Japanese FrameNet: Lexicon and Constructicon Building for Japanese

Kyoko Hirose Ohara
Keio University
ohara@hc.st.keio.ac.jp

1st Dec, 2012
The 13th Korea-Japan Workshop
on Linguistics and Language Processing:
Corpora, Annotation and Human Language Processing
Waseda University
Outline

1. Introduction
2. What is Japanese FrameNet
3. What are the theories behind Japanese FrameNet
4. How is Japanese FrameNet different from other linguistic resources
5. Summary
6. Conclusions
Introduction
Purpose of the talk

- Corpus
  - Analyzing and annotating corpus data
- Annotation
  - Word meaning, sentence meaning
- “Human” Language Processing
  - Speaker’s understanding
<table>
<thead>
<tr>
<th>Name</th>
<th>University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toshio Ohori</td>
<td>(The University of Tokyo)</td>
</tr>
<tr>
<td>Seiko Fujii</td>
<td>(The University of Tokyo)</td>
</tr>
<tr>
<td>Ryoko Suzuki</td>
<td>(Keio University)</td>
</tr>
<tr>
<td>Hiroaki Saito</td>
<td>(Keio University)</td>
</tr>
<tr>
<td>Hiroaki Sato</td>
<td>(Senshu University)</td>
</tr>
<tr>
<td>Shun Ishizaki</td>
<td>(Keio University)</td>
</tr>
</tbody>
</table>
Current annotators & programmers

Toshiko Kigoshi
Anna Gladkova
Hiroya Hagino
Naoko Kurokawa
Hidetoshi Kobori
Antoine Mousnier
Benoit Eudier
What is Japanese FrameNet
Japanese FrameNet (JFN): Input

• Balanced Corpus of Contemporary Written Japanese (BCCWJ) by National Institute for Japanese Language and Linguistics (NINJAL)
  – the first available balanced and representative corpus of Modern Written Japanese (2011-)
  – Copyright-free
  – Contains 143-million words of texts taken from:
    • Magazines, Newspapers, Government white papers, Books, Congress proceedings, Internet, and Textbooks

<table>
<thead>
<tr>
<th>Publication Subcorpus</th>
<th>Library Subcorpus</th>
<th>Special-Purpose Subcorpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books, magazines, and newspapers published during 2001-2005</td>
<td>Books catalogued at more than 13 public libraries in Tokyo area, and published after 1985</td>
<td>Whitepaper text, Internet text, Diet minutes, Best selling books, etc.</td>
</tr>
<tr>
<td>35 million words</td>
<td>30 million words</td>
<td>35 million words</td>
</tr>
</tbody>
</table>

Structure of the BCCWJ
JFN-KWIC Concordance Program

Display of parsed sentence
Japanese FrameNet (JFN): Method

• Analyze and annotate a word meaning in a sentence,
• based on **Frame Semantics** (Fillmore 1985, Fillmore & Baker 2010, etc.)

  – Frames
    • “[A] script-like conceptual structure that describes a particular type of situation, object, or event along with its participants and props” (Ruppenhofer et al. 2010)
    • Related through **frame-to-frame relations**
      • **Frame Elements (FE)**
        – Participant (or prop) roles of the frames are identified and defined
    • Words are grouped based on the frame they *evoke*
    • A **Lexical Unit (LU)** is the pairing of a word and frame
Japanese FrameNet (JFN): Method

• Creating a **prototype of an on-line Japanese lexical resource** following FrameNet methodology and practice
  – Describes the **sense of each lexical unit** with respect to the **semantic frame** it evokes
  – **Annotates corpus examples** of each word analyzed with frame elements

• Compatibility with FrameNet
  – JFN databases and annotation tool
  – JFN frames: imported from FN (the Expand approach)
  – Annotation methods

• Lexicon building > Constructicon Building
Japanese FrameNet (JFN): Output Lexical Entry Report

Valence Patterns:

These frame elements occur in the following syntactic patterns:

<table>
<thead>
<tr>
<th>Number Annotated</th>
<th>Number Annotated</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) TOTAL</td>
<td>(1) TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TOTAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TOTAL</td>
<td></td>
</tr>
</tbody>
</table>

- そして 彼はあごひげなどで変装して、モーティマー医師をここまで尾行してみたり、ヘンリー親の着
- 目だったときは駅へも行ったし、ノーサンブランド・ホテルへ尾行したりして出歩いている間、細君は
- ホテルの部屋に監禁しておいたのだ。DN
- やっと眉間かな、あ、かわ
- 相互タンケー（大阪市東区）でも、依頼事務に到着した際、「車が着きました」という連絡が
- 省けず、使者「無事を願うのか」といった言葉が数件寄せられたという。DN
- 花子がタケエに飛び乗られてくれたんですか、なかなか 防音 ないから 彼は覚悟を決めた。DN
- 着信のメールの言うことに、相手に自動連絡するよう 新手のシリーズだろう。DN
- いつもの 間にから、骨董（つぼ）を持って道楽が過場するまで通常約1時間半と見込まれて
Japanese FrameNet (JFN): Output

• A semantically annotated corpus
Experiencer_focus

Definition:

The words in this frame describe an Experiencer’s emotions with respect to some Content. A Explanation for the emotion in current state of affairs, quite often it refers to a general situation which causes the emotion.

• My ENJOYMENT of the movie was considerably impaired by the seven-foot guy sitting in front of me. [Yahoo!Japan]

• Smithers takes great PLEASURE in collecting matchboxes. [Yahoo!Japan]
Japanese FrameNet (JFN): Output Frame-to-Frame relations

- Ontology
What are the theories behind Japanese FrameNet
Frame Elements vs. Semantic Roles

Frame Elements (FEs) are relativized to frames and much more fine-grained than Semantic Roles

- Replacing frame
  
  • An Agent changes the filler of a Role by placing a New filler in the position after the Old filler ceases to occupy the position. In most cases the Role is implicit.

- If you REPLACE me with a robot, who's gonna make excuses to your wife for you?

- If you SUBSTITUTE a 15" arm for the 50cm one, it works pretty well.
Valence Patterns vs. Case Frames

Valence Patterns are multi-layers consisting of:

- Frame Elements (FEs)
- Grammatical Functions (GFs: Subj, Obj, ...)
- Phrase Types (PTs)
- Case markers
Valence Patterns vs. Case Frames

– hanako wa **taroo ni** tyokoreeto o AGETA

  TOP  DAT chocolate  ACC gave

  “Hanako gave chocolate **to** Taro.”

– hanako wa **taroo ni** tyokoreeto o MORATTA

  TOP  DAT  ACC received

  “Hanako received chocolate **from** Taro.”
Valence Patterns vs. Case Frames

– Giving frame

• A Donor transfers a Theme from a Donor to a Recipient.

• hanako wa taroo ni tyokoreeto o AGETA
  TOP DAT chocolate ACC gave

  “Hanako gave chocolate to Taro.”

– Receiving frame

• A Recipient comes into possession of the Theme as a result of the joint action of the Donor and the Recipient.

  hanako wa taroo ni tyokoreeto o MORATTA
  TOP DAT ACC received

  “Hanako received chocolate from Taro.”
How is Japanese FrameNet different from other linguistic resources
Annotating
dictionary example phrases


a. *seken o odorokasetziken*
   public ACC surprised incident
   ‘the incident which surprised the public’

b. *zimoku o odorokasu*
   many_people’s_attention ACC surprise
   ‘to surprise people’
Relevant entries in VAST and JFN

(2) VAST entry for *odorokasu* (Takeuchi et al. 2008)

a.  
\[
\text{<Agent>} \ ga \ \text{<Person>} o \ odorokasu
\]
    
    NOM     ACC  surprise

b.  
\[
\text{<Causer>} \ ga \ \text{<Person>} o \ odorokasu
\]

(3) The *Experiencer_obj* frame in JFN
Some phenomenon (the *STIMULUS*) provokes a particular emotion in an *EXPERIENCER.*
Annotations of (1) in VAST and JFN

(2’) VAST annotations

a. \(<\textbf{Person}> seken o] \ odorokaseta [<\textbf{Causer}> ziken]\)
   public ACC surprised incident
   ‘the incident which surprised the public’

b. \(<\textbf{Person}> zimoku o] \ odorokasu\)
   many_people’s_attention ACC surprise
   ‘to surprise people’

(3’) JFN annotations

a. \(<\textbf{EXPERIENCER seken o]} \ odorokaseta [\textbf{STIMULUS ziken}]\)

b. \(<\textbf{EXPERIENCER zimoku o]} \ odorokasu\)
Annotating Corpus Sentences

(4) Sentence from the BCCWJ corpus

Sadako ga dansu o suru siin o soozosita koto mo
   NOM dance ACC do scene ACC imagined thing PART
nakatta tame, Sadako no odori o mite,
did.not.exist SUB GEN dance ACC see-TE
Tooyama wa kanari odorokasareta.
   TOP much be.surprised

‘Since (he) had not imagined a scene in which Sadako performs a dance, seeing her dance, Toyama was much surprised.’
Treatment of ‘peripheral’ phrases

• JFN assigns FEs to adjunct phrases, which are often disregarded as ‘peripheral’ in VAST.

• Many sentences in the corpus contain adjunct phrases, and JFN uses the framework of Frame Semantics to annotate them properly, just as English FN does.
  – JFN: [STIMULUS Seeing Sadako’s dance] (Sadako no odori o mi-Te), Tooyama was much surprized.
  – BFN: ... it always surprises me [STIMULUS when people turn out to be such bad listeners]
Annotations of (4) in VAST and JFN

(5) VAST annotation

Sadako ga dansu o suru siin o soozosita koto mo nakatta tame,
Sadako no odori o mite,
[< Person > Tooyama wa]
kanari odorokasareta.

(6) JFN annotation

[EXPLANATION Sadako ga dansu o suru siin o soozosita koto mo nakatta tame]
[STIMULUS Sadako no odori o mite],
[EXPERIENCER Tooyama wa]
[DEGREE kanari] odorokasareta.
Entries in VAST and JFN

(7) Entry for kangaeru.v ‘think’ in VAST (Takeuchi et al. 2008)
   a.  <Experiencer> ga  <Content> o  kangaeru
       NOM         ACC think
   b.  [<Content> kono syoosetu no teema    o] kangaeru
       this novel  GEN theme ACC think
       ‘(I) think about the theme of this novel.’

(8) Cogitation frame in JFN
   A person, the COGNIZER, thinks about a TOPIC over a period of time.

(8’) Cogitation.kangaeru.v in JFN
     [COGNIZER External NP ga] [TOPIC Dependent NP/S o/ni.tuite] kangaeru
     NOM         ACC/about think
JFN entries for *kangaeru.v*

(8) Cogitation frame in JFN
A person, the **Cognizer**, thinks about a **Topic** over a period of time.

(8') Cogitation. *kangaeru.v* in JFN

[**Cognizer** External NP *ga*] [**Topic** Dependent NP/S *o/ni.tuite*] *kangaeru*

NOM ACC/about think

(9) Coming_up_with frame
Words in this frame have to do with a **Cognizer** creating a new intellectual entity, the **idea**.

(9') Coming_up_with. *kangaeru.v* in JFN

[**Cognizer** External NP *ga*] [**Idea** Dependent NP *o*] *kangaeru*

NOM ACC think

(10) Opinion frame
A **Cognizer** holds a particular **Opinion**, which may be portrayed as being about a particular **Topic**.

(10') Opinion. *kangaeru.v* in JFN

[**Cognizer** External NP *ga*] [**Topic** Dependent NP *o*] [**Opinion** Dependent VP *to*] *kangaeru*

NOM ACC QUOTE think
JFN entries for *kangaeru.v*

(8) Cogitation frame in JFN
A person, the **Cognizer**, thinks about a **topic** over a period of time.

(8’)
Cogitation.*kangaeru.v* in JFN

[**Cognizer** External NP *ga*] [**Topic** Dependent NP/S *o/ni.tuite*] *kangaeru*

NOM ACC/about think

(8’’)
*Zen.sekai ni.watatte* [**Topic** *kono mondai o*] *kangaeru beki da*

all.world throughout this problem ACC think should COP

‘Throughout the whole world, (we) should consider this problem.’

(8’’’)
[**Topic** *dare ga daihyoo to.site husawasii ka*] *kangaeru*

who NOM representative as appropriate Q think

‘think who would be appropriate as the representative’
Summary

• JFN is a linguistic resource containing semantic annotation of corpus data, based on frame semantics and construction grammar.

• Its theoretical foundation is solid and can describe word meanings and sentence meanings, at least in some respects, better than others.

• It is possible to use JFN together with other lexical resources for Japanese depending on applications
  – cf. Matsubayashi et al 2010
Conclusions

- JFN is about ...
  - corpus
    - => BCCWJ
  - annotation
    - => Frame-Semantic annotation
  - "human" language processing
    - => How Japanese speakers understand word & sentence meanings
Thank You!

- The research reported here is supported in part by Grant-in-Aid for Scientific Research (C) #24520437
Selected References

URLs

• Japanese FrameNet http://jfn.st.hc.keio.ac.jp/

• JFN data on FrameSQL
  http://framenet2.icsi.berkeley.edu/frameSQL/jfn23/notes/index.html

• Japanese FrameNet on YouTube
  http://www.youtube.com/watch?v=kfqR9aUcp1c