Veredas: Frame Semantics and Its Technological Applications
Vol. 17, nº 1, 2013

Toward Constructicon Building for Japanese in Japanese FrameNet

Kyoko Hirose Ohara (Japanese FrameNet / Keio University)

ABSTRACT: This paper discusses Japanese FrameNet and focuses on the kinds of problems that arise when building a FrameNet for a language totally unrelated to English[*]. It discusses differences and similarities in frames between English and Japanese, as exhibited by the current versions of FrameNet and Japanese FrameNet. Through lexicographic and full-text annotation, the Japanese FrameNet project has been led to describe the relations between sentential, lexical, and constructional meanings. The project has thus started to build a "constructicon" of Japanese. Based on examination of to what extent existing FrameNet frames account for full-text annotations of Japanese texts, this paper suggests that the organization of frames for Japanese may be different from that of English. It will furthermore discuss implications of lexicon and constructicon building to Frame Semantics and Construction Grammar.

KEYWORDS: Japanese FrameNet, constructicon, frame-based lexicon, frames, constructions

Introduction

This paper discusses the following three issues concerning the Japanese FrameNet project: 1) difficulties of building Japanese FrameNet while relying on a language resource of English, which is typologically unrelated to Japanese; 2) differences and similarities in the frames and constructions in English and Japanese, as exhibited by the current versions of FrameNet and Japanese FrameNet; 3) implications of lexicon building and constructicon building in the FrameNet style to the theories of Frame Semantics and Construction Grammar. Based on examination of to what extent existing FrameNet frames account for full-text annotations of Japanese texts, it will be suggested that organization of frames for Japanese may be different from that of English. This paper will furthermore discuss implications of lexicon and constructicon building to Frame Semantics and Construction Grammar.

1. FrameNet and Japanese FrameNet

Let us first examine what FrameNet and Japanese FrameNet are about, by focusing on the important features of FrameNet, which are reflected in Japanese FrameNet.

1.1. What is FrameNet?

FrameNet is an online linguistic resource based on Frame Semantics and supported with data from corpora. Put differently, "Frame Semantics has been operationalized in the FrameNet project, and its product, the FrameNet lexicon." (BAKER, 2006, p.34).

1.1.1. Lexicon

In FrameNet, a frame is "[a] script-like conceptual structure that describes a particular type of situation, object, or event along with its participants and props" (RUPPENHOFER ET AL.,

---

[*] This paper is a revised version of the plenary lecture given at the 7th International Conference on Construction Grammar (ICCG-7) on August 11th, 2012 in Seoul, Korea (OHARA, 2012b).

2010)[1]. Frames are related to one another through frame-to-frame relations. Participant (or prop) roles of the frames, called *Frame Elements* (FEs), are identified and defined. They are much more fine-grained than traditional thematic roles or semantic roles and are defined relative to the scenes[2]. Words in FrameNet are grouped based on the frame they evoke.

In FrameNet, a Lexical Unit (LU) is the pairing of a word and a frame. Sentences that represent clearly the meaning and combinatory possibilities, or *valence patterns*, of Lexical Units are collected from a corpus; and FrameNet lexicographers annotate the sentences with FEs.

In some of the early articles on Frame Semantics, Fillmore distinguished between *scenes* and *frames* (e.g. FILLMORE, 1977). Scenes were thought to be basically cognitive, while frames were linguistic. In other words, "the scenes can be more or less language-independent and hence more general, while the frames are likely to be more specific" (BAKER, 2006, p.36)[3]. FrameNet only uses the term *frame*, and NOT *scene*. While developing FrameNet, however, the researchers found that more than two levels of generality were needed. As a consequence, FrameNet developed and implemented a set of frame-to-frame relations to link frames to each other[4] and Japanese FrameNet does the same.

1.1.2. Constructicon

Originally, FrameNet only consisted of a frame-based lexicon but recently it started to incorporate a "constructicon", a registry of English constructions. When the FrameNet team went "beyond 'mere' lexicography" and started analysis and annotation of continuous text, they found that "what is missing from ordinary valence representations is the means of including the combinatorial properties of words as they function in the so-called 'peripheral' structures of the language" (FILLMORE, 2006, p.35). According to Fillmore,

> [a] **lexicon** should specify the grammatical affordances of its entries; a **grammar** should specify the kinds of lexical units capable of occurring in specifiable positions within grammatical constructions. The most consistent way to represent such mutual dependencies would be to provide both kinds of information in a single well-articulated **grammar** + **lexicon** (FILLMORE, 2006, p.35, emphasis in the original).

Noting the parallelism between constructicon building and lexicon building, the FrameNet team has been cataloguing English constructions by listing the following for each: a schematic representation, prototypical examples taken from corpus evidence, layered annotations, and descriptions in prose (Cf. Section 3.2).

---

[1] "[I]n much of NLP, the term *frame* is used to mean a syntactic frame in which a verb occurs, which defines the class of words occurring in it." (BAKER, 2006, p.33)

[2] Traditional thematic roles do not fit some frames, such as the `Similarity` frame (e.g. *The children are very similar to each other*; *John resembles his father*) and `Replacing` frame (e.g. *The coach replaced Smith with Jones*; *France replaced Brazil as world champions*) (BAKER, 2006, p.37).

[3] Furthermore, it was considered that "scenes and frames activate each other, and learning a language consists in large part of learning these associations" (BAKER, 2006, p.36). Fillmore and Baker (2009) make a distinction between *cognitive* frames, used to interpret experiences independently of whether such experiences are delivered through language and *linguistically anchored* frames, which contribute to interpreting a passage (FILLMORE & BAKER, 2009, p.313-314, 316). They claim that Frame Semantics is the study of how people associate linguistic forms with the cognitive frames, which largely determine the process and the result of interpreting the linguistic forms (FILLMORE & BAKER, 2009, p.314).

[4] For some FrameNet frames that define complex events, FrameNet uses the term *scenario* as part of their frame names.

1.2. What is Japanese FrameNet?

The Japanese FrameNet project is creating a prototype of an online Japanese lexical resource in the style of FrameNet, describing the sense of each lexical unit with respect to the frame it evokes and annotating corpus examples for each word with its FEs.

The Japanese FrameNet database structure and the software, including the online annotation tool and various web reports for viewing the results of annotation, are compatible with those of the FrameNet project. The Japanese FrameNet project imported the database and the software and modified the software accordingly for use with Japanese. The contents of the Japanese FrameNet database, including frames, FEs, and frame-to-frame relations, were imported from FrameNet as well (cf. LONNEKER-RODMAN, 2007). Our annotation methods are more or less the same as those of the FrameNet project, relying mostly on manual annotation using online annotation tools.

The Japanese FrameNet project started in 2003 and until now, we have worked on lexicon building through two annotation modes, namely, lexicographic annotation and full-text annotation. The Japanese FrameNet database currently has about 8,500 LUs (tokens; cf. 2825 types) and 60,480 annotated sentences (cf. 565 frames). Recently, the project began a pilot study about creating a constructicon of Japanese.

However, the Japanese FrameNet project also attends to contrasts between Japanese and English (cf. OHARA & SATO, 2010), thereby addressing the following research questions: 1) to what extent is the Frame-semantic approach suitable for analyzing the Japanese lexicon; and 2) to what extent the existing English-based frames are applicable to characterizing Japanese lexical units.

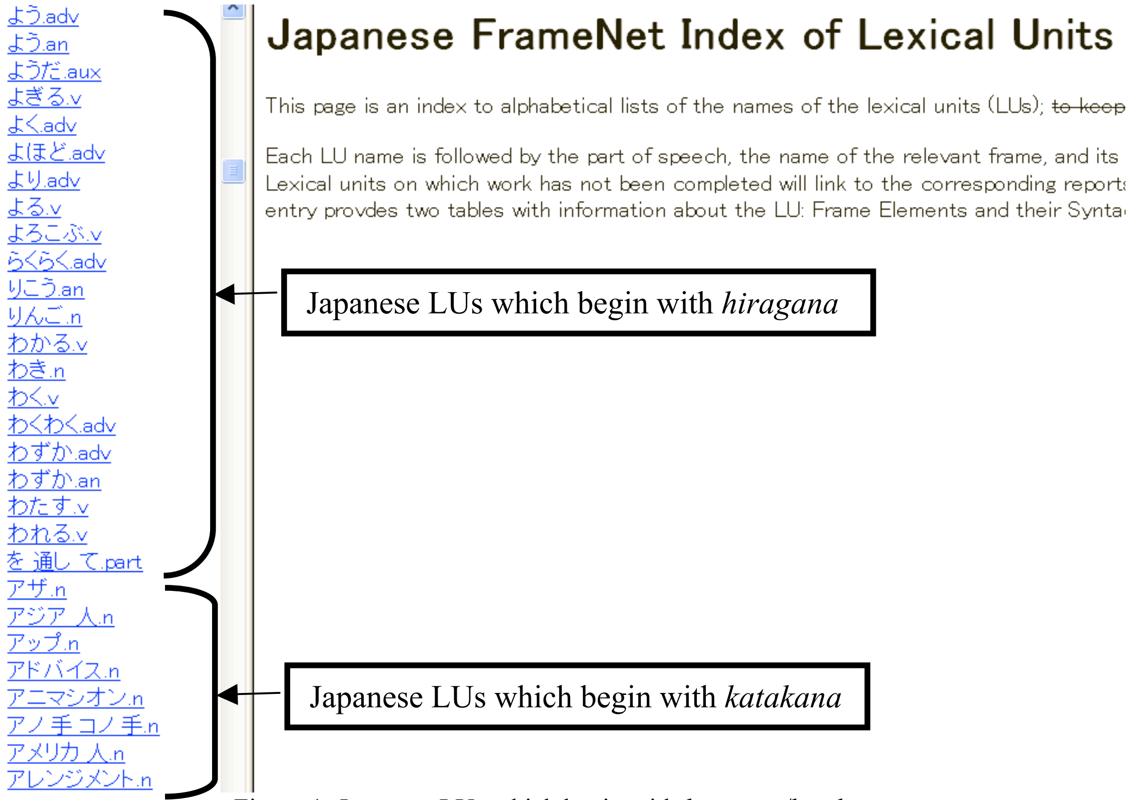
I will now discuss challenges the project faced when starting to build Japanese FrameNet.

2. Difficulties in building Japanese FrameNet

Starting to build Japanese FrameNet based on the original FrameNet for English was difficult because of the typological differences between English and Japanese. Here, I will list some of the problems that we encountered.

2.1. Writing system and character issues

FrameNet provides an alphabetical listing of all the Lexical Units that have been annotated in its Lexical Unit index, an automatically generated report. The Japanese writing system, however, does not use an alphabet, and instead employs three kinds of scripts: *kanji* (adopted Chinese characters) and two syllabaries. *Hiragana* is used, along with *kanji*, for native or naturalized Japanese words, as well as for grammatical elements; and *katakana* is used for foreign words, names, loanwords, onomatopoeia, scientific names, and sometimes to replace *kanji* or *hiragana* for emphasis. The ideal way to list Japanese LUs is to sort them according to their “readings” or pronunciations. However, the Japanese FrameNet project imported the FN database and retained as much of its original structure as possible. As a result, the project currently has no means to store information about the readings of Japanese LUs in the database. Consequently, we have listed LUs according to the order of the character code in the Japanese FrameNet Lexical Unit index report, even though this report is not user-friendly to human annotators (Figures 1, 2).



よう.adv
よう.an
ようだ.aux
よぎる.v
よく.adv
よほど.adv
より.adv
よる.v
よろこぶ.v
らくらく.adv
りこう.an
りんご.n
わかる.v
わき.n
わく.v
わくわく.adv
わずか.adv
わずか.an
わたす.v
われる.v
を 通し て.part
アザ.n
アジア 人.n
アップ.n
アドバイス.n
アニマシオン.n
アノ 手 コノ 手.n
アメリカ 人.n
アレンジメント.n

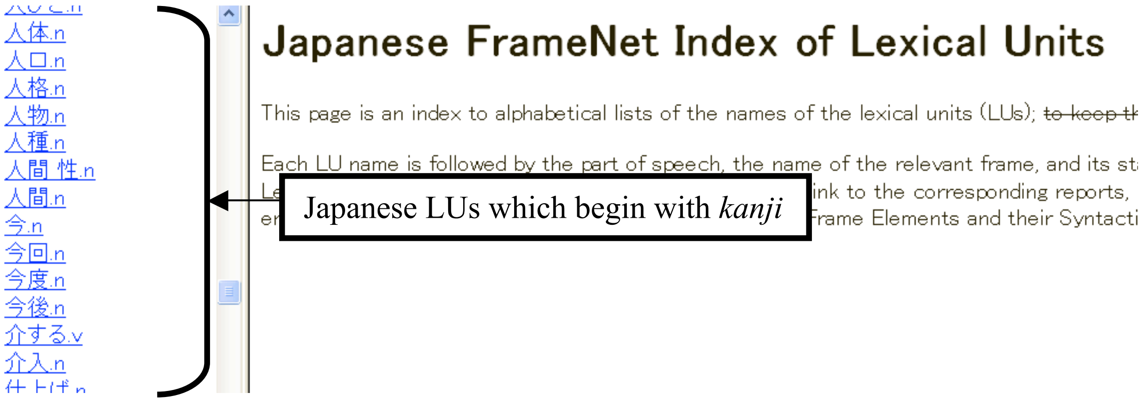
Japanese FrameNet Index of Lexical Units

This page is an index to alphabetical lists of the names of the lexical units (LUs); ~~to keep~~

Each LU name is followed by the part of speech, the name of the relevant frame, and its
Lexical units on which work has not been completed will link to the corresponding report
entry provdes two tables with information about the LU: Frame Elements and their Synta

Japanese LUs which begin with *hiragana*
Japanese LUs which begin with *katakana*

Figure 1: Japanese LUs which begin with *hiragana/katakana*



人体.n
人口.n
人格.n
人物.n
人種.n
人間 性.n
人間.n
今.n
今回.n
今度.n
今後.n
介する.v
介入.n

Japanese FrameNet Index of Lexical Units

This page is an index to alphabetical lists of the names of the lexical units (LUs); ~~to keep t~~

Each LU name is followed by the part of speech, the name of the relevant frame, and its st
Le ink to the corresponding reports,
er rame Elements and their Syntacti

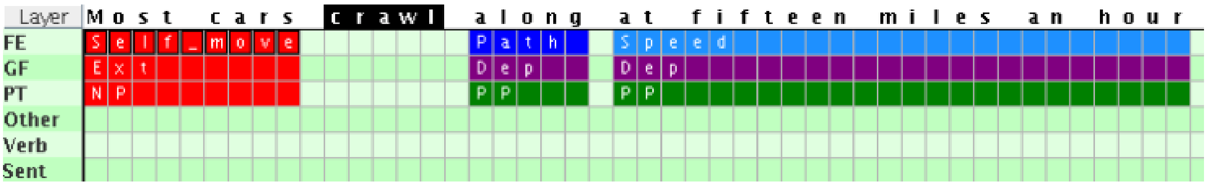
Figure 2: Japanese LUs which begin with *kanji*

2.2. Word boundary issues

Japanese is an agglutinating language, and is written with no spaces between words in sentences. Like the original FrameNet database, the Japanese FrameNet database required having spaces between words, so we developed additional programs to insert such spaces when Japanese sentences are stored in our database.

(1) Example of how Japanese sentences are stored in the JFN database
Original Japanese sentence: 彼の家へ行く
In JFN database: 彼 の 家 へ 行 く

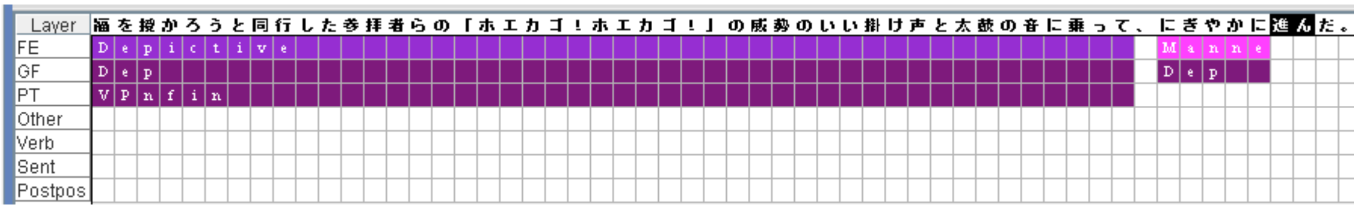
However, when Japanese sentences are displayed on the screen of the JFN annotation tool, spaces in between words are eliminated, so that it is easier for Japanese annotators to view sentences (Figures 3, 4). The cursor, however, moves word by word, not character by character, to make the user-interface more user-friendly for human annotators.



Most cars crawl along at fifteen miles an hour

Layer
FE
GF
PT
Other
Verb
Sent
Self_move
Ext
NP
Path
Dep
PP
Speed
Dep
PP

Figure 3: FN annotation tool (FNDesktop)
(LEE-GOLDMAN & RHODES, 2009)



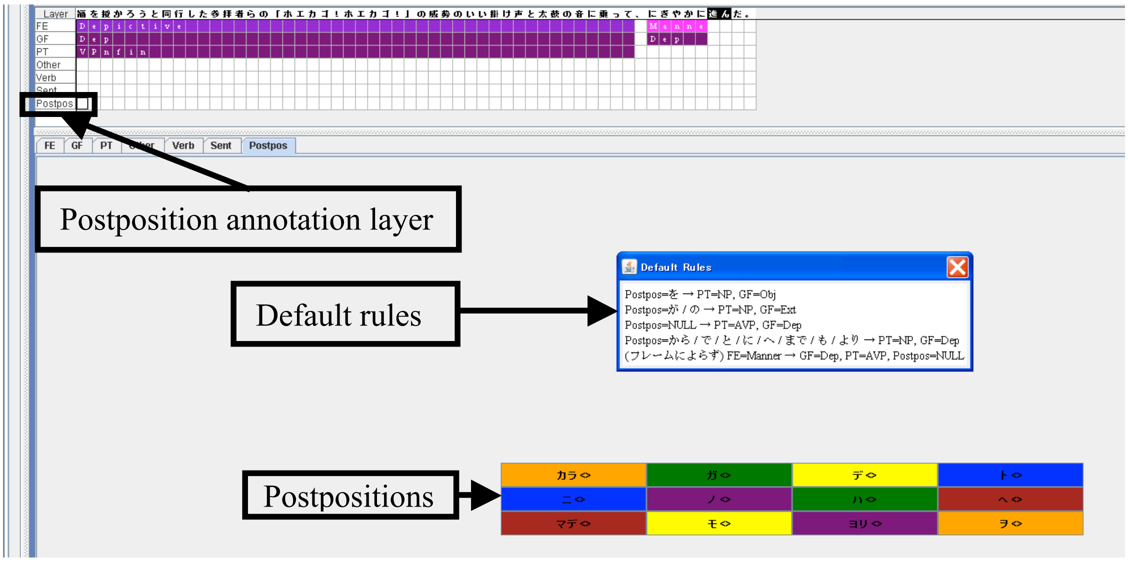
福を授かろうと同行した参拝者らの「ホエカゴ！ホエカゴ！」の威勢のいい掛け声と太鼓の音に乗って、にぎやかに進んだ。

Layer
FE
GF
PT
Other
Verb
Sent
Postpos
Depictive
Dep
VPnfin
Manne
Dep

Figure 4: JFN annotation tool (JFNDesktop)

2.3. Postpositions

Japanese has postpositions that follow a noun. They are somewhat comparable to prepositions in English in that they may provide information about the FE and/or Grammatical Function (GF) of complements of Frame-Evoking Elements (FEEs). Japanese, however, unlike English, has a relatively free word order, so sometimes postpositions are the only source of information regarding the FE and GF. JFNDesktop, the JFN annotation tool, therefore has an additional layer called "Postpos(ition)", in addition to the layers in FNDesktop, as shown in Figure 5.



Default Rules

Postpos=を → PT=NP, GF=Obj
Postpos=が / の → PT=NP, GF=Ext
Postpos=NULL → PT=AVP, GF=Dep
Postpos=から / で / と / に / へ / まで / も / より → PT=NP, GF=Dep
(フレームによらず) FE=Manner → GF=Dep, PT=AVP, Postpos=NULL

Postposition annotation layer
Default rules
Postpositions
Layer
FE
GF
PT
Other
Verb
Sent
Postpos
FE
GF
PT
Other
Verb
Sent
Postpos
カラ<>
ガ<>
デ<>
ト<>
ニ<>
ノ<>
ハ<>
ヘ<>
マデ<>
モ<>
ヨリ<>
ヲ<>

Figure 5: Postposition layer and default rules in JFNDesktop

Furthermore JFNDesktop displays suggestions based on "default rules", so that the annotator need not always annotate all the layers manually. Instead the annotator checks the system's suggestions and only when the suggested annotation is incorrect, does s/he corrects it manually, using the online annotation tool. (2) below shows the current default rules concerning the combinatorial patterns between FE GF, PT and Postpos.

(2) Default rules concerning combinatorial patterns between FE, GF, PT, and Postpos

a. Postpos=*o* → PT=NP, GF=Obj
Acc

Ex. *kono koto ga mai tosi* [EXPERIENCER.**NP.Obj.***o* *watasi* ***o***]
this thing NOM every year me ACC
kurusime-ta$_{\texttt{Experiencer\_obj}}$
torture-Perfect
"This thing has tortured me every year"

b. Postpos=*ga / no* → PT=NP, GF=Ext
Nom Gen

Ex. [STIMULUS.**NP.Ext.***ga* *kono koto* ***ga***] *mai ¥tosi watasi o*
this thing NOM every year me ACC
kurusime-ta$_{\texttt{Experiencer\_obj}}$
torture-Perfect
"This thing has tortured me every year"

c. Postpos=NULL → PT=AVP, GF=Dep

Ex. *kono koto ga* [TIME.**AVP.Dep** *mai tosi*] *watasi o*
kurusime-ta$_{\texttt{Experiencer\_obj}}$
this thing NOM every year me ACC torture-Perfect
"This thing has tortured me every year"

d. Postpos=*kara / de / to / ni / e / made / mo / yori* → PT=NP, GF=Dep
Abl Loc Com Dat Goal All also Abl

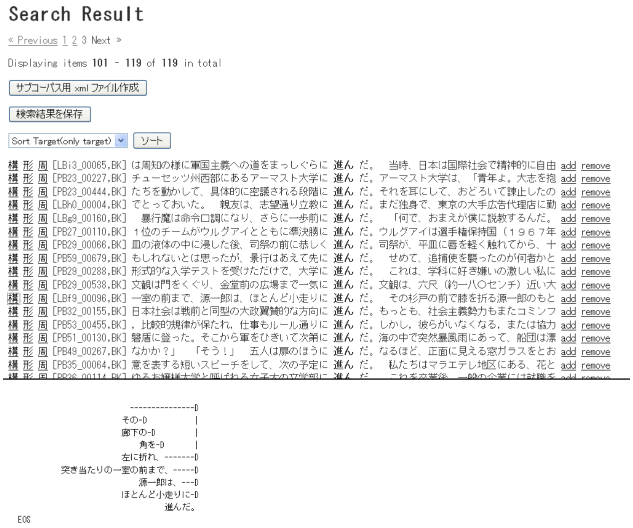
Ex. [GOAL.**NP.Dep.***ni* *yama* ***ni***] *tadori-tuku*$_{\texttt{Arriving}}$
Mountain DAT arrive
"arrive at a mountain"

e. FE=Manner → GF=Dep, PT=AVP, Postpos=NULL
Ex. *syoku-in ga* [**MANNER.AVP.Dep** *hisoka-ni*] *yama ni haitta*$_{\texttt{Arriving}}$
staff NOM secretly mountain DAT enter-Perfect
"The staff went inside the mountain secretly"

2.4. Corpus

It is ideal if the corpus used for annotation is as balanced and representative as possible. Also, having a copyright-free corpus to annotate is essential for making the annotation publicly available later. During the early phase of the project, Japanese FrameNet used copyrighted newspaper texts, because there was no copyright-free Japanese corpus, let alone a balanced corpus at that time. Since 2008, however, we have been privileged to be able to use the Balanced Corpus of Contemporary Written Japanese (BCCWJ), the first available balanced and representative corpus of Modern Written Japanese (MAEKAWA ET AL., 2010). It contains 143-milllion words of texts taken from Magazines, Newspapers, Government white papers, Books, Congress proceedings, Internet, and Textbooks.

Using a parsed corpus for annotation is helpful, because it makes it easier for annotators to find the valence of an LU. Since the annotation tool that Japanese FrameNet imported from FrameNet did not assume using parsed texts, so far we have not used parsed data for annotation. Human annotators look at the parsed data when they select sentences for annotation, i.e., before actual annotation, using our own concordance program called JFN-KWIC (SONE, OHARA, & SAITO, 2010). When the annotator wants to see the parsed structure of a sentence in the search result, s/he clicks the button to get the structure.



Displaying items 101 - 119 of 119 in total

Sort Target(only target)

```
構 形 周 [LBi3_00065,BK] は周知の様に軍国主義への道をまっしぐらに 進ん だ。　当時、日本は国際社会で精神的に自由 add remove
構 形 周 [PB23_00227,BK] チューセッツ州西部にあるアーマスト大学に 進ん だ。アーマスト大学は、「青年よ。大志を抱 add remove
構 形 周 [PB23_00444,BK] たちを動かして、具体的に密議される段階に 進ん だ。それを耳にして、おどろいて諫止したの add remove
構 形 周 [LBh0_00004,BK] でとっておいた。　親友は、志望通り立教に 進ん だ。まだ独身で、東京の大手広告代理店に勤 add remove
構 形 周 [LBg9_00160,BK] 　暴行魔は命令口調になり、さらに一歩前に 進ん だ。　「何で、おまえが僕に説教するんだ。 add remove
構 形 周 [PB27_00110,BK] １位のチームがウルグアイとともに準決勝に 進ん だ。ウルグアイは選手権保持国（１９６７年 add remove
構 形 周 [PB29_00066,BK] 皿の液体の中に浸した後、司祭の前に恭しく 進ん だ。司祭が、平皿に唇を軽く触れてから、十 add remove
構 形 周 [PB59_00679,BK] もしれないとは思ったが、景行はあえて先に 進ん だ。　せめて、追捕使を襲ったのが何者かと add remove
構 形 周 [PB29_00288,BK] 形式的な入学テストを受けただけで、大学に 進ん だ。　これは、学科に好き嫌いの激しい私に add remove
構 形 周 [PB29_00538,BK] 文観は門をくぐり、金堂前の広場まで一気に 進ん だ。文観は、六尺（約一八〇センチ）近い大 add remove
構 形 周 [LBf9_00096,BK] 一室の前まで、源一郎は、ほとんど小走りに 進ん だ。　その杉戸の前で膝を折る源一郎のもと add remove
構 形 周 [PB32_00155,BK] 日本社会は戦前と同型の大政翼賛的な方向に 進ん だ。もっとも、社会主義勢力もまたコミンフ add remove
構 形 周 [PB53_00455,BK] ，比較的規律が保たれ，仕事もルール通りに 進ん だ。しかし，彼らがいなくなる，または協力 add remove
構 形 周 [PB51_00130,BK] 磐盾に登った。そこから軍をひきいて次第に 進ん だ。海の中で突然暴風雨にあって、船団は漂 add remove
構 形 周 [PB49_00267,BK] なかか？」　「そう！」　五人は扉のほうに 進ん だ。なるほど、正面に見える窓ガラスをとお add remove
構 形 周 [PB35_00064,BK] 意を表する短いスピーチをして、次の予定に 進ん だ。　私たちはマラエテレ地区にある、花と add remove
```

```
                     ---------------D
                  その-D           |
                廊下の-D           |
                    角を-D         |
                左に折れ、-------D
突き当たりの一室の前まで、-----D
                  源一郎は、---D
          ほとんど小走りに-D
                          進んだ。
EOS
```

Figure 6: JFN-KWIC Dependency-Structure Display of a Search Result

2.5. Establishing interrelatedness between FrameNet and Japanese FrameNet

Lonneker-Rodman (2007) discusses two methods with respect to the way in which links between interrelated language resources are established (LONNEKER-RODMAN, 2007, p.3-5; cf. BURCHARDT ET AL., 2009; PETRUCK, 2009; SCHMIDT, 2009; SUBIRATS, 2009). When the *merge* approach is adopted, independent resources for different languages are first built independently from scratch and then links that relate selected types of components cross-linguistically are added (LONNEKER-RODMAN, 2007, p.4). In contrast, "[w]ith the *expand* approach, a resource for one language, which is regarded as stable at that time, is transferred to another language" (LONNEKER-RODMAN, 2007, p.4-5). Japanese FrameNet, just like the Spanish and German FrameNet projects, has taken the expand approach (cf. Section 3.1).

So far I have discussed the difficulties the Japanese FrameNet project had when starting to build Japanese FrameNet, which had to do with the Japanese writing system, word boundaries, postpositions, corpus and establishing interrelatedness with the existing FrameNet.

3. Differences in frames and constructions between FrameNet and Japanese FrameNet

Let us now turn to the discussion of some of the differences the Japanese FrameNet project has found to date between English and Japanese with respect to frames and constructions.

3.1. Differences in Frames

First, frames in Japanese and English will be discussed. The discussion is based on comparison of frames in the current versions of FrameNet and Japanese FrameNet. Since FrameNet is still under construction and so is Japanese FrameNet, the findings regarding differences in frames between them discussed below are tentative and may not be due to differences in the English and Japanese languages. The findings so far, however, seem to deserve further investigation. Section 3.1.1 discusses to what extent existing FrameNet frames are applicable to full-text annotation of Japanese in Japanese FrameNet, while Section 3.1.2 points out that judging from the actual kinds and definitions of the existing frames in FrameNet, the organizing principle of the network of frames for English and that of Japanese may be different.

3.1.1. Coverage

We examined the extent to which the Japanese FrameNet uses existing FrameNet frames originally defined for analyzing English words (cf. OHARA, 2012a). We conducted a full text annotation of a portion of the so-called core data in the book genre of the BCCWJ corpus. The task was to find a frame from the FrameNet frame repertoire for each of the Japanese nouns, verbs, adjectives, adverbs which appeared in 810 sentences in the corpus: we tried to find a frame for each of the Japanese content words, excluding named entities. Out of 4,587 tokens which appeared in the 810 sentences, we were able to assign frames to approximately 4,000 tokens. That is, of all the content words in the sentences, we were able to identify frames for 87 per cent of them. This means that the frames defined so far in FrameNet for English could be used for 87 per cent of the Japanese words[5]. The ratio seems relatively high, compared to the ratio in FrameNet. The FrameNet team conducted a similar survey, based on the annotation of one file from the ANC for nouns, verbs, and adjectives[6]. The coverage ratio was about 69 per cent for tokens.

We have yet to find out the reasons why the coverage ratio is higher in Japanese than in English. It may be due to the fact that the coverage varies depending on the text, so it may be worth further investigation with comparable texts in the two languages.

Let us now look at what kinds of Japanese words are not currently covered by existing frames in FrameNet. (3) lists some instances of Japanese words to which we could not assign an existing FN frame. The words are sorted by parts of speech in Japanese.

(3) Examples of Japanese words in the book genre of BCCWJ, to which no 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

---

[5] We recently started to investigate the coverage ratio with respect to types as well. According to our preliminary study, in a text in which the coverage ratio for tokens was 87%, the coverage ratio for types was 85%. The ratios were obtained based on annotation of the first ten sentences of PB54_00015 in the book genre of the BCCWJ corpus.

[6] The text is a life history as narrated by a woman named Sarah Reid. According to Collin Baker, the project manager of FrameNet, the vocabulary is quite ordinary (Collin Baker, personal communication).

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 the current FrameNet does not yet have appropriate frames for the conjunctions in (3b) and the nouns such as *kami* – 'god', *gangu* – 'toy', *tan'i* – 'unit', *wariai* – 'ratio', *inu* – 'dog' in (3g) is that for the most part, the FrameNet project has been annotating verbs, adjectives, and event nouns, but not conjunctions and non-event nouns (For the underlined nouns in (3g), see below). Also, since conjunctions express relations between propositions, describing their meanings with respect to various participants in situations, i.e. frames, may be difficult.

(3) also suggests that there are still many missing frames for very general words in English as well. In some of such cases, a one-to-one correspondence seems to exist between frames needed for a Japanese word and its English counterpart. Examples include frames having to do with Japanese event nouns such as *taiken*.n – 'experience', *tuukoo*.n – 'crossing', *syuppan*.n – 'publication' in (3f). At other times, the correspondences between frames needed for Japanese words and their English counterparts are more complex (cf. BOAS, 2005; FILLMORE & ATKINS 2000; OHARA, 2009). For instance, although Japanese annotators would define only one frame for *simeru*.v in (3d), it turns out that *simeru*.v corresponds not only to *make up* or *account for* as in *Students make up 5% of the population*, but also to *take up* as in *The picture takes up the entire wall* in English. That is, there may be one-to-many correspondences between the frame needed for *simeru*.v in Japanese and the frames evoked by the corresponding words in English such as *make up/account for* and *take up*. It is not the main purpose of the paper to present detailed analyses of contrastive lexical semantics, but at least it is clear that FrameNet still needs to create frames for general English words such as *make up, account for*, and *take up.*

Furthermore, the reason why we could not find an appropriate frame in the current frame repertoire in FrameNet for the underlined nouns in (3g), namely, *tatami* – 'straw mat', *syoozi* – 'sliding paper', *husuma* – 'sliding door', *kyookaku* – 'knight of the town', which refer to various elements that concern the Japanese culture, was NOT because the Japanese nouns involved Japanese culture-specific scenes but rather having to do with the fact that they are nouns. As I mentioned above, except for event nouns, FrameNet has not defined many frames pertaining to nouns so far (cf. FILLMORE, 1994). Thus, currently there are not many FrameNet frames for concepts having to do with (parts of) buildings and rooms, and with people. If there were, it would be possible to find frames for the underlined Japanese nouns in (3g). In other words, we do not have to define a culture-specific frame every time we encounter a culture-specific word. It therefore seems to confirm the validity of the current Japanese FrameNet methodology of annotating Japanese texts using the existing frames in FrameNet, while supplementing them with new frames whenever it is necessary.

3.1.2. Frame definitions

The discussion above appears to support the interlingual usefulness of the current English FrameNet frames. In what follows it is argued that there is, however, one frame-to-frame relation which seems more prominent in Japanese than in English: Intransitive_of relation. The Intransitive_of relation has not been defined in the English FrameNet database so far. In Japanese there are many intransitive verbs whose transitive counterparts are derived by adding a causative suffix. In that sense, transitive verbs are more basic in Japanese. In the current FrameNet frame database, there are many frames defined from the transitive perspective, while there are few frames defined from the intransitive perspective[7]. It may be argued that it is a case of lacking coverage in FrameNet, but the fact that there are few intransitive-transitive pairs of frames in the current FrameNet may be due to the fact that the intransitive perspective is not as prominent in English as in Japanese.

Consider the pair of sentences in (4), which pertains to a contrast between an intransitive/inchoative verb (4a) and a transitive verb (4b). Example (4a) depicts a scene in which petals of cherry blossoms get scattered. The intransitive verb *tiru* – 'get scattered' in (4a) is an inchoative verb used to describe particles or small objects falling.

(4) a. *sakura* *no* *hanabira ga* *tiru* `Motion`
cherry.blossom GEN petals NOM be.scattered
'Petals of cherry blossoms get scattered.'
b. *sakura* *no* *hanabira o* *tirasu* `Dispersal`
cherry.blossom GEN petals ACC scatter
'(Somebody) scatters petals of cherry blossoms.'

It was impossible to find a frame in the current FrameNet database relevant to the meaning of *tiru* in (4a). In contrast, for the morphologically related transitive counterpart *tirasu* – 'scatter' in (4b), we assume that the `Dispersal` frame is involved. The `Dispersal` frame is defined in FrameNet as "an AGENT or a CAUSE disperses or scatters INDIVIDUALS from the SOURCE, a relatively confined space, to the GOAL_AREA, a broader space". The only existing frame that seems relevant to the intransitive verb *tiru* – 'be scattered' in (4a) is the `Motion` frame. The `Motion` frame is 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 a very general situation involving motion. The frame-to-frame relation between the `Dispersal` frame and the `Motion` frame is represented in Figure 7.

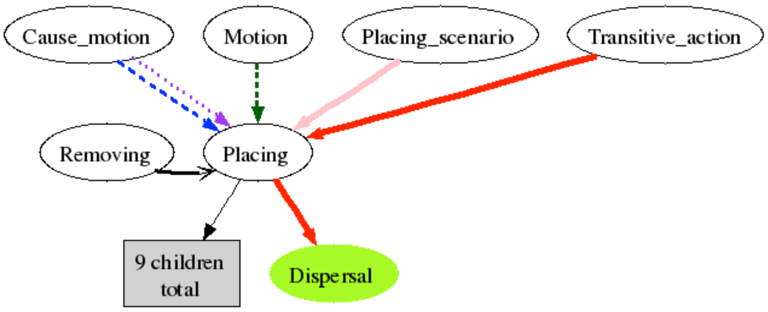


Figure 7. Frame-to-frame relations pertaining to `Dispersal` and `Motion` frames

---

[7] As of January 31, 2013, FrameNet frames defined from the intransitive perspective include the following: `Become_silent`, `Become_triggered`, `Becoming`, `Becoming_a_member`, `Becoming_aware`, `Becoming_detached`, `Becoming_dry`, `Becoming_separated`, `Becoming_visible`, and `Expansion`.

To accurately describe the contrast between the intransitive *tiru* – 'get scattered' and the transitive *tirasu* – 'scatter', we need to define the `Becoming_Dispersed` frame, which characterizes a situation in which INDIVIDUALS get scattered from the SOURCE to the GOAL_AREA, in a downward movement and the Intransitive_of relation (cf. Figure 8).

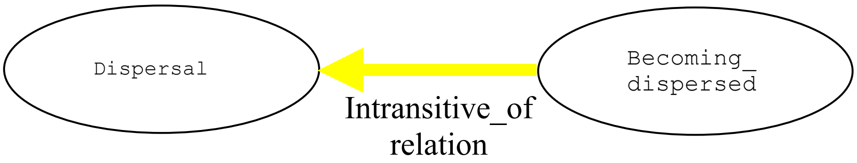


Figure 8: Suggested New Frame and Frame-to-Frame Relation for Japanese

Many other pairs of intransitive/inchoative and transitive verbs in Japanese exist that are often morphologically related. We have determined, however, that many of the existing frames originally defined for analyzing the semantics of English words, involve the transitive perspective rather than the intransitive perspective. Few cases exist in which pairs of FrameNet 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 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)[8].

As noted above, many existing frames capture the transitive perspective, rather than the intransitive. In other words, existing frames seem to assume perspectives and lexical aspects (*aktionsart*) of English words, which are not necessarily the same as those of Japanese words[9]. To some extent, however, this situation seems to be unavoidable. Whereas scenes are thought to be basically cognitive and language-independent, frames are linguistic and language-dependent. It is beyond the scope of the paper but the hypothesis that the intransitive perspective is more prominent in Japanese than in English seems worth further research.

3.2. Differences in constructions?

The Japanese FrameNet project began a pilot study about creating a constructicon of Japanese. The project is in the process of modifying the FrameNet constructicon annotation tool, for use in annotating Japanese texts. I will thus describe what Japanese FrameNet has done in the pilot study and suggest a possible research question for cross-linguistic comparison of constructions and, hence also for constructicon building.

When Japanese FrameNet project started full-text annotation work, we discovered our existing annotation practices do not suffice for at least the following types of multiword expressions in Japanese: complex particles; complex auxiliaries; and clause-level constructions.

---

[8] Note that for situations characterizing detachment, a three-way distinction in semantic frames exists. 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, FrameNet defined 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.

[9] There seem to be cross-linguistic variations with respect to kinds of FEs, coreness statuses of FEs, and Frame Element relations (cf. OHARA ET AL., 2004; LONNEKER-RODMAN, 2007, p.19-21).

For example, (5a) describes hearsay, but it is not possible to compositionally derive the meaning. The underlined segment literally means ‘depending on (the) place (where) V’ or ‘based on (the) place (where) V’. Hearsay is a complex notion, having to do with a kind of proposition attributed to a medium and an attitude about the proposition. Furthermore, statements of hearsay have a special syntax: a verb describing the medium precedes the whole segment; and, an auxiliary expressing the speaker’s attitude follows the clause expressing the proposition.

(5) a. Complex particles
kiku tokoro ni yoru to Toyota 86 wa subarasii rasii
hear place LOC depend QUOTE TOP superb seem
literal. ‘Depending/Based on (the) place (where I) hear, Toyota 86 seems superb.’
= ‘Judging from (what I) hear, Toyota 86 seems superb.’

b. Complex auxiliaries
koi kiri ga numa no ue ni
thick fog NOM mire GEN top LOC
ori te iru
fall
‘A thick fog has fallen over the mire.’

c. Clause-level constructions
kore (*no hoo*) *ga are yori nagai*
This GEN side NOM that than long
‘This is longer than that.’

The framework that the FrameNet constructicon-building project laid out seems suitable for handling the three types of multiword expressions (FILLMORE ET AL., 2012; LEE-GOLDMAN & RHODES, 2009). The goal of the constructicon project is developing a repository of constructions, not writing a full-fledged construction grammar (LEE-GOLDMAN & RHODES, 2009). The constructicon includes the following parts: a list of constructions; construction definitions; and annotation of sentences illustrating the constructions. An annotation of each construction identifies and annotates the following: the CONSTRUCT, i.e., a phrase licensed by the rules of a construction; CONSTRUCT ELEMENTS (CEs, or components of the construct); a special CONSTRUCTION-EVOKING ELEMENT (CEE); and relevant features of the context (FILLMORE ET AL., 2012, p.321).

(5’a) is a constructional annotation of (5a). It lists: a mnemonic name for the construction; informal descriptions of an interpretation of the manner in which properties of the CEs participate in the resulting sign; mnemonic names for the CEs; and the bracketing formula with the CEE underlined and with the mnemonic names for the CEs. That is, (5a) can be analyzed as a construct of the `Hearsay` construction, whose interpretation is that the speaker has an attitude about a proposition which is attributed to a medium. The CEE is the underlined phrase *tokoro ni yoru to*. The CEs include: *Medium, Proposition*, and *Attitude*.

(5’a) `Hearsay` construction

- The `Hearsay` construction indicates that the speaker has an *Attitude* about the *Proposition* attributed to a *Medium*.
- Construct Elements: *Medium, Proposition, Attitude*
- [$_{\textit{Medium}}$ *kiku*] [$_{\textit{CEE}}$ *tokoro ni yoru to*] [$_{\textit{Proposition}}$ *Toyota 86 wa subarasii*]

[*Attitude* *rasii*]
‘Judging from (what I) hear, Toyota 86 seems superb.’

It turns out that the *`Hearsay`* construction evokes the `Attribute_information` frame as shown in (5’’a). (5’’a) lists: the name of the frame; informal descriptions of the frame; mnemonic names for the Frame Elements (FEs); and the bracketing formula with the underlined Frame-Evoking Element (FEE) and with the mnemonic names for the FEs. Note that the FE SPEAKER, one of the core FEs of the `Attribute_information` frame, does not appear in the sentence and is annotated as an instance of the definite null instantiation (DNI):

(5’’a) *`Hearsay`* construction evokes the `Attribute_information` frame:

- A PROPOSITION is attributed to a SPEAKER or a TEXT.
- Frame Elements: PROPOSITION, SPEAKER, TEXT
- *kiku* [FEE *tokoro ni yoru to*] [PROPOSITION *Toyota 86 wa subarasii*] *rasii*
  SPEAKER:DNI
  ‘Judging from (what I) hear, Toyota 86 seems superb.’

Similarly, the constructional annotation of (5b) would look like (5’b):

(5’b) *`Resultant_state`* construction

- The *`Resultant_state`* construction describes a *State* after an *Event* pertaining to an *Entity* has occurred.
- Construct Elements: *Entity, Event, State*
- [*Entity* *koi kiri ga*] *numa no ue ni* [*State* [*Event* *orite*] [*CEE* *iru*]]
  ‘A thick fog has fallen over the mire.’

The *`Resultant_state`* construction is another frame-bearing construction. That is, the expressions which correspond to the CE *Event* and the CEE *teiru* together evoke the `State_continue` frame.

(5’’b) *`Resultant_state`* construction evokes the `State_continue` frame:

- The ENTITY remains in the specified STATE.
- Frame Elements: ENTITY, STATE
- [ENTITY *koi kiri ga*] *numa no ue ni* [STATE= FEE [*orite*] [*iru*]]
  ‘A thick fog has fallen over the mire.’

Finally, (5c) can be annotated for constructions as shown in (5’c):

(5’c) *`Comparative_inequality`* construction (cf. HASEGAWA ET AL., 2010, 2012)

- The *`Comparative_inequality`* construction reports inequalities between two *Entities* as arguments of a plain adjective.
- Construct Elements: *Entity1, Entity2, Feature*
- [*Entity1* *kore* (*no hoo*) *ga*] [*Entity2* *are*] [*CEE* *yori*] [*Feature* *nagai*]
  ‘This is longer than that.’

The `Comparative_inequality` construction evokes the `Comparison_inequality` frame as shown in (5''c). Here, the expressions corresponding to the CE *Entity2* and the CEE *yori* as a whole evoke the `Comparison_inequality` frame.

(5''c) `Comparative_inequality` construction evokes the `Comparison_inequality` frame:

- The ENTITY is compared against some STANDARD with respect to their values for some FEATURE.
- Frame Elements: ENTITY, STANDARD, FEATURE
- [ENTITY *kore* (*no hoo*) *ga*] [STANDARD=FEE [*are*] [*yori*]] [FEATURE *nagai*]
  'This is longer than that.'

I have briefly described how each of the three types of Japanese multiword expressions that cannot be handled in the Japanese FrameNet lexicon can be annotated using the construction-annotation procedures that FrameNet proposed for building its constructicon. As shown above, each of the three constructions seems to evoke a frame. However, Fillmore et al. list "constructions without meanings" as one type of construction (FILLMORE ET AL., 2012, p.325). Citing Subject Auxiliary Inversion (or in the FrameNet Constructicon terminology `Aux-initial`) as an example, Fillmore et al. (2012) point out:

> [w]hile construction grammars originated in the recognition of conventional pairings between specific formal patterns and the meanings they contribute to the expressions that contain them – against a contrasting view that syntactic principles should be stated independently of questions of meaning and use – there remains the question of whether all constructions should be seen as meaning-bearing. (FILLMORE ET AL., 2012, p.325)

They go on to say, "the actual work of building the FrameNet Constructicon is proceeding under an assumption of the legitimacy of semantically null constructions" (FILLMORE ET AL., 2012, p.328)[10]. It remains to be seen whether or not Japanese does indeed have semantically null constructions.

4. Relating Lexicon and Constructicon building to Frame Semantics and Construction Grammar

Let us now turn to the topic of implications of lexicon building and constructicon building for the theories of Frame Semantics and Construction Grammar.

4.1. Testing ground for Frame Semantics and Construction Grammar

How do lexicon building and constructicon building relate to Frame Semantics and Construction Grammar? One possible answer would be to say that we engage in lexicon building and constructicon building of the sort I have described here to test the theories of Frame Semantics and Construction Grammar. Frame Semantics is "the study of how linguistic forms *evoke* or activate frame knowledge, and how the frames thus activated can be integrated into an understanding of the passages that contain these forms" (FILLMORE & BAKER 2009, p.317). Thus, moving into the semantics of grammar by building a constructicon in addition to a lexicon would mean testing the methods of the two theories to integrate lexical meanings and grammatical meanings into a complete account of the language-based interpretations of texts (FILLMORE & BAKER, 2009, p.339).

---

[10] It may have to do with the following remark: "[h]ow do we convince ourselves that we've come upon a set of phenomena that require separate treatment in terms of constructions that go beyond what our grammar already has; and how do we talk about the difference between constructions that have no particular lexical requirements – constructions that are wholly syntactic – from those that require descriptions in terms of specific words, or specific word classes?" (FILLMORE, 2006, p.57)

4.2. Refining frameworks and concepts in Frame Semantics and Construction Grammar

Second, by building a FrameNet-style lexicon and constructicon, we refine both the frameworks and concepts of Frame Semantics and Construction Grammar. Baker (2006) points out "[a]n important part of the development of FrameNet that was foreseen only in the most general terms in Fillmore's writing on Frame Semantics was the crucial role of frame-to-frame relations" (BAKER, 2006, p.37). One of the roles of the JFN project would be to see how frame-to-frame relations would differ from those of English.

Also, the current debate on whether or not to allow purely syntactic constructions, which I mentioned in Section 3.2, may lead to refining the concept of "construction" in Construction Grammar. The JFN project can contribute to the discussion by giving evidence for or against the existence of purely syntactic constructions, as we go.

Conclusions

This paper discussed the findings of the ongoing language resource building project of Japanese FrameNet, focusing on two research questions: to what extent the existing English-based frames are applicable to characterizing Japanese words; and to what extent the Frame Semantic and Construction Grammar approaches are suitable for analyzing the Japanese lexicon and grammar. So far the Frame Semantic and Construction Grammar approaches seem suitable for analyzing Japanese. We have yet to find out whether the organization of frames for Japanese is different from that of English.

References

BAKER, C. W. Frame Semantics in Operation: The FrameNet Lexicon as an Implementation of Frame Semantics. In: *ICCG4 The Fourth International Conference on Construction Grammar Plenary Lectures*. Tokyo, Japan, 2006. p. 34-43.

BOAS, H. C. Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography,* vol. 18, n.4, 2005. p. 445-478.

BOAS, H. C. (Ed.). *Multilingual FrameNets in Computational Lexicography: Methods and Applications.* Berlin/New York: Mouton de Gruyter, 2009.

BURCHARDT, A.; ERK, K.; FRANK, A.; KOWALSKI, A.; PADO, S.; & PINKAL, M. Using FrameNet for the semantic analysis of German: annotation, representation, and automation. In: BOAS, H. C. (Ed.), *Mulitilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter, 2009. p. 209-244.

FILLMORE, C. J. Scenes-and-frames semantics, Linguistic Structures Processing. In: ZAMPOLLI, A. (Ed.). *Fundamental Studies in Computer Science*, n. 59, North Holland Publishing, 1977. p. 55-88.

FILLMORE, C. J. The Hard Road from Verbs to Nouns. In: CHEN, M. Y. & TZENG, O. J. L. (Ed.) *In Honor of Wong: Interdisciplinary Studies on Language and Language Change.* Pyramid Press, 1994. p. 105-129.

FILLMORE, C. J. The Articulation of Lexicon and Constructicon. In: *ICCG4 The Fourth International Conference on Construction Grammar Plenary Lectures*. Tokyo, Japan, 2006. p. 53-63.

FILLMORE, C. J. Border Conflicts: FrameNet meets construction grammar. In: BERNAL, E. & DECESARIS, J. (Eds.). *Proceedings of the XIII Euralex International Congress*. Barcelona: Institut Universitari de Lingüistica Aplicada, Spain, 2008. p. 49-68.

FILLMORE, C. J. and ATKINS, B.T.S. Describing polysemy: The case of 'Crawl'. In.: RAVIN, Y. And LEACOCK, C. (Eds.) *Polseymy: Theoretical and Computational Approaches.* Oxford: Oxford University Press, 2000. p.91-110.

FILLMORE, C. J. & BAKER, C. F. A Frames Approach to Semantic Analysis. In.: HEINE, B. & NARROG, H. (Eds.). *The Oxford Handbook of Linguistic Analysis*. Oxford: OUP, 2009. p. 313-339.

FILLMORE, C. J., LEE-GOLDMAN, R. & ROMIEUX, R. The FrameNet Constructicon. In: BOAS, H. C. and SAG, I. A. (Eds.). *Sign-based Construction Grammar*. Stanford: Center for the Study of Language and Information, 2012. p.309-372.

GOLDBERG, A. *Constructions at Work: the nature of generalization in language.* Oxford: OUP, 2006.

HASEGAWA, Y., LEE-GOLDMAN, R., OHARA, K., FUJII, S. & FILLMORE, C. J. On expressing measurement and comparison in Japanese and English. In: BOAS, H. C. (Ed.). *Contrastive Construction Grammars*. Amsterdam: John Benjamins, 2010. p. 169-200.

HASEGAWA, Y., LEE-GOLDMAN, R., OHARA, K., ELLSWORTH, M. & FILLMORE, C. J. The Frames-and-Constructions Approach to Paraphrase. Paper presented at *ICCG7*. Seoul, Korea, 2012.

LEE-GOLDMAN, R. & RHODES, R. Corpus-based analysis and annotation of constructions. *Frames and Constructions Conference*. Berkeley, California, USA, 31 July-2 August 2009.

LONNEKER-RODMAN, B. Multilinguality and FrameNet. *International Computer Science Institute Technical Report*, 2007.

MAEKAWA, K., YAMAZAKI, M., MARUYAMA, T., YAMAGUCHI, M., OGURA, H., KASHINO, W., OGISO, T., KOISO, H. & DEN, Y. Design, Compilation, and Preliminary Analyses of Balanced Corpus of Contemporary Written Japanese. In: *Proceedings of the 7th*

International Conference on Language Resources and Evaluation (LREC2010). Malta, 2010. p.1483-1486.

OHARA, K. H. Frame-based contrastive lexical semantics in Japanese FrameNet: The case of *risk* and *kakeru*. In.: BOAS, H. C. (Ed.), *Mulitilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter, 2009. p. 163-182.

OHARA, K. H. Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English. In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12).* Istanbul, Turkey, 2012a.

OHARA, K. H. Japanese FrameNet: Toward Constructicon Building for Japanese. Plenary lecture given at *ICCG7*. Seoul, Korea, 2012b.

OHARA, K. H., FUJII, S., OHORI, T., SUZUKI, R., SAITO, H. & ISHIZAKI, S. The Japanese FrameNet project: An introduction. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC2004).* Lisbon, Portugal, 2004.

OHARA, K. H. & SATO, H. Investigating Japanese FrameNet Data with FrameSQL. *Sixth International Conference on Construction Grammar (ICCG6).* Charles University, Prague, Czech Replublic. September 5th, 2010.

PETRUCK, M. R. L. 2009. Typological considerations in constructing a Hebrew FrameNet. In: BOAS, H. C. (Ed.), *Mulitilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter, 2009. p. 183-205.

RUPPENHOFER, J., ELLSWORTH, M., PETRUCK, M. R. L., JOHNSON, C. R. & SCHEFFICZYK, J. *FrameNet II: Extended Theory and Practice*. 2010.

SCHMIDT, T. 2009. The Kicktionary A multilingual lexical resource of football language. In: BOAS, H. C. (Ed.), *Mulitilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter, 2009. p. 101-132.

SONE, T., OHARA, K. H. & SAITO, H. JFN-KWIC: A Concordancer for BCCWJ and Japanese FrameNet. (In Japanese). In: *Proceedings of the 16th Annual Meeting of the Association for Natural Language Processing*, 2010. p.506-509.

SUBIRATS, C. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In: BOAS, H. C. (Ed.), *Mulitilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter, 2009. p. 135-162.

Received: 31/01/2013
Revised: 23/06/2013
Accepted: 10/07/2013
Published: 15/10/2013