

Dictionary Usage and Needs of Francophone Learners of Japanese

Louis Lecailliez

Independent Researcher
Strasbourg, France
louis.lecailliez@outlook.fr

Abstract

The body of lexicography and dictionary use research is growing but a lot of languages and language pairs are understudied. So is the Japanese-French pair. This paper draws an overview of the dictionary offer, usage and needs of French-speaking university students majoring in Japanese language and civilization. Data was collected using an electronic survey which was filled by more than a hundred respondents from three universities. It sheds light on dictionary ownership, usage and sought-after lexicographic data. Results show that digital transition is done, while paper dictionary ownership is still significant. However, despite the apparent plentifulness of applications, the offer for French as a target language is lacking and the field is plagued by a data source monoculture, to the point a significant portion of students are relying on Japanese-English bilingual dictionaries. Finally, respondents voiced a need for more example sentences and a broader coverage of neologism in their dictionaries.

Keywords: dictionary use, Japanese-French dictionaries, survey

1. Introduction

1.1 Research Questions

With the democratization of personal computers and world-wide networks, both dictionaries making and usage have been profoundly transformed. As new software capabilities appear, dictionary offer and usage evolve, and previous research results are in need of reevaluation. However, a lot of languages or language pairs are under-studied, to the point many are the subject of no study at all. The Japanese-French pair is in this situation.

While both French and Japanese have a rich lexicographic history and production of their own, the pair itself is the target of only a few paper dictionaries. Coincidentally, the situation is the same in NLP: both languages are well endowed, with a decent collection of resources (such as corpora and dictionaries) and tools like morphological analyzers, whereas the pair itself is considered poorly endowed (Mangeot, 2018). With digitization and user expectation for free applications it is seriously questionable that the dictionary offering for francophone learners of Japanese is adequate. On the research side, the Japanese-French pair is the subject of a single article in *International Journal of Lexicography* (Mangeot, 2018) and none in *Lexicography* (journal of ASIALEX) nor *Lexikos*. As such, there is a real need for a contemporary picture of the dictionary offering and usage and users' needs of francophone learners of Japanese.

The present article aims to answer the following research questions:

- What is the dictionary offer and how learners make use of it?
- What is the state of the transition to digital dictionaries for French learners of Japanese?
- Which information are sought by learners and which ones are they missing?

Since Japanese has a very particular writing system that impact its dictionary offer and use, the second subsection of introduction is dedicated to explaining this specificity using paper dictionaries as reference. The first section provides an overview of the Japanese-French paper and digital dictionary offering. In a second time, the methodology used to collect and analyze data through a questionnaire is explained, as well as the one used for an experiment about example sentences quality. After that, the next section exposes and discusses the results of the survey and the experiment. The antepenultimate section discusses the implications of the findings, before the concluding section.

1.2 Specificities of the Japanese Writing System and its Impact on Dictionary Use

Japanese is written using a complex writing system that mix one logographic script (Chinese characters), two domestic syllabaries (hiragana & katakana) and one alphabet (the Latin script). The presence of logograms has a profound impact on lexicography and dictionary use: in such a system a character refers to a couple of *signifié/signifiant* as a whole, and the significant (pronunciation of the word) usually cannot be inferred without prior knowledge. This property makes text harder to read for learners, which have to rely more on dictionary in comparison to the study of European languages. In addition, it makes the lookup of a word more complex as well, often requiring two distinct types of dictionaries: one for the characters themselves, and one for the words present in the language.

In Chinese, a character can have multiples pronunciations (often also referred to as readings) that denote different meanings. In Japanese, a Chinese character can have multiple Sino-Japanese readings borrowed from Chinese, often coming from one of the three distinct waves of borrowings (Frellesvig, 2010). In addition, it can have zero, one or more readings that are “pure” Japanese words (大和言葉 *yamato kotoba*). In both Japanese and Chinese, Chinese characters are considered a distinct lexicographic object and thus being listed in their own dictionaries such as the *The New Nelson Japanese-English Character Dictionary* (Haig & Nelson, 1997) for Japanese.

The process of looking up a word in a text written in Japanese is more contrived than in a language written with an

alphabet. In French for instance, the lookup process is relatively straightforward: the learner has to lemmatize the word and then search for the lemma in the dictionary. On the contrary, spaces aren't used to separate words in a Japanese text. The reader thus needs to perform a segmentation of the relevant part of the text as the first step of looking up a word. A lemmatization step is required for verbs and verbal adjectives. Then, if the word contains any Chinese character unknown to the user those have to be lookup up. In case the dictionary used to search for characters doesn't list the compound word the student is interested in, he has to search them in a language dictionary (国語辞典 *kokugo jiten*).

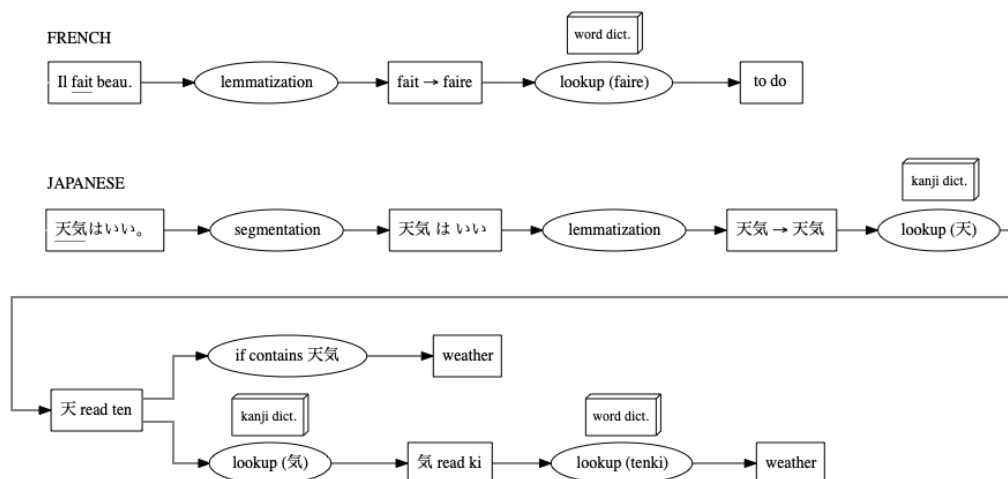


Table 1: Comparison of word lookup workflow between French and Japanese

Each step (lemmatization, segmentation or any lookup) has a potential for failure, either because of missing knowledge or error from the student, or from missing content in the dictionary. While the Figure 1 use “lookup” as a generic action in both the French and Japanese case, looking up a kanji is often more time consuming than looking up a word. Since searching a Japanese involve more steps than for a European language, the process is naturally more error prone and time consuming. This situation made the digitization even more attractive than it is for other languages. Digitization offers news user friendly ways of searching for words such as handwriting input or optical characters recognition (OCR). However, those are advanced features technology-wise and not available in every dictionary, so using a digital dictionary for Japanese could still be more challenging than for other languages, especially when copy-pasting can't be used. Moreover, even a handwriting input can be challenging to use for a beginner.

1.3 Terminology

The fast technical evolution of last decades have led to the proliferation of digital media on which dictionaries are available, which gave birth to a luxuriant and arguably confusing terminology. It is thus important to define precisely the meaning of words used in the remaining of the paper. The most overloaded term is probably electronic dictionary which is the hypernym of different kind of dictionaries that (Lew & De Schryver, 2014) advocated to call *digital dictionary*. Electronic dictionaries can be divided into pocket electronic dictionary (PED) which are popular in East-Asia, online dictionaries, desktop applications and mobile dictionaries. The line between mobile applications and online dictionaries is now blurry owing to the democratization of mobile internet: some applications store their data on the local device while others make use of a web service, and thus are more akin to an online dictionary while still being distributed in an application store.

An important distinction is often lost to the user of electronic dictionaries between the software and the data. In the “paper age” (Verlinde & Binon, 2010), a given dictionary was inseparable from the information it contained: revisions of the content led to a new edition, major changes gave birth to new dictionaries and the particular content of a dictionary was the property of its publisher and usually used exclusively by it. On the contrary, a given dictionary software can be used with different data sources. For example, the software PLECO, a dictionary application for Chinese, provides access to multiples dictionary databases from distinct vendors. Conversely, a given lexicographic database can be published as part of different applications: the Japanese-English dictionary JMDict (Breen, 2004), which is freely re-usable, is used by dozens of applications and websites. It is highly probable that most users are unaware of this distinction.

2. Dictionary Offering for French-speaking Learners of Japanese

2.1 Japanese-French Paper Dictionaries

As mentioned in Section 1.2, Japanese is a very particular language for a learner because of its partial use of a

logographic script. This means that in addition to look up and learn words, a learner has to search for Chinese characters (漢字 *kanji*) and learn them too to become a literate individual.

The Japanese-French pair has a few dictionaries, most of them being published by Japanese companies. A famous one still on the market is the Japanese-French Royal. Now out of print was the 和仏大辞典 *Dictionnaire Japonais-Français* (Cesselin, 1940). Mangeot (2018) lists only a single dictionary printed after 1945: *Assimil Japanese dictionary* (Hisamatsu et al., 2009) and 3 before that date. It is notable that there is no Japanese-French kanji dictionary, save for the dictionary-like kanji manual *Kanji et Kana* (Hadamitzky & Durmou, 1987) which only lists the 2141 characters standardized as of common usage (常用漢字 *jōyō kanji*). A decade ago, the Japanese-English dictionary *New Nelson* (Haig & Nelson, 1997) was recommended to students majoring in Japanese, and it may still be the case. This situation is in striking contrast with the Chinese-French pair for which exist both a character dictionary and a language dictionary.

No academic review exists for the all available dictionaries but a French website dedicated to Japanese language teaching rated a bunch of minor Japanese-French dictionaries, the *Assimil Dictionnaire Japonais*, three dictionaries published by the printing house YouFeng, *Le Petit Fuju Diko Français-Japonais Japonais-Français*, *Le Dicoyama français-japonais et japonais-français*, *Le Jisho 01 et 02 : Dictionnaire Français-Japonais et Japonais-Français*, *Dictionnaire japonais-français Concorde* and the dictionary integrated to the Kindle e-book reader. The conclusion of the review is that “paper dictionaries for francophone beginners are almost all bad”¹⁰. While this source is informal, its authors nonetheless used a common protocol for comparing each dictionary, in addition to giving its own opinion on the quality of the dictionaries, which is relevant as he was part of the demographic this study is targeting as a former student at Inalco. Among the main defaults mentioned are: the choice of ad-hoc romanizations, the translations, the lack of examples and explanations of usage, and the poor coverage. Those dictionaries are thus deemed not recommendable for learning.

Finally, the content of some paper dictionaries produced in Japan is available as PED. PED are not generally available in the francophony but the Japanese-French learners are a specific crowd that is advised by teaching staff to invest in one: import by buying on the internet, or occasion sold by former students are the main way to buy one. This solution is however quite costly, 200€-300€ depending on the model and being an occasion or not, so only a part of around 35% students invests as this study discovered (see section 4.1).

2.2 Japanese-French Dictionary Databases

2.2.1 JMDict & KanjiDic

Regarding the digital space, there are three main Japanese-French databases accessible on the web. The most used one is JMDict (Breen, 2004). This project initiated by J. Breen was originally a Japanese-English dictionary. It now also contains glosses in German, Russian, Hungarian, Spanish, Italian, Dutch, Swedish, French (15,000) and Slovenian. Most entries in French are very likely translations of English glosses instead of being translated from Japanese. This fact is observable from the numbers of meanings listed, their order and cognate status with English. For instance, 天気 (*tenki*) is defined by “weather, the elements” in English and “*temps (qu’il fait), les éléments*” while German list 4 distinct meanings.

JMDict is often coupled with KanjiDic – a kanji dictionary – originally targeting English but now also containing some entries in Portuguese, Spanish, and French for the *jōyō kanji*. Those 2 dictionaries are accessible for free and can be freely reused. The revision 1.08 of JMDict contains 176,874 entries, 14,892 of which has a least one translation in French.

2.2.2 Dictionnaire-japonais.com

Dictionnaire-japonais.com is a crowd-sourced online dictionary created by G. Chaslerie, accessible for free. Its database is used for three mobile applications, two made by the creator of the website (Android and iOS) and one licensed for free to a third party (Windows Phone). The database is not available for download and the dictionary content can’t be reused by third parties. It contains 61,019 entries as of writing.

2.2.3 Jibiki.fr

The Jibiki project digitized the Cesselin dictionary which content fell into public domain and modernized it in a crowd-sourced fashion by placing it in an open-source editing platform (Mangeot, 2006; 2018). The modernization steps notably comprise the use of simplified characters (新字体 *shinjitai*), modern use of the *kana* (現代仮名遣い *gendai kanazukai*) and modernization of stem-trailing characters (送り仮名 *okurigana*). The content of the French-Japanese dictionary of Raguet & Martin has been added to the platform recently, albeit a campaign of human revisions has not yet started. The revised content of the Cesselin is available to download as a public domain work. All the headword containing kanji have been humanly reviewed by contributors of the project. While an important amount

¹⁰ Translated from French.

work has been done, the crowd-sourced aspect of the project is to be relativized since most of the content have been revised by a handful of persons (Mangeot & Tomokiyo, 2018).

2.3 French-Japanese and Monolingual Dictionaries

While, dictionaries for the French-Japanese pair may be useful for students when writing in Japanese, university curriculum is heavily skewed towards translation from Japanese thus reducing the need for such a dictionary. In addition, the offer is even slimmer than of the reverse pair, with none offered by established lexicographic publishing house in France. In Japan, the Petit Royal has a French-Japanese edition.

Monolingual Japanese dictionaries are another tool that can be used for learning Japanese, despite the existing shortcomings for this usage (Gally, 2012). While the controversy between bilingual and monolingual dictionary for language education has long lasted, the difficulty of Japanese script makes it uncontroversial that monolingual dictionaries would be too hard to use for beginners. Advanced learners, or in some case intermediate ones, on the other hand could benefit from using them since it contains content rarely present in their bilingual homologues. Acquiring a monolingual Japanese dictionary was historically difficult, but is now made easier thanks to the possibility of online shopping from the few libraries specialized on Japanese in Paris. However, they remain quite costly and being paper-based, less attractive than free applications.

3. Methodology

3.1 Questionnaire Dissemination

The main data set used for this study is the answers to an electronic survey communicated to French-speaking students learning Japanese. While survey has their shortcomings, some of which can be alleviated (Lew, 2002), there are a fast and cheap way to acquire a panoramic view of dictionary landscape for a language or language pair, as recently illustrated by (Kosem et al., 2019). The present survey was disseminated in 2019 in three universities that offers Japanese degree at bachelor, master and doctorate level. It got 135 respondents. The survey, formatted as a Google Form, was first posted in a social network group of students and alumni of the University of Strasbourg from which 27 respondents participated.

Secondly, a Japanese teacher-researcher from University Paris 7 Diderot asked her students to take the survey. This resulted in 71 more participants. Finally, a former Inalco (National Institute for Oriental Languages and Civilizations) student posted the survey in a social media group of students currently attending Inalco from which got it 36 participants.

Japanese studies are offered in two distinct curricula in France: *LLCER* (foreign and regional languages literatures and cultures) and *LEA* (applied foreign languages). The former is focused on majoring a foreign language, while the latter incorporates more diverse teachings at the expense of language and civilization subjects. The duration of a bachelor degree in both cases is of 3 years (abbreviated as L1, L2 and L3), and the master degree comprises 2 years (M1 & M2).

The overwhelming percentage of respondents (94.8%) were from a *LLCE* cursus, the remaining from *LEA*. The level of study was split almost evenly over the 3 years of bachelor (L1: 23.7%, L2: 25.9%, L3: 25.2%), and the remaining quarter is divided among the two years of master (M1: 10.4%, M2: 14.8%). There was no student of doctoral level responding to the questionnaire. The results are thus very representative of the needs of students regardless of their level, and relatively close to the ratio of undergraduate to graduate students in universities.

Six people outside those universities replied to the survey as well and at least one person from Inalco is majoring in Chinese and one in Mongolian as revealed by answers in the poll. Those have been kept in the results they represent a very minor portion of the total, and doing so allows for easy use of results as displayed by the survey tool.

3.2 Encoding of Free Text Answers

Question 9 of the survey asked users to write about “the point you would like to improve in the dictionary you use the more often” to which a large textbox was allocated. Given the free writing nature of the question, this is where the analysis can be tainted with subjectivity. This question required the most effort from the respondent, so concern voiced there are to be taken seriously, as users took time to write them in plain text.

Each answer was associated to up to 3 categories in a coding scheme that evolved alongside the coded answers and then reapplied to the relevant answers. Example of categories are “coverage”, “neology”, “example sentences” etc. Each category was then counted and ranked in a decreasing way.

3.3 Experiment on Example Sentences

Analysis of questionnaire responses showed that the Tateoba Corpus is the main resource used in Japanese dictionary applications. A quick use of said resource is enough to notice that personal pronouns, which are quite rare in Japanese, are used in a significantly large part of the example sentences provided. This triggered an additional investigation related to the first research question, using the angle of quality of dictionary offer in relation to example sentences. Much like other languages in the area like Thai and Chinese, pronouns can be omitted in Japanese when they can be

inferred from context (Polio, 1995). The tendency of using null or zero pronouns is particularly strong in Japanese (Taira, Sudoh & Nagata, 2012). Therefore, a high use of pronouns is characteristic of sentences crafted by native speakers of languages where pronouns are present in most sentences such as French or English. While a few sentences containing pronouns isn't problematic *per se*, it become an issue at scale when a large number of sentences forms a corpus (here the example sentences contained in a dictionary) because it skews the frequency of apparition in comparison to natural speech. The use of null pronoun being a skill to be acquired by L2 learners, a skewed distribution could hinder the development of said skill.

In one hand the frequency of personal pronouns (first and second personal singular) is computed over the whole Tatoeba corpus. This was done using a script that counted occurrences in sentence of each pronoun exactly as written in input. On the other hand, the frequency of the same pronouns was counted manually in two pedagogical books that contains example sentences in Japanese. In both cases, a page number was semi-randomly selected and all examples of the two pages at the opened position were collected until a threshold of a hundred sentences was reached. Randomization was partial as a way to get pages from all over the books.

The selected printed material is first the Japanese manual to learn kanji named *Kanji 2200* (Tokuhiko, 2014). This manual, associated to the eponym course offered at Nagoya University, contains one example sentence for a selection of words appearing in a kanji entry. The second manual is *N1, N2, N3, N4, N5 文法大全* (Matsuyoshi et al., 2013), is a Taiwanese manual for learning Japanese grammar that contains multiple example sentence in Japanese per grammar point, accompanied by a Chinese translation. While published in Taiwan, 5 of 6 of authors have Japanese names, so it qualifies as native-level material.

The pronouns were searched for in their kanji and hiragana forms, the full list being: 1SG: *watashi* (私, わたし), *boku* (僕, ぼく), *ore* (俺, おれ), 2SG: *anata* (貴方, あなた), *kimi* (君, きみ), 3SG.MASC: *kare* (彼, かれ), 3SG.FEM: *kanojo* (彼女, かのじょ). More personal pronouns exist in Japanese but are likely of less frequent occurrence and no pronoun outside the list were found in the printed material (except one occurrence of *jibun* 自分).

4. Results

4.1 Digital Transition

The state of digital transition was evaluated based on three questions: dictionary ownership (multiple choices allowed), the most frequent type of dictionary used (single choice) and the three most frequently used dictionaries where respondents could write up to three names. All the respondents answered the first two questions and 107 answered the third one.

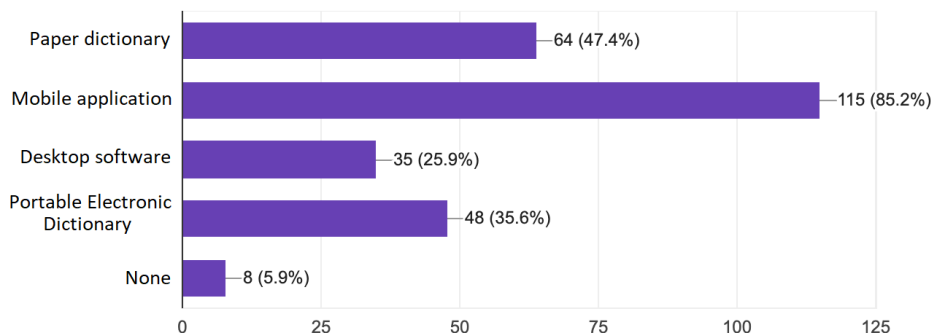


Figure 2: Dictionary media ownership

Regarding dictionary ownership the overwhelming majority of respondents own a mobile application. This result is not surprising given the penetration of smartphones, how convenient they are and the existence of a lot of free mobile applications. More surprisingly, almost half of the respondents still own a paper dictionary and a little more than a third own a PED. While the paper dictionary owned could be a French monolingual as the question was not precise enough, PED is assured to be in relation to Japanese (either in monolingual, bilingual or kanji dictionary) as those are only readily available in East-Asia. This may be the effect of teacher recommendations in university. The least popular kind of dictionary is the PC software, which is still owned by a quarter of students.

Dictionary ownership however does not inform on dictionary usage and the next question asked what is the type of dictionary the student use the most.

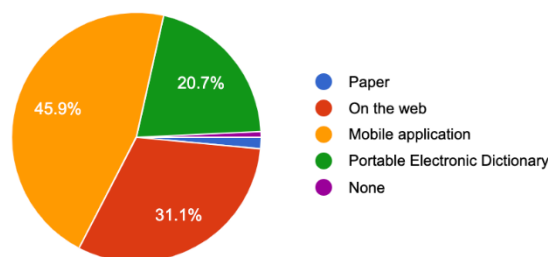


Figure 3: Breakdown of most commonly used dictionary media

This question included web dictionaries which was absent from the previous one because ownership made no sense in this case. Only two respondents use a paper dictionary more than any form of digital one: while paper dictionaries might see occasional use, this is strong evidence that the digital transition have been completed, as digital media dominate the answers, either in the form of mobile application, website or PED.

With 45.9% of users using a mobile application, 20.7% a pocket electronic dictionary and presumably a fair share of web searches being made from a mobile device as well, it is clear that the portability and mobility of dictionaries is a major concern for users.

4.2 Dictionaries and their Popularity

Question 6 of the survey asked the write up to three dictionaries used by the learner, in order of frequency of use. The field was a free text input so results had to be manually cleaned up. In a few cases, the referenced dictionary was not clear. Only 107 people answered this question. The top-10 dictionaries are listed in Table 1. The table is augmented with columns indicating their type and the target language of bilingual dictionaries. “Multilingual” means multiple languages are available (see Section 2.2.1, and “Japanese” that the dictionary is monolingual.

Rank	Dictionary	Count	Type	Language
1	Jisho.org	36	website	English
2	Dictionnaire-japonais.com	20	website; offers applications	French
3	Imiwa?	14	iOS application	Multilingual
3	Takoboto	14	website; offers applications	English
5	Yomiwa	7	iOS & Android application	Multilingual
6	Petit Royal	6	paper	French
6	Weblio	6	website	Japanese
6	Kanji Study	6	Android application	English
9	Shirabe Jisho	4	iOS application	English
9	Assimil	4	paper	French
9	Google Translate	4	translation service (web)	Multilingual
9	Japanese	4	iOS / Android application	English
9	Japanese Dictionary	4	Android application	English

Table 1: Most used dictionaries mentioned by respondents

The most used dictionary, and a majority (5 others dictionaries) present in the list are Japanese-English dictionaries. Given the relatively average English proficiency in France, this is a surprising result. Moreover, the large number of *faux amis* (false cognate) in-between the two languages, this is a concerning observation and research could be done to evaluate the impact on translation quality obtained from such users. It is also notable that Google Translate is used as a dictionary despite it being an automated translation tool by a fair number of users, and has a rank similar to as specialized applications.

4.3 Information Sought by Students

The final research question addressed by the survey pertains to the information that learners are searching for in dictionaries. The data to answer it was collected via two questions: the first one asked to choose between different data often present in a dictionary (definition, part of speech, etc.) It was possible to check multiple answers as well as add one or more answers. The second question was on open question about which points the respondent would like to be improved in the dictionary they use the most.

Result for frequently sought after information is displayed in Figure 4. Every respondent filled this question. Without surprise, the most common usage of a dictionary for this population is to get a translation or definition of a word (97.8% users). The second most sought information by most of users (80.7%) is the reading or pronunciation of a

word. This is a specificity of Japanese studies, stemming from the writing system, as explained in Section 1.2. The top three answers are also coherent with findings by Janjić & Kocijan (2022) from a survey distributed to European students of Asian languages, with the exception of grammatical information. This absence can be explained by the fact Japanese has a straightforward grammar, which is not the case of Sanskrit and Hindi that were the languages studied by Janjić & Kocijan (2022).

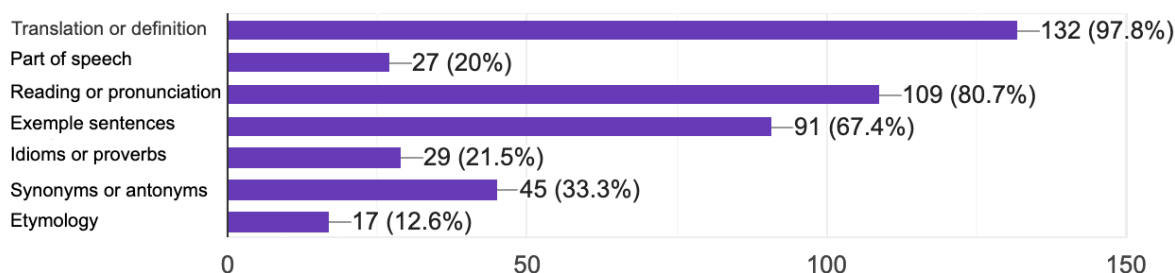


Figure 4: Most sought-after lexicographical information

To be sure not to miss any type of information that could have been overlooked, it was possible for the respondent to add additional answers. Only 9 respondents made use of this option, adding the following 9 type of information: (a) “kanji of a word”, (b) “the writing [of a word]”, (c) “conjugation, usage context”, (d) “frequency of use”, (e) “expressions”, (f) “verb declensions”, (g) “traditional writing of the word”, (h) “compound word using a given kanji” and (i) “how to write a kanji”.

Answers (a), (b), (g) and (i) all arise because of the Japanese writing system, and for (g) how it was reformed in 1946. Studies on languages written using an alphabet will miss those problematics, save for (b) which is still applicable to language where the reading and writing rules are complex, such as French and English.

Another information sought by more than half of the respondents are examples sentences. Not all dictionaries feature this, especially the paper ones but it is true of applications as well. When present in a free electronic-dictionary, the data is invariably taken from the Tatoeba corpus, which for the Japanese language includes a revised version of the Tanaka corpus.

The 4th most sought-after information is related to synonyms and antonyms, raised by a third of respondents. Synonyms is in itself a lexicographic question large enough to warrant dedicated dictionaries. The data is absent from JMDict.

Amongst the proposed choices, the remaining three are consulted by a minority of users. Idioms and proverbs are relevant to a small user base of 21.5%. It is notable that Japanese lexicographic culture provides 2 types of dictionaries for proverbs: in one hand “Japanese proverbs” (ことわざ辞典, *kotowaza jiten*), and the other hand “four characters idioms” (四字熟語辞典, *yojijukugo jiten*) which lists phrases or proverbs written with 4 Chinese characters. While a small part are read using Japanese readings, most were borrowed from Chinese, where those are called *chengyu* (成語, *chéngyǔ*). Again, this is not present in JMDict.

Part of speech is consulted frequently by one fifth of learners. This information exists in JMDict, and is often inferable from the word form so it isn’t much of an issue. Finally, etymology is interesting for a fraction of users. Again, this is very specialized knowledge, especially in Japanese where not only the word itself but also its written form in kanji may warrant explanations can be the subject of dedicated dictionaries. Those exists in Japanese: [ex] While this information can be interesting to learn for general culture, it is not necessary for any subject a student might encounter at university except for people working directly on the topic at master level.

4.4 Users Wishes

The free text section was very informative but also the more difficult to interpret. The number of answers was 85 (~63%), which is high considered this require a lot more efforts to fill than other items. Most answers were short (often a single sentence) but a few were very elaborate, one user having written 97 words.

For each answer a category was applied such as “example” or “coverage”. The categories were crafted while reading the answers. Any answer could be tagged in up to 3 categories. As some answers were difficult to categorize, someone interested in working on improvement of Japanese dictionaries for French learners would be advised to look at the raw data directly. Some answers were pretty straightforward to understand and encode, such as ¹¹ “more vocabulary!”, “more example”, “be more reliable” while others spawned multiple sentences.

The Table 2 lists the encoding categories along the number of responses encoded using it.

Category	Frequency
Example sentences	43

¹¹ Replies are written in French and are translated here freely by myself.

Context	10
Coverage	9
Manuscript character recognition	9
Precision	7
Search	7
Nuances	5
Synonym	5
French	4
Nothing	3
Price	3

Table 2: Categories appearing in user wishes

Users desire for example sentence is striking: a third of all survey respondents wants improvement in this area, with more examples. The first category is also related to the next most popular: context. The context category is distinct because the word itself appear a lot in user responses, usually but not always associated with example sentences. It could mean that reader wants to know when to use a given word instead of a synonym.

Coverage encodes requests pertaining to the number of entries. Since JMDict is not produced by professional lexicographers but compiled from user submissions, it lags in specialized areas such as neology. This has effect on content: one respondent asked for “More vocabulary!” while another said: “Adding slang words”.

Besides the number of entries, about as much users complained about *precision*, which encodes the quality of translations. Remarks such as “to have a translation in French as efficient as the English one”, “Clarity of definition”, and “Add nuances for some Japanese words that are translated the same in French but have different nuance in Japanese” denote that a part of users are dissatisfied with the quality of glosses.

We already discussed the fact that Table 1 contained as top dictionary a Japanese-English one and that most dictionaries in the list were Japanese-English as well. This has been noticed by some users as well as they requested French translations, for instance: “Moreover, the dictionary is only in Japanese-English”. It is baffling that the said individual didn’t search for a dictionary better suited to his needs. The effect choosing the first meaning in a dictionary entry described by (Nesi & Haill, 2002) seems to happen with search engine as well: even if the top result in a Japanese-English dictionary website or application, Japanese-French dictionaries indeed appeared on the first result page at the time of data collection.

Two categories pertained to lookup: “manuscript character recognition” and “search”. The first one refers to request concerning handwritten recognition of Chinese characters, which is a very useful feature of beginners. The second category regrouped suggestions related to search in general, both for French words (one user noted that compounds words are hard to find) or Japanese. As noted in Section 1.2, handwriting recognition can be difficult to implement for developers and is not offered in all apps. This can be circumvented using native platform keyboard when available.

4.5 Example Sentences Quality

As example sentences were often mentioned as a lacking point in the survey, it is legitimate to investigate their quality. Table 2 lists the number of times each pronoun under investigation was found in the Tatoeba Corpus. Note that *kanojo* also has the meaning of “girlfriend” and kanji used to write *kimi* can be a name suffix read as “kun”. None of those usages have been disambiguated nor filtered out, hence the magnitude of the result is more significant than the precise details.

Pronoun	Kanji	Hiragana
<i>watashi</i>	30,475	486
<i>boku</i>	2,125	216
<i>ore</i>	455	25
<i>anata</i>	95	5,958
<i>kimi</i>	4,927	143
<i>kare</i>	49,368	830
<i>kanojo</i>	14,907	1

Table 3: Pronouns appearance in Tatoeba corpus of Japanese sentences

In total, 84,591 sentences out of the 105,896, that is 44%, contained on the pronoun under search.

4.5.1 N5-N1 All Grammar

The count from the Taiwanese grammar manual is presented in table 3. The first column indicates the page from which the examples were taken and the final column the number of example sentence considered in the page range. The second column lists the pronouns found in the range.

Page range	Pronouns	Sentences count
40-41	<i>anata</i> 2	15
72-73	<i>kare</i> 1	18
112-113	<i>watashi</i> 8, <i>kare</i> 1	14
206-207	<i>watashi</i> 1, <i>anata</i> 1, <i>kare</i> 1	17
258-259	<i>kare</i> 1, <i>kanojo</i> 1	18
310-311	<i>kare</i> 3, <i>kanojo</i> 1	16
376-377	—	15
Total	<i>watashi</i> 9, <i>kare</i> 7, <i>anata</i> 3, <i>kanojo</i> 2	113

Table 4: pronouns found in the book N5-N1 All Grammar

All the personal pronouns listed are written in kanji, except *anata* which is always written in hiragana. No sentence features more than one pronoun. The pages 112 and 113 contains a higher-than-normal pronoun count, which is explained by the fact the grammar point under discussion are related to expressing giving and receiving things. In total, 19% of the sentences contain a personal pronoun; 12% if the pages 112-113 were removed.

4.5.2 Kanji 2200

The last source used is a Japanese manual for learning kanji, which also features vocabulary and for some of them example sentences. In that source only 3 pronouns were found in the page observed. This is the lowest of all sources, and also the only one published in Japan from a Japanese author. Pages 30, 78, 110, 200, 273, 303 and 428 were used, and *watashi* is found once in pages 30, 78 and 200. This amount to about 3% of the example sentences containing a pronoun.

5. Discussion

5.1 Digital Transition & Dictionary Monoculture

Data from sections 4.1 & 4.2 show the digital transition is completed for the Japanese-French pair. Ownership of printed dictionary is still a thing for almost half of people, day to day use is done almost exclusively from an electronic device. A small portion of learners (20.7%) use a PED which contains dictionaries from established Japanese publishers, while the remaining 79.3% relies on freely accessible dictionaries. Numerous different applications are used, but all but one relies on a single resource provider. Those numbers are taken from a cohort of learners majoring in the language which thus have a strong incentive to invest in tool of quality in spite of their price: the possession and use of PED of present and past learners of Japanese outside higher-education institutions is likely to be way below the 20% presented here.

The Japanese-French landscape is dominated by a monoculture where Breen's JMDict and KanjiDict dictionaries are used by almost every free available application. The only concurrent known to learners is dictionnaire-japonais.com, a crowd-sourced dictionary. The alternative Jibiki.fr was poorly referenced by search engines at the time of data collection and was unknown to students. The most cited online dictionary used by student doesn't even contains the French translations from JMDict. Those glosses were often translated from English in the first place. This fact, added to the below-average proficiency of English by French in comparison to their Europeans neighbors, raise the question of the quality of understanding and resulting translations by students who relies on those applications. This is especially problematic since researcher on dictionary use shown that student often select the first meaning they encounter (Nesi & Haill, 2002) or feel constrained to choose in between definitions present in an entry. The same seems to apply to search engine results: users prefer to complain about the lack of French content in the application they use than to search for an alternative.

On the lexical content itself, the dictionary is lacking some of data used as quality evaluation criterion used by (Lew & Szarowska, 2017), to the extend those overlapped with the data present in the survey. Coverage itself in particular is mentioned by a sheer number of respondents. Other data consulted are specialized enough to warrant a dictionary of their own. It is unclear if this fact is understood by readers as the survey didn't investigate it directly, but the fact that no specialized dictionaries (e.g. etymology dictionary) are mentioned by users, especially those interested in specialized information, may indicate that learners aren't fully aware of the large range of available specialized Japanese dictionaries. Those being published only in Japan, often only on paper, does not help awareness nor acquisition.

5.2 Examples Sentences

Example sentences are a big concern of the students. It is reflected by having 67.4% of respondents claiming to consult them, but also by being the topic of most requests for dictionary improvement. In effect, more than half of the respondents (41 people) expressed wanting more examples of word usage; this is 5 times more than the second ranked request which is about dictionary coverage. The need for more sentence examples is thus glaring.

In current electronic dictionaries used by the respondents, the monoculture of dictionary content (word, kanji and their definitions) is also present for example sentences content. When examples are present, they are systematically taken from Tatoeba corpus, a subset of which is included in JMDict. The Tatoeba corpus has quality issues, in particular because it doesn't vouch contributor language skill and a lot of content for Japanese is created by non-native speakers.

It is particularly clear from the result of our investigation: native content features low count of pronouns (from less than 3% up to 19% when the situation at hands calls for it in the material, we examined) in the overall set of sentences. On the other hand, content outsourced to a mostly non-native speakers overuse personal pronouns, to the point that almost half (44%) of the corpus of sentences feature them. Applications makers could select subset of example sentences from the corpus but more investigation would be required to see if this is effectively the case in practice.

6. Conclusion

Similarly, to the field of NLP two issues are plaguing lexicography research: on one hand only a handful of languages or pairs, often English or related to it, are investigated. On the other hand, research results are generalized based on the restricted typology of studied languages. The Asian languages making use of Chinese character are posing different challenges to users to warrant more dedicated research.

The present study tried to shed a light on usage and needs of such an understudied language pair. A surprising finding is the overreliance on English-based dictionaries while French-based alternatives exist. Both however are lacking in the sentence examples department as mentioned by users and conformed by experimentation. The language pair is pretty much in the "shoestring lexicography" state described by Klegr (2004).

The lack of commercial incentive let the development of good Japanese-French dictionaries in the hands of community, as road paved by the Jibiki project. However, the problem of funding is too deep to be addressed by academia alone. Creative solutions must be tried if the sad state of Japanese-French lexicography is to be redressed. I therefore advocate for the creation of an association to federate interested individuals, secure funding from the private sector and use it to create lexicographic content and research related to the Japanese-French pair.

7. Acknowledgements

Research was done at Kyoto University under Japanese MEXT scholarship. This study would not have been possible without the support of Akiko Nakajima and Nodoka Nakamura who helped disseminate the survey.

8. Bibliographical References

- Breen, J. (2004). JMDict: a Japanese-multilingual dictionary. In *Proceedings of the workshop on multilingual linguistic resources* (pp. 65-72).
- Cesselin, G. (1940). *Dictionnaire japonais-français*. Maruzen.
- Frellesvig, B. (2010). *A history of the Japanese language*. Cambridge University Press.
- Gally, T. (2012). Kokugo Dictionaries as Tools for Learners: Problems and Potential. *Acta Linguistica Asiatica*, 2(2), 9-20.
- Hadamitzky, W., & Durmou, P. (1987). *Kanji & Kana: manuel de l'écriture japonaise et dictionnaire des 1945 caractères officiels; suivi de caractères composés importants formant un vocabulaire de 11000 mots*. Jean Maisonneuve successeur.
- Haig, J. H., & Nelson, A. N. (1997). *The New Nelson Japanese-English Character Dictionary*. Tuttle Publishing.
- Hisamatsu K., Y. Obataya, F. Hayakawa, et al. (2009). *Dictionnaire Japonais-Français / Français-Japonais*. Assimil.
- Janjić, M., & Kocijan, K. (2022). Creation of multilingual learners'e-dictionary for learners of Asian languages. *Lexicography*, 9(2), 194-224.
- Kleg, A. (2004). Shoestring Lexicography. In F. Čermáka (Ed), *Jazyky a jazykověda* (pp. 445-456). Filozofická Fakulta Univerzity Karlovy.
- Kosem, I., Lew, R., Müller-Spitzer, C., Ribeiro Silveira, M., Wolfer, S., Dorn, A., ... & Nesi, H. (2019). The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture. *International Journal of Lexicography*, 32(1), 92-114.
- Lew, R. (2002). Questionnaires in dictionary use research: A reexamination. In *Proceedings of the Tenth EURALEX International Congress* (pp. 267-271). Copenhagen University.
- Lew, R., & De Schryver, G. M. (2014). Dictionary users in the digital revolution. *International Journal of Lexicography*, 27(4), 341-359.

- Lew, R., & Szarowska, A. (2017). Evaluating online bilingual dictionaries: The case of popular free English-Polish dictionaries. *ReCALL*, 29(2), 138-159.
- Mangeot, M. (2006). Papillon project: Retrospective and Perspectives. In *Acquiring and Representing Multilingual, Specialized Lexicons: the Case of Biomedicine*, LREC workshop (pp. 15-20).
- Mangeot, M. (2018). Collaborative construction of a good quality, broad coverage and copyright free Japanese-French dictionary. *International Journal of Lexicography*, 31(1), 78-112.
- Mangeot, M., & Tomokiyo, M. (2018). Contributions et corrections dans le dictionnaire japonais-français Jibiki.fr. In *11e journées du réseau "Lexicologie, terminologie, traduction"*.
- Yoshimatsu, Y., Tanaka, Y., Nishimura, K., Chida, H., Ōyama, W., & Huang, M. [吉松由美・田中陽子・西村恵子・千田晴夫・大山和佳子・黄茗楚] (2013). Xīnzhì rì jiǎn juéduì hégé N1,N2,N3,N4,N5 wénfǎ dàquán [新制日檢 絕對合格 N1,N2,N3,N4,N5 文法大全 New system of daily test. Absolute pass. N1, N2, N3, N4, N5 All the grammar]. Shāntián shè wénhuà shìyè yǒuxiàn gōngsī [山田社文化事業有限公司 Yamada society for cultural undertakings Limited Co].
- Nesi, H., & Haill, R. (2002). A study of dictionary use by international students at a British university. *International journal of Lexicography*, 15(4), 277-305.
- Polio, C. (1995). Acquiring nothing?: The use of zero pronouns by nonnative speakers of Chinese and the implications for the acquisition of nominal reference. *Studies in Second Language Acquisition*, 17(3), 353-377.
- Taira, H., Sudoh, K., & Nagata, M. (2012). Zero pronoun resolution can improve the quality of J-E translation. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation* (pp. 111-118).
- Tanaka, Y. (2001). Compilation of a multilingual parallel corpus. *Proceedings of PACLING 2001*, 265-268.
- Tokuhiro, Y. [徳弘康代] (2014). Nihongo gakushū no tame no yoku tsukau jun kanji 2200 [日本語学習のためのよく使う 順漢字 The 2200 frequently used in Japanese learning]. Sanseidō [三省堂].
- Verlinde, S., & Binon, J. (2010). Monitoring dictionary use in the electronic age. In *Proceedings of the XIV Euralex International Congress* (pp. 1144-1151). Ljouwert: Afûk