Design of a Japanese-English Graph Dictionary by character correlation

A comprehensive report on the design of a Japanese-English Learner's Dictionary application implementing an indexing system based on graph theory

Author

José Vicente Tomás Pérez

idu18585@usal.es

Supervisor

Alfonso Falero Folgoso

Department of Modern Philology

Master's degree in East Asian Studies - Specialization in Japanese Studies

Academic year 2021/2022

VNiVERSiDAD D SALAMANCA

CAMPUS OF INTERNATIONAL EXCELLENCE

Submission date: September 3rd, 2022

List of Acronyms y Abbreviations

BCCWJ	Balanced Corpus of Contemporary Written Japanese.
CH	in Chinese.
EDICT	Electronic DICTionary.
EDRDG	Electronic Dictionary Research and Development Group.
JMdict	Japanese-Multilingual Dictionary.
JP	in Japanese.
NDC	Nippon Decimal Classification.
NINJAL	National Institute for Japanese Language and Linguistics.
SRS	Spaced Repetition Systems.
TFG	Trabajo Final de Grado.
TFM	Trabajo Final de Máster.
VDRJ	Vocabulary Database for Reading Japanese.

Contents

List of Figures

1. Introduction

1.1. Overview

Electronic Japanese dictionaries have been available to the public since the early 2000s, but their usability and accessibility for different target users remain a field with considerable room for improvement. In the last decade, with the emergence of smartphone technology for the general public, electronic Japanese dictionaries have experienced a new resurgence in the form of mobile applications, although their core usability model has seen less change and improvement than expected. This especially impacts new learners that are not familiar nor accustomed to the core aspects of the language, having to learn a new indexing system on top of the language itself to properly use each dictionary in the first place.

The main objective of this Master's Thesis is to completely design in all respects a new Japanese learner's dictionary mobile application for modern devices to partly overcome the aforementioned issues. In order to understand this task's challenge, we will first research the models used in traditional Japanese dictionaries and, more specifically, their digital variants. With the knowledge of these past experiences, we will outline a new visual word indexing system, specifically tailored for our use case. Once we have clearly specified these core aspects of our model, we can start designing the application itself, starting from defining the linguistic and technical resources publicly available that we would use to create the back end of our dictionary. With all this material in place, we will be able to finally design the actual application user workflow and visual design, which should be intrinsically related to the indexing model defined for it. Using this final product proposal and a first design demon-

stration, we will be able to receive some feedback from potential target users to improve the overall experience of our digital dictionary.

1.2. Motivation

As a lasting Japanese learner myself, I have needed to look up countless words, individual kanji or even complete sentences in different digital dictionaries throughout my learning process. Using such dictionaries as a beginner student has always been tedious if you don't have the possibility to copy and paste the word directly into the search engine. Learners don't necessarily have knowledge of the phonetics of a certain character, and therefore have to rely on complicated on-screen kanji radical keyboards, or, in some of the most modern ones, calligraphy recognition engines, which could be less than ideal if you don't use a touchscreen device and your writing tool is the computer mouse directly. After my first year of study, I gained some proficiency in these techniques, being able from thereafter to look up words directly by inputting their characters with the help of a phonetic or *hiragana* keyboard.

Nevertheless, no matter how much experience I have using Japanese language digital dictionaries, I still encounter over and over again with a specific stumbling navigation problem. Japanese digital dictionaries are typically indexed on the *kanji* level, offering words containing one concrete character. This *kanji* indexing system is not replicated on the word level though, and typically it is required to access the individual entry for each character to look up the words containing it. Moreover, usually if a word is inputted into the search engine, the returned results only include compound words containing the whole word itself, not words related by individual *kanji*. As a Japanese learner, I have always feel the need to navigate from word to word in order to easily discover related vocabulary on the *kanji* level, bookmark them and keep the search from where I left without constraints. This project serves as my own personal proposal to tackle this issue, by developing an adapted indexing model based on graph theory for this purpose and an interface that properly reflects the intended user workflow.

Furthermore, I want to use my interdisciplinary academical background as an engineer to foster further the relationship between Japanese language studies and multimedia technology, as I also intended to do with my Bachelor's Thesis, centred around the recognition of Japanese handwritten characters using machine learning techniques. This way, I wish to close a circle by combining my two fields of study during my Bachelor's and Master's years, at first glance areas very far apart from each other, yet sharing multiple research possibilities when they get together. A very good example of this collaboration is the recent improvements in the field of the Japanese Digital Humanities, such as the use of optical recognition technologies to transliterate old Japanese cursive script to modern script in a swift manner Clanuwat et al. (2018). Another great example of this close cooperation between Japanese studies and informatics is the curation of large collections in the form of automated databases, such as the recent Database of Pre-Modern of Japanese Works, compiled by the National Institute of Japanese Literature (JP: 国文学研究資料館) NIJL-NW Database (2020).

1.3. Proposal and goals

The main purpose of this report is to fully design the project for a Japanese learners' digital dictionary in the form of a mobile application, using an indexing system specially defined for its use on said application. Therefore, the main objectives of this Master' s thesis are:

- Research the current state of the art of Japanese Digital Dictionaries, both traditional and learner-oriented, first by looking at its origins and then by compiling the main features available in each case study analysed, in order to then determine the average capabilities present in these modern digital dictionaries.
- Create a proposal for a new Japanese digital dictionary application, elaborating on its indexing system from a linguistic perspective and the resources needed to implement it from a technical perspective.
- Design the actual aforementioned application from end-to-end, with an interface built

around the expected user workflow while using the dictionary functions.

- Review the project design by taking feedback from potential users, outlining a development roadmap to follow after the completion of this essay.

2. State of the Art

2.1. Brief overview of traditional Japanese Dictionaries history and their collation systems

Note: As for this section and the rest of the essay standard, transcriptions of Japanese words to Latin alphabet will be written in italics, along with the original word preceded by the initials JP.

Before we begin to outline our dictionary project proposal, we must first research and understand the main features and workflows of traditional Japanese dictionaries and their origins, so that when the time comes to go digital, we do not lose any fundamental aspects along the way. In Japanese, there are several words that can express in some way or another our western concept of ”dictionary”. The word we commonly use in modern Japanese to refer to a traditional dictionary is *jisho* (JP: 辞書), meaning literally ”word book”, along with the almost synonym *jiten* (JP: 辞典), with this last one more commonly used in compound words or dictionary names, such as *kokugo jiten* (JP: 国語辞典), meaning ”national language (Japanese) dictionary”. However, there are also a group of related homophones which share some of the meaning, such as *jisho* (JP: 字書) and *jiten* (JP: 字典), referred specifically to character or *kanji* dictionaries, which can also be expressed with the less common *jibiki* (JP: 字引). These words related to the lexical dictionary concept should not be confused with yet another homophone, *jiten* (JP: 事典), which is used in Japanese to refer to the concept of ”encyclopedia”, commonly translated into Japanese with the compound word *hyakka jiten*

(JP: 百科事典), literally meaning ”encyclopedia of many subjects of study”.

The history and origins of Japanese Dictionaries are deeply influenced by the foundations already established by Classical Chinese Lexicography. Japanese lexicography evolved as its writing system did, first by taking its references from Chinese sources and then progressively becoming an independent tradition. For instance, the earliest Japanese dictionaries were not for the Japanese language itself, but rather dictionaries of Chinese characters in their original form, sometimes annotated with their Japanese reading. As we will review in the following paragraphs, the convoluted Japanese writing system, nowadays composed of three different scripts, creates several complications for dictionary ordering. We can divide the different ordering methods used in Japanese lexicography into three main categories: the *bunruitai* (JP: 分類体), referred to semantic ordering, the *jikeibiki* (JP: 字形引き), referred to ordering by logographic means, and lastly, the *onbiki* (JP: 音引き), meaning ordering by phonetic collation. These three systems were borrowed from Chinese character dictionaries, and therefore each of them can be traced back from Chinese original works before being used in actual Japanese dictionaries.

Apart from the Chinese influence in the establishment of these three collation systems, the structure of Japanese dictionaries was also influenced by the works originated from contacts during the Nanban (JP: 南蛮) trade period. A very good example of this is the *Rakuyōshū* (JP: 落葉集), which was printed by the Jesuit Mission Press brought to Japan by the Italian vicar-general Alessandro Valignano in 1590 for the Katsusa’s society college, with one later installed in Nagasaki around the year 1600 (Loureiro, 2006, p. 143). The *Rakuyōshū*, as pointed by its compilers and unlike the other contemporary dictionaries, introduces a system in which the *kanji* compounds are arranged so that they can be looked up by both its Chinese *on* reading and its Japanese *kun* reading in the *iroha* order (Bailey, 1960b, p. 9). However, the most well-known work produced by this contact is undoubtedly the 1603 *Nippo Jisho* (JP: 日葡辞書) or *Vocabulario da Lingoa de Iapam*, a unique Japanese-Portuguese dictionary also printed in the Jesuit Press of Nagasaki. It holds a special interest to us, as it has the credit to be the first bilingual Japanese dictionary with a Western language. The *Vocabulario* contains

almost 33000 entries, enriched by examples, proverbs, idioms, synonyms, Buddhist terms, proper nouns and vulgar expressions. Because of its quality, it is believed to have been compiled with the invaluable help of the new Japanese members who joined the Society of Jesus during the early years of evangelization. It is still today a reference work for consulting Japanese vulgar and dialectal variants of the time (Rojo-Mejuto, 2019). As a side note, a Spanish translation of this work was published in Manila in 1630.

Modern Japanese dictionaries structures are a result of both the evolution presented in this overview and, in part, of the external influences assimilated during different periods. The title of the first modern Japanese dictionary is often credited to the 1981's *Genkai* (JP: 言海), a "Sea of Words", by the lexicographer Ōtsuki Fumihiko (JP: 大槻文彦). The *Genkai* itself is not free of foreign influence. Ōtsuki was an English translator himself, and, according to Nagashima Daisuke, he admitted to having modelled his *Genkai* in the form of the *Webster's* Dictionary. Ōtsuki took the *Webster's* modern dictionary concepts and translated them into his work, such as with the inclusion of idiomatic phrases and formal expressions (Takayanagi, 1971, p. 485). As previously mentioned, monolingual Japanese dictionaries are commonly referred as *kokugo jiten* (JP: 国語辞典). Although Ōtsuki did not call his work a *kokugo jiten*, it did serve as a model for many later *kokugo jiten* dictionaries. Initially, Ōtsuki described his work as simply a dictionary of "common Japanese", but later works started using the *kokugo* denomination, which, as (Yeounsuk & Hubbard, 1996, p. 62) points out, is part of a process of language "nationalization" and its establishment as a part of the new Japanese nation-state identity during the late *Meiji* era.

Nowadays, the available *kokugo jiten* are usually divided into two categories for practical reasons: small-sized or *kogata* (JP: 小型) and medium-sized or *chūgata* (JP: 中型). Starting with the latter, there are three main exponents of this kind in the current market. The most famous work of this type is the *Kōjien* (JP: 広辞苑), a "Wide Garden of Words", by publisher Iwanami Shoten, followed by the *Daijirin* (JP: 大辞林), a "Great Forest of Words", by publisher Sanseidō, and lastly the *Daijisen* (JP: 大辞泉), a "Great Fountain of Words", by Shōgakukan. Dictionary reviewer and book author Nagasawa (2019) gives us a succinct

overview of the differences between each other. First, *chūgata* sized dictionaries have in common the inclusion of ancient vocabulary and abundant examples, along with lists of proper nouns and technical terms, which are often absent from *kogata* formats. This ”encyclopedic” component is not addressed in the same way in the three dictionaries though. For example, the *Daijirin* emphasizes the inclusion of modern language words and expressions. It is also the only one of the three that includes accents in the words. The *Daijisen*, on the other hand, is more focused on its frequently updated electronic version, and it is the only one of the three to contain illustrations. On the entries level, the *Kōjien* is the most differentiated of the three, as its definitions are ordered from the meaning closest to the etymology, while the *Daijisen* and *Daijirin* order them starting by its more modern meaning. In contrast, *kogata* sized dictionaries are designed for a more practical and daily use, usually intended for high school students. This kind of dictionary contains around 70000 entries, therefore losing the ”encyclopedic” aspect of *chūgata* sized dictionaries, which usually surpass the 150000-200000 entries. However, such a format makes the market for this type of dictionary much more competitive and filled with options. The most popular one, at the time of drafting this report and according to the own publisher data (SANSEIDO Co., Ltd. Publishers, 2020), is the *Shin Meikai Kokugo Jiten* (JP: 新明解国語辞典). This dictionary contains around 77500 entries, with a special emphasis on word use case explanations and verb conjugations. Another prominent example from this market segment is the *Sanseido Kokugo Jiten* (JP: 三省堂国語辞典), by the same publisher. It contains around 82000 entries and is especially prolific with modern language. New commonly used words or new meanings for pre-existing entries are usually swiftly recorded in this dictionary. On the other hand, it does not include words with no modern usage examples. Many *kogata* sized dictionaries can be set apart by the inclusion or not of new modern entries, which is a common problem because some of these works still cling to their 20th-century edition (Nagasawa, 2017).

Characters or *kanji* dictionaries are still an important part of the modern Japanese dictionary market, rooting back its origins to the early Chinese character dictionaries already mentioned at the start of this overview. Undoubtedly, the most recognized work of this kind is the *Dai*

Kan-Wa Jiten (JP: 大漢和辞典), a monumental 13-volume dictionary edited by the sinologist and lexicographer Morohashi Tetsuji (JP: 諸橋轍次) starting from 1955. It contains 49964 entries for characters, which are linked in turn with over 370000 *kanji* compound words. Apart from the main entries, there are also multiple proper noun examples, Buddhist terms and poetry fragments. Needless to say, this is one of the most encyclopedic works in the genre, and much more abridged dictionaries can be easily found in the market, with less than 10000 or 5000 entries, usually centred in modern usage.

Apart from these main exponents, a wide variety of specialized dictionaries are also available nowadays. Under this category, we can mainly encounter dictionaries focused on technical terms, such as science, engineering or economics. More on the humanities side, we have for instance pre-modern Japanese dictionaries or *Kogo jiten* (JP: 古語辞典), collocations dictionaries, loanwords or *gairaigo* dictionaries and thesauruses, known in Japanese as *Ruigo jiten* (JP: 類語辞典) or with its *katakana* transcription *shisōrasu* (JP: シソーラス).

2.2. Electronic and Digital Japanese Dictionaries

Although traditional paperback dictionary sales are still prevalent in the Japanese market, it is clear that nowadays most dictionary lookups are done first on digital devices. In the beginning, electronic Japanese dictionaries started as an extension of traditional major dictionaries, in the form of CD-ROMs included inside the paper edition. These contain a desktop computer program with the content already present in the paper edition plus some new entries exclusive to the digital version. In addition, such programs usually come with a guaranteed update period during which any new word entries will be automatically added to the user's license. However, a much more common way to use a dictionary digital license is through the so-called *denshi jisho* (JP: 電子辞書). These devices, which are almost entirely exclusive to the East Asian market, are particularly popular with high school and university students. Their form factor often consists of a clamshell-like body which opens to reveal a small keyboard and a screen. This LCD panel can be monochrome or in full colour and

touch compatible, in which case a stylus is often included. Professor Smith (2005) gives us a succinct review of the current Japanese *denshi jisho* market, which is mainly dominated by four brands: Canon, Casio, Sharp and Seiko. However, since the creation of this report, Seiko has withdrawn from the market. Most of these devices work with the same business model: each brand signs license deals with famous publishers, allowing them to include the original entries from recognized dictionaries on their devices. This way, a user can use his *denshi jisho* to look up a word and obtain entries from multiple dictionaries at once, which is a function almost exclusive to these devices. For example, most of them include the complete *Kōjien* or the *Daijirin* as main *kokugo jiten* for reference. The second most important factor usually is the Japanese-English dictionary licenses, with titles such as the *Kenkyūsha New Japanese-English Dictionary*. The most modern models include some innovations, such as Text-to-Speech, handwriting recognition and language-learning programs. However, what often makes a user choose one brand or another is the exclusivity of the licence for a particular dictionary, especially with technical terms ones. Currently, the market leader is the brand Casio with 55.8% of the market, followed by Sharp with 29.9% (data obtained from the IT market agency BCN (2022)).

However, the market share for these devices has been shrinking gradually during the last decade due to the popularization of smartphones and mobile applications, which are the focus of this essay. In fact, the term *denshi jisho* is more and more used nowadays to refer to other types of digital software, such as online web dictionaries and dictionary applications. Following this trend, dictionary publishers have started to release their own mobile applications for the Android[1] and iOS[2] operating systems. These applications are not developed by the publishers themselves, but by a handful of individual software companies that have made this niche market their business. Their names are *Monokakidou* (JP: 物書堂), *LogoVista* and *BIGLOBE Inc*. Of these three, *Monokakidou* has the software most recognized for its

[1] The Android operating system is developed by Google Inc. with an open source core. Nowadays is present in thousands of devices.

[2] The iOS operating system is developed by Apple Inc. Currently, is a closed environment only present in the mobile devices created by Apple itself. Nevertheless, applications targeted to iOS can also be compatible with Apple's desktop devices.

Figure 2.1: An example *denshi jisho* device. This is specifically a *Casio EX-word XD-A6600BK* model. We can clearly see on the screen how this particular model includes a license for the *Kōjien* as main dictionary. Image retrieved from Casio (2010).

quality, almost exclusively developed for the iOS environment. All three, however, most of the time do not develop software exclusively for each product or publisher, but create what could be defined as ”dictionary viewers”. Their business model is similar to that used by the manufacturers of physical *denshi jisho* devices discussed above. They first develop a common application, a dictionary ”aggregator” that already implements all the basic search and input features that are expected from a user. These are usually also inherited from the ones used in traditional *denshi jisho* devices, such as the ability to select a word within an example and jump to its own definition without having to input it. After this, the user can then purchase different dictionary modules from within the same application. *LogoVista* and *BIGLOBE* make individual applications available for each of their dictionary licences, but they all use the company’s same content viewer. Following the same trend as Japanese web

content, these dictionary viewers are overly textual, losing the chance to model a user interface around a concrete dictionary. On the other hand, this helps to maintain uniformity between dictionaries when the user wants to swiftly change between two of them. Because of this, for instance, *Monokakidou*'s viewer interface is regarded as very clean. In the case of touch devices, some of the viewers also include a handwriting recognition engine, so that users can directly write their query using their finger or a stylus, character by character. As for secondary features, most of them allow the user to save the entries as bookmarks for later. Text styling settings are also common, such as font size and colour.

The dictionary licenses inside these applications have a wide range of prices. They are expected to contain exactly the same entries as the paper edition, but with a small discount compared to it due to savings in printing costs. Despite this, prices for popular common dictionaries range between 1900 to 4800 yen, with technical dictionaries easily surpassing the 15000 yen barrier.

Apart from these applications, which are natively created for mobile devices, there are also a smaller number of Japanese web services that provide similar functions. These websites also work as dictionary aggregators, but this time usually without payment required. They use text databases provided by publishers as the data source. Another advantage of these sites is that they are inherently multiplatform, which means they can be accessed from any device with an internet connection. When a term is looked up on these websites, a batch search is performed, retrieving the available references to that term within their database. These usually include its entry from a particular *kokugo jiten* and from an encyclopedia. Two well-known Japanese websites of this kind are *Goo Dictionary*[3] and *Kotobank*[4].

2.3. Japanese-English Learner's Dictionary Applications

As one can imagine, the compilation of a truly bilingual and bidirectional English-Japanese dictionary is a challenge difficult to achieve. English and Japanese are languages very far

[3] https://dictionary.goo.ne.jp/
[4] https://kotobank.jp/

Figure 2.2: Comparison between the three main dictionary viewers applications and their UI. **Left**: *Monokakidou*'s viewer interface with the *Sanseido Kokugo Jiten* on screen. **Center**: *LogoVista*'s viewer using the *Shin Meikai Kokugo Jiten*. **Right**: *BIGLOBE*'s viewer showing the *Daijirin* dictionary. Images respectively retrieved from 西練馬 (2019), LogoVista (2022) and BIGLOBE (2020)

apart both linguistically and culturally. For this reason, Japanese-English bilingual dictionaries are usually designed for learners of one of the two sides only. English-Japanese dictionaries intended for Japanese-speaking learners are usually referred to as *ei-wa jisho* (JP: 英和辞書) or *wa-ei jisho* (JP: 和英辞書). On the other hand, dictionaries intended for English-speaking learners are usually called *nichi-ei jisho* (JP: 日英辞書), intentionally changing the *kanji* used to refer to Japan (Nakao, 1989, p. 2). The history and evolution of the first ones have been much more extensive and prolific, due to the obvious necessity of millions of Japanese students to have a solid word reference during their compulsory English lessons. The history of these formally started during the late 1800s, when Dutch trade influence started to decrease in Japan in favour of the new American pressures. In fact, the first English-Japanese dictionary, the *Eiwa Taiyaku Shuchin Jisho* (JP: 英和対訳袖珍辞書) or *A Pocket Dictionary of the English and Japanese Language*, was crafted in 1862 by several Dutch-Japanese interpreters

using an English-Dutch dictionary and some Japanese-Dutch ones they already owned. The evolution of the later ones, however, is much more modest. These dictionaries also have an early origin. The 1867's *A Japanese and English dictionary: with an English and Japanese index* by the American physician and missionary James Curtis Hepburn is considered the first of its kind. Hepburn is more widely remembered nowadays for the *Hepburn romanization*, the most popular transcription system to Latin script used in the Japanese language. Hepburn popularized its use by including it in his dictionary. Unfortunately, dictionaries for Japanese language learners have had a much slower evolution from their first conception. Professor Humblé (n.d.) gives us a brief review of 2 of the few paper dictionaries available for learners of Japanese: the *Kenkyusha Japanese-English Learner's Dictionary*, first published in 1996 with 11000 headwords, and the *Basic Japanese-English Dictionary of the Japan Foundation*, first published in 1986 with 2873 entries. For the learner's sake, these two dictionaries have their headwords written in both *romaji* and normal Japanese, along with *furigana* transcriptions. Still, they suffer from the wide differences between the two languages. Many examples present in its entries are perfectly translated into English, with the same words a native would use to sound natural. However, the lack of more "literal" translations does not help learners, as one needs to distinguish the different elements of an example in order to truly understand it. Another important issue for learners is how these dictionaries deal with functional vocabulary and grammar structures. In this sense, one more recent work designed to tackle with this issue is the *A Handbook of Japanese Grammar Patterns for Teachers and Learners*, published by Kurosio in 1998. This is a popular sentence pattern dictionary containing more than 3000 entries, each one with diverse examples created by experienced Japanese teachers. One prominent achievement of this work is that it successfully collates thousands of grammar patterns containing multiple words, by extensive cross-referencing between them and the use of different index systems (Sunakawa, 2017). The advent of the new digital platforms is helping to iron out the disadvantages of traditional paper dictionaries, providing students with new tools for optimized Japanese language learning. Of these tools, smartphone dictionary applications are arguably one of the cornerstones for Japanese students, combining

traditional dictionary headwords with grammar pattern examples and practice tests.

The applications presented until now in the previous section were created primarily for the Japanese public, with a similar form factor and visuals. However, the applications created for foreign Japanese learners are very different. Although Japanese-English dictionaries modules can be purchased for the applications mentioned in the last section, they share the same interface and search engine used for Japanese-Japanese dictionaries. On the contrary, English-Japanese dictionary applications are developed with the learner in mind, with a unique interface tailored for each application and some additional features to help the user through his learning process. Unlike Japanese dictionary viewers, these applications are usually free of charge. Since our project falls into the same category, we will take a closer look into some of these applications and their features. In this case, we have selected three of the most downloaded and rated English-Japanese dictionaries on the Google Play Store[5] platform, as Android is the primary target for our application. These apps are: *Japanese Dictionary Takoboto*, *Jsho - Japanese Dictionary* and *Yomiwa - Japanese Dictionary and OCR*. These are all applications that have been fully tested by thousands of users. Some of them have more unique features, but they all share a few common dictionary aspects. This review will help to determine the core features expected of this type of software, so that they can be taken into account during the design of our application.

Fig 2.3: Takoboto's Logo

The first on the list is ***Japanese Dictionary Takoboto***. This application, which also has a website[6] equivalent, is currently the first result returned when searching the terms "Japanese dictionary" on Google Play Store. According to Google itself, these results are ordered by the relevance, user interactions and quality of each of the items. The application itself could be summarized as a "straight to the point" dictionary. The design is clean and simple, yet plenty of content. The search engine recognizes at the same time *kanji*, *kana*, *romaji* and alphabet inputs in the two languages.

[5] The Google Play Store is the main authorized platform for uploading and downloading new mobile applications for the Android operating system.

[6] https://takoboto.jp/

Full sentences and conjugations are broken down and searched in the same order in which they appear in the sentence. *Kanji* can also be searched using an on-screen radical keyboard. As we can see in figure 2.4, each entry contains three separated sections distributed in tabs: the main word definition, the *kanji* of the word and a handful of example phrases. Also, when a verb entry is selected, an additional conjugations section becomes available. The *kanji* section offers some preview information about each character, such as its readings, JLPT exam level and stroke order. There are no individual entries for each *kanji* in Takoboto, which means *kanji* compounds are only indexed and can only be searched through the main search function. Each word entry has also some additional features, such as text-to-speech reading, adding to favourites or to a list. The lists feature allows the user to create personal word lists for any purpose. There are also some simple official lists already included inside the app, such as word and *kanji* list by their JLPT exam level. In addition, Takoboto offers a paid subscription plan with which you can search online lists created by other users, mainly about semantic categories and technical fields. The subscription also unlocks lists synchronization between multiple devices and cloud saving. As for secondary features of the standard version, Takoboto also offers a grammar section with content contributed by the application community. Each grammar pattern entry has a brief meaning explanation and some example sentences. Users can leave comments with additional remarks about the entry. The application settings include some options such as font styling, background colour, *romaji* mode instead of *kana* and showing pitch accent when available.

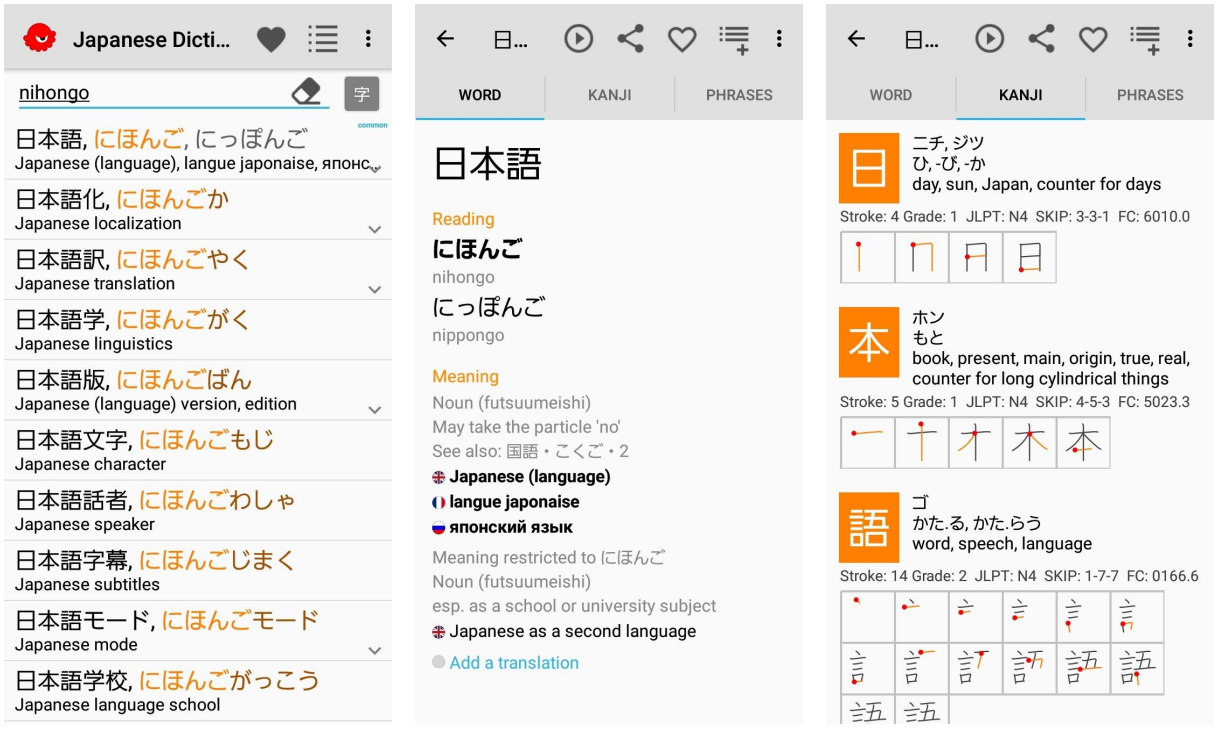


Figure 2.4: Left: Takoboto's search engine showing the results for "nihongo" written in *romaji*. **Center**: Entry example for 日本語, we can see Takoboto offers extra translations in a few languages apart from English. **Right**: *Kanji* section tab inside the entry for 日本語. Images retrieved from Takoboto (2022)

Fig 2.5: Jsho's Logo

The second on the list is ***Jsho - Japanese Dictionary***. This application was created by the mobile developer Richard L, who has published other dictionary applications with a similar format. It was updated for the last time in 2019. Although it is arguably an old application, its simple format keeps attracting thousands of users even today. Its design is simpler and more straightforward than Takoboto's, and it does not implement as many features, but it has some advantages of its own. Unlike in Takoboto, the search engine in Jsho requires you to select the language you are writing the query in. If Japanese is selected and you are using an alphabet keyboard, syllables will be converted into *kana* as you write. A *kanji* radical keyboard is also available. As we can see in figure 2.6, Jsho does not implement a word entry page itself, as selecting a word from the search results takes you directly into a *kanji* details page. Therefore, the word meaning is only visible from the search results page. If the word has multiple inflection forms, such as in verbs or adjectives, an additional conjugation section is displayed. The *kanji* details page also offers shortcut buttons to automatically search more words containing the *kanji*, instead of including them directly on the page. The *kanji* stroke order is also available if the character image is touched, almost as a hidden feature. Each word and *kanji* entry can be saved separately as bookmarks for later. The application settings include just a few options, such as font styling, background colour, showing pitch accent and an interesting filter to limit dictionary results only to common words or *kanji*.

Fig 2.7: Yomiwa's Logo

The last from our list is ***Yomiwa - Japanese Dictionary and OCR***. This application is the most complete of the three, being what could be defined as a "complete learning tool" instead of just a dictionary. Yomiwa was developed by Nomad AI, a group founded by former PhD students in Artificial Intelligence at Kyoto University. Their applications implement machine learning algorithms and natural language processing to improve their functionalities. In this case, these technologies are applied to some of Yomiwa's advanced features, such as OCR, handwriting recognition and

Figure 2.6: Left: Jsho's search results for " はなす", we can see the meanings are directly displayed on the list. **Center**: Conjugation section for the verb to use, *tsukau* (JP: 使う). **Right**: *Kanji* entries for " 辞書", we can see the shortcut buttons for words containing the character. Images retrieved from Richard L. (2019)

a text analyzer. The OCR feature, which stands for Optical Character Recognition, allows the user to detect text in any picture from your device storage or camera, with a high success rate. This feature works offline and the detected characters or words can be instantly looked up in the dictionary. The handwritten recognition feature is offered as another input method for the dictionary, along with the traditional and radical keyboard. This recognition allows the user to draw easily any *kanji* in case he does not know its reading. The text analyzer is, in technical terms, a Japanese text parser. This function involves splitting the parts of any sentence pasted on the analyzer, allowing the learner to distinguish between words more easily. The parsed text also contains *furigana* for each kanji and dictionary meanings when tapping a word. This parser can also be used within a small web browser included in the application, allowing the user to parse text contained in any Japanese website. Of course, Yomiwa also contains a standard Japanese-English dictionary much like the ones we have

already reviewed in the last two applications. One of the main differences in this aspect is that word entries include meanings from more than one database, showing them one after another. Interestingly enough, Yomiwa asks you to download additional offline databases if you want to look up entry examples or search for proper nouns. There are also individual *kanji* entries with the usual information, including compounds for each *kanji*. In tandem with the dictionary, Yomiwa includes two other common functions: lists and flashcards. Words can be added to both functions from the dictionary entries. In the case of lists, there are several free ones already included in the application, from the most diverse fields. The last of this plethora of features is called ”Wall”. Here, users can post small requests with pictures related to the application or the Japanese language in general. Other users can accordingly answer these posts in the form of comments. This effectively makes Yomiwa a small ”social network for learning Japanese”. Some of the features mentioned in this review require a subscription fee to be paid in order to use them after a weekly limit, notably OCR recognition and the web browser parser. Lastly, the basic settings offered are similar to the other applications, with some additional ones added for Yomiwa’s special features.

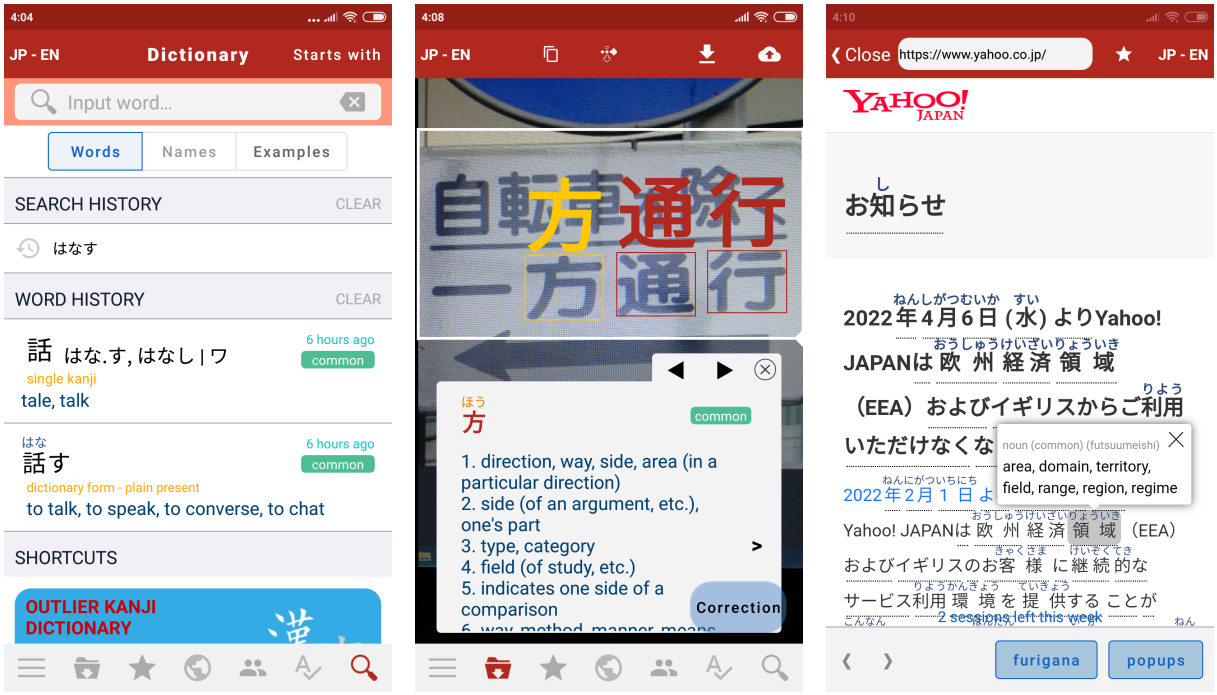


Figure 2.8: Left: Yomiwa's main dictionary interface. **Center**: OCR engine example. The recognition algorithm has detected the characters on the picture, showing the meaning of the selected one. **Right**: Example of Yomiwa's text parser over a website, displayed within the application itself. Images retrieved from NomadAI (2022)

All in all, we can conclude Japanese-English learners' dictionary applications can vary considerably in terms of the features included in each of them. However, core dictionary features are significantly similar between the applications, using almost the same navigation workflow in all of them. The user typically looks up a word and retrieves the meaning instantly, sometimes even without the need to access the full word entry. Independent *kanji* entries are not implemented in all of them, but the information they include can be obtained most of the times from other parts of the interface. As a side note, every application of the list offers its basic dictionary data without an internet connection, which is still a feature very demanded by users. "Learning" features, on the other hand, are much more diverse between applications, with traditional word lists and flashcards as the only common denominator.

3. Proposal

3.1. Linguistic perspective

Vocabulary plays a crucial role in the performance of L2 Japanese language learners. Once the essential grammatical patterns have been studied in class and correctly internalised, their effective use depends largely on the learner's vocabulary skills. For English-speaking students this is not a trivial task, as the Japanese language does not share any phylogenetic origin with English. On the other hand, for instance, Chinese students take advantage of the Japanese relations with their mother tongue and follow a learning process much more guided by intuition. It is clear that many of the differences in the rate of Japanese vocabulary acquisition among students can be attributed to the amount of exposure they receive to the language. In Dewey (2008), for instance, three groups of Japanese language learners are compared: regular students in their home country, exchange students in Japan and intensive program students outside Japan. Naturally, the last two were exposed to the language for a longer period of time, ergo their vocabulary skills grew accordingly. However, intensive program students showed a more internalised command of complex vocabulary on certain occasions, as they had access to more specific learning materials than exchange students, who learned their vocabulary intuitively mostly through social interactions. Some popular textbooks used in Japanese language classes, such as *Marugoto: Japanese Language and Culture* (The Japan Foundation 2013-2017, published by Sanshusha Ltd.) and *Minna no Nihongo* (1998-2016, published by 3A Corporation), introduce the vocabulary used in each unit regardless of the level of the *kanji* they contain. It is therefore important to determine just how optimal are

the average vocabulary materials and what is the correlation between the rate of Japanese vocabulary acquisition and the study materials used for this purpose. Most of the Japanese vocabulary is made up of *kanji* compounds, meaning you have to recognize the characters which made up the words to properly read them at least. Luckily, individual *kanji* learning has been a regular focus of research into acquisition techniques and we can use them as a reference point for our project. In fact, among the results yielded by research databases when searching for Japanese ”vocabulary learning” or ”lexical development”, over 80% are related with *kanji* learning (Mori et al., 2020, p. 3). One of the most popular trends in recent decades has been the use of mnemonic techniques by attributing an intuitive meaning to the components that made up each *kanji*. *Kanji* components should not be confused with radicals. There are exactly 214 radicals, and each *kanji* has only one radical assigned, even when there is more than one radical contained inside a *kanji* structure. This is because radicals were conceived as an index for Japanese dictionaries. The rest of the components of a *kanji* can have multiple origins, such as pictographs representing its meaning (like in horse ” 馬” or fire ” 火”), which are the focus of mnemonic techniques. These techniques can be useful at the start of the learning process, but, according to Rose (2013), surveys show that mnemonic strategies make associations in most cases only with the meaning of the *kanji*, causing an inability to read them in real situations. They become less helpful when associations are convoluted or when the *kanji* is rarely used in isolation from other kanji. This is the case for many advanced *kanji*, and an issue we want to approach with our project. Other studies, such as Yamashita & Maru (2000), have a more granular conception of the types of components that can make up a *kanji*, and the effectiveness of mnemonic techniques when applied to each of them. The results show that pictographs were perceived as easier to learn, followed by *katakana* composites. *Kanji* that are made up of *katakana* components were perceived as such because beginner students typically learn the *katakana* syllabary before any *kanji* character. On the other hand, components that refer to phonetics of the *kanji* were perceived as much more difficult by the students, because most of these readings are used in vocabulary compounds that are not the focus of teaching at this level. This ties in directly with the main claim of the study, which

is that any technique proposed only has significant results when the concepts are relatable to the existing cognitive structure of the learners. One of the future directions proposed is especially relevant to us, which is that acquisition of compound words must be investigated in order to obtain a more complete picture of *kanji* acquisition (Yamashita & Maru, 2000, p. 171). This is precisely the starting point from where our application concept is based. We want to propose an indexing system for our learner's dictionary in which every word entry will be directly linked to the words containing at least one of the *kanji* that compounds the first word. This intends to be a middle way through which students can learn the vocabulary they purposely looked up, but at the same time create a correlation between related words by their common *kanji*. This way, the student gets to recognize and internalize each *kanji* as a unit that is part of the vocabulary he or she already knows, but more importantly, the student gets to consciously learn both the meaning and the reading of each character involved, unlike with mnemonic techniques.

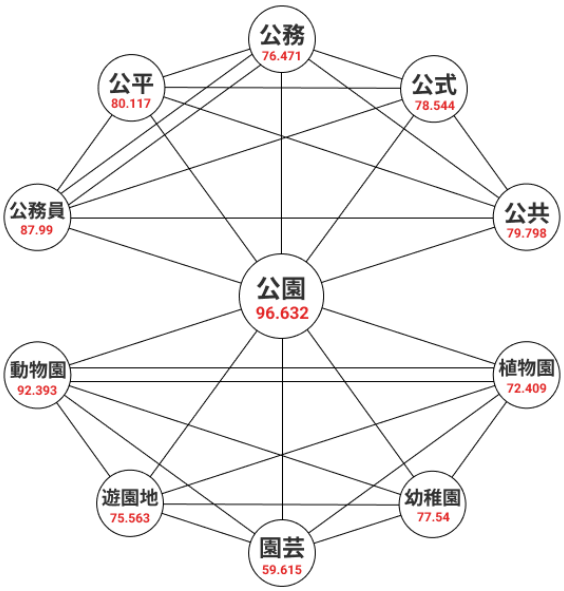
We have used the structures present in graph theory to implement this indexation system. Graph theory is a branch of mathematics, particularly important for discrete mathematics, dedicated to the study of graph structures used to model pairwise relations between concepts or objects. When describing these graphs, we refer to the objects depicted as *vertices* or *nodes*, and to the connections between them as *edges* or *links*. These *edges* can be undirected, meaning they link vertices symmetrically, or directed, meaning they link vertices asymmetrically with an orientation. Additionally, edges can have an associated numerical value, called a *weight*. Weighted graphs usually use non-negative integer weights to represent the "cost" of each connection. Graph theory has several fields of applications, where it helps to model the relations between elements and to represent problems. In these fields, graphs usually have some attributes (like names or types) associated with their vertices or edges, in which case they are referred to as *networks*. Graph theory and network science have proven to be particularly useful in linguistics. Linguistic structures like the one we will be implementing are usually discrete in nature. For example, standard semantic networks represent the relations between words in semantic triples (subject-predicate-object), in which subjects and objects

are represented as graph vertices, with the edges as predicates. These kinds of networks are fairly important for computational linguistics, with large projects publicly available such as Wordnet[1] (Miller, 1995). Social sciences is another field where graph theory has proven effective to analyze particular phenomena, such as social network relations and topic spreading. For our case, we will first generate an undirected graph where each node attribute represents a word entry from our dictionary. The edges will connect words that share at least one of the *kanji* with each other. As two words can have more than one *kanji* in common, so can the edges between the nodes of our graph. This makes our graph also a multigraph, meaning it allows the presence of multiple edges that have the same end nodes. In order to visually explain our model, we have created an example graph containing only a limited set of word nodes which share common *kanji* with the word 公園, meaning public park. We can see it on the left side of figure 3.1. To save page space, these nodes do not comprise all the words containing said *kanji*, just a representative subset of 10 of them. Apart from this, every aspect of our graph is properly represented here, within the limits we just stated. Inside our real database, the entirety of the dictionary headwords will be included inside the graph, but we will use this example for illustrative purposes. Here, the multigraph aspect of our network can clearly be seen on two pairs of nodes, 公務員-公務 and 動物園-植物園, which both share more than one *kanji* and therefore are connected by multiple edges. Regarding the overall topological structure of our graph, we can already see some patterns. For instance, we can deduce that the graph complexity will fluctuate depending on the frequency of the *kanji* making up each word. More popular *kanji* means more edge connections for a node containing it, ergo more complexity. This property of nodes is known as the vertex degree, and has already been used as a complexity measure for graphs in the linguistic field (Piperski, 2014, p. 93). For instance, the average vertex degree of our example graph, which is only a fragment of the real database, is around 5.45 ($2 \times$ number of edges / number of nodes). Another common graph pattern is the so-called ”small-world” property. Small-world graphs are defined by a high clustering coefficient and small average shortest path length. This, in

[1] Available through Princeton University portal: https://wordnet.princeton.edu/

simple terms, means that the nodes of such a graph can be reached from any other node in a relatively low number of steps, without most of them being neighbours to each other. This pattern has already been found in past studies related to the Japanese language. For instance, graphs composed of Japanese *kanji* connected through its components have been found to follow small word features. It is believed that this behaviour may be a result of the successive elimination of components during the evolution of the logographic writing system, extensive also to Chinese characters (Jeronimus et al., 2017). Coincidentally, small-worldness has also already been found in networks similar to ours, but limited to two-*kanji* compounds and with a structure slightly different, with nodes representing individual *kanji* and edges representing compounds (Yamamoto & Yamazaki, 2009). It is therefore logical to expect our complete graph model to have similar features once the complete database is generated. For instance, our example graph already complies with the properties expected from small-world networks. This represents a great advantage for learners using our model, as it means that our dictionary can be quickly traversed to and from any node that is needed. Students can reach unrelated terms through a small number of steps and use the in-between *kanji* connections to consciously remember new vocabulary.

Looking up a word in our dictionary, in general terms, would consist of taking a shortcut to the node representing that word, and from that point let the user continue exploring the graph. However, our objective is not to merely show the user every headword connected to its looked-up term, but to prioritize the most useful words from a learner's perspective. It is also important to note that we can not efficiently show the complete dictionary graph on screen to the user due to the impracticality of its size. For these reasons, we have elaborated a scoring system divided into 2 distinct phases. The first phase consists of the inclusion of individual scores for each node, whose values are determined by the external properties of each word, completely independent of the network structure itself. These individual scores are calculated taking into account a number of relevant parameters, each with a percentage of the score assigned according to its importance. 50% of the score is given to its frequency rank inside a Japanese corpus from multiple sources. 30% is allocated to the frequency of



Lemma	Individual Score	Average Connected Score	Total Cumulative Score
動物園	92.39	19.088	116.481
公務員	87.99	20.578	113.568
公平	80.12	20.971	106.089
公共	79.80	20.972	105.770
公式	78.54	21.050	104.594
公務	76.47	21.154	102.625
幼稚園	77.54	19.831	102.371
遊園地	75.56	19.929	100.492
植物園	72.41	20.087	97.177
園芸	59.62	20.727	85.342

Figure 3.1: An example graph following our model, limited to a set of words containing common *kanji* with the word 公園, which is the search term in this case.

each of the *kanji* that make up the word, of which 25% is divided between the frequency rank inside a corpus of each of the characters and the remaining 5% is awarded when all the characters are *hiragana*, *katakana* or belong to the official list of 2136 *jōyō kanji* (JP: 常用漢字), meaning commonly used *kanji*, maintained by the Japanese Ministry of Education. The remaining 20% is achieved depending on the JLPT exam level to which the word belongs, with the lower the level to which the word belongs, the higher the percentage. The aim is to encourage the discovery of words belonging to the more basic curricular levels before those of more advanced levels. When all these parameters are added together, we obtain a theoretical maximum individual score of 100 points. For instance, we can see the individual scores written in red in each of the nodes of the example graph in figure 3.1 and in the second column of the table in the same figure. All of the data sources required for the implementation of these parameters, such as corpus, will be specified in the next section. There are also many

occasions when headwords can be written in *kanji*, but are usually written in *kana*. In such cases, the frequency of the *kanji* will still be taken into account for the score, but they will preferably be displayed in *kana* on the interface. In cases where a word can really only be written in *kana*, as is the case with particles, loanwords and onomatopoeia, the frequency percentage will be adjusted to 75% to replace the frequency of the individual *kanji*.

Naturally, these scores only represent the value of the word for the learner by itself. The second phase of the scoring system takes into account the knowledge provided by our graph and the value added to each word by its neighbours. In other words, a headword is most useful to the learner when he or she can use its characters to learn other words which, in turn, are also relatively useful and frequent. To reflect this, an additional 30% is added to the individual score of each word, of which 25% corresponds to the average individual score of the neighbouring nodes of that word. We have called this new property ”average connected score”, and can be found in the table of figure 3.1. The remaining 5% depends on the number of neighbouring nodes that share the same reading for the *kanji* by which they are connected to the searched word. This brings the total maximum score to 130 points. In this way, a word that by itself does not have a higher than average individual score can stand out with additional points if it is connected to several particularly useful words. Following this scoring system, we can see the total cumulative scores in the table of figure 3.1. The most useful word other than the search term would be 動物園, helped by a particularly high frequency rate. Nevertheless, these scores are not definitive, as the example does not include all the words that share *kanji* with the ones shown, which could slightly affect the average connected score.

In summary, this scoring system will help us determine which headwords should be preferential from a learner’s perspective, and therefore which ones should be prioritized inside the application interface.

3.2. Technical perspective

One of the most important factors in digital dictionary development is the suitability of the data for the required purpose. For this reason, we have selected data from sources that have already been tested by some software projects of diverse characteristics related to the Japanese language. The most important of these is the **JMdict** database, used for the main dictionary content of our application. The Japanese-Multilingual Dictionary (JMdict) database was created by professor Jim Breen and is currently managed by the Electronic Dictionary Research and Development Group (EDRDG). The creation of this database dates back to 1991 with the release of the first version of the Electronic DICTionary (EDICT) project, as part of an early Japanese word-processor for the DOS operating system (Breen, 2000). The goal was to create a Japanese-English dictionary file that could be used both as a traditional dictionary and as an assistant for reading Japanese text. By 1999 the EDICT file contained over 60000 entries and a decision was made to convert its format to the more standardized XML format, due to the limitations that the EDICT file had with some of the complexities of the language (Breen, 2004). This is how the current JMdict file was born. The EDICT format was expanded in 2003 and is still maintained by the EDRDG, as part of the support for older software released before the creation of the JMdict. Nowadays JMdict and EDICT share the same data source, with over 170,000 entries, most of them contributed by volunteers over the years. Contributions have slowed down in recent years as the file already has covered a large proportion of the Japanese lexicon. However, the JMdict format also supports the inclusion of translations to other languages apart from English, which keeps growing year after year, though still far from the English version. Each entry of the JMdict file format contains the headword itself, its reading, its variants (such as alternative *kanji* for the headword), part of the speech (such as noun, verb ...), the field of application if specified (such as ”mathematics” or ”engineering”), miscellaneous markings (such as ”usually written in *kana*”), the translations to English (other languages are labelled separately) and some simple frequency markings from a handful of user-contributed corpora. Dialects are also marked when necessary, and loanwords have labels to indicate the source language. All the content from the JMdict file was made available

to the public in 2003 by the EDRDG, for both commercial and non-commercial purposes, subject to ensuring appropriate acknowledgement. One of the most famous projects based on JMdict is the Japanese-English dictionary website Jisho[2], which uses this database as the backbone of its operation. As a side note, the EDRDG also manages addition databases with similar formats for Japanese proper nouns and individual *kanji*.

Although JMdict contains simple frequency data, it only divides the headwords into two levels, effectively ”common” or not. As we need much more granularity in this type of data, we have used a different source for our headword and *kanji* score frequency. For the frequency of complete headwords, we will use the Vocabulary Database for Reading Japanese (VDRJ), compiled by professor Matsushita Tatsuhiko (JP: 松下達彦). This database offers a word ranking list method that removes terms with a biased usage range from the range of high-frequency vocabulary by taking into account the degree of dispersion. The database is extensively labelled with the frequency of each word in different genres and has list variations for international students and teachers (Matsushita, 2021). It is freely available for download from his laboratory website[3]. In turn, the corpus used for the creation of VDRJ was the so-called Balanced Corpus of Contemporary Written Japanese (BCCWJ), compiled by the Center for Language Resource Development from the National Institute for Japanese Language and Linguistics (NINJAL). The BCCWJ aims to solve the lack of balanced corpora from modern Japanese texts by compiling roughly 104.3 million words from a diverse range of sources. These include newspapers, magazines, textbooks, business reports, web message boards, legal documents, books and Japanese Diet minutes. All these contents span through a 30-year period (1976 - 2006), and are divided into sub-corpora that represent the whole body of published works and the share of them that had a sizable impact on readers (Maekawa, 2007). The corpora are also divided according to the genres specified by the Nippon Decimal Classification (NDC).

Sample sentences from our application headword entries will be retrieved from the Tatoeba Project, an open-source database that contains translated sentences from hundreds of lan-

[2] https://jisho.org/
[3] http://www17408ui.sakura.ne.jp/tatsum/database.html#vdrj

guages, freely available from their official website[4]. It was originally created by developer Trang Ho in 2006 under the name ”multilangdict”. Almost all the contents of the database are contributed by voluntary users, however, most of the Japanese-English pair sentences come from the so-called Tanaka Corpus. This corpus was initially compiled by professor Yasuhito Tanaka and his students from Hyogo University. The sentences were taken from student textbooks, popular books and song lyrics. Students were tasked with collecting 300 data items each, which after 4 years of efforts amounted to more than 212,000 sentence pairs (Tanaka, 2001). Professor Tanaka released the corpus to the public domain after its announcement. The original corpus file contained a large number of spelling and transcription errors, most of which have been corrected over the years by members of the EDRDG. The Japanese section of the Tatoeba Project is also actively maintained by members of the EDRDG, to keep a proper indexing with the entries of the JMdict.

Last but not least, our individual *kanji* frequency data will be retrieved from the open-source *Kanji* Usage Frequency project, created by Russian developer Dmitry Shpika. This tool generates an automatic *kanji* frequency ranking based on the corpus you previously selected. There are 4 corpora available for testing: 51.5 million *kanji* from books extracted from the famous Aozora Bunko website[5], 10.3 million *kanji* from articles published between 2014 and 2015 in newspapers (Yomiuri, Mainichi, Asahi and Saga Shimbun), 10.0 million from Twitter posts written in June 2015 and 784.6 millions from the Japanese Wikipedia in May 2015. The tool can be easily tested using its website[6] simple interface.

All these databases will be stored in the local file of our application, in order to offer the complete dictionary functionality without the need for an internet connection, which, as we saw in the previous chapter, is a feature highly demanded by users.

[4] https://tatoeba.org/ja
[5] https://www.aozora.gr.jp/
[6] https://scriptin.github.io/kanji-frequency/

4. Design of our application

4.1. Drafts and Mockups

The design of an application always starts with basic prototyping. As expected, the main design problem we faced when making the first mockups of our application was the graph interface. This is the interface that will be responsible for displaying the headwords graph to the user at all times. Therefore, we needed to ensure that it could be used intuitively with as little prior knowledge as possible. The graph of our dictionary will be a large structure made up of thousands of nodes and connections that are very difficult to manage visually. Moreover, navigating through it becomes even more difficult if it is displayed on smaller screens, such as those of mobile devices. We quickly opted to subdivide the network structure into small subgroups, displaying only eight of the nodes that were connected to the term entered in the search bar. This solution allows us to design a much cleaner and easier to manage interface. Instead of sharing these 8 node slots among the multiple *kanji* that the word may contain, we have chosen to display at once only the 8 most important ones that contain one of the *kanji* of the searched word. As we have already mentioned in the previous chapter, these 8 words will be selected by our scoring system, which focus on determining the usefulness of these words for the learner. Consequently, the main design problem we face when using this model is the way the user must select the *kanji* he or she wants to use in each moment to generate the 8 words related to the term entered in the search engine.

In figure 4.1 we can see the first prototypes we designed to tackle this problem, using 花火 as the test search word. Initially, we thought of converting each of the characters within the

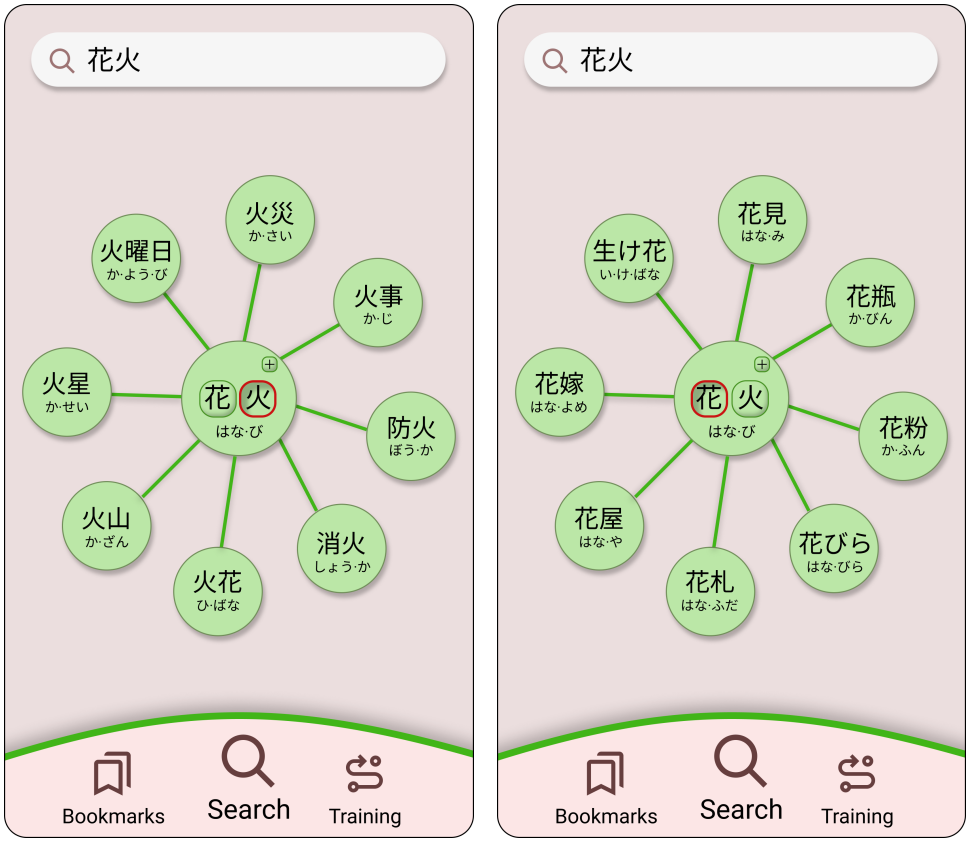


Figure 4.1: Left: Prototype of the graph interface showing the 8 most important related terms containing the 花 character from 花火. **Right**: Same interface case but using the 火 character.

central node into individual buttons. Clicking them would display the 8 most important nodes containing the same *kanji* as the one clicked. To see the full entry of the word, one would have to click the ”+” button within the node itself. However, we subsequently discarded this design, due to the poor scalability it would cause in cases where words contain more than two characters. In these cases the font would have to be reduced to fit the size of the node, making the buttons incrementally smaller and cumbersome to press. This difficulty is compounded by the possibility that some users may use the application on devices with a small screen. It is for this reason that the final graph interface uses another system to implement character selection, as we will see in the following sections.

Apart from the direct search via the search bar at the top, the navigation through the graph

will be done by simply touching one of the 8 surrounding nodes. By touching one of them, this node will become the new central node, with 8 new nodes containing shared *kanji*. The edges of the graph that connect these new nodes to each other are also omitted, leaving only those connected to the central node for the sake of visual clarity. At the bottom of the mockups in figure 4.1 we can also see the first version of our navigation bar, which will give access to the additional features that will be described in the next section. As an additional note, we can see how at this point in the development the palette was made up of green tones, something that would change in the final design.

4.2. Application content and workflow

The next step in the design of our application is to define exactly what final functionalities will be available for the user. As we have seen in the mockups of the previous section, the main functionality of our dictionary will be the graph interface, around which the rest of the components of our application will be organized. From it, it will be possible to access the complete entry of each of the words displayed. Taking as a reference the content of the entries present in other dictionary applications available for mobile, our entries will include details such as the part of the speech, the JLPT level and some example sentences. Although not all the applications analysed in the state-of-the-art chapter contained them, our dictionary will also have individual entries for each *kanji*, with information about their readings, the compound words containing them, their radicals and their stroke order. In order to not limit the results of the graph to only the 8 main words, we also propose to give the user access to an alternative list of all the words that would be connected to the searched word in the graph without the limits of the main interface. Each item in the list will be accompanied by its assigned score within our system, as well as a short explanatory text with basic information on how these scores are calculated.

Although the graph interface and dictionary entries are the core of our application, we have seen that many dictionary applications contain multiple additional functions focused on

language study without necessarily being related to the dictionary functions. In our case, we have decided to include two common additional features to assist the learner, while retaining the dictionary-focused character of our application. The first one is the possibility to save dictionary entries in bookmark folders, the contents of which can be easily consulted via a list interface. Also, the user will be able to create new folders with customised names from this same list interface. The second functionality, on the other hand, will take each of these folders and turn them into flashcard decks. Flashcards with Spaced Repetition Systems (SRS) algorithms are common tools in Japanese language study, in our case taking the form of self-generated questions. Each item within the bookmark folders will be converted into a question with four options. These options are based on the related words connected to this headword within the graph, making it a little more difficult for the user to discern between the answers. These questions can have two variants, asking for the meaning or the reading of the word. When accessing one of the flashcard/bookmark folders, these questions will be repeated in a spaced pattern, until the user gets the answer right 3 times in a row, in which case it will be considered learned. The word will still occasionally appear in the questions of that folder, and will be considered ”unlearned” again if the user fails the flashcard question. The number of words learned will be reflected in the interface of each folder. This way, the user receives a small visual motivation to keep the number of words learned in each of their own folders as high as possible.

All these functionalities will be reflected in individual interfaces that will require appropriate interconnections. To define them, we first need to assess the most optimal paths to access the functionalities according to the needs of the average user at any given time. Usually, this assessment is done through the elaboration of a user flowchart. This type of diagram helps us to visually define exactly what the user’s typical path will be when using our application. Our user flowchart can be found in Annex I. As we can see in this document, the centre of the application is the graph interface, from which the other main functionalities can be accessed through the navigation bar. However, the user’s path during a session will start at a simple welcome screen, from which he or she can use the search bar and thus quickly access the

network interface, from which the rest of the session will pivot. The graph interface in turn gives access to the full entry of the word or *kanji* being looked up. As we can see, the rest of the secondary screens are in turn connected to the main interfaces through buttons or lists.

4.3. Final Design

In this phase of the project design, we already have all the elements to define the final form of all the interfaces of our application, following the description of the functionalities given in the previous section and the structure specified in the user flowchart. Therefore, we must create final interfaces for each of the states represented in the user flowchart. Due to the large number of designs involved and the large amount of space it would require within the body of this report, we have chosen to attach a complete design guide for our application in Annex 2. This design guide is the document that would be used by the application developers to implement exactly each of the interfaces as they are designed. When consulting it, first of all, we see the final name of our application, "**Kanji Cells**", in reference to the cell shape of the graph nodes in our design. On the second page, we find some common details shared by all the designs, such as the font and the colour palette. The font chosen is Roboto, which is a *sans-serif* font designed by Google for the Android operating system. It is therefore easily recognisable for mobile device users who are used to reading it. The final colour palette combines light pastel shades of orange, green and purple to facilitate the reading of the texts in the application and avoid tiring the user's eyes, which is essential for a dictionary. On the other hand, when we need to highlight some parts of the interface, such as the buttons, we use gradients of the same colours to draw the user's attention. The second page also contains the designs in three different sizes of our application's icon. They all share references to *kanji* and the graph model, using the same colours from our palette. The rest of the pages contain the actual designs of each of the interfaces, with a small text annotation next to them to detail their structure and operation.

We can highlight from all of them the final design and functioning of our graph interface,

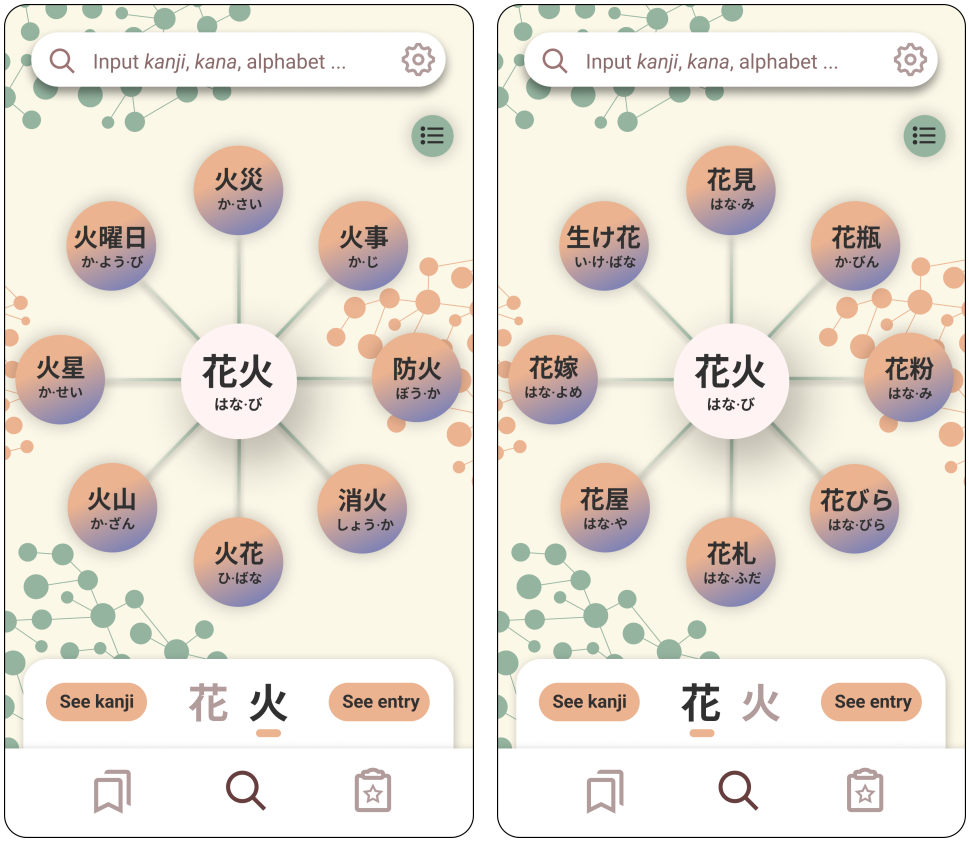


Figure 4.2: Left: Final design of the graph interface showing the 8 most important related terms containing the 花 character from 花火. **Right**: Same interface case but using the 火 character.

as we see in figure 4.2. Unlike the mockups of the same interface that we saw in figure 4.1, the final interface uses an additional options bar located just above the navigation bar, to control the graph. In the centre of the bar are the different *kanji* that make up the searched word, also reflected in the central node of the graph. By selecting one of these, the graph generates the 8 corresponding nodes containing the same *kanji*. At the same time, we have a button to access the entry of the complete headword and the entry of the *kanji* selected at that moment. This system solves the scalability problem present in the mockups and allows us to display a much cleaner graph. The final designs also include a settings screen that can be used to change some visual aspects of the application, in particular the font size and colour, as well as some more functional aspects. The latter include displaying pitch accents

on entries, using or not using our score system to order the words connected by the graph, or selecting the preferred JLPT level. This last setting is particularly useful, as it allows the user to change the exam level that receives the highest rating within our scoring system. This way, the user can prioritise in the network the words that best fit his or her level, which by default is defined as N5. The last option ”About this app”, will lead to a small text window with contact information of the developers and some attributions to the resources used in the application, as is required for example by the JMdict licence[1].

4.4. User Experience Testing

After the final design phase of our application, all the foundations of the project have been laid and we have all the necessary information for its possible implementation. However, prior to the closure of the design phase, it is always a good practice to subject our design to a round of testing to get other perspectives on our concept and the user experiences while using it. For this, we have gathered 10 students of Japanese who have studied the language for at least 1 year, and who have previously used other websites and applications with Japanese-English dictionary functions. None of them had prior knowledge of any other East Asian language, so their possible knowledge of *kanji* is derived only from their study of Japanese.

For the test, an interactive prototype has been developed using the Figma interface design tool, available via a permanent web link[2]. This prototype allows you to navigate through all the interfaces of the application as if it were a fully functional implementation of the application, but only limited to a specific use case. In this case, the test session consists of looking up the term ” 花火” in the dictionary. First, a short initial presentation was given to all students, explaining the main concept of the application and its functionalities. After this, everyone was left to navigate through the prototype individually for a few minutes without any guidance from us. Finally, everyone was asked to share their experiences while using

[1] http://www.edrdg.org/edrdg/licence.html

[2] https://www.figma.com/proto/bWfNm1F0lOXQG0oilyJ9Vv/TFM-Kanji-Cells?node-id=460%3A8652&scaling=scale-down&page-id=0%3A1&starting-point-node-id=460%3A8652

it and any suggestions they had for improving the design according to their needs. This last section was divided into suggestions on visual design and suggestions for improving functionality. Regarding the first part, students expressed satisfaction with the visual design of the application, suggesting only some minor stylistic changes. These included: underlining the word of each entry in each of their example sentences to identify them more easily, underlining the *kun* and *on* readings in the compounds that include each *kanji* entry in order to differentiate them more quickly, and giving more contrast to the colour of the progress bars in the flashcard lists.

On the other hand, students provided a greater number of suggestions regarding functionality. A common suggestion was the inclusion of more ”back” buttons in the bookmarks section, so that one could return to the list of folders from the contents of a folder without having to click on the bookmarks icon within the main navigation bar again. It was also proposed to expand the length of the results list display frame in the search bar interface, to reduce the scrolling required to go through them all. The usefulness of the graph interface as a form of indexing in the dictionary was very well received. However, in order to improve the navigation through the interface, it was suggested to include ”Undo” and ”Redo” buttons next to the graph itself. These buttons would allow the graph to return to a state before or after the current one during navigation. This way, if the user gets lost while navigating from node to node and wants to easily return to a word he or she had consulted earlier in the same session, they would only have to use these new buttons. Another important suggestion given by the students was to include the meaning of the words as a subtitle within the node itself, so that the user does not have to access the individual entry for each word to find out its meaning. The only problem with this suggestion is that the reading of the word in *hiragana* already occupied the subtitle of each node. To solve this, a button was proposed next to the graph interface that would serve as a ”switch”, swapping the subtitle of the nodes between the reading and the meaning of the word. Finally, they proposed the inclusion in the bookmarks section of some default folders containing the words corresponding to each level of the JLPT exam, to facilitate the study of these words and take advantage of the fact that this

information is already available in each of the dictionary entries.

All these improvements have also been included in the final interfaces that can be consulted in the design guide of Annex 2. In the case of new features, these are also explained in the text annotations next to each interface.

5. Conclusions

5.1. Review of the project state

At this point, we can proceed to review the status of the project before the end of this report. First, both traditional and digital backgrounds of Japanese language dictionaries have been analysed, with a special focus on Japanese-English dictionary applications. Taking the features of these as a reference, both the theoretical and technical foundations on which our application will operate have been established. This includes, on the one hand, an indexing system based on discrete mathematical graph structures, and on the other hand, all the databases necessary to supply our dictionary content. On this basis, we have built a complete set of interfaces that are adapted to the requirements of our dictionary concept. After this first iteration, an interactive design prototype was built and tested by a group of Japanese language learners. Taking into account their suggestions, some interfaces and functionalities were modified and finally translated into the final design guide, which is included in one of the annexes of this essay.

We can therefore say that the application has been completely defined and designed, ready for its possible implementation in any mobile operating system for which it could be considered. In that case, this essay can serve as a complete design document to carry out such a development. Because of this, we can say that we have covered all the objectives proposed at the beginning of the project.

Personally, I consider that this project has helped me to considerably improve my research skills, in order to design a new application concept practically from scratch. Although the

Japanese digital dictionary market has a long history and is full of competitors, our dictionary interface based on network science is rather unique and has required additional research to turn into a viable project.

Finally, as I mentioned in the introduction, I hope that this project has contributed to the approach of two very different disciplines, such as multimedia engineering and the Japanese language, so that projects like ours will become more and more common in the future and allow us to achieve together advances that benefit both in their fields.

5.2. Future improvements and research

During the development of this project we have encountered several questions and ideas that were outside the main objective of this essay, but which could still be worth implementing or researching in the future if conditions allow.

In terms of possible improvements to our network and usefulness scoring system, one of the most interesting could be to take into account student tastes in order to prioritise related words within our graph. This would require first defining what kind of genres or types of media could be part of this consideration, such as magazines, newspapers or books, and in turn, different genres of them. Once the selection has been defined, a completely new interface would have to be designed to ask the user for this information, as well as having the possibility to change it later. We would have to delimit only the part of our available corpus that fits the user's priorities and from these, recalculate the frequency of the words in the dictionary and their individual scores. As we have already seen in the section on the technical perspective of the project, Japanese corpora with genres and types of media labels already exist, so it would only be necessary to subdivide them correctly. Similarly, whether or not a word has already been learned by the user within the flashcards section of the application could also be taken into account for scoring, giving a small increase in the final score of the headword if it has been saved in bookmarks but not studied yet.

On the other hand, although in this project we have focused on *kanji* compounds or words

containing *kanji*, we should not lose sight of the fact that in Japanese there are a large number of words that do not have any *kanji* within their characters, such as particles. This automatically makes them isolated nodes in our graph and therefore chronically disadvantaged in terms of their final score. To overcome this, we can propose replacing the score for connectivity within the graph with a score for grammatical pattern usage. The more common the headword is in these type of patterns, the more points it would receive. We can even propose the creation of a secondary graph within the application in which the nodes are these words without any *kanji* and the edges between them are the grammatical structures in which both are used. In this way, for example, verbs would also be connected in our interface to the particles with which they are commonly used. Implementing features such as these would require the inclusion of a properly indexed database of grammatical patterns, which are less common than general dictionary ones. Another useful feature of easier implementation would be the inclusion of different languages in our dictionary translations, subject to the expectation of achieving a number of translated entries similar to those for English in public databases.

At the research level, it would be interesting to investigate the statistical aspects of the structure of our network, comparing the small-world features of our graph with those of *kanji*-only graphs reported in previous studies. In other words, the question would be whether *kanji* are more interconnected by their components than words are interconnected by their full *kanji*, or vice versa. If these conditions were not met, it could be assumed that there are areas of our graph with much less dense connections than others, and if so, the question would be what kind of words would make up these areas. It can be assumed that these words will contain *kanji* that are much less frequent on their own, such as obsolete characters. In this way, parallels could be drawn between the low-density areas of our network and the frequency of the *kanji* that make up these words.

Bibliography

Bailey, D. C. (1960a). Early Japanese Lexicography. *Monumenta Nipponica*, *16*(1/2), 1–52. Retrieved 2022-06-26, from http://www.jstor.org/stable/2383355

Bailey, D. C. (1960b). The Rakuyōshū. *Monumenta Nipponica*, *16*(3/4), 289–376. Retrieved 2022-07-06, from http://www.jstor.org/stable/2383204

BCN. (2022). *BCN AWARD 2022* 事務・照明関連機器. Retrieved from https://www.bcnaward.jp/award/gallery/detail/contents_type=276 ([Online; accessed July 17, 2022])

BIGLOBE. (2020). 大辞林第四版｜ビッグローブ辞書. Retrieved from https://play.google.com/store/apps/details?id=jp.ne.biglobe.daijirin.gp4 ([Online; accessed July 20, 2022])

Breen, J. (2000). A WWW Japanese Dictionary. *Japanese Studies, Japanese Studies Association of Australia*, *20*(3), 313-317. Retrieved from https://doi.org/10.1080/713683789 doi: 10.1080/713683789

Breen, J. (2004, August 28). JMdict: a Japanese-multilingual dictionary. , 65–72. Retrieved from https://aclanthology.org/W04-2209

Casio. (2010). エクスワード *xd-a6600*. Retrieved from http://arch.casio.jp/exword/products/XD-A6600/ ([Online; accessed July 17, 2022])

Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., & Ha, D. (2018). Deep Learning for Classical Japanese Literature. *CoRR*, *abs/1812.01718*. Retrieved from http://arxiv.org/abs/1812.01718

Dewey, D. (2008, 01). Japanese Vocabulary Acquisition by Learners in Three Contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *XV*, 127–148. doi: 10.36366/frontiers.v15i1.223

Humblé, P. (n.d.). Why Japanese is So Difficult and How Dictionaries Could Help. Retrieved from https://www.academia.edu/387230/Why_Japanese_is_So_Difficult_and_How_Dictionaries_Could_Help ([Online; accessed July 22, 2022])

Jeronimus, M., Westerveld, S., van Leeuwen, C. C., Bhulai, S., & van den Berg, D. (2017, nov). Japanese Kanji characters are small-world connected through shared components. *The Sixth International Conference on Data Analytics : November 12-16, 2017, Barcelona, Spain*, *1*, 53 - -58.

LogoVista. (2022). 新明解国語辞典第七版. Retrieved from https://play.google.com/store/apps/details?id=jp.co.logovista.dic.SINMEI7&hl=es ([Online; accessed July 20, 2022])

Loureiro, R. M. (2006). Kirishitan bunko: Alessandro Valignano and the Christian press in Japan. *Rc. Revista De Cultura (international Ed.)*, *19*, 135–153.

Maekawa, K. (2007). KOTONOHA and BCCWJ : Development of a Balanced Corpus of Contemporary Written Japanese..

Matsushita, T. (2021, 11). In What Order Should Learners Learn Japanese Vocabulary? A Corpus-based Approach. Retrieved from https://openaccess.wgtn.ac.nz/articles/thesis/In_What_Order_Should_Learners_Learn_Japanese_Vocabulary_A_Corpus-based_Approach/17011514 doi: 10.26686/wgtn.17011514.v1

Miller, G. A. (1995, nov). WordNet: A Lexical Database for English. *Commun. ACM*, *38*(11), 39--41. Retrieved from https://doi.org/10.1145/219717.219748 doi: 10.1145/219717.219748

Mori, Y., Hasegawa, A., & Mori, J. (2020, 07). The trends and developments of L2 Japanese research in the 2010s. *Language Teaching*, *54*, 1–38. doi: 10.1017/S0261444820000336

Nagasawa. (2017). 国語辞書の購入の手引き　市販 15 種徹底比較【 1 種追加】. Retrieved from https://fngsw.hatenablog.com/entry/2017/02/17/195320 ([Online; accessed July 17, 2022])

Nagasawa. (2019). 『広辞苑』『大辞林』『大辞泉』はどう違う？　中型国語辞典徹底比較. Retrieved from https://fngsw.hatenablog.com/entry/2019/09/26/234935 ([Online; accessed July 12, 2022])

Nakao, K. (1989, 12). English-Japanese Learners' Dictionaries. *International Journal of Lexicography*, *2*(4), 295-314. Retrieved from https://doi.org/10.1093/ijl/2.4.295 doi: 10.1093/ijl/2.4.295

NIJL-NW Database. (2020). *Database of Pre-Modern Japanese Works*. Retrieved from https://kotenseki.nijl.ac.jp/?ln=en#/pickup ([Online; accessed May 20, 2022])

NomadAI. (2022). *Yomiwa - japanese dictionary and ocr*. Retrieved from https://play.google.com/store/apps/details?id=com.yomiwa.yomiwa&hl=es&gl=US ([Online; accessed July 26, 2022])

Piperski, A. (2014, 12). An application of graph theory to linguistic complexity. *Yearbook of the Poznan Linguistic Meeting*, *1*, 89–102. doi: 10.1515/yplm-2015-0005

Richard L. (2019). *Jsho - japanese dictionary*. Retrieved from https://play.google.com/store/apps/details?id=ric.Jsho&hl=es&gl=US ([Online; accessed July 25, 2022])

Rojo-Mejuto, N. (2019, July). Los inicios de la lexicografía hispano-japonesa. *Revista de Lexicografía*, *24*, 143–169. Retrieved 2022-07-11, from https://revistas.udc.es/index.php/rlex/article/view/rlex.2018.24.0.5522 doi: 10.17979/rlex.2018.24.0.5522

Rose, H. (2013, 12). L2 learners' attitudes toward, and use of, mnemonic strategies when learning Japanese Kanji. *The Modern Language Journal*, *97*, 981–992. doi: 10.1111/j.1540-4781.2013.12040.x

SANSEIDO Co., Ltd. Publishers. (2020). 日本で一番売れている国語辞典. Retrieved from https://dictionary.sanseido-publ.co.jp/smk8/no1/ ([Online; accessed July 16, 2022])

Smith, H. (2005). Report on the Current Generation of Japanese denshi jisho 電子辞書. Based on catalogs assembled at Yodobashi Camera in Shinjuku. Retrieved from http://www.columbia.edu/~hds2/denshi_jisho.html ([Online; accessed July 18, 2022])

Sunakawa, Y. (2017). Compilation of Japanese learners' dictionaries. *Journal of Japanese Linguistics*, *33*(1), 15–26. Retrieved 2022-07-23, from https://doi.org/10.1515/jjl-2017-0102 doi: doi:10.1515/jjl-2017-0102

Takayanagi, S. (1971). Review of Ranwa-eiwa Jisho Hattatsu-shi (History of Dutch-Japanese and English- Japanese Lexicography) by N. Daisuke. *Monumenta Nipponica*, *26*(3/4), 480–488. Retrieved 2022-07-11, from http://www.jstor.org/stable/2383664

Takoboto. (2022). *Japanese dictionary takoboto.* Retrieved from https://play.google.com/store/apps/details?id=jp.takoboto&hl=es ([Online; accessed July 24, 2022])

Tanaka, Y. (2001). Compilation of A Multilingual Parallel Corpus..

Yamamoto, K., & Yamazaki, Y. (2009, jun). A network of two-Chinese-character compound words in the Japanese language. *Physica A: Statistical Mechanics and its Appli-*

cations, *388*(12), 2555–2560. Retrieved from https://doi.org/10.1016%2Fj.physa.2009.02.032 doi: 10.1016/j.physa.2009.02.032

Yamashita, H., & Maru, Y. (2000). Compositional Features of Kanji for Effective Instruction. *The Journal of the Association of Teachers of Japanese*, *34*(2), 159–178. Retrieved 2022-08-08, from http://www.jstor.org/stable/489552

Yeounsuk, L., & Hubbard, M. H. (1996). *The Ideology of Kokugo: Nationalizing Language in Modern Japan*. University of Hawai'i Press. Retrieved 2022-07-11, from http://www.jstor.org/stable/j.ctt6wqwgz

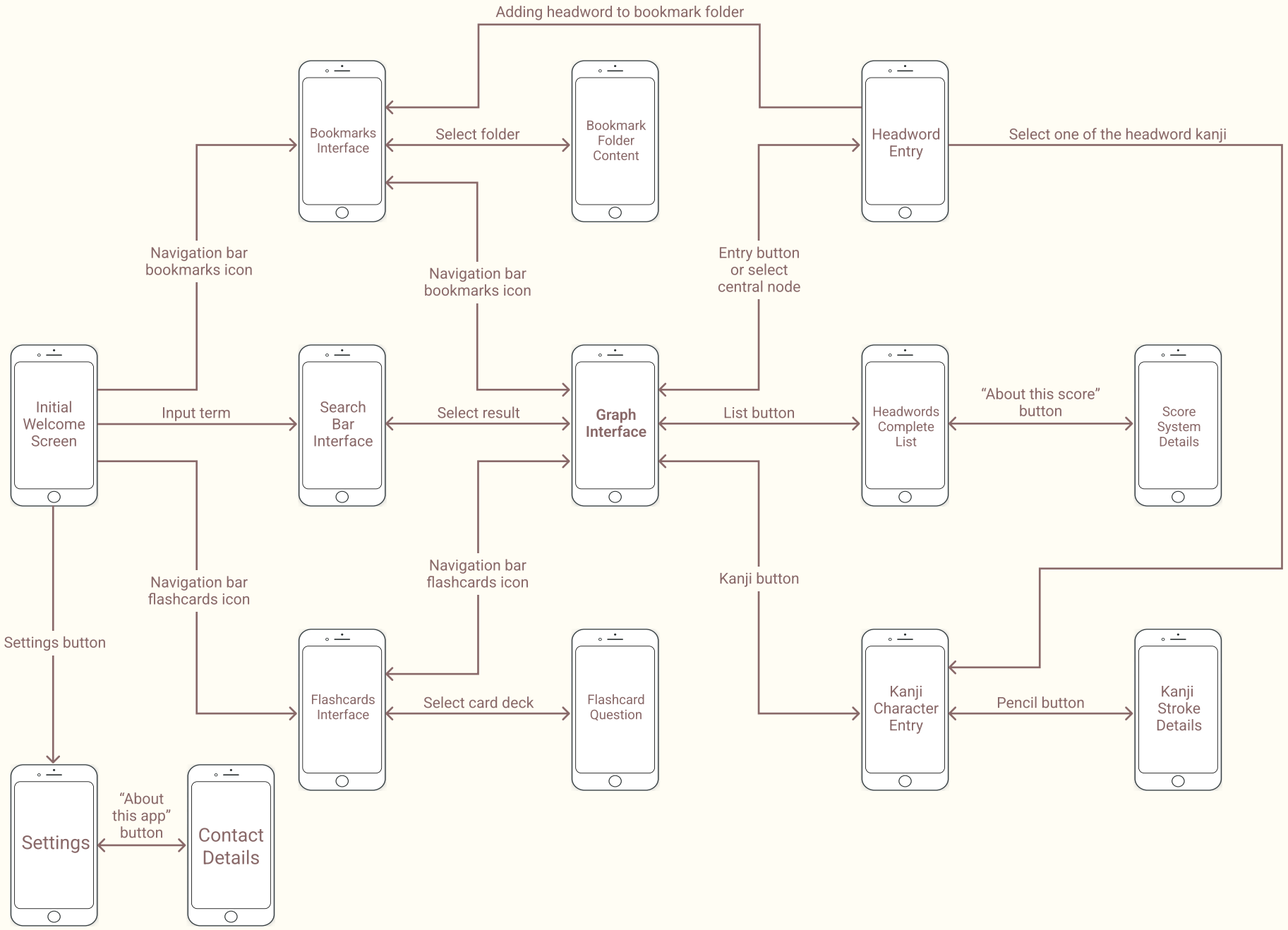
西練馬. (2019). 物書堂の辞書アプリはこれを買え：国語辞書編【*2021* 年版】. Retrieved from https://note.com/nishinerima/n/nfec6266931d2 ([Online; accessed July 20, 2022])

高山寺. (2022). 国宝・重要文化財. Retrieved from https://kosanji.com/about/national_treasure/#a_01 ([Online; accessed July 2, 2022])

A. Annex I

This annex includes from the next page the user flowchart for our application.

Kanji Cells User Flowchart



B. Annex 2

This annex includes from the next page the design guide for our application.

Kanji Cells

A Graph based Japanese-English Dictionary

Design Guide

Kanji Cells Application Icons

Large Size
200x200

Medium Size
144x144

Small Size
96x96

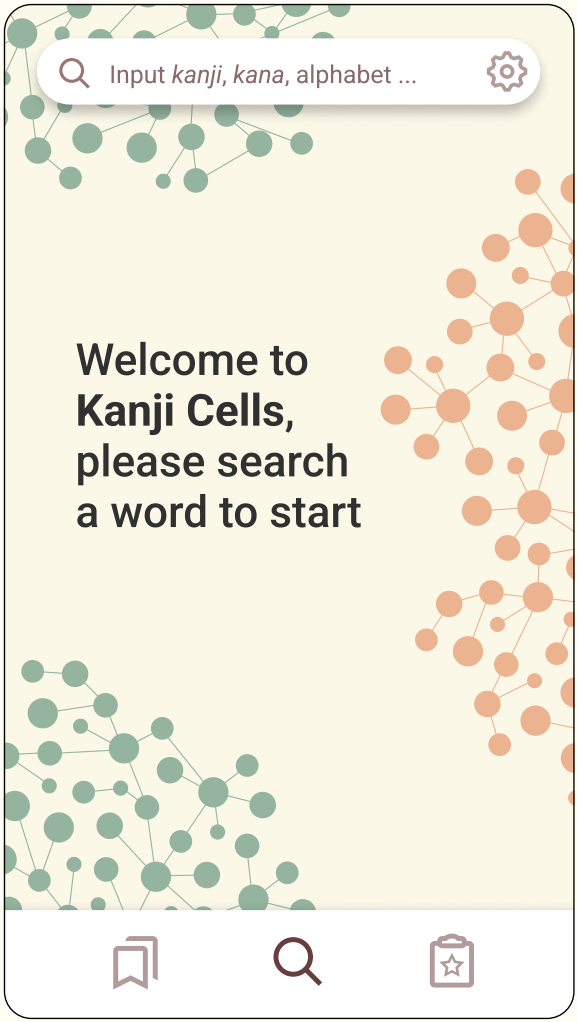
The font family used for alphabet, kana and kanji is

Roboto Medium

Roboto Medium

Prototype Color Palette

#ECB390	#94B49F	#8284B5	#FCF8E8	#303030

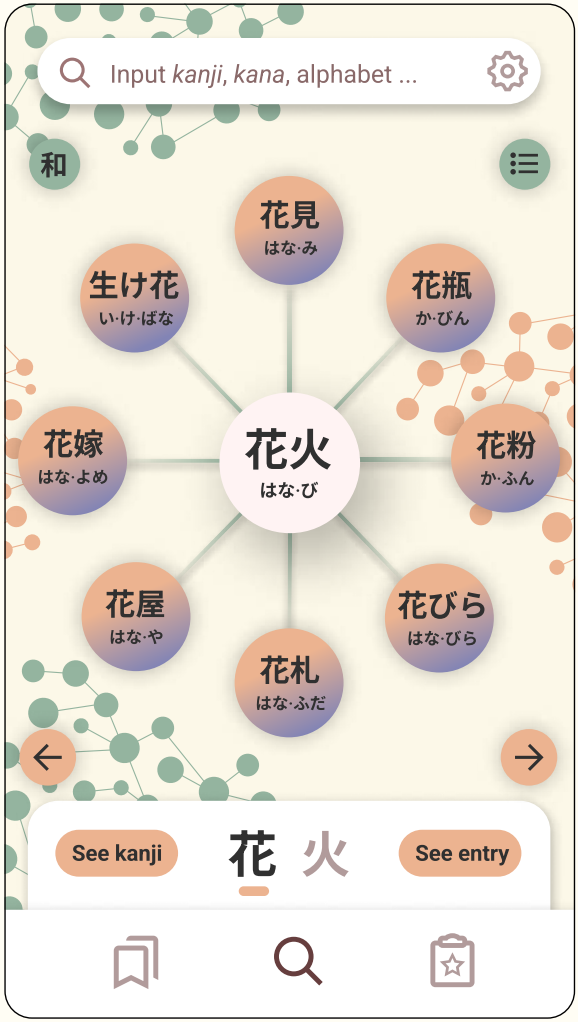


Main Page

This is the first page every user encounter when opening the application for the first time. A welcome text offers you to look up a word in the search bar, which is in the top screen area. The main options of the app are displayed in the bottom area: Search, Bookmarks and Training.

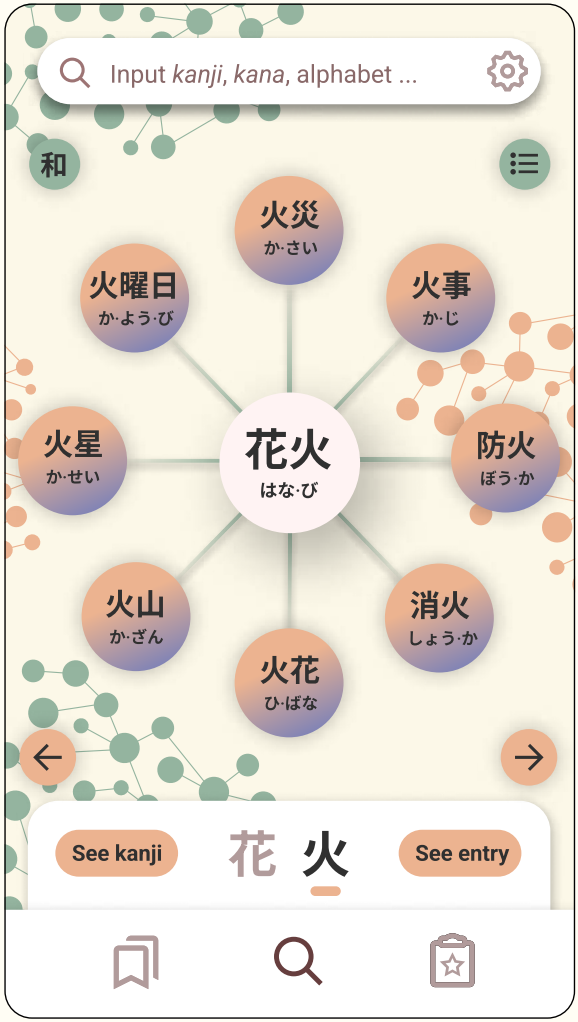
Search Bar Interface

Once the user enters a word in the top search bar, a list of results is displayed over the background interface. The user then is expected to navigate through it and select one of the options in order to start the application workflow.



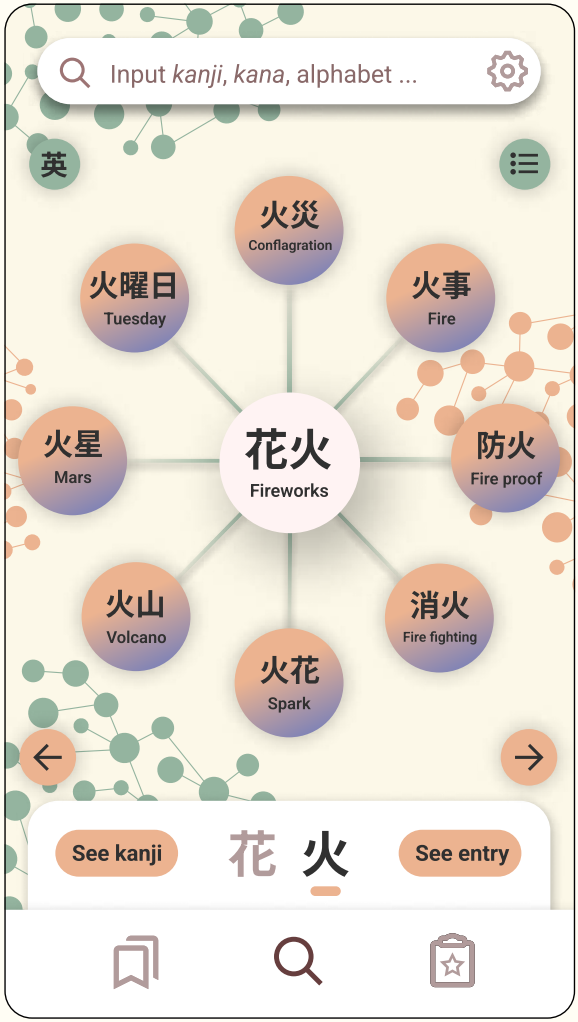
Search Graph Interface

The search graph interface is the pivot around which the entire application is built. The searched word is positioned inside a cell in the center of the screen, around which a graph of related words is created. These words contain one of the characters of the searched term, and they are selected by frequency criteria inside corpus databases. This way, we ensure that the user discovers related common vocabulary without having to search for it explicitly.



Search Graph Interface

The user can change which character from the central cell is used for generating the related terms by selecting it on the bar below the graph. If the user touches one of the surrounding cells, it will be transferred to the center, and the graph will be regenerated according to the new word. This way, users can navigate through the graph seamlessly without having to write more words than the original one. The bar below the graph also includes two buttons, one for showing the headword entry and other for showing the currently selected *kanji* entry.

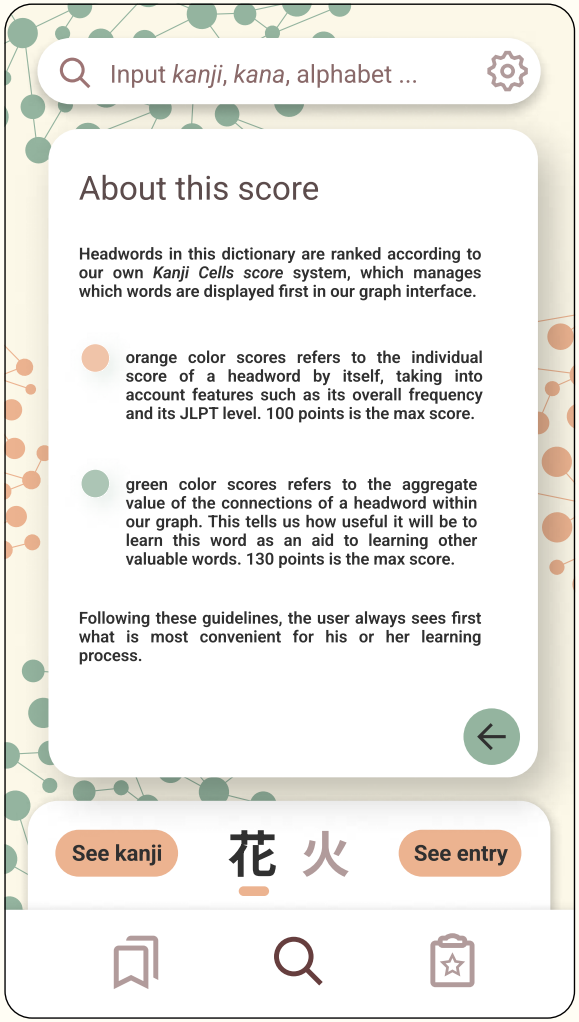


Search Graph Interface

The buttons around the graph give access to some aditional features. In case you want to return the graph to a previous state after some navigation, you can touch the arrow buttons over the bottom bar to undo the current state. Also, in the top left side of the interface there is a small button to change the subtitle of each node between its reading and its English meaning, with the former one being the case on this page's screenshot. The button on the top right side of the screen gives access to the connected headwords list, detailed in the next slide.

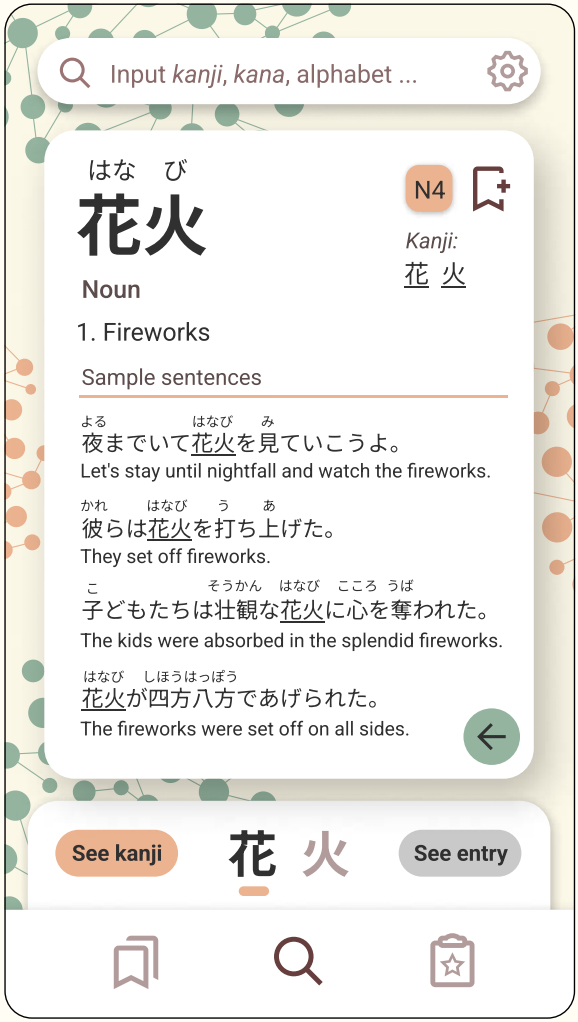
Connected Headwords List

This interface shows all the graph connected words by the currently selected *kanji*, ordered by a score which determines which of these headwords are displayed on the main graph interface. Below this list is an "About this score" button which, when clicked, takes the user to the next slide screen.



Score Information

This screen gives a brief explanation about the color code used for showing the scores in the Connected Headwords List interface. These scores are designed to highlight the most interesting words to students, both individually and in relation to their surrounding words in the graph.



Word Entry

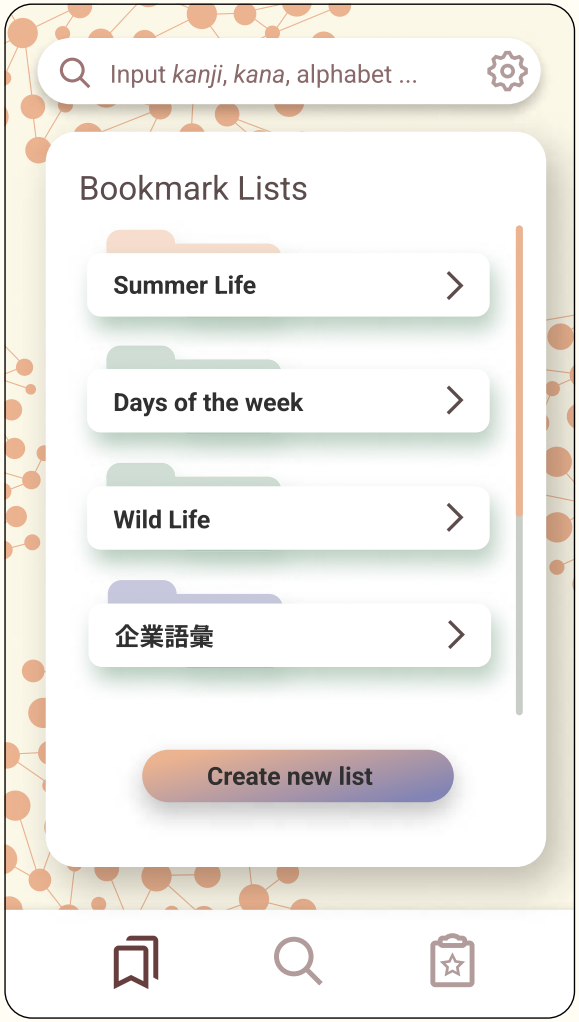
Word entries contain the usual information: the meaning or meanings, the word type, *furigana* reading, JLPT exam level and example sentences. Additionally, in the top right corner there is a button to add the entry to bookmarks, along with shortcuts to the kanji entries which compound this word.

Kanji Entry

The *kanji* entry of our application contains the *kun* and *on* readings, the meaning or meanings, the JLPT exam level and shortcut buttons to the entries of words containing this character, ordered by their readings. More details about the character strokes can be found by touching the pencil button on the top right corner, as we will see in the next slide.

Additional Kanji Information

This *kanji* strokes page serves as a subpage for the *kanji* entry section. Here we can see the actual number of strokes and their writing order. Current *kanji* radicals are also indicated at the bottom of the panel.



Bookmarks

This section opens when touching the “Bookmarks” button from the main bottom bar. Here, we can see the word entries added to bookmarks, organized in lists also created by the user. New lists can also be created easily using the button situated below the list.

Bookmarks Folder Detail

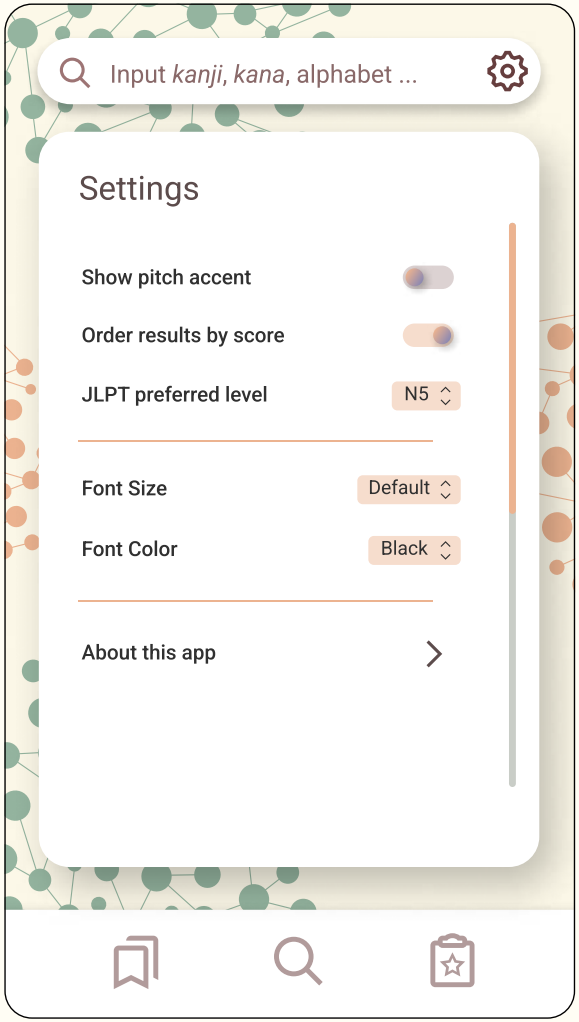
When touching one element from the bookmark list showed in the previous slide, a preview list of its contents is displayed like this. Each item includes its meaning and reading, but for more details the user can touch it to easily jump to that word full entry.

Flashcards

This section opens when touching the “Training” button from the main bottom bar. Flashcard lists are created with the same content as bookmark lists. The user has to pass through each word flashcard multiple times to be considered a learned word. Each item in the list reflects the number of words learned in a small text and throught a progress bar above each folder.

Flashcard Question

Actual flashcards in our application are multiple choice questions. There are a certain number of variants for the question, such as asking for the meaning or the reading of the word. The options displayed are based on the related words in the main graph interface, making it a little more difficult for the user to discern between the answers.



Settings

A small settings menu can be opened by touching the gear icon inside the search bar. Here, the user can change the font style and some preferences, such as showing pitch accent within word entries or adjusting the user's default JLPT level to give preference to vocabulary above that level. Selecting the “About” button will open a small text with credicts and contact information.