Semantic Annotations in Japanese FrameNet:
Comparing Frames in Japanese and English

Kyoko Hirose Ohara

Keio University

4-4-1 Hiyoshi, Kohoku-ku, Yokohama City 223-8521, Japan

E-mail: ohara@hc.st.keio.ac.jp

Abstract

Since 2008, the Japanese FrameNet (JFN, http://jfn.st.hc.keio.ac.jp/) project has been annotating the Balanced Corpus of Contemporary Written Japanese (BCCWJ), the first such corpus, officially released in October 2011. This paper reports annotation results of the book genre of BCCWJ (Ohara 2011, Ohara, Saito, Fujii & Sato 2011). Comparing the semantic frames needed to annotate BCCWJ with those that the FrameNet (FN) project (Fillmore and Baker 2009, Fillmore 2006) already has defined revealed that: 1) differences in the Japanese and English semantic frames often concern different perspectives and different lexical aspects exhibited by the two lexicons; and 2) in most of the cases where JFN defined new semantic frame for a word, the frame did not involve culture-specific scenes. We investigated the extent to which existing semantic frames originally defined for analyzing English words were used, annotating 810 sentences of the so-called core data of the book genre of BCCWJ. In the 810 sentences we were able to assign semantic frames to approximately 4000 words, although we could not assign any to 587 words. That is, of all the LUs in the sentences, we were able to identify semantic frames to about 87 per cent of them. In other words, the semantic frames already defined in FN for English could be used for 87 per cent of the Japanese LUs.

Keywords: Frame Semantics, Semantic Annotation, Japanese

1. Introduction

Since 2008, the Japanese FrameNet (JFN, http://jfn.st.hc.keio.ac.jp/) project has been annotating the Balanced Corpus of Contemporary Written Japanese (BCCWJ), the first such corpus, officially released in October 2011. This paper reports annotation results of the book genre of BCCWJ (Ohara 2011, Ohara, Saito, Fujii & Sato 2011). Comparing the semantic frames needed to annotate BCCWJ with those that the FrameNet (FN) project (Fillmore and Baker 2009, Fillmore 2006) already has defined revealed that: 1) differences in the Japanese and English semantic frames often concern different perspectives and different lexical aspects exhibited by the two lexicons; and 2) in most of the cases where JFN defined new semantic frame for a word, the frame did not involve culture-specific scenes.

The JFN project is building a lexical resource of Japanese by annotating corpus data using the framework of frame semantics (e.g. Fillmore 1987). Since the start of the project, the project has worked closely with the FN project (http://framenet.icsi.berkeley.edu/, cf. Hasegawa et al. 2010). The JFN database and software have been imported from the FN project and the set of semantic frames in JFN is basically the same as the one in FN. Additionally, JFN pays very close attention to contrasts between Japanese and English (cf. Ohara and Sato 2010, Lönneker-Rodman 2007).

Semantic annotation based on frame semantics involves identifying the structure of the knowledge necessary for producing and understanding speech acts (semantic frames) and its semantic components (frame elements, hereafter FEs). FEs are defined relative to semantic frames, so they are much more fine-grained than often used semantic roles (e.g. Agent, Instrument, Object)

based on so-called case frames[1]. In frame semantics, lexical units (LUs) evoke semantic frames, which are related to one another through frame-to-frame relations (Ruppenhofer et al. 2010).

There are two modes of semantic annotation in the JFN project: one is called lexical annotation mode and the other is called full text annotation mode. The result of the JFN lexical annotation will be reported in Section 2 and that of the JFN full text annotation will be discussed in Section 3, with an emphasis on comparison and contrast between Japanese and English semantic frames. Finally, Section 4 will summarize the result and state the implications for future research.

2. Investigating Semantic Frames in JFN Lexical Annotation

For lexical annotation, the JFN project currently focuses on annotating the most frequently occurring verbs, adjectives, adverbials, and event nouns in BCCWJ. In this annotation mode, the annotator chooses a word and annotates selected sentences in which it occurs. The current JFN lexical annotation mode involves the following: 1) decide on the word to be annotated; 2) identify the semantic frames (the structure of knowledge needed for producing and understanding speech acts) that the word in question evokes. Here, the annotator looks at the semantic frames that the FN project already has defined for annotating the English lexicon. When no existing semantic frames in FN evoke the word in question, the annotator determines whether the semantic

---

[1] VerbNet (http://verbs.colorado.edu/~mpalmer/projects/verbnet.html) employs abstract semantic roles. See Matsubayashi et al. (2010) for an attempt to generalize FrameNet FEs and PropBank semantic roles using the abstract semantic roles in VerbNet.

frame yet to be defined is necessary for annotating English words as well or whether it is necessary for annotating Japanese words only; 3) Select sentences from BCCWJ that contain the word in question using a concordance called JFN-KWIC. When choosing example sentences for annotation, the annotator also takes collocation and valence patterns into account; 4) Annotate the relevant phrases with frame element (FE), phrase type (PT), and grammatical function (GF) in the selected sentences. Figure 1 is a screen shot of the JFN lexical annotation report.
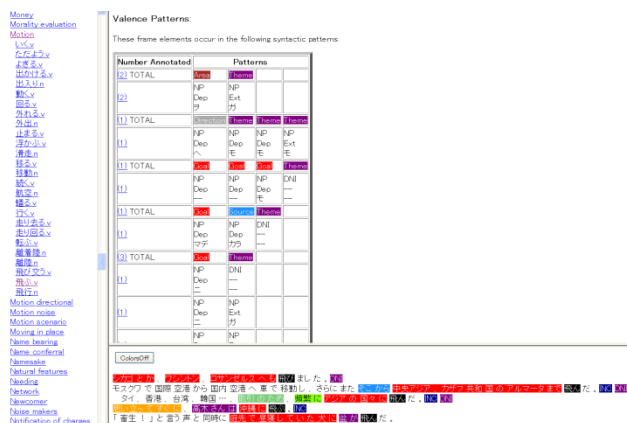

Figure 1: JFN Lexical Annotation Report

The JFN project is investigating the extent to which semantic frames defined in FN for analyzing the English lexicon are appropriate for describing lexical meanings of the Japanese lexicon as well. Therefore, in JFN lexical annotation, it is important to ensure the necessity of defining semantic frames and FEs specifically for Japanese.

Consider the pair of sentences in (1), which pertains to a contrast between an intransitive/inchoative verb (1a) and a transitive verb (1b). (1a) depicts a scene in which petals of cherry blossoms get scattered. The intransitive verb *tiru* 'get scattered' in (1a) is an inchoative verb used to describe particles or small objects falling. It is difficult to find a semantic frame in the current FN database with the meaning. As for the morphologically-related transitive counterpart *tirasu* 'scatter' in (1b), on the other hand, we assume that the Dispersal frame is involved, since the corresponding English verb *scatter* evokes the Dispersal frame (defined in FN as "an AGENT or a CAUSE disperses or scatters INDIVIDUALS from the SOURCE, a relatively confined space, to a the GOAL_AREA, a broader space"). The only existing semantic frame that seems relevant to the intransitive verb *tiru* 'be scattered' in (1a) is the Motion frame (defined as "some entity (THEME) starts out in one place (SOURCE) and ends up in some other place (GOAL), having covered some space between the two PATHs"), which pertains to the background knowledge of a very general situation involving motion. In order to accurately describe the contrast between the intransitive *tiru* 'get scattered' and the transitive *tirasu* 'scatter', it might be worthwhile defining the Inchoative_Dispersal frame, which

has to do with the situation in which INDIVIDUALS get scattered from the SOURCE to the GOAL_AREA, in a downward movement.

(1) a. sakura   no   hanabira ga
cherry.blossom GEN petals   NOM
<u>tiru</u> Motion
be.scattered
'Petals of cherry blossoms get scattered.'

   b. sakura   no   hanabira o
cherry.blossom GEN petals   ACC
<u>tirasu</u> Dispersal
scatter
'(Somebody) scatters petals of cherry blossoms.'

(2) is another example of a contrast between an intransitive/inchoative verb *keisi suru* 'die of capital punishment' in (2a), and a transitive verb *syokei suru* 'execute' in (2b). It is possible to assume that the transitive verb *syokei suru* 'execute' in (2b) evokes the Execution frame ("An EXECUTIONER punishes an individual (EXECUTED) with death as a consequence of some action of the Evaluee (the REASON)"). However, as for the intransitive verb in (2a), the only existing frame which seems relevant to it is the Death frame ("The words in this frame describe the death of a PROTAGONIST"), which has to do with a general background knowledge pertaining to death. In order to describe the background knowledge that Japanese speakers have which enable them to understand the meaning of the verb *keisi suru*, it is necessary to define the Die_of_Execution frame, which involves the death as a result of an execution.

(2) a. si.kei   syuu   ga
death.penalty prisoner NOM
<u>kei.si   suru</u> Death
penalty.death do
'A death-row prisoner dies of capital punishment.'

   b. si.kei   syuu   o
death.penalty prisoner ACC
<u>syokei   suru</u> Execution
execute do
'(An executioner) puts a death-row prisoner to death.'

There are many other pairs of intransitive/inchoative and transitive verbs in Japanese, which are often morphologically related. We have determined, however, that many of the existing semantic frames originally defined for analyzing the semantics of English words, involve the transitive perspective rather than the intransitive perspective. Few cases exist in which FN semantic frames are defined from both intransitive/inchoative and transitive perspectives. Exceptions include Becoming_detached (involving either of the two situations: a scene in which one thing comes to be physically detached from something else; or a

1560

scene in which two things come to be disconnected from each other) and `Detaching` (defined for either of the following two situations: a scene in which somebody causes one thing to be physically detached from something else; or a scene in which somebody causes two things to be disconnected). One might think that the `Fullness` frame (A CONTAINER is in a state of fullness/emptiness with respect to some CONTENTS) and the `Filling` frame (relating to filling containers and covering areas with some thing, things or substance, the THEME) may be another example of pairs of frames from intransitive/inchoative and transitive perspectives. However, that is not the case. The `Fullness` frame has to do with the intransitive perspective and the stative aspect, not with the inchoative aspect[2].

There are many intransitive verbs in Japanese and their transitive counterparts are often derived by suffixing a causative morpheme. On the other hand, as noted above, many existing semantic frames have the transitive perspective, rather than the intransitive[3]. In addition, whereas "Causative Of" is one of the current 9 frame to frame relations, "Intransitive Of" frame to frame relation is yet to be defined. In other words, existing semantic frames may assume perspectives and lexical aspects (*aktionsart*) of English words, which are not necessarily the same as those of Japanese words. It is thus necessary to take this into account in the lexical annotation process, especially for frame identification and frame definition.

3. Investigating Semantic Frames in JFN Full Text Annotation

For full text annotation, JFN annotates all the LUs in a text (excluding named entities) which evoke semantic frames. The merits of full text annotation include the following. First, a semantically-tagged Japanese corpus based on frame semantics can be achieved. At the moment, there are not many semantically-tagged corpora of Japanese available. Secondly, discovering the distributions of semantic frames (i.e. senses), valence patterns, and zero pronouns becomes possible with full text annotation. Figure 2 is a screen shot of the JFN full text annotation report.

---

[2] It is worth noting that as for situations having to do with detachment, there is a three-way distinction in existing semantic frames. That is, in addition to the `Becoming_detached` frame (defined from the intransitive perspective and the inchoative aspect) and `Detaching` frame (defined from the transitive perspective) mentioned above, there exists the `Being_detached` frame ("An ITEM is detached from a SOURCE, or ITEMS are detached from each other"), which involves the stative aspect in addition to the intransitive perspective.

[3] Few existing FN semantic frames defined from the intransitive perspective include the following 10 semantic frames: `Become_silent`, `Become_triggered`, `Becoming`, `Becoming_a_member`, `Becoming_aware`, `Becoming_detached`, `Becoming_dry`, `Becoming_separated`, `Becoming_visible`, and `Expansion` frames (As of the 13th of March, 2012).

Figure 2: JFN Full Text Annotation Report

The JFN project has been annotating the book genre of BCCWJ in full text annotation mode. We investigated the extent to which existing semantic frames originally defined for analyzing English words were used (cf. Ohara 2011). We examined the so-called core data of the book genre of BCCWJ. There were 81 files and we annotated the first 10 sentences of each file. In the 810 sentences we were able to assign semantic frames to approximately 4000 words, although we could not assign any to 587 words. That is, of all the LUs in the sentences, we were able to identify semantic frames to about 87 per cent of them. In other words, the semantic frames already defined in FN for English could be used for 87 per cent of the Japanese LUs. In calculating the ratio, the number of tokens rather than the number of types was used. Example (3) shows the LUs to which we could not assign an existing FN semantic frame.

(3) Examples of the LUs in the book genre of BCCWJ, to which no semantic frame has been assigned
  a. Adjective
     arai 'coarse'
  b. Conjunction
     dakara 'therefore', *sikasi* 'but', *naraba* 'then', *sunawati* 'thus'
  c. Adjectival noun
     kooiteki 'favorable', *toozen* 'naturally', *noroma* 'stupid'
  d. Verb
     asobu 'play', *muku* 'face', *simeru* 'make up, take up', *ki o tukeru* 'be careful'
  e. Adverb
     sikkari 'firmly', *tatoeba* 'for example', *ippan ni* 'in general'
  f. Event noun
     otukai 'errand', *taiken* 'experience', *tuukoo* 'crossing', *syuppan* 'publication'
  g. Noun
     kami 'god', *gangu* 'toy', *tan'i* 'unit', *wariai* 'ratio', *inu* 'dog', *tatami* 'straw mat', *syoozi* 'sliding paper', *husuma* 'sliding door', *kyookaku* 'knight of the town'

One of the reasons why there are no appropriate semantic frames for the conjunctions in (3b) and the nouns such as *kami* 'god', *gangu* 'toy', *tan'i* 'unit', *wariai* 'ratio', *inu* 'dog' in (3g)[4] is that so far the FN project has been annotating verbs, adjectives, and event nouns but not

---

[4] For the other nouns listed in (3g), see below.

conjunctions and nouns (For the other nouns listed in (3g), see below). Also, since the conjunctions express relations between propositions, it may be difficult to describe their meanings with respect to various participants in situations, i.e. semantic frames.

When entirely new frames are needed for Japanese, they are often needed for English as well. Sometimes there is one to one correspondence between frames needed for a Japanese word and its English counterpart. Examples include frames having to do with *asobu*.v 'play', *muku*.v 'face', *ki o tukeru*.v 'be careful' in (3d), and *otukai*.n 'errand' in (3f). At other times, there are more complex correspondences between frames needed for Japanese words and their English counterparts (cf. Ohara 2009). For instance, there currently exists no semantic frame for *simeru*.v listed in (3d). The Japanese verb *simeru*.v corresponds not only to *make up* or *account for* as in 'Google makes up only 3% of all advertising revenue' but also to *take up* as in 'My kids take up my time'. In other words, when defining the frames needed for *simeru*.v, it is necessary to define not only the frame for *make up/account for* and but also the one for *take up*.

On the other hand, very few frames actually involved Japanese culturally specific scenes, shown with the underlining in (3g). Examples include nouns such as *tatami* "straw mat", *syoozi* "sliding paper", *husuma* "sliding door", *kyookaku* "knight of the town", which refer to various elements that concern the Japanese culture.

4. Conclusion

This paper discussed the results of JFN annotation of the book genre in BCCWJ. Comparing the semantic frames needed to annotate Japanese words with those that the FN project already has defined in, the paper showed that: 1) the perspectives and *aktionsart*s of English words reflected in existing FN semantic frames may be different from those of Japanese words; and 2) in most cases where JFN defined a new semantic frame for a Japanese word, the frame did not involve culture-specific scenes. The two findings must be taken into account in building a multilingual FrameNet and in using FN and JFN for natural language processing applications.

5. Acknowledgements

6. References

Fillmore, Charles J. (2006). The Articulation of Lexicon and Constructicon. Plenary Lecture at the Fourth International Conference on Construction Grammar (ICCG4). September 3, 2006. The University of Tokyo, Japan.

Fillmore, Charles J. (1987). A private history of the concept "frame". Dirven, Rene and Radden, Gunter. (Eds.), *Concept of Case*. Gunter Narr Verlag, Tubingen. pp.28-36.

Fillmore, Charles J. and Collin Baker (2009). A frames approach to semantic analysis. In Heine, Bernd and Heiko Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*. pp.313-339. Oxford University Press.

Hasegawa, Yoko, Russell Lee-Goldman, Kyoko Hirose Ohara, Seiko Fujii, and Charles J. Fillmore (2010). On expressing measurement and comparison in English and Japanese. In Boas, Hans C. (Ed.), *Contrastive Studies in Construction Grammar*. pp.169–200. Amsterdam: John Benjamins Publishing.

Hasegawa, Yoko, Russell Lee-Goldman, Albert Kong, Kimi Akita (2011). FrameNet as a resource for paraphrase research. *Constructions and Frames* 3(1). 104–127.

Lönneker-Rodman, Birte (2007). Multilinguality and FrameNet. ICSI Technical Report TR-07-001.

Matsubayashi, Yuichiro, Naoaki Okazaki and Jun'ichi Tsujii (2010). Generalization of semantic roles in Automatic Semantic Role Labeling. Natural Language Processing, 17 (4). 59-90 (in Japanese).

Ohara, Kyoko Hirose (2009) Frame-based contrastive lexical semantics in Japanese FrameNet: The case of *risk* and *kakeru*. In Boas, Hans C. (Ed.), *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. pp. 163-182. Berlin & New York: Mouton de Gruyter.

Ohara, Kyoko Hirose (2011) Full Text Annotation in Japanese FrameNet. Proceedings of the 17th annual conference of Language Processing Society. (in Japanese).

Ohara, Kyoko Hirose, Hiroaki Saito, Seiko Fujii, and Hiroaki Sato (2011). The Project Report, Proceedings of the "Japanese Corpus" Project, pp. 231-240 (in Japanese).

Ohara, Kyoko Hirose, Junya Kato, and Hiroaki Saito (2011). The JFN annotation of BCCWJ, Proceedings of the "Japanese Corpus" Project, pp. 513-518 (in Japanese).

Ohara, Kyoko Hirose and Hiroaki Sato (2010). Investigating Japanese FrameNet Data with FrameSQL. Sixth International Conference on Construction Grammar (ICCG-6). Charles University, Prague, Czech Republic. September 5th, 2010.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Jan Scheffczyk (2010). FrameNet II: Extended Theory and Practice. Berkeley, California: International Computer Science Institute.