

Constructing a Bilingual Japanese-English Collocations Dictionary: A Corpus-based Approach

Uchida, Satoru
Kyushu University

Ishii, Yasutake
Seijo University

Kudo, Yoji
Tamagawa University

Haswell, Christopher G.
Kyushu University

他

<https://hdl.handle.net/2324/4480695>

出版情報 : 2020 Conference Proceedings of the 18th Annual Hawaii International Conference on Education, pp.879-882, 2020-01-04. Hawaii International Conference on Education
バージョン :
権利関係 : (c) The Authors.



Constructing a Bilingual Japanese-English Collocations Dictionary: A Corpus-based Approach

Satoru Uchida¹, Yasutake Ishii², Yoji Kudo³, Christopher G. Haswell⁴,
Danny Minn⁵, Seiji Uchida⁶, Ichiro Akano⁷

^{1, 4} Kyushu University, ² Seijo University, ³ Tamagawa University,
⁵ The University of Kitakyushu, ⁶ Nara University, ⁷ Kyoto University of Foreign Studies

¹ 744, Motooka, Nishi-ku, Fukuoka, Japan

¹ uchida@flc.kyushu-u.ac.jp

1. Introduction

Dictionaries are undoubtedly one of the most important and useful companions for language learners. Due to the advances in corpus linguistics and lexicography, English learners' dictionaries have been significantly improved, and new types of dictionaries have emerged. Among them are collocations dictionaries which provide information on the possible combinations of words (e.g., an entry for "question" would include collocational adjectives such as "awkward," "hard," and "tricky"), which are particularly useful for encoding purposes. However, the current collocations dictionaries are compiled on a monolingual basis, that is, English headwords (e.g., "question") are given English collocations (e.g., "awkward," "hard," and "tricky").

The present study aims to construct a bilingual collocations dictionary; Japanese headwords (e.g., "*mondai*" [problem/question]) are given English translations of their collocations (e.g., "*toku*" (solve) is combined into "*mondai wo toku*" and translated as "solve a problem"). We will take a corpus-based approach, and large-sized corpora will be employed for finding collocations of the headwords. Each collocation is grouped based on cases (e.g., *wo* case (accusative), *ga* case (nominative), etc.) so that users can search their target collocations quickly, which are then followed by English translations and usage notes.

2 Previous studies

Ellis (2003) and Chen and Baker (2010) point out the importance of collocations for English learners. Specifically, collocations play a vital role for writing activity and hence a dictionary of collocations is of great use for learners. In addition, according to Horsfall (1997), dictionaries are important study companions for language learners and thus are helpful for autonomous learning.

Against the backdrop of these findings and the advances in corpus linguistics, a number of collocations dictionaries have been published including *Oxford Collocations Dictionary for Students of English*, 2nd edition (Oxford University Press, 2009), *Macmillan Collocations Dictionary* (Macmillan Education, 2010), and *Longman Collocations Dictionary and Thesaurus* (Pearson, 2013). However, these dictionaries are monolingual, that is, English headwords and their English collocations are

presented in each entry, which makes these dictionaries less useful for encoding purposes than bilingual dictionaries, especially for elementary learners.

There are two types of bilingual dictionaries. The first type is the cases where second (or foreign) language headwords are given translations in the first language (henceforth L2-L1 type). For Japanese learners of English, for example, English-Japanese dictionaries fall into this category. The other type refers to the bilingual dictionaries whose headwords are in the first language, and translations are provided in a second language (henceforth L1-L2 type). Again, to use Japanese learners of English as an example, Japanese-English dictionaries are the cases in point. In contrast to the active publications and wide use of English-Japanese dictionaries, Japanese-English dictionaries are considered less useful and less popular in English classes. Some teachers even intentionally avoid introducing Japanese-English dictionaries in their classes. One of the reasons is that it is hard to find an appropriate word in the target context in L1-L2 type bilingual dictionaries (East, 2008). Also, it is said that this type lacks necessary information on the second language in terms of grammar and usage even when a user locates the right headword (cf. Laufer, 1995). In addition, according to Thompson (1987), L1-L2 type dictionaries undesirably promote word-to-word understanding in the first and second languages.

Laufer (1995) is a pioneering attempt to provide a framework for improving L1-L2 dictionaries. She argues the importance of examples, usage notes, and the list of related words in the entries, and Laufer and Levitzky-Aviad (2006) examine the usefulness of her framework in the writing task using revised Hebrew (L1)-English (L2) entries. The results showed that enriching the information in L1-L2 dictionaries is effective, although collocations are beyond of the scope of their study.

3 Overview of our approach

In order to create a useful L1-L2 dictionary that overcomes the drawbacks of the existing Japanese-English dictionaries, this study takes a corpus-based approach to lay the foundation of an L1-L2 collocations dictionary. Specifically, we use two large-sized Japanese corpora (the Balanced Corpus of Contemporary Written Japanese¹ and Tsukuba Web Corpus²) to select headwords and collocations, which are then translated into English.

One of the advantages of our approach is that it makes the coverage of the dictionary comprehensive. Our preliminary survey showed that the coverages of the existing Japanese-English dictionaries are limited with regard to the collocations of headwords. For example, only 2.5–3.75% out of the top 20 frequent collocations of the noun “*kankei*” (relationship) is covered in three major Japanese-English dictionaries. Table 1 shows the top five collocations of “*kankei*” in Tsukuba Web Corpus and its coverage in major Japanese-English dictionaries (OJE2: *O-LEX Japanese-English Dictionary*, 2nd edition (Obunsha, 2016), GJE3: *GENIUS Japanese-English Dictionary*, 3rd edition

1 <http://nlb.ninjal.ac.jp/>

2 <http://nlt.tsukuba.lagoinst.info/>

(Taishukan, 2011), WJE2: *WISDOM Japanese-English Dictionary*, 2nd edition (Sanseido, 2012)).

Table 1: Top five collocations of wo-case for “*kankei*” [relationship]

Case	Rank	Collocation	Freq	OJE2	GJE3	WJE2
ヲ[wo]	1	{関係}を築く [build a relationship]	7336			✓
ヲ[wo]	2	{関係}を持つ [have a relationship]	5150			
ヲ[wo]	3	{関係}を作る [create a relationship]	2595			
ヲ[wo]	4	{関係}を構築する [build a relationship]	1963			
ヲ[wo]	5	{関係}を結ぶ [establish a relationship]	1660			✓

It is clear from this table that even highly frequent collocations are not covered in the current Japanese-English dictionaries. However, this coverage can be significantly improved if we create entries based on frequencies in corpora.

The entries of our dictionary are organized on the basis of Japanese semantic cases. Also, we plan for our dictionary to be online and publicly available, which would increase searchability of the dictionary. For example, the collocations of “*kankei*” is classified into “ga-case + verb”, “ni-case + verb”, “wo-case + verb” and so on. By users clicking each pattern, collocations and their translations are displayed (see Figure 1).

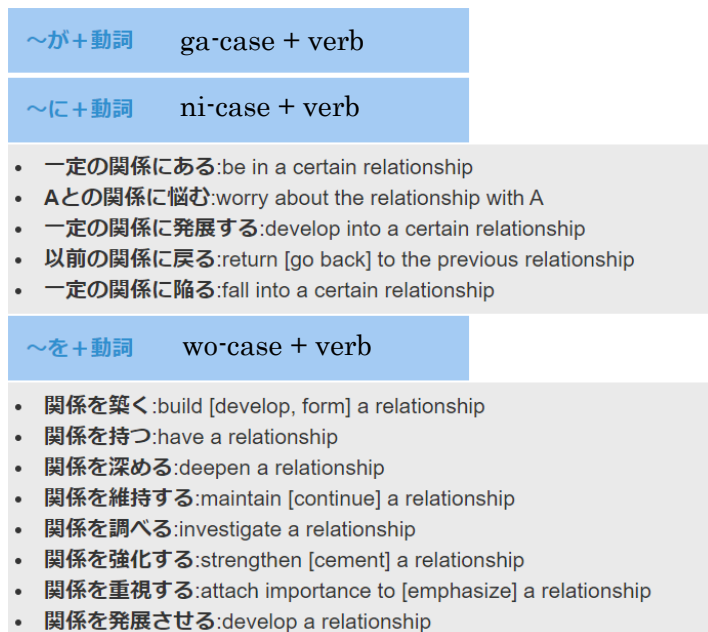


Figure 1: A sample screenshot of the dictionary

Since the semantic cases serve as signposts to the target translation, this way of presentation, we believe, is more user-friendly than the ones in existing Japanese-English dictionaries. It should be also noted that the collocations are sometimes expanded to a phrase in order to provide more natural

and useful wordings. For example, “*kankei ni modoru*” (return to a relationship) is not immediately usable. Therefore, “*izenno*” (previous) is added, and the combination of “*kankei*” and “*modoru*” is finally expanded to a phrase “*izenno kankei ni modoru*” (return to the previous relationship).

Another potential advantage of this approach is overcoming language gaps. A Japanese word “*mondai*”, for example, can be translated into English either as “problem” or “question” depending on the context, but when to use which may remain unclear for a learner. If a bilingual collocations dictionary is available, it would give him/her a clear answer because it provides translations for the combinations of words, not just the translations for each single word. This results in translating “*mondai wo toku*” and “*mondai ni kotaeru*” into “solve a problem” and “answer a question” respectively, and hence a user can immediately understand which English word to use.

4 Concluding remarks

Features of our new collocations dictionary can be summarized as follows: it is an online L1-L2 type bilingual collocations dictionary based on corpora to improve the coverage, whose entries are organized based on semantic cases to improve searchability. It is currently being constructed, and the present paper thus remains as a progress report, but we hope to receive feedback from researchers in various fields. We also plan to conduct experiments to examine the usefulness of our dictionary for writing tasks after the release of the prototype edition.

Acknowledgement

This study is supported by JSPS KAKENHI Grant Number JP18H00693.

References

- Chen, Y. H., & P. Baker (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- East, M. (2008). *Dictionary use in foreign language writing exams: Impact and implications*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ellis, N. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Oxford: Blackwell, 63–103.
- Horsfall, P. (1997). Dictionary skills in MFL 11–16. *Language Learning Journal*, 15, 3–9.
- Laufer, B. (1995). A case for semi-bilingual dictionary for production purposes. *Kernerman Dictionary News*, 3.
- Laufer, B., & T. Levitzky-Aviad (2006). Examining the effectiveness of “bilingual dictionary plus”—A dictionary for production in a foreign language. *International Journal of Lexicography*, 19(2), 135–155.
- Thompson, G. (1987). Using bilingual dictionaries. *ELT Journal*, 41(4), 282–286.