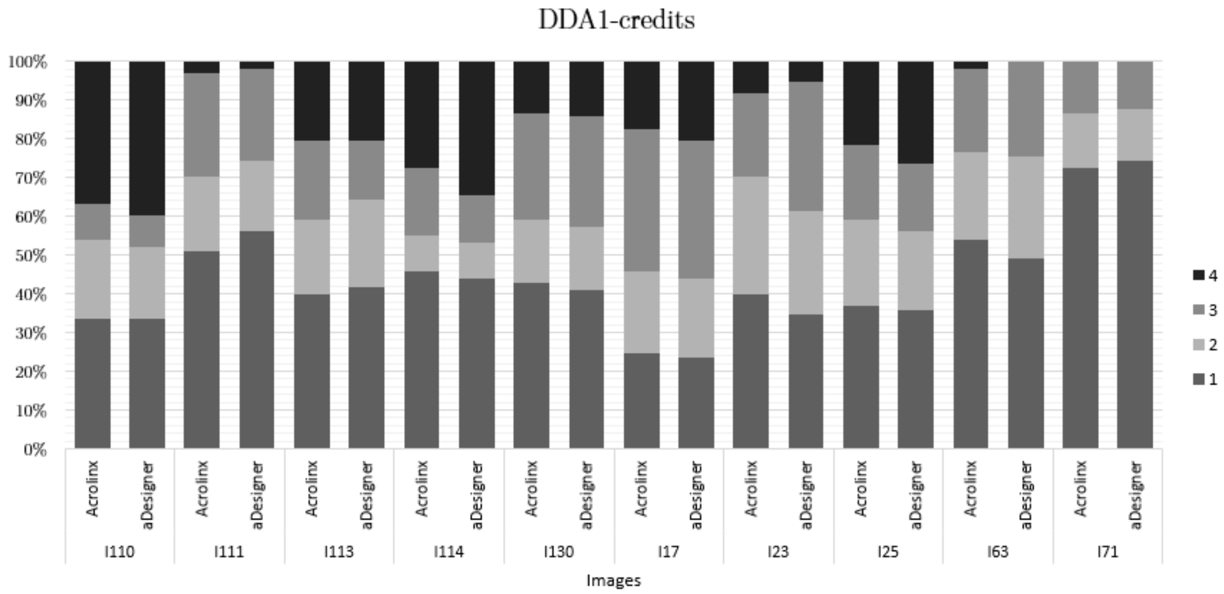


Figure K.14. Distribution of scores by image for the rule DDA1-credits: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.21. Score mean, standard deviation and proportions (%) by image for the rule DDR2-image in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDR2-image	**I122**	Acrolinx	2.04	0.96	36	36.73	29	29.59	26	26.53	7	7.14
		aDesigner	2.01	0.92	36	36.73	30	30.61	27	27.55	5	5.10
	I19	Acrolinx	2.07	1.09	42	42.86	19	19.39	25	25.51	12	12.24
		aDesigner	2.13	1.12	41	41.84	17	17.35	26	26.53	14	14.29
	I22	Acrolinx	2.31	1.20	36	36.73	20	20.41	18	18.37	24	24.49
		aDesigner	2.23	1.21	39	39.80	20	20.41	16	16.33	23	23.47
	I26	Acrolinx	1.96	0.90	38	38.78	29	29.59	28	28.57	3	3.06
		aDesigner	2.05	0.88	32	32.65	32	32.65	31	31.63	3	3.06
	I62	Acrolinx	1.87	0.95	43	43.88	33	33.67	14	14.29	8	8.16
		aDesigner	1.87	0.95	43	43.88	33	33.67	14	14.29	8	8.16
	I65	Acrolinx	2.03	1.00	38	38.78	28	28.57	23	23.47	9	9.18
		aDesigner	2.06	1.03	39	39.80	24	24.49	25	25.51	10	10.2
	I66	Acrolinx	2.58	1.23	27	27.55	22	22.45	14	14.29	35	35.71
		aDesigner	2.58	1.23	27	27.55	22	22.45	14	14.29	35	35.71
	I67	Acrolinx	2.35	1.07	26	26.53	30	30.61	24	24.49	18	18.37
		aDesigner	2.39	1.07	26	26.53	26	26.53	28	28.57	18	18.37
	I84	Acrolinx	1.48	0.69	62	63.27	25	25.51	11	11.22	.	.
		aDesigner	1.50	0.72	62	63.27	23	23.47	13	13.27	.	.
	I88	Acrolinx	1.86	1.10	57	58.16	8	8.16	23	23.47	10	10.20
		aDesigner	1.83	1.11	58	59.18	11	11.22	17	17.35	12	12.24

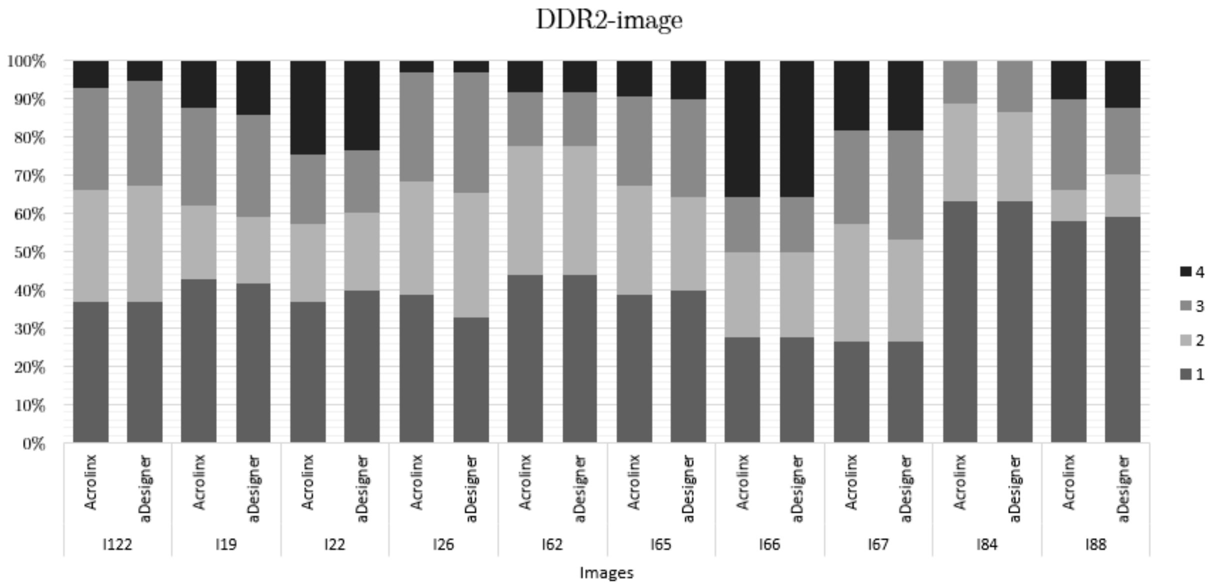


Figure K.15. Distribution of scores by image for the rule DDR2-image: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.22. Score mean, standard deviation and proportions (%) by image for the rule DDR3-logo in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
DDR3-logo	**I10**	Acrolinx	1.60	0.64	47	47.96	43	43.88	8	8.16	.	.
		aDesigner	1.68	0.60	38	38.78	53	54.08	7	7.14	.	.
	I102	Acrolinx	2.07	1.04	36	36.73	32	32.65	17	17.35	13	13.27
		aDesigner	1.94	1.08	47	47.96	23	23.47	15	15.31	13	13.27
	I12	Acrolinx	1.36	0.68	72	73.47	19	19.39	5	5.10	2	2.04
		aDesigner	1.35	0.54	67	68.37	28	28.57	3	3.06	.	.
	I125	Acrolinx	1.35	0.59	70	71.43	22	22.45	6	6.12	.	.
		aDesigner	1.17	0.38	81	82.65	17	17.35	.	.	.	.
	I45	Acrolinx	1.22	0.51	80	81.63	14	14.29	4	4.08	.	.
		aDesigner	1.14	0.45	87	88.78	9	9.18	1	1.02	1	1.02
	I47	Acrolinx	2.55	1.06	18	18.37	32	32.65	24	24.49	24	24.49
		aDesigner	2.16	1.05	30	30.61	39	39.8	12	12.24	17	17.35
	I54	Acrolinx	1.39	0.82	76	77.55	11	11.22	6	6.12	5	5.10
		aDesigner	1.10	0.42	91	92.86	5	5.10	1	1.02	1	1.02
	I72	Acrolinx	1.59	0.82	57	58.16	28	28.57	9	9.18	4	4.08
		aDesigner	1.46	0.75	66	67.35	21	21.43	9	9.18	2	2.04
	I78	Acrolinx	1.36	0.79	80	81.63	3	3.06	13	13.27	2	2.04
		aDesigner	1.09	0.43	93	94.9	2	2.04	2	2.04	1	1.02
	I98	Acrolinx	2.02	0.96	33	33.67	41	41.84	13	13.27	11	11.22
		aDesigner	1.59	0.51	41	41.84	56	57.14	1	1.02	.	.

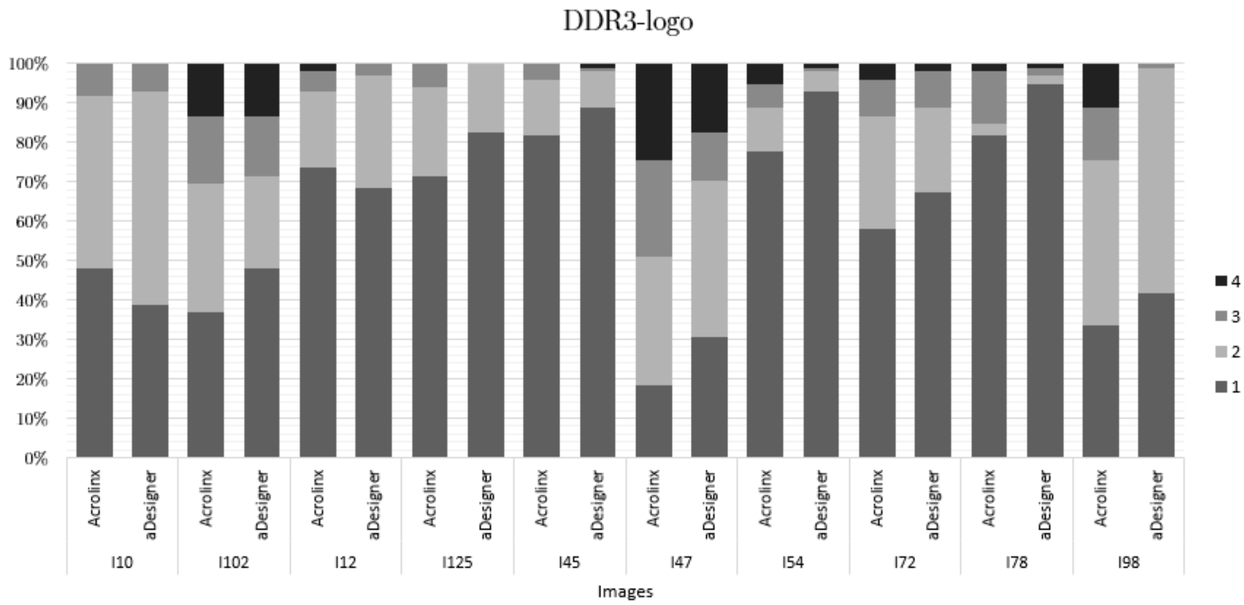


Figure K.16. Distribution of scores by image for the rule DDR3-logo: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K. 23. Score mean, standard deviation and proportions (%) by image for the rule FAA1-files in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAA1-files	**I115**	Acrolinx	2.00	0.96	37	37.76	32	32.65	21	21.43	8	8.16
		aDesigner	2.07	0.97	34	34.69	31	31.63	25	25.51	8	8.16
	I116	Acrolinx	2.00	0.98	39	39.8	28	28.57	23	23.47	8	8.16
		aDesigner	1.96	1.02	43	43.88	26	26.53	19	19.39	10	10.20
	I117	Acrolinx	2.32	1.09	30	30.61	24	24.49	27	27.55	17	17.35
		aDesigner	2.36	1.14	32	32.65	19	19.39	27	27.55	20	20.41
	I119	Acrolinx	1.54	0.73	57	58.16	31	31.63	8	8.16	2	2.04
		aDesigner	1.32	0.47	67	68.37	31	31.63	.	.	.	.
	I28	Acrolinx	2.37	1.00	26	26.53	21	21.43	40	40.82	11	11.22
		aDesigner	2.28	0.91	26	26.53	23	23.47	45	45.92	4	4.08
	I31	Acrolinx	1.94	0.86	34	34.69	41	41.84	18	18.37	5	5.10
		aDesigner	1.79	0.79	43	43.88	33	33.67	22	22.45	.	.
	I34	Acrolinx	1.64	0.85	56	57.14	24	24.49	15	15.31	3	3.06
		aDesigner	1.48	0.78	67	68.37	16	16.33	14	14.29	1	1.02
	I73	Acrolinx	2.02	0.95	35	35.71	34	34.69	21	21.43	8	8.16
		aDesigner	1.95	1.02	42	42.86	30	30.61	15	15.31	11	11.22
	I76	Acrolinx	2.21	0.98	26	26.53	37	37.76	23	23.47	12	12.24
		aDesigner	2.14	0.89	23	23.47	47	47.96	19	19.39	9	9.18
	I79	Acrolinx	1.92	0.98	45	45.92	22	22.45	25	25.51	6	6.12
		aDesigner	1.88	1.00	47	47.96	24	24.49	19	19.39	8	8.16

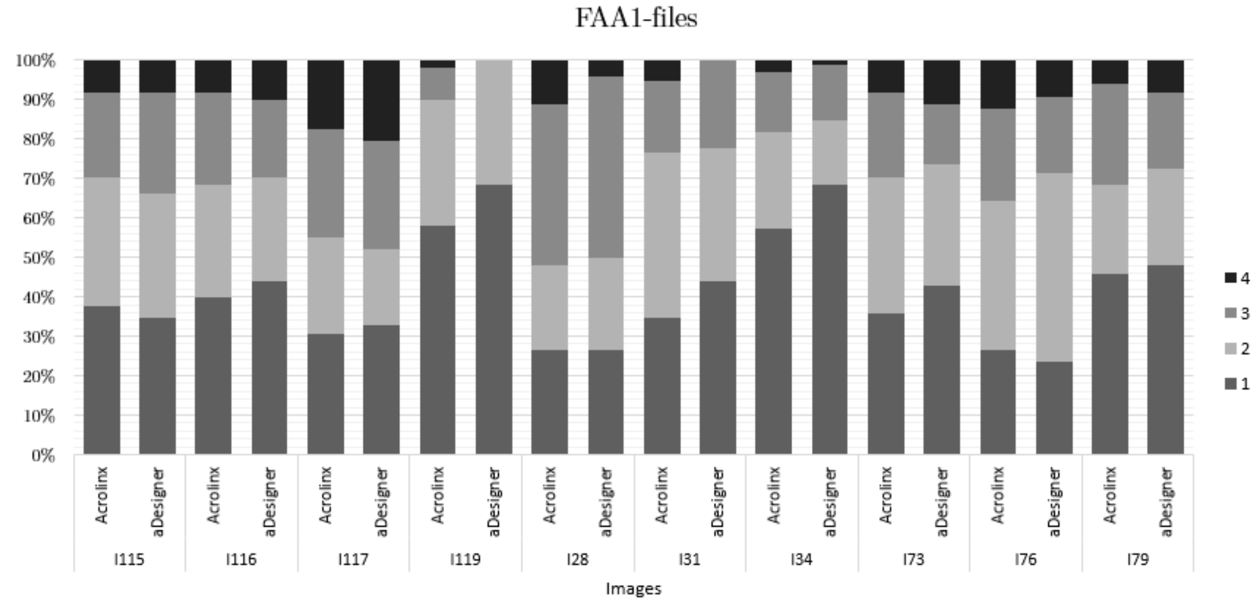


Figure K.17. Distribution of scores by image for the rule FAA1-files: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.24. Score mean, standard deviation and proportions (%) by image for the rule FAA3-social in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAA3-social	I2	Acrolinx	2.77	0.96	11	11.22	26	26.53	36	36.73	25	25.51
		aDesigner	2.19	0.76	16	16.33	51	52.04	27	27.55	4	4.08
	I3	Acrolinx	2.35	1.13	31	31.63	22	22.45	25	25.51	20	20.41
		aDesigner	1.63	0.89	58	59.18	23	23.47	12	12.24	5	5.10
	I4	Acrolinx	2.43	1.06	25	25.51	24	24.49	31	31.63	18	18.37
		aDesigner	1.82	0.82	40	40.82	39	39.8	16	16.33	3	3.06
	I49	Acrolinx	2.18	1.08	34	34.69	27	27.55	22	22.45	15	15.31
		aDesigner	1.63	0.82	55	56.12	26	26.53	15	15.31	2	2.04
	I5	Acrolinx	2.57	0.97	19	19.39	19	19.39	45	45.92	15	15.31
		aDesigner	2.24	0.77	19	19.39	37	37.76	41	41.84	1	1.02
	I50	Acrolinx	2.33	1.11	32	32.65	19	19.39	30	30.61	17	17.35
		aDesigner	1.79	0.88	49	50.00	22	22.45	26	26.53	1	1.02
	I51	Acrolinx	2.19	0.97	30	30.61	27	27.55	33	33.67	8	8.16
		aDesigner	1.78	0.73	39	39.8	42	42.86	17	17.35	.	.
	I6	Acrolinx	2.04	0.93	33	33.67	35	35.71	23	23.47	7	7.14
		aDesigner	1.55	0.63	50	51.02	43	43.88	4	4.08	1	1.02
	I94	Acrolinx	2.09	1.09	38	38.78	29	29.59	15	15.31	16	16.33
		aDesigner	1.47	0.61	56	57.14	40	40.82	.	.	2	2.04
	I95	Acrolinx	1.72	0.96	55	56.12	22	22.45	14	14.29	7	7.14
		aDesigner	1.33	0.51	68	69.39	28	28.57	2	2.04	.	.

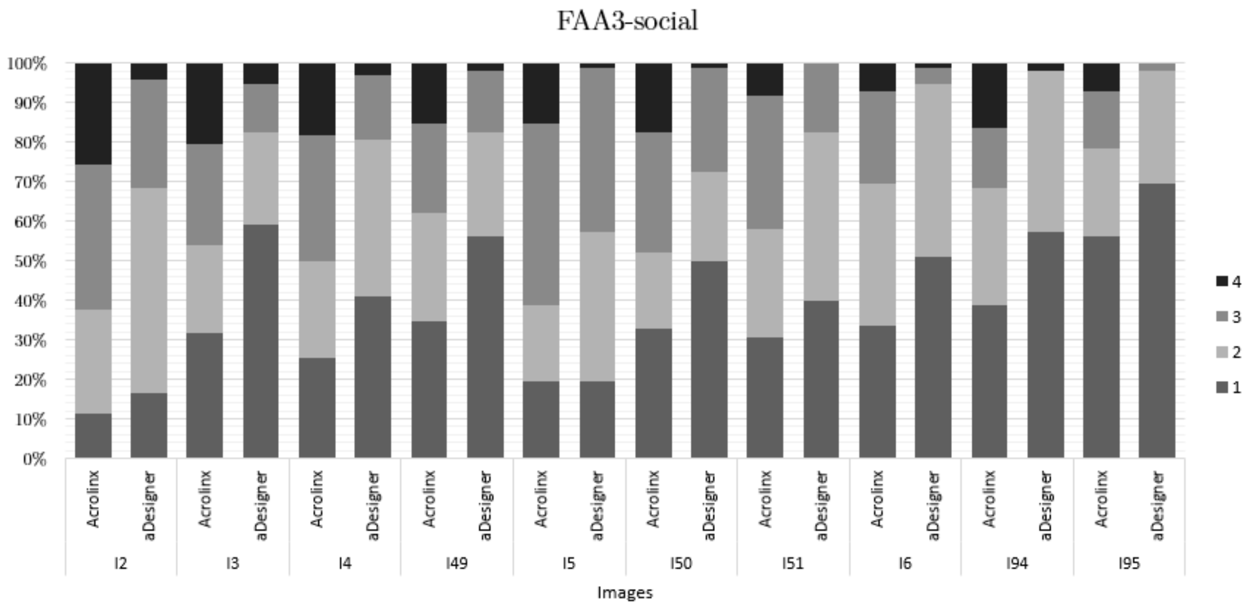


Figure K.18. Distribution of scores by image for the rule FAA3-social: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K. 25. Score mean, standard deviation and proportions (%) by image for the rule FAA4-print in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAA4-print	**I107**	Acrolinx	2.12	1.03	34	34.69	30	30.61	22	22.45	12	12.24
		aDesigner	1.94	0.97	41	41.84	30	30.61	19	19.39	8	8.16
	I120	Acrolinx	2.26	1.07	32	32.65	23	23.47	29	29.59	14	14.29
		aDesigner	2.01	1.01	40	40.82	26	26.53	23	23.47	9	9.18
	I14	Acrolinx	3.16	0.93	7	7.14	14	14.29	33	33.67	44	44.9
		aDesigner	3.09	0.93	8	8.16	14	14.29	37	37.76	39	39.8
	I29	Acrolinx	2.03	1.11	44	44.90	21	21.43	19	19.39	14	14.29
		aDesigner	1.66	0.88	55	56.12	26	26.53	12	12.24	5	5.10
	I32	Acrolinx	2.44	0.96	18	18.37	34	34.69	31	31.63	15	15.31
		aDesigner	1.99	0.82	31	31.63	39	39.80	26	26.53	2	2.04
	I35	Acrolinx	2.31	1.13	35	35.71	14	14.29	33	33.67	16	16.33
		aDesigner	1.89	0.93	46	46.94	19	19.39	31	31.63	2	2.04
	I59	Acrolinx	2.56	1.04	20	20.41	23	23.47	35	35.71	20	20.41
		aDesigner	2.22	0.91	29	29.59	21	21.43	45	45.92	3	3.06
	I74	Acrolinx	2.56	1.07	21	21.43	23	23.47	32	32.65	22	22.45
		aDesigner	2.59	1.18	26	26.53	18	18.37	24	24.49	30	30.61
	I77	Acrolinx	2.43	0.94	22	22.45	21	21.43	46	46.94	9	9.18
		aDesigner	2.36	0.93	23	23.47	25	25.51	42	42.86	8	8.16
	I80	Acrolinx	2.30	0.88	20	20.41	36	36.73	35	35.71	7	7.14
		aDesigner	2.16	0.89	23	23.47	45	45.92	21	21.43	9	9.18

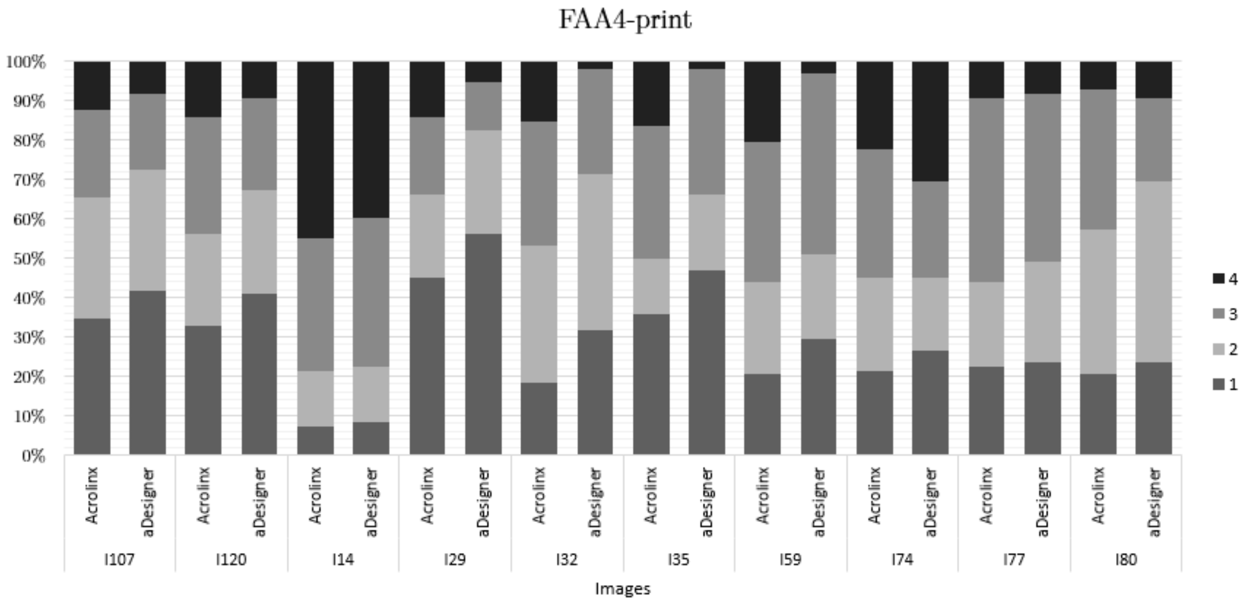


Figure K.19. Distribution of scores by image for the rule FAA4-print: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.26. Score mean, standard deviation and proportions (%) by image for the rule FAC1-accueil in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FAC1-accueil	**I103**	Acrolinx	2.04	1.07	42	42.86	22	22.45	22	22.45	12	12.24
		aDesigner	1.88	1.03	48	48.98	24	24.49	16	16.33	10	10.20
	I106	Acrolinx	1.39	0.77	75	76.53	10	10.2	11	11.22	2	2.04
		aDesigner	1.14	0.48	88	89.80	7	7.14	2	2.04	1	1.02
	I52	Acrolinx	1.66	1.07	67	68.37	8	8.16	12	12.24	11	11.22
		aDesigner	1.23	0.69	87	88.78	1	1.02	8	8.16	2	2.04
	I53	Acrolinx	1.54	0.81	64	65.31	16	16.33	17	17.35	1	1.02
		aDesigner	1.41	0.70	70	71.43	16	16.33	12	12.24	.	.
	I55	Acrolinx	1.47	0.78	66	67.35	21	21.43	8	8.16	3	3.06
		aDesigner	1.34	0.59	70	71.43	24	24.49	3	3.06	1	1.02
	I57	Acrolinx	1.64	0.84	55	56.12	26	26.53	14	14.29	3	3.06
		aDesigner	1.59	0.72	53	54.08	32	32.65	13	13.27	.	.
	I8	Acrolinx	1.68	0.81	48	48.98	37	37.76	9	9.18	4	4.08
		aDesigner	1.45	0.52	55	56.12	42	42.86	1	1.02	.	.
	I9	Acrolinx	1.58	0.80	57	58.16	28	28.57	10	10.2	3	3.06
		aDesigner	1.34	0.54	68	69.39	27	27.55	3	3.06	.	.
	I92	Acrolinx	1.77	0.99	54	55.10	21	21.43	15	15.31	8	8.16
		aDesigner	1.45	0.72	67	68.37	18	18.37	13	13.27	.	.
	I93	Acrolinx	1.54	0.99	73	74.49	4	4.08	14	14.29	7	7.14
		aDesigner	1.08	0.31	91	92.86	6	6.12	1	1.02	.	.

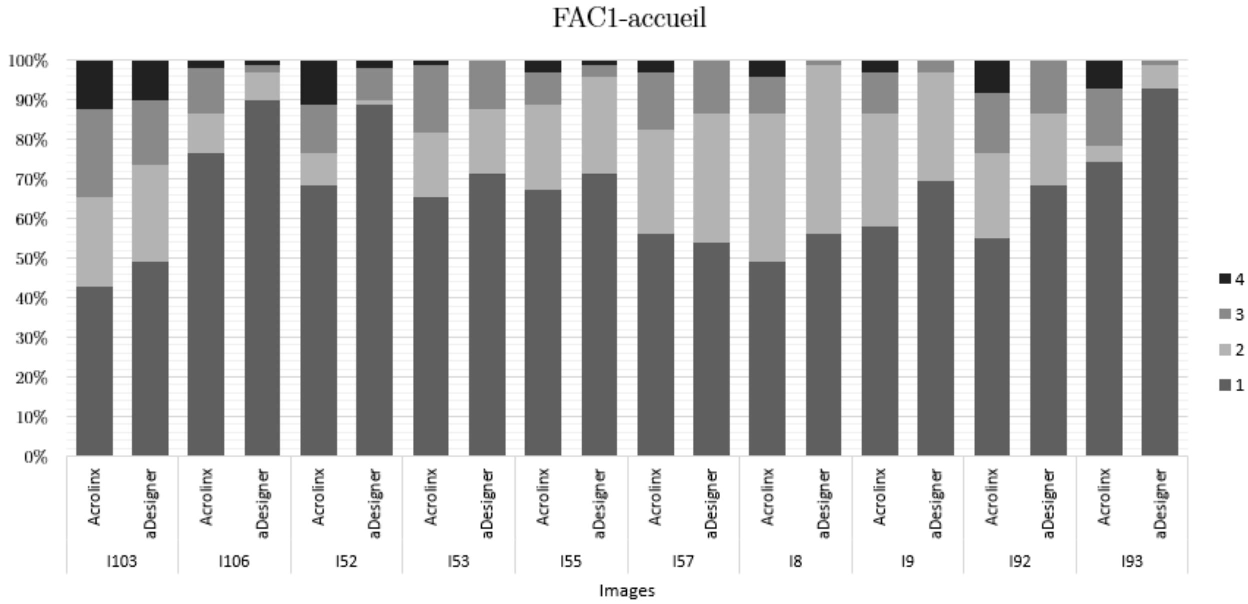


Figure K.20. Distribution of scores by image for the rule FAC1-accueil: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.27. Score mean, standard deviation and proportions (%) by image for the rule FDA3-structure in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
FDA3-structure	**I105**	Acrolinx	1.21	0.66	87	88.78	4	4.08	4	4.08	3	3.06
		aDesigner	1.04	0.2	94	95.92	4	4.08	.	.	.	.
	I11	Acrolinx	2.09	1.14	43	43.88	19	19.39	20	20.41	16	16.33
		aDesigner	1.80	1.00	51	52.04	26	26.53	11	11.22	10	10.20
	I129	Acrolinx	1.40	0.93	81	82.65	3	3.06	6	6.12	8	8.16
		aDesigner	1.03	0.17	95	96.94	3	3.06	.	.	.	.
	I13	Acrolinx	2.18	1.20	44	44.9	11	11.22	24	24.49	19	19.39
		aDesigner	1.61	0.92	60	61.22	23	23.47	8	8.16	7	7.14
	I16	Acrolinx	2.13	1.23	49	50.00	6	6.12	24	24.49	19	19.39
		aDesigner	1.26	0.72	85	86.73	5	5.10	4	4.08	4	4.08
	I44	Acrolinx	1.86	0.95	46	46.94	26	26.53	20	20.41	6	6.12
		aDesigner	1.46	0.61	59	60.20	33	33.67	6	6.12	.	.
	I56	Acrolinx	1.84	0.99	51	52.04	18	18.37	23	23.47	6	6.12
		aDesigner	1.40	0.65	68	69.39	21	21.43	9	9.18	.	.
	I58	Acrolinx	2.16	1.16	39	39.80	24	24.49	15	15.31	20	20.41
		aDesigner	1.55	0.75	55	56.12	36	36.73	3	3.06	4	4.08
	I60	Acrolinx	1.92	1.15	52	53.06	18	18.37	12	12.24	16	16.33
		aDesigner	1.27	0.44	72	73.47	26	26.53	.	.	.	.
	I89	Acrolinx	1.79	0.90	52	53.06	15	15.31	31	31.63	.	.
		aDesigner	1.22	0.42	76	77.55	22	22.45	.	.	.	.

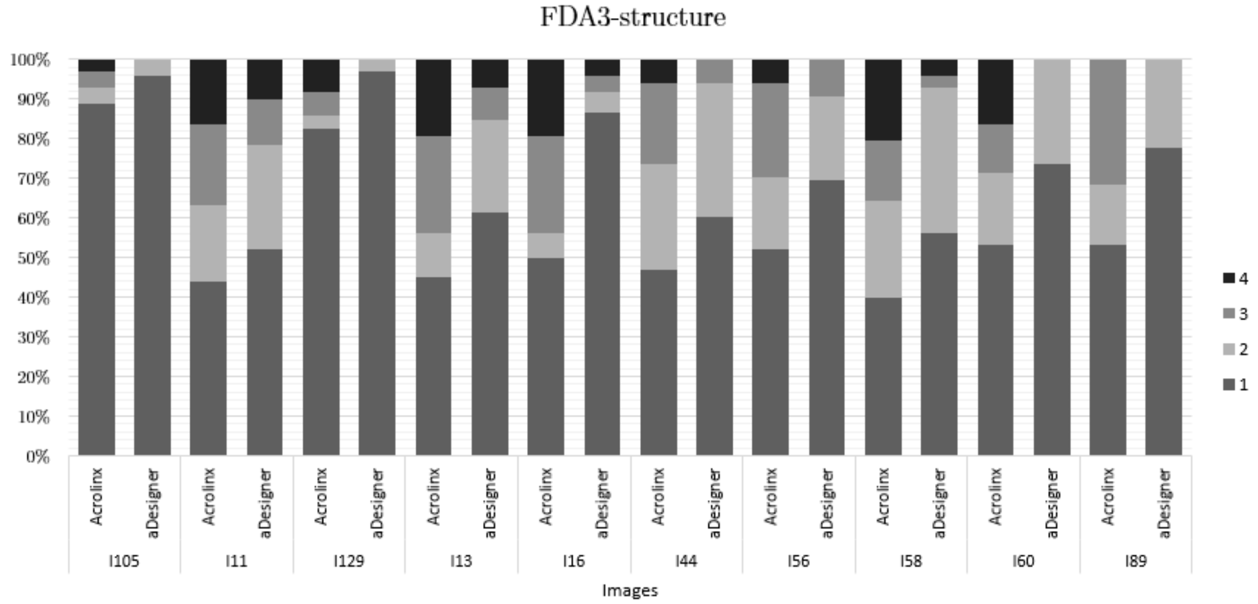


Figure K.21. Distribution of scores by image for the rule FDA3-structure: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.28. Score mean, standard deviation and proportions (%) by image for the rule U2-placeholder in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
U2-placeholder	**I109**	Acrolinx	1.65	1.07	70	71.43	.	.	20	20.41	8	8.16
		aDesigner	1.33	0.82	84	85.71	.	.	10	10.20	4	4.08
	I112	Acrolinx	1.87	1.21	63	64.29	.	.	20	20.41	15	15.31
		aDesigner	1.17	0.64	91	92.86	.	.	4	4.08	3	3.06
	I18	Acrolinx	1.47	0.80	70	71.43	11	11.22	16	16.33	1	1.02
		aDesigner	1.13	0.47	89	90.82	6	6.12	2	2.04	1	1.02
	I24	Acrolinx	1.35	0.81	81	82.65	4	4.08	9	9.18	4	4.08
		aDesigner	1.13	0.49	90	91.84	4	4.08	3	3.06	1	1.02
	I41	Acrolinx	1.44	0.80	70	71.43	17	17.35	7	7.14	4	4.08
		aDesigner	1.30	0.75	82	83.67	7	7.14	5	5.10	4	4.08
	I43	Acrolinx	1.32	0.55	71	72.45	23	23.47	4	4.08	.	.
		aDesigner	1.07	0.30	92	93.88	5	5.10	1	1.02	.	.
	I48	Acrolinx	1.58	0.92	66	67.35	11	11.22	17	17.35	4	4.08
		aDesigner	1.40	0.80	74	75.51	13	13.27	7	7.14	4	4.08
	I75	Acrolinx	1.41	0.84	75	76.53	11	11.22	7	7.14	5	5.10
		aDesigner	1.21	0.56	83	84.69	10	10.20	4	4.08	1	1.02
	I83	Acrolinx	1.35	0.73	77	78.57	10	10.20	9	9.18	2	2.04
		aDesigner	1.27	0.67	82	83.67	8	8.16	6	6.12	2	2.04
	I86	Acrolinx	1.43	0.73	69	70.41	17	17.35	11	11.22	1	1.02
		aDesigner	1.10	0.44	92	93.88	3	3.06	2	2.04	1	1.02

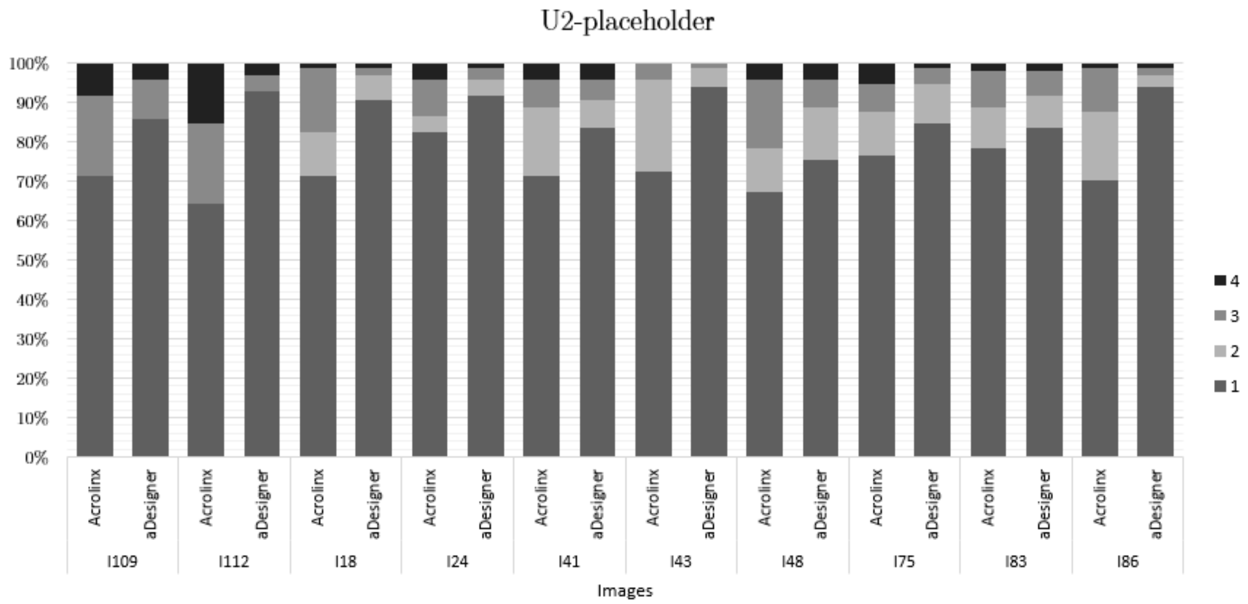


Figure K.22. Distribution of scores by image for the rule U2-placeholder: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.29. Score mean, standard deviation and proportions (%) by image for the rule U4-decor in the pairwise comparison of translation conditions: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Rule	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
U4-decor	**I1**	Acrolinx	2.11	1.01	34	34.69	30	30.61	23	23.47	11	11.22
		aDesigner	2.00	0.94	34	34.69	39	39.8	16	16.33	9	9.18
	I100	Acrolinx	1.63	0.99	66	67.35	8	8.16	18	18.37	6	6.12
		aDesigner	1.19	0.57	87	88.78	3	3.06	8	8.16	.	.
	I37	Acrolinx	1.32	0.62	75	76.53	15	15.31	8	8.16	.	.
		aDesigner	1.17	0.48	85	86.73	9	9.18	4	4.08	.	.
	I38	Acrolinx	1.49	0.68	60	61.22	28	28.57	10	10.20	.	.
		aDesigner	1.21	0.54	83	84.69	9	9.18	6	6.12	.	.
	I39	Acrolinx	1.44	0.80	69	70.41	20	20.41	4	4.08	5	5.10
		aDesigner	1.15	0.66	93	94.90	.	.	.	.	5	5.10
	I42	Acrolinx	1.58	0.80	57	58.16	28	28.57	10	10.20	3	3.06
		aDesigner	1.34	0.73	77	78.57	12	12.24	6	6.12	3	3.06
	I46	Acrolinx	1.88	1.01	48	48.98	22	22.45	20	20.41	8	8.16
		aDesigner	1.60	0.78	55	56.12	29	29.59	12	12.24	2	2.04
	I85	Acrolinx	1.64	0.56	39	39.80	55	56.12	4	4.08	.	.
		aDesigner	1.62	0.49	37	37.76	61	62.24	.	.	.	.
	I87	Acrolinx	1.52	0.68	57	58.16	31	31.63	10	10.20	.	.
		aDesigner	1.40	0.49	59	60.20	39	39.80	.	.	.	.
	I90	Acrolinx	1.35	0.75	78	79.59	8	8.16	10	10.20	2	2.04
		aDesigner	1.08	0.37	93	94.90	2	2.04	3	3.06	.	.

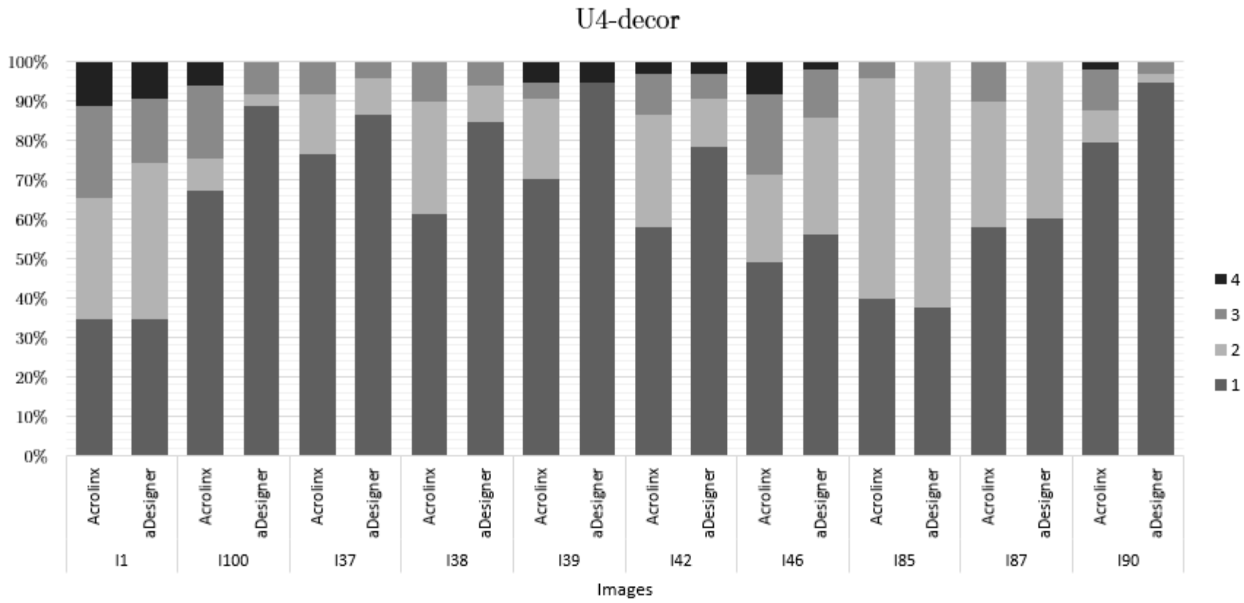


Figure K.23. Distribution of scores by image for the rule U4-decor: T3 (Both-last-Acrolinx) - T2 (aDesigner)

Table K.30. Score mean, standard deviation and proportions (%) of images which originally had an empty `alt` attribute in the source website, in the pairwise comparison of translation conditions: T2 (aDesigner) - T1 (none)

	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
alt-empty	**I108**	aDesigner	3.14	0.89	10	10.20	2	2.04	50	51.02	36	36.73
		none	3.43	0.50	.	.	.	.	56	57.14	42	42.86
	I124	aDesigner	1.99	0.75	27	27.55	46	46.94	24	24.49	1	1.02
		none	2.14	0.64	14	14.29	56	57.14	28	28.57	.	.
	I27	aDesigner	1.38	0.79	79	80.61	2	2.04	16	16.33	1	1.02
		none	1.29	0.70	84	85.71	.	.	14	14.29	.	.
	I30	aDesigner	1.34	0.75	81	82.65	1	1.02	16	16.33	.	.
		none	1.29	0.70	84	85.71	.	.	14	14.29	.	.
	I40	aDesigner	1.94	0.55	18	18.37	68	69.39	12	12.24	.	.
		none	2.14	0.35	.	.	84	85.71	14	14.29	.	.
	I61	aDesigner	2.96	1.11	18	18.37	8	8.16	32	32.65	40	40.82
		none	3.57	0.50	.	.	.	.	42	42.86	56	57.14
	I81	aDesigner	1.95	0.88	39	39.80	26	26.53	32	32.65	1	1.02
		none	2.14	0.84	28	28.57	28	28.57	42	42.86	.	.
	I82	aDesigner	1.97	0.88	38	38.78	26	26.53	33	33.67	1	1.02
		none	2.14	0.84	28	28.57	28	28.57	42	42.86	.	.
	I91	aDesigner	2.61	0.67	8	8.16	24	24.49	64	65.31	2	2.04
		none	2.71	0.45	.	.	28	28.57	70	71.43	.	.
	I96	aDesigner	1.51	0.76	64	65.31	18	18.37	16	16.33	.	.
		none	1.29	0.70	84	85.71	.	.	14	14.29	.	.

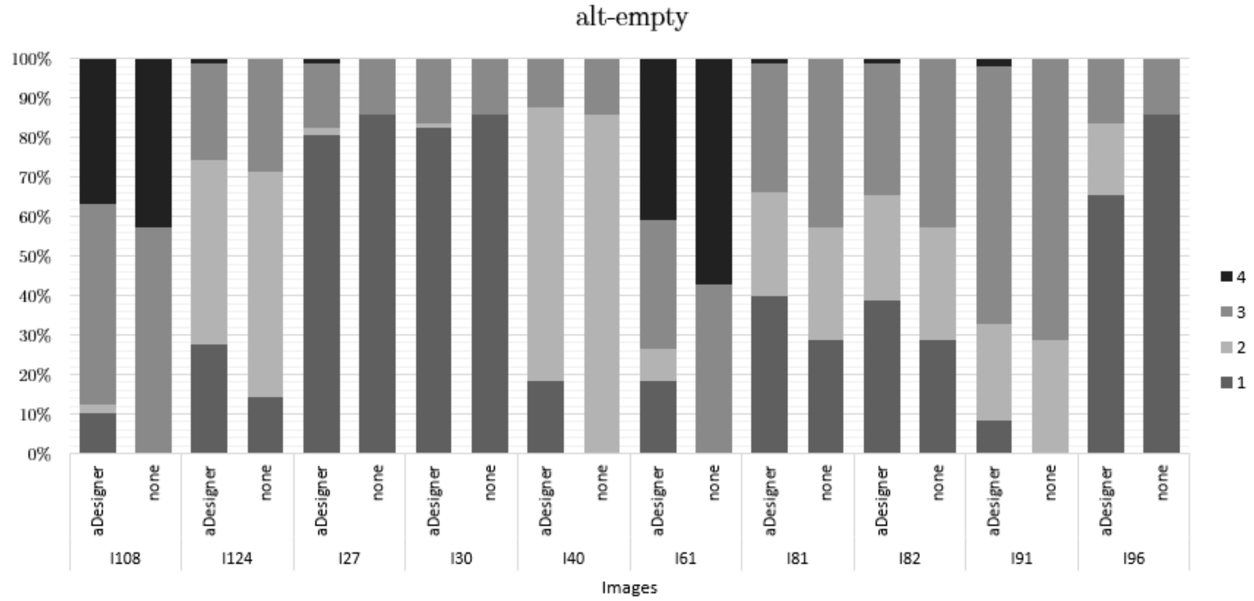


Figure K.24. Score distribution of images which originally had an empty `alt` attribute: T2 (aDesigner) - T1 (none)

Table K.31. Score mean, standard deviation and proportions (%) of images which originally had an empty `alt` attribute in the source website, in the pairwise comparison of translation conditions: T3 (Both-last-aDesigner) - T2 (Acrolinx) [5]

	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
alt-empty	**I108**	Acrolinx	3.08	0.97	14	14.29	.	.	48	48.98	36	36.73
		aDesigner	3.43	0.50	.	.	.	.	56	57.14	42	42.86
	I124	Acrolinx	2.12	0.80	22	22.45	46	46.94	26	26.53	4	4.08
		aDesigner	2.20	0.85	21	21.43	42	42.86	29	29.59	6	6.12
	I27	Acrolinx	1.48	0.91	74	75.51	6	6.12	13	13.27	5	5.10
		aDesigner	1.44	0.91	78	79.59	2	2.04	13	13.27	5	5.10
	I30	Acrolinx	1.44	0.90	78	79.59	1	1.02	15	15.31	4	4.08
		aDesigner	1.54	0.98	73	74.49	3	3.06	16	16.33	6	6.12
	I40	Acrolinx	2.01	0.68	20	20.41	59	60.2	17	17.35	2	2.04
		aDesigner	2.16	0.68	12	12.24	62	63.27	20	20.41	4	4.08
	I61	Acrolinx	3.05	1.11	18	18.37	3	3.06	33	33.67	44	44.90
		aDesigner	3.29	0.91	9	9.18	3	3.06	37	37.76	49	50
	I81	Acrolinx	2.15	0.91	29	29.59	30	30.61	34	34.69	5	5.10
		aDesigner	2.27	0.91	23	23.47	34	34.69	33	33.67	8	8.16
	I82	Acrolinx	2.04	0.91	35	35.71	27	27.55	33	33.67	3	3.06
		aDesigner	2.10	0.90	31	31.63	30	30.61	33	33.67	4	4.08
	I91	Acrolinx	2.47	0.74	14	14.29	24	24.49	60	61.22	.	.
		aDesigner	2.71	0.45	.	.	28	28.57	70	71.43	.	.
	I96	Acrolinx	1.97	1.04	46	46.94	17	17.35	27	27.55	8	8.16
		aDesigner	2.00	1.03	43	43.88	20	20.41	27	27.55	8	8.16

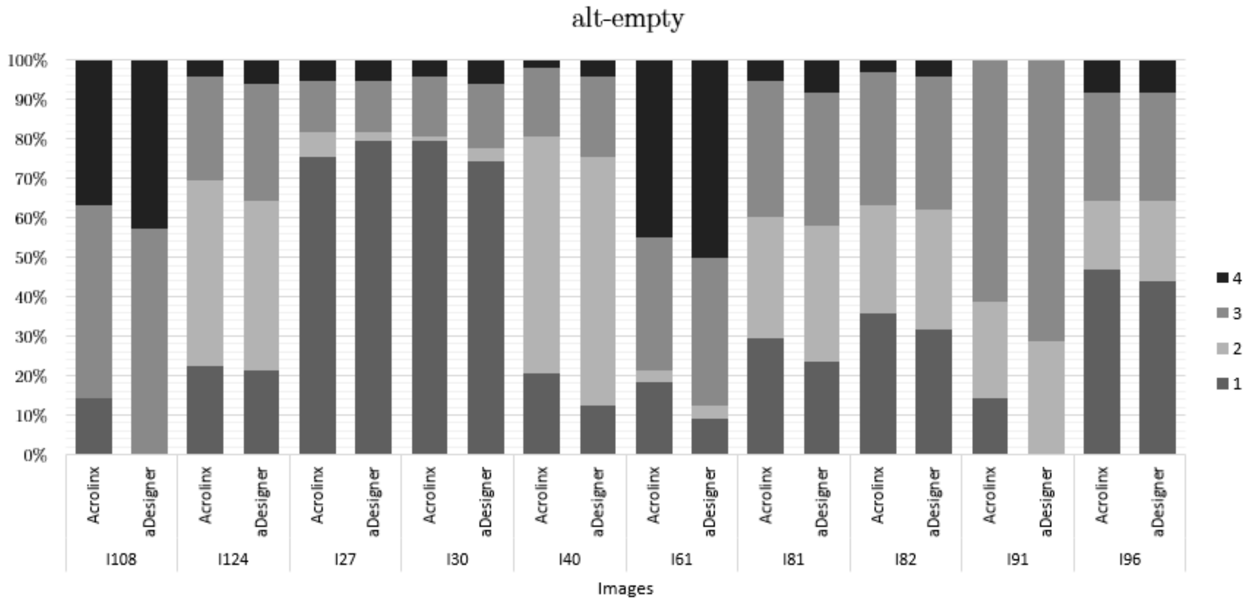


Figure K.25. Score distribution of images which originally had an empty `alt` attribute: T3 (Both-last-aDesigner) - T2 (Acrolinx)

Table K.32. Score mean, standard deviation and proportions (%) of images which originally had no `alt` attribute in the source website, in the pairwise comparison of translation conditions: T2 (aDesigner) - T1 (none)

	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
no-alt	**I101**	aDesigner	1.22	0.60	84	85.71	7	7.14	6	6.12	1	1.02
		none	1.00	0	98	100	.	.	.	.	.	.
	I104	aDesigner	1.19	0.51	84	85.71	9	9.18	5	5.1	.	.
		none	1.00	0	98	100	.	.	.	.	.	.
	I118	aDesigner	1.18	0.58	88	89.8	3	3.06	6	6.12	1	1.02
		none	1.00	0	98	100	.	.	.	.	.	.
	I123	aDesigner	1.00	0	98	100	.	.	.	.	.	.
		none	1.00	0	98	100	.	.	.	.	.	.
	I15	aDesigner	1.39	0.92	82	83.67	1	1.02	8	8.16	7	7.14
		none	1.00	0	98	100	.	.	.	.	.	.
	I20	aDesigner	1.20	0.56	85	86.73	6	6.12	7	7.14	.	.
		none	1.00	0	98	100	.	.	.	.	.	.
	I33	aDesigner	1.23	0.67	85	86.73	6	6.12	4	4.08	3	3.06
		none	1.00	0	98	100	.	.	.	.	.	.
	I36	aDesigner	1.37	0.78	77	78.57	9	9.18	9	9.18	3	3.06
		none	1.00	0	98	100	.	.	.	.	.	.
	I68	aDesigner	1.39	0.78	74	75.51	14	14.29	6	6.12	4	4.08
		none	1.00	0	98	100	.	.	.	.	.	.
	I97	aDesigner	1.21	0.56	84	85.71	7	7.14	7	7.14	.	.
		none	1.00	0	98	100	.	.	.	.	.	.

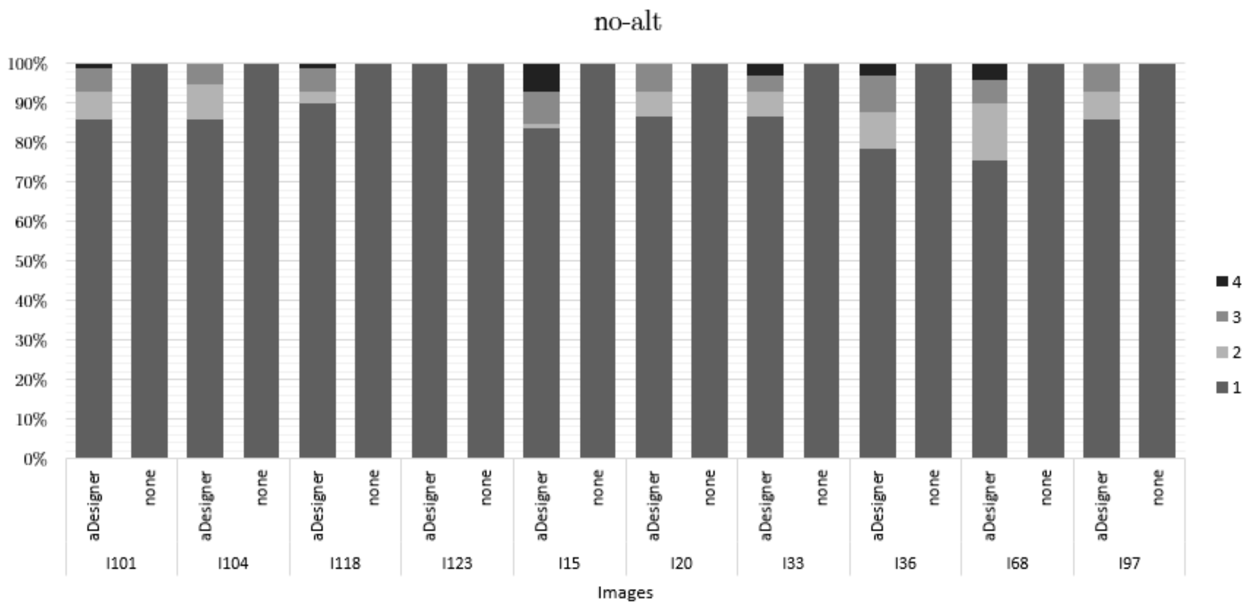


Figure K.26. Score distribution of images which originally had no `alt` attribute: T2 (aDesigner) - T1 (none)

Table K.33. Score mean, standard deviation and proportions (%) of images which originally had no `alt` attribute in the source website, in the pairwise comparison of translation conditions: T3 (Both-last-aDesigner) - T2 (Acrolinx)

	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
no-alt	**I101**	Acrolinx	1.76	1.08	61	62.24	11	11.22	15	15.31	11	11.22
		aDesigner	1.96	1.09	48	48.98	18	18.37	20	20.41	12	12.24
	I104	Acrolinx	1.47	0.92	74	75.51	9	9.18	8	8.16	7	7.14
		aDesigner	1.78	1.06	58	59.18	14	14.29	16	16.33	10	10.20
	I118	Acrolinx	1.00	0	98	100	.	.	.	.	.	.
		aDesigner	1.54	0.93	70	71.43	8	8.16	15	15.31	5	5.10
	I123	Acrolinx	1.24	0.69	86	87.76	2	2.04	8	8.16	2	2.04
		aDesigner	1.39	0.77	75	76.53	10	10.20	11	11.22	2	2.04
	I15	Acrolinx	1.35	0.87	84	85.71	.	.	8	8.16	6	6.12
		aDesigner	2.73	1.18	28	28.57	.	.	40	40.82	30	30.61
	I20	Acrolinx	1.35	0.70	76	77.55	11	11.22	10	10.20	1	1.02
		aDesigner	2.00	0.92	38	38.78	25	25.51	32	32.65	3	3.06
	I33	Acrolinx	1.18	0.66	90	91.84	2	2.04	2	2.04	4	4.08
		aDesigner	1.62	1.08	69	70.41	10	10.20	6	6.12	13	13.27
	I36	Acrolinx	1.22	0.67	86	87.76	5	5.10	4	4.08	3	3.06
		aDesigner	1.56	0.86	64	65.31	16	16.33	15	15.31	3	3.06
	I68	Acrolinx	1.19	0.62	87	88.78	6	6.12	2	2.04	3	3.06
		aDesigner	1.89	1.02	47	47.96	25	25.51	16	16.33	10	10.20
	I97	Acrolinx	1.73	1.06	61	62.24	12	12.24	15	15.31	10	10.20
		aDesigner	2.00	1.11	47	47.96	17	17.35	21	21.43	13	13.27

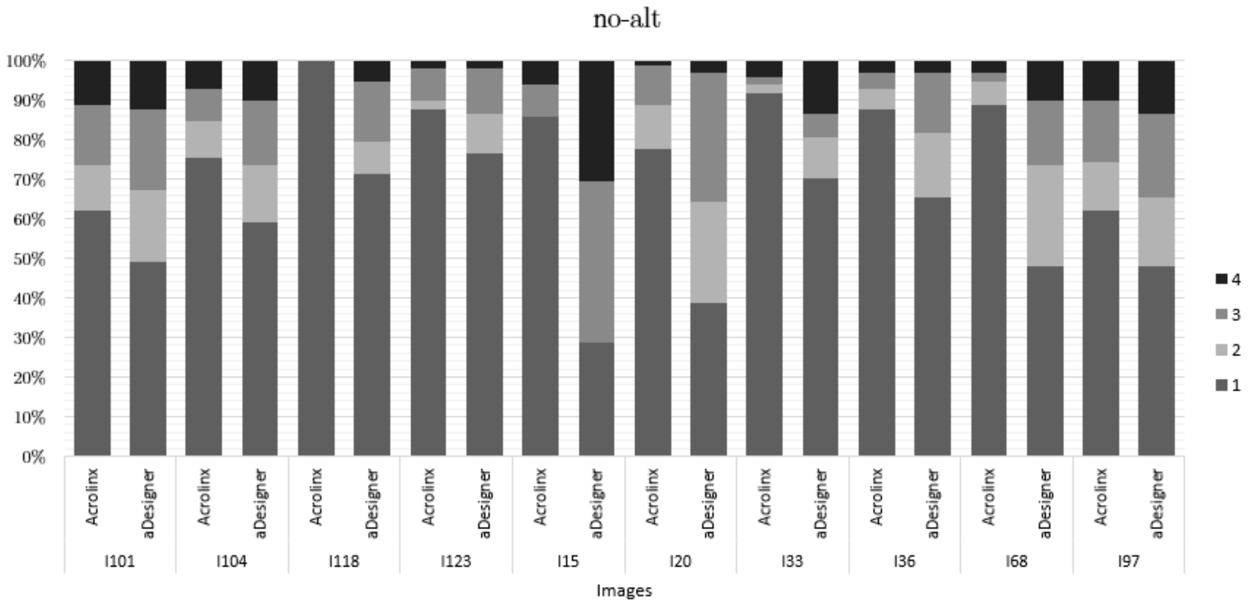


Figure K.27. Score distribution of images which originally had no `alt` attribute: T3 (Both-last-aDesigner) - T2 (Acrolinx)

Table K.34. Score mean, standard deviation and proportions (%) of images which originally had an appropriate alt text in the source website, in the pairwise comparison of translation conditions: T2 (aDesigner) - T1 (none)

	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
app-alt	**I121**	aDesigner	1.96	0.95	38	38.78	34	34.69	18	18.37	8	8.16
		none	2.00	0.93	34	34.69	38	38.78	18	18.37	8	8.16
	I126	aDesigner	2.61	1.12	23	23.47	18	18.37	31	31.63	26	26.53
		none	2.69	1.08	19	19.39	19	19.39	33	33.67	27	27.55
	I127	aDesigner	2.17	1.05	37	37.76	16	16.33	36	36.73	9	9.18
		none	2.27	1.04	33	33.67	16	16.33	39	39.80	10	10.20
	I128	aDesigner	2.31	1.13	35	35.71	14	14.29	33	33.67	16	16.33
		none	2.37	1.10	31	31.63	16	16.33	35	35.71	16	16.33
	I21	aDesigner	2.53	1.22	32	32.65	10	10.20	28	28.57	28	28.57
		none	2.62	1.20	28	28.57	11	11.22	29	29.59	30	30.61
	I64	aDesigner	2.08	0.90	34	34.69	24	24.49	38	38.78	2	2.04
		none	2.14	0.89	30	30.61	26	26.53	40	40.82	2	2.04
	I69	aDesigner	2.18	1.07	37	37.76	17	17.35	33	33.67	11	11.22
		none	2.23	1.05	34	34.69	18	18.37	35	35.71	11	11.22
	I7	aDesigner	3.04	0.95	6	6.12	24	24.49	28	28.57	40	40.82
		none	3.10	0.91	5	5.10	21	21.43	31	31.63	41	41.84
	I70	aDesigner	2.44	1.12	29	29.59	17	17.35	32	32.65	20	20.41
		none	2.52	1.11	26	26.53	17	17.35	33	33.67	22	22.45
	I99	aDesigner	2.90	1.21	22	22.45	11	11.22	20	20.41	45	45.92
		none	2.98	1.19	20	20.41	10	10.20	20	20.41	48	48.98

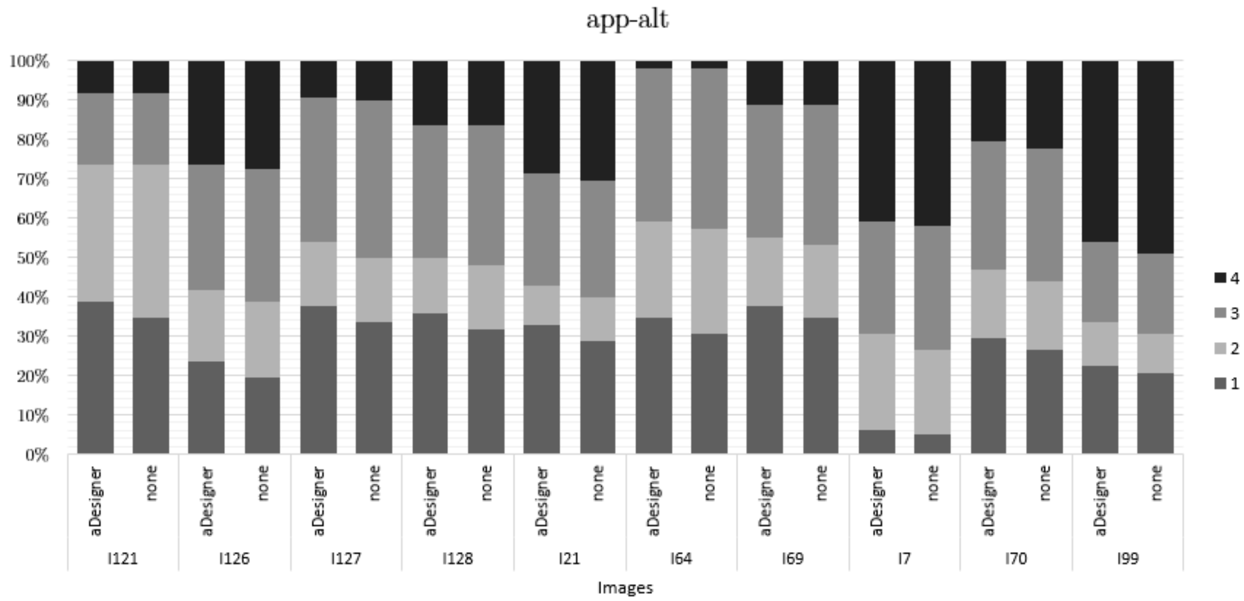


Figure K. 28. Score distribution of images which originally had an appropiate alt text: T2 (aDesigner) - T1 (none)

Table K.35. Score mean, standard deviation and proportions (%) of images which originally had an appropriate alt text in the source website, in the pairwise comparison of translation conditions: T3 (Both-last-aDesigner) - T2 (Acrolinx)

	Image	Tool	Mean	SD	1	%	2	%	3	%	4	%
app-alt	**I121**	Acrolinx	2.00	0.94	35	35.71	36	36.73	19	19.39	8	8.16
		aDesigner	2.13	0.95	28	28.57	39	39.80	21	21.43	10	10.20
	I126	Acrolinx	2.31	1.10	29	29.59	29	29.59	21	21.43	19	19.39
		aDesigner	2.41	1.06	23	23.47	32	32.65	23	23.47	20	20.41
	I127	Acrolinx	2.01	1.06	42	42.86	25	25.51	19	19.39	12	12.24
		aDesigner	2.14	1.07	35	35.71	29	29.59	19	19.39	15	15.31
	I128	Acrolinx	2.15	1.04	34	34.69	27	27.55	25	25.51	12	12.24
		aDesigner	2.24	1.02	28	28.57	31	31.63	26	26.53	13	13.27
	I21	Acrolinx	2.42	1.21	36	36.73	8	8.16	31	31.63	23	23.47
		aDesigner	2.42	1.21	36	36.73	8	8.16	31	31.63	23	23.47
	I64	Acrolinx	2.38	1.10	31	31.63	15	15.31	36	36.73	16	16.33
		aDesigner	2.38	1.10	31	31.63	15	15.31	36	36.73	16	16.33
	I69	Acrolinx	2.30	1.02	28	28.57	25	25.51	33	33.67	12	12.24
		aDesigner	2.39	1.02	25	25.51	24	24.49	35	35.71	14	14.29
	I7	Acrolinx	3.11	0.91	5	5.10	20	20.41	32	32.65	41	41.84
		aDesigner	3.11	0.91	5	5.10	20	20.41	32	32.65	41	41.84
	I70	Acrolinx	2.37	1.08	30	30.61	17	17.35	36	36.73	15	15.31
		aDesigner	2.37	1.08	30	30.61	17	17.35	36	36.73	15	15.31
	I99	Acrolinx	2.90	1.21	22	22.45	10	10.20	22	22.45	44	44.90
		aDesigner	2.90	1.21	22	22.45	10	10.20	22	22.45	44	44.90

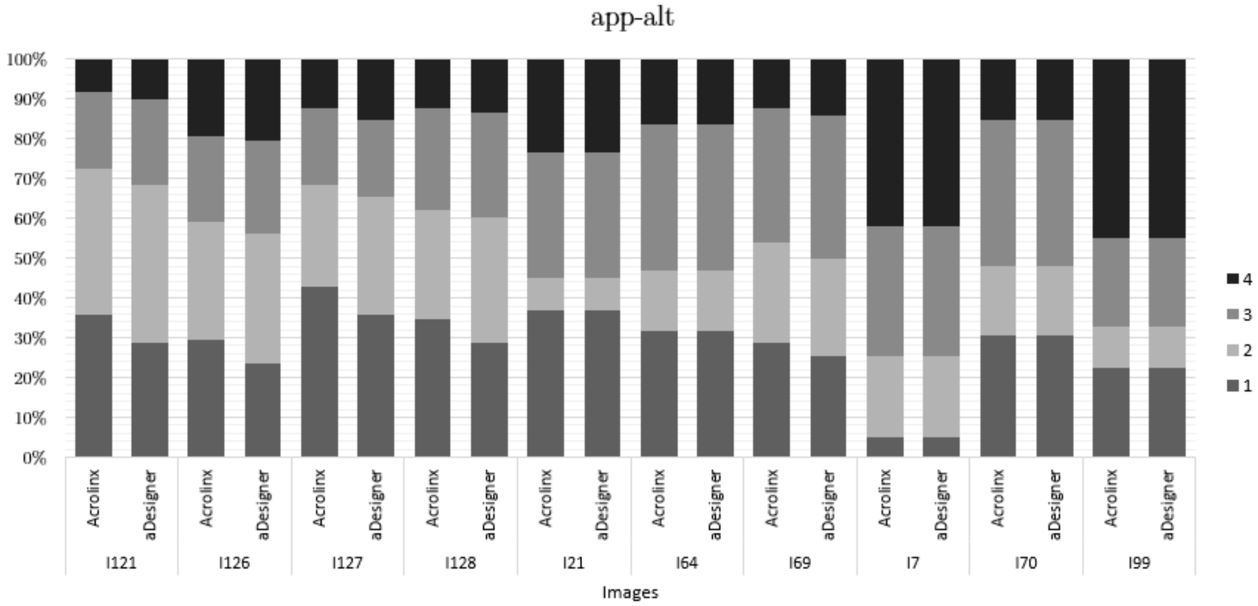


Figure K.29. Score distribution of images which originally had an appropriate alt text: T3 (Both-last-aDesigner) - T2 (Acrolinx)

Appendix L. Resumen ampliado de la tesis en español

1. Contextualización del trabajo de investigación

Las tecnologías de la información y la comunicación (TIC) se encuentran presentes en prácticamente todos los aspectos de nuestra vida cotidiana. Han cambiado la forma en la que se genera y se comparte conocimiento en las esferas profesional, política, educativa, económica y social, entre otras. En esta era tecnológica en la que la alfabetización digital es ya la norma (Folaron 2012), la World Wide Web (la web) está considerada hoy en día como un producto básico. Tal y como afirma su creador, el poder de la web reside en el hecho de que infinidad de documentos se publican en línea siguiendo un formato estandarizado y se vinculan entre ellos, y «la universalidad y la flexibilidad de este tipo de arquitectura de nodos tiene la capacidad de romper las barreras impuestas por la distancia, el idioma y los distintos campos de conocimiento»[156] (Berners-Lee 2007).

El World Wide Web Consortium (W3C) es una organización internacional que desarrolla protocolos y normas para garantizar la evolución de la web a largo plazo, tomando como referencia su visión de «una sola web» (*One Web*), que, a su vez, se asienta en dos principios de diseño web fundamentales: «la web en todo» (*Web on Everything*) y «la web para todos» (*Web for All*). El primero de ellos persigue la promoción del desarrollo de tecnologías que permitan el acceso a Internet desde cualquier lugar, en cualquier momento, a través de cualquier dispositivo; el segundo, por su parte, busca conseguir el pleno potencial de la web como un constructo social. El objetivo del paradigma de la «Web para todos» es que la información contenida en la web sea accesible para cualquier persona, independientemente de la tecnología que utilice (hardware y/o software), su lengua o cultura de origen, su ubicación geográfica o sus diferentes (dis)capacidades.

Los avances logrados a lo largo de los últimos treinta años en lo que respecta a las tecnologías destinadas a facilitar la interacción persona-ordenador (IPO) para las personas con discapacidad (véase la sección 1.1), así como la proliferación de pautas de accesibilidad web (AW) (véase la sección 1.2) y la creciente demanda de contenido web internacionalizado y localizado (véase sección 1.3) han sido cruciales para el desarrollo de sitios web más inclusivos para todos. Siguiendo un enfoque interdisciplinar, esta tesis se asienta sobre estos tres aspectos fundamentales de la

[156] Traducción propia.

sociedad de la información para abogar por un mayor nivel de concienciación sobre la accesibilidad en la producción de sitios web multilingües y, especialmente, durante la adaptación de elementos gráficos que se produce en todo proceso de localización web (véase la sección 2).

1.1 El acceso a la web por las personas con discapacidad

La discapacidad forma parte de la condición humana. La Organización Mundial de la Salud (OMS) afirma que casi todas las personas experimentarán algún tipo de discapacidad en algún momento de su vida (OMS 2011, 3). Esto se debe a que la el concepto de discapacidad no se entiende únicamente como una característica inherente a la persona, causada por una enfermedad o cualquier otro problema de salud, sino que también constituye un problema social, ajeno a la naturaleza del individuo (OMS 2002, 8-9), que depende del entorno que le rodea. Si dicha definición se aplicase en el campo de la IPO, diríamos que la discapacidad de las personas viene dada no solo por sus posibles limitaciones físicas a la hora de interactuar con la web, sino también por la tecnología utilizada en un principio para crear contenido y las ayudas técnicas de las que dichas personas dependen para percibirlo (Harper y Yesilada 2008a, xv). En el presente trabajo, nuestro interés se centra en la interacción entre la tecnología (para la) web y las personas con discapacidad visual.

1.1.1 La discapacidad visual

De acuerdo con la clasificación de la OMS, la función visual se subdivide en cuatro niveles: visión normal, discapacidad visual moderada, discapacidad visual grave y ceguera. Los dos últimos niveles se reagrupan comúnmente bajo el término «baja visión». A su vez, la baja visión y la ceguera representan conjuntamente el total de casos de discapacidad visual (OMS 2014). A nivel mundial, hay 39 millones de personas ciegas y 246 millones con baja visión (ibid). En España, el último informe al respecto indica que casi un millón de personas cuentan con algún tipo de discapacidad visual, de las cuales 71.000 son ciegas (Ernst & Young 2012, 28). En proporción, las cifras también son relativamente elevadas en Suiza: de las 325.000 personas con discapacidad visual del país, aproximadamente 10.000 sufren de ceguera (SNAB 2012, 4).

Hasta los años 80, las fuentes de información que existían para los ciegos se limitaban a unos cuantos libros en papel Braille o en audio (Asakawa 2014). Con la llegada de los ordenadores personales como un «medio de comunicación dinámico estructurado» (Winograd y Flores, 1987, 176) y de la web como un «universo global de información» (Berners-Lee, 1992), las personas con discapacidad visual, y muy especialmente los ciegos, comenzaron a disfrutar de una mayor autonomía: a día de hoy, los usuarios ciegos pueden llevar a cabo actividades propias de la vida cotidiana sin la ayuda de una persona vidente, como hacer compras en línea, leer el periódicos realizar transacciones bancarias o simplemente comunicarse más fácilmente con los amigos. En la actualidad, la web es para muchos de ellos una fuente vital de información, empleo y entretenimiento (Harper y Yesilada 2008a, 1).

Todo ello ha sido posible gracias a los esfuerzos emprendidos desde la década de los 90 por el sector privado y el mundo académico para facilitar el acceso de los usuarios ciegos tanto a los ordenadores como a la web. Dado que esta última es, por definición, un medio visual, el reto principal consistió en buscar cómo ofrecer a este grupo de población una experiencia de usuario similar a la de aquellos que podían ver la pantalla, de manera que pudieran navegar por la web de forma intuitiva y disfrutar de los beneficios de Internet (Asakawa 2005). En este contexto, la mayoría de las soluciones se centraron en tratar de presentar la información visualizada en la pantalla a través del canal auditivo y mediante dispositivos hápticos (Barreto 2008, 10). Todos estos esfuerzos dieron lugar, finalmente, a dos innovaciones tecnológicas importantes: la digitalización del sistema braille y el acceso a la web a través de sintetizadores de voz (Asakawa 2014).

1.1.2 Ayudas técnicas para usuarios ciegos

El término «ayudas técnicas» se utiliza en el campo de la accesibilidad para referirse a hardware y/o software que facilita el uso del ordenador para las personas con discapacidad (RDC 2004, 1) que no pueden utilizar las aplicaciones de usuario convencionales necesarias para acceder a una interfaz gráfica de usuario (GUI); por ejemplo, un ratón o una pantalla. La funcionalidad proporcionada por las ayudas técnicas incluye presentaciones alternativas (por ejemplo, voz sintetizada o contenido ampliado), métodos de entrada de datos alternativos (por ejemplo, voz), navegación adicional o mecanismos de orientación, y transformaciones de contenido (Caldwell et al. 2008).

Las personas con baja visión presentan dificultades diferentes en relación con el acceso a la web que los usuarios ciegos. Las primeras suelen buscar la adaptación de la información digital a su nivel de visión, mientras que los ciegos necesitan alternativas no visuales para el contenido que aparece en la pantalla (Edwards 2008, 150-151). En este sentido, los usuarios con baja visión utilizan software destinados a aumentar el tamaño del texto o de las imágenes, comúnmente conocidos como magnificadores de pantalla (Paciello 2000, 71), mientras que las personas ciegas utilizan una mayor variedad de ayudas técnicas, entre las que se encuentran las líneas braille, los navegadores de voz y los lectores de pantalla.

Estos últimos son los más utilizados entre dicha comunidad. Se trata de un programa informático que, mediante un sintetizador de voz, retrascribe lo que se muestra en la pantalla de un ordenador en cuanto a contenido y estructura, proporcionando a los usuarios una serie de mecanismos para simplificar la búsqueda de información (Barreto 2008, 10). El principio de diseño visual de la web se basa en el hecho de que las personas sin discapacidad visual pueden elegir en cualquier momento la parte del texto en la que desean centrar su atención (Edwards 2008, 154). Los lectores de pantalla reemplazan este tipo de funcionalidad a través de una interfaz de acceso directo que permite a los usuarios navegar entre los diferentes elementos de una página de forma secuencial. Por ejemplo, para acceder a los enlaces de una página web, el usuario puede hacer, a través de un determinada combinación de teclas, que se abra una ventana auxiliar con una lista de todos los enlaces o títulos en la página (Borodin et al. 2010), lo que le permite crear un modelo mental de la misma (Theofanos y Redish 2006). Para los usuarios ciegos, por tanto, el poder navegar por la web depende del diseño de la página y de la pertinencia de su estructura semántica (Connor 2012, 37). De la misma manera, conseguir que el contenido sea suficientemente robusto dependerá, a su vez, de los conocimientos de la persona que lo crea (o de las características de la herramienta utilizadas para ello) en cuanto a las buenas prácticas de diseño web. La interdependencia y la complementariedad entre los diferentes elementos del ciclo de desarrollo web se tratarán en la sección 1.2.

1.1.3 El acceso al contenido gráfico de la web

La capacidad del ser humano en términos de agudeza visual, sensibilidad al contraste y campo de visión, así como los avances en la tecnología, han dado rienda suelta a los desarrolladores para explotar la riqueza de la web como interfaz visual (Barreto 2008, 8). Un estudio realizado por Asakawa (2005) indica que el número de imágenes en la web se ha cuadruplicado entre 1996 y 2005, un hecho que fue confirmado más tarde por Chen y Harper (2008) al observar un crecimiento significativo en el uso de formatos gráficos en los documentos web a lo largo de un periodo de diez años (1999-2008). En estas circunstancias, el uso de lectores de pantalla plantea cada vez más desafíos.

Los lectores de pantalla fueron diseñados para transformar un documento de dos dimensiones en una cadena de texto unidimensional, convirtiendo así la GUI convencional en una «interfaz auditiva de usuario» (Freitas 2010, 273). Este «linearización» de la página web (Thatcher et al. 2006, 105), que implica el paso del canal visual al auditivo, puede dar lugar a una pérdida de información considerable si el contenido gráfico que se presenta al usuario no dispone de alternativas textuales que cumplan una función similar. La precisión y la eficacia del acto de comunicación entre el usuario ciego y el ordenador dependen, así pues, de (i) la existencia de dichas alternativas textuales, y (ii) de la capacidad del lector de pantalla para identificarlas como tal y transmitir la información correspondiente al usuario.

Para que las personas ciegas puedan percibir contenido gráfico como, por ejemplo, las imágenes, éstas deben ir acompañadas de un texto alternativo. El atributo `alt`, introducido por primera vez en 1995 como parte de la especificación HTML 2.0 para proporcionar este tipo de funcionalidad (Berners-Lee y Connolly 1995) es el mecanismo más utilizado a día de hoy para que las imágenes sean accesibles. Por lo general, las imágenes se introducen en una página web mediante la etiqueta <img>. Al encontrar este elemento, el lector de pantalla señala la presencia de la imagen y lee el valor del atributo alt correspondiente. Por ejemplo, si un usuario ciego daría con la imagen mostrada en la figura 1, escucharía lo siguiente: «Gráfico: La playa de arena blanca de Rodas, en Galicia, España, en un día de verano con cielo azul». Si no se utilizase el atributo alt, el lector de pantalla indicaría la presencia de la imagen, pero no le proporcionaría al usuario ninguna alternativa textual, lo que afectaría negativamente a la experiencia de navegación. Del mismo modo, la interacción entre la web y el usuario ciego no sería eficaz si se ofreciese un texto alternativo irrelevante, que no describiese lo que se muestra verdaderamente en la imagen.

```
<body>
...
<img
src="myweb-en/images/praia-rodas-
cies.jpg" alt=" La playa de arena
blanca de Rodas, en Galicia, España,
en un día de verano con cielo
azul.">
...
</body>
```

Figura 1. Presentación visual y no visual (HTML) de una imagen en el contexto de la web

En nuestro trabajo de investigación, se parte del principio de que los textos alternativos son la clave para conseguir imágenes más accesibles, y constituyen nuestro objeto de estudio principal (ver sección 2.2). En el capítulo 3 de la tesis, se abordan, de manera más detallada, cuestiones relacionadas con estos elementos, como su composición, formulación y evaluación.

1.2 Los estándares de accesibilidad web

Estar familiarizado con las necesidades de las personas con discapacidad y con cómo perciben la web es crucial para el diseño de sitios web accesibles. No obstante, esto no siempre es suficiente para garantizar que la interacción persona-ordenador para este grupo de usuarios sea eficaz. La accesibilidad web también depende de aspectos técnicos, tales como el cumplimiento de las especificaciones técnicas establecidas por el W3C y las recomendaciones en materia de accesibilidad elaboradas por la *Web Accessibility Initiative* (WAI), una iniciativa fundada en 1997 para desarrollar pautas y recursos para hacer la web más accesible.

El término «accesibilidad web» significa que las personas con discapacidad puedan percibir, entender, navegar e interactuar con la Web (Lawton Henry 2005b). De acuerdo con la WAI, la accesibilidad web depende de la interacción entre diferentes componentes técnicos y humanos (Henry Chisholm 2005, 32), tal y como como se muestra en la figura 2.[157] Por un lado, los componentes técnicos incluyen aplicaciones de usuario (navegadores, ayudas técnicas, etc.), las herramientas de edición y evaluación utilizadas por los profesionales que participan en el ciclo de producción web, y las especificaciones técnicas que describen los lenguajes que se utilizan para crear contenido. Por otro lado, los componentes humanos incluyen no solo los usuarios finales, sino también las personas que diseñan, codifican, crean y editan contenido web (ibid). Thatcher et al. (2006, 14-15) utilizan el caso de los textos alternativos para ilustrar las interdependencias entre dichos componentes, como se resume a continuación:

- Las especificaciones técnicas proporcionan normas para utilizar un lenguaje de marcado que permita la introducción de textos alternativos para las imágenes (atributo `alt`).
- Las herramientas de edición facilitan y promueven el uso del atributo `alt`.
- Las herramientas de evaluación permiten a los creadores de contenido comprobar si se han utilizado los atributos `alt` y les ayudan a determinar si son pertinentes.
- Las aplicaciones de usuario proporcionan interfaces para transmitir los textos alternativos al usuario.
- Los creadores de contenido formulan textos alternativos y los introducen en la página a través del atributo `alt`, utilizando habitualmente para ello herramientas de edición y evaluación.
- Los usuarios finales perciben los textos alternativos gracias al uso de un navegador y ayudas técnicas.

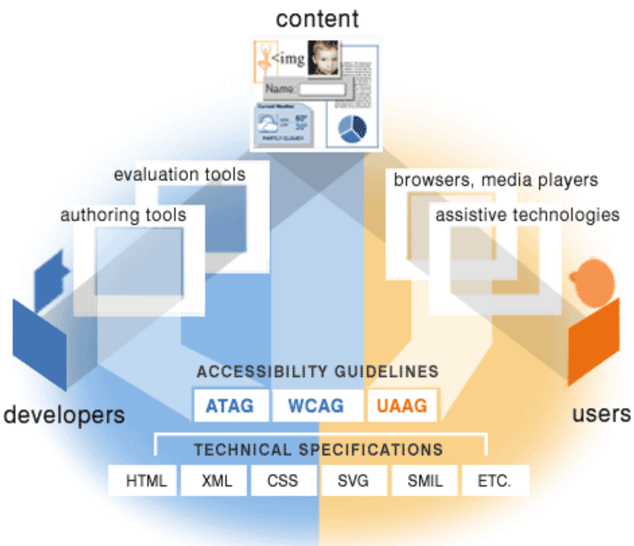


Figura 2. Componentes esenciales de accesibilidad web

[157] Imagen de Michael Duffy: «Essential Components of Web Accessibility». S.L. Henry, ed. W3C (MIT, ERCIM, Keio). Última actualización: agosto de 2005. www.w3.org/WAI/intro/components. Fecha de consulta: 2 de marzo de 2016.

Cuando uno de dichos componentes pone en práctica de forma adecuada un mecanismo para la mejora de la accesibilidad, como el atributo `alt`, hay una mayor probabilidad de que los otros también lo hagan (ibid). Con el fin de fomentar esta visión y facilitar la interacción entre todos estos componentes, la WAI publicó entre 1999 y 2000 una serie de normas de accesibilidad basadas en las especificaciones técnicas de la web: las Pautas de accesibilidad para herramientas de autor (ATAG), las Pautas de accesibilidad para aplicaciones de usuario (UAAG) y las Pautas de accesibilidad para el contenido web (WCAG). En el marco de esta tesis, nos centramos en este último documento (WCAG), ya que prescribe cómo crear contenido web accesible (incluidas las imágenes) para las personas con discapacidad.

La primera versión de las WCAG (1.0) fue publicada como una Recomendación del W3C en mayo de 1999 (Chisholm et al. 1999). Después de un proceso de revisión de cinco años, la versión 2.0 fue presentada en mayo de 2008 (Caldwell et al. 2008) con cambios significativos. Transformadas ahora en una norma ISO/IEC (40500: 2012), las WCAG están consideradas como un documento de referencia fundamental para el diseño web. Las pautas en esta nueva y última versión están organizadas en torno a cuatro principios: 1. El contenido debe ser *perceptible;* 2. Los componentes de la interfaz de usuario y la navegación deben ser *operables;* 3. La información y el manejo de la interfaz de usuario deben ser *comprensibles;* y 4. El contenido debe ser lo suficientemente *robusto* como para ser interpretado de forma fiable por una amplia variedad de aplicaciones de usuario. Existen 12 pautas asociadas a estos principios[158] y, para cada una de ellas, se proporcionan criterios de conformidad (un total de 61) en tres niveles (A, el más bajo; AA, y AAA, el más alto) para determinar su grado de cumplimiento.

En términos generales, esta tesis persigue investigar por qué los localizadores, en su calidad de creadores de contenido web en el marco del ciclo de desarrollo de sitios web multilingües (véase sección 1.3), también deben ser responsables de garantizar que el producto meta cuenta con un nivel de accesibilidad aceptable, no sólo gracias a la aplicación de las WCAG 2.0, sino también al uso de herramientas de control de calidad destinadas a la evaluación de la accesibilidad de documentos web. A un nivel más específico, buscamos estudiar la aplicación y la evaluación de la pauta 1.1 de las WCAG 2.0 en los sitios web localizados, la cual establece lo siguiente: *Proporcionar alternativas textuales para todo contenido no textual de modo que se pueda convertir a otros formatos que las personas necesiten, tales como textos ampliados, braille, voz, símbolos o en un lenguaje más simple.* (Caldwell et al. 2008).

[158] En la tabla 2.1 del capítulo 2 de la tesis se presenta una lista completa de las pautas.

1.3 Desarrollo de sitios web multilingües

En general, la localización web se puede describir como la traducción, ingeniería y testeo de contenido web (Esselink 2000, 3). Al proporcionar una experiencia de navegación en la lengua materna de los usuarios, la localización también contribuye al principio de una «web para todos» del W3C y representa la actividad principal de todo proceso de desarrollo de sitios web multilingües. En el capítulo 2 de la tesis, se aborda el tema de la localización web en mayor profundidad. A continuación presentamos tan solo una breve introducción para argumentar, más tarde, la justificación de nuestro objeto de estudio (véase la sección 2).

1.3.1 El modelo GILT

En una sociedad digital cada vez más globalizada, el multilingüismo de la web no ha pasado desapercibido. Atrás quedan ya los primeros días de la era de Internet, cuando los usuarios tenían que leer el contenido en inglés porque apenas había otras alternativas. En la actualidad, no obstante, los sitios se adaptan a los idiomas hablados por los consumidores de información y servicios digitales de todo el mundo. Este fenómeno tiene su origen en el paradigma de la globalización, entendida como «la transformación de procesos y negocios para satisfacer las necesidades de los clientes, sea cual sea su idioma, país o cultura»[159] (LISA en Jiménez Crespo 2013, 25). La globalización, por su parte, desencadena una sucesión de procesos interrelacionados a los que se suele referir bajo el acrónimo GILT (Cadieux y Esselink 2002): la internacionalización, la localización y la traducción de productos digitales, como los sitios web.

Pym (2014, 119) afirma que internacionalizar consiste en «generalizar un producto digital de manera que se pueda adaptar más fácilmente a diversos idiomas y convenciones culturales sin tener que replantearse su diseño inicial». Cuando la internacionalización forma parte integral del proceso de desarrollo de un sitio web (y no se deja para las últimas fases del mismo) contribuye a facilitar la tarea de localización. La idea que se plantea es la de generar un sitio suficientemente «neutro» para que, durante el proceso de localización, se pueda modificar de forma sencilla y adecuarse, así, a las necesidades de los usuarios de cada *locale*[160] (Pym 2011, 413). Algunas de las buenas prácticas de internacionalización consisten en páginas web compatibles con distintas normas de codificación de caracteres, o facilitar la identificación y modificación de cadenas de texto traducibles (por

[159] El texto entrecomillado en esta página es una traducción propia.

[160] En el campo de la localización, el término *locale* hace referencia a un grupo de personas que comparten el mismo idioma y convenciones culturales, pero que no necesariamente se encuentran el la misma ubicación física. De ahí la necesidad de definir parámetros regionales por pares de lengua y país al configurar productos digitales (Cadieux y Esselink 2002). Por ejemplo, francés-Francia (fr-FR) francés-Suiza (fr-CH), y francés-Canadá (fr-CA) son tres *locales* diferentes.

ejemplo, evitando textos incrustados no editables en las imágenes) (Esselink 2006, 23).

Tras la internacionalización del producto web, la fase de localización abarcaría no sólo la traducción del contenido textual, sino también la adaptación de otros elementos no textuales (por ejemplo: imágenes, colores, estructura) para satisfacer las necesidades del público meta (Fernández Costales 2009). Aunque también es habitual que una persona se encargue, de forma autónoma, de todo el proceso de localización (Gouadec 2007, 43), la localización se considera en el modelo GILT como una fase independiente de la traducción en sí, e incluye la preparación y gestión del contenido web que se debe traducir, las tareas de ingeniería en la fase posterior a la traducción (como la edición de imágenes o la adaptación de código HTML) y la etapa de control de calidad final (Jiménez Crespo 2013, 26). Con todo, la separación de las tareas de traducción y localización depende, en última instancia, de un conjunto de factores mucho más complejos, como el volumen y el alcance del proyecto, su complejidad técnica, y los recursos humanos y económicos disponibles para llevarlo a cabo.

Esta tesis se aleja de la estructura jerárquica y lineal impuesta originalmente por el modelo GILT (Montalt i Resurrecció 2003) en dos frentes. En primer lugar, entendemos que la traducción y la ingeniería de localización forman, en su conjunto, una labor continua y entrelazada que puede ser realizada por una sola persona, el localizador, del que se espera que cuente con las competencias lingüísticas, técnicas e instrumentales necesarias para conseguir un producto web meta funcional (desde un punto de vista técnico, cultural y lingüístico). Este solapamiento entre las tareas de traducción y localización, que ha sido estudiado por muchos investigadores en nuestro campo (Sandrini 2005; Nauert 2007; Fernández Costales 2009; Schäler 2010), especialmente al tratar el carácter multimodal de la web, también constituye un eje central en nuestra investigación. En segundo lugar, siguiendo la propuesta de Jiménez Crespo (2013, 27), consideramos que el ciclo GILT debería ser más interactivo, y promover la comunicación y el intercambio de conocimientos entre el localizador y los diferentes actores que participan en el proceso de desarrollo de sitios web.

1.3.2 La localización web como un proceso centrado en el usuario

En lugar de considerar la localización web como una actividad motivada por intereses puramente económicos con el único objetivo de llegar a los mercados internacionales, como era el caso en los años 80 (Dunne 2015), creemos que el contexto de recepción y el usuario final son dos aspectos fundamentales. En los estudios de traducción, esta visión de la localización centrada en los destinatarios del producto meta se ha abordado a menudo desde un enfoque funcionalista (Jiménez Crespo 2009b; Costales Fernández 2009). El funcionalismo se aleja de la estricta noción de equivalencia y de importancia prestada tradicionalmente al texto origen. Por otra parte, las teorías funcionalistas establecen que el objetivo o *Skopos* del acto de traducción es crucial, y consideran que una traducción es funcional cuando tiene en cuenta las expectativas, las necesidades, los conocimientos previos y las circunstancias situacionales del receptor (Nord 1997). En nuestro trabajo, el concepto de funcionalidad también está relacionado con el entorno del usuario final, pero se amplía para abarcar, además, la eficacia de la interacción entre el usuario y el contenido web desde un punto de vista técnico.

En este sentido, adoptamos la definición de localización propuesta por Sandrini (2008, 9), que la describe como «el proceso de modificar un sitio web existente para que sea accesible, usable y culturalmente adecuado para un determinado público meta»[161]. Haciendo especial hincapié en el usuario final, Sandrini apunta que el objetivo de toda tarea de localización debe consistir en que las personas de un determinado *locale* pueden utilizar el producto localizado en su propio idioma sin ningún tipo de dificultad (ibid). En el marco de esta tesis, entendemos que los usuarios de un mismo *locale* comparten el mismo idioma y siguen convenciones culturales similares, pero no necesariamente tienen las mismas capacidades físicas y sensoriales, las cuales ejercen una influencia en su modo de interacción con la web. Asimismo, consideramos que las dificultades con las que puedan encontrarse los usuarios al navegar por un sitio localizado pueden deberse no sólo a problemas en términos de adecuación y corrección lingüística y cultural, sino también a obstáculos relacionados con la funcionalidad del sitio que el comisario de la tarea no pudo identificar en el producto origen y/o que el localizador no pudo (o supo) corregir en el producto meta. Si trasladamos esta reflexión a nuestro objeto de estudio principal (las imágenes y sus textos alternativos), dichas dificultades podrían aparecer si una página web contiene imágenes localizadas (i) que no tienen un atributo `alt`, (ii) cuyos textos alternativos no han sido traducidos, o (iii) cuyos textos alternativos no se adecúan al valor comunicativo de las imágenes a las que acompañan.

2. Justificación del objeto de estudio

Tras quince años desde la publicación de la primera versión de las WCAG, los estudios demuestran que, a día de hoy, muy pocos sitios registran un nivel alto de conformidad con las pautas de accesibilidad internacionales (Lopes et al. 2010; Harper y Chen 2012; Power et al. 2012). Si bien a lo largo de las dos últimas décadas, la web ha experimentado mejoras en términos de accesibilidad, los investigadores creen que no todas esas mejoras se deben a un interés más pronunciado en la aplicación de mejores prácticas de accesibilidad, sino más bien a (i) los cambios en la codificación de estilos para lograr un mayor grado de consistencia y compatibilidad entre diferentes dispositivos, así como a (ii) las

[161] Traducción propia.

nuevas funcionalidades y características de los navegadores para mejorar la presentación y el diseño web (Richards et al. 2012; Hanson y Richards 2013). El estudio bibliográfico revela que los webmasters, los desarrolladores y los diseñadores web son conscientes de los beneficios para la sociedad en general que puede aportar la adopción de buenas prácticas de accesibilidad (Yesilada al. 2012; Putnam et al. 2012). Sin embargo, reconocen que no suelen tener el tiempo ni incluso la formación necesarios para aplicarlas (Lazar et al. 2004; Trewin et al. 2010). Del mismo modo, algunos afirman que las pautas de accesibilidad para el contenido web son difíciles de entender y las herramientas para la evaluación de la accesibilidad web (herramientas WAE) no ofrecen una ayuda suficiente (ibid).

Con esta tesis, se pretende comprender si estos obstáculos también se dan en el caso de los sitios web multilingües, y si los localizadores deben participar activamente en la creación de contenido web más accesible. A nuestro modo de ver, la adopción de mecanismos de accesibilidad debería formar parte integral de todas las fases del ciclo de vida de un producto web (Cooper et al. 2012), desde el diseño y el desarrollo del sitio, hasta su operacionalización, mantenimiento y localización. En nuestro trabajo, abordamos esta cuestión primero desde una perspectiva general, examinando las medidas que se toman en la actualidad para desarrollar sitios multilingües accesibles (véase la sección 2.1) para, a continuación, centrarnos en un aspecto más concreto: la accesibilidad de las imágenes y su localización (véase la sección 2.2).

2.1 Accesibilidad de los sitios web multilingües

El análisis de los procesos de ejecución y evaluación de la accesibilidad de sitios web multilingües no ha recibido hasta la fecha demasiada atención en la literatura, a pesar del hecho de que, como apunta Folaron (2012, 25), la web es ahora un «espacio de traducción» por excelencia, en donde las prácticas de traducción desempeñan un papel fundamental para el mantenimiento de redes locales y mundiales al facilitar una comunicación fluida entre los usuarios.

El W3C considera que los sitios web que están disponibles en varias versiones y éstas son independientes entre sí (por ejemplo, los sitios web en dos idiomas y con URL diferentes) la evaluación de la accesibilidad de las mismas debe realizarse de forma individual (Velleman y Abou-Zahra 2014). Se entiende, por tanto, que en el caso de sitios localizados o sitios culturalmente personalizados (Singh y Pereira 2005), los desarrolladores del producto origen no son necesariamente responsables de la accesibilidad del producto meta. De la misma manera, una de las pocas referencias que se hacen en las WCAG 2.0 a la accesibilidad web en un contexto multilingüe es la posibilidad de reconocer que un sitio (o página) web es conforme a las pautas de accesibilidad parcialmente cuando el contenido solo ha sido verificado en una de las lenguas disponibles (Caldwell et al. 2008). Por lo demás, e independientemente de la labor desempeñada por la *Internationalisation (I18n)*

Activity del W3C para fomentar el principio de una web para todos, no se ha encontrado ningún documento ni grupo oficial que aborde, de forma explícita, como facilitar el acceso a contenido multilingüe en la web para personas con discapacidad. Por otro lado, hemos constatado que, hasta la fecha, no se ha llevado a cabo ningún estudio empírico para investigar este tema.

En el campo de la localización, se ha apuntado a la accesibilidad web como uno de los paradigmas que forman parte de la interdisciplinariedad de los llamados *Localisation Studies* (véase la figura 1.5 en el capítulo 1 de la tesis). Algunos autores ya han resaltado los aspectos que tienen en común la localización y la accesibilidad desde una perspectiva teórica, y han afirmado que, cuando un sitio web inicialmente accesible se transforma en un sitio multilingüe, los localizadores deben cerciorarse de que los mecanismos de accesibilidad adoptados en el producto origen se mantienen en todas las versiones lingüísticas del sitio final (Gutiérrez y Restrepo y Martínez Normand 2010; Tercedor Sánchez 2010). Además, algunos investigadores han sugerido que la versión localizada nunca debería ser menos accesible que la original, y que dependiendo del grado de libertad otorgado al equipo de localización, la versión localizada podría ser incluso más funcional que la original siempre y cuando los cambios realizados se lleven a cabo para satisfacer las necesidades de un mayor número de usuarios (ibid).

Pym (2011, 424) considera que la accesibilidad es uno de los problemas éticos a los que la localización debe hacer frente hoy en día, tanto en la esfera profesional como en la académica. La necesidad (y la obligación) del localizador de solucionar los obstáculos que dificultan la IPO ya ha sido reconocida por algunos investigadores de este campo de investigación (Tercedor Sánchez 2010; Jiménez Crespo 2009a), pero sólo unos pocos han apostado por introducir la accesibilidad en los procesos de control de calidad (Gibb y Matthaiakis 2007; Jiménez Crespo 2013). Tras constatar el éxito de una serie de seminarios sobre accesibilidad web impartidos en varias universidades (Rodríguez Vázquez 2014), con esta tesis perseguimos aumentar la sensibilización sobre cuestiones relacionadas con la accesibilidad en el seno de comunidad de localización web, donde todavía se registran niveles muy bajos (Ó Broin 2004).

2.2 Accesibilidad y localización de imágenes

La importancia de la web para la difusión de la información es incuestionable, pero la proliferación de contenido no textual ha puesto a las personas con discapacidad visual en una situación de desventaja (Harper y Chen 2012). Los usuarios ciegos experimentan angustia y frustración cuando navegan por la web debido a las barreras de accesibilidad, que les hacen perder, en promedio, un 30,4% del tiempo que pasan delante del ordenador (Lazar et al. 2007). Una de las principales causas de la frustración en este grupo de población es la falta de accesibilidad de las imágenes (Paciello 2000; Asakawa 2005; Petrie et al. 2005;

Lazar et al. 2007). La investigación presentada en esta tesis refleja nuestro deseo de contribuir a la eliminación de dicha barrera, y viene motivada por (i) los problemas a los que todavía se enfrentan los profesionales de la web para formular textos alternativos pertinentes, y (ii) la escasez de estudios empíricos que se centren específicamente en cómo se aborda esta cuestión durante el proceso de localización web.

2.2.1 Elaboración de textos alternativos

A pesar de que garantizar la existencia de alternativas textuales para todo contenido no textual resulta de vital importancia para que las personas ciegas pueden participar con éxito en la sociedad de la información, estudios recientes apuntan a unos índices de conformidad relativamente bajos en lo que respecta a la accesibilidad de imágenes (Acces for all 2011; Hanson y Richards 2013). Aunque la proporción de imágenes sin atributo `alt` se ha ido reduciendo gradualmente a lo largo de los últimos años (Asakawa 2005; Richards et al. 2012), el mayor problema reside ahora en la presencia de textos alternativos no informativos (ibid). Creemos que esto podría deberse a tres problemas en concreto:

Falta de pautas detalladas sobre cómo escribir textos alternativos pertinentes

Introducir un atributo `alt` en un elemento `<img>` es, a priori, un paso fácil de ejecutar. No obstante, la elaboración de un texto alternativo pertinente puede llegar a resultar una tarea complicada. Según el W3C, el valor del atributo `alt` debe comunicar el objetivo de la imagen o explicar su significado; en otras palabras, los textos alternativos deben substituir a las imágenes, y no incluir simplemente una descripción sencilla de su contenido. La formulación de textos alternativos de calidad requiere no sólo conocimientos lingüísticos y una buena capacidad de síntesis, sino también una importante inversión de tiempo. Las recomendaciones que existen sobre qué tipo de información incluir en el atributo `alt` (es decir, acerca de la composición de textos alternativos) forman parte, por lo general, de documentos oficiales cuya asimilación puede resultar difícil para los profesionales de la web con un perfil más técnico; por ejemplo, la norma ISO/TS 20071-11 (Organización Internacional de Normalización 2012b). Por otra parte, las orientaciones lingüísticas de las que disponen en la actualidad para escribir textos alternativos (es decir, recomendaciones acerca de su formulación) son muy generales y apenas se hace referencia alguna a preferencias de tipo léxico y/o sintáctico.

Limitaciones de las herramientas utilizadas para evaluar la accesibilidad de las imágenes

Cuando los creadores de contenido no tienen el tiempo o la formación suficientes para poner en práctica las pautas existentes para la accesibilización de las imágenes, estos suelen apoyarse en herramientas de evaluación para encontrar una solución rápida o simplemente para compensar la falta de conocimientos necesarios. Cuando se trata de comprobar que un sitio web satisface la pauta 1.1 de las WCAG 2.0, las herramientas WAE son capaces de identificar los elementos `<img>` que no tienen atributo `alt`, pero en raras ocasiones proporcionan información relativa a la pertinencia de los textos alternativos presentes en la página. Los mensajes de advertencia relacionados con la accesibilidad de las imágenes son a menudo demasiado vagos (por ejemplo, «Asegúrese de que el texto alternativo transmite el contenido de la imagen») y no son considerados de utilidad (Petrie et al. 2005). Si bien se han llevado a cabo estudios para tratar de ofrecer alternativas más eficaces a estas herramientas (Bigham et al. 2006; Bigham 2007; Olsen et al. 2010), la mayoría de ellos se ha centrado en la detección de textos alternativos no informativos (cadenas de caracteres alfanuméricos, extensiones de archivo, información sobre el tamaño del archivo) en lugar proponer una solución para mejorar la calidad de los textos alternativos formulados con construcciones típicas de lenguaje natural.

Falta de acuerdo en cuanto a la responsabilidad de los creadores de contenido en materia de accesibilidad de imágenes

Independientemente de la falta de recomendaciones exhaustivas para la correcta formulación de textos alternativos y las deficiencias de las herramientas de evaluación actuales para evaluar su pertinencia, hay una tercera barrera que dificulta la consecución de altos niveles de accesibilidad en las imágenes: no existe un acuerdo con respecto a quién debe ser responsable de proporcionar textos alternativos de calidad para las imágenes contenidas en un sitio web. Son muchos los profesionales que participan en las actividades llevadas a cabo a lo largo de todo el ciclo de vida de un sitio web: comisarios, ingenieros, diseñadores gráficos y editores de contenido, entre otros. En teoría, son los desarrolladores quienes deberían asegurarse de que existe un mecanismo para introducir un texto alternativo (el atributo `alt`), mientras que los editores, por su parte, tendrían que proponer una formulación adecuada. En la práctica, sin embargo, seguir un flujo de trabajo tan estricto resulta, cuanto menos, complicado, ya que las imágenes pueden añadirse, modificarse o eliminarse durante las diversas etapas del ciclo de desarrollo web. Esta situación se complica todavía más, si cabe, cuando se solicita la localización de un sitio web, dado que, por norma general, se duplica en número de actores implicados en el ciclo de producción.

2.2.2 Traducción y adaptación de textos alternativos

El valor del atributo `alt` suele considerarse como una cadena de texto que debe ser traducida durante la tarea de localización web (Gibb y Matthaiakis 2007, 668; Mata Pastor 2009b, 552; Roturier 2015, 88). La importancia de este recurso para facilitar una mejor experiencia de navegación para los usuarios ciegos ha sido reconocida incluso por expertos en el campo de la accesibilidad (Clark 2002, 99). Sin embargo, el estudio bibliográfico realizado indica que sólo dos estudios de localización han investigado la presencia de textos alternativos para las imágenes en sitios web localizados.

Siguiendo un enfoque descriptivo, Jiménez Crespo (2008) analizó un corpus paralelo español compuesto por sitios web corporativos originales (N=172) y localizados (N=95). El autor indicó en su trabajo que el sub-corpus localizado contenían un mayor número de atributos `alt`, alegando que esto se debía posiblemente al uso de sistemas de gestión de contenidos (CMS), que suelen insertar dichos atributos automáticamente en la página. Además de presentar el promedio de palabras por texto alternativo, la presencia de caracteres no alfabéticos (espacios, *, /) y el predominio de formas nominales, no se aportaron datos adicionales en cuanto a la composición y formulación de los textos alternativos. Asimismo, Fernández Costales (2010) llevó a cabo un estudio a gran escala para investigar la localización de sitios web de universidades de 27 países de la Unión Europea. Tras realizar un análisis cualitativo sobre el contenido textual y no textual (incluidos los aspectos culturales) de los mismos, concluyó que la traducción de las alternativas textuales para las imágenes recibía poca atención durante el proceso de localización de este tipo de sitios institucionales.

Si bien estos dos estudios constituyen dos valiosas contribuciones para nuestro campo, en ninguno de ellos se abordó de forma exhaustiva la cuestión de la pertinencia de los textos alternativos, probablemente porque el análisis de estos elementos no constituía su objeto de estudio principal. En los estudios de traducción, las consideraciones relativas a la composición y formulación de textos alternativos se han examinado con relación a los textos técnicos (Prieto Velasco, 2009). Aunque algunas de las pautas definidas se extrapolaron más tarde al caso particular de las imágenes en la web (Tercedor Sánchez Prieto Velasco y 2009, 84), la investigación de estos autores no incluyó ningún tipo de estudio destinado a examinar, por ejemplo, si los localizadores eran capaces de poner en práctica estas recomendaciones o si su aplicación se podía automatizar por medio de herramientas destinadas al control de calidad.

3. Objetivos de la tesis

Si se tiene en cuenta todo lo expuesto hasta ahora, se infiere que el objetivo general de esta tesis ha sido el de explorar el papel del localizador en la accesibilización de sitios web multilingües. El alcance de nuestra investigación se ha delimitado estableciendo tres objetivos más específicos relacionados con el proceso de evaluación de la accesibilidad. Estos objetivos se presentan a continuación en orden creciente de importancia:

Objetivo 1: Determinar si existen procedimientos ya establecidos para garantizar la accesibilidad de los sitios web multilingües en general.

Objetivo 2: Proponer una solución, basada en una herramienta de lenguaje controlado (LC), para facilitar la evaluación de la calidad de los textos alternativos para las imágenes durante el proceso de localización, a fin de cubrir algunas de las limitaciones observadas en las recomendaciones existentes para su formulación, así como las deficiencias de las herramientas utilizadas actualmente para dicho fin.

Objetivo 3: Determinar la medida en que los localizadores son capaces de conseguir un nivel aceptable de accesibilidad en el producto web meta en lo que respecta a las imágenes.

En nuestra investigación, el punto de partida (o hipótesis inicial) es que la evaluación de la accesibilidad debe formar parte del proceso de control de calidad de toda tarea de localización web para así garantizar la creación de un producto meta funcional. Tomando como referencia la definición de Saldanha y O'Brien (2014, 95), por control de calidad (CC) entendemos la suma de todos los sistemas y procesos utilizados para ayudar a conseguir o mantener la calidad de un producto. En este sentido, el proceso de CC abarcaría toda acción emprendida para asegurar (de forma prospectiva) y evaluar (a posteriori) la accesibilidad. Es importante indicar que en esta tesis no se han analizado en profundidad los modelos de control de calidad utilizados en la industria de la localización ni se ha elaborado un estudio teórico sobre la noción de calidad, tan controvertida en los estudios de traducción. A través de nuestro trabajo, hemos tratado más bien de mostrar, por medio de datos empíricos (véase la sección 4), que ciertos aspectos relativos a la accesibilidad, como la pertinencia de los textos alternativos localizados, pueden llegar a ser considerados por los usuarios finales como un factor determinante de la calidad del producto web meta.

4. Metodología y preguntas de investigación

Para alcanzar los tres objetivos señalados en el apartado anterior, hemos adoptado un enfoque empírico multimetódico, combinando estrategias de investigación de carácter descriptivo y experimental. En el campo de la IPO, los estudios descriptivos conforman a menudo la primera etapa de un proyecto de investigación, pues permiten a los investigadores examinar lo que ocurre en un área de estudio en concreto, y son la base sobre las que se asientan las siguientes etapas del trabajo. Los estudios experimentales, por su parte, facilitan el examen de relaciones de causalidad (Lazar et al. 2010, 20-22). Esta tesis se basa en un enfoque principalmente descriptivo para cumplir los dos primeros objetivos, y en un enfoque experimental para lograr la consecución del objetivo 3, el cual constituye el eje central de nuestro trabajo de investigación. La figura 3 ilustra la triangulación metodológica adoptada en esta tesis.

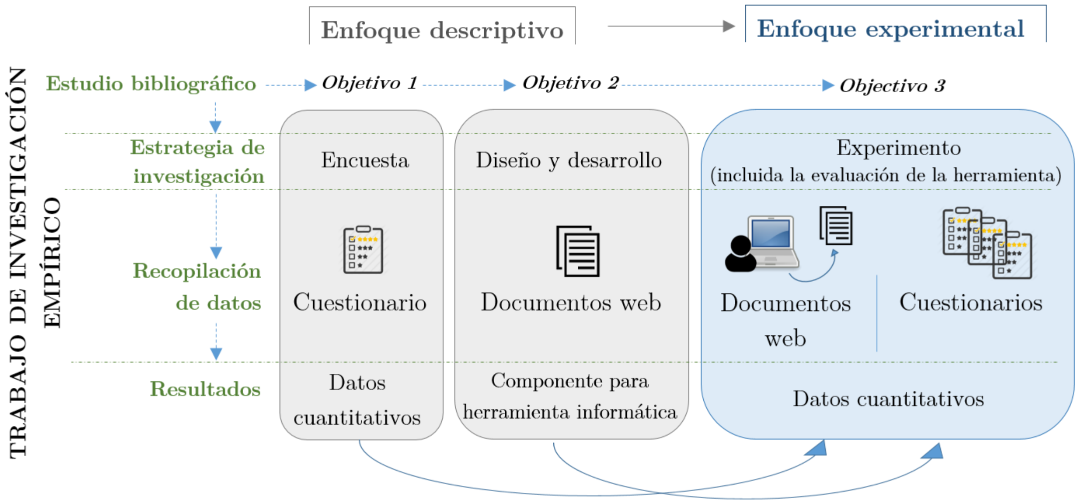


Figura 3. Marco metodológico de nuestro trabajo de investigación

Para lograr el **objetivo 1,** se llevó a cabo un estudio bibliográfico de los trabajos existentes sobre la aplicación y la evaluación de pautas de accesibilidad web, así como de los pasos más importantes del proceso de localización web. Esta revisión de la literatura se complementó con una encuesta realizada a 67 expertos en AW, con el fin de obtener información acerca de (i) los procedimientos que siguen a la hora de evaluar la accesibilidad de los sitios web multilingües, y (ii) su opinión acerca de la potencial contribución de los localizadores a la consecución de sitios web multilingües más accesibles. Los datos cuantitativos sobre estos dos aspectos se obtuvieron a través de un cuestionario en línea.

Guiados por nuestro **objetivo 2,** y tras un examen exhaustivo de los estudios emprendidos hasta la fecha en lo que respecta a la accesibilidad de las imágenes, hemos desarrollado 40 reglas de lenguaje controlado (LC), siguiendo un formalismo declarativo (Bredenkamp et al. 2000), para facilitar la identificación de textos alternativos en francés no pertinentes durante el proceso de localización. La definición de las reglas, que se fundamentó en el análisis de dos corpus (1. las recomendaciones existentes sobre la formulación de textos alternativos, y 2. un corpus web compilado por la fundación Access for All), se llevó a cabo utilizando la tecnología de Acrolinx, uno de los programas informáticos líderes para el control de calidad que, además, ofrece un cliente para aplicarlas en archivos HTML. Al crear un componente para una herramienta ya existente, hemos seguido una estrategia de investigación que en el campo de la informática y de los sistemas de información que se conoce como «diseño y desarrollo» (Oates 2005, 107). La evaluación Acrolinx y las reglas se incluyó en la estrategia metodológica seguida para alcanzar el **objetivo 3.**

Con el fin de determinar la medida en que los localizadores son capaces de garantizar la accesibilidad de las imágenes en el producto web meta (Objetivo 3), se llevó a cabo un estudio experimental, dividido en dos partes: (i) un experimento sobre localización web (fase 1) en el cual participaron 28 localizadores (14 de los cuales tenían conocimientos básicos sobre accesibilidad); y (ii) una evaluación con usuarios finales (fase 2), que contó con la participación de siete personas ciegas (véase la figura 4). El estudio tuvo como objetivo responder a las siguientes preguntas de investigación:

R1. ¿Los textos alternativos de las imágenes son considerados por los localizadores como elementos traducibles durante el proceso de localización web?

R2. ¿Facilita el uso de herramientas de control de calidad destinadas a la evaluación de la accesibilidad la creación de textos alternativos más pertinentes durante el proceso de localización web?

R3. ¿Tener conocimientos básicos sobre accesibilidad web ayuda a los localizadores a crear textos alternativos más pertinentes durante el proceso de localización web?

A lo largo del experimento, se pidió a los participantes que localizasen, en primer lugar, un sitio web compuesto por tres páginas con 130 imágenes, y a continuación, que evaluasen la accesibilidad de las imágenes del sitio localizado utilizando dos herramientas de control de calidad: (i) Acrolinx, la herramienta de evaluación semi-automática que facilita la aplicación de las reglas de LC que desarrollamos, y (ii) aDesigner, una herramienta WAE general (Asakawa 2005). A fin de garantizar la validez de la investigación, el orden en el que se utilizaron las herramientas no fue el mismo para todos los participantes. Por tanto, se propusieron dos escenarios diferentes de CC: aDesigner-Acrolinx (escenario A) y Acrolinx-aDesigner (escenario B). En una segunda fase, siete usuarios de lectores de pantalla evaluaron la pertinencia de los textos alternativos producidos por los localizadores a través de un cuestionario en línea.

Con el fin de responder a las preguntas de investigación expuestas anteriormente, se analizaron los datos cuantitativos recopilados durante las dos fases del estudio para estimar el efecto de tener conocimientos sobre AW y de usar herramientas de CC especiales para evaluar la accesibilidad (variables independiente (VI) principales) en la traducción y la pertinencia de los textos alternativos del producto web meta (variables dependientes (VD)) (véase la figura 4). La influencia de otras variables independientes secundarias en el producto localizado, tales como el escenario de CC seguido por los participantes, también se tuvo en cuenta durante el procesamiento y análisis de los datos[162]. En el capítulo 5 de la tesis, se explica de forma detallada la metodología adoptada para alcanzar el **objetivo 3**. Las hipótesis

[162] Además del escenario de CC, se investigaron otras seis variables independientes secundarias. La tabla 6.1 del capítulo 6 de la tesis contiene una lista completa de todas las variables analizadas en este trabajo de investigación.

y sub-hipótesis asociadas a nuestras preguntas de investigación están recogidas en ese mismo capítulo, en la sección 5.4.

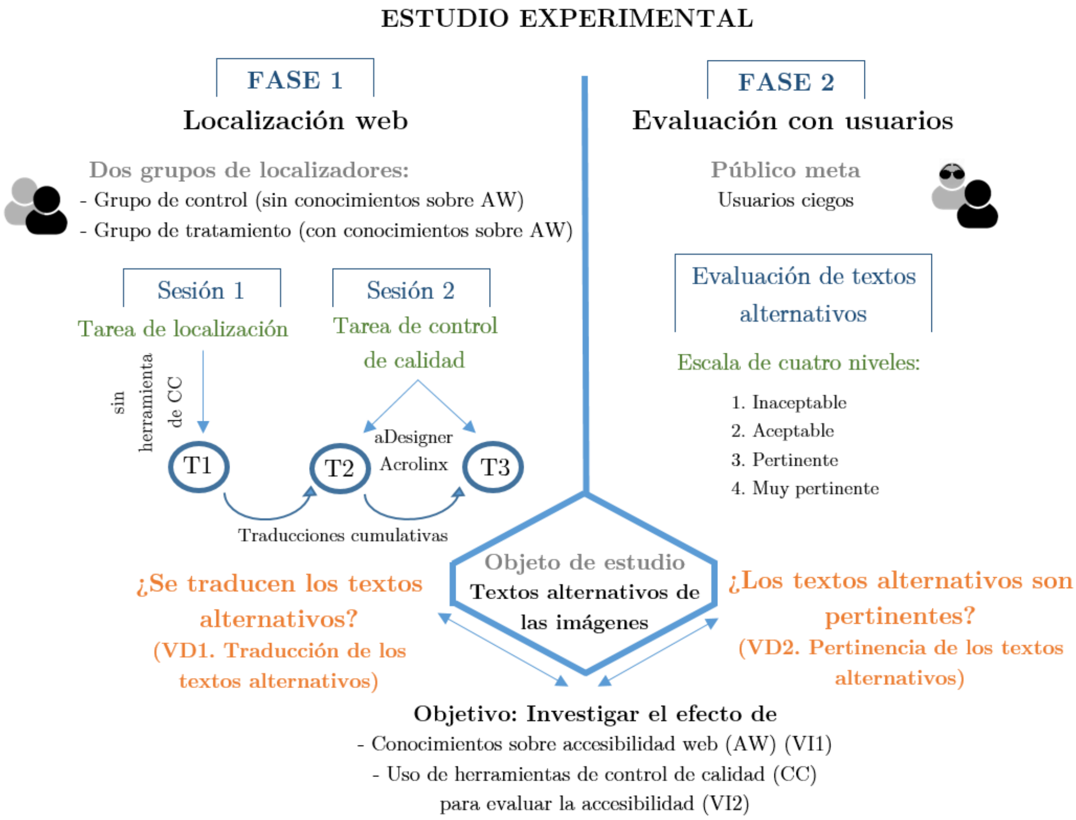


Figura 4. Resumen esquemático del estudio experimental (fases, variables, participantes, objetivos)

5. Resultados

En términos generales, el análisis de todos los datos recogidos nos ha ayudado a comprender mejor cómo y por qué la accesibilidad debe tenerse en cuenta durante proceso de localización web. Esta es, de hecho, una de las principales aportaciones de nuestra investigación, la cual constituye el primer trabajo empírico en el que se han examinado de forma más exhaustiva los puntos de conexión entre la localización accesibilidad. Consideramos que, a pesar de las limitaciones de nuestro trabajo (véase la sección 7), hemos logrado cumplir nuestros objetivos. A continuación se presentan los resultados principales obtenidos, se ponen de relieve sus implicaciones prácticas y se exponen las contribuciones realizadas a los diferentes campos de investigación o áreas de estudio que han delimitado el marco teórico de esta tesis (véase la sección 6).

Objetivo 1

Los resultados relacionados con el Objetivo 1 se han conseguido a partir de un estudio bibliográfico y de la encuesta presentada en el capítulo 2 de la tesis (véase la sección 2.5 del mismo).

- *El W3C no propone una metodología o recomendaciones normalizadas para garantizar la accesibilidad de los sitios web multilingües.*

Las recomendaciones que existen en la literatura sobre la forma de aplicar buenas prácticas en materia de accesibilidad o evaluar el contenido web para ver si se éstas se han seguido correctamente no ofrecen ningún tipo de consejo sobre cómo proceder en el caso de sitios web localizados. El W3C es el único organismo que hace alusión a sitios web multilingües, indicando que si las diferentes versiones lingüísticas de los mismos se pueden utilizar independientemente (por ejemplo, cuando disponen de dos URL diferentes), cada una de ellas debería ser analizada de forma individual durante el proceso de evaluación de la accesibilidad. Sin embargo, ningún documento oficial aporta información sobre quién debería hacerlo o en qué pautas habría que basarse para ello.

- *Los expertos en AW no siguen un procedimiento estándar durante la evaluación de la accesibilidad de sitios web multilingües.*

Los profesionales de la Web con experiencia en cuestiones de accesibilidad consideran que la solución ideal para garantizar que los sitios web multilingües sean accesibles es asegurarse de que cada versión lingüística registra un nivel de conformidad mínimo con las pautas de accesibilidad establecidas por el W3C. No obstante, reconocen que, en el día a día, esto casi nunca sucede debido a la falta de tiempo, recursos humanos y competencias lingüísticas adicionales. En consecuencia, y aunque no es lo más conveniente, la práctica habitual consiste en centrarse únicamente en un idioma específico y asumir que en el resto de las versiones del sitio se ha conseguido un nivel similar de accesibilidad. Por otra parte, los datos recogidos apuntan a que, durante las evaluaciones de AW, no se suele prestar demasiada atención a aspectos culturales ni al contenido textual, a pesar de su relevancia en el caso de los sitios web localizados.

- *Los expertos en AW coinciden en que crear contenido accesible debería ser uno de los objetivos principales de toda tarea de localización web.*

Los profesionales que participan en el ciclo de desarrollo web parecen estar convencidos de que los localizadores deben asumir la responsabilidad de crear contenidos accesibles cuando desempeñen su labor, al igual que los desarrolladores y diseñadores web, e incluso en un grado mayor que los webmasters o gestores de contenido web. Del mismo modo, la comunidad de AW considera que el perfil tan completo del localizador puede ser de especial ayuda durante la última fase de evaluación de la accesibilidad en sitios web multilingües.

Objetivo 2

En relación con el Objetivo 2, se resumen los resultados de la estrategia de diseño y desarrollo adoptada en el presente trabajo (véase el capítulo 4 de la tesis), así como las conclusiones derivadas de su evaluación.

- *Utilizando la tecnología de Acrolinx, se han desarrollado 40 reglas de LC para identificar textos alternativos no pertinentes.*

Tomando como referencia ejemplos concretos de textos alternativos, hemos definido tres grupos de reglas para garantizar la pertinencia de dichos elementos en sitios web localizados al francés. En concreto, nos hemos centrado en la formalización de patrones utilizados frecuentemente en textos alternativos para imágenes con un valor descriptivo (grupo A, de 14 reglas; véase la sección 4.4.1 del capítulo 4) y un valor funcional (grupo B, de 18 reglas; véase la sección 4.4.2 del mismo capítulo). Además, hemos desarrollado ocho reglas para identificar textos alternativos no informativos (grupo C, de 8 reglas; véase la sección 4.4.3 del mismo capítulo). Estas reglas ayudan a evitar el uso de ciertos términos, y proponen estructuras sintácticas específicas de acuerdo con el valor comunicativo de la imagen. Las reglas se formalizaron en el entorno de desarrollo de Acrolinx, que también ofrece un cliente para automatizar su aplicación en el caso de archivos HTML: Acrolinx IQ Batch Checker (2.7.0).

- *Cada regla definida viene acompañada de documentación relevante para ayudar a los localizadores a entender los problemas detectados gracias a la misma y cómo se podrían solucionar.*

Cuando los usuarios utilizan Acrolinx y nuestro conjunto de reglas para evaluar la accesibilidad de las imágenes contenidas en archivos HTML, se genera automáticamente un informe de errores. A través de dicho informe, se puede consultar la documentación para cada regla, la cual facilita la interpretación de los problemas identificados por la herramienta. Cada archivo de ayuda incluye una breve introducción sobre la accesibilidad de las imágenes en la Web, una explicación del error señalado, ejemplos que ilustran dicha explicación, una serie de recomendaciones para mejorar la calidad del texto alternativo marcado como erróneo. La documentación de las reglas también contiene enlaces a materiales proporcionados por el W3C en materia de accesibilidad. En la figura 4.4 del capítulo 4 de la tesis se muestra un ejemplo de archivo de ayuda.

- *Ocho de las diez reglas de LC evaluadas facilitan la elaboración de textos alternativos significativamente más pertinentes durante el proceso de localización.*

Cuando el sitio web origen contiene textos alternativos de baja calidad, la probabilidad de elaborar un texto alternativo pertinente para la versión localizada del sitio es entre dos y tres veces mayor cuando se aplican las reglas que hemos desarrollado. En particular, las reglas destinadas a mejorar la calidad de los textos alternativos para las imágenes funcionales, a través de los cuales se debe describir la acción que desencadena la interacción entre la imagen y el usuario, permiten conseguir resultados significativamente mejores en cuanto a la pertinencia del valor del atributo `alt`. El efecto significativo de las reglas es más notable cuando Acrolinx es la única herramienta que se utiliza para mejorar la accesibilidad de las imágenes; no obstante, cuando las reglas se aplican tras haber utilizado otra herramienta (en este caso, aDesigner), la significatividad de los resultados se mantiene. Dos de las diez reglas que no han resultado ser muy eficaces están relacionadas con (i) la introducción de información sobre los derechos de autor de las imágenes en el atributo `alt`, y (ii) el uso de expresiones redundantes como «imagen de un ... » o «La imagen muestra ... ».

- *En términos generales, una herramienta de CC basada en reglas de LC como Acrolinx puede proporcionar un mejor apoyo para la consecución de imágenes accesibles en sitios web localizados que una herramienta WAE general como aDesigner.*

Nuestra propuesta de adoptar un enfoque basado en herramientas de lenguaje controlado para mejorar la accesibilidad de las imágenes perseguía dar respuesta a algunas de las limitaciones observadas en las herramientas utilizadas a día de hoy para dicho fin, como la imposibilidad de distinguir textos alternativos relevantes de aquellos que no lo son, o la de verificar cadenas de texto en idiomas distintos del inglés. El estudio llevado a cabo ha confirmado que, efectivamente, aplicar dicho enfoque durante el proceso de localización web puede dar lugar a la producción de textos alternativos significativamente más pertinentes que cuando se utiliza una herramienta general de evaluación de la accesibilidad. Además, los datos recopilados demuestran que utilizar Acrolinx no solo es ventajoso a la hora de identificar textos alternativos de baja calidad, sino también para guiar a los localizadores sobre cómo aumentar el grado de pertinencia de dichos elementos. De hecho, los resultados de la evaluación con usuarios indican que cuando los localizadores consultaron la documentación de las reglas que hemos presentado anteriormente en relación con el Objetivo 2, la calidad de los textos alternativos en sus sitios web aumentó de forma significativa.

Objetivo 3

El objetivo principal de este trabajo se ha alcanzado dando respuesta a tres preguntas de investigación específicas. A continuación, resumimos (i) las principales conclusiones extraídas de la evaluación de nuestras hipótesis, descritas en la tabla

6.22 (véase el capítulo 6 de la tesis), y (ii) los resultados de los análisis adicionales presentados en el capítulo 7.

R1. *¿Los textos alternativos de las imágenes son considerados por los localizadores como elementos traducibles durante el proceso de localización web?*

Contrariamente a lo esperado, hemos constatado que los localizadores no traducen los textos alternativos de forma sistemática. Gracias a los datos cuantitativos recopilados durante el experimento y los datos subjetivos recogidos a través de los cuestionarios administrados a los participantes, hemos observado un notable grado de variabilidad con respecto a la importancia que se le da a la existencia de estos elementos y a la necesidad de traducirlos. Resulta de interés señalar que los localizadores con conocimientos básicos sobre AW fueron capaces de identificar y traducir un número de textos alternativos significativamente mayor que los que no habían recibido una formación previa en materia de accesibilidad. Además, el estudio experimental realizado revela que los localizadores no se sienten cómodos cuando se trata de introducir atributos `alt` en elementos `<img>` que no tenían un texto alternativo asociado en el documento origen. En lo que al tratamiento de los atributos `alt` vacíos se refiere, parece que lo localizadores tienen una mayor tendencia a proponer nuevos atributos vacíos para señalar el valor decorativo de las imágenes (es decir, convertir textos alternativos presentes en el sitio web original en atributos `alt` vacíos) que a reconsiderar la pertinencia de aquellos atributos de estas características que ya estaban en el documento origen.

Asimismo, nuestro estudio también ha demostrado que el uso de herramientas de CC destinadas a la evaluación de la accesibilidad facilita la identificación de los textos alternativos como elementos traducibles. Hemos observado que los localizadores que no habían editado ningún texto alternativo en la primera versión de traducción que entregaron fueron capaces de detectar luego su presencia durante la tarea de CC. Concretamente, hemos constatado que, en términos generales, Acrolinx dio lugar a un mayor número de cambios en los textos alternativos que aDesigner. Para localizadores sin ningún conocimiento previo sobre la AW, la primera resultó ser, de hecho, mucho más eficaz que la segunda en este sentido. Por otro lado, los resultados indican que, para este grupo de participantes, el uso combinado de ambas herramientas ha facilitado la realización de cambios en el doble de textos alternativos con respecto a la primera versión de traducción producida durante el ejercicio de localización.

R2. *¿Facilita el uso de herramientas de control de calidad destinadas a la evaluación de la accesibilidad la creación de textos alternativos más pertinentes durante el proceso de localización web?*

Una de las principales conclusiones de nuestro trabajo de investigación es que el uso de herramientas de CC puede ayudar a localizadores a conseguir un nivel de accesibilidad en el producto meta significativamente mayor en lo que concierne a las imágenes que el nivel que se obtendría si dichas herramientas no estuviesen disponibles. Los datos recogidos revelan que esto ocurre no sólo cuando se combinan dos herramientas con funcionalidades complementarias, sino también cuando los localizadores cuentan solo con una de ellas (es decir, Acrolinx o aDesigner). La elaboración de un análisis más detallado nos ha permitido concluir que utilizar un programa informático como Acrolinx, basado en la tecnología más puntera para el procesamiento del lenguaje natural (PLN), es mucho más eficaz que utilizar una herramienta WAE como aDesigner cuando solo una de ellas está a disposición de los localizadores. Además, hemos comprobado que aDesigner da lugar a resultados significativos con respecto a la calidad de los textos alternativos únicamente cuando se utiliza después de Acrolinx. Por ello, consideramos que la mejor solución para asegurar la pertinencia de estos elementos en la versión localizada de un sitio web es utilizar Acrolinx en primer lugar, a fin de solucionar cuestiones de adecuación lingüística, y a continuación usar aDesigner, el cual permite al localizador identificar rápidamente los elementos `<img>` que no cuentan con un atributo `alt`, así como aquellos con un texto alternativo compuesto por cadenas de caracteres de un modo no informativo.

Por otra parte, hemos observado que el uso de herramientas de CC para asegurar la accesibilidad de las imágenes es eficaz a la hora de mejorar la calidad de los textos alternativos que en el documento origen no eran pertinentes, pero cuando las imágenes ya cuentan con un texto alternativo original aceptable, dichas herramientas no contribuyen a conseguir un nivel más alto de pertinencia ni tampoco a reducirlo. Del mismo modo, y a diferencia de las herramientas de CC, conviene destacar que el uso de herramientas de traducción asistida por ordenador (TAO) no ejerció ninguna influencia significativa sobre el nivel de accesibilidad de las imágenes en el producto web meta.

R3. *¿Tener conocimientos básicos sobre accesibilidad web ayuda a los localizadores a crear textos alternativos más pertinentes durante el proceso de localización web?*

Al igual que en el caso de la traducción de los textos alternativos, se ha constatado que tener conocimientos básicos en materia de AW da lugar a resultados significativamente mejores, esta vez en términos de calidad. Nuestros datos indican que el hecho de haber recibido una hora de formación sobre buenas prácticas de accesibilidad ha tenido un efecto significativo en la pertinencia de los textos alternativos producidos por los localizadores. Con todo, uno de los hallazgos más sorprendentes de nuestro estudio es que esta ventaja sobre los localizadores sin conocimientos previos sobre el tema desaparece cuando se utilizan las herramientas adecuadas. De hecho, nuestro estudio demostró que cuando los dos grupos de localizadores siguieron las recomendaciones proporcionadas por Acrolinx, las diferencias registradas entre los mismos en cuanto a la pertinencia de los textos alternativos creados no fueron significativas. Además, hemos constatado que el escenario de CC Acrolinx-aDesigner ha sido especialmente eficaz para los localizadores no familiarizados con los problemas de accesibilidad web. Por último, conviene apuntar que, durante el análisis de los datos sobre la pertinencia de los textos alternativos, también tuvimos en cuenta los conocimientos de los localizadores sobre cuestiones de HTML. Al comparar los datos de los participantes, se observó que, de hecho, tener un buen dominio de este lenguaje de marcado ayuda a los localizadores a producir textos equivalentes significativamente más relevantes.

6. Contribuciones realizadas a los diferentes campos de investigación

Si se tienen en cuenta los resultados descritos en la sección anterior, así como las conclusiones adicionales presentadas gradualmente a lo largo de la tesis, podríamos decir que hemos podido confirmar nuestra hipótesis inicial: tanto las opiniones personales de los profesionales de la web (diseñadores, desarrolladores, consultores de accesibilidad, localizadores) como los datos objetivos que hemos recopilado a través de nuestro estudio experimental indican que los localizadores deberían participar activamente en la consecución de sitios web multilingües más accesibles. A un nivel más específico, nuestra investigación ha realizado diversas contribuciones a los campos de la accesibilidad web, el lenguaje controlado y la localización web.

Accesibilidad web

En esta tesis se ha seguido la tendencia creciente observada en el campo de la accesibilidad a estudiar cuáles son los métodos más eficaces y sostenibles para garantizar una Web más inclusiva para todos. La novedad de nuestra investigación radica en el hecho de que (i) se ha estudiado el comportamiento de un grupo de profesionales que, hasta la fecha, nunca había sido considerado como relevante para la consecución de dicho objetivo universal, y de que (ii) se ha demostrado que las acciones de estos profesionales pueden, de hecho, ejercer una influencia significativa en el nivel de accesibilidad general de la Web. Asimismo, conviene destacar que hemos abierto una nueva línea de debate dentro de la comunidad científica interesada en el campo de la AW al traer a un primer plano la distinción entre sitios monolingües y multilingües. Desde el punto de vista de la accesibilidad, consideramos que los segundos plantean muchos más retos que los primeros, no sólo porque el contenido se presenta en varios idiomas y puede haber variaciones entre las distintas versiones lingüísticas, sino también porque su producción implica la participación de mayor número de actores, con diferentes competencias y percepciones acerca de las necesidades de las personas con discapacidad. Nuestros hallazgos con respecto a la falta de procedimientos normalizados para garantizar la accesibilidad de los sitios web multilingües han puesto de relieve la necesidad apremiante de desarrollar nuevos métodos de evaluación para este tipo de sitios web en particular.

En lo que se refiere a la accesibilidad de imágenes, nuestro trabajo puede tener varias aplicaciones prácticas. En primer lugar, hemos demostrado que la ineficacia de las herramientas de evaluación utilizadas actualmente puede solucionarse recurriendo a las características y funcionalidades de algunos programas informáticos utilizados en otras áreas de especialización, tales como el PNL. En este sentido, tomando como referencia las ventajas identificadas en relación con el uso de Acrolinx (y su combinación con herramientas como aDesigner), se podría sugerir la integración de herramientas basadas en LC en los procesos de evaluación de la accesibilidad. En segundo lugar, el corpus utilizado para el desarrollo de las reglas de CL también nos ha ayudado a comprobar que, de hecho, los textos alternativos con un alto nivel de pertinencia siguen a menudo patrones similares, por lo que es posible extraer recomendaciones de tipo lingüístico sobre cómo escribir textos alternativos de calidad. Aunque se han identificado únicamente los patrones más comunes en francés, creemos que los resultados de nuestra investigación podrían servir de gran ayuda para creadores de contenido en otros idiomas, ya que dan pistas sobre las formulaciones preferidas por los usuarios ciegos en función del valor comunicativo de la imagen.

Lenguaje controlado

Creemos que nuestro trabajo también es innovador desde el punto de vista de la investigación en lenguaje controlado. Constituye el primer estudio centrado en la aplicación de software de LC con fines de accesibilidad web, y los resultados en este sentido parecen prometedores. En el caso particular de la lengua francesa, hemos demostrado que ciertos aspectos de accesibilidad se pueden formalizar mediante la creación de reglas de LC relativamente simples. Esto podría atraer la atención de los investigadores en este ámbito hacia el campo de la AW, quienes podrían examinar si es posible abordar otros problemas de accesibilidad siguiendo un enfoque similar. Por otra parte, en los estudios de localización, la aplicación de LC se ha centrado principalmente en distintas dimensiones de la calidad del producto meta a nivel textual, como la legibilidad o su traducibilidad. Con nuestro estudio, hemos demostrado que controlar la producción de cadenas de texto también puede facilitar el acceso a la información visual contenida en un documento web.

Localización web

Por último, creemos que nuestro trabajo de investigación puede tener implicaciones directas en la definición de buenas prácticas de localización en el futuro. Al tratarse del primer estudio científico que muestra cómo se abordan cuestiones de accesibilidad durante el proceso de localización, hemos podido observar que la familiarización de los profesionales que participan en dicho proceso acerca de las necesidades de las personas con discapacidad sigue siendo relativamente baja en esta comunidad. La aplicación de mecanismos para promover la accesibilidad no se considera todavía como una obligación o una necesidad, tal y como hemos demostrado con el caso especial de los textos alternativos, lo que puede tener consecuencias negativas para de la accesibilidad global de la Web. En concreto, nuestros resultados ponen de manifiesto la necesidad de capacitar a los localizadores para que participen de manera más activa en el ciclo de desarrollo web. Parece que no tienen la suficiente confianza como para realizar cambios en el código fuente, incluso cuando esos cambios pueden facilitar el acceso al contenido web por parte de un mayor número de usuarios.

Por otro lado, nuestra investigación ha realizado contribuciones al campo de los estudios de localización, al presentar el primer trabajo empírico en el que se presentaron documentos HTML a los participantes como producto origen, y el primer estudio en involucrar a usuarios finales con discapacidad en calidad de evaluadores o expertos para verificar la calidad del producto meta. Además, al señalar la necesidad de prestar más atención a los problemas de accesibilidad durante el proceso de localización (algo que también ha sido reconocido por los expertos de la comunidad de accesibilidad y que se refleja en los resultados de nuestro estudio empírico), se han hecho reflexiones interesantes sobre la accesibilidad como un potencial indicador de la calidad del producto localizado. Esto podría tener un impacto significativo en tres aspectos específicos del campo: (i) p*roceso de CC en localización:* nuestros resultados indican que la evaluación de la accesibilidad debería formar parte de la cadena de CC en localización, basada tanto en métodos de validación automáticos como manuales; (*ii*) *las competencias generales del localizador:* nuestro trabajo ha puesto de manifiesto que, además de las habilidades instrumentales, técnicas y lingüísticas tradicionales, otro tipo de competencias pueden resultar especialmente útiles en la consecución de sitios web plenamente funcionales, tales como conocer las pautas de accesibilidad web y las ayudas técnicas utilizadas por las personas con discapacidad para interactuar con la web; y *(iii) la formación en localización:* nuestra investigación ha demostrado los efectos positivos de conocer nociones básicas en materia de accesibilidad y el uso de herramientas de evaluación de la accesibilidad; esto sugiere que la introducción de aspectos relacionados con la AW en la enseñanza de la localización podría presentar un valor añadido para el perfil de las nuevas generaciones de localizadores.

7. Limitaciones del trabajo realizado y futuras vías de investigación

Dado que se trata del primer estudio sobre la accesibilidad de los sitios web multilingües que intentó examinar los vínculos entre las tres áreas de especialización mencionadas anteriormente, el alcance de esta tesis se ha visto limitado en lo que respecta al tipo de contenido web analizado y el número de lenguas utilizadas. En este sentido, nuestro objeto central de estudio ha sido la accesibilidad de las imágenes, y solo hemos hecho referencia ocasionalmente a ejemplos relativos a otros elementos web de interés tanto para la localización como la accesibilidad. En cuanto a los idiomas, si bien en los capítulos más teóricos se abordaron las diferentes áreas de estudio de manera general, nuestro trabajo empírico (el desarrollo de reglas de LC y el estudio experimental) giró entorno a la combinación inglés-francés.

A lo largo de la tesis, hemos tratado de describir las medidas adoptadas para compensar estas limitaciones generales, especialmente para fomentar la validez interna y ecológica de nuestros estudios, así como para asegurar su replicabilidad. A continuación se presentan las principales limitaciones de los enfoques adoptados para lograr nuestros tres objetivos. Esta última parte del resumen servirá también para presentar posibles vías de investigación para futuros trabajos.

El **objetivo 1** se abordó desde una perspectiva relativamnte amplia. El hecho de que nos basamos fundamentalmente en el estudio de la literatura y en las opiniones personales de los expertos en accesibilidad web podría hacer que nuestras conclusiones sean menos generalizables. Teniendo en cuenta nuestros resultados preliminares, se podrían llevar a cabo otros estudios de observación directa para examinar más en detalle cómo evalúan estos profesionales el nivel de accesibilidad del contenido web al trabajar con sitios multilingües. También sería interesante comparar las políticas y la legislación relativas a cuestiones relacionadas con la accesibilidad web en países plurilingües y monolingües, con el fin de comprender si existen métodos o directrices internas para garantizar que la información de interés público sea accesible a los hablantes de todas las lenguas utilizadas en el territorio nacional, independientemente de su diversidad funcional. Nuestros resultados plantean, asimismo, cuestiones importantes con respecto a la situación actual de sitios web multilingües en cuanto a la conformidad con las pautas WCAG 2.0. En este sentido, se deberían llevar a cabo estudios adicionales para averiguar si cada versión lingüística tiene un nivel de accesibilidad diferente, así como para estimar el impacto potencial que estas diferencias podrían tener, por ejemplo, en la experiencia de navegación de usuarios finales bilingües con algún tipo de discapacidad. A pesar de estas limitaciones, consideramos que este primer estudio sobre la accesibilidad de sitios multilingües puede servir como un interesante punto de partida para futuras investigaciones en el ámbito de la accesibilidad y la localización.

Para lograr el **objetivo 2,** se tuvieron que asumir también ciertas limitaciones. La primera de ellas está relacionada con las características del corpus web elegido para el desarrollo de reglas de LC. Somos conscientes de que el tipo y la cantidad de las imágenes usadas en un sitio web pueden variar mucho entre los diferentes géneros textuales. En este sentido, la formalización de las recomendaciones lingüísticas sobre cómo escribir textos alternativos más pertinentes podría ser haber resultado más complicada si, por ejemplo, se hubiesen tomado como referencia sitios web de turismo o comerciales en lugar de sitios web institucionales, especialmente porque, por lo general, estos últimos contienen más imágenes funcionales, mientras que los primeros cuentan con un mayor número de imágenes con un valor descriptivo. Para garantizar la portabilidad de las reglas que hemos definido, futuros estudios podrían evaluar si éstas pueden ser útiles para evaluar la accesibilidad de las imágenes en otros tipos de sitios web. En cuanto al método de evaluación adoptado, conviene destacar que otra limitación de nuestro estudio ha sido que la eficacia de las reglas solo se ha evaluado teniendo en cuenta la calidad del producto final localizado. Creemos que un análisis más exhaustivo de los datos podría dar lugar a resultados interesantes con respecto a la usabilidad de dichas reglas en función del perfil de la persona que las aplica. Del mismo modo, los informes de errores generados automáticamente por Acrolinx y los vídeos de las sesiones de localización y CC del estudio experimental podrían proporcionar nuevas pistas sobre las estructuras lingüísticas que no hemos cubierto en nuestras reglas y que podrían haber sido propuestas espontáneamente por los localizadores. Por último, pero no por ello menos importante, sería pertinente considerar si Acrolinx y las reglas son igualmente eficaces si son utilizados por otros profesionales con un perfil más técnico, como los desarrolladores o diseñadores de sitios web.

Finalmente, conviene destacar otra serie de limitaciones al interpretar los resultados relacionados con **el objetivo 3.** Algunas de ellas están relacionadas con el diseño del propio estudio experimental. Por ejemplo, hemos desarrollado un sitio experimental fundamentalmente de carácter estático, con el fin de contar con documentos HTML sencillos y así la evaluación del mayor número de casos problemáticos posible relativos a la accesibilidad de las imágenes. Asimismo, dada la gran cantidad de datos recogidos, se decidió adoptar un enfoque principalmente cuantitativo para analizar la traducción y la pertinencia de los textos alternativos. A pesar de que, al mismo tiempo, esto constituye uno de los aspectos más destacados de nuestro trabajo, puesto que hemos aplicado análisis estadísticos inferenciales avanzados que nos permitieron hacer conclusiones sobre la significatividad de los resultados, sería aconsejable considerar también aspectos cualitativos en análisis futuros. Esto nos permitiría ampliar nuestras conclusiones con respecto al nivel de accesibilidad de las imágenes obtenidas en el producto web meta.

El perfil de los localizadores que participaron en el experimento también se podría considerar como una limitación de nuestra investigación, debido a que la mayoría de ellos trabajaba como autónomos (freelance). Se podría pensar que los equipos de localización de las grandes empresas están más especializados y prestan más atención a los problemas de accesibilidad. En este sentido, es posible que si expertos en AW con un perfil de localización hubiesen participado en el estudio, nuestras conclusiones con respecto a la importancia de tener conocimientos sobre AW podrían haber sido diferentes. Del mismo modo, el impacto positivo de la utilización de herramientas de CC para verificar la accesibilidad podría haber sido menos concluyente. Sin embargo, creemos que nuestro estudio y sus conclusiones siguen siendo pertinentes, sobre todo si tenemos en cuenta la realidad del sector de la localización en la actualidad, donde la tendencia es a externacionalizar cada vez más los servicios. Además, nuestro protocolo experimental también nos ha permitido descubrir el potencial didáctico de las herramientas utilizadas, lo que podría ser útil para aumentar la sensibilización de los localizadores sobre cuestiones de accesibilidad durante las fases de CC.

Esto está directamente relacionado con otra limitación de nuestro estudio que, al mismo tiempo, alberga un gran potencial para futuras investigaciones: la poca atención prestada en general a los datos sobre el proceso de localización recogidos. En nuestra opinión, los vídeos constituirán una valiosa fuente de información para continuar nuestra investigación sobre localización y accesibilidad. Basándonos en estos datos (por ejemplo, el tiempo invertido en inspeccionar el sitio durante la etapa de traducción o las estrategias adoptadas para resolver los problemas del código fuente), podríamos tratar de determinar cómo las buenas prácticas de accesibilidad podrían integrarse de forma más sencilla en el proceso de localización, evitando esfuerzos innecesarios de rediseño en las últimas etapas del ciclo de desarrollo web multilingüe. Las conclusiones derivadas del análisis de estos registros podrían entonces ser comparados, por ejemplo, con otros métodos típicamente utilizados en la investigación del proceso de traducción, tales como el seguimiento ocular.

Esperamos que estos análisis adicionales también ayudarán a definir mejor futuros modelos de formación en accesibilidad para profesionales de la localización. Si bien la importancia de adquirir conocimientos sobre AW fue bien recibida por los estudiantes que participaron en los seminarios sobre el tema impartidos por la investigadora en diferentes universidades (Vázquez Rodríguez 2014), la selección de materiales no se ha basado hasta ahora en estudios científicos. Además, la enseñanza de la accesibilidad web como un módulo independiente se considera a menudo como una técnica ineficaz. Consideramos que la enseñanza de la accesibilidad debe ser integrada gradualmente a lo largo del programa, como ya se ha hecho en cursos de informática y ciencias de la información (Waller et al. 2009; Wang 2012). En este sentido, en el futuro, nuestra investigación sobre la enseñanza de la accesibilidad podría basarse en el marco didáctico que actualmente estamos desarrollando, el cual aboga por un enfoque comunicativo y social en el aprendizaje de tareas de localización (Torres del Rey et al. 2014):

> "Localisation students must learn to play a three-fold role, vis-a-vis the digital product and the production process, as: **mediators** of the communicative value generated by the digital product as a cultural object and as a technical extension of human physical and intellectual abilities; **stakeholders** in the

distribution of the localisable object's functional and informative values; and **negotiators** of their own (as professionals) and of their localisable technological product's social position and responsibility."

Appendix M. Résumé étendu de la thèse en français

1. Contexte de la thèse

Les technologies de l'information et de la communication (TIC) imprègnent presque tous les aspects de notre vie quotidienne. Elles ont façonné la manière dont les connaissances sont produites et partagées au sein de contextes multiples et variés, y compris les sphères professionnelle, politique, éducative, économique et sociale. Dans cette époque de technologie où l'alphabétisation numérique est devenue la norme (Folaron 2012), le World Wide Web (le Web) est désormais considéré comme un produit indispensable. Comme le dit son inventeur, le pouvoir du Web découle du fait que des documents sont mis en ligne de manière standardisée et sont reliés entre eux, et « the universality and flexibility of this linking architecture has a unique capacity to break down boundaries of distance, language, and domains of knowledge » (Berners-Lee 2007).

Le World Wide Web Consortium (W3C) est l'organisme international qui élabore des protocoles et des règles pour assurer l'évolution du Web à long terme en suivant sa vision de *One Web* (« Un Web »), qui, à son tour, dépend de deux grands principes de conception web : *Web on Everything* (« le Web partout ») et *Web for all* (« le Web pour tous »). Le premier est fondé sur la promotion du développement des technologies qui permettent l'accès au Web partout, à tout moment, en utilisant n'importe quel dispositif ; le deuxième vise à poursuivre le plein potentiel du Web en tant que construction sociale. L'objectif du paradigme du « Web pour tous » est que les informations contenues sur le Web soient accessibles à toutes les personnes, quel que soit le matériel et le logiciel qu'ils utilisent, leur langue ou leur culture d'origine, leur localisation géographique ou leurs modes sensoriels ou d'interaction (qui, en même temps, peuvent dépendre de la capacité physique ou mentale de chaque personne).

Au cours des trente dernières années, les progrès réalisés dans le domaine des technologies pour les personnes handicapées visant à faciliter l'interaction homme-machine (IHM) (voir section 1.1), ainsi que la prolifération des normes sur l'accessibilité web (AW) (voir section 1.2) et la demande croissante de contenus web internationalisés et localisés (voir section 1.3) ont été cruciaux pour la création de sites web plus ouverts à tous. En adoptant une perspective interdisciplinaire, cette thèse repose sur ces trois aspects importants de la société de l'information pour plaider en faveur d'un niveau plus élevé de sensibilisation à l'accessibilité pendant la production de sites web multilingues et, plus particulièrement, pendant la modification de contenu graphique qui se produit tout au long du processus de localisation web (voir section 2).

1.1 L'accès au Web par les personnes en situation de handicap

Le handicap fait partie de la condition humaine. Selon l'Organisation mondiale de la santé (OMS), presque tout individu vivra avec une forme ou une autre de handicap à un moment donné dans sa vie (OMS 2011, 3). Cette affirmation est fondée sur le fait que l'on comprend le handicap non seulement comme une caractéristique de la personne, directement causée par une maladie, un traumatisme ou un autre problème de santé, mais aussi comme un problème social qui n'est pas un attribut inhérent à l'individu (OMS 2002, 8-9), mais qui dépend plutôt de l'environnement global dans lequel il vit. Lorsque cette définition est appliquée au contexte de l'IHM, il est possible de dire que les utilisateurs sont handicapés non seulement à cause des modalités restreintes par lesquelles ils interagissent avec les contenus web, mais aussi à cause de la technologie utilisée à la base pour transmettre cette information et des technologies dont ils dépendent pour la percevoir (Harper et Yesilada 2008a, xv). Dans le cadre de cette thèse, nous nous intéressons particulièrement à l'interaction entre la technologie et les personnes qui présentent une déficience visuelle.

1.1.1 Le Web et les personnes avec un handicap visuel

Selon l'OMS, la fonction visuelle peut être classée en quatre catégories : vision normale, déficience visuelle modérée, déficience visuelle grave et cécité. Le deuxième et le troisième niveaux sont souvent regroupés sous le terme de « baisse de la vision », qui, avec la cécité, représente l'ensemble des déficiences visuelles (OMS 2014). Dans le monde, il y a 39 millions de personnes qui sont aveugles, et 246 millions présentent une baisse de l’acuité visuelle (ibid). En Espagne, le dernier rapport publié sur cette question indique que près d'un million de personnes sont enregistrées comme ayant une déficience visuelle, parmi lesquelles 71 000 sont aveugles (Ernst & Young 2012, 28). Proportionnellement, les chiffres en Suisse sont relativement élevés aussi : sur les 325 000 personnes qui sont atteintes d’un handicap visuel dans le pays, environ 10 000 sont aveugles (UCBA 2012, 4).

Avant les années 1980, les principales ressources d'information pour les personnes aveugles étaient limitées à quelques livres : certains étaient en papier braille et d’autres enregistrés sur cassettes audio (Asakawa 2014). Avec l'avènement des ordinateurs personnels comme « moyen structuré de communication dynamique » (Winograd et Flores 1987, 176) et du Web comme « univers d'information globale » (Berners-Lee 1992), les personnes avec un handicap visuel, en particulier les aveugles, sont devenues plus autonomes : les utilisateurs aveugles peuvent maintenant effectuer eux-mêmes, sans l'aide d'une personne voyante, des activités quotidiennes, comme faire des achats en ligne, lire des journaux, effectuer des transactions bancaires ou simplement communiquer plus facilement avec leurs amis. Aujourd'hui, le Web est pour beaucoup d'entre eux une source essentielle pour s'informer, travailler et se divertir (Harper et Yesilada 2008a, 1).

Tout cela a été possible grâce aux efforts déployés par l'industrie et le milieu académique depuis le début des années 1990 pour faciliter l'accès des utilisateurs aveugles à des ordinateurs et au Web. Le Web étant un moyen visuel par excellence, le défi était de proposer aux utilisateurs aveugles une expérience semblable à celle des personnes qui pouvaient voir l'écran, en leur permettant de naviguer de manière intuitive à travers les pages web et de profiter d'Internet, mais sans utiliser les yeux (Asakawa 2005). Dans ce contexte, la plupart des approches ont visé à présenter l'information montrée sur l'écran à travers le canal auditif et à travers des dispositifs tactiles (Barreto 2008, 10). Tous ces efforts ont finalement abouti à deux innovations technologiques importantes : la digitalisation du langage braille et l'accès au Web par synthèse vocale (Asakawa 2014).

1.1.2 Les technologies d'assistance utilisées par les utilisateurs aveugles

Le terme « technologie d'assistance (TA) » est utilisé dans le cadre des études d'IHM pour désigner tout logiciel ou matériel destiné à faciliter l'utilisation des ordinateurs par des personnes qui sont touchées d'un handicap (RDC 2004, 1) et qui ne peuvent pas utiliser les technologies conventionnelles nécessaires pour avoir accès à une interface graphique (GUI), par exemple, une souris ou un écran. Les fonctionnalités fournies par les technologies d'assistance comprennent des présentations de remplacement (par exemple de la synthèse vocale ou du contenu agrandi), des méthodes de saisie alternatives (par exemple la voix), des mécanismes de navigation ou d'orientation supplémentaires, et des transformations de contenu (Caldwell et al. 2008).

Les difficultés rencontrées par les personnes ayant une baisse de la vision en ce qui concerne l'accès au Web sont différentes de celles des utilisateurs aveugles. Les malvoyants doivent chercher des techniques pour adapter l'information visuelle à leur niveau de vision, tandis que les aveugles ont besoin d'alternatives non-visuelles au contenu affiché à l'écran (Edwards 2008, 150-151). Les utilisateurs ayant des limitations de la vision utilisent normalement des logiciels qui augmentent la taille du texte ou des images, communément appelés agrandisseurs d'écran (Paciello 2000, 71). Les personnes aveugles, quant à elles, utilisent une plus grande variété de TA dont des afficheurs braille, des navigateurs vocaux et des lecteurs d'écran, entre autres.

Les lecteurs d'écran sont les TA les plus populaires parmi les utilisateurs aveugles. Ces logiciels retranscrivent par synthèse vocale ce qui est affiché sur l'écran d'un ordinateur tant en termes de contenu que de structure, en fournissant aux utilisateurs un certain nombre de fonctionnalités pour accélérer et simplifier la recherche d'informations (Barreto 2008, 10). La conception visuelle (« visual design ») du Web repose sur le fait que les utilisateurs sans déficience visuelle peuvent choisir à tout moment la partie du texte sur laquelle se concentrer (Edwards 2008, 154). Les lecteurs d'écran remplacent cette fonctionnalité via une interface de raccourcis qui permet aux utilisateurs de naviguer entre les différents éléments de la page de manière séquentielle. Par exemple, l'utilisateur peut demander, par le biais d'une commande clavier, d'ouvrir une fenêtre auxiliaire contenant une liste de tous les liens ou les titres présents sur la page (Borodin et al. 2010), en lui permettant ainsi d'obtenir un modèle mental ou « hors écran » de la page en question (Theofanos et Redish 2006). Par conséquent, pour les utilisateurs aveugles, le fait de pouvoir surfer sur le Web est entièrement dépendant de la bonne conception de la page, qui doit avoir une structure sémantique pertinente (Connor 2012, 37). La robustesse de la page dépendra, à son tour, des capacités de son créateur (ou les capacités de l'outil qu'il utilise) d'appliquer correctement les bonnes pratiques de conception web. L'interdépendance et la complémentarité entre les différents éléments de la chaîne de développement web seront abordées dans la section 1.2.

1.1.3 L'accès non visuel au contenu graphique des sites web

Les capacités humaines en termes d'acuité visuelle, de sensibilité au contraste et de champ de vision, ainsi que la performance accrue des technologies ont encouragé les développeurs à exploiter pleinement le Web en tant qu'interface visuellement très riche (Barreto 2008, 8). Une étude menée par Asakawa (2005) a révélé que le nombre d'images sur le Web a quadruplé entre 1996 et 2005, un fait confirmé plus tard par Chen et Harper (2008), qui ont observé un accroissement significatif de l'utilisation de formats graphiques dans les documents web sur une période de dix ans (1999-2008). Dans un Web de plus en plus peuplé de contenus graphiques, le processus de lecture d'écran pose des défis.

Les lecteurs d'écran convertissent un document conçu en deux dimensions en une chaîne de texte unidimensionnelle, en transformant l'interface graphique traditionnelle en une interface utilisateur sonore, où l'audio est le moyen dominant (Freitas 2010, 273). Cette « linéarisation » de la page web (Thatcher et al. 2006, 105), – qui implique un changement de canal, visuel en auditif –, peut entraîner une perte d'information considérable si le contenu graphique présenté à l'utilisateur n'a pas d'équivalent textuel qui remplit la même fonction. La précision et l'efficacité de l'acte de communication entre l'utilisateur aveugle et l'ordinateur dépendent donc (i) de l'existence d'un tel équivalent textuel, et (ii) de la capacité du lecteur d'écran de l'identifier en tant que tel et de transmettre les informations correspondantes à l'utilisateur.

Pour que les personnes aveugles puissent percevoir le contenu graphique comme, par exemple, des images, celles-ci doivent être accompagnées d'un équivalent. L'attribut `alt` du langage HTML, introduit pour la première fois en 1995 dans le cadre de la spécification HTML 2.0 afin de fournir cette fonctionnalité (Berners-Lee et Connolly 1995), est la technique la plus largement adoptée pour rendre les images accessibles sur le Web. Les images sont habituellement insérées dans une page web grâce à l'élément `<img>`. Quand un lecteur d'écran trouve cet élément, le logiciel signale la présence de l'image et lit la valeur de l'attribut `alt` correspondant. Par exemple, si un utilisateur aveugle tombait sur l'image représentée à la figure 1, il entendrait les informations suivantes : « Graphique : La plage de Rodas au sable blanc, à Galice, Espagne, un jour d'été avec un ciel bleu ». Si l'attribut `alt` n'était pas utilisé, le lecteur d'écran indiquerait toujours que l'image existe, mais aucun équivalent textuel ne serait fourni, ce qui aurait un effet négatif sur l'expérience de navigation. De même, l'interaction entre le Web et l'utilisateur aveugle serait de moins bonne qualité si l'équivalent textuel n'était pas pertinent, c'est-à-dire, s'il ne représentait pas ce qui est vraiment montré sur l'image.

```
<body>
...
<img
src="myweb-en/images/praia-rodas-
cies.jpg" alt=" La plage de Rodas au
sable blanc, à Galice, Espagne, un
jour d'été avec un ciel bleu.">
...
</body>
```

Figure 1. Présentation visuelle et non visuelle (langage HTML) d'une image.

Dans cette thèse, les équivalents textuels sont considérés comme un élément essentiel pour obtenir un haut niveau d'accessibilité en ce qui concerne les images et ils forment notre objet d'étude principal (voir section 2.2). Le sujet des équivalents textuels (composition, formulation, évaluation) est abordé plus en détail dans le chapitre 3 de la thèse.

1.2 Conformité aux normes d'accessibilité web

Comprendre les besoins des personnes handicapées et la façon dont elles interagissent avec le Web est crucial pour la conception de sites web accessibles, mais cela n'est pas toujours suffisant pour garantir une interaction homme-ordinateur agréable et efficace pour ce groupe de la population. L'accessibilité du Web est également déterminée par des aspects techniques, tels que la conformité avec les spécifications techniques établies par le W3C et les bonnes pratiques d'accessibilité recommandées par cette organisation. Celles-ci sont l'objectif principal des activités menées par la *Web Accessibility Initiative* du W3C (WAI), lancée en 1997 et visant à élaborer des règles et des ressources pour rendre le Web accessible.

Le terme « accessibilité web » signifie que les personnes handicapées peuvent percevoir, comprendre, naviguer et interagir avec le Web (Lawton Henry 2005b). Selon la WAI, l'accessibilité web dépend de l'interaction entre différents composants techniques et humains (Chisholm et Henry 2005, 32), comme illustré sur la figure 2. D'une part, les composants techniques comprennent les agents utilisateurs (navigateurs, technologies d'assistance, etc.), les outils d'édition et d'évaluation utilisés par les professionnels impliqués dans le cycle de production du Web, et les spécifications techniques qui décrivent les caractéristiques des langages qui sont utilisés pour créer des contenus Web. D'autre part, les composants humains englobent non seulement les utilisateurs finaux, mais aussi les producteurs de contenu : ceux qui conçoivent, codent, créent et modifient des contenus Web (ibid). Thatcher et al. (2006, 14-15) utilisent le cas des équivalents textuels des images pour illustrer les interdépendances entre tous ces composants, résumées ci-dessous :

- Les spécifications techniques fournissent des normes pour utiliser un langage de balisage permettant l'introduction des équivalents textuels pour les images (l'attribut `alt`).
- Les outils d'édition permettent de faciliter et de promouvoir l'utilisation de l'attribut `alt`.
- Les outils d'évaluation permettent aux créateurs de contenu de vérifier que les attributs `alt` ont été bien introduits et les aident à déterminer s'ils sont pertinents.
- Les agents utilisateurs fournissent des interfaces homme-machine pour présenter les équivalents textuels.
- Les producteurs de contenu formulent des équivalents textuels et les introduisent via l'attribut `alt`, souvent en utilisant des outils d'édition et d'évaluation.
- Les utilisateurs perçoivent les équivalents textuels grâce à leurs technologies d'assistance et à un navigateur.

Quand un mécanisme d'amélioration de l'accessibilité, tel que l'attribut `alt`, est bien implémenté par un composant, les autres composants sont plus susceptibles de l'adopter eux aussi (ibid). Afin de promouvoir cette vision et d'améliorer l'interaction entre tous les composants mentionnés ci-dessus, la WAI a publié, entre les années 1999 et 2000, une série de règles d'accessibilité sur la base des spécifications techniques fondamentales du Web : des Règles pour l'accessibilité des outils d'édition (ATAG), des Règles pour l'accessibilité des agents utilisateurs (UAAG) et des Règles pour l'accessibilité des contenus web (WCAG). Dans le cadre de cette thèse, nous nous intéressons à ce dernier document (WCAG), étant donné qu'il prescrit des manières de rendre des contenus web accessibles (y compris les images) pour les personnes handicapées.

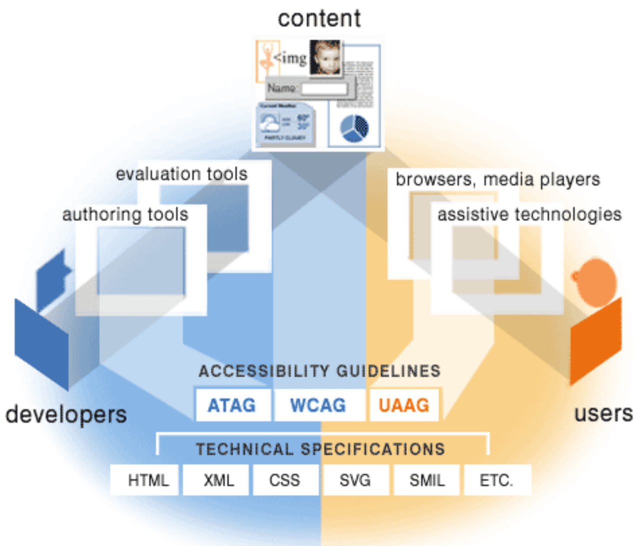


Figure 2. Composants essentiels de l'accessibilité web.

La première version des WCAG (1.0) a été publiée en tant que recommandations du W3C en mai 1999 (Chisholm et al. 1999). Après un processus de révision de cinq ans, la version 2.0 a été présentée en mai 2008 (Caldwell et al. 2008) avec des changements importants. Transformée en une norme ISO/IEC (40500 : 2012), elle est considérée actuellement comme un ouvrage de référence fondamental. Cette nouvelle version (la plus récente) est conçue autour de quatre principes : 1. Le contenu doit être *perceptible*; 2. Les composants de l'interface utilisateur et de navigation doivent être *utilisables*; 3. Les informations et l'utilisation de l'interface utilisateur doivent être *compréhensibles*; et 4. Le contenu doit être suffisamment *robuste* pour être interprété de manière fiable par une large variété d'agents utilisateurs. Au total, 12 règles sont désormais associées à ces principes[163], qui, à leur tour, fournissent des critères de succès (61) servant à déterminer la mesure dans laquelle chaque règle est satisfaite, selon le niveau de conformité défini : A (le plus bas), AA et AAA (le plus élevé).

De manière générale, cette thèse vise à expliquer pourquoi les professionnels de la localisation, en tant que producteurs des contenus web impliqués dans le cycle de développement des sites web multilingues (voir section 1.3), devraient également avoir pour tâche d'assurer que le produit cible a un niveau d'accessibilité acceptable, non seulement en suivant les WCAG 2.0, mais aussi en utilisant les outils d'évaluation destinés à contrôler l'accessibilité des documents web. Sur un plan

[163] La liste complète des règles est présentée dans le chapitre 2 de la thèse, tableau 2.1.

plus spécifique, nous cherchons à étudier l'implémentation et l'évaluation, dans des sites localisés, de la règle 1.1 des WCAG 2.0, qui recommande la chose suivante : *Proposer des équivalents textuels à tout contenu non textuel qui pourra alors être présenté sous d'autres formes selon les besoins de l'utilisateur : grands caractères, braille, synthèse vocale, symboles ou langage simplifié* (Caldwell et al. 2008).

1.3 La production de sites web multilingues

D'une manière générale, la localisation web peut être décrite comme la traduction, l'ingénierie et la vérification (« testing ») de contenus Web (Esselink 2000, 3). En offrant une expérience de navigation dans la langue maternelle des utilisateurs, la localisation contribue aussi au principe du W3C d'un « Web pour tous » et représente l'activité centrale de tout processus de développement de sites web multilingues. Dans le chapitre 2 de la thèse, nous abordons le sujet de la localisation de manière plus détaillée. Dans ce qui suit, nous ne faisons qu'une brève introduction avant d'exposer la motivation de notre travail de recherche (section 2).

1.3.1 Le modèle « GILT »

Dans une société numérique de plus en plus mondialisée, le multilinguisme du Web n'a pas passé inaperçu. Nous avons laissé derrière nous les premiers jours de l'ère de l'Internet, pendant lesquels les utilisateurs devaient lire les contenus en anglais car il y avait peu d'alternatives. Aujourd'hui, les sites web s'adaptent aux langues parlées par les consommateurs de l'information et de services numériques dans le monde entier. Ce phénomène tire sa source du paradigme de la mondialisation, vu comme « la transformation des activités d'affaires et des processus pour satisfaire les besoins des clients du monde entier, quels que soient leur langue, leur pays ou leur culture » (LISA dans Jiménez Crespo 2013, 25). Les efforts de mondialisation déclenchent inévitablement une succession de processus interdépendants communément appelés GILT (Cadieux et Esselink 2002) : l'internationalisation, la localisation et la traduction des produits numériques, tels que les sites web.

Selon Pym (2014, 119), l'internationalisation est « le processus de généralisation d'un produit afin que celui-ci puisse gérer plusieurs langues et conventions culturelles, sans que sa conception initiale doive être remaniée ». Lorsque l'internationalisation est considérée comme faisant partie intégrante du développement d'un site, et non comme une réflexion après coup, elle contribue à faciliter la localisation. L'idée est de générer un site suffisamment « neutre » de sorte qu'il puisse être facilement adapté plus tard, pendant le processus de localisation proprement dit, afin de le rendre attrayant pour les utilisateurs dans des *locales*[164] particuliers (Pym 2011, 413). Cela peut se faire, par exemple, en créant un site web compatible avec différents jeux de caractères ou en facilitant l'identification et la modification des chaînes de texte traduisibles (par exemple en évitant le texte incrusté, non-éditable, dans les images), pour ne citer que quelques techniques (Esselink 2006, 23).

Après l'internationalisation du produit web source, la phase de localisation couvre non seulement la traduction du texte, mais aussi l'adaptation d'autres éléments non textuels (par exemple, les images, les couleurs) afin de répondre aux exigences du public cible (Fernández Costales 2009). Bien qu'il ne soit pas rare pour une personne d'être entièrement responsable de la tâche de localisation (Gouadec 2007, 43), la localisation est considérée, dans le modèle GILT, comme une étape indépendante de la traduction elle-même, englobant la préparation et la gestion des contenus web à traduire, les tâches d'ingénierie postérieures à la phase de traduction (comme l'édition des images ou l'adaptation du code HTML) et les tests de contrôle qualité finaux (Jiménez Crespo 2013, 26). Néanmoins, le degré de séparation entre les activités de localisation et de traduction est déterminé finalement par un ensemble de facteurs assez complexe, y compris la taille et la portée du projet, sa complexité technique et les ressources humaines et financières disponibles.

Cette thèse s'éloigne de la structure hiérarchique et linéaire initialement imposée par le modèle GILT (Montalt i Resurrecció 2003) de deux manières différentes. Premièrement, nous traitons les tâches de traduction et d'ingénierie de la localisation comme un effort continu qui peut être effectué par une seule personne, le localisateur, qui est censé avoir les compétences linguistiques, techniques et instrumentales nécessaires pour délivrer un produit web cible fonctionnel (d'un point de vue linguistique, culturel et technique). L'inévitable chevauchement entre la localisation et la traduction, qui a déjà été traité par plusieurs chercheurs du domaine (Sandrini 2005 ; Nauert 2007 ; Fernández Costales 2009 ; Schäler 2010), ceux-ci ayant en particulier analysé le caractère multimodal du Web, a été également un aspect central de notre recherche. Deuxièmement, en suivant la proposition de Jiménez Crespo (2013, 27), nous estimons que le cycle GILT doit être plus interactif et favoriser la communication et l'échange de connaissances entre le localisateur et les différents acteurs impliqués dans le processus de développement de sites web.

---

[164] Dans le domaine de la localisation, le terme *locale* fait référence à un groupe de personnes qui partagent la même langue et les mêmes conventions culturelles. Le groupe ne se trouve pas nécessairement au même emplacement physique, d'où la nécessité de définir des paramètres régionaux par des paires langue-pays lors de la configuration de produits numériques (Cadieux et Esselink 2002). Par exemple, les paramètres régionaux français-France (FR-FR), français-Suisse (FR-CH), français-Canada (FR-CA) ou français-Belgique (FR-BE) sont des *locales* différents.

1.3.2 La localisation web comme un processus axé sur l'utilisateur

Loin de la considérer comme une activité menée par des intérêts purement économiques visant à pénétrer les marchés internationaux, comme ce fut le cas quand elle est apparue dans les années 1980 (Dunne 2015), nous voyons la localisation comme un processus où le contexte de réception et l'utilisateur final jouent un rôle crucial. En traductologie, les chercheurs ont souvent expliqué cette vision de la localisation plutôt orientée vers les destinataires du produit cible en adoptant une approche fonctionnaliste (Jiménez Crespo 2009b ; Fernández Costales 2009). Le fonctionnalisme rejet la notion stricte d'équivalence et l'importance extrême traditionnellement donnée au texte source ; par ailleurs, il établit que le but ou *Skopos* de l'acte de traduction est d'une importance capitale, et il considère que la fonctionnalité est obtenue lorsque la traduction répond aux attentes du récepteur, à ses besoins, à ses connaissances préalables et aux conditions situationnelles (Nord 1997). Dans le cadre de notre travail de recherche, la notion de fonctionnalité est toujours liée à l'environnement de l'utilisateur final, mais elle est élargie de façon à couvrir également l'efficacité de l'interaction entre l'utilisateur et les contenus web d'un point de vue technique.

Dans ce sens, nous adoptons la définition de localisation proposée par Sandrini (2008, 9), qui la décrit comme « le processus de modification d'un site web existant pour le rendre accessible, utilisable et culturellement approprié pour un certain public cible » ; en d'autres termes, nous sommes d'accord avec l'idée que le but ultime de la localisation est d'assurer le bon fonctionnement du site cible qui en résulte. En mettant l'accent sur l'utilisateur final, Sandrini affirme que l'objectif de toute tâche de localisation doit être que les gens d'un *locale* spécifique puissent utiliser le produit cible dans leur propre langue sans aucune difficulté (ibid). Dans notre travail, nous partons du principe que les utilisateurs qui appartiennent au même *locale* partagent la même langue et ont des conventions culturelles similaires, mais qu'ils n'ont pas nécessairement les mêmes capacités physiques et sensorielles, ce qui influence leur manière d'interagir avec le Web. En outre, nous considérons que les difficultés rencontrées par les utilisateurs lors de la navigation d'un site localisé peuvent être associées non seulement à des problèmes en termes d'adéquation et de correction linguistique et culturelle, mais aussi à des obstacles liés à la fonctionnalité du site, que le commissaire de la tâche n'a pas réussi à identifier dans la source et/ou que le localisateur n'a pas supprimés dans le produit cible. Si nous prenons notre objet d'étude principal — les images et leurs équivalents textuels —, de telles difficultés peuvent se produire si un site web localisé contient des images (i) qui n'ont pas d'attribut `alt`, (ii) qui ont des équivalents textuels non traduits, ou (iii) qui ont des équivalents textuels non pertinents par rapport au but et au contexte des images auxquelles ils sont associés.

2. Motivation de la recherche

Après plus de quinze ans depuis la publication de la première version des WCAG, les études montrent que très peu de sites sont entièrement conformes aux normes internationales d'accessibilité (Lopes et al. 2010 ; Harper et Chen 2012 ; Power et al. 2012). Même si, au cours des deux dernières décennies, le Web a connu des améliorations en termes d'accessibilité, les chercheurs croient qu'une partie seulement des progrès sont dus à un intérêt accru pour les bonnes pratiques d'accessibilité de la part des professionnels du Web et qu'ils s'expliquent plutôt par (i) des changements dans les styles de codage visant à atteindre un degré plus élevé de cohérence et de compatibilité entre différents appareils informatiques, ainsi que par (ii) l'exploitation des nouvelles fonctionnalités des navigateurs pour améliorer la conception initiale et la mise en page (Richards et al.2012 ; Hanson et Richards 2013). Selon la littérature, les webmasters, les développeurs web et les graphistes sont bien conscients des avantages d'adopter de bonnes pratiques d'accessibilité pour la société en général (Yesilada et al.2012 ; Putnam et al. 2012). Cependant, ils manquent souvent du temps et même de la formation nécessaires pour les appliquer (Lazar et al. 2004 ; Trewin et al. 2010). Certains allèguent également que les règles pour l'accessibilité des contenus web sont difficiles à comprendre et que les outils destinés à l'évaluation de l'accessibilité web (ci-après, outils WAE) n'offrent pas de conseils pertinents (ibid).

Cette thèse est motivée par le désir de comprendre si ces difficultés se présentent également dans le cas des sites web multilingues et si les localisateurs ont un rôle à jouer dans la création de contenus web plus accessibles. Nous sommes d'avis que, idéalement, les mécanismes d'accessibilité doivent être intégrés dans les pratiques quotidiennes, tout au long du cycle de vie du produit web (Cooper et al. 2012), de la conception et du développement du site à sa mise en ligne, son entretien et sa localisation. Dans notre travail, nous abordons ce sujet d'abord d'un point de vue général, en examinant les mesures adoptées actuellement pour rendre les sites multilingues accessibles (voir section 2.1), pour passer ensuite à l'étude d'un aspect plus concret : l'accessibilité des images sur le Web et sa localisation (voir section 2.2).

2.1 Accessibilité des sites web multilingues

Jusqu'à présent, l'analyse de l'implémentation et de l'évaluation de l'accessibilité sur des sites web multilingues n'a pas été beaucoup abordée dans la littérature, en dépit du fait que, comme le suggère Folaron (2012, 25), le Web est maintenant un « espace de traduction » par excellence, où les pratiques de traduction assurent une communication fluide entre les utilisateurs et jouent ainsi un rôle clé dans l'entretien des réseaux locaux et mondiaux.

Le W3C indique que les sites web avec des versions multiples indépendantes les unes des autres (par exemple, un site web en langues différentes, avec des URL différentes) devraient être évalués de façon indépendante en ce qui concerne leur niveau d'accessibilité (Velleman et Abou-Zahra 2014). Nous pourrions donc déduire que, dans le cas de sites localisés ou de sites culturellement personnalisés (Singh et Pereira 2005), les développeurs du produit source ne sont pas nécessairement responsables de l'accessibilité du produit cible. De même, dans le document des WCAG 2.0, l'une des rares références faites à l'accessibilité dans un contexte multilingue concerne la possibilité de reconnaître qu'un site web (ou une page) est conforme aux normes uniquement de manière partielle dans les cas où seuls les contenus dans l'une des langues disponibles ont été vérifiés (Caldwell et al. 2008). À part cela, et abstraction faite du travail effectué par l'équipe chargée des activités menées par la « W3C Internationalisation (I18n) Activity », destinées à promouvoir le principe d'un web pour tous, aucun document ou groupe de travail officiel du W3C n'aborde explicitement la façon dont l'accessibilité peut être assurée dans les sites web multilingues. Par ailleurs, aucune étude portant sur l'accessibilité web n'a fourni, à ce jour, des données empiriques ou théoriques sur cette question.

Dans le domaine de la localisation, l'accessibilité a été considérée comme l'un des paradigmes qui nourrissent l'interdisciplinarité des *Localisation Studies* (voir Figure 1.5 dans le chapitre 1 de la thèse). Certains auteurs ont déjà parlé des intérêts communs entre la localisation web et l'accessibilité d'un point de vue théorique, en affirmant que, quand un site web accessible est rendu multilingue, les localisateurs devraient s'assurer que les mêmes mécanismes d'accessibilité sont maintenus dans toutes les versions linguistiques du site sur lequel ils travaillent (Gutiérrez y Restrepo et Martínez Normand 2010 ; Tercedor Sánchez 2010). En outre, des chercheurs dans le domaine ont suggéré que la version localisée ne devrait jamais être moins accessible que l'original et que, selon le degré de liberté donné à l'équipe de localisation, la version localisée d'un site pourrait même se caractériser par davantage de fonctionnalité que la source et satisfaire les besoins d'un plus grand nombre d'utilisateurs (ibid).

Pym (2011, 424) considère que l'accessibilité est l'un des problèmes éthiques qu'affronte aujourd'hui la localisation web, comme discipline académique et comme profession. La nécessité (et l'obligation) du localisateur d'aborder les obstacles qui rendent l'interaction avec le Web difficile pour une personne a été reconnue par certains chercheurs intéressés par le sujet (Tercedor Sánchez 2010 ; Jiménez Crespo 2009a), mais seulement quelques-uns ont considéré qu'il s'agit d'une étape nécessaire du processus d'assurance qualité (Gibb et Matthaiakis 2007 ; Jiménez Crespo 2013). Après avoir constaté le succès d'une série de séminaires portant sur l'accessibilité web donnés dans le cadre de différents cours de localisation (Rodríguez Vázquez 2014), nous sommes déterminée à améliorer la sensibilisation aux questions liées à l'accessibilité au sein de la communauté de localisation web, où la prise de conscience à cet égard est encore considérée comme très faible (Ó Broin 2004).

2.2 L'accessibilité des images et sa localisation

L'importance du Web pour la diffusion de l'information est incontestable, mais la prédominance du contenu non textuel place les personnes ayant une déficience visuelle dans une situation de désavantage (Harper et Chen 2012). Les utilisateurs aveugles éprouvent de la détresse et de la frustration en naviguant sur le web en raison des obstacles liés à l'accessibilité, qui leur font perdre, en moyenne, 30,4% du temps qu'ils passent devant l'ordinateur (Lazar et al. 2007). L'une des principales sources de frustration pour ce groupe de la population est l'inaccessibilité des images (Paciello 2000 ; Asakawa 2005 ; Petrie et al. 2005 ; Lazar et al. 2007). Le travail de recherche présenté dans cette thèse reflète notre volonté de contribuer à l'élimination de cette barrière et est motivé par (i) les problèmes qui existent encore pour formuler des équivalents textuels pertinents pour les images et par (ii) le manque d'études empiriques se concentrant spécifiquement sur la manière dont ces problèmes sont abordés lors du processus de localisation web.

2.2.1 Production des équivalents textuels pour les images

Malgré le fait qu'assurer l'existence des équivalents textuels pour tout contenu non textuel est essentiel pour que les personnes aveugles puissent participer pleinement à la société de l'information, des travaux récents rapportent encore des taux de conformité assez faibles concernant l'accessibilité des images (Accès pour tous 2011 ; Hanson et Richards 2013). Même si le ratio d'attributs `alt` manquants a diminué graduellement sur le Web au cours des dernières années (Asakawa 2005 ; Richards et al. 2012), le principal problème est maintenant la présence d'équivalents textuels non informatifs (ibid). Nous pensons qu'il pourrait y avoir trois raisons principales à cette situation :

Manque de conseils détaillés sur la manière de formuler des équivalents textuels

Insérer un attribut `alt` dans un élément `<img>` lorsque des images doivent être introduites dans une page web semble être une technique très simple à appliquer. Cependant, élaborer un équivalent textuel pertinent peut devenir une tâche compliquée. Selon le W3C, la valeur de l'attribut `alt` doit communiquer le but de l'image ou expliquer sa signification ; en d'autres termes, un équivalent textuel devrait servir à remplacer l'image, et non inclure une simple description de son contenu. La formulation d'un équivalent textuel pertinent exige donc non seulement des compétences linguistiques et de synthèse, mais aussi un investissement de temps important. Les recommandations qui existent sur le type d'informations à fournir dans l'attribut `alt` (c'est-à-dire les conseils sur la composition des équivalents textuels) sont généralement incluses dans des documents officiels, longs et difficiles à comprendre pour des professionnels du Web ayant un profil plutôt technique (par exemple, la norme ISO/TS 20071-11 (Organisation internationale de normalisation 2012b)). En outre, des conseils de type linguistique axés sur la manière d'écrire des équivalents textuels (c'est-à-dire des indications sur leur formulation) sont rares et presque aucune référence ne peut être trouvée dans la littérature concernant le lexique préféré par les utilisateurs ou les structures syntaxiques à privilégier.

Limites des outils pour vérifier automatiquement l'accessibilité des images

Quand le temps ou la formation nécessaires pour appliquer les règles existantes sur la manière de rendre les images accessibles font défaut aux créateurs de contenus web, ceux-ci peuvent compter sur des outils d'évaluation pour trouver une solution rapide ou simplement combler le manque de connaissances. Quand il s'agit de vérifier qu'un site web est conforme à la règle 1.1 des WCAG 1.0, les outils WAE génériques sont capables d'identifier les éléments `<img>` qui n'ont pas d'attribut `alt` comme des cas problématiques, mais presque aucune information n'est fournie par rapport à la pertinence des équivalents textuels contenus dans la page. Les avertissements liés à l'accessibilité des images restent souvent trop vagues (par exemple « Veiller à ce que l'équivalent textuel transmette le contenu de l'image ») et sont souvent considérés comme peu utiles (Petrie et al. 2005). Des études ont tenté de proposer des alternatives à ces outils (Bigham et al. 2006 ; Bigham 2007 ; Olsen et al. 2010), mais la plupart d'entre eux ont mis l'accent sur la détection des équivalents textuels non informatifs (par exemple, des chaînes de caractères alphanumériques, des extensions de fichier, des informations sur la taille des fichiers) au lieu de proposer une solution visant à améliorer la qualité des équivalents textuels formulés avec des constructions typiques du langage naturel.

Distribution peu claire de la responsabilité concernant l'accessibilité des images

Indépendamment de l'absence d'une collection complète de recommandations sur la formulation d'équivalents textuels et de la non-spécificité des outils d'évaluation actuels en ce qui concerne la vérification de leur pertinence, il y a un troisième obstacle à l'accessibilité des images : il existe toujours une incertitude quant à la personne qui a la responsabilité de fournir des équivalents textuels de qualité pour les images. Plusieurs professionnels participent aux activités effectuées tout au long du cycle de vie d'un site web : les commissaires, les ingénieurs, les graphistes et les éditeurs de contenu, pour n'en citer que quelques-uns. En théorie, si l'on prend en compte les connaissances de ces différents acteurs, ce sont les développeurs qui devraient garantir que le mécanisme permettant d'ajouter un équivalent textuel existe (l'attribut `alt`), et ce sont les éditeurs qui devraient créer le texte à insérer dans cet attribut. En pratique, cependant, suivre un tel flux de travail strict est difficile, car les images peuvent être ajoutées, modifiées ou supprimées pendant des étapes différentes du cycle de développement web. Nous considérons que cette situation devient encore plus complexe quand la localisation d'un site est demandée, étant donné que le nombre d'acteurs participant au cycle est normalement doublé.

2.2.2 Traduction et adaptation des équivalents textuels pour les images

La valeur de l'attribut `alt` est généralement considérée comme une chaîne de texte qui doit être traduite au cours du processus de localisation web (Gibb et Matthaiakis 2007, 668 ; Mata Pastor 2009b, 552 ; Roturier 2015, 88). L'importance d'une telle action pour faciliter une meilleure expérience de navigation web aux utilisateurs aveugles a même été reconnue par des experts dans le domaine de l'accessibilité (Clark 2002, 99). Néanmoins, à notre connaissance, seules deux études de localisation ont porté sur la présence des équivalents textuels pour les images sur des sites web localisés.

En suivant une approche descriptive, Jiménez Crespo (2008) a analysé un corpus parallèle espagnol, comprenant des sites web originaux (N=172) et localisés (N=95) du secteur privé. L'auteur a constaté que le sous-corpus de textes traduits contenait un nombre plus élevé d'attributs `alt`, indiquant que cela était probablement dû à une utilisation croissante de systèmes de gestion de contenu (CMS), qui les insèrent automatiquement dans la page. Outre les rapports de la moyenne de mots par équivalent textuel, la présence de caractères non alphabétiques (espaces, *, /) et la prédominance des formes nominales, aucune donnée supplémentaire n'a été livrée concernant la composition et la formulation des équivalents textuels. De même, Fernández Costales (2010) a mené une étude à grande échelle pour étudier la localisation de sites universitaires de 27 pays de l'Union européenne. Dans le cadre d'une analyse qualitative complète du contenu textuel et non-textuel (y compris des éléments culturels) dans la collection de documents web, il a conclu que la traduction des équivalents textuels reçoit peu d'attention pendant le processus de localisation de ce type de sites institutionnels.

Malgré les précieuses contributions apportées par ces deux études, la pertinence des équivalents textuels n'a pas été examinée en détail, très probablement parce que l'observation de ces éléments ne représentait que l'un des multiples aspects analysés dans le travail de ces auteurs. En traductologie, des considérations relatives à la composition et à la formulation des équivalents textuels ont été explorées seulement dans le contexte des textes techniques (Prieto Velasco 2009). Alors que certains critères pratiques ont été extrapolés au cas particulier des images sur le Web (Tercedor Sánchez et Prieto Velasco 2009, 84), aucune autre étude n'a été menée pour déterminer, par exemple, si ces recommandations pourraient être appliquées par des localisateurs ou si leur mise en œuvre pourrait être automatisée.

3. Objectifs de la thèse

Compte tenu de ce qui vient d'être présenté, l'objectif général de cette thèse a été d'étudier le rôle des localisateurs en tant que fournisseurs d'accessibilité dans les sites web multilingues. La portée de notre recherche a été, par contre, délimitée par trois objectifs plus spécifiques liés aux méthodes d'assurance et d'évaluation de l'accessibilité. Ces objectifs sont présentés ci-dessous par ordre d'importance croissante :

Objectif 1 : Déterminer s'il y a des procédures déjà établies pour assurer l'accessibilité des sites web multilingues en général.

Objectif 2 : Proposer une solution pour contrôler la qualité des équivalents textuels pour les images lors du processus de localisation, fondée sur le langage contrôlé (LC), en vue de remédier à certaines lacunes observées dans les recommandations déjà existantes et dans les outils conçus pour l'évaluation de l'accessibilité des images.

Objectif 3 : Déterminer dans quelle mesure les localisateurs sont capables d'assurer un niveau d'accessibilité acceptable dans le produit web cible en ce qui concerne les images.

Dans notre travail de recherche, notre hypothèse de départ a été que l'évaluation de l'accessibilité devrait faire partie des mesures d'assurance qualité prises au cours du processus de localisation web, afin de garantir un produit cible fonctionnel. En suivant la définition de Saldanha et O'Brien (2014, 95), nous comprenons l'assurance qualité (AQ) comme la somme de tous les systèmes et processus utilisés pour aider à créer ou à maintenir la qualité. En ce sens, l'AQ englobe toutes les actions effectuées ou mises en place pour assurer (de manière prospective) et évaluer (a posteriori) l'accessibilité. Nous aimerions signaler que cette thèse n'a pas pour objectif d'étudier en profondeur des modèles d'assurance qualité utilisés dans l'industrie de la localisation ou de théoriser sur la notion complexe de qualité, si controversée en traductologie. Avec notre travail, nous avons cherché plutôt à montrer, avec des données empiriques (voir section 4), que les aspects liés à l'accessibilité, tels que la pertinence des équivalents textuels localisés, peuvent être considérés par les utilisateurs finaux comme un facteur déterminant de la qualité du produit web cible.

4. Méthodologie et questions de recherche

Pour atteindre les trois objectifs énoncés dans la section précédente, nous avons adopté une approche empirique, en combinant des stratégies de recherche descriptives et expérimentales. Dans le domaine de l'IHM, les études descriptives représentent souvent la première étape d'un projet de recherche ; elles permettent aux chercheurs d'examiner concrètement ce qui se passe dans un domaine d'étude et elles servent de base aux étapes de recherche suivantes. Les études expérimentales, d'autre part, permettent d'explorer des relations de causalité (Lazar et al. 2010, 20-22). Cette thèse s'appuie sur une approche descriptive pour répondre aux deux premiers objectifs et sur une approche expérimentale pour atteindre le troisième objectif, qui constitue la partie la plus importante de notre travail de recherche. La figure 3 donne un aperçu de la triangulation méthodologique suivie dans cette thèse.

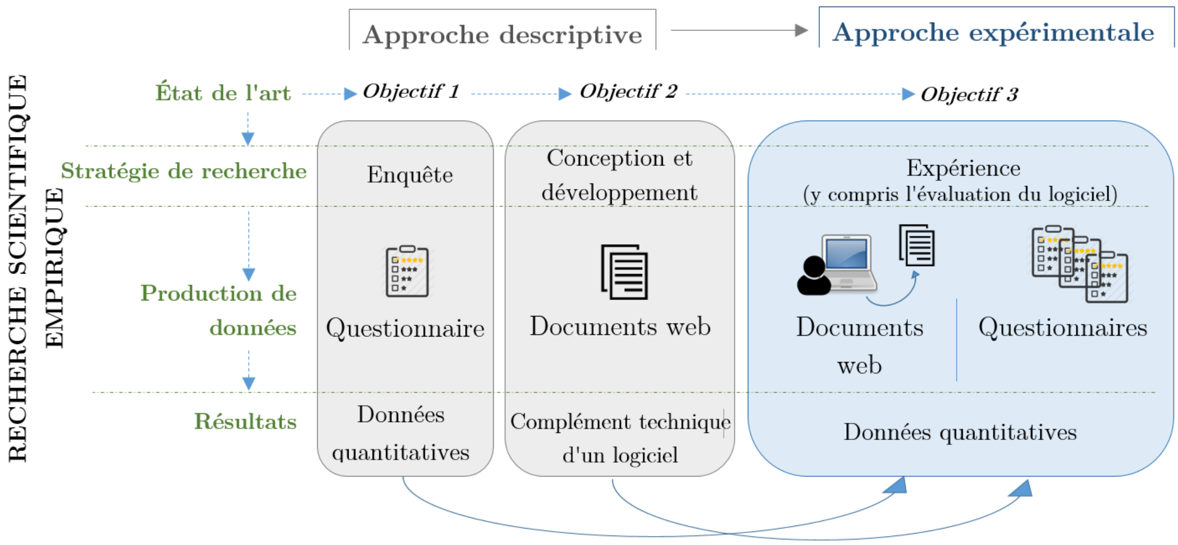


Figure 3. Aperçu du cadre méthodologique de notre travail de recherche

Pour atteindre l'**objectif 1,** nous avons passé en revue la littérature existante sur l'implémentation et l'évaluation de l'accessibilité web, ainsi que les étapes les plus importantes du processus de localisation web. Cette étude de la littérature a été complétée par une enquête avec 67 experts en AW, visant à obtenir des informations sur (i) les procédures qu'ils suivent lors de l'évaluation de l'accessibilité des sites web multilingues et (ii) leurs perceptions de la contribution potentielle des localisateurs quant à la réalisation de sites web multilingues plus accessibles. Les données quantitatives relatives à ces deux aspects ont été collectées à l'aide d'un questionnaire accessible en ligne.

Guidée par notre **objectif 2**, après un examen complet des études menées jusqu'à présent en ce qui concerne l'accessibilité des images, nous avons développé 40 règles de langage contrôlé (LC) pour faciliter l'identification des équivalents textuels non pertinents en français pendant le processus de localisation, en utilisant un formalisme déclaratif propriétaire (Bredenkamp et al. 2000). Les règles en question, dont la définition a été basée sur l'analyse de deux corpus de documents — 1. les recommandations trouvées dans la littérature sur la formulation d'équivalents textuels et 2. un corpus web déjà existant (voir section 4.3 du chapitre 4 de la thèse) − peuvent être appliquées en utilisant Acrolinx, l'un des logiciels de pointe pour le contrôle-qualité. En créant un complément technique pour un logiciel, nous avons suivi une stratégie de recherche qui, dans le domaine de l'informatique et des systèmes d'information, est connue comme « conception et développement » (Oates 2005, 107). L'évaluation d'Acrolinx et des règles a été intégrée dans la stratégie de recherche suivie pour atteindre l'**objectif 3.**

Afin de déterminer la capacité des localisateurs à assurer l'accessibilité des images dans le produit web cible (objectif 3), nous avons mené une étude expérimentale qui a été divisée en deux étapes : (i) une expérience de localisation web (étape 1) pour laquelle nous avons recruté 28 localisateurs (dont 14 avaient des connaissances de base sur l'accessibilité web) ; et (ii) une évaluation par des utilisateurs finaux (étape 2), qui a impliqué la participation de sept utilisateurs aveugles (voir figure 4). L'étude visait à répondre aux questions de recherche suivantes :

R1. Les équivalents textuels pour les images sont-ils considérés par les localisateurs comme éléments traduisibles pendant le processus de localisation web ?

R2. L'utilisation d'outils d'assurance qualité destinés à contrôler l'accessibilité pendant le processus de localisation web permet-elle de créer des équivalents textuels pour les images plus pertinents ?

R3. Le fait d'avoir des connaissances sur l'accessibilité web aide-t-il les localisateurs à créer des équivalents textuels pour les images plus pertinents ?

Au cours de l'expérience, les localisateurs ont tout d'abord été invités à localiser un site web de trois pages contenant 130 images, puis à évaluer l'accessibilité des images du site localisé à l'aide de deux outils d'AQ : (i) Acrolinx, l'outil de validation semi-automatique pour lequel nous avons développé les règles de LC, et (ii) aDesigner, un outil WAE (Asakawa 2005). Afin de ne pas biaiser les résultats, les outils n'ont pas toujours été utilisés dans le même ordre. Deux scénarios d'AQ ont donc été testés : aDesigner-Acrolinx (scénario A) et Acrolinx-aDesigner (scénario B). La pertinence des équivalents textuels a ensuite été évaluée par des utilisateurs de lecteurs d'écran via un questionnaire.

En vue de répondre aux questions de recherche mentionnées ci-dessus, nous avons analysé les données quantitatives collectées pendant les deux étapes de l'étude dans le but de mesurer l'impact du fait d'avoir des connaissances sur l'AW et de l'utilisation des outils d'assurance qualité pour contrôler l'accessibilité – variables indépendantes (VI) primaires – sur la traduction et la pertinence des équivalents textuels produits par les localisateurs – variables dépendantes (VD) – (voir figure 4). L'impact des autres variables indépendantes secondaires sur le résultat de la localisation, telles que le scénario d'AQ suivi par les participants, a également été mesuré.[165] Des informations plus détaillées sur la méthodologie adoptée pour atteindre l'**objectif 3** se trouvent dans le chapitre 5 de la thèse. Les hypothèses et les sous-hypothèses associées à nos questions de recherche sont incluses dans le même chapitre, dans la section 5.4.

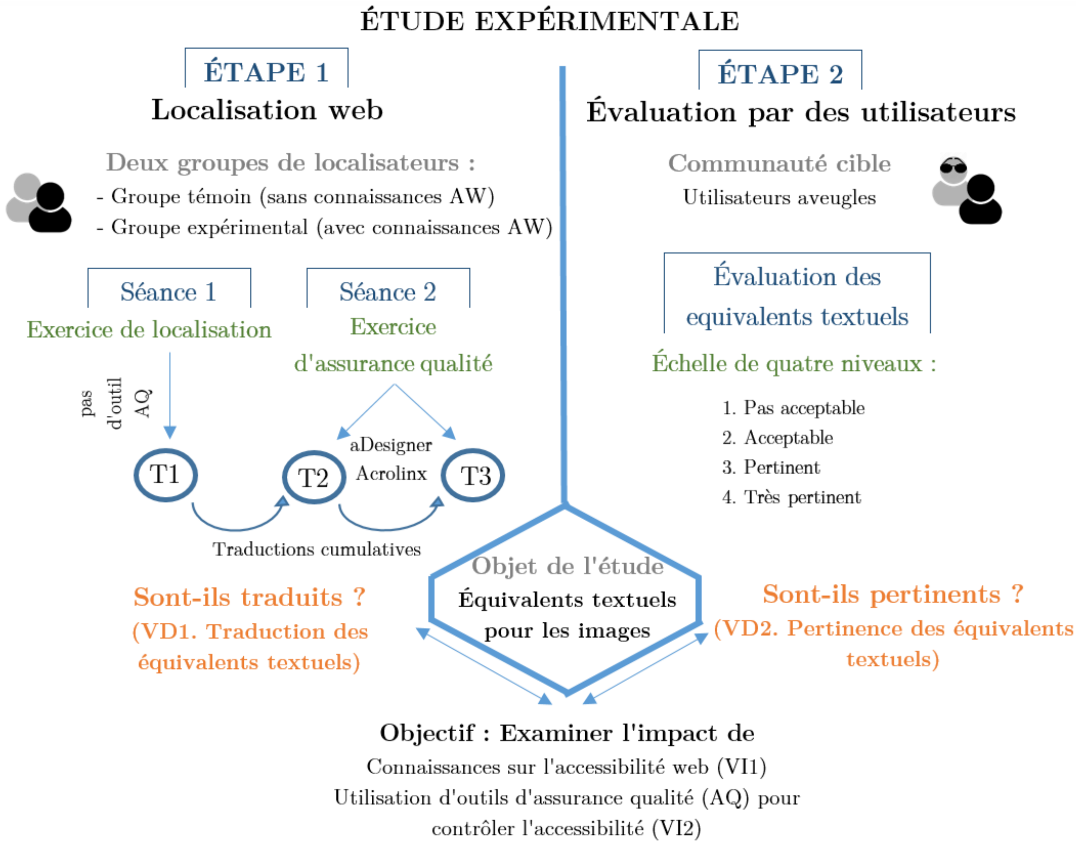


Figure 4. Aperçu de l'étude expérimentale (étapes, variables, participants, objectifs)

5. Résultats principaux

De manière générale, l'analyse de toutes les données collectées a contribué à mieux comprendre comment et pourquoi l'accessibilité doit être prise en compte lors du processus de localisation web. En fait, il s'agit de l'une des contributions principales de notre recherche, qui est d'ailleurs le premier travail empirique à avoir étudié plus en profondeur les convergences entre la localisation et l'accessibilité. Nous considérons que, malgré les limites de notre travail de recherche (voir section 7), nous avons atteint les objectifs initialement fixés. Dans ce qui suit, nous allons présenter nos résultats principaux, ainsi que souligner leur impact et leur

[165] Outre le scénario d'AQ, six autres variables indépendantes secondaires ont également été explorées. Un résumé de toutes les variables considérées lors de l'analyse des données peut être consulté dans le tableau 6.1 du chapitre 6 de la thèse.

valeur ajoutée pour les domaines de spécialisation qui ont délimité le cadre théorique de notre recherche (voir section 6).

Objectif 1

Les résultats liés à l'objectif 1 sont tirés de la revue de la littérature et de l'enquête décrite dans le chapitre 2 de la thèse (section 2.5).

- *Le W3C ne propose pas des méthodes ou des recommandations homogénéisées pour assurer l'accessibilité des sites web multilingues.*

Les conseils qui existent dans la littérature sur la manière d'implémenter les meilleures pratiques d'accessibilité ou de vérifier les contenus web pour vérifier si ces pratiques ont été suivies n'offrent pas de pistes sur la façon de procéder dans le cas des sites web localisés. Le W3C ne reconnaît que les cas où les différentes versions d'un site web multilingue peuvent être utilisées indépendamment les unes des autres (par exemple, quand elles sont disponibles sous différentes URL), chaque version doit être contrôlée individuellement. Cependant, il n'y a pas d'information sur qui devrait le faire ou comment.

- *Les experts en AW ne suivent pas une procédure standardisée lors de l'évaluation des sites web multilingues pour contrôler l'accessibilité.*

Les professionnels du Web avec expertise dans l'accessibilité considèrent que la solution idéale pour garantir l'accessibilité des sites web multilingues est d'assurer que chaque version linguistique est conforme aux règles d'accessibilité du W3C. Pourtant, ils reconnaissent que, dans des situations réelles, cela ne se produit presque jamais en raison du manque de temps, de compétences linguistiques supplémentaires et des ressources humaines nécessaires. En conséquence, il semble que, même si elles ne sont pas considérées comme souhaitables, les pratiques actuelles consistent à se concentrer uniquement sur une version dans une langue déterminée et à supposer que les autres ont un niveau d'accessibilité similaire. En outre, il s'avère que les éléments culturels et le contenu textuel ne reçoivent habituellement que peu d'attention lors du contrôle de l'accessibilité, en dépit du fait que ces éléments sont particulièrement importants dans le cas des sites web localisés.

- *Les experts en AW conviennent que le contrôle de l'accessibilité devrait faire partie des tâches quotidiennes des localisateurs.*

Les professionnels impliqués dans le processus de développement web partagent notre conviction que les localisateurs doivent être responsables du niveau d'accessibilité du contenu qu'ils produisent, dans la même mesure que les concepteurs de sites web, les développeurs et plus encore que les webmestres. De même, la communauté d'accessibilité estime que le profil complet du localisateur peut être particulièrement avantageux lors des contrôles d'AW réalisés au cours de la dernière phase d'assurance qualité avant qu'un site web multilingue devienne opérationnel.

Objectif 2

Les résultats liés à l'objectif 2 comprennent l'issue de la stratégie de conception et de développement adoptée (voir chapitre 4 de la thèse), ainsi que les conclusions tirées de son évaluation.

- *Un ensemble de 40 règles de LC visant à identifier des équivalents textuels non pertinents en français a été développé en utilisant la technologie d'Acrolinx.*

En prenant compte d'exemples concrets d'équivalents textuels de haute et basse qualité, nous avons défini trois sous-ensembles de règles pour assurer la pertinence de ces éléments en français pour des images avec une valeur descriptive (sous-ensemble A, 14 règles, voir la section 4.4.1 du chapitre 4) et une valeur fonctionnelle (sous-ensemble B, 18 règles, voir la section 4.4.2 du même chapitre). De plus, nous avons développé 8 règles pour identifier des équivalents textuels contenant des chaînes de texte non informatif (sous-ensemble C, 8 règles, voir la section 4.4.3 du chapitre mentionné ci-dessus). Ces règles interdisent l'utilisation de certains termes et expressions et proposent des structures syntaxiques spécifiques selon la fonction communicative de l'image. Ces règles ont été formalisées en utilisant la technologie d'Acrolinx, qui offre également un client pour automatiser leur application dans le cas des fichiers HTML : Acrolinx IQ Batch Checker (2.7.0).

- *Pour chaque règle développée, nous avons créé aussi des fichiers de documentation pour aider les utilisateurs à comprendre pourquoi un certain équivalent textuel n'est pas pertinent et comment il pourrait être amélioré.*

Quand les utilisateurs utilisent Acrolinx et notre ensemble de règles pour valider des fichiers HTML, un rapport d'erreurs est généré. Grâce à ce rapport, les utilisateurs peuvent consulter la documentation de chaque règle pour mieux interpréter les problèmes identifiés par l'outil. Plus précisément, chaque fichier d'aide comprend une brève introduction au sujet de l'accessibilité des images, une explication des raisons pour lesquelles un certain équivalent textuel a été signalé, des exemples qui illustrent cette explication et des recommandations pour améliorer la qualité de l'équivalent textuel. La documentation des règles contient également des liens vers les informations officielles fournies par le W3C. Un exemple est montré à la figure 4.4 du chapitre 4 de la thèse.

- *Des dix règles de LC évaluées, nous avons constaté qu'au moins huit d'entre elles conduisent à formuler des équivalents textuels significativement plus pertinents pendant le processus de localisation.*

Lorsque le site web source contient des équivalents textuels de mauvaise qualité, la probabilité de formuler une valeur pertinente pour l'attribut `alt` dans la version localisée du site est entre deux et trois fois plus élevée quand les règles que nous avons développées sont appliquées. En particulier, les règles visant à améliorer la pertinence des équivalents textuels pour des images fonctionnelles, qui devraient décrire l'action qu'elles déclenchent lors de l'interaction avec l'utilisateur, conduisent à des résultats d'une pertinence significativement plus élevée. L'effet significatif de la plupart des règles testées est plus remarquable lorsqu'Acrolinx est le seul outil utilisé pour renforcer la pertinence des équivalents textuels localisés, mais il est généralement maintenu aussi quand l'application des règles se produit après avoir effectué des contrôles avec un autre outil. Les deux règles qui n'ont pas montré un haut niveau d'efficacité étaient liées à (i) l'insertion d'informations sur les droits d'auteur des images dans l'équivalent textuel et (ii) l'utilisation d'expressions redondantes telles que « Image d'un... » ou « Une image qui montre ... ».

- *De manière générale, un outil de validation basé sur des règles de LC comme Acrolinx peut offrir un meilleur soutien à la réalisation d'images plus accessibles dans des sites web localisés qu'un outil WAE général.*

La proposition d'adopter une approche fondée sur le langage contrôlé pour améliorer l'accessibilité des images a été motivée par les limitations observées dans les outils actuellement utilisés à cette fin, y compris leur incapacité de distinguer les équivalents textuels pertinents de ceux qui ne le sont pas, ainsi que de valider des chaînes de texte dans des langues autres que l'anglais. Notre travail de recherche a confirmé que l'adoption d'une telle approche pendant le processus de localisation web peut entraîner la production d'équivalents textuels plus pertinents dans le site cible que lorsqu'un outil général de validation de l'accessibilité est utilisé. En outre, notre proposition s'est révélée significativement avantageuse non seulement en termes de détection d'équivalents textuels non pertinents, mais aussi à l'égard des conseils fournis pour leur amélioration. Nous avons constaté que, quand les localisateurs ont consulté la documentation des règles dont nous venons de parler par rapport à l'objectif 2, la qualité de leurs équivalents textuels a augmenté de manière significative.

Objectif 3

Le dernier objectif de cette thèse a été atteint en donnant réponse à trois questions de recherche spécifiques. Ci-dessous, nous résumons (i) les conclusions principales auxquelles nous sommes parvenue après avoir testé nos hypothèses, décrites dans le tableau 6.22 (chapitre 6 de la thèse), ainsi que (ii) les résultats des analyses supplémentaires présentées au chapitre 7.

R1. *Les équivalents textuels pour les images sont-ils considérés par les localisateurs comme éléments traduisibles pendant le processus de localisation web ?*

Contrairement aux attentes, les équivalents textuels ne sont pas systématiquement traduits par les localisateurs. En analysant les données quantitatives collectées pendant l'expérience et les données subjectives recueillies à travers les questionnaires, nous avons observé un degré de variabilité assez remarquable à l'égard de l'importance attribuée à l'existence de ces éléments et à la nécessité de les traduire. Il est intéressant de noter que les localisateurs ayant des connaissances de base sur l'AW ont été en mesure d'identifier et de traduire un nombre significativement plus élevé d'équivalents textuels que ceux qui n'avaient pas reçu une formation préalable sur l'accessibilité. En outre, l'étude expérimentale que nous avons menée a révélé que les localisateurs ne sont pas à l'aise quand il s'agit d'introduire des attributs `alt` sur des éléments `<img>` qui n'avaient pas d'équivalents textuels associés dans le document source. En ce qui concerne le traitement des attributs `alt` vides pendant le processus de localisation, les localisateurs semblent être plus enclins à proposer de nouveaux attributs vides pour marquer la valeur décorative des images (c'est-à-dire, transformer certains équivalents textuels trouvés dans le site web source en `alt` vides) qu'à reconsidérer la pertinence des attributs `alt` déjà vides dans la source.

Notre étude a également montré que l'utilisation d'outils d'assurance qualité pour contrôler l'accessibilité facilite l'identification des équivalents textuels comme éléments traduisibles. Nous avons constaté que les localisateurs qui n'avaient touché aucun équivalent textuel dans la première version de leur traduction ont ensuite repéré leur présence au cours du processus d'assurance qualité. Plus précisément, nous avons observé que, de manière générale, Acrolinx a déclenché un nombre plus élevé de modifications dans les équivalents textuels qu'aDesigner (l'outil WAE) pendant l'exercice d'AQ. Pour les localisateurs sans connaissance préalable sur l'AW, le premier s'est en fait révélé beaucoup plus efficace que le second à cet égard. De même, nous avons remarqué que, pour ce groupe de participants, l'utilisation combinée de ces deux outils a donné lieu au double de modifications dans les équivalents textuels par rapport à la version de la traduction qu'ils avaient produite au cours de l'exercice de localisation.

R2. *L'utilisation d'outils d'assurance qualité destinés à contrôler l'accessibilité pendant le processus de localisation web permet-elle de créer des équivalents textuels pour les images plus pertinents ?*

L'une des principales conclusions tirées de notre recherche est que l'utilisation d'outils de vérification peut aider les localisateurs à produire un niveau d'accessibilité des images dans le site cible significativement plus élevé que dans les cas où aucune technologie d'AQ n'est utilisée. Nous avons constaté que cela se passe non seulement lorsque deux outils avec des fonctionnalités complémentaires sont combinés, mais aussi quand les localisateurs n'ont qu'un seul outil à leur disposition (soit Acrolinx soit aDesigner). Une analyse plus détaillée des données a révélé qu'un logiciel comme Acrolinx, orienté pour le traitement automatique du langage naturel (TALN), est nettement plus performant qu'un outil WAE comme aDesigner lorsque les localisateurs utilisent seulement un outil de validation. C'est uniquement quand aDesigner est utilisé après Acrolinx que des résultats significatifs sont obtenus par rapport à la qualité des équivalents textuels. Sur la base de ces données, nous avons déduit que la meilleure solution pour assurer la pertinence de ces éléments dans la version localisée est d'utiliser tout d'abord Acrolinx, qui se concentre davantage sur la correction linguistique, et ensuite utiliser aDesigner, qui permet au localisateur d'identifier rapidement des attributs `alt` manquants et les équivalents textuels non pertinents les plus courants.

En outre, nous avons observé que l'utilisation d'outils d'assurance qualité pour contrôler l'accessibilité s'est révélée efficace pour améliorer la qualité des équivalents textuels dans les cas où le texte source n'était pas pertinent, mais elle n'a pas contribué à améliorer ou réduire le niveau de pertinence des équivalents textuels qui, à l'origine, étaient conformes aux normes d'accessibilité dans le site source. De même, il nous semble important de noter que, contrairement aux outils d'AQ, l'utilisation de logiciels de traduction (notamment des outils de TAO) n'a eu aucun impact significatif sur le niveau d'accessibilité des images.

R3. *Le fait d'avoir des connaissances sur l'accessibilité web aide-t-il les localisateurs à créer des équivalents textuels pour les images plus pertinents ?*

Comme dans le cas de la traduction des équivalents textuels, nous avons trouvé que le fait d'avoir des connaissances de base sur l'AW conduit à des résultats significativement meilleurs par rapport aux équivalents textuels des images, cette fois-ci en termes de qualité. Nos données indiquent que le fait de suivre une formation d'une heure sur les meilleures pratiques d'accessibilité a un impact tout à fait significatif sur la pertinence des équivalents textuels produits par les localisateurs. Cependant, l'un des résultats les plus frappants de notre étude est que cet avantage face aux localisateurs sans aucune connaissance préalable sur le sujet disparait lorsque les outils adéquats sont utilisés. En effet, notre étude a montré que lorsque les deux groupes de localisateurs ont suivi les recommandations fournies par Acrolinx, les différences entre eux en ce qui concerne la pertinence des équivalents textuels créés n'ont pas été significatives. En outre, lorsque le scénario Acrolinx-aDesigner a été appliqué, nous avons constaté que les localisateurs qui ne sont pas au courant des problèmes d'accessibilité ont obtenu des résultats légèrement meilleurs. Dans notre étude, lors de l'analyse du niveau d'accessibilité des images que les localisateurs avaient réussi à atteindre dans le site cible, nous avons également pris en compte leurs connaissances en HTML. En comparant les données des participants, nous avons observé qu'en fait, une bonne maîtrise de ce langage de balisage aide les localisateurs à produire des équivalents textuels significativement plus pertinents.

6. Impact de la recherche et contributions aux différents domaines de spécialité

Si nous prenons en considération les résultats présentés dans la section précédente, ainsi que les autres conclusions auxquelles nous sommes progressivement parvenue tout au long de la thèse, nous pourrions affirmer que notre hypothèse de départ a été confirmée : les opinions subjectives des professionnels du Web (concepteurs, développeurs, consultants en matière d'accessibilité, localisateurs) et les données objectives que nous avons recueillies à partir de notre étude expérimentale montrent que les localisateurs devraient effectivement contribuer à la réalisation de sites web multilingues plus accessibles. À un niveau plus spécifique, notre travail de recherche a apporté plusieurs contributions aux domaines de l'accessibilité, du langage contrôlé et de la localisation web.

Accessibilité

Cette thèse a suivi la tendance observée dans le domaine de l'accessibilité à étudier quels sont les moyens les plus efficaces et durables pour assurer un Web plus inclusif pour tous. La nouveauté de notre recherche réside dans le fait que (i) nous avons étudié le comportement d'un groupe de professionnels qui, à ce jour, n'avait jamais été considéré en tant que contributeur potentiel à l'objectif d'accès universel mentionné ci-dessus et que (ii) nous avons démontré que leurs actions peuvent en effet avoir un impact sur l'accessibilité globale du Web. En outre, nous avons ouvert une nouvelle ligne de discussion au sein de la communauté scientifique intéressée à l'accessibilité en mettant au premier plan la distinction entre sites monolingues et multilingues. Du point de vue de l'accessibilité, nous croyons que ces derniers présentent beaucoup plus de défis que les premiers, non seulement parce que le contenu est disponible dans plusieurs langues et que les différents éléments web peuvent varier selon les versions, mais aussi parce que leur production implique un plus grand nombre d'acteurs, avec des compétences et perceptions différentes sur les besoins des personnes handicapées. Nos découvertes en ce qui concerne l'absence de procédures normalisées pour assurer l'accessibilité des sites web multilingues ont mis en lumière le besoin urgent de développer de nouvelles méthodes d'évaluation pour ce type de sites en particulier.

Sur le plan de l'accessibilité des images, notre travail peut avoir des applications pratiques. Tout d'abord, nous avons démontré que l'inefficacité des outils de validation actuels peut être couverte en identifiant les avantages des logiciels utilisés dans d'autres domaines, tels que le TALN. Sur la base des avantages que nous avons mis en évidence par rapport à l'utilisation d'Acrolinx (et sa combinaison avec des outils comme aDesigner, axés sur la validation de la syntaxe des langages de balisage), nous sommes maintenant en mesure de suggérer l'intégration des outils de LC dans les flux d'évaluation de l'accessibilité. Deuxièmement, l'étude de corpus qui nous a permis de définir des règles de LC nous a aussi aidé à comprendre qu'en fait, les équivalents textuels pertinents suivent des patrons similaires et qu'il est possible d'en tirer des recommandations de type linguistique pour l'élaboration d'équivalents textuels de qualité. Même si nous avons identifié les patrons les plus fréquents en français, nous considérons qu'ils peuvent également servir de guide pour les créateurs de contenu travaillant avec d'autres langues, étant donné qu'ils donnent des pistes sur les formulations préférées par les utilisateurs aveugles en fonction de la valeur communicative de l'image.

Langage contrôlé

Notre travail pourrait également être considéré comme innovant du point de vue de la recherche en langage contrôlé. Il représente la première étude qui a porté sur l'applicabilité des logiciels de LC à des fins d'accessibilité et nos résultats à cet égard semblent prometteurs. En particulier dans le cas de la langue française, nous avons montré que certains aspects liés à l'accessibilité peuvent être formalisés en créant des règles de LC relativement simples. Cela pourrait contribuer à attirer l'attention des chercheurs en LC sur le domaine de l'accessibilité, lesquels pourraient examiner si d'autres questions d'accessibilité peuvent être abordées en suivant une approche similaire. En outre, dans les études de localisation, l'application du LC a généralement mis l'accent sur différentes dimensions de la qualité du produit cible au niveau textuel, comme sa lisibilité ou sa traduisibilité. Avec notre étude, nous avons montré que de contrôler la façon dont le texte est produit peut également faciliter l'accès à l'information visuelle dans un document Web.

Localisation web

Finalement, nous croyons que notre travail de recherche peut avoir des conséquences directes sur la définition de meilleures pratiques en matière de localisation dans le futur. En rapportant la première étude scientifique qui montre la manière dont l'accessibilité est abordée au cours du processus de localisation, nous avons pu observer que la prise de conscience concernant les besoins des personnes handicapées est encore relativement faible au sein de la communauté de localisation. L'implémentation de mécanismes favorisant l'accessibilité n'est pas encore considérée comme une obligation ou une nécessité – comme nous l'avons illustré avec le cas particulier de l'adaptation des équivalents textuels –, ce qui peut avoir des conséquences négatives au niveau de l'accessibilité globale du Web. En ce sens, nous avons également montré qu‘avec un peu d'intérêt de la part des localisateurs, le contraire peut aussi se produire. Concrètement, nos résultats mettent en évidence la nécessité de donner aux localisateurs les moyens nécessaires pour participer plus activement au cycle de développement web. Il semble qu'ils ne se sentent pas suffisamment à l'aise pour faire des modifications dans le code source, même si ces changements sont dans l'intérêt de faciliter l'accès aux contenus web localisés par une plus grande variété d'utilisateurs.

De même, notre recherche a contribué au domaine des *Localisation Studies* en présentant la première étude empirique où le produit source était composé de documents HTML et en faisant appel à des utilisateurs finaux souffrant d'un handicap pour jouer le rôle d'évaluateurs ou d'experts dans le but de vérifier la qualité du produit cible. De plus, la nécessité d'accorder plus d'attention aux problèmes d'accessibilité au cours du processus de localisation, qui a été reconnue par les experts de la communauté de l'accessibilité et reflétée dans les résultats de notre étude empirique, a fourni des indications intéressantes sur l'accessibilité comme un indicateur de qualité du produit localisé. Cela pourrait avoir un impact remarquable sur trois dimensions spécifiques. (i) *Le processus d'assurance qualité en localisation* : nos résultats indiquent qu'une étape d'évaluation de l'accessibilité devrait être introduite dans le cadre du processus d'assurance qualité en localisation, en suivant les techniques de validation automatiques et manuelles ; (ii) *les compétences générales de localisation :* outre les compétences instrumentales, techniques et linguistiques traditionnelles, notre travail a mis en évidence d'autres compétences qui pourraient se révéler avantageuses pour la réalisation de sites web localisés entièrement fonctionnels, telles que comprendre comment les normes d'accessibilité sont implémentées et comment les personnes handicapées utilisent leurs technologies d'assistance pour percevoir et interagir avec le Web ; et (iii) *la formation en localisation* : notre recherche a montré les effets positifs de l'augmentation de la sensibilisation en matière d'accessibilité et de l'utilisation des outils de validation pour le contrôle de l'accessibilité, suggérant ainsi que les aspects liés à l'accessibilité du cycle de développement web pourraient présenter une valeur ajoutée s'ils étaient introduits dans le programme des cours de localisation.

7. Limites et pistes de recherche pour l'avenir

Étant la première étude sur l'accessibilité des sites web multilingues qui a tenté d'examiner les liens entre les trois domaines de spécialisation mentionnés ci-dessus, cette thèse a vu sa portée limitée en ce qui concerne le type de contenus web analysés et le nombre de langues étudiées. En ce sens, nous avons examiné plus en profondeur l'accessibilité des images, ne faisant référence que de temps en temps à des exemples liés à d'autres éléments web, pertinents à la fois pour l'accessibilité et pour la localisation. De plus, le choix des images comme objet central d'étude nous a inévitablement mené à restreindre notre intérêt aux personnes aveugles comme communauté cible principale. En ce qui concerne les langues, même si dans les chapitres plus théoriques, la plupart de nos arguments ont été présentés sans faire référence à aucune paire de langues, notre travail empirique (l'élaboration des règles de LC et l'étude expérimentale) a porté principalement sur la combinaison anglais-français. Ces deux limites, ainsi que la taille relativement petite des échantillons des populations que nous avons étudiées, pourraient être considérées comme des faiblesses de notre recherche.

Tout au long de notre travail de thèse, nous avons essayé de décrire les mesures adoptées pour compenser ces limites générales, notamment en vue de favoriser la validité interne et écologique de nos études, ainsi que de garantir leur reproductibilité. Ce faisant, nous avons aussi souligné régulièrement les limites associées à ces mesures. Dans ce qui suit, nous présentons les principales lacunes des approches adoptées pour atteindre nos trois objectifs. Cette dernière partie du résumé servira également à présenter d'éventuelles pistes pour de futurs travaux de recherche.

L'**objectif 1** a été abordé d'un point de vue assez général. Le fait que nous avons seulement compté sur l'étude de la littérature et sur les opinions subjectives des experts en accessibilité web pourrait rendre nos conclusions moins généralisables. D'autres études basées sur des observations directes pourraient être menées afin d'examiner de plus près les pratiques d'évaluation de l'accessibilité suivies par ces professionnels lorsqu'ils travaillent avec des sites web d'au moins deux versions linguistiques différentes. En outre, il serait intéressant de comparer les politiques et les législations portant sur des aspects associés à l'accessibilité du Web des pays plurilingues et celles des pays monolingues, en vue de comprendre s'il existe des méthodes ou des directives internes qui ont pour but de veiller à ce que l'information d'intérêt public soit accessible pour les locuteurs de toutes les langues utilisées dans les territoires nationaux. Nos résultats ont également soulevé des questions importantes en ce qui concerne l'état actuel des sites web multilingues en termes de conformité aux WCAG 2.0. À cet égard, des études devraient être menées pour savoir si chaque version linguistique a un niveau de conformité différent, ainsi que pour estimer l'impact potentiel que ces différences pourraient avoir, par exemple, sur l'expérience de navigation des utilisateurs finaux bilingues. Malgré ces limitations, nous voyons notre rapport général sur l'assurance de l'accessibilité des sites multilingues comme un point de départ bien motivé pour de futurs travaux de recherche dans le domaine de l'accessibilité et de la localisation.

Pour atteindre l'**objectif 2,** certains compromis ont dû être faits. Le premier est lié aux caractéristiques du corpus web choisi pour l'élaboration des règles de LC. Nous sommes consciente que le type et la quantité d'images utilisées dans un site web peuvent varier considérablement entre différents genres de texte. En ce sens, la formalisation des recommandations linguistiques sur la manière d'écrire des équivalents textuels plus pertinents aurait pu se révéler plus difficile si, par exemple, nous avions pris comme référence des sites de tourisme ou commerciaux, au lieu de sites web gouvernementaux et du secteur public, surtout parce que ces derniers contiennent généralement des images fonctionnelles, tandis que les premiers ont régulièrement un nombre plus élevé d'images avec une valeur descriptive. Pour assurer la portabilité des règles élaborées, des études futures pourraient évaluer dans quelle mesure nos règles aident les évaluateurs lors de la validation de l'accessibilité des images contenues dans d'autres types de sites web. Précisément en lien avec l'approche d'évaluation des règles adoptée, il nous semble important de noter qu'une autre limite de notre étude est que l'efficacité des règles a été mesurée en prenant compte uniquement de la qualité du produit final localisé. Nous pensons qu'une analyse plus exhaustive des données pourrait rapporter des résultats intéressants en ce qui concerne l'utilisabilité de chaque règle en fonction des connaissances sur l'accessibilité web de la personne qui l'applique. De même, l'étude des rapports générés par Acrolinx lors de chaque test effectué et les vidéos résultant de l'enregistrement des écrans des participants pourrait donner de nouvelles pistes sur des structures linguistiques que nous n'avons pas couvertes dans nos règles et qui pourraient avoir été proposées spontanément par les localisateurs. *Last but not least*, il serait pertinent d'examiner si le même niveau de performance d'Acrolinx et de ses règles est observé lorsque l'outil est utilisé par d'autres professionnels avec un profil plus technique, tels que les développeurs ou les concepteurs de sites web.

Enfin, un certain nombre de limites doivent être prises en compte lors de l'interprétation des résultats associés à l'**objectif 3.** Certaines limites sont liées à la conception de l'étude expérimentale elle-même. Par exemple, nous avons développé un site expérimental essentiellement statique (ou semi-statique si l'on tient compte de l'introduction de certains morceaux de code JavaScript), afin de produire des documents HTML simples pour ensuite faciliter l'évaluation d'autant de cas problématiques liés à l'accessibilité des images que possible. À l'avenir, il serait important d'explorer également comment l'accessibilité de ces éléments est abordée par les localisateurs qui travaillent sur des sites web dynamiques. De même, compte tenu de la grande quantité de données recueillies, il a été décidé d'adopter une approche principalement quantitative pour opérationnaliser la traduction et la pertinence des équivalents textuels. Bien que cela soit, en même temps, l'un des points forts de notre travail – étant donné que nous avons appliqué des tests statistiques inférentiels avancés et rapporté sur la signification de nos résultats–, il serait souhaitable d'examiner aussi les aspects qualitatifs. Cela nous permettrait d'étendre nos conclusions par rapport au niveau d'accessibilité des images obtenu dans le produit cible.

Le profil des localisateurs engagés pour notre étude a pu être aussi une source de biais puisque la majorité d'entre eux travaillaient à titre d'indépendant (*freelance*). On pourrait penser que les équipes de localisation des grandes entreprises sont plus spécialisées et prêtent plus d'attention aux questions d'accessibilité (peut-être en raison des avantages (au niveau juridique et financier) de l'accessibilité pour leurs patrons). En ce sens, il est possible que si des évaluateurs experts avec un profil de localisation avaient participé à l'étude, nos résultats en ce qui concerne l'importance d'avoir des connaissances sur l'AW auraient pu être différents. De même, l'impact positif de l'utilisation des outils d'assurance qualité pour contrôler l'accessibilité pourrait avoir été moins concluante. Néanmoins, nous pensons que notre étude et ses conclusions sont toujours pertinentes, en particulier si l'on tient compte de la réalité du secteur de la localisation actuelle, où la tendance est de plus en plus à externaliser les services. En outre, notre protocole expérimental nous a également permis de découvrir le potentiel didactique des deux logiciels testés, qui pourraient être utiles pour sensibiliser les localisateurs à l'accessibilité pendant le processus d'assurance qualité en localisation.

Cela est directement lié à une autre limite de notre étude qui, en même temps, détient le plus grand potentiel pour nos travaux de recherche futurs : peu d'attention a été accordée, en général, aux données collectées sur les processus de localisation. À notre avis, les enregistrements d'écran en particulier seront une source précieuse d'information pour poursuivre nos recherches en matière de localisation et d'accessibilité. Sur la base des données tirées de ces vidéos (par exemple, le temps consacré à inspecter l'aperçu visuel du site lors de l'étape de traduction ou les stratégies suivies pour résoudre les problèmes du code source), nous pourrions essayer de déterminer comment les meilleures pratiques d'accessibilité pourraient être mieux intégrées dans le processus de localisation, ce qui éviterait des efforts inutiles de reconception à des stades ultérieurs du cycle de développement web multilingue. Les conclusions tirées de l'analyse de ces enregistrements pourraient ensuite être comparées et étendues, par exemple, avec d'autres méthodes typiques de la recherche axée sur le processus de traduction, telles que des études de suivi oculaire (*eye-tracking*).

Nous croyons que ces analyses supplémentaires aideront également à mieux définir la formation en accessibilité des futurs professionnels de la localisation. Bien que la pertinence de l'acquisition de connaissances sur l'AW ait généralement été appréciée par les étudiants qui ont participé à nos séminaires sur le sujet et que nos matériels du cours pilote semblent avoir été efficaces (Rodríguez Vázquez 2014), la sélection des contenus montrés aux étudiants jusqu'à présent n'était pas basée sur des études scientifiques. En outre, l'enseignement de l'accessibilité web comme un module indépendant est souvent considéré comme une technique pédagogique inefficace. Nous considérons que l'enseignement de l'accessibilité devrait être intégré peu à peu tout au long des cours de localisation, comme il a déjà été fait dans des cours d'informatique et de sciences de l'information (Waller et al. 2009; Wang 2012). En ce sens, à l'avenir, notre recherche pourrait être fondée sur le cadre pédagogique que nous sommes en train de développer actuellement et qui défend une approche communicative et sociale dans l'enseignement de localisation (Torres del Rey et al. 2014) :

> « Localisation students must learn to play a three-fold role, vis-a-vis the digital product and the production process, as: **mediators** of the communicative value generated by the digital product as a cultural object and

as a technical extension of human physical and intellectual abilities; **stakeholders** in the distribution of the localisable object's functional and informative values; and **negotiators** of their own (as professionals) and of their localisable technological product's social position and responsibility ».