The Japanese FrameNet Project: An Introduction

**Kyoko Hirose Ohara, Seiko Fujii[*], Toshio Ohori[*], Ryoko Suzuki,
Hiroaki Saito[**], Shun Ishizaki[***]**

Keio University
4-1-1 Hiyoshi, Kohoku-Ku, Yokohama City, 223-8521, Japan
ohara@hc.cc.keio.ac.jp

Abstract

This paper presents an overview of the Japanese FrameNet (JFN) research project, which began in July 2002. The goal of JFN is to create a corpus-based lexicon of Japanese described in terms of frame semantics. An important question being asked by JFN is whether Japanese words can be described in FrameNet style, i.e., along the same lines as English words. This point is illustrated in this paper with an example of preliminary analysis of Japanese motion verbs. The Japanese FrameNet can be described as a lexicographic project with an eye to finding out similarities and differences between Japanese and English pertaining to their lexical and grammatical structures.

1. Introduction

This paper gives an overview of the Japanese FrameNet (hereafter JFN) research project, which was launched in July 2002. JFN tries to create a corpus-based lexicon of Japanese described in terms of frame semantics (Fillmore, 1982).

JFN is headquartered on Hiyoshi Campus of Keio University and includes researchers from Keio University and University of Tokyo. So far, a corpus has been chosen and is being expanded and a tool for extracting data from the corpus has been implemented. A pilot study is being undertaken to analyze motion and communication verbs in Japanese.

The rest of the paper is structured as follows. Section 2 describes the goals of JFN. Section 3 illustrates the corpus and computational tools used in the project. Section 4 gives an example of preliminary analysis of Japanese motion verbs.

2. Project Goals

The ultimate goal of JFN is to produce a FrameNet-style database of Japanese words (cf. Fillmore et al., 2003). The resulting database will thus contain valence descriptions of Japanese words and a collection of annotated corpus attestations. In producing this database we will explore whether Japanese words can be described along the same lines as English words, employing the same frame semantic approach.

In the first phase of the project, which lasts until March 2005, we are trying to build a prototype of such a lexicon, focusing on analyzing and annotating Japanese motion and communication verbs. In the second phase, however, we intend to analyze words in other semantic domains as well.

3. The Corpus and the Tools for Analyzing Japanese

The JFN corpus currently contains 8 million sentences taken from the Mainichi Newspaper (CD-Mainichi Newspaper) and texts taken from novels and essays.

The search tool has been developed in JFN (cf. Ohara et al., 2003). The tool searches for both the root form and conjugated forms of a verb, adjective, or adjectival noun at the same time. Another key feature of the JFN search tool is the fact that it can be used with a dependency structure analyzer called CaboCha. CaboCha was developed at Nara Institute of Science and Technology and it performs morphological analysis as well as syntactic parsing of any Japanese sentence. Although CaboCha sometimes parses colloquial sentences incorrectly, using our search tool together with CaboCha enables us to add any text to our corpus.

Currently there are three display modes in the JFN search tool: the parse tree mode; the morphological analysis mode; and context display mode. Figure 1 shows a snapshot of the parse tree mode:

FIGURE 1. A Screenshot of the parse tree mode of the JFN search tool

The entire screen consists of five windows: the search input window to input a keyword to be searched; the file window to specify file(s) in which a keyword is searched; the KWIC window displaying all the sentences containing the keyword and allowing the user to highlight any sentence by clicking on it; the sentence window showing

the highlighted sentence; and the parse tree window, which displays the tree of a highlighted sentence.

The JFN search tool is written in Ruby script language, and runs on Linux and Solaris operating systems as well as on various Windows platforms. We plan to make it publicly available.

4. Applicability of English-based Frames to Japanese

An important question being asked by JFN is whether Japanese words can be described in FrameNet style, i.e., along the same line with English words. Although our analysis and annotation of corpus sentences are based on relevant frames and frame elements established for English in the English FrameNet, we constantly ask whether it is necessary to establish frames and frame elements separately for Japanese. What follows is a preliminary analysis of Japanese motion verbs. As we have done elsewhere for Japanese communication verbs (ibid.), the analysis below attempts to deal with recognized differences in senses among Japanese motion verbs related in meaning, by refining frame elements already established in the English FrameNet.

Japanese has been characterized as a "path-type verb-framed language," since in order to encode a Path of motion, Japanese employs motion verbs unlike English, which employs prepositions and verb particles for the same purpose (cf. Talmy, 1985; 1991; 2000; Matsumoto, 1997; Kageyama, 2001). Japanese is thus abundant in motion verbs which at the same time describe various paths. *Wataru* 'go across' and *koeru* 'go beyond, go over,' are examples of such motion+path verbs. Both of the verbs can be analyzed as evoking the Path_Shape frame, since the two verbs describe motion in terms of the shape of the Path traversed by the Theme that moves. Thus, *wataru* 'go across' can be used with an accusative-marked direct object NP denoting a Path as in (1a) and (1b) (cf. Kunihiro et al., 1982). In (1a) *kawa* 'river' denotes an area that lies between two points in space, while in (1b) *hasi* 'bridge' refers to a medium or a passage that is constructed between the two points:

(1) a. *nanmin   ga   kawa   o*
         refugees   NOM   river   ACC
         watatta
         went.across
         'The refugees went across (crossed, traversed) the river.'
     b. *nanmin   ga   hasi   o*
         refugees   NOM   bridge   ACC
         watatta
         went.across
         'The refugees crossed the bridge.'

On the other hand, although the verb *koeru* 'go beyond' takes an accusative-marked direct object NP such as *kawa* 'river' in (2a) just like *wataru* 'go across' does, *hasi* 'bridge' typically cannot be its direct object as shown in (2b):

(2) a. *nanmin   ga   kawa   o*
         refugees   NOM   river   ACC
         koeta
         went.beyond

'The refugees went beyond (passed) the river.'
     b. **nanmin   ga   hasi   o*
         refugees   NOM   bridge   ACC
         koeta
         went.beyond
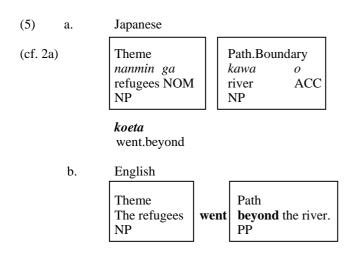   (Intended meaning) 'The refugees passed the bridge.'

Furthermore, *koosaten* 'intersection' can be the direct object of *wataru* as in (3a), but not of *koeru* as shown in (3b). However, when (3b) is used to depict a situation in which the child not only crossed the intersection but also went beyond it, then the sentence becomes acceptable:

(3) a. *kodomo ga   zitensya de*
         child   NOM   bike   INSTR
         koosaten   o   watatta
         intersection ACC   went.across
         'The child crossed the intersection by bike.'
     b. **kodomo ga   zitensya de*
         child   NOM   bike   INSTR
         koosaten   o   koeta
         intersection ACC   went.beyond
         (Unacceptable with the reading) 'The child passed the intersection by bike (and stopped there).'
         (Acceptable with the reading) 'The child went by bike past the intersection.'

It thus seems necessary to identify sub-categories of the frame element Path such as Route and Boundary, in order to describe the different kinds of Path that the two verbs above and others take. That is, *wataru* 'go across' may be described as taking an accusative-marked Route, while *koeru* 'go beyond' may be characterized as taking an accusative-marked Boundary as the direct object. Therefore, the annotations of (1a), (2a), and their equivalents in English would be as follows:

(4) a. Japanese

(cf. 1a)

Theme	Path.Route
nanmin ga	*kawa o*
refugees NOM	river ACC
NP	NP

watatta
went.across

     b. English

Theme		Path
The refugees	**went**	**across** the river.
NP		PP

(5)   a.   Japanese

(cf. 2a)

Theme	Path.Boundary
nanmin ga	*kawa o*
refugees NOM	river ACC
NP	NP

koeta
went.beyond

    b.   English

Theme		Path
The refugees	**went**	**beyond** the river.
NP		PP

   With these contrastive analyses of Japanese and English, we hope to eventually build a bilingual lexicon, to be used by Japanese learners of English as well as by machine translation (cf. Boas, 2002). Such a lexicon will especially be effective as an encoding dictionary for Japanese learners of English.

5. Conclusion

   This paper has outlined the goal, computational environments, and a preliminary analysis of JFN. In conclusion, our current effort can be described as a lexicographic project with an eye to finding out similarities and differences between Japanese and English pertaining to their lexical and grammatical structures.

Aknowledgements

References

Boas, H.C. (2002). Bilingual FrameNet Dictionaries for Machine Translation. In Proceedings of the Third International Conference on Language Resources and Evaluation (LREC2002). Vol. IV. (pp. 1364-1371).

CaboCha
   http://cl.aist-nara.ac.jp/~taku-ku/software/cobocha

CD-Mainichi Newspaper 1995-2002.

Fillmore, C.J. (1982). Frame semantics. In Linguistics in the Morning Calm. (pp. 111-137). Seoul: Hanshin Publishing Co.

Fillmore, C.J., Johnson, C.R., and Petruck, M.R.L. (2003). Background to Framenet, International Journal of Lexicography, Vol. 16.3: 235-250.

Kageyama, T. ed. (2001). Nichi Ei Taishoo Dooshi no Imi to Koobun [Contrastive Analysis of Japanese and English: Semantics of Verbs and Constructions]. Tokyo: Taishukan Shoten.

Kunihiro, T., Shibata, T., Nagashima, Y., Yamada, S., and Asano, Y. (1982). Kotoba no Imi. [Meanings of Words]. Vol.3. Tokyo: Heibonsha.

Matsumoto, Y. (1997). Kuukan Idoo no Gengo Hyoogen to Sono Kakuchoo [Linguistic Expressions of Motion in Space and Their Extensions]. In Tanaka, S. and Matsumoto, Y. Nichi Eigo Hikaku Sensho. Roku: Kuukan to Idoo no Hyoogen [A Series on Contrastive Analysis of Japanese and English, 6: Expressions of Space and Motion]. (pp. 125-230). Tokyo: Kenkyusha.

Ohara, K.H., Fujii, S., Saito, H., Ishizaki, S., Ohori, T., and Suzuki, R. (2003). The Japanese FrameNet Project: A Preliminary Report. In Proceedings of Pacific Association for Computational Linguistics (PACLING'03) (pp. 249-254).

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In Shopen, T. (Ed.), Language typology and syntactic description, Vol. 3, (pp. 57-149). Cambridge: Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. Berkeley Linguistics Society 17 (pp. 480-519).

Talmy, L. (2000). Toward Cognitive Semantics. Cambridge, MIT Press.