

Japanese One-Word Sentences and their English Translations in a Parallel Corpus

大矢 政徳
Masanori OYA

Keywords : one-word sentences, Japanese-English parallel corpus, typed-dependency trees

キーワード : 一語文、日英対訳コーパス、タイプ分けされた依存関係木

1. Introduction

This paper investigates the types and properties of Japanese one-word sentences with respect to their English translations in an English-Japanese parallel corpus. The structure of this paper is as follows: Section 2 introduces the issue of correspondences between one-word sentences in Japanese and English with respect to their forms and meanings. Section 3 refers to some of the related research on one-word sentences. Sections 4, 5, and 6 explain the data used in this study, how these data were analyzed, and the description of the result, respectively. Section 7 covers the discussion of some of the correspondence patterns between Japanese one-word sentences and their English translations, and Section 8 concludes this paper.

2. One-word sentences in Japanese and English

One-word sentences are often found in spoken varieties of Japanese, and yet we find some instances of one-word sentences in written varieties too. They belong to different parts of speech (e.g., noun, verb, adjective) and have different discourse functions (e.g., directing the listener's attention to an entity, describing a property of an entity). The syntactic structure of their English translations is determined by two factors: the discourse functions of the Japanese one-word sentences on the one hand, and the parts of speech of the English translation equivalents of the Japanese word on the other. For example, the Japanese one-word sentence, *Tori* 'a bird' has a number of different translation possibilities. When we are to translate such Japanese one-word sentences into English, we need to take into consideration (1) the function of the sentence in the original discourse and (2) the

elements that should be added to the English translation. For the example above, if the function of this one-word sentence is to direct the listener's attention to a particular bird, it would be best translated into English as the imperative sentence *Look at the bird*. However, if the same one-word sentence functions as the answer to the question *Nani wo mitano?* 'What did you see,' it would be best translated into English by the declarative sentence *I saw a bird*.

As these simple examples indicate, the discourse function of a one-word Japanese sentence determines the elements to be added to the English equivalent of the single Japanese word, and therefore, the syntactic structure of the full English translation. Consider the Japanese one-word sentence *Tori* again. The syntactic dependency tree for this sentence is shown in Figure 1 below.¹⁾ The word *Tori* depends on the abstract word *Root*, which represents the discourse, and the dependency type is "ROOT," an all-inclusive abstract dependency type. This typed-dependency tree represents the dependency type and the fact that the word *Tori* depends on the discourse.

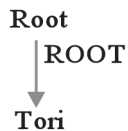


Figure 1. The syntactic dependency tree for *Tori*.

In the first English translation mentioned above, namely *Look at the bird*, the elements to be added are the verb *look*, the preposition *at*, and the definite determiner *the*. The syntactic dependency tree for this sentence is shown below.

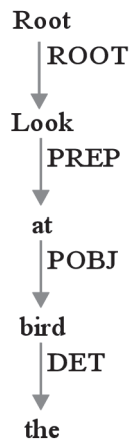


Figure 2. The typed-dependency tree for *Look at the bird*.

In the second English translation mentioned above, namely *I saw a bird*, the elements to be added are the pronoun *I*, the verb *saw*, and the indefinite determiner *a*.

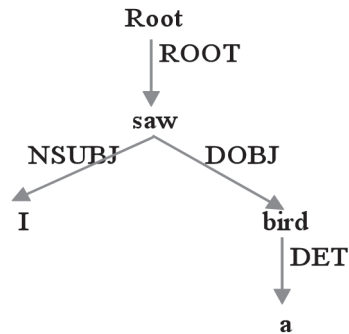


Figure 3. The typed-dependency tree for *I saw a bird*.

The patterns in which Japanese one-word sentences are translated into English sentences are the simplest patterns of Japanese-English correspondence. If we are to deal with a complex problem, we need to divide the problem into smaller and simpler parts, and deal with each of them in turn. Similarly, when we are to deal with the correspondence between English and Japanese, Japanese one-word sentences are a good starting point.

It must be pointed out that these examples illustrate only two of the English-Japanese correspondence patterns, and there are other possibilities. An English-Japanese parallel corpus provides us with the patterns according to which Japanese sentences are translated into English. By focusing on how Japanese one-word sentences are translated into English sentences in a parallel corpus, we can observe the various patterns of their correspondence, not just in a small number of examples, but on a larger scale. These patterns are expected to give us interesting insight into the difference between these two languages, and will be applied into the other fields of research. For example, the insight into pattern matching of Japanese one-word sentence and its English translation will contribute to machine translation, and the insight into how Japanese one-word sentences are translated into English sentences in a given context will contribute to teaching English as L2 to native speakers of Japanese.

3. Related research

According to Ishikawa (2008), it was Yamada (1936, p. 1116) who proposed the idea of one-word sentences, arguing that a one-word sentence “is a sentence composed of one word, in which an idea suddenly expresses itself while it has not fully developed.” Onoe (2006) argues that a single noun uttered without reference to any context has the function of either accepting or wishing for the existence of the referent of the noun, and one-word sentences in different contextual situations contain different sentential meanings. He refers to Onoe (2001), in which the usages of one-word sentences are categorized. Ishikawa (2008) compared the Japanese one-word sentences in daily conversation with those in TV conversation and found some differences between them in terms of parts of speech and functions.

4. Data

The corpus used in this study is the Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles ver.2.01 (NICT 2011). This corpus contains about 500,000 Japanese-English pairs of sentences on topics related to Kyoto, which were translated manually. This corpus is divided into sixteen subcorpora according to their topics (such as religion, famous people, or famous buildings). For this study, 5,000 sentences in the subcorpus on Buddhism (26890 sentences) were chosen randomly.²⁾

5. Analysis

The 5,000 English and Japanese sentences in the subset of the parallel corpus explained in the previous section (henceforth, subset 5000) were parsed by two parsers: Stanford Parser (de Marneffe and Manning 2008, 2011) for the English sentences, and Juman/KNP (Kurohashi and Nagao 1992, 1994, 1998; Kawahara and Kurohashi 2007) for the Japanese sentences. In order to focus on sentences and to exclude sentence fragments, only the Japanese sentences that end with the sentence-ending mark (。) and their English translation were parsed; this process was facilitated by my original Ruby script. In this study, we assume that the sentencehood of Japanese sentences should be ensured by the fact that they end with the Japanese sentence-ending punctuation (。). If they do not end with this punctuation, they are considered sentence fragments. We also assume that a Japanese one-word sentence contains neither case particles (e.g., *ga*, *wo*, *no*) nor phrase-ending punctuation (、). From the parse output, the one-word Japanese sentences and their English translations were extracted using my original Ruby script and categorized manually according to the syntactic patterns of these English translations. The patterns are based on the part of

speech of the English word that corresponds to the Japanese word and the syntactic environment in which the English word appears. The categories found in subset 5000 will be shown in the next section. In the first example above, namely the pair of the Japanese *Tori* and the English *Look at the bird*, the Japanese word *Tori* corresponds to the English word *bird*, which is the object of the preposition *at*. Therefore, this English sentence is categorized as “prepositional complement.” In the second example above, namely the pair of the Japanese *Tori* and the English *I saw a bird*, the Japanese word *Tori* corresponds to the English word *bird*, which is the object of the verb *saw*. Therefore, this English sentence is categorized as “verbal complement.” Unlike the extraction of one-word Japanese sentences from the parse output of subset 5000, the process of syntactic categorization is yet to be automated.

6. Result

Out of the 5,000 sentences in subset 5000, 105 Japanese sentences are one-word sentences. Table 1 presents the syntactic patterns and the numbers of their instances among the English translations of the Japanese one-word sentences in subset 5000.

categories	Instances
adjectival complement	3
adjectival complement with prepositional phrase	8
adjective	1
adjective with prepositional phrase	6
noun	10
noun with prepositional phrase	4
nouns	4
prepositional complement	23
subject	3
subject and nominal complement	1
subject and verbal complement	2
verb	1
verb (imperative)	2
verb and adverb	1
verb with prepositional phrase	2
verbal complement	4
verbal complement with adverb	5
verbal complement with prepositional phrase	24
vocative	1
	105

Table 1. The syntactic patterns and the numbers of their instances among the English translations

of the Japanese one-word sentences in subset 5000

The English sentences that fall into the following categories are verbless: “adjective (e.g., *Utsukushii* ‘Beautiful!’),” “adjective with prepositional phrase,”³⁾ “noun (e.g., *Kokuhou* ‘A national treasure’),” “noun with prepositional phrase,” “nouns,” or “vocative”; the English sentences that fall into all the other categories are not verbless. For example, the category “adjectival complement” comprises cases where the Japanese word is translated into an English adjective, and it functions as a complement to a verb (in all three instances of this category in our data, the verb is *is*).

7. Discussion

What is common among the English sentences in all these categories is that they make explicit what is not expressed in the original Japanese sentences. This information can be identified through the context in which each of the Japanese one-word sentences is used. In this respect, it must be pointed out that it is not only possible, but also necessary, to take into consideration the contextual information which can be identified by examining the sentences before and after the Japanese one-word sentence and its English translation. The parallel corpus used in this study allows us to consider this type of information; for each of the Japanese one-word sentences, we can look at the sentences before or after it, so that we can identify what is not expressed in the original Japanese sentence, and what is made explicit in its English translation. In addition to this, it is possible to say that a Japanese one-word sentence used in a context of a certain type is translated into an English sentence of a certain syntactic structure, such as those summarized in the list of categories above.

Let us consider the top four of the most frequent categories in subset 5000: “verbal complement with prepositional phrase,” “prepositional complement,” “adjectival complement with prepositional phrase,” and “noun.” The fact that they appear frequently in subset 5000 does not necessarily indicate that they also appear frequently in a larger corpus in different genres; they are chosen here for a preliminary investigation into the variety of structural correspondences between Japanese typed-dependency trees and their English translations on a larger scale.

7.1 Verbal complement with prepositional phrase

First, let us consider the category “verbal complement with prepositional phrase,” where a Japanese one-word sentence is translated into an English sentence in which the Japanese word is translated into an English word that functions as a verbal complement that is accompanied by a prepositional phrase or prepositional phrases. Out of the twenty-four

instances in subset 5000, thirteen sentences indicate the titles of notable people in Buddhism, such as the chief priest or chief abbot, and they constitute parts of texts that introduce these people. For example, the Japanese one-word sentence (1) appears in a text that introduces a notable imperial prince. The English gloss is shown in the second line of the example.

- (1)
Tendai-zasu. (天台座主。)
Tendai sect-the head priest

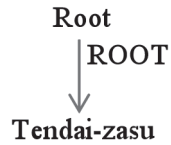
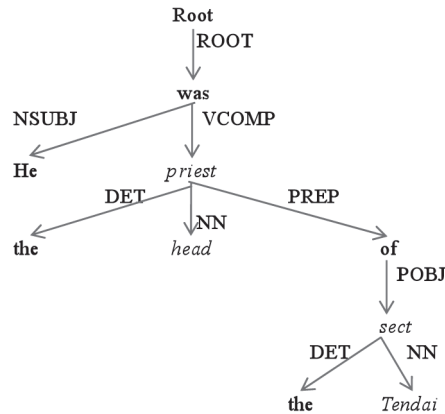
In such a context, the English translation of (1) inevitably includes the pronoun *he* and the copula *was*, which indicates that the person is dead, as shown in (2) below.

- (2)
He was the head priest of the Tendai sect.

All the other eleven sentences belonging to this category are part of an introduction of a history book, describing the content of each of the volumes in the book (e.g., *Nangokumonka* (南獄門下) 'Volume 6 covers the disciples of Nangoku'). The Japanese words in these sentences are all translated into English noun phrases which are the object of the verb *covers*, whose subject is *Volume x*.

As we can see from the example above, Japanese one-word sentences in this category introduce a detail of the property of the discourse topic of the text concisely. What is not expressed in this type of Japanese one-word sentences is the pronoun that refers to the discourse topic and the verbal element that indicates the tense, aspect, and mood (henceforth, TAM) properties.

The typed-dependency trees for (1) and (2) are shown below.

Figure 4. The typed-dependency tree for *Tendai-zasu*.Figure 5. The typed-dependency tree for *He was the head priest of the Tendai sect*.

7.2 Prepositional complement

Second, let us consider the category “prepositional complement,” where a Japanese one-word sentence is translated into an English sentence in which the Japanese word is translated into an English word that functions as the complement to a preposition. Out of the twenty-three instances in subset 5000, twelve of the Japanese one-word sentences are translated into English nouns or noun phrases that are complements to the preposition *of*, seven are complements to the preposition *as*, and the rest are complements to *in or from*. In terms of their functions, eight of them indicate titles of notable people. For example, the Japanese one-word sentence in (3) below appears in a context where a notable priest is being introduced.

(3)

Hakushaku. (伯爵。)

“(the title of) count.”

In such a context, the English sentence inevitably includes the pronoun *he* and the

verbal phrase *held the title of*, as shown in (4) below.

(4)

He held the title of count.

As for the other instances, six of them indicate the age of death of a notable person (e.g., *Rokujyu-san-sai* 'He died at the age of 63'). Four of them indicate the number of volumes of books (e.g., *Jukkan* 'It comes in ten volumes'). Two of them indicate another name of an entity (e.g., *Haibutsu-kishaku-tomo* 'It is also written as 廃仏毀釈 in Chinese characters'). The rest indicate the birthplace of a person (*Hyuganokuni-shussin* 'He was from Hyuga Province'), the school a person graduated from (*Nara-kenritsu-kooriyama-engei-koukou-sotsu* 'He graduated from Nara Prefectural School of Horticulture'), and that the entity is one of the important cultural properties (*Juyo-bunka-zai* 'It is one of the important cultural properties').

Japanese one-word sentences in this category also introduce a detail of the property of the discourse topic of the text concisely. What should be added in the English translation of this type of Japanese one-word sentences is the pronoun that refers to the discourse topic, the preposition, and the element that the preposition depends on. Which preposition should be chosen depends on the meaning of the word, and the context in which it appears. For example, if the word is a place name and the text describes someone, the preposition will probably be *from*.

7.3 Adjectival complement with prepositional phrase

Third, let us move on to the category "adjectival complement with prepositional phrase." A Japanese one-word sentence in this category is translated into an English sentence in which the Japanese word is translated into an English adjective that functions as the complement to the predicate. Out of the eight instances in subset 5000, six use "was born in," and the other two use "was compiled by," "is kept in." For example, the Japanese one-word sentence in (5) appears in a text that introduces a notable person.

(5)

Kyushu-shussin. (九州出身。)

Kyushu born

This sentence is translated by the English sentence in (6) below.

(6)

He was born in the Kyushu region.

Some of the one-word sentences in this category raise the issue of morphological analysis of the Japanese language because they can be analyzed not as one-word sentences but as multiple-word sentences. For example, consider the Japanese sentence in (7), which introduces the birthplace of a notable person, and its English translation (8).

(7)

Yamanashiken-kitakomagun-nagasakacho-shussin (山梨県北巨摩郡長坂町出身。)

(8)

He was born in Nagasakacho, Kitakoma district, Yamanashi Prefecture.

The problem is to determine whether sentences like (7) should be considered one-word sentences. As mentioned in Section 5, we assume that Japanese sentences end with the sentence-ending punctuation (。). We also assume that a Japanese one-word sentences contains neither case particles nor phrase-ending punctuation (、). With these assumptions, the sentence (7) is considered one-word sentence. To be precise, this type of one-word sentences should be called *one-compound-noun sentences*. What should be added to the English translation of this type of Japanese one-word sentences is the pronoun that refers to the discourse topic, and the TAM properties.

7.4 Noun

Lastly, let us finish this account with the category “noun,” where a Japanese one-word sentence is translated into an English verbless sentence in which a single noun phrase constitutes the whole sentence (or sentence fragment). These translation pairs are simpler than the others in terms of structural correspondence. Japanese one-word sentences in this category also introduce a detail of the property of the discourse topic of the text concisely, as is the case in other categories, and nothing needs to be added to their English translations.

8. Conclusion

This paper investigated the types and properties of Japanese one-word sentences with respect to their English translations in an English-Japanese parallel corpus. Since the genre of the texts used in this study is explanatory text (Wikipedia articles on Buddhism), the four

patterns (“verbal complement with prepositional phrase,” “adjectival complement with prepositional phrase,” “prepositional complement,” and “noun”) discussed in this paper may not be necessarily representative of the correspondence between Japanese one-word sentences and their English translations in other genres of text. For example, only one instance of the vocative usage is found in the data, yet this type of one-word sentences must be frequently used in colloquial styles of Japanese, as investigated in Ishikawa (2008). The issue of investigating the correspondence patterns between them in terms of forms and meanings must be addressed more extensively using texts of other genres, in order to apply the insight into other fields of research mentioned in section 2, which will be the objective of future research.

[Notes]

- 1) The format of dependency trees in this study is based on Stanford Dependencies (de Marneffe and Manning 2006). The dependency types used in this paper is as follows: DET for determiner; DOBJ for direct object; NN for noun compound; NSUBJ for nominal subject; PREP for preposition; POBJ for prepositional object; and VCOMP for verbal complement. The definition of each dependency type is discussed in Oya (2014a).
- 2) This subset of the corpus is the same as the one used in Oya (2014b), for a different purpose.
- 3) In this study, the participles are treated as adjectives.

[References]

- De Marneffe, M. C., MacCartney, B. & Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *LREC 2006*. Retrieved on June 23, 2012, from http://nlp.stanford.edu/manning/papers/LREC_2.pdf.
- De Marneffe, M. C., & Manning C. D. (2008). The Stanford typed dependencies representation. *COLING Workshop on Cross-framework and Cross-domain Parser Evaluation*. Retrieved on June 23, 2012, from <http://nlp.stanford.edu/pubs/dependencies-coling08.pdf>.
- De Marneffe, M. C., & Manning, C. D. (2011). *Stanford Typed Dependency Manual*. Retrieved June 21, 2012, from http://nlp.stanford.edu/software/dependencies_manual.pdf.
- Ishikawa, S. (2008). Hanashi kotobano ichigobunto danwa shiryouno seikaku [Relation between One-word-sentence in Spoken language and the Sort of Conversation material]. *Bulletin of the Graduate Division of Literature of Waseda University*, vol. 3, pp. 5–16.
- Kawahara, D. & Kurohashi, S. (2007). Daikibokakufurameni motozuku koubun kakukaisekino tougouteki kakuritudoderu [An integrated probabilistic model for syntactic and case analyses based on a large-scale case frames]. *Natural Language Processing* vol. 14, no. 4, pp. 67–81.
- Kurohashi, S. & Nagao, M. (1992). A method for analyzing conjunctive structures in Japanese. *Journal of Information Processing Society of Japan*, vol. 33, no. 8, pp. 1022–1031.
- Kurohashi, S. & Nagao, M. (1994). A syntactic analysis method of long Japanese sentences based on

- coordinate structure detection. *Journal of Natural Language Processing* vol. 1, no. 1, pp. 35–57.
- Kurohashi, S. & Nagao, M. (1998). Building a Japanese parsed corpus while improving the parsing system. *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pp. 719–724.
- Onoe, K. (2001). *Bunpo to Imi* [Grammar and its meaning]. Tokyo: Kuroshio Publishers.
- Onoe, K. (2006). Sonzai kikyū to shōnin-shūgo jutsugo hasseino gennri [Wishing for and accepting the existencethe principle of the generation of subjects and predicates]. *Kokugo to Kokubungaku*, pp. 1–13.
- Oya, M. (2014a). *A Study of Syntactic Typed-Dependency Trees for English and Japanese and Graph-centrality Measures*. Doctoral dissertation, Waseda University.
- Oya, M. (2014b). An English-Japanese bilingual corpus-based comparison of their syntactic dependency structures. *Proceedings of the 19th Conference of Pan-Pacific Association of Applied Linguistics*, pp.69–70.
- Yamada, Y. (1936). *Nihon bunpogaku gairon* [Introduction to Japanese Grammar]. Tokyo: Hobunkan.

(平成26年11月4日受理)