

UNIVERSIDAD DE SALAMANCA  
FACULTAD DE TRADUCCIÓN Y DOCUMENTACIÓN  
GRADO EN TRADUCCIÓN E INTERPRETACIÓN  
Trabajo de Fin de Grado

El uso de herramientas de corpus en la  
preparación del intérprete:

Guía básica y aplicaciones prácticas de Sketch Engine

Autor:

Jesús Melcón Vega

Tutor:

Joaquín García Palacios

Salamanca, 2019

## **RESUMEN**

La terminología y la interpretación de conferencias forman un vínculo indisoluble. Dadas las particularidades de la interpretación y los condicionantes a los que está sujeto el intérprete, es indispensable emprender una buena preparación terminológica para garantizar la calidad de la interpretación. En el presente trabajo hacemos hincapié en dicho vínculo, revisamos los flujos de trabajo terminológicos descritos en la bibliografía y proponemos el nuestro propio, en el que se emplea el programa de gestión de corpus Sketch Engine. La finalidad de este trabajo es proporcionar al intérprete nuevas herramientas para gestionar la preparación terminológica previa al encargo de manera ágil y automatizada.

**Palabras clave:** interpretación de conferencias, terminología, corpus, Sketch Engine, extracción automática de términos.

## **ABSTRACT**

It is safe to say that terminology and conference interpreting are inextricably linked. Given the special features of interpreting and the constraints to which the interpreter is subject, it is essential to undertake a good terminological preparation in order to guarantee the quality of interpretation. In this work we emphasize the aforementioned link, review the terminological workflows described in the literature and propose our own, in which the Sketch Engine corpus management software is used. The purpose of this work is to provide the interpreter with new tools to manage his or her advance terminological preparation in a fast and automated way.

**Keywords:** conference interpreting, terminology, corpora, Sketch Engine, automatic term extraction.

# ÍNDICE

INTRODUCCIÓN.....	1
<b>CAPÍTULO 1: TERMINOLOGÍA E INTERPRETACIÓN DE CONFERENCIAS.....</b>	<b>3</b>
1.1 La importancia de la terminología para el intérprete.....	3
1.2 La preparación terminológica del intérprete.....	6
1.2.1 Preparación terminológica vs. preparación temática.....	7
1.2.2 Beneficios de la preparación previa.....	8
<b>CAPÍTULO 2: ENFOQUES Y PROCESOS PARA LA PREPARACIÓN TERMINOLÓGICA DEL INTÉRPRETE.....</b>	<b>13</b>
2.1 Fuentes de documentación y recursos para los intérpretes de conferencias.....	13
2.1.1 Información por parte del cliente.....	13
2.1.2 Fuentes de información alternativas.....	14
2.2 Flujos de trabajo para la preparación terminológica.....	15
2.2.1 El flujo de trabajo del intérprete para Moser-Mercer.....	16
2.2.2 El flujo de trabajo del intérprete para Will.....	16
2.2.3 La propuesta de flujo de trabajo en interpretación de Seghiri.....	18
2.2.4 La propuesta de flujo de trabajo en interpretación de Xu.....	19
<b>CAPÍTULO 3: EL USO DE CORPUS EN EL FLUJO DE TRABAJO DEL INTÉRPRETE.....</b>	<b>20</b>
3.1 Tipos de corpus que se pueden emplear.....	20
3.2 Beneficios del uso de corpus durante la preparación terminológica.....	22
3.3 El uso de Sketch Engine para la labor terminológica previa del intérprete.....	24
3.3.1. Pasos preliminares.....	25
3.3.2 Primera fase: creación de la lista de términos.....	30
3.3.2.1 Extracción de la fraseología.....	33
3.3.3 Segunda fase: consulta de información.....	35
3.3.4 Tercera y cuarta fase.....	41
3.4 Empleo de Sketch Engine para un encargo real.....	42
3.4.1 Creación de un corpus paralelo.....	42
3.4.2 Extracción de términos.....	46
3.4.3 Concordancia bilingüe y búsqueda de equivalentes terminológicos.....	48
CONCLUSIÓN.....	50
BIBLIOGRAFÍA.....	52
<b>ANEXO: GLOSARIO REALIZADO CON SKETCH ENGINE MEDIANTE EL FLUJO DE TRABAJO DESCRITO PARA INTERPRETAR EN LA ONU.....</b>	<b>57</b>

## **INTRODUCCIÓN**

Como apuntan Rodríguez y Schnell (2009), no cabe duda de que la terminología en el campo de la interpretación no ha recibido una atención proporcional a su importancia en la labor del intérprete; las autoras explican que el ligero interés que suscitó la terminología en el campo de la interpretación a finales de los años 80 de la mano de estudiosos como Gile no se vio avivado hasta comienzos del siglo XXI, en esta ocasión ligado al uso de la tecnología. Además, las autoras afirman que los estudiosos siguen centrándose en estudiar la terminología desde el punto de vista del traductor, dejando de lado las necesidades del intérprete. En efecto, en el campo de los estudios de traducción encontramos una extensa bibliografía sobre el solapamiento entre la terminología y la labor del traductor (*cf.* Cabré 2000a, Cabré 2010, Alcina Caudet y Gamero Pérez 2002, Faber 2010). No es así en el campo de los estudios de interpretación, en el que la atención que ha suscitado esta relación no ha sido tan acusada. Por ello, en el primer capítulo del presente trabajo presentaremos el marco teórico existente que explica las particularidades de la labor del intérprete y su relación con la terminología. Debido a estas particularidades de la interpretación y a los condicionantes a los que se ve sometido el profesional, la labor terminológica previa a la conferencia adquiere una importancia máxima para asegurar la calidad, salvar escollos y evitar problemas en cabina. En consecuencia, en el segundo capítulo se presentarán los flujos de trabajo terminológicos para intérpretes descritos por la bibliografía y los enfoques y procesos que adoptan los intérpretes profesionales para esta preparación. Llama la atención, en este sentido, que dichos procesos sigan siendo eminentemente tradicionales, sobre todo si tenemos en cuenta que algunos autores (Fantinuoli 2018a) ya anuncian el inminente “giro tecnológico de la interpretación”. La realidad, no obstante, parece ser muy distinta. Aunque algunos estudiosos (Costa *et al.* 2014, Costa *et al.* 2015, Corpas Pastor 2018) se estén centrado en la intersección entre interpretación y tecnología, no parece que los profesionales hayan adoptado estas herramientas, e incluso es posible que las desconozcan (Corpas Pastor 2016). El caso de la preparación terminológica no es una excepción, pues la bibliografía muestra que el uso de herramientas tecnológicas sigue siendo mínimo. El presente trabajo se enmarca en una línea de investigación reciente que pone de manifiesto los beneficios del uso de corpus por parte de los intérpretes y que aboga por su uso durante el estudio de la terminología. Este trabajo pretende contribuir a subsanar algunas de las carencias tecnológicas del profesional actual presentando una herramienta de gestión de corpus (Sketch Engine) y

un flujo de trabajo con la misma para ayudarle a adoptar en su labor previa de preparación nuevas tecnologías que aumenten su eficiencia.

## **CAPÍTULO 1: TERMINOLOGÍA E INTERPRETACIÓN DE CONFERENCIAS**

### **1.1 La importancia de la terminología para el intérprete**

Resulta llamativo el vacío bibliográfico mencionado sobre todo si tenemos en cuenta que la relación existente entre terminología e interpretación es prácticamente indisoluble. En este capítulo pretendemos hacer hincapié en este vínculo exponiendo las razones que hacen de la terminología un elemento esencial de la interpretación. Con ello, queremos resaltar las necesidades terminológicas específicas del intérprete y poner de manifiesto las condiciones inherentes a su labor relacionadas con la terminología.

En primer lugar, por la evolución de un mercado cada vez más exigente, se aprecia que el intérprete se ve forzado a interpretar en conferencias internacionales o reuniones muy especializadas y de temática muy diversa (Kalina 2006 en Fantinuoli 2017). Es decir, aunque algunos intérpretes se especialicen en un ámbito concreto (Valentini 2002) es común que el intérprete autónomo interprete en congresos de temáticas muy dispares, cada cual con su propia terminología. De hecho, como revela una encuesta reciente de Pöchhaker y Zwischenberger (2010), los intérpretes autónomos componen el grupo más común (89 %). Como *freelance*, se enfrentan a un mercado mucho más variado que los intérpretes de plantilla y, por lo tanto, se encuentran en contacto con la terminología de numerosos ámbitos de especialidad.

Por otro lado, durante la interpretación simultánea de una conferencia, nuestro cometido es transmitir a un auditorio que desconoce el idioma original el conocimiento especializado que expone el orador. De hecho, la mayoría de conferencias y reuniones del sector privado se caracterizan por lo general por su densidad informativa y la dificultad del contenido (Gile, 1986). Y como explica Cabré (2000b), son las unidades terminológicas las que se encargan precisamente de representar y transmitir este conocimiento especializado. Como expone la autora, con el uso del término en unas condiciones discursivas concretas se activarán una serie de informaciones pragmáticas y un sentido unívoco que comparte y sanciona la comunidad de expertos. Por ello, para conseguir una transmisión unívoca y clara del conocimiento, el intérprete ha de emplear correctamente las unidades terminológicas que lo vehiculan. De hecho, a lo largo de la conferencia son principalmente los términos los que ponen trabas a la comunicación. Por

un lado, son imprescindibles durante la fase de comprensión para entender y transmitir correctamente el discurso original. Por otro, es necesario conocerlos en la fase de producción en lengua meta para generar frases cortas y concisas y evitar los hiperónimos, paráfrasis y otras estrategias que atentan contra la economía del lenguaje (Fantinuoli, 2017).

Por otra parte, la bibliografía nos muestra las altas expectativas que albergan los clientes, pues exigen un dominio elevado de la terminología por parte del intérprete. Es decir, no solo empleamos la terminología para expresar correctamente el mensaje, sino también para satisfacer las expectativas de nuestros oyentes y, por ende, del cliente. Como revela Kurz (1993) en una encuesta a usuarios de interpretación (participantes de dos grandes congresos internacionales y miembros del Consejo de Europa), las expectativas son, en efecto, muy altas. Kurz expone que “[t]hey expect the interpreter to use the technical jargon they are used to and seem to be less tolerant toward any deviations in this respect<sup>1</sup>”. De hecho, para los encuestados la terminología era de media el tercer factor más importante. No obstante, para los miembros del Consejo de Europa se trataba de la característica más importante de una buena interpretación (por encima de rasgos como la fluidez o la restitución íntegra del contenido). Esto parece indicar que las diferentes comunidades de hablantes podrían tener diferentes expectativas terminológicas, aspecto que podría ser relevante en la preparación del intérprete. Por otro lado, Moser (1995) alcanza unos resultados comparables en su informe para la Asociación Internacional de Intérpretes de Conferencias (AIIC). Las respuestas de los encuestados (usuarios de interpretación en congresos generales y técnicos) revelan, de nuevo, el papel crucial de la terminología a la hora de satisfacer las expectativas del cliente. Como es de esperar, los asistentes a conferencias técnicas otorgan gran importancia a la terminología, como se puede observar en la figura 1 (la consideran “bastante importante” o “muy importante”). No obstante, en las conferencias menos especializadas también se le otorga cierta relevancia, como se deduce del reducido número de respuestas que consideran que “no tiene importancia”. Por otro lado, se puede apreciar en la figura 2 que para los asistentes a conferencias técnicas la terminología sería incluso más importante que la transmisión íntegra (“*completeness*”) del mensaje. A la vista de estos datos queda claro que para satisfacer las expectativas del cliente resulta imprescindible el dominio de la

---

<sup>1</sup> “Pretenden que el intérprete emplee los tecnicismos con los que están familiarizados y, en este sentido, parecen ser menos tolerantes con cualquier desviación”. (La traducción es mía).

terminología. No en vano, usar una terminología deficiente equivale a perder la credibilidad como profesionales (Gile, 2009).

## AIIC

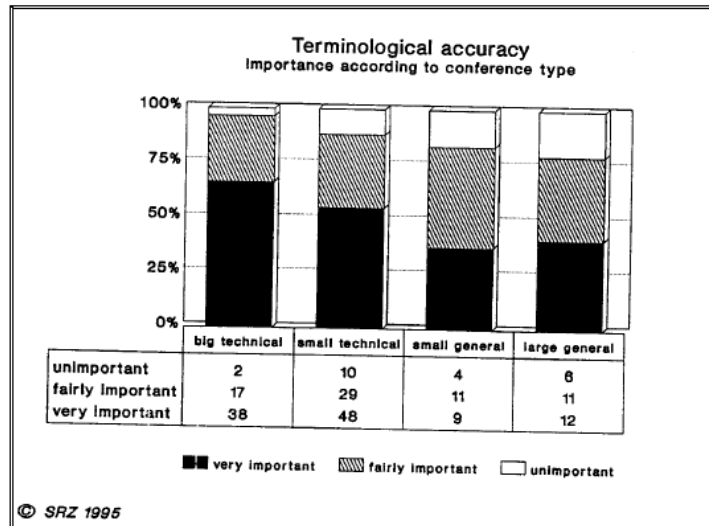


Figura 1, Moser (1995)

## AIIC

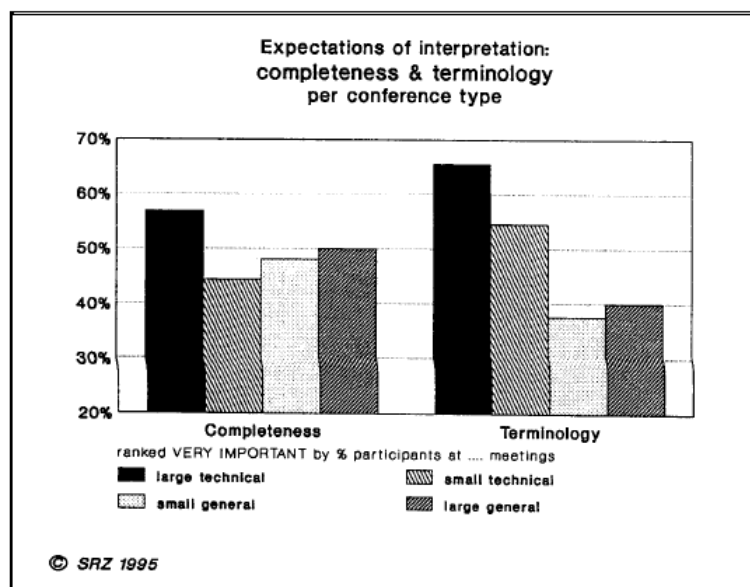


Figura 2, Moser (1995)

Otra dificultad añadida en lo referente a la terminología es la inmediatez inherente a la interpretación. De hecho, es aquí donde radica una de las principales diferencias entre

traducción e interpretación a todos los niveles, incluido el terminológico. Los traductores pueden construir su conocimiento *ad hoc* a medida que surgen las dificultades terminológicas y de comprensión para poder superar estas barreras. En el caso del intérprete existe una incertidumbre mucho mayor; pues debe preparar la terminología de antemano y sin saber a ciencia cierta qué problemas terminológicos le planteará la conferencia o si los términos en que se está centrando le acabarán resultando útiles (Fantinuoli, 2017). En resumen, en interpretación no queda más remedio que adquirir conocimientos *ad hoc* antes del encargo (con la incertidumbre que ello conlleva), mientras que en traducción podemos permitirnos hacerlo sobre la marcha.

## **1.2 La preparación terminológica del intérprete**

Ha quedado claro que la terminología forma un vínculo inextricable con la interpretación y que puede suponer un escollo debido a la transmisión de conocimiento muy especializado, a las expectativas de los usuarios y a la inmediatez de la interpretación, que obliga a una adquisición de conocimiento previa. Es decir, la terminología le puede plantear al intérprete una serie de problemas, y dada la gran diversidad de temas y su complejidad, el único modo de solventarlos y salvar la brecha de conocimiento entre el intérprete y la comunidad de expertos es una preparación intensiva *ad hoc*. Como recuerda Gile (1986), sin la preparación el intérprete solo puede apoyarse en su técnica y en las estrategias interiorizadas (simplificación del discurso, naturalización, omisión etc.). Ahora bien, al ser tácticas que entrañan una pérdida informativa, es preferible centrarse en una buena preparación. El mismo autor (2009) crea una división de las diferentes fases de la preparación del intérprete: la preparación previa, la preparación de último minuto (por ejemplo, la obtención de documentos o de diapositivas a última hora) y la preparación durante la conferencia (por ejemplo, escuchar al resto de oradores para captar terminología y fraseología). No obstante, recalca Gile, la preparación previa es la más relevante de todas. Si bien es cierto que parte de la bibliografía se centra en la consulta de terminología durante la conferencia mediante diversas herramientas (por ejemplo Costa *et al.*, 2015 y Costa *et al.*, 2014), la mayor parte estudia esta fase previa. En el presente trabajo trataremos esta fase, pues en ella se concentra el grueso del trabajo del intérprete.

### **1.2.1 Preparación terminológica vs. preparación temática**

En lo referente a esta primera fase previa, existe un debate sobre si el intérprete debería centrarse en la información intralingüística (la terminología y fraseología) o la información extralingüística (información enciclopédica). Gile (1986) sostiene que en la documentación extralingüística empezaríamos consultando obras de divulgación para informarnos sobre el tema. Se consultarían progresivamente obras de mayor dificultad hasta llegar al nivel de especialización de nuestra conferencia, incluido el vocabulario. El enfoque alternativo es meramente lingüístico: detectar los términos y buscar sus equivalentes en lengua meta, así como la información fraseológica y estilística “*for the purpose of reformulating the message along the same stylistic lines and phraseology*”<sup>2</sup> (*ibidem*). Si bien autores como Seleskovitch (1968) afirman que para interpretar sería necesario al menos haber interiorizado unos conceptos básicos sobre la disciplina en cuestión, Gile (1986) concluye que una comprensión completa es raramente posible y que ante la falta de tiempo un enfoque terminológico resultaría mucho más efectivo. Para el autor, conocer las equivalencias terminológicas es suficiente para permitir la interpretación. Gile (2009) postula asimismo que, dado que la mayoría de las dificultades del intérprete son de orden léxico o terminológico, interesa centrarse en un enfoque lingüístico. Otro punto a favor de la preparación terminológica es el hecho de que en ocasiones la preparación temática resulta simplemente imposible. Como revela el estudio de Valentini (2002), el 56,2 % de los intérpretes encuestados recibe documentación del congreso con menos de cinco días de antelación. Según la autora, ante esta perspectiva a los intérpretes no les queda otra opción que emprender una preparación estrictamente terminológica. Nosotros pensamos que, si bien es cierto que en ocasiones se puede interpretar sin comprender, conociendo tan solo los equivalentes terminológicos (Gile 2009), la información extralingüística entraña una gran importancia para evitar contrasentidos en la prestación, facilitar la comprensión y anticipar información. En consecuencia, sería deseable combinar estos dos enfoques a la hora de preparar un encargo. Ahora bien, en la mayor parte de ocasiones deberemos priorizar el enfoque terminológico por razones pragmáticas (falta de tiempo). En el tercer capítulo presentaremos un flujo de trabajo con corpus que permita combinar estos dos enfoques sin que ello resulte en un aumento del tiempo de preparación.

### **1.2.2 Beneficios de la preparación previa**

Aunque la importancia de la preparación terminológica pueda resultar obvia en el contexto de la interpretación de conferencias, lo cierto es que existen pocos estudios que aporten pruebas empíricas al respecto. Una de las pocas autoras que han investigado empíricamente los efectos beneficiosos de la preparación sobre el rendimiento del intérprete ha sido Díaz Galaz (2012) en su tesis doctoral. La autora prueba la validez de la hipótesis clásica de que el conocimiento previo ayuda a un procesamiento más rápido de la información y a una mejora del rendimiento y de la precisión. Para ello, realiza un estudio experimental con estudiantes y profesionales de la interpretación a los que proporciona 30 minutos antes del ejercicio las diapositivas de la presentación del ponente, un breve resumen del discurso, el programa de la conferencia, la biografía de los ponentes y un glosario bilingüe. Como se puede apreciar, la autora proporciona información de los dos tipos mencionados (extralingüística e intralingüística), pero el peso de esta última es ciertamente bastante importante (estaría presente en el glosario, las diapositivas y la biografía del ponente). Los resultados muestran que la preparación previa mejoró en gran medida el rendimiento de estudiantes y profesionales en comparación con el grupo de control. En el caso de los estudiantes, la preparación permitió el uso de un vocabulario más rico y un uso correcto de la terminología en segmentos de todo tipo de dificultad. Por el contrario, la falta de preparación estaba directamente relacionada con una disminución de la precisión léxica y un mal rendimiento en segmentos con cierta densidad terminológica. En general, la preparación desembocó en una mayor precisión a todos los niveles. El análisis de la autora del rendimiento de intérpretes profesionales es más detallado y muestra un *décalage* menor y una mayor precisión léxica, sintáctica y semántica. El grupo que obtuvo los documentos de preparación pudo hacer frente con más facilidad a los segmentos con gran densidad terminológica. Además, este grupo recurrió más frecuentemente a los equivalentes directos en lengua meta, empleando en menor medida que el grupo de control estrategias como la paráfrasis o la omisión de la información. En resumen, este trabajo nos proporciona una prueba empírica de que la consulta de información previa mejora el rendimiento y la precisión en interpretación.

---

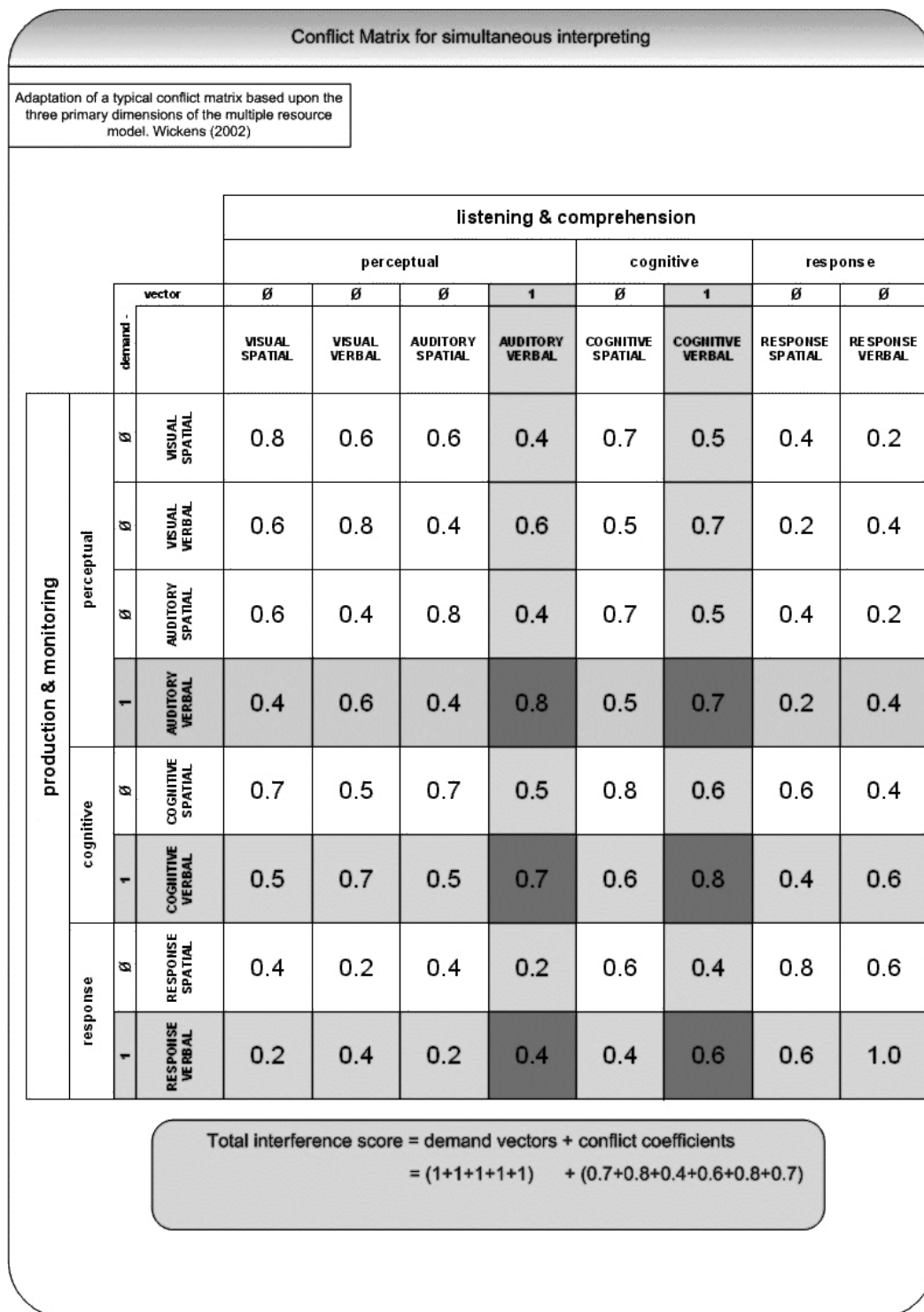
<sup>2</sup> “Con vistas a reformular el mensaje empleando el mismo estilo y la misma fraseología”. (La traducción es mía).

Otra conclusión interesante que se desprende del estudio es que la preparación previa aligeraría la carga cognitiva del intérprete<sup>3</sup>. Resulta fructífero ahondar en esta relación, pues en muchas ocasiones la densidad terminológica en el discurso original es la responsable de sobrecargar los procesos cognitivos del intérprete. De estudiar la carga cognitiva del intérprete durante su labor se han encargado autores como Seeber (2011), que ha propuesto un modelo de la carga cognitiva en interpretación simultánea. El autor se distancia de otros modelos de recursos cognitivos centralizados y se centra en el modelo de recursos múltiples de Wickens para predecir la carga cognitiva y la interferencia en interpretación simultánea, en la que para el autor se produce una combinación de una tarea de comprensión lingüística y de otra de producción lingüística. El autor expone que la combinación de varias tareas entraña más limitaciones y genera dificultades que no son inherentes a la tarea en cuestión (por ejemplo, la escucha), sino que surgen precisamente de la simultaneidad de los procesos. La razón es que, como expone el autor, *“the combination of two (or more) tasks requires more processing capacity than either (or any) of the tasks performed individually<sup>4</sup>”*. Para sustentar su afirmación de que con la simultaneidad de tareas surgen complicaciones no inherentes a las tareas por separado, el autor menciona los estudios realizados hasta el momento, que muestran que la comprensión del discurso no se ve alterada significativamente por su ritmo en palabras por minuto. De hecho, incluso a velocidades de 300-500 palabras por minuto la comprensión no se resiente significativamente. Ahora bien, si la comprensión se conjuga con la producción verbal (interpretación simultánea), se crean nuevas dificultades que no existían individualmente: como muestran los estudios, cuanto mayor sea la velocidad, más errores de pronunciación, omisiones y sustituciones realiza el intérprete. Además de la velocidad del discurso, otros factores que incrementarían la carga cognitiva serían la densidad semántica o ciertas cuestiones sintácticas. Otro aspecto importante del modelo es el hecho de que la interferencia entre tareas será aún mayor si sumamos a la simultaneidad el uso del mismo canal de procesamiento. Con base en estas afirmaciones, Seeber predice mediante una matriz computacional la cantidad de recursos que emplea cada tarea y su interferencia con otras que tratan de emplear los mismos recursos comunes

---

<sup>3</sup> La afirmación se desprende, como comenta la autora, del hecho de que el *ear-to-voice span (décalage)* es menor en el grupo de estudio que en el grupo de control, lo cual evidencia una capacidad de procesamiento más rápida.

<sup>4</sup> “La combinación de dos (o más) tareas requiere una mayor capacidad de procesamiento que cualquiera de las tareas realizada de forma individual”. (La traducción es mía).



*Figura 3. Matriz de conflicto de Seeber (2011) en interpretación simultánea*

Como se aprecia, la tarea de escucha y comprensión y la tarea de producción se dividen en sus componentes de demanda cognitiva (líneas en gris claro). La escucha se divide en un esfuerzo cognitivo (comprender) y un esfuerzo auditivo, mientras que la producción se compone de un esfuerzo de respuesta verbal (reproducir el discurso), uno cognitivo (comprender lo que decimos) y otro auditivo (escuchar nuestra prestación).

Como se puede apreciar en las casillas en gris oscuro, se puede producir un solapamiento entre los esfuerzos que provoquen una interferencia; los esfuerzos entran en conflicto al usar un mismo canal de procesamiento, pues compiten por unos recursos cognitivos comunes. En este sentido, Seeber se centra en estudiar la carga cognitiva que produce el contraste sintáctico sujeto-objeto-verbo vs. sujeto-verbo-objeto en la combinación DE > EN. Ahora bien, este mismo modelo se podría aplicar a una vertiente terminológica.

El autor expone que la simultaneidad de tareas genera dificultades extra que no son inherentes a las tareas por separado, por lo que se podría afirmar que si el intérprete se enfrentase a segmentos terminológicamente densos o a terminología inesperada, se producirían dificultades adicionales que no existirían en una escucha monolingüe. Con la simultaneidad de los procesos, el esfuerzo perceptual y cognitivo extra que puede suponer la terminología interferiría con otros procesos que empleen el mismo canal de procesamiento y que compitan por unos recursos compartidos. De este modo, se podría resentir la escucha y control de la propia producción y la respuesta verbal. Cabe decir, por lo tanto, que con una buena preparación terminológica el esfuerzo cognitivo y perceptual en la fase de escucha sería menor, lo que generaría menos interferencia con el resto de procesos y aligeraría la carga cognitiva del intérprete. Esta afirmación no es meramente hipotética. De hecho, ya hemos comentado cómo los participantes del estudio de Díaz Galaz (2012) que contaron con preparación terminológica evidenciaron un menor *décalage*, lo cual pone de manifiesto una menor carga cognitiva.

Una alternativa al modelo de Seeber que también se ha empleado para predecir la carga cognitiva de los intérpretes es el modelo de esfuerzos de Gile (2009). El postulado fundamental del modelo es que durante toda interpretación coexisten cuatro esfuerzos, cada uno de los cuales emplea parte del conjunto común de recursos cognitivos del intérprete. El primer esfuerzo (esfuerzo de escucha y análisis) se trataría, según Gile, de todas las operaciones mentales encaminadas a permitir la escucha del mensaje y la comprensión de su contenido. El segundo esfuerzo (esfuerzo de producción) se define como las operaciones mentales que permiten al intérprete repetir en lengua meta el discurso original. Por su parte, el esfuerzo de memoria a corto plazo comprende el proceso cognitivo que realiza el intérprete para retener en la memoria información mientras realiza otras actividades (escucha o producción de la prestación). A estos tres esfuerzos Gile añade un cuarto, el de coordinación, que es el resultado de la simultaneidad de los

procesos. Las necesidades totales (T) equivalen a la suma de estos cuatro procesos (E + P + M + C). Según Gile, para una buena interpretación sería preciso que las necesidades totales fueran menores que la capacidad de tratamiento total disponible ( $T \leq D$ ). Teniendo en cuenta esta inecuación, las fuentes de problemas para el intérprete pueden ser diversas. En unos casos, el intérprete puede experimentar saturación; es decir, el total de necesidades superaría la capacidad de tratamiento disponible. En otros casos se experimentarían déficits individuales: la capacidad de tratamiento total es mayor que las necesidades totales, pero en un esfuerzo concreto la capacidad de tratamiento necesaria para la realización de la tarea no basta. Por otro lado, Gile denomina *déclencheurs* (“desencadenantes”) a los factores que generan saturación o déficits. Estos factores están ligados a un aumento de las necesidades de tratamiento, y pueden estar causados por una tasa de palabras por minuto elevada o por la densidad del discurso. En el caso de un discurso con gran densidad terminológica, por ejemplo, aumentarían las necesidades de capacidad de tratamiento en dos esfuerzos: escucha y análisis y producción. El segundo aspecto clave de la teoría de Gile son los denominados “encadenamientos deficitarios”. Según el autor, los esfuerzos se pueden descompensar. Por ejemplo, si escuchamos un acento fuerte, dedicamos mucha capacidad de tratamiento al esfuerzo de escucha. En consecuencia, se produce un déficit en el esfuerzo de producción (menos fluidez, prestación más lenta). El retraso se va acumulando y el orador sigue hablando, por lo que aumenta la información retenida en la memoria a corto plazo. Llegado a un cierto punto la información retenida sobrepasa ciertos límites y se produce la saturación de la capacidad cognitiva. En nuestro caso, la densidad terminológica tendría un efecto similar en la labor del intérprete, pues obligaría a concentrar esfuerzos en el esfuerzo de escucha y análisis (identificar y comprender el término) y en el de producción (recordar y pronunciar el equivalente). El encadenamiento de retrasos podría acabar conduciendo, como se ha visto, a una saturación. De ahí la importancia de una preparación óptima, pues aligeraría la carga de los esfuerzos ligados a la terminología y permitiría destinar más recursos a otros esfuerzos.

## **CAPÍTULO 2: ENFOQUES Y PROCESOS PARA LA PREPARACIÓN TERMINOLÓGICA DEL INTÉRPRETE**

En vista de lo expuesto en el capítulo anterior, ha quedado claro que la preparación terminológica es indispensable para el intérprete si queremos asegurar la calidad y aligerar la carga cognitiva dados los rasgos inherentes de su labor (gran especialización, transmisión de conocimiento especializado, altas expectativas por parte de los usuarios etc). Ahora bien, no nos hemos adentrado en las fuentes y métodos a los que puede recurrir el intérprete para gestionar su preparación. Tras poner de manifiesto la importancia de la preparación terminológica, revisaremos a lo largo de este capítulo cómo se aborda esta preparación y qué enfoques, flujos de trabajo y herramientas se emplean.

### **2.1 Fuentes de documentación y recursos para los intérpretes de conferencias.**

#### **2.1.1 Información por parte del cliente**

En teoría, el punto de partida de la documentación del intérprete debería ser siempre el cliente. Según Gile (2009) el organizador debería facilitar a los intérpretes, entre otros, el programa de la conferencia, la lista de participantes y, sobre todo, documentos relacionados con el contenido de las ponencias, ya sean resúmenes, borradores de documentos o documentos finalizados que se vayan a leer durante la conferencia. El autor resalta asimismo la importancia de las reuniones informativas previas (*briefings*), aunque advierte que no son frecuentes. Por su parte, la AIIC (Asociación Internacional de Intérpretes de Conferencias) también nos recomienda centrarnos en los documentos de la conferencia, y enumera una lista mucho más extensa que la de Gile<sup>5</sup>; a saber: el programa del congreso, documentos con información sobre los temas y organizaciones involucradas, los documentos que se tratarán, los textos que se leerán, las presentaciones de los oradores (a poder ser con notas), actas de reuniones o congresos previos, la lista de oradores o de delegados y las notas biográficas de los ponentes. La asociación subraya además el valor de las reuniones informativas previas en las que se puedan resolver las dudas terminológicas del intérprete. Por otro lado, la AIIC advierte que, si bien en su contrato estándar se obliga a los organizadores a proporcionar documentación a los intérpretes con suficiente antelación, en la práctica casi nunca ocurre. Por último, Lucarelli (2013) recomienda a los contratantes de servicios de

---

<sup>5</sup> <https://aiic.net/page/628/practical-guide-for-professional-conference-interpreters/lang/1#2>. Última consulta: 18/06/2019

interpretación que proporcionen a los profesionales los siguientes recursos: el programa de la conferencia, los materiales a los que tengan acceso el resto de participantes, el sitio web del congreso o conferencia y los discursos y presentaciones de los oradores. Como el resto de autores, hace hincapié una vez más en la importancia de las reuniones informativas previas a la interpretación. Al margen de estas recomendaciones, la situación real dista mucho de la ideal. Gile (2009) advierte que la entrega de documentos al intérprete no suele ocurrir siempre, y si se los proporcionan, suelen ser insuficientes, hecho que reflejan otros autores. Por ejemplo, la encuesta de Moser-Mercer (1992) muestra que el único documento que los organizadores del congreso suelen enviar sistemáticamente es el programa; otros, como las actas de conferencias anteriores, documentos o copias de la presentación, son poco comunes, y los glosarios multilingües y monolingües se reciben con una frecuencia extremadamente baja. Por su parte, Valentini (2002) revela en su estudio que, si bien el 86 % de los encuestados afirman recibir algo de documentación por parte de la organización del congreso, la información obtenida no es de calidad: los documentos que se reciben más a menudo son el programa (el 42 % de los encuestados dice recibirlo siempre), la intervención del ponente (el 56,8 % de los intérpretes afirma recibirlo a menudo) y su presentación (el 46,4 % dice recibirlo a menudo), mientras que se reciben “raramente” las actas de congresos anteriores, glosarios e información sobre los participantes. Por último, un estudio reciente de Corpas (2016) revela que la situación no ha cambiado con el transcurso de los años. De nuevo, un gran número de los intérpretes encuestados revelan que no se suele contar con la información necesaria para preparar con garantías el encargo, lo cual prueba, según la autora, el desconocimiento de la profesión por parte de quienes son ajenos a ella.

### **2.1.2 Fuentes de información alternativas**

Debido a esta falta generalizada de información proporcionada por el cliente, los intérpretes no tienen más remedio que recurrir a fuentes alternativas para constituir su conocimiento *ad hoc* y preparar la terminología de la reunión o congreso. Si los organizadores no entregan al intérprete un conjunto de documentos pertinentes en las dos lenguas, la AICC aconseja a los profesionales recurrir a Internet, que según ellos supone una herramienta de gran utilidad para la adquisición de conocimiento temático y terminológico<sup>6</sup>. Dadas las posibilidades infinitas de este recurso, la asociación

---

<sup>6</sup> <https://aiic.net/page/628/practical-guide-for-professional-conference-interpreters/lang/1#2> Última consulta: 13/05/2019

recomienda a los intérpretes afinar sus habilidades documentales en la web. La AIIC señala igualmente la utilidad de enciclopedias, obras de divulgación, glosarios temáticos preexistentes y otras obras de consulta básica como fuentes de partida para obtener conocimiento terminológico y temático. Existen, por otro lado, datos empíricos que revelan las fuentes de documentación predilectas de los intérpretes. Según las respuestas de profesionales del sector que obtiene Valentini (2002), los recursos empleados son variados y van desde diccionarios bilingües (el 44,8 % de los encuestados dice emplearlos “a menudo”) hasta los monolingües (empleados “a menudo” por el 44,1 % de los intérpretes encuestados) pasando por las bases de datos terminológicas en línea (el 33,1 % de los intérpretes encuestados las emplean “a menudo”). Los resultados que arrojan las encuestas de Corpas Pastor (2016) son bastante similares. Los recursos más empleados por los intérpretes de conferencias vuelven a ser los diccionarios bilingües y monolingües, además de los glosarios y tesauros. En un segundo escalafón se encuentran recursos como las búsquedas en la web, los textos paralelos, libros, revistas y publicaciones periódicas electrónicas y bases de datos. Entre los recursos menos utilizados (mencionados solo por el 20 % de los encuestados) se encuentran las bases de datos terminológicas, los sistemas TAO, los sistemas de memorias de traducción, la traducción automática, los extractores de términos, las herramientas de concordancias y los corpus. De hecho, en lo referente al tema de nuestro trabajo, tan solo 22 entrevistados de 133 afirmaron emplear corpus durante el proceso de preparación. De estos resultados se infiere que a día de hoy todavía prima una preparación terminológica “tradicional” en la que generalmente no se emplea ningún tipo de herramienta especializada y se va recopilando terminología manualmente de textos paralelos, diccionarios, bases de datos y glosarios. Valentini (2002) corrobora este planteamiento al afirmar que el tratamiento de los documentos suele ser tradicional. Los datos de la autora revelan que el 91,3 % de los intérpretes encuestados extraen la terminología de la conferencia manualmente, mientras el 11,3 % de los encuestados afirma emplear “alguna vez” programas automáticos, entre los que mencionan Termino o Wordsmith Tools.

## **2.2 Flujos de trabajo para la preparación terminológica**

Aunque hayamos presentado en el apartado anterior las fuentes y herramientas a las que recurre el intérprete para la preparación, lo cierto es que la gestión de la terminología abarca un proceso más amplio, es decir, todo un flujo de trabajo. En esta sección pretendemos realizar una revisión de lo que expone la bibliografía al respecto.

### **2.2.1 El flujo de trabajo del intérprete para Moser-Mercer**

Para Moser-Mercer (1992) la labor del intérprete comienza a partir de la aceptación de un contrato. La autora nos recuerda que, como hemos comentado, el contrato estándar de la AIIC obliga al organizador a enviar al intérprete todos los documentos pertinentes en las dos lenguas de trabajo con al menos 15 días de antelación. A partir de ese momento, el intérprete emprende un trabajo terminológico que podríamos considerar tradicional. Durante la lectura de la documentación entregada, o en su defecto, de obras de consulta relevantes, el intérprete se irá percatando de la existencia de términos y expresiones con las que no está familiarizado. El intérprete buscará los equivalentes de la terminología problemática en la lengua de destino. A continuación, realizará un glosario en papel o mediante medios informáticos y, por último, lo estudiará. Este glosario constituye para Moser-Mercer, junto con la preparación temática básica, la piedra angular de la preparación para la conferencia. Durante una sesión informativa previa el intérprete puede ampliar o mejorar su glosario, que también puede actualizar durante la conferencia al escuchar al resto de ponentes o al hablar con compañeros y expertos. Este paso equivaldría a la preparación durante la conferencia que menciona Gile (2009). Por último, tras la conferencia el intérprete debería ultimar y almacenar la lista de términos para un posible uso futuro. En resumen, la construcción *ad hoc* de conocimiento terminológico se considera “*a continuous process, with terminological equivalents refined until the very last moment before embarking on the interpretation [...]*”<sup>7</sup>. La autora describe, en resumen, un proceso de trabajo tradicional y que podría mejorarse mediante los avances tecnológicos con los que contamos en la actualidad. No cabe duda de que, por ejemplo, no resulta del todo eficiente leer documentos e ir apuntando términos manualmente si tenemos en cuenta que el intérprete suele contar con relativamente poco tiempo para prepararse.

### **2.2.2 El flujo de trabajo del intérprete para Will**

Will (2007) enfoca la adquisición de terminología por parte del intérprete como un proceso más amplio y que tiene en cuenta los sistemas de conocimiento subyacentes. Como explica el autor, se piensa que el intérprete no solo necesita conocer los términos más importantes, sino también el modo en el que todos ellos encajan en una estructura ontológica. Es decir, para Will el flujo de trabajo del intérprete es un modelo

---

<sup>7</sup> “Un proceso continuo en el que se refinan los equivalentes terminológicos hasta el último minuto antes de comenzar la interpretación” (La traducción es mía).

de gestión del conocimiento. Para describir el enfoque del intérprete, el autor divide el proceso en tres etapas (recepción, transferencia y producción) que se subdividen a su vez en diez pasos diferentes, encadenando el final de cada uno de ellos con el inicio del siguiente. En el primer paso se identifica un término en un texto relativo a la conferencia (“*texterm*”) que se registrará mediante diferentes medios. En el segundo paso, el objetivo es reconocer el concepto que subyace al término. A continuación, se vincula el “*texterm*” a un sistema de conocimientos hipotético para formar lo que el autor denomina una *Terminological Knowledge Entity* (TKE). Cada una de estas entidades se compone de un término individual encontrado en contexto (denominación + concepto), un término de referencia en fuentes autorizadas/autoritativas (denominación + concepto) con el que comparamos el término inicial, y un sistema de conocimiento interrelacionado y jerárquico. En el paso cuatro se juntan todas las TKE y se forma una TKC no estructurada (*Non-structured Terminological Knowledge Constellation*) que se analiza en base a ciertos parámetros: la cantidad (el número de sistemas de conocimiento que representan), la calidad (el grado de conocimiento externo necesario para constituir la constelación) y la valencia (el número de apariciones en el texto de un sistema de conocimiento determinado mediante estas constelaciones). Cuanto mayores sean los valores combinados de estos parámetros, más atención habrá que prestar al sistema de conocimiento en cuestión y a su terminología. En el paso 5, nos apoyamos en obras y material de referencia para establecer el sistema de conocimiento, lo cual nos conduce al paso 6 (vincular los conceptos con los términos del texto para comprender). El autor denomina estos primeros siete pasos la fase de recepción. La fase de transferencia ocupa los pasos 7, 8 y 9, en la que se constituye un sistema de conocimiento en lengua meta siguiendo los mismos principios, se comparan la estructura jerárquica de ambos sistemas y se realizan modificaciones en la lengua meta si fuera necesario por diferencias o carencias léxicas. La última fase (paso 10) es de reproducción: es decir, la interpretación en cabina propiamente dicha. Si bien es cierto que este modelo presenta una descripción pormenorizada de la adquisición de conocimientos por parte del intérprete, puede que no resulte el mejor a efectos prácticos para mejorar el rendimiento real del intérprete. No se describen en detalle los procesos y no se expone, en nuestra opinión, un flujo de trabajo que resulte beneficioso y eficaz para una adquisición de conocimiento rápida. Tampoco se menciona, por otra parte, el uso de medios tecnológicos.

### **2.2.3 La propuesta de flujo de trabajo en interpretación de Seghiri**

En el enfoque de Seghiri (2017) se aprecia un giro pragmático, pues se tienen en cuenta las necesidades y los condicionantes del intérprete. Dada la falta de glosarios disponibles en línea, que resultan de un valor inmenso para el traductor, Seghiri describe un proceso de creación de glosarios *ad hoc* durante la fase documental previa para conseguir un mejor rendimiento durante la fase de interpretación. Como la autora misma lo define, se trata de “una metodología protocolizada para la extracción terminológica basada en bitextos con objeto de generar un glosario”. La metodología, según Seghiri, resulta gratuita, rápida y aplicable a cualquier temática y par de lenguas. En primer lugar se buscan textos para constituir un corpus (la autora emplea los resúmenes de artículos de una revista del ámbito biosanitario y recomienda descargarlos por lotes con programas como HTTTRACK). A continuación, se tienen que convertir todos los textos en formato HTML y PDF a TXT, un paso que incluye la revisión de errores de formato, faltas tipográficas, etc. De nuevo, los procesos se pueden automatizar mediante programas como Scleaner o UniPdf. En un cuarto paso se renombran los textos con un código alfanumérico para marcar los idiomas y poder clasificarlos y encontrarlos fácilmente en las carpetas y subcarpetas. Por último, se alinean los textos con la herramienta gratuita LFAAligner y se exporta la alineación en formato de texto plano con codificación UTF-8 para poder extraer la terminología con Lxterm. Este programa extrae automáticamente los n-gramas en una columna y nos da la opción de extraer los equivalentes más probables en la columna siguiente. Si el programa se equivoca o no los encuentra, podemos introducirlos manualmente en esta segunda columna. Además, el programa nos permitiría consultar las concordancias de los términos. Por último, exportamos el resultado en formato TSV y pegamos las columnas en un documento Excel para finalizar nuestro glosario. En este flujo de trabajo percibimos un enfoque más pragmático y un mejor conocimiento de las necesidades del intérprete. Además, se emplean medios tecnológicos para automatizar procesos y trabajar de forma eficiente, lo cual es indispensable para el intérprete teniendo en cuenta la falta de tiempo. No obstante, nos parece discutible la afirmación de la autora de que se trate de un proceso “rápido”. Sin duda cuenta con numerosas etapas, muchas de ellas tediosas y con otros subprocesos involucrados. Además, si bien es cierto que todos los procesos se pueden automatizar, para ello el intérprete tendría que estar familiarizado con multitud de programas diferentes. Consideramos que se podría optar por un proceso más simple.

#### **2.2.4 La propuesta de flujo de trabajo en interpretación de Xu**

Por último, Xu (2018) identifica los 4 elementos clave que debería tener la preparación terminológica del intérprete: (1) la creación de una lista de términos relativos al encargo, ya sea de manera automática o manual, (2) la búsqueda de información para entender los conceptos que vehiculan los términos y para crear una estructura de conocimiento, (3) la búsqueda de equivalentes en lengua meta para las palabras de nuestra lista y (4) la consecución de una velocidad de recuperación y consulta rápida. Para la autora, estos cuatro elementos clave son el punto de partida para poder analizar y diseñar los procedimientos de preparación terminológica del intérprete. En consecuencia, para poder cumplir estos cuatro requisitos con un mismo flujo de trabajo, la autora diseña un modelo de cinco pasos en el que se utilizan corpus y herramientas de corpus. En el primer paso se elabora un corpus mediante documentos de la conferencia u otros archivos encontrados en internet. En este paso la autora menciona las herramientas para limpiar archivos del corpus (*corpus clean tools*), los tokenizadores y las herramientas de etiquetado gramatical (*POS taggers*). Tras elaborar el corpus, se realiza una extracción terminológica monolingüe que nos proporciona una lista de términos. El intérprete revisa y refina esta lista generada automáticamente durante el paso 3. En el paso 4 se comienza a emplear una herramienta de concordancias para ver los términos extraídos en su contexto y para consolidar el aprendizaje. En el quinto paso se gestiona la terminología mediante programas como Excel y otras herramientas de gestión terminológica. En el estudio, la autora no describe en detalle todos los pasos, sino que se centra en los pasos 2, 3 y 4. En el paso 3 Xu utiliza la herramienta de extracción Syllab Tools, mientras que para las concordancias el programa elegido es Sketch Engine. En este caso, encontramos un modelo que define claramente los elementos clave de la preparación (los 4 requisitos mencionados) y propone un proceso que asegura la inclusión de los mismos. No obstante, la autora solo se centra en los pasos 2-4, y no menciona el método o el programa idóneo para la constitución del corpus. Por otro lado, llama la atención que se empleen dos programas distintos para la extracción de términos y para la revisión de concordancias. A pesar de este hecho, nos encontramos ante un flujo de trabajo automatizado casi en su totalidad y que emplea las herramientas tecnológicas al alcance del intérprete para mejorar su productividad y eficiencia.

## **CAPÍTULO 3: EL USO DE CORPUS EN EL FLUJO DE TRABAJO DEL INTÉRPRETE**

Como hemos visto en el capítulo anterior, los intérpretes de conferencias siguen recurriendo en su mayoría a métodos tradicionales para constituir su conocimiento terminológico *ad hoc* antes de una conferencia. Además, la bibliografía describe unos métodos de preparación en su mayoría manuales, con ciertas excepciones (Seghiri 2017, Xu 2018) que presentan enfoques en los que se emplean herramientas automáticas y corpus. A la vista de esta preparación eminentemente tradicional, nos parece muy prometedora la línea de investigación que se centra en el uso de herramientas de corpus durante la preparación terminológica del intérprete. Si bien es cierto que una parte mayoritaria de la investigación sobre corpus se centra en su uso y compilación por parte del traductor para facilitar su labor (*cf.* Sánchez Gijón 2009, Varantola 2002, Zanettin 2002, Bernardini y Ferraresi 2013, por citar algunos), existe una tendencia reciente que incluye estas herramientas en los estudios de interpretación. Ahora bien, es preciso recalcar que la atención que se ha prestado a los corpus desde el ámbito de la interpretación ha sido bastante limitada (Xu 2018). Por ello, en este capítulo plantaremos las nuevas tendencias de investigación sobre los beneficios de los corpus en la preparación del intérprete y propondremos un programa específico (Sketch Engine) y un flujo de trabajo para gestionar esta fase previa a la conferencia. El propósito es claro: colmar la laguna bibliográfica existente y presentar a los intérpretes las nuevas herramientas a su disposición para automatizar y agilizar los procesos en vista del desequilibrio de uso y disponibilidad de herramientas tecnológicas entre traductores e intérpretes (Corpas Pastor 2016). Es necesario además “educar” a los intérpretes sobre la existencia de estos recursos, pues en muchos casos parecen desconocer su existencia (Costa *et al.* 2014, Corpas Pastor 2016).

### **3.1 Tipos de corpus que se pueden emplear**

Bernardini y Ferraresi (2013) se centran en el uso de corpus para satisfacer las necesidades terminológicas de los traductores. No obstante, y siendo conscientes de que sus necesidades no son exactamente las mismas, podemos extrapolar sus conclusiones a los estudios de interpretación. Los autores comparan los diferentes tipos de corpus para concluir cuáles podrían resultar más útiles. La primera opción sería construir un corpus manualmente, buscando textos fiables y que sean pertinentes para el encargo en cuestión.

Los autores consideran que no estamos ante la mejor opción para los traductores (y, por extensión, los intérpretes): aunque nos proporcionen un alto grado de fiabilidad, Bernardini y Ferraresi aclaran que no pueden garantizar un alto grado de certeza sobre los usos detectados en el corpus. Además, por su reducido tamaño no recogerán términos o expresiones que podrían ser de gran utilidad, afirmación con la cual concuerda Fantinuoli (2018b). Teniendo en cuenta las deficiencias mencionadas y el factor tiempo (búsqueda de textos fiables y útiles en internet, posible conversión de formato, familiarizarse con las herramientas etc.) es posible que el profesional considere que no merezca la pena optar por este enfoque. En el otro extremo del espectro contamos con los grandes corpus de referencia disponibles en línea, que podrían ser de utilidad para el intérprete al realizar su labor terminológica previa a la conferencia. Ahora bien, es posible que en estos corpus no esté presente el campo de especialidad que nos interesa o que nuestras búsquedas arrojen muy pocos resultados útiles. Por otro lado, nos podemos encontrar con el caso contrario: nuestra búsqueda generaría demasiados resultados, muchos de ellos irrelevantes y provenientes de tipologías textuales diversas. Los autores proponen una solución intermedia: el uso de corpus *ad hoc* compilados semiautomáticamente, que son la mejor solución para los profesionales en términos del ratio calidad/esfuerzo, ya que en poco tiempo es posible obtener resultados aceptables. Ahora bien, recalcan que estos corpus no tienen por qué ser un sustituto, sino un complemento. Proponen emplear BootCat, pero otros autores se centran en programas como CorpusMode (Fantinuoli 2018b). Existe además el programa de gestión de corpus Sketch Engine, mencionado por Xu (2018), quien no ahonda en sus funcionalidades ni en su empleo. Todos estos programas comparten una característica: la compilación semiautomática de corpus con documentos y páginas web de internet mediante una serie de semillas (*seeds*) o palabras clave que se introducen en la interfaz del programa (Fantinuoli, 2006). Este método de búsqueda mediante semillas puede resultar muy conveniente, ya que en muchas ocasiones las palabras clave de los resúmenes de las ponencias son lo único con lo que cuenta el traductor. En función de estas palabras y sus combinaciones (tuplas), el rastreador web (*web crawler*) indexará páginas web (y sus respectivas URL) en las que aparezcan las palabras seleccionadas y compilará un corpus. Más tarde, estos programas usarán su propio algoritmo interno para extraer los n-gramas y los términos más representativos mediante la comparación con un corpus de referencia.

Estamos ante un método muy útil para generar un corpus “*quick-and-dirty*”<sup>8</sup> (Bernardini y Ferraresi, 2013) que resulte útil para el intérprete si no contamos con documentación o información por parte del cliente o si las limitaciones temporales nos impiden realizar búsquedas más metódicas o la lectura atenta de documentos, lo cual suele ser común.

### **3.2 Beneficios del uso de corpus durante la preparación terminológica**

La ventaja más evidente de emplear estos corpus semiautomáticos durante el proceso de preparación terminológica es el ahorro de tiempo. Debido a limitaciones temporales, en muchas ocasiones los intérpretes no tienen tiempo de compilar corpus especializados, analizar los datos y gestionar glosarios (Gorjanc 2009). Además, como expone Fantinuoli (2018b), la consulta de información en internet no deja de ser un proceso lento en el que se manejan múltiples páginas cuya fiabilidad tenemos que evaluar. Con este método podemos “simular” la consulta de documentos que realizaría el intérprete en línea, solo que en mucho menos tiempo, ya que el *webcrawler* encontraría automáticamente las páginas web más apropiadas. Por otro lado, con el enfoque tradicional el intérprete consultaría los documentos obtenidos de manera lineal para obtener la terminología, mientras que con el uso de corpus podemos realizar una consulta no lineal, incluyendo el uso de expresiones regulares y comodines para detectar con más facilidad patrones (Bernardini y Ferrarese, 2013) y agilizar de esa manera la labor del intérprete. En conclusión, el uso de corpus nos permitiría conseguir de manera rápida una gran cantidad de información que de otro modo tendríamos que haber extraído manualmente (Pérez Pérez 2018). Ya no tenemos que leer y analizar cada documento o página individualmente: con el programa de corpus se añadirán automáticamente y se nos proporcionará el vocabulario principal.

Otra ventaja clara de este enfoque se relaciona con la falta de recursos para el intérprete. Tengamos en cuenta que en muchos casos el profesional tendrá que enfrentarse a temas punteros o muy novedosos en los que exista dispersión léxica o una terminología todavía difusa. En estos casos es posible que no abunden los trabajos terminológicos y glosarios en línea, por lo que el intérprete no tendrá más remedio que informarse por otros medios (en especial mediante los textos de los expertos). De hecho, suele ser común que los intérpretes no puedan recurrir a los recursos terminológicos ya establecidos, dado que su situación de disponibilidad en todos los ámbitos temáticos suele ser limitada a nivel

---

<sup>8</sup> “Un corpus rápido y sin refinar” (La traducción es mía).

europeo (Gorjanc 2009). Por otro lado, existen bancos de datos terminológicos como Termium o Unterm, pero no resultan lo suficientemente específicos para la preparación de una conferencia (Xu 2018). Ante la falta de otro tipo de recursos, constituir un corpus semiautomático *ad hoc* con textos y páginas web nos podría proporcionar de manera rápida orientaciones terminológicas muy valiosas, pues nos permitiría analizar de manera ágil qué términos emplean los expertos para poder fijar nosotros mismos soluciones concretas ante los nuevos problemas terminológicos.

Por otro lado, el empleo de corpus podría optimizar la adquisición de conocimientos por parte del intérprete, ya que permitiría realizar un proceso “*bottom-up*” (Fantinuoli 2006) en el que, empezando por los términos, se iría ascendiendo progresivamente hasta llegar a la estructura conceptual del ámbito temático. En este sentido, consideramos que el empleo de herramientas de gestión de corpus podría ser un enfoque muy útil para conseguir de una forma práctica la estructuración del sistema de conocimiento por la que aboga Will (2009) en la preparación de los intérpretes. El intérprete partiría de los términos extraídos automáticamente por programas como BootCat o Sketch Engine (información intralingüística). A continuación, la exploración progresiva del corpus mediante las concordancias nos permitiría adquirir conocimientos específicos (términos y expresiones) en un contexto semántico real (Pérez Pérez 2018, Fantinuoli 2018b), lo cual ayudaría a consolidar los conocimientos. De hecho, autores como Pérez Pérez (2018) hacen hincapié en la importancia de que el intérprete aprenda estos términos en un contexto real (concordancias), pues en la conferencia no se emplearán de manera aislada. El uso de las concordancias, en conclusión, nos permitiría adquirir al mismo tiempo conocimientos intralingüísticos y extralingüísticos mediante una sola plataforma gracias a la posibilidad de ver los términos y la fraseología en su contexto.

Al margen de estos beneficios hipotéticos, contamos con estudios empíricos (si bien escasos) que han conseguido probar una mejora del rendimiento del intérprete gracias al uso de estas herramientas. En primer lugar, Xu (2018) comprueba que el uso de una herramienta para navegar por las concordancias (Sketch Engine) junto con el empleo de un extractor terminológico (SyLabTools) entraña una mejora del rendimiento. Un grupo de control emprende una preparación tradicional, mientras que el grupo de estudio emplea las herramientas mencionadas. Como resultado, la autora comprueba que el grupo de estudio muestra una mayor precisión terminológica que el grupo de control.

El grupo de estudio presentó además menos errores graves en términos especializados, nombres propios de organismos y términos generales. Por otro lado, el grupo de control evidenció bastantes más errores en párrafos con alta densidad terminológica. Además, el tiempo de preparación fue menor para el grupo de estudio y dos meses después quienes usaron el programa se acordaban de más términos (*“higher recall rate”*) que el grupo de control (lo cual podría probar que el uso de concordancias ayuda a consolidar y fijar la terminología). Una última faceta positiva fue que los participantes en el estudio (estudiantes de Máster) tuvieron una opinión positiva de Sketch Engine, pues consideraron que les permitió aprender patrones y colocaciones en ambas lenguas, aprender términos relevantes y activar los términos mediante la traducción a la vista.

Por otro lado, Pérez Pérez (2018) realiza un estudio que prueba empíricamente la mejora del rendimiento gracias al uso de sistemas de gestión de corpus (en su caso, AntConc). Para ello, el autor manda interpretar sobre tres ámbitos temáticos distintos a un grupo de control y un grupo de estudio. Los temas se anunciaron con siete días de antelación, durante los cuales un grupo se preparó de manera tradicional y el otro mediante el empleo de AntConc. El autor analiza las transcripciones de los alumnos y calcula su tasa de acierto comparando su prestación con el listado de términos especializados que aparecieron en cada uno de los tres discursos. En el primer tema, la tasa de acierto fue de un 34 % de media para el grupo de control y de un 39 % para el grupo de estudio. En el resto de discursos el grupo de estudio obtuvo también mejores resultados terminológicos (26 % vs. 34 % y 26 % vs. 38 %), lo cual podría probar que el uso de estas herramientas ayuda a consolidar el conocimiento terminológica y mejora la precisión.

### **3.3 El uso de Sketch Engine para la labor terminológica previa del intérprete**

A la luz de lo expuesto, quedan claros los beneficios de emplear herramientas de corpus durante la preparación previa al encargo. Ahora bien, hemos mencionado que el uso de estos recursos es moderado y que muchos profesionales incluso desconocen su existencia. Para remediar esta situación y la falta de herramientas tecnológicas que faciliten la labor del intérprete (Corpas Pastor 2016) presentaremos en este apartado la herramienta Sketch Engine y expondremos el flujo de trabajo que seguiría el profesional. En este sentido, hemos revisado en el segundo capítulo los diferentes flujos de trabajo que describen los estudiosos para la labor terminológica del intérprete y hemos puesto de manifiesto ciertas carencias. En consecuencia, en lugar de centrarnos en alguno de ellos

en concreto, propondremos uno nuevo que se base en la herramienta Sketch Engine y que tenga en cuenta los cuatro pilares sobre los que se debe basar el trabajo terminológico del intérprete (Xu 2018): (1) establecer un listado de términos, (2) buscar información para comprender los términos y conceptos y llegar a un sistema de conocimientos organizado, (3) encontrar los equivalentes terminológicos en la lengua de destino y (4) asegurar la rapidez de acceso y de consulta del glosario. En los siguientes apartados explicaremos cómo completar estas fases mediante la herramienta de gestión de corpus Sketch Engine. Consideramos importante esta labor expositiva o didáctica, puesto que si el intérprete percibe que el uso de estas herramientas cuesta demasiado trabajo, podría verse disuadido de usarlas (Fantinuoli 2018b). En la misma línea, Bernardini y Ferraresi (2013) también recalcan la importancia de la percepción de estas herramientas por parte de los profesionales.

También es preciso comentar porqué se ha elegido la herramienta mencionada. Algunos autores (Gorjac 2009, Fantinuoli 2006) proponen emplear herramienta BootCat, que cuenta con menos funcionalidades que Sketch Engine y no permite consultar las concordancias. Otros (Pérez Pérez 2018) emplean AntConc, que no cuenta con la opción de creación semiautomática de corpus mediante semillas y que presenta una interfaz menos intuitiva. Otro gran punto a favor de Sketch Engine es que nos permitiría tener acceso a los tres tipos de corpus mencionados por Bernardini y Ferraresi (2013) que pueden complementarse entre sí: los grandes corpus de referencia<sup>9</sup>, los corpus *ad hoc* creados manualmente con nuestros propios archivos y los corpus de tamaño medio creados semiautomáticamente mediante semillas (*seeds*). Por otro lado, hemos elegido esta herramienta para agilizar el flujo de trabajo del intérprete ya que consideramos que aún en una sola interfaz todo lo que necesita para realizar su labor terminológica y realizar un glosario. Además, el estudio de Xu (2018) muestra que los estudiantes que emplearon Sketch Engine para su preparación tuvieron una visión positiva de la herramienta, lo cual podría indicar que la experiencia para un usuario novel sería positiva.

### **3.3.1. Pasos preliminares**

Antes de iniciar el primero de los cuatro pasos de cualquier flujo de trabajo, que consiste la creación de una lista de términos (Xu, 2018), es preciso decidir qué tipo de

---

<sup>9</sup> En Sketch Engine contamos, por ejemplo, con los corpus EnTenTen y EnTenTen, así como corpus de referencia en muchas otras lenguas de trabajo.

corpus emplear<sup>10</sup>: corpus de referencia, corpus compilado mediante semillas o corpus *ad hoc* compilado manualmente (Bernardini y Ferraresi 2013). Debemos tener en cuenta sus posibles desventajas (mencionadas anteriormente) y otros factores como la temática del encargo. Si el intérprete cuenta con poco tiempo o debe afrontar un tema novedoso o sobre el que apenas existan recursos en línea, es recomendable que compile semiautomáticamente un corpus mediante semillas. Por otro lado, si el intérprete cuenta con actas de congresos pasados y documentos temáticos y el encargo es similar a otro anterior o se trata de una nueva edición de un congreso pasado, es posible que el intérprete quiera compilar un corpus manualmente con sus archivos. Por otra parte, los corpus de referencia como EnTenTen no resultarían en general tan útiles para el profesional (Bernardini y Ferraresi 2013), pero en algunos casos podrían ser de ayuda. En el menú principal (*dashboard*) podemos elegir una lengua para emplear su corpus de referencia (en inglés tendremos acceso por defecto a EnTenTen2013).

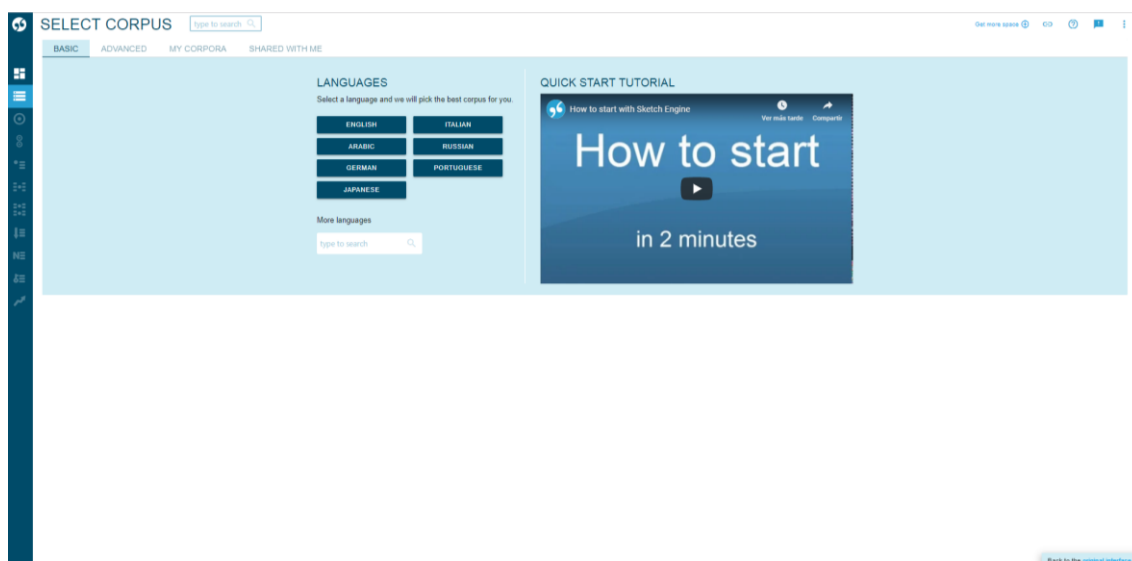


Figura 4. Pestaña “basic” del menú principal (“dashboard”).

En la pestaña “*advanced*” podemos encontrar aún más corpus que podrían resultar útiles en función de la temática del encargo. El programa incluye, entre otros, el corpus del Parlamento Europeo, el corpus de la Dirección General de Traducción de la Unión Europea y el corpus EUR-Lex, que contiene el derecho de la UE). El corpus EUR-Lex, por ejemplo, resultaría útil si nos encargasen interpretar en el Tribunal de Justicia de la

<sup>10</sup> Para hacerlo el intérprete deberá dirigirse a la pestaña *dashboard* mediante el menú desplegable de la izquierda de la pantalla.

Unión Europea. Por otro lado, existe un gran número de corpus en diversas lenguas de trabajo, incluidas las consideradas más minoritarias. Si cualquiera de estos corpus de referencia no satisficiese las necesidades del intérprete<sup>11</sup> debido a la especificidad de la temática del encargo, el profesional contaría con la opción de crear un corpus *ad hoc* pulsando el botón “*new corpus*” en la pestaña “*my corpora*”.

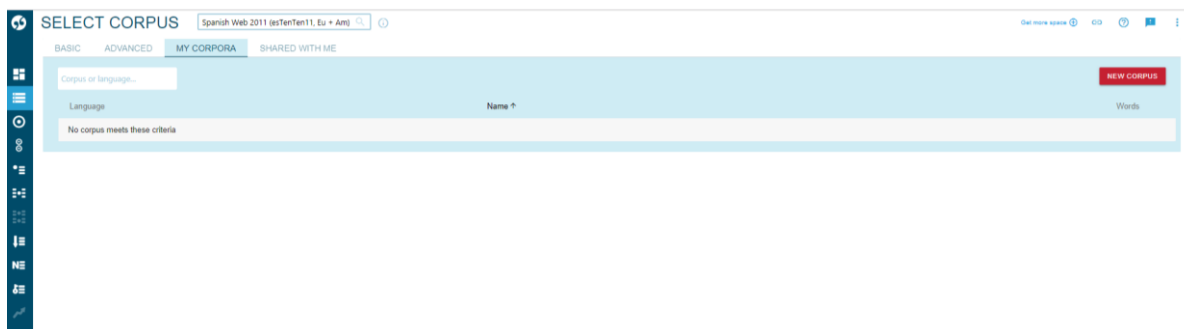


Figura 5. Creación de un corpus *ad hoc* en la pestaña “*My corpora*”

En la siguiente pantalla (figura 6) se debe elegir un nombre para el corpus y un idioma (también se puede realizar un corpus multilingüe seleccionando la opción “*multilingual*”, proceso que ilustraremos más adelante).

Build your own private corpus from texts on the web or from your own documents.

Name

Language

Description

Storage used: 972,608 of 1,000,000 words (97%)

Available features

Figura 6. Menú de creación del corpus

<sup>11</sup> Tengamos en mente las desventajas de los corpus de referencia según Bernardini y Ferraresi (2013). Puede darse una ausencia de resultados pertinentes para nuestro campo de especialidad o el caso contrario: un exceso de resultados (muchos de ellos irrelevantes) pertenecientes a muchas tipologías textuales diferentes.

Tras pulsar *next*, aparecerá una pantalla (figura 7) en la se presentará al intérprete dos opciones: crear un corpus *ad hoc* con sus propios documentos<sup>12</sup> o crear un corpus mediante semillas a partir de documentos y páginas web procedentes de internet. Para temas muy concretos o especializados, ambas opciones proporcionarán resultados mejores que los grandes corpus de referencia (Bernardini y Ferraresi, 2013).

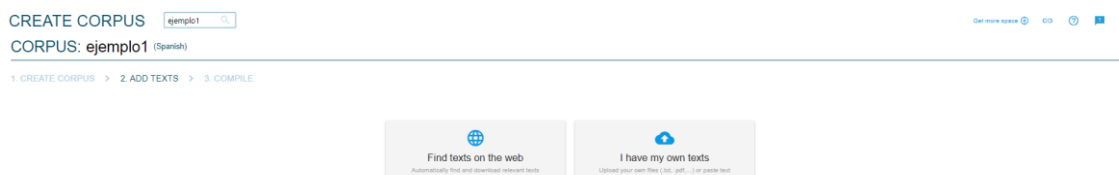


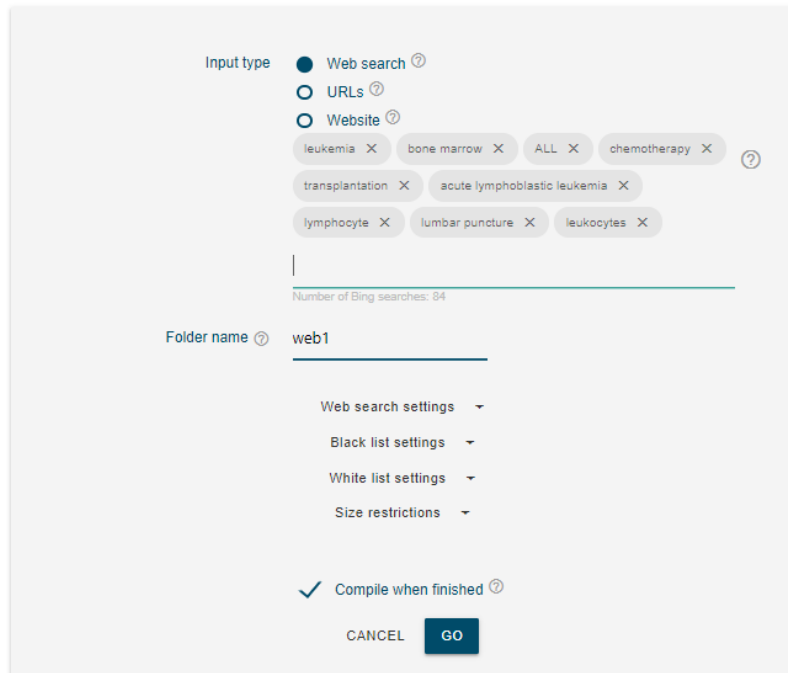
Figura 7. Creación de corpus mediante la web (izquierda) y mediante nuestros propios archivos (derecha)

Ahora bien, la opción más conveniente para el intérprete será normalmente la compilación automática de un corpus mediante semillas debido a su comodidad y su rapidez, la falta de documentos para compilar un corpus manual y la escasez de recursos en línea (glosarios, bases de datos terminológicas etc.). Hemos utilizado a modo de ejemplo las semillas que introduce Fantinuoli (2006) en Bootcat para compilar un corpus de una posible conferencia sobre leucemia. Según el autor, estas semillas podrían obtenerse, por ejemplo, de los resúmenes de la conferencia proporcionados al intérprete. Tenemos que tener en cuenta que si queremos introducir en Sketch Engine semillas de más de una palabra debemos hacerlo entre comillas (figura 8). Además, cuantas más palabras clave se introduzcan, más tamaño tendrá el corpus final. Como referencia, con las 9 palabras clave de Fantinuoli (2006) hemos compilado un corpus de 972.000 palabras.

---

<sup>12</sup> El programa admite un gran número de tipos de archivo como TXT, DOCX, XLS, CSV, TSV o incluso archivos bilingües como TMX.

## ← TEXTS FROM WEB



The screenshot shows a web-based interface for generating a corpus from web searches. It features the following elements:

- Input type:** Three radio buttons are present: 'Web search' (selected), 'URLs', and 'Website'. Each has a help icon.
- Search terms:** A horizontal list of tags with 'X' icons for removal: 'leukemia', 'bone marrow', 'ALL', 'chemotherapy', 'transplantation', 'acute lymphoblastic leukemia', 'lymphocyte', 'lumbar puncture', and 'leukocytes'. A help icon is on the right.
- Search count:** A text input field is empty, with the text 'Number of Bing searches: 84' below it.
- Folder name:** A text input field containing 'web1'.
- Settings:** Four dropdown menus: 'Web search settings', 'Black list settings', 'White list settings', and 'Size restrictions'.
- Options:** A checked checkbox labeled 'Compile when finished' with a help icon.
- Buttons:** 'CANCEL' and 'GO' buttons at the bottom.

Figura 8. Semillas de Fantinuoli (2006) para generar un corpus sobre la leucemia

Si bien es cierto que podrían caber dudas sobre la fiabilidad de las páginas y documentos extraídos automáticamente por la herramienta, existen estudios (Fantinuoli 2018b con CorpusMode, Bernardini y Ferraresi 2013 con BootCat) que indican que la calidad de los textos (evaluada por humanos) presenta un nivel adecuado. En el estudio de Bernardini y Ferraresi, por ejemplo, el nivel de calidad de los textos extraídos y su relación con el tema variaba entre en 44 y el 70 % en función del corpus. Aunque no se hayan encontrado estudios sobre Sketch Engine, los precedentes en otras herramientas son prometedores. Sí que es cierto que el corpus podría estar sesgado (demasiado peso de algunas fuentes en detrimento de otras, Fantinuoli 2006), pero es un problema que se podría solucionar manualmente eliminando las URL de documentos que contengan demasiadas palabras en comparación con la media. Como se aprecia en la figura 9, se podrían analizar, seleccionar y borrar las fuentes que contengan demasiadas palabras o que no resulten fiables para que el corpus creado sea representativo.

<input type="checkbox"/>	Document	Type	Words
<input type="checkbox"/>	bone-marrow-anatomy-373236	html	~1,102 ...
<input type="checkbox"/>	autologous-transplant	html	~1,232 ...
<input type="checkbox"/>	index.html	html	~477 ...
<input type="checkbox"/>	285666.php	html	~6,154 ...
<input type="checkbox"/>	Basics	html	~1,265 ...
<input type="checkbox"/>	blood-cell-disorders	html	~1,610 ...
<input type="checkbox"/>	erythrocytes-leukocytes-thrombocytes.html	html	~1,752 ...
<input type="checkbox"/>	blood-cell-formation	html	~680 ...
<input type="checkbox"/>	bone-marrow-transplantations-for-leukemia	html	~615 ...
<input type="checkbox"/>	cml-treatment-pdq	html	~2,674 ...
<input type="checkbox"/>	adult-aml-treatment-pdq	html	~3,266 ...
<input type="checkbox"/>	leukemia-bone-marrow-transplant	html	~352 ...
<input type="checkbox"/>	High-dose_chemotherapy_and_bone_marrow_transplant	html	~4,338 ...
<input type="checkbox"/>	bone-marrow-stem-cell-transplant.html	html	~1,155 ...
<input type="checkbox"/>	s41375-019-0494-9	html	~1,987 ...
<input type="checkbox"/>	bone-marrow-suppression-during-chemotherapy-2249318	html	~1,330 ...
<input type="checkbox"/>	152504	html	~382 ...
<input type="checkbox"/>	stem-cell-transplantation-with-high-dose-chemotherapy	html	~371 ...
<input type="checkbox"/>	what-s-bone-marrow-transplant.asp	html	~2,195 ...
<input type="checkbox"/>	s41409-018-0184-7	html	~2,051 ...

BULK ACTIONS ▾ Rows per page: 20 ▾ 181-200 of 428 |< < 10 > >|

Figura 9. Gestión de las fuentes del corpus (“My corpora”>“manage corpora”>“browse”)

### **3.3.2 Primera fase: creación de la lista de términos**

Una vez seleccionado el corpus (ya sea uno de referencia, uno creado manualmente o uno compilado a partir de la web) es posible explorarlo gracias a las funcionalidades que ofrece el programa<sup>13</sup>. Siguiendo a Xu (2018) el primer paso que debemos realizar es la elaboración de una lista de términos. En este sentido, Sketch Engine resultaría de gran ayuda para el intérprete, pues la herramienta elabora estas listas de manera automática. En lugar de revisar manualmente los documentos relevantes (actas de congresos, artículos científicos, páginas web, etc.) en busca de candidatos a término, podemos hacer que Sketch Engine extraiga automáticamente los términos más relevantes usando un algoritmo en el que compara su frecuencia de aparición en el corpus compilado con su frecuencia en un corpus de referencia<sup>14</sup> (Kilgarriff *et al.* 2014a). De este modo, detectará los términos más importantes del ámbito en cuestión. Para la extracción, solo tenemos que dirigirnos a la pestaña “keywords”. En la pestaña “basic” solo tenemos que pulsar el botón “go” para que comience la extracción. En la pestaña “advanced” (figura 10) el intérprete podrá cambiar ciertos parámetros de la extracción, como el corpus de

<sup>13</sup> Para explicar estas funcionalidades emplearemos el corpus creado con las semillas de Fantinuoli. Es decir, el supuesto sería prepararse para una conferencia especializada sobre leucemia.

<sup>14</sup> Por defecto, EnTenTen13 para inglés y EsTenTen11 para español.

referencia, la frecuencia mínima de aparición (por defecto 1) o el número máximo de palabras clave a extraer (por defecto 1000). En estas opciones sería interesante para el intérprete aumentar la frecuencia mínima para que se extraigan términos que aparezcan con cierta frecuencia y no solo una vez en todo el corpus. Además, podría ser conveniente para el intérprete reducir el número máximo de términos extraídos, ya que después deberá revisar y limpiar la lista.

The image shows a web interface for advanced search settings. At the top, there are three tabs: 'BASIC', 'ADVANCED' (which is selected), and 'ABOUT'. Below the tabs, there are several settings sections. The first section is 'Subcorpus', with a dropdown menu set to 'none (the whole corpus)'. To its right is a 'Focus on' slider ranging from 'rare' to 'common', with a value of '1' in the middle. Below this are two checkboxes: 'At least one alphanumeric' and 'Only alphanumeric', both of which are checked. The 'Minimum frequency' is set to '5'. Below these are two columns of settings. The left column is titled 'Keywords (single-words) settings' and includes fields for 'Reference corpus' (set to 'English Web 2013 (enTenTen13)'), 'Reference subcorpus' (set to 'none (the whole corpus)'), 'Maximum items' (set to '800'), 'Attribute for keywords' (set to 'lemma'), and 'Matching regex' (set to '.\*'). The right column is titled 'Terms (multi-words) settings' and includes fields for 'Reference corpus' (set to 'English Web 2013 sample'), 'Reference subcorpus' (set to 'none (the whole corpus)'), 'Maximum items' (set to '800'), and 'Matching regex' (set to '.\*'). At the bottom center of the interface is a red button labeled 'GO'.

Figura 30. Opciones avanzadas para la extracción de términos

Tras aceptar, los resultados se mostrarán en dos columnas (figura 11): a la izquierda, las extracciones formadas por una sola palabra (*single-words*), y a la derecha, extracciones formadas por dos o más palabras (*multi-words*). Los términos más útiles para el intérprete se encontrarán en la columna de la derecha, y por ello deberá dedicar más tiempo a revisar y depurar dicha lista. Ahora bien, es conveniente que no deje de lado la lista de unigramas, pues también podría proporcionarle términos interesantes. En nuestro caso, hemos encontrado, entre otros, siglas como CSF (*cerebrospinal fluid*), GVHD (*graft-versus-host disease*) o CLL (*chronic lymphocytic leukemia*).

Word	Focus corpus ?	Reference corpus ?
1 leukemia	5,374	53,223 ...
2 allogeneic	1,594	3,195 ...
3 lymphoblastic	1,443	2,397 ...
4 marrow	4,463	70,650 ...
5 lymphocyte	2,034	20,788 ...
6 transplantation	2,896	44,169 ...
7 CSF	1,125	15,055 ...
8 GVHD	738	2,386 ...
9 CLL	794	4,302 ...
10 hematopoietic	810	6,649 ...
11 leukocyte	839	7,973 ...
12 lymphocytic	659	3,853 ...
13 HSCT	574	1,132 ...
14 Marrow	676	6,292 ...
15 myeloid	653	5,422 ...

Word	Focus corpus ?	Reference corpus ?
1 lymphoblastic leukemia	1,211	22 ...
2 bone marrow	3,480	512 ...
3 acute lymphoblastic leukemia	1,155	22 ...
4 lumbar puncture	891	10 ...
5 cell transplantation	921	23 ...
6 stem cell transplantation	669	16 ...
7 marrow transplantation	528	12 ...
8 lymphocytic leukemia	506	18 ...
9 stem cell	1,369	475 ...
10 white blood	944	296 ...
11 bone marrow transplantation	411	12 ...
12 hematopoietic stem	339	11 ...
13 cell transplant	378	41 ...
14 peripheral blood	381	46 ...
15 blood cell	540	180 ...

Figura 11. Muestra de los términos extraídos

Como se aprecia en la figura anterior, el programa compara la frecuencia de aparición de los términos en el corpus que hemos compilado (*focus corpus*) con el corpus de referencia (*reference corpus*) para determinar cuáles son los términos más relevantes (cuanto más aparezca en el *focus corpus* y menos aparezca en el *reference corpus*, más específico será del ámbito en cuestión y, por lo tanto, más le interesará al intérprete). Es decir, *[c]andidate terms are [...] presented to the user in a sorted list, with the best candidates – those with the highest domain:reference ratio – at the top<sup>15</sup>*” (Kilgarriff *et al.* 2014a). Con este algoritmo se asegura que los términos extraídos sean relevantes para preparar el encargo. Además, como se puede apreciar en las figuras 11 y 12, la extracción parece ser de buena calidad; no se observan palabras del lenguaje común ni términos incompletos, mal formado o mal extraídos entre las primeras posiciones. Por ello, esta lista podría ser de gran utilidad al intérprete como punto de partida para realizar el resto de fases de la labor terminológica previa al encargo, sobre todo si tenemos en cuenta la buena calidad de la extracción. A este respecto, Kilgarriff *et al.* (2014a) realizan un estudio sobre la precisión de la extracción de términos con Sketch Engine comparando los términos extraídos manualmente del corpus GENIA con los términos extraídos por la herramienta. Los autores observaron que Sketch Engine extrajo 61 de los 100 términos

<sup>15</sup> “[s]e proporciona al usuario una lista con los candidatos a término, con los mejores candidatos (es decir, los que presentan un mayor ratio corpus de trabajo:corpus de referencia) en las primeras posiciones”. (La traducción es mía).

principales del corpus que habían sido identificados previamente por humanos. La mayoría de palabras no identificadas eran acrónimos. Concluyen además que aunque la extracción puede ser fructífera en corpus pequeños, los resultados mejoran con varias decenas de millones de palabras. Podemos inferir de estos datos que la calidad de la extracción es buena, lo cual concordaría con investigaciones sobre otras herramientas<sup>16</sup>.

MULTI-WORDS ●

Word	Focus corpus ?	Reference corpus ?
21 remission	1,371	46,091 ...
22 AML	695	13,594 ...
23 intrathecal	467	2,353 ...
24 chemotherapy	2,989	154,017 ...
25 Leukemia	563	10,812 ...
26 lumbar	1,095	44,022 ...
27 Oncol	357	578 ...
28 puncture	1,218	58,592 ...
29 Transplant	484	9,826 ...
30 CNS	789	30,421 ...
31 leukaemia	437	8,120 ...
32 Clin	387	5,148 ...
33 MRD	340	1,903 ...
34 cerebrospinal	401	7,495 ...
35 platelet	842	41,567 ...

Word	Focus corpus ?	Reference corpus ?
21 allogeneic stem cell	281	1 ...
22 marrow transplant	358	73 ...
23 cerebrospinal fluid	312	34 ...
24 acute leukemia	275	7 ...
25 myelogenous leukemia	267	7 ...
26 bone marrow transplant	327	71 ...
27 hematopoietic stem cell	247	4 ...
28 allogeneic stem cell transplantation	220	1 ...
29 graft-versus-host disease	221	3 ...
30 allogeneic bone	214	2 ...
31 donor lymphocyte	210	0 ...
32 allogeneic bone marrow	209	2 ...
33 white blood cell	271	81 ...
34 acute myeloid leukemia	213	15 ...
35 residual disease	206	7 ...

Figura 12. Muestra de algunas de los términos extraídos considerados más relevantes por el algoritmo

### 3.3.2.1 Extracción de la fraseología

Si bien Xu solo menciona la identificación de términos en su primera etapa de preparación, consideramos que sería imprescindible incluir la fraseología en esta primera fase. De hecho, autores como Fantinuoli (2017) recalcan la importancia de la fraseología en interpretación tanto en la fase de comprensión como en la de producción. Tampoco olvidemos que las expectativas de los usuarios de interpretación son altas y que pretenden que empleemos su misma terminología y fraseología. Afortunadamente, las *words sketches*, una funcionalidad característica de Sketch Engine y de la que carecen otros programas, nos permitirían extraer esta fraseología. Kilgariff *et al.* (2014b), definen esta funcionalidad como “*a one-page summary of a word’s grammatical and collocational*

<sup>16</sup> Fantinuoli (2006) compara los términos extraídos automáticamente por BootCat con los términos extraídos manualmente por un profesional. Concluye que la calidad de la extracción es comparable, aunque el sistema no tuvo en cuenta bastantes términos que sí que detectó el terminólogo.

*behaviour*<sup>17</sup>”. Según los autores, el sistema se encarga de encontrar y organizar todos los patrones asociados a una palabra. Al intérprete le resultaría extremadamente útil consultar los patrones gramaticales y de colocación de los términos más relevantes para el encargo o de los que considere que puedan tener mayor importancia. En la siguiente figura, por ejemplo se presentan todos los patrones asociados a “*bone marrow*” (médula ósea).

modifiers of "bone marrow"	nouns modified by "bone marrow"	verbs with "bone marrow" as object	verbs with "bone marrow" as subject	"bone marrow" and/or ...	prepositional phrases	adjective predicates of "bone marrow"	"bone marrow" is a ...	possessors of "bone marrow"	pronominal possessors of "bone marrow"
healthy ...	blood ...	replace ...	produce ...	blood ...	"marrow" "bone"	unable ...	tissue ...	recipient ...	your ...
liquid ...	infiltrate ...	donate ...	make ...	tissue ...	... in "bone"	due ...	blast ...	donor ...	her ...
red ...	smear ...	view ...	stem ...	system ...	marrow"		material ...	person ...	their ...
human ...	chimerism ...	destroy ...	contain ...	node ...	... of "bone"		site ...	patient ...	
diseased ...	node ...	harvest ...	recover ...	liver ...	marrow"		part ...		
yellow ...	cell ...	restore ...	be ...	thymus ...	... from ...				
autologous ...	system ...	damage ...	replace ...	spleen ...	"bone marrow"				
normal ...		remove ...	stimulate ...	sample ...	... to "bone"				
new ...		collect ...	do ...	cell ...	marrow"				
own ...		function ...	have ...	transplantation ...	marrow"				
liver ...		affect ...	show ...	engraftment ...	"bone marrow"				
homologous ...		stimulate ...		organ ...	from ...				
damaged ...		take ...		transplant ...	... by "bone"				
blood ...		purge ...		production ...	marrow"				
unhealthy ...		leave ...		lymphocyte ...	"bone marrow"				
Red ...		use ...		leukemia ...	marrow" of ...				
donor ...		help ...			"bone marrow" in ...				
allogenic ...		obtain ...			... with ...				
unrelated ...		include ...			"bone marrow"				
		receive ...			"bone marrow" for ...				
		make ...							
		find ...							
		do ...							
		have ...							
		be ...							

Figura 13. Patrones gramaticales y de colocación del término “*bone marrow*”

Resultan de extrema utilidad, por ejemplo, todos los adjetivos asociados a “*bone marrow*”, pues nos revelan nuevos términos como *diseased bone marrow*, *autologous bone marrow*, *homologous bone marrow* o *unrelated bone marrow*. También son importantes los verbos que forman patrones con el término, como *replace the bone marrow*, *donate bone marrow*, *harvest bone marrow*, *restore bone marrow*, *collect bone marrow*, *purge bone marrow*, *stimulate bone marrow* o *produce bone marrow*. Resultan asimismo útiles otras columnas como “*possessors of bone marrow*”, en la que se nos indica que la médula ósea puede tener poseedores como *person*, *patient*, *donor* o *recipient*. Es muy conveniente para el intérprete contar en una sola página con los principales patrones que forman los términos más importantes. Estos patrones adquieren una importancia aún mayor si el intérprete trabaja hacia su lengua B (lengua extranjera), ya que aunque pueda estar familiarizado con los términos, es menos probable que conozca los verbos o adjetivos que formen colocaciones con ellos. Y como interpretar no se basa en repetir mecánicamente términos, sino en expresar ideas, el intérprete deberá conocer a

<sup>17</sup> “[U]n resumen en una página de los patrones gramaticales y de colocación de una palabra” (La traducción es mía).

fondo estas colocaciones. En este sentido, las *word sketches* del programa le proporcionarían una solución muy rápida para colmar sus lagunas lingüísticas en lengua extranjera<sup>18</sup>. En resumen, con este método el intérprete podrá aprender de manera ágil toda la constelación fraseológica y gramatical en torno a los términos principales y a las palabras clave de la conferencia. Por otro lado, es evidente que aunque la extracción sea extremadamente útil dista de ser perfecta, por lo que es importante que el intérprete revise las concordancias haciendo clic en el icono a la derecha de cada colocación para asegurarse de su fiabilidad.

### **3.3.3 Segunda fase: consulta de información**

Una vez extraídos los términos principales de manera automática, el intérprete se adentraría en el siguiente paso del flujo de trabajo terminológico: la consulta de información para crear un sistema de conocimiento organizado de términos y conceptos interrelacionados (Xu 2018). Para ello, Sketch Engine nos permite realizar búsquedas de concordancia. Como hemos expuesto, combinar la extracción terminológica con la consulta de concordancias podría resultar de gran utilidad al intérprete para adquirir al mismo tiempo conocimiento intralingüístico y extralingüístico y para aprender los términos en su contexto real. Este último punto es indispensable, pues en una situación real los términos se emplearán en un contexto concreto y no aislados (como los vemos en la lista de términos). Asimismo, las concordancias permitirán al intérprete entender el significado y el uso del término y empezar a tener una visión general del ámbito (Fantinuoli 2017). El intérprete puede acceder directamente a las concordancias desde la lista de términos extraídos: tan solo debe hacer clic en la casilla con tres puntos a la derecha del término y seleccionar “*Concordance (focus corpus)*”. Pongamos que el intérprete quisiera saber más sobre el término “*graft-versus-host disease*” (enfermedad de injerto contra huésped). En las concordancias podría ver su uso en contexto e ir detectando ciertos patrones (figura 14). Además, podría encontrar concordancias provenientes de textos más divulgativos que esclarecieran su significado (ver figura 15, en la que se proporciona una definición).

---

<sup>18</sup> Evidentemente, también le ayudaría a aprender colocaciones en su lengua nativa. La diferencia es que en lengua A el intérprete cuenta con bastantes más recursos lingüísticos para salir del paso en comparación con su lengua B. Por ello, interiorizar las colocaciones es aún más importante si trabajamos hacia una lengua extranjera.

Source	Snippet
bloodjournal.org	and cytogenetic remission, which persisted 32 to 91 weeks after treatment. <b>graft-versus-host disease</b> developed and was treated by immunosuppression. These results are an example of adoptive immunotherapy without
bloodjournal.org	? The combination of cyclosporine A (CSA) and methotrexate was administered for prophylaxis of <b>graft-versus-host disease</b> (GVHD). Chimerism was studied by the use of restriction fragment length polymorphism (RFLP) and polymorphic
bloodjournal.org	Walden PL, Floumoy N, Thomas ED, Francis R, Feller A, Buckner CD, Storb R. Antileukemic effect of <b>graft-versus-host disease</b> in human recipients of allogeneic marrow grafts. N Engl J Med 300:1068, 1979. 18. Walden PL, Sullivan KM, Floumoy N,
bloodjournal.org	N, Storb R, Thomas ED, and the Seattle Marrow Transplant Team. Antileukemic effect of chronic <b>graft-versus-host disease</b> . Contribution to improved survival after allogeneic marrow transplantation. N Engl J Med 304:1529, 1981. 19. Sullivan
bloodjournal.org	P, Neiman P, Petersen FB, Sanders J, Singer J, Stewart P, Thomas ED. Influence of acute and chronic <b>graft-versus-host disease</b> on relapse and survival after bone marrow transplantation from HLA-identical siblings as treatment of acute and
bloodjournal.org	Walden P, Witherspoon R, Appelbaum F, Banaji M, Hansen J, Martin P, Sanders J, Singer J, Thomas ED. <b>Graft-versus-host disease</b> as adoptive immunotherapy in patients with advanced hematologic neoplasms. N Engl J Med 320:28, 1989. 22. Vignani F,
semanticscholar.org	and 4 National Yang-Ming University, School of Medicine, Taipei, Taiwan, R.O.C. 2 Hepatic <b>graft-versus-host disease</b> (GVHD) post allogeneic hematopoietic stem cell transplantation generally presents as cholestatic jaundice and
semanticscholar.org	peripheral blood stem cell transplantation (allo-PBSCT), donor lymphocyte infusion (DLI), <b>graft-versus-host disease</b> (GVHD) liver dysfunction is a common complication after hematopoietic stem cell transplantation (HSCT), which may be
semanticscholar.org	may be complicated by the concurrent adverse effects of pretransplant chemoradiotherapy, <b>graft-versus-host disease</b> (GVHD), GVHD prophylactic medication, veno-occlusive liver disease (VOD) and various viral hepatitis. The hepatic
semanticscholar.org	and alkaline phosphatase highly predictive liver function tests for the diagnosis of <b>graft-versus-host disease</b> in bone marrow transplantation recipients. Transplantation 1989;48:809-14. Miacarek M, Martin PJ, Lessening W,
semanticscholar.org	1984;4:123-30. Ghossein CF, Yau JC, LeMaistre CF, Cleary KR. Steroid-responsive chronic hepatic <b>graft-versus-host disease</b> without extrahepatic graft-versus-host disease. Am J Gastroenterol 1989;10:1306-9. Strasser SI, Shulman HM,
semanticscholar.org	KR. Steroid-responsive chronic hepatic graft-versus-host disease without extrahepatic <b>graft-versus-host disease</b> . Am J Gastroenterol 1989;10:1306-9. Strasser SI, Shulman HM, McDonald GB. Chronic graft-versus-host disease (GVHD) of the
semanticscholar.org	with a large pericardial effusion and constrictive pericarditis as manifestations of chronic <b>graft-versus-host disease</b> after non-myeloablative peripheral stem cell transplantation and subsequent donor lymphocyte infusion. Bone
semanticscholar.org	<b>graft-versus-host disease</b> . Atzkep G, Bolhoff JK, Lee LA, Hallick JP, Torbensohn M, Jacobsohn DA, et al. Hepatic variant of <b>graft-versus-host disease</b> after donor lymphocyte infusion. Blood 2002;100:3903-7. 415. We use cookies to make interactions with our website
medicinenet.com	cells of the graft. When graft cells attack host cells, the result is a dangerous condition called <b>graft-versus-host disease</b> (GVHD), which may be acute or chronic and may manifest as a skin rash, gastrointestinal illness, or liver disease. The
researchgate.net	. On multivariate analysis, low LC30 (HR, 2.44, 95% CI, 1.02-5.68; p=0.046) and grade II-IV acute <b>graft-versus-host disease</b> (HR, 2.41, 95% CI, 0.99-5.90, p=0.053) were significantly associated with worse OS. However, LC30 was not a risk;
researchgate.net	ALC independent of the infused stem cell or lymphocyte doses. Regulatory T cells (Tregs) prevent <b>graft-versus-host disease</b> by inhibiting the proliferation and function of conventional T cells (Tcons). However, the impact of Tregs on
bloodjournal.org	in nine (13%). Thirteen (19%) patients received methotrexate posttransplant to prevent <b>graft-versus-host disease</b> (GVHD). 29 (43%) Blood, Vol 79, No 1 (June 1), 1992; pp 3067-3070 From the International Bone Marrow Transplant
bloodjournal.org	, Kolb HJ, Rimm AA, Ringden O, Rozman C, Sobocinski K, Zwaan FE, Bortin MM. Risk factors for chronic <b>graft-versus-host disease</b> after HLA-identical sibling bone marrow transplantation. Blood 75:2458, 1990. 17. Warner RS, Bortin MM, Gale RP,
ptt.edu	A, Wilson DB. Analysis of T cell populations that induce and mediate specific resistance to <b>graft-versus-host disease</b> in rats. 1994. J Exp Med 160:652-658, 386. 139. M. Salo K, Tamura A, Naniwata H, Tadajima T, Nagashima Chied

Figura 14. Concordancias del término “graft-versus-host disease”

Si fuera necesario, el intérprete podría hacer clic en la palabra resaltada en rojo para visualizar la concordancia ampliada y obtener información adicional. Puede resultar muy útil para que comprenda e interiorice el concepto que subyace al término, como se aprecia en la figura 15. Si el intérprete continúa consultando las concordancias de este modo, podría empezar a intuir las interrelaciones entre términos y su estructura epistemológica.

fluid, which can lead to pneumonia and liver dysfunction. A mismatch between donor and recipient tissues can lead to an immune reaction between cells of the host and cells of the graft. When graft cells attack host cells, the result is a dangerous condition called graft-versus-host **disease** (GVHD), which may be acute or chronic and may manifest as a skin rash, gastrointestinal illness, or liver disease. The risk of GVHD can be minimized through careful tissue matching. Even when a donor antigen match is identical, roughly 40 percent of recipients still develop GVHD, rising to 60 to 80 percent when only a single antigen is mismatched. Because of the danger of this complication, autologous transplants are more commonly performed. Bone marrow transplantation was not previously recommended for patients aged over 50 years, due to

Figura 15. Concordancia ampliada

Por otro lado, si el profesional quisiera consultar el documento original para informarse más sobre el término, puede hacer clic en la URL que aparece a la izquierda de la línea de concordancia para ir a la página web o documento original (figuras 16 y 17). Tener acceso a estas páginas o archivos originales podría ser de gran utilidad para el intérprete, pues podrían emplearse para realizar traducción a vista de cara a prepararse para la conferencia. Los alumnos que emplearon Sketch Engine en el estudio de Xu (2018) valoraron positivamente que la herramienta les permitiera realizar traducción a vista con facilidad.

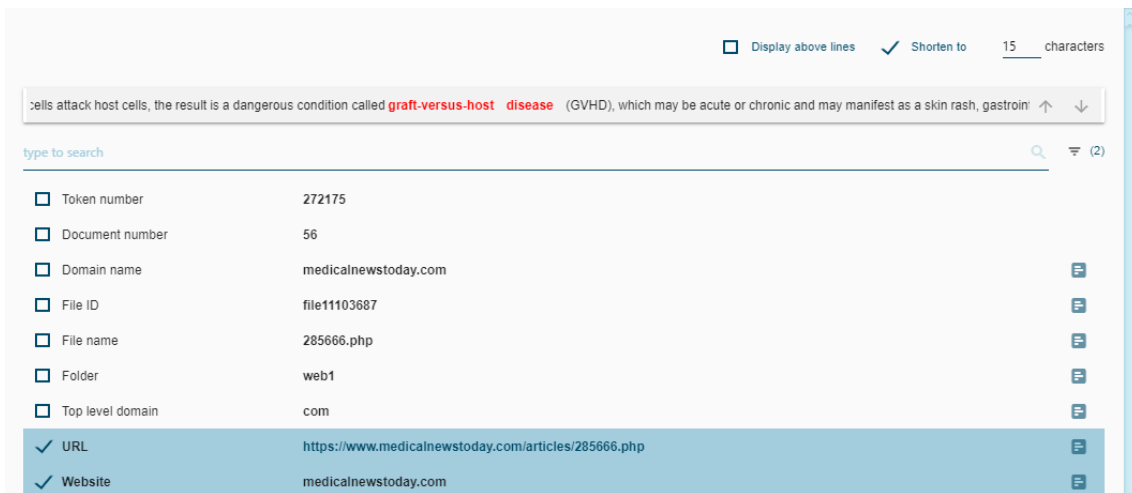


Figura 16. Menú que aparece tras hacer clic en la URL a la izquierda de la concordancia

## All you need to know about bone marrow

By [Hannah Nichols](#) | Last updated Fri 15 December 2017

[What is bone marrow?](#) | [Function](#) | [Transplants](#) | [Diseases](#) | [Donation](#)

Bone marrow is the spongy tissue inside some of the bones in the body, including the hip and thigh bones. Bone marrow contains immature cells, called stem cells.

Many people with blood [cancers](#), such as [leukemia](#) and [lymphoma](#), sickle cell [anemia](#), and other life-threatening diseases, rely on bone marrow or cord blood transplants to survive.

Healthy bone marrow and blood cells are needed in order to live. When disease affects bone marrow so that it can no longer function effectively, a marrow or cord blood transplant could be the best treatment option; for some patients it is the only potential cure.

Figura 17. Tras hacer clic en la URL se nos dirige a la página original

Por otra parte, el intérprete también puede acceder a las concordancias mediante el menú desplegable de la izquierda de la pantalla, que le conducirá al buscador (pestaña “basic”). En la pestaña “advanced” podrá afinar la búsqueda mediante comodines y expresiones regulares, entre otros, pero a nivel de usuario basta con la búsqueda básica de concordancia. En las siguientes figuras se aprecian los resultados de concordancia para el término “*engraftment*” (injerto) realizando la búsqueda desde la pestaña “basic”.



Word	Cooccurrences	Candidates	T-score	MI	↓ LogDice
1 repopulation	5	18	2.24	11.17	9.94 ---
2 permits	5	21	2.23	10.95	9.91 ---
3 Neutrophil	4	15	2.00	11.11	9.64 ---
4 demonstrate	5	62	2.23	9.39	9.60 ---
5 develops	5	62	2.23	9.39	9.60 ---
6 subsequent	6	118	2.44	8.72	9.52 ---
7 prompt	3	22	1.73	10.14	9.17 ---
8 kinetics	3	23	1.73	10.08	9.16 ---
9 incomplete	3	24	1.73	10.02	9.15 ---
10 opportunistic	3	50	1.73	8.96	8.95 ---
11 durable	3	78	1.73	8.32	8.76 ---
12 CD34	3	101	1.73	7.94	8.62 ---
13 achieved	4	212	1.99	7.29	8.50 ---
14 known	6	419	2.43	6.89	8.44 ---
15 infections	7	584	2.62	6.64	8.29 ---
16 until	3	200	1.72	6.96	8.14 ---
17 process	4	325	1.98	6.67	8.11 ---
18 defined	3	267	1.71	6.54	7.88 ---
19 donor	13	1,965	3.54	5.78	7.65 ---
20 syndrome	3	345	1.71	6.17	7.64 ---

Figura 21. Colocaciones del término “engraftment” obtenidas a partir de las concordancias.

Left context	KWIC	Right context
, busulfan and 4–8 Gy TBI were added in 28 patients. <S><S> These regimens might have been beneficial for establishing durable engraftment	in RIST for advanced ALL, and might have contributed to temporary control of disease. <S><S> The role of additional	
These studies showed that, although inclusion of alemtuzumab in the preparative regimen was associated with durable engraftment	and a low incidence of GVHD, this did not necessarily translate into a survival benefit when retrospectively compared	
+ cells: CD34+Lin-Selectin+ is the best predictor of engraftment rapidity and CD34+Thy-1+ correlates with durable engraftment (Pratt et al., 2001). <S><S> Recovery of lymphocyte subsets after ASCT: CD3+ cells returns to normal after 2 years in HL and NHL		

Figura 22. Concordancia de la combinación “durable” + “engraftment”

Por otro lado, la función “tesauro” de Sketch Engine también podría resultar útil al intérprete en esta segunda fase de obtención de información y de creación de un sistema de términos interrelacionados. Cuando introduzca un término en el campo de búsqueda, el programa le mostrará como resultado todas las palabras que compartan con él patrones de colocación (Kilgarriff *et al* 2014b). Gracias a este algoritmo, el intérprete podrá en ocasiones encontrar sinónimos, aunque no siempre será posible. Sería interesante introducir los términos más relevantes para comprobar la existencia de sinónimos. En la figura 22 se aprecia cómo tras introducir en el campo de búsqueda el término “transplant” aparece “transplantation” en primer lugar. Esto podría indicar que estamos ante un sinónimo, pero es necesario asegurarse mediante documentación externa o con las concordancias.

	Word	Frequency ?
1	transplantation	2,896 ...
2	BMT	569 ...
3	chemotherapy	2,989 ...
4	treatment	4,889 ...
5	therapy	2,267 ...
6	HSCT	551 ...
7	procedure	1,309 ...
8	donor	2,532 ...
9	DLI	546 ...
10	SCT	245 ...

Figura 22. Resultados del tesauro para el término “transplant”

Para documentarse el intérprete cuenta además con la herramienta *word sketch difference* si hace clic en el icono de los tres puntos al lado del término en cuestión. Con esta opción se pueden visualizar las colocaciones de cada término y cuáles son comunes a ambos (figura 23). De nuevo, contamos con la opción de acceder a las concordancias de cada combinación (por ejemplo *blood + transplant* y *blood + transplantation*).

"transplant/transplantation" and/or ...		verbs with "transplant/transplantation" as object		verbs with "transplant/transplantation" as subject		adjective predicates of "transplant/transplantation"	
donor	6 0 ...	need	21 0 ...	help	10 0 ...	necessary	0 4 ...
regimen	5 0 ...	purge	12 0 ...	replace	4 0 ...	effective	0 12 ...
cell	32 4 ...	lymphodepleting	10 0 ...	take	5 0 ...		
transplant	32 5 ...	do	27 5 ...	depend	12 4 ...		
patient	21 8 ...	receive	51 13 ...	use	17 9 ...		
blood	19 15 ...	stem	45 39 ...	be	257 193 ...		
infusion	4 7 ...	perform	20 19 ...	involve	9 7 ...		
chemotherapy	14 41 ...	undergo	18 51 ...	have	22 40 ...		
transplantation	5 20 ...	follow	17 101 ...	follow	12 24 ...		
harvesting	0 6 ...	recommend	0 8 ...	develop	0 5 ...		
remission	0 9 ...	compare	0 19 ...	induce	0 4 ...		
conditioning	0 8 ...	deplete	0 12 ...	improve	0 8 ...		

"transplant/transplantation" is a ...		modifiers of "transplant/transplantation"		nouns modified by "transplant/transplantation"		pronominal possessors of "transplant/transplantation"	
process	4 0 ...	Autologous	39 25 ...	center	50 0 ...	my	4 0 ...
month	4 0 ...	organ	23 31 ...	Proc	17 0 ...	your	7 0 ...
type	5 0 ...	bone	357 415 ...	team	20 0 ...		
procedure	16 4 ...	marrow	390 530 ...	doctor	10 0 ...		
treatment	17 5 ...	Allogeneic	48 102 ...	patient	43 0 ...		

Figura 23. "Word sketch difference" entre "transplant" y "transplantation".

### 3.3.4 Tercera y cuarta fase

La tercera fase que establece Xu (2018) es la búsqueda de equivalentes terminológicos en lengua meta. En nuestro caso, el intérprete podría crear un corpus sobre la leucemia como hemos hecho en inglés pero introduciendo las palabras clave en español. Tras realizar la extracción y revisar las concordancias como se ha explicado, el intérprete encontrará en contexto gran parte de los equivalentes terminológicos. Como explican Fantinuoli y Prandi (2018): *"if the extraction [...] is performed for both languages, the user can try to establish terminology connections between languages"*<sup>19</sup>. Por otro lado, Sketch Engine también puede ayudar al intérprete a encontrar fácilmente equivalentes en lengua meta si se crea un corpus paralelo. Se explicará el proceso en el siguiente capítulo. Para la cuarta fase (mejorar la velocidad de acceso y búsqueda de los términos) se deberá recurrir a herramientas externas, pues Sketch Engine no cuenta con un sistema de creación de glosarios. El intérprete podría optar por una herramienta simple como Excel. De hecho, es posible descargar directamente la lista de términos desde Sketch Engine en formato XLS. Esto resultaría de gran utilidad para el intérprete, puesto

<sup>19</sup> "Si se realiza una extracción en cada una de las dos lenguas, es posible que el usuario establezca conexiones terminológicas entre ambas" (La traducción es mía).

que podría ir modificando y editando la lista en Excel para elaborar su glosario final. También existen opciones más avanzadas y diseñadas para intérpretes que permiten una consulta rápida en cabina. Contamos, por ejemplo, con herramientas como LookUp, Intragloss, InterpretBank o BoothMate (Corpas Pastor 2018), aunque estas quedan fuera de nuestro campo de estudio.

### **3.4 Empleo de Sketch Engine para un encargo real**

Para describir nuestro flujo de trabajo e ilustrar cómo se puede trabajar con Sketch Engine emplearemos el caso real de unas prácticas de interpretación realizadas en el 28º período de sesiones de la Comisión de Prevención del Delito y Justicia Penal de las Naciones Unidas. Consideramos que se trata de un buen ejemplo para ejemplificar el proceso de preparación, pues la ONU cuenta con una fraseología característica y la temática (derecho y prevención del delito) es terminológicamente densa. Además, otros intérpretes más experimentados nos advirtieron de la velocidad de lectura de discursos en las Naciones Unidas, que es uno de los factores que pueden causar dificultades al intérprete (Seeber 2011). A este escollo se tienen que añadir otros como los acentos, un lenguaje poco oralizado y sintácticamente complejo<sup>20</sup> y la presencia de fragmentos densos desde el punto de vista terminológico. Todas estas circunstancias aumentan las necesidades de procesamiento y la demanda cognitiva del intérprete, por lo que resultaba más necesario que nunca interiorizar la terminología para reducir la carga cognitiva y poder destinar más esfuerzos a salvar las dificultades mencionadas.

#### **3.4.1 Creación de un corpus paralelo**

Una ventaja de trabajar para la ONU es la existencia de mucha documentación previa de los anteriores períodos de sesiones. Además, una parte de los documentos de la sesión se colgó con una semana de antelación, por lo que se pudo emplear para compilar un corpus. Como mencionamos en apartados anteriores, la selección del tipo de corpus se deberá realizar en función de las características del encargo. Si bien hemos puesto de manifiesto las ventajas de los corpus compilados mediante internet, consideramos que para este encargo lo óptimo sería emplear un corpus pequeño *ad hoc* con los documentos específicos de la reunión. La razón es que el uso del *webcrawler* no garantiza que las páginas y archivos extraídos sean relativos a la ONU; necesitamos manejar documentos muy específicos para obtener la terminología y fraseología de las Naciones Unidas, que

---

<sup>20</sup> En la ONU, leer los discursos es la norma, por lo que el lenguaje empleado se acerca más al escrito que al oral.

es muy concreta. Los documentos empleados fueron los seis proyectos de resolución colgados en la página web de la ONU (EN/ES)<sup>21</sup> una semana antes del inicio del período de sesiones y el resto de documentos (EN/ES) relativos al período de sesiones<sup>22</sup>. También se emplearon los mismos documentos pero del período de sesiones anterior<sup>23</sup>. Estimamos que emplear este flujo de trabajo con corpus puede ser una alternativa mucho más eficiente a leer y extraer manualmente los términos de las decenas de documentos de la sesión. De hecho, Xu y Sharoff (2014) exponen que los estudiantes de interpretación que emplearon en su preparación el método tradicional de lectura de documentos y extracción de términos tardaron 9 horas en generar las listas de términos iniciales, mientras que quienes emplearon la extracción automática tardaron de media 4 horas en generar sus listas de términos bilingües. Además, la retroalimentación de los alumnos fue positiva: *“most of them expressed the opinion that if they are given very limited time, they would prefer to use the automatically-generated list for their preparation<sup>24</sup>”*.

Una vez descargados los archivos mencionados en inglés y en español, se clasificaron en carpetas y se empleó un código alfanumérico para facilitar su identificación (Seghiri 2017). El siguiente paso fue crear un archivo bilingüe (en nuestro caso, TMX) con los documentos descargados para importarlo a Sketch Engine. Algunos autores (Seghiri 2017) emplean LF Aligner, pero esta herramienta no permite alinear documentos por lotes. Por ello, la herramienta empleada fue SDL Trados Studio 2019, que nos permite alinear todos los archivos a la vez y ahorrar tiempo. Debemos acceder al menú de memorias de traducción del programa y seleccionar el icono “alinear documentos”. Se mostrará un menú desplegable en el que seleccionaremos la opción “alinear varios archivos”. En el siguiente menú (figura 24) crearemos una memoria de traducción vacía. Es importante que desmarquemos la casilla “guardar los archivos resultantes de la alineación para revisarlos más tarde”. De no hacerlo, se creará un archivo SDLALIGN por cada par de ficheros, y lo que queremos es que todas las alineaciones

---

<sup>21</sup> Disponibles en [https://www.unodc.org/unodc/en/commissions/CCPCJ/session/28\\_Session\\_2019/resolutions-28-of-the-ccpcj.html](https://www.unodc.org/unodc/en/commissions/CCPCJ/session/28_Session_2019/resolutions-28-of-the-ccpcj.html). Estos documentos son importantes puesto que su contenido se debate en la Comisión Plenaria.

<sup>22</sup> Disponibles en [https://www.unodc.org/unodc/en/commissions/CCPCJ/session/28\\_Session\\_2019/session-28-of-the-ccpcj.html](https://www.unodc.org/unodc/en/commissions/CCPCJ/session/28_Session_2019/session-28-of-the-ccpcj.html).

<sup>23</sup> Disponibles en [https://www.unodc.org/unodc/en/commissions/CCPCJ/session/27\\_Session\\_2018/session-27-of-the-ccpcj.html](https://www.unodc.org/unodc/en/commissions/CCPCJ/session/27_Session_2018/session-27-of-the-ccpcj.html)

<sup>24</sup> “La mayoría opinó que, de contar con poco tiempo, emplearían para prepararse las listas generadas automáticamente” (La traducción es mía).

pasen directamente a la memoria sin revisión. Dado que trabajamos con traducciones exactas sin ninguna frase o párrafo extra en ninguna de las dos versiones, la alineación no debería causar problemas.

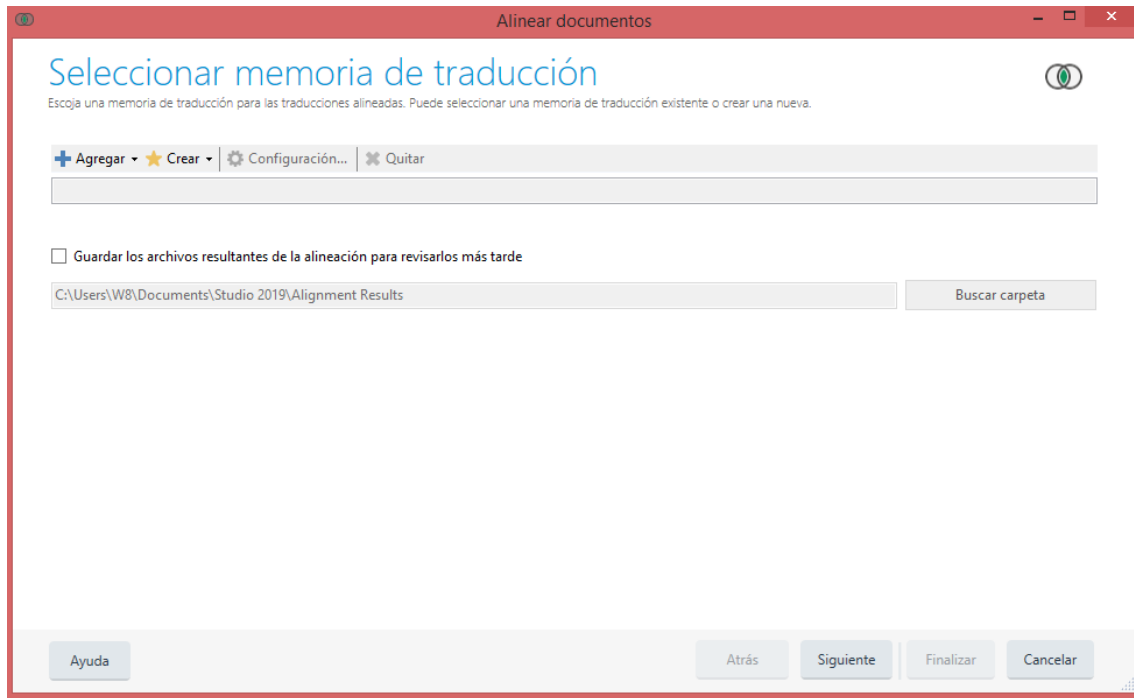


Figura 24. Menú para crear la memoria de traducción

Tras aceptar, en el siguiente menú (figura 25) se deben añadir las carpetas con los archivos que queramos alinear. Haremos clic en “agregar” y “archivos de origen por carpeta”/ “archivos de destino por carpeta”. A continuación aceptaremos y los segmentos alineados pasarán automáticamente a nuestra memoria de traducción.

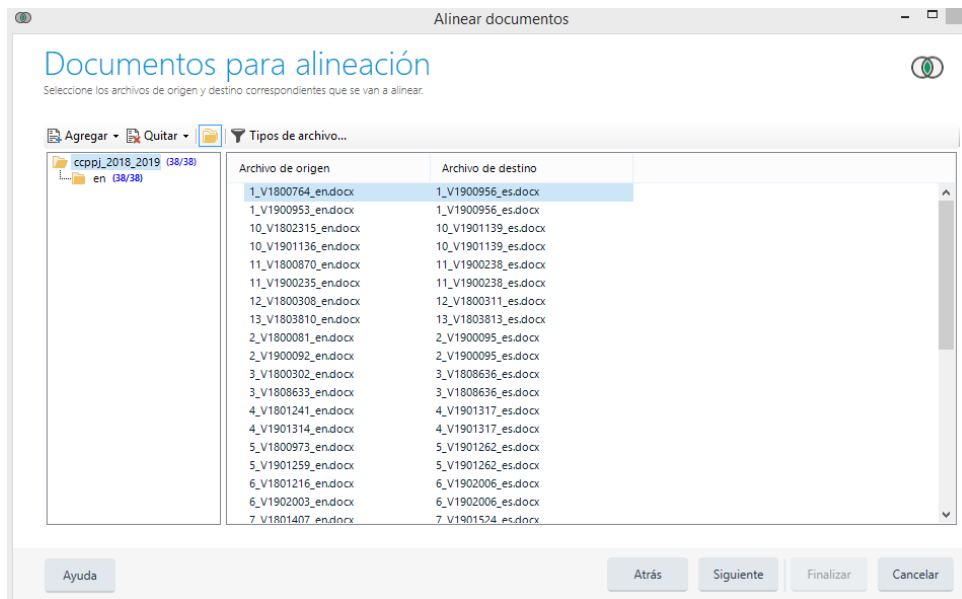


Figura 25. Menú de alineación de documentos por lotes

En el siguiente paso exportaremos la memoria en formato TMX para poder importarla a Sketch Engine. Para ello, la abrimos con SDL Trados Studio y en la cinta de opciones seleccionamos “exportar” (figura 26) y elegimos un directorio. A continuación, volveremos a Sketch Engine y crearemos un corpus como se explicó en apartados anteriores pero seleccionando la opción “*multilingual*” e importando nuestro archivo TMX.

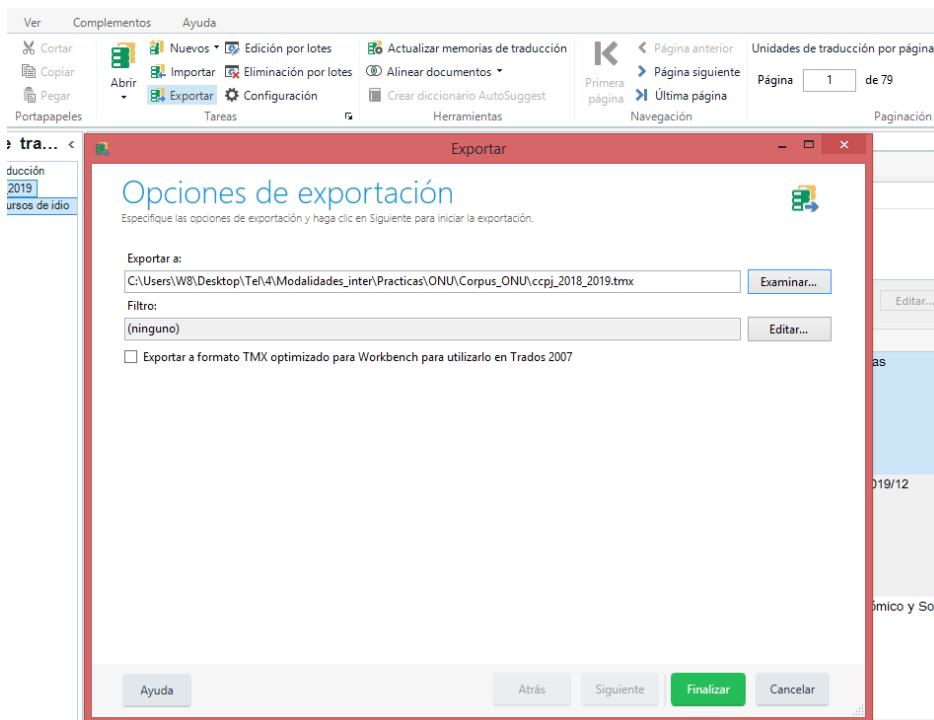


Figura 46. Exportación de la memoria al formato TMX

### 3.4.2 Extracción de términos

Como hemos visto, la primera fase de la preparación siempre debe ser la creación de una lista de términos. Para ello, consideramos que el flujo de trabajo más eficaz sería el siguiente<sup>25</sup>. En primer lugar, el intérprete seleccionaría el corpus en inglés (al crear un corpus paralelo este se subdivide en las dos lenguas) y realizaría la extracción terminológica como se ha explicado. Los términos extraídos se descargarían en formato XLS pulsando en el símbolo de descarga de encima de la lista de términos. En la siguiente figura podemos ver una muestra de los resultados de la extracción del corpus inglés y comprobar que presenta una calidad aceptable.

	A	B	C	D
1	Term	Score	Freq	Ref freq
2	crime prevention	1.371.080	235	125
3	criminal justice	688.760	357	859
4	transnational organized crime	487.170	56	6
5	twenty-seventh session	445.950	50	0
6	international cooperation	401.750	83	200
7	twenty-eighth session	374.760	42	0
8	organized crime	362.480	104	370
9	criminal justice programme	356.960	40	0
10	justice programme	355.540	40	1
11	violent extremism	326.890	40	22
12	provisional agenda	325.060	37	4
13	cultural property	318.890	44	55
14	restorative justice	314.200	49	93
15	criminal misuse	307.540	35	4
16	draft resolution	272.480	43	97
17	mutual legal assistance	263.750	30	4
18	criminal justice programme network	250.170	28	0
19	justice programme network	250.170	28	0
20	programme network	249.180	28	1
21	technical assistance	235.960	116	803
22	thematic discussion	232.370	26	0
23	field of crime prevention	201.450	23	5
24	electronic evidence	192.730	22	5
25	intergovernmental working group	187.880	21	0
26	open-ended intergovernmental working group	187.880	21	0
27	standing open-ended intergovernmental working group	187.880	21	0
28	legal assistance	182.570	33	146
29	gender-related killing	170.080	19	0
30	programme support	159.270	18	3

Figura 27. Muestra de los términos extraídos

<sup>25</sup> Se han considerado otras opciones como la extracción terminológica bilingüe, que combinaría la primera fase de creación de una lista de términos con la tercera etapa de búsqueda de equivalentes en lengua meta. Para ello se ha empleado el extractor bilingüe de la interfaz antigua de Sketch Engine (<https://old.sketchengine.co.uk/login/>) y el extractor de la nueva interfaz (<https://terms.sketchengine.eu/>). No obstante, los resultados no han sido satisfactorios, puesto que la calidad de los términos extraídos dejaba que desear y en muchas ocasiones se extraía un equivalente incorrecto en lengua meta. Por ello, consideremos que este enfoque, pese a ser completamente automático, resultaría contraproducente.

Tras esta extracción el intérprete podría eliminar de la lista en Excel los términos mal formados o irrelevantes. No debemos descuidar la lista de unigramas extraídos, pues nos puede aportar términos interesantes (en nuestro caso “*perpetrators*”, “*peacebuilding*”, “*intersex*”, “*recidivism*”, “*rappporteur*”, “*femicide*”, “*prosecution*”, “*opiate*” o “*cybercrime*”, entre otros). Recordemos además que es posible realizar una extracción de la fraseología en esta etapa. Sería conveniente extraer los patrones de las palabras clave del encargo, como “*crime*”. La *word sketch* nos muestra los adjetivos que modifican al término (figura 28) y nos aporta un gran número de términos nuevos como “*hate crime*”, “*transnational organized crime*”, “*environmental crime*”, “*wildlife crime*”, “*maritime crime*”, “*drug-related crime*” o “*youth crime*”. Resultan asimismo de gran utilidad los verbos que forman patrones con “*crime*”, como “*combat crime*”, “*counter crime*”, “*adress crime*”, “*commit a crime*” o “*constitute a crime*”. Es vital que el intérprete interiorice estos patrones y sus equivalentes en español, pues el término “*crime*” aparecerá continuamente durante las reuniones.

modifiers of "crime"	nouns modified by "crime"	verbs with "crime" as object
<b>organized</b> ... transnational organized crime	<b>prevention</b> ... crime prevention and criminal justice	<b>motivate</b> ... crime motivated by intolerance or
<b>transnational</b> ... transnational organized crime and	<b>trend</b> ... World crime trends and emerging issues	<b>combat</b> ... cooperation in combating transnational organized crime and corruption
<b>hate</b> ... hate crime	<b>strategy</b> ... crime prevention strategies , policies and	<b>prevent</b> ... prevent crime
<b>serious</b> ... a serious crime	<b>convention</b> ... of the Organized Crime Convention	<b>counter</b> ... in preventing and countering crime motivated by intolerance
<b>environmental</b> ... environmental crime and the	<b>violence</b> ... and combating organized crime , violence , trafficking in	<b>address</b> ... order to effectively address transnational organized crime
<b>wildlife</b> ... of wildlife crime	<b>policy</b> ... crime prevention policies , strategies and	<b>fight</b> ... international community for fighting transnational organized crime , and noted
<b>other</b> ... other crimes	<b>survey</b> ... such as the crime trends survey , the	<b>hate</b> ... related to hate crime
<b>drug-related</b> ... responses to drug-related crime ; and countering	<b>effort</b> ... youth participation in crime prevention efforts	<b>commit</b> ... crime committed
<b>maritime</b> ... maritime crime		<b>constitute</b> ... constitute crimes
<b>corruption</b> ... corruption crimes		<b>include</b> ... including more serious crimes
<b>youth</b> ... preventing youth crime through		

Figura 28. Patrones gramaticales y de colocación de “*crime*”.

Otra palabra clave cuyos patrones de colocación convendría aprender sería *law*, dado que las reuniones versarán sobre justicia penal. En este caso obtenemos resultados interesantes como “*domestic law*”, “*applicable law*”, “*procedural law*”, “*cybercrime law*” o “*draft law*”, entre otras. La *word sketch* muestra además verbos que forman colocaciones con el término como “*update a law*” o “*adopt a law*”, así como términos relacionados de gran importancia: “*law enforcement officials*”, “*law enforcement authorities*”, “*law enforcement bodies*” o “*law assistance*”, entre otros.

### **3.4.3 Concordancia bilingüe y búsqueda de equivalentes terminológicos**

Si contamos con un corpus paralelo, podemos aunar la segunda y tercera fase del proceso, puesto que las concordancias bilingües nos ofrecerían una solución muy sencilla para obtener los equivalentes de los términos en inglés que no conozcamos a la vez que nos proporcionarían información extralingüística y contextual. Por ejemplo, si el intérprete desconociera el equivalente español del término “*wildlife crime*”, solo tendría que realizar una búsqueda de concordancia paralela, que nos revelará que en español el concepto se expresa como “delitos contra la vida silvestre” (figura 29). En la figura 30 encontramos otro ejemplo de cómo se puede encontrar equivalentes terminológicos mediante estas concordancias.

<S> A number of speakers expressed their concern about the global increase in <b>wildlife crime</b> and its links to organized crime and money-laundering. </S>	<S> Varios oradores expresaron su preocupación por el aumento a nivel mundial de los <b>delitos</b> contra la vida <b>silvestre</b> y sus vínculos con la delincuencia organizada y el blanqueo de dinero. </S>
<S> Owing to its technical assistance, the impact UNODC has had on combating <b>wildlife crime</b> has been increasing. </S>	<S> Gracias a su asistencia técnica, la UNODC ha tenido mayor repercusión en las actividades de lucha contra los <b>delitos</b> que afectan a la vida <b>silvestre</b> . </S>
<S> Rapid reference guides for investigators and prosecutors of <b>wildlife crime</b> were developed for Botswana, Mozambique and Uganda. </S>	<S> En África Oriental y Meridional, se han reforzado las capacidades de enjuiciamiento. </S>
<S> Work started on the second World <b>Wildlife Crime</b> Report. </S>	<S> Se inició la labor de preparación del segundo Informe Mundial sobre los <b>Delitos</b> contra la Vida <b>Silvestre</b> . </S>
<S> Moreover, a number of publications were under preparation and were expected to be published in 2019, including guides on corruption and fraud risk assessment in government organizations, addressing corruption in the fisheries sector and addressing corruption in relation to <b>wildlife crime</b> . </S>	<S> Además, se estaban preparando varios documentos cuya publicación estaba prevista para 2019, entre otros, guías sobre la corrupción y la evaluación del riesgo de fraude en las organizaciones gubernamentales, la lucha contra la corrupción en el sector pesquero y la lucha contra la corrupción en relación con los <b>delitos</b> contra la vida silvestre. </S>

Figura 29. Búsqueda de concordancia paralela para obtener equivalentes terminológicos

<p>① &lt;S&gt; (a) Approved, provisionally, the projected use of general-purpose funds in the biennium 2018–2019 and requested the Executive Director of the United Nations Office on Drugs and Crime (UNODC) to revise the funding arrangement proposed for the Research and Trend Analysis Branch, as well as the Justice Section, and to present to the <b>standing open-ended intergovernmental working group</b> on improving the governance and financial situation of UNODC, at its formal meeting in early 2018, a revised budget that addressed that funding arrangement by redistributing resources throughout the Office, for approval by the Commissions at their regular sessions; &lt;/S&gt;</p>	<p>&lt;S&gt; a) Aprobaron, con carácter provisional, la utilización prevista de los fondos para fines generales en el bienio 2018-2019 y solicitaron al Director Ejecutivo de la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC) que revisara los arreglos de financiación propuestos para la Subdivisión de Investigación y Análisis de Tendencias, así como la Sección de Justicia, y que presentara al <b>grupo de trabajo intergubernamental permanente de composición</b> abierta sobre el mejoramiento de la gobernanza y la situación financiera de la UNODC, en el período de sesiones oficial que celebraría a principios de 2018, un presupuesto revisado que tuviera en cuenta ese mecanismo de financiación mediante la redistribución de recursos en toda la Oficina, para que lo aprobaran las comisiones en sus períodos ordinarios de sesiones; &lt;/S&gt;</p>
<p>① &lt;S&gt; (a) Work of the <b>standing open-ended intergovernmental working group</b> on improving the governance and financial situation of the United Nations Office on Drugs and Crime; &lt;/S&gt;</p>	<p>&lt;S&gt; a) Labor del <b>grupo de trabajo intergubernamental permanente de composición</b> abierta sobre el mejoramiento de la gobernanza y la situación financiera de la Oficina de las Naciones Unidas contra la Droga y el Delito; &lt;/S&gt;</p>
<p>① &lt;S&gt; (b) Note by the Secretariat on the work of the <b>standing open-ended intergovernmental working group</b> on improving the governance and financial situation of the United Nations Office on Drugs and Crime (E/CN.7/2018/3–E/CN.15/2018/3); &lt;/S&gt;</p>	<p>&lt;S&gt; b) Nota de la Secretaría sobre la labor del <b>grupo de trabajo intergubernamental permanente de composición</b> abierta sobre el mejoramiento de la gobernanza y la situación financiera de la Oficina de las Naciones Unidas contra la Droga y el Delito (E/CN.7/2018/3–E/CN.15/2018/3); &lt;/S&gt;</p>

Figura 30. Búsqueda de concordancia paralela para encontrar el equivalente de “standing open-ended intergovernmental working group”

Además, las concordancias proporcionarían al intérprete información fraseológica muy valiosa en español. Por ejemplo, pongamos que mediante las *word sketches* el intérprete encontrase *enact* como verbo que forma una colocación con *law*. Si quisiera saber cómo se expresa este verbo en español, podría recurrir a la concordancia bilingüe, que le revelará que el equivalente es *promulgar una ley*. Otra información fraseológica crucial para el intérprete sería el equivalente de verbos omnipresentes en los proyectos de resolución como “welcome”, “call upon”, “request”, “urge” o “encourage”<sup>26</sup>. Las concordancias le revelarán al profesional los equivalentes fraseológicos que se emplean siempre en la redacción de proyectos de resolución en español: “acoger con beneplácito”, “exhortar”, “solicitar”, “instar” y “alentar” respectivamente. La concordancia paralela nos ayudará asimismo a aprender la fraseología del derecho, que aparecerá muy a menudo durante las reuniones, como “pursuant to” (“con arreglo a” o “de conformidad con”) o “in accordance with” (“de conformidad con”). Para la última fase de la labor previa (asegurar la rapidez de búsqueda y consulta de los términos) se ha empleado Excel para realizar un glosario, aunque ya hemos mencionado que existen otras alternativas más especializadas (Corpas Pastor 2018). En el anexo del presente trabajo se muestra la lista bilingüe de términos fruto de todo el proceso descrito que fue empleada para preparar las reuniones.

<sup>26</sup> Es imprescindible que el intérprete conozca la fraseología de los proyectos de resolución en ambas lenguas, puesto que en la Commission on the Whole (COW) se ultima la redacción del proyecto y se debaten cuestiones lingüísticas muy concretas.

## CONCLUSIÓN

A la luz de lo expuesto en el presente trabajo, es innegable que la interpretación de conferencias y la terminología forman un vínculo indisoluble, ya que el intérprete se enfrenta en su trayectoria profesional a numerosos campos de especialidad, es el responsable de transmitir conocimientos altamente especializados y debe cumplir las expectativas de los clientes en un mercado cada vez más especializado y exigente. En este sentido, y dados los condicionantes inherentes a la labor del intérprete (velocidad del discurso, densidad terminológica, inmediatez, simultaneidad de los procesos, riesgo de sobrecarga cognitiva etc.), es imprescindible emprender una preparación terminológica previa a la conferencia. El objetivo de esta labor previa sería asegurar una transmisión clara del conocimiento y reducir la carga cognitiva ligada a la identificación y reproducción de terminología para poder emplear el poder de procesamiento liberado en otras operaciones (escucha, análisis, reformulación). Los estudios muestran que tanto la preparación intralingüística como la extralingüística mejoran el rendimiento del intérprete (Díaz Galaz 2013). No obstante, por motivos pragmáticos (falta de tiempo), el intérprete debe realizar en la mayoría de los casos una labor meramente terminológica.

Pese a la importancia crucial de esta labor, los flujos de trabajo descritos por los estudiosos hasta la fecha presentan carencias y suelen ser tradicionales, con algunas excepciones. Por otro lado, las encuestas a intérpretes revelan que su preparación sigue siendo a día de hoy eminentemente tradicional, empleando diccionarios monolingües y bilingües y búsquedas en internet y extrayendo los términos de los documentos pertinentes de manera manual. Dada esta preparación poco eficiente, la escasa disponibilidad de herramientas terminológicas que agilicen la labor del intérprete y el desconocimiento de las mismas por parte de los profesionales (Corpas Pastor 2016), en este trabajo hemos presentado el programa de gestión de corpus Sketch Engine y hemos expuesto un flujo de trabajo para agilizar la labor terminológica previa del intérprete según las recomendaciones de Xu (2018). Se ha optado por este enfoque de corpus debido a su repercusión positiva en el rendimiento del intérprete, que ha sido probada empíricamente. En el flujo de trabajo presentado, el intérprete compilaría un corpus automáticamente mediante semillas o crearía un corpus bilingüe o monolingüe manualmente con sus propios documentos. Más tarde, extraería la terminología y la fraseología automáticamente y revisaría las concordancias para afianzar el aprendizaje y para obtener información intralingüística y extralingüística. A continuación, buscaría los

equivalentes terminológicos en lengua meta mediante la consulta de concordancias en el caso de un corpus paralelo o mediante la constitución de otro corpus monolingüe en la lengua de destino. Finalmente, crearía y gestionaría su glosario mediante las herramientas existentes a tal efecto. Con este trabajo, en resumen, se ha pretendido suplir el vacío bibliográfico en el ámbito de convergencia entre terminología e interpretación y proporcionar al intérprete un flujo de trabajo automatizado con corpus para gestionar su preparación terminológica de manera más ágil y eficiente.

## BIBLIOGRAFÍA

- Alcina Caudet, M<sup>o</sup> Amparo y Gamero Pérez, Silvia (eds.). 2002. *La traducción científico-técnica y la terminología en la sociedad de la información*. Castellón: Universitat Jaume I.
- Bernardini, Silvia y Ferraresi, Adriano. 2013. “Old needs, new solutions: Comparable corpora for language professionals”. En *Building and Using Comparable Corpora*, ed. Por Serge Sharoff, Reinhard Rapp, Pierre Zweigenbaum y Pascale Fung. Berlin, Heidelberg: Springer, 303–319.
- Cabré, María Teresa. 2000a. “El traductor y la terminología: necesidad y compromiso”. *Panacea@: Revista de Medicina, Lenguaje y Traducción* 2: 2-3.
- Cabré, María Teresa. 2000b. “Terminologie et linguistique: La théorie des portes”. *Terminologies Nouvelles* 21: 10-15.
- Cabré, María Teresa. 2010. “Terminology and Translation”. En *Handbook of Translation Studies I*, ed. por Y.Gambier y L. Van Doorslaer. Amsterdam: John Benjamins, 356-365.
- Corpas Pastor, Gloria y Lily Fern. 2016. “A survey of interpreters’ needs and practices related to language technology”. Documento técnico [FFI2012-38881-MINECO/TI-DT-2016-1]. Málaga: Universidad de Málaga.
- Corpas Pastor, Gloria. 2018. “Tools for interpreters: the challenges that lie ahead”. *Current Trends in Translation Teaching and Learning E* 5: 157-182.
- Costa, Hernani, Corpas Pastor, Gloria y Durán Muñoz, Isabel. 2014. “A comparative user evaluation of terminology management tools for interpreters”. *Proceedings of the 4th International Workshop on Computational Terminology*: 68-76.
- Costa, Hernani, Corpas Pastor, Gloria y Durán Muñoz, Isabel. 2015. “An interpreter’s guide to selecting terminology management tools”. *NATO Conference on Terminology Management*. Disponible en <https://eden.dei.uc.pt/~hpcosta/docs/papers/201511-NATO.pdf>

- Díaz Galaz, Stephanie. 2013. *La influencia del conocimiento previo en la interpretación simultánea de discursos especializados : un estudio empírico*. Tesis doctoral. Universidad Granada.
- Faber, Pamela. 2010. “Terminología, traducción especializada y adquisición de conocimiento”. En *La traducción en contextos especializados. Propuestas didácticas*, ed. por E. Alarcón Navío. Granada: Atrio, 87-95.
- Fantinuoli, Claudio y Prandi, Bianca. 2018. “Teaching information and communication technologies A proposal for the interpreting classroom”. *Trans-kom* 11 (2): 162-182.
- Fantinuoli, Claudio. 2006. “Specialized corpora from the Web and term extraction for simultaneous interpreters”. En *WaCky! Working papers on the Web as Corpus*, ed. por Marco Baroni y Silvia Bernardini. Bologna: GEDIT, 173-190.
- Fantinuoli, Claudio. 2017. “Computer-assisted preparation in conference interpreting”. *Translation & Interpreting*, 9(2): 24-37
- Fantinuoli, Claudio. 2018a. “Interpreting and technology: The upcoming technological turn”. En *Interpreting and Technology*, ed. por Claudio Fantinuoli. Berlin: Language Science Press, 1-12.
- Fantinuoli, Claudio. 2018b. “The use of comparable corpora in interpreting practice and training”. *The Interpreters' Newsletter* 23: 133–149.
- Gile, Daniel. 1986. “Le travail terminologique en interprétation de conférence”. *Multilingua* 5 (1): 31–36.
- Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training. Revised edition*. Amsterdam, Philadelphia: John Benjamins
- Gorjanc, Vojko. 2009. “Terminology resources and terminological data management for medical interpreters”. En *Spürst Du, wie der Bauch rauf-runter? Fachdolmetschen im Gesundheitsbereich. Is everything all topsy turvy in your tummy? Healthcare Interpreting*, ed. por D. Andres y S. Pöllabauer. München: Meidenbauer, 85-95.

- Kalina, Sylvia. 2006. "Zur Dokumentation von Maßnahmen der Qualitätssicherung beim Konferenzdolmetschen". En *Text and Translation. Theory and Methodology of Translation*, ed. por Carmen Heine, Klaus Schubert y Heidrun Gerzymisch-Arbogast. Tübingen: Narr, 253-268.
- Kilgarrif, Adam *et al.* 2014a. "Finding Terms in Corpora for Many Languages with the Sketch Engine". *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*: 53-56.
- Kilgarriff, Adam *et al.* 2014b. "The Sketch Engine: ten years on". *Lexicography* 1 (1): 7-36.
- Kurz, Ingrid. 1993. "Conference Interpretation: Expectations of Different User Groups". *The Interpreters' Newsletter* 5: 13-21.
- Lucarelli, Luigi. 2013. "Tips on helping interpreters prepare for your meeting". *aiic.net*. 9 de mayo de 2013. <https://aiic.net/blog/post/6533/>.
- Moser, Peter. 1995. "Survey on expectations of users of conference interpretation". Geneva: International Association of Conference Interpreters (AIIC).
- Moser-Mercer, Barbara. 1992. "Banking on Terminology: Conference Interpreters in the Electronic Age". *Meta: Journal des traducteurs* 37 (3): 507-522.
- Pérez Pérez, Pablo Salvador. 2018. "The use of a corpus management tool for the preparation of interpreting assignments: A case study". *Translation and Interpreting* 10 (1): 137-151.
- Pöschhacker, Franz y Zwischenberger, Cornelia. 2010. "Survey on quality and role: conference interpreters' expectations and self-perceptions". *aiic.net*. 15 de marzo de 2010. <https://aiic.net/page/3405/>.
- Rodríguez, Nadia y Schnell, Bettina. 2009. "A Look at Terminology Adapted to the Requirements of Interpretation". *Language Update* 6 (1): 21-27.
- Sánchez Gijón, Pilar. 2009. "Developing documentation skills to build do-it-yourself corpora in the specialized translation course". En *Corpus use and translating*.

- Corpus use for learning to translate and learning corpus use to translate*, ed, por A. Beeby, P. Rodríguez-Inés y P. Sánchez Gijón. Amsterdam: John Benjamins, 109-127.
- Seeber, Kilian G. 2011. “Cognitive load in simultaneous interpreting: Existing theories — new models”. *Interpreting. International Journal of Research and Practice in Interpreting* 13 (2): 176–204.
- Seghiri, Míriam. 2017. “Corpus e interpretación biosanitaria: extracción terminológica basada en bitextos del campo de la Neurología para la fase documental del intérprete”. *Panacea* 18 (45): 123-132
- Seleskovitch, Danica. 1968. *L'interprète dans les conférences internationales. Problèmes de langue et de communication*. Paris: Minard.
- Valentini, Cristina. 2002. “Uso del computer in cabina de interpretazione”. *aiic.net*. 4 de Febrero de 2002. <http://aiic.net/p/656>.
- Varantola, Krista. 2002. “Disposable corpora as intelligent tools in translation”. *Cadernos de Tradução IX — Tradução e Corpora* 1 (9): 171-189.
- Will, Martin. 2007. “Terminology work for simultaneous interpreters in LSP Conferences: Model and method”. *MuTra 2007 – LSP translation scenarios. Conference proceedings*. Disponible en [http://www.euroconferences.info/proceedings/2007\\_Proceedings/2007\\_Will\\_Martin.pdf](http://www.euroconferences.info/proceedings/2007_Proceedings/2007_Will_Martin.pdf)
- Xu, Ran y Sharoff, Serge. 2015. “Evaluating Term Extraction Methods for Interpreters”. *Proceedings of the 4th International Workshop on Computational Terminology*: 86-93.
- Xu, Ran. 2018. “Corpus-based terminological preparation for simultaneous interpreting». *Interpreting. International Journal of Research and Practice in Interpreting* 20 (1): 29–58.

Zanettin, Federico. 2002. "DIY Corpora: the WWW and the translator". En *Training the language services provider for the new millennium*, ed. por B. Maia, J. Haller y M. Ulrych. Porto: Faculdade de Letras, Universidade do Porto, 239-248.

**ANEXO: GLOSARIO REALIZADO CON SKETCH ENGINE MEDIANTE EL  
FLUJO DE TRABAJO DESCRITO PARA INTERPRETAR EN LA ONU**

<b>Términos EN</b>	<b>Términos ES</b>
adoption of	aprobación de
agenda item	tema del programa
aggravating circumstance	agravante (de un delito)
asset recovery	recuperación de activos
blockchain technology	tecnología de cadenas de bloques
building capacity	fomentar/crear capacidad
child abuse	abuso de menores
child exploitation and sexual abuse online	explotación y abusos sexuales de niños, niñas y adolescentes en línea
child sex offenders	delincuentes pedófilos
cybercrime	ciberdelito/ delito cibernético
confiscate	decomisar
crime	Delincuencia (en general)/delito (en particular)
crime prevention	prevención del delito
criminal justice	justicia penal
criminal misuse	uso delictivo
criminalize	tipificar (como delito)
culture of lawfulness	cultura de legalidad
customs	aduanas
cybercrime	delito cibernético
disaggregated by	desglosada(o)s por
dismantle	desarticular
draft resolution	proyecto de resolución
drug control	fiscalización de drogas
drug prevention	prevención de la drogodependencia
drug trafficking	tráfico de drogas
electronic evidence	pruebas electrónicas
enabler of	facilitador de
environmental crime	delitos contra el medio ambiente
evidence-based	con base empírica
evidence-informed	basado en datos empíricos
extension	prórroga
femicide	feminicidio
field office	oficina extrasede
foreign terrorist fighters	combatientes terroristas extranjeros
freeze/freezing	embargo/embargar
gender-based violence	violencia por razón de género/ violencia de género
gender-related killing	asesinato por razones de género
grantee	beneficiario

grooming	captación con fines sexuales
hate crime	delito de odio
hate speech	discurso de odio
human trafficking	trata de personas
imprisonment	encarcelamiento
in accordance with	con arreglo a
informal	oficioso (por ejemplo, las reuniones)
judicial process	proceso judicial
law enforcement	aplicación de la ley
legal aid	asistencia jurídica
legal assistance	asistencia jurídica
legal system	ordenamiento jurídico
maritime crime	delincuencia marítima
migrant smuggling	tráfico ilícito de migrantes
money-laundering	blanqueo de dinero
monitoring	vigilancia
mutual legal assistance	asistencia judicial recíproca
narcotic drugs	estupefacientes
open-ended intergovernmental working group	grupo de trabajo intergubernamental permanente de composición abierta
opiate	opíaceo
organized criminal group	grupo delictivo organizado
overcrowding (en prisiones)	hacinamiento
peacebuilding	consolidación de la paz
penalize	tipificar (como delito)
perpetrators	autores
policy guidance	orientación normativa
prison management	gestión penitenciaria
prison reform	reforma penitenciaria
prison staff	personal penitenciario
prison-based rehabilitation programmes	programas de rehabilitación en cárceles
prisoner file management	gestión de expedientes de reclusos
prisoners	reclusos
probation	libertad vigilada
procedural law	derecho procesal
prosecution of	enjuiciamiento de
prosecution office	fiscalía
provision of technical assistance	prestación de asistencia técnica
provisional agenda	programa provisional
provisions	disposiciones
public procurement	contratación pública
pursuant to	con arreglo a/ de conformidad con
rapporteur	relator
recidivism	reincidencia
reporting period	período sobre el que se informa/período que abarca el informe
restorative justice	justicia restaurativa
security governance	gobernanza de la seguridad
seize	incautar(se) de (vb). El adjetivo es "incautado".

session	periodo de sesiones
sexual exploitation	explotación sexual
stakeholders	partes interesadas ( <b>siempre</b> )
standing open-ended intergovernmental working group	grupo de trabajo intergubernamental permanente
stem	detener
task forces	grupos de tareas
technical assistance	asistencia técnica
trafficking in cultural property	tráfico de bienes culturales
training programme	programa de capacitación
transnational organized crime	delincuencia organizada transnacional
under (programa, iniciativa etc.)	en el marco del (programa, iniciativa etc.)
United Nations Crime Prevention and Criminal Justice Programme	Programa de las Naciones Unidas en materia de Prevención del Delito y Justicia Penal
victimization	victimización
vocational training	formación profesional
whistle-blower protection	protección de los denunciantes de irregularidades
wildlife crime	delitos contra la vida silvestre
witness protection	protección de testigos
world drug problem	problema mundial de las drogas
youth crime	delincuencia juvenil

<b>Siglas EN</b>	<b>Siglas ES</b>
UNODC (United Nations Office on drugs and crime)	UNODC (Oficina de las Naciones Unidas contra la Droga y el Delito)
UN-WOMEN (Nations Entity for Gender Equality and the Empowerment of Women)	ONU Mujeres
UNFPA (United Nations Population Fund)	UNFPA (Fondo de Población de las Naciones Unidas)
WHO	OMS
WCO (World Customs Organization)	OMA (Organización Mundial de Aduanas)
UNDP (United Nations Development Programme)	PNUD (Programa de las Naciones Unidas para el Desarrollo )
Office of the United Nations High Commissioner for Refugees (UNHCR)	ACNUR (Alto Comisionado de las Naciones Unidas para los Refugiados)
UNAIDS (Joint United Nations Programme on HIV/AIDS )	ONUSIDA
COW (Committee of the Whole)	Comité Plenario
United Nations Convention against Corruption (UNCAC)	Convención de las Naciones Unidas contra la Corrupción (UNCAC)
VIC (Vienna International Centre)	CIV (Centro Internacional de Viena)
World Customs Organization	Organización Mundial de Aduanas
Optional Protocol to the Convention on the Rights of the Child	Protocolo Facultativo de la Convención sobre los Derechos del Niño

United Nations Interregional Crime and  
Justice Research Institute

Instituto Interregional de las Naciones Unidas  
para Investigaciones sobre la Delincuencia y la  
Justicia