Japanese FrameNet:
Lexicon and Constructicon Building for Japanese

JAPANESE FRAMENET

Kyoko Hirose Ohara

Keio University

ohara@hc.st.keio.ac.jp

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

Japanese FrameNet

- Corpus
  - ➢ Analyzing and Annotating corpus data
- Annotation
  - ➢ Word meaning, Sentence meaning
- "Human" Language Processing
  - ➢ Speaker's understanding

Japanese FrameNet project

Toshio Ohori   (The University of Tokyo)

Seiko Fujii   (The University of Tokyo)

Ryoko Suzuki   (Keio University)

Hiroaki Saito   (Keio University)

Hiroaki Sato   (Senshu University)

Shun Ishizaki   (Keio University)

Current annotators & programmers

Toshiko Kigoshi
Anna Gladkova
Hiroya Hagino
Naoko Kurokawa
Hidetoshi Kobori
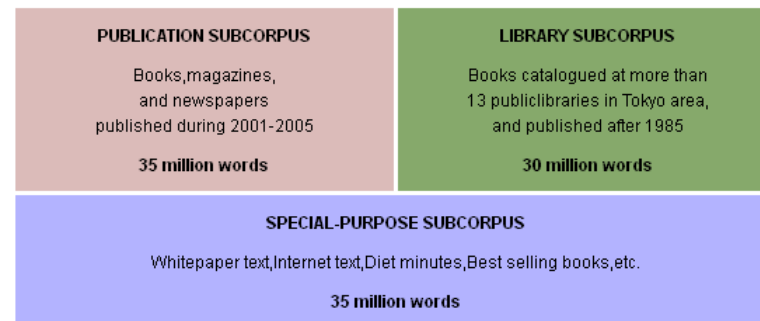Antoine Mousnier
Benoit Eudier

What is Japanese FrameNet
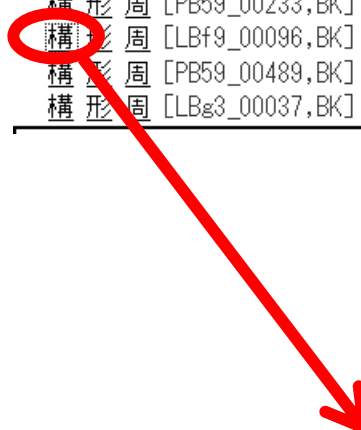
Japanese FrameNet (JFN): Input

- 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-milllion words of texts taken from:
        - Magazines, Newspapers, Government white papers, Books, Congress proceedings, Internet, and Textbooks

PUBLICATION SUBCORPUS	LIBRARY SUBCORPUS
Books,magazines, and newspapers published during 2001-2005	Books catalogued at more than 13 publiclibraries in Tokyo area, and published after 1985
35 million words	30 million words

SPECIAL-PURPOSE SUBCORPUS
Whitepaper text,Internet text,Diet minutes,Best selling books,etc.
35 million words

Structure of the BCCWJ

JFN-KWIC Concordance Program

構 形 周 [PB59_00233,BK] 一歩 **進む** ごとに頭がクラクラして、和葉はほとんど朦朧としながら導かれるままに連れて行かれた。 add remove

構 形 周 [LBf9_00096,BK] その廊下の角を左に折れ、突き当たりの一室の前まで、源一郎は、ほとんど小走りに **進ん** だ。 add remove

構 形 周 [PB59_00489,BK] 「ほとんどが名門の私立小学校へ **進む** から、地元の公立へ通う子供たちとは接点がない。 add remove

構 形 周 [LBg3_00037,BK] ２．市町村別データによる分析　地域コミュニティの基盤は全国にある約３４００の市区町村である．ここでは最近に

```
              --------------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。
```

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)**   *cf. Semantic Roles*

    – Participant (or prop) roles <u>of the frames</u> 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

Annotation Tool

Japanese FrameNet (JFN): Output Lexical Entry Report

Japanese FrameNet (JFN): Output
Full Text Annotation

- A semantically annotated corpus

14

Japanese FrameNet (JFN): Output FrameSQL

Sato (2012)

• Thesaurus

Experiencer_focus

Definition:

The words in this frame describe an Experiencer's emotions with respect to some Content. A Explanation for the emotion m 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翻訳 [EJ]]

Japanese FrameNet (JFN): Output Frame-to-Frame relations

- Ontology
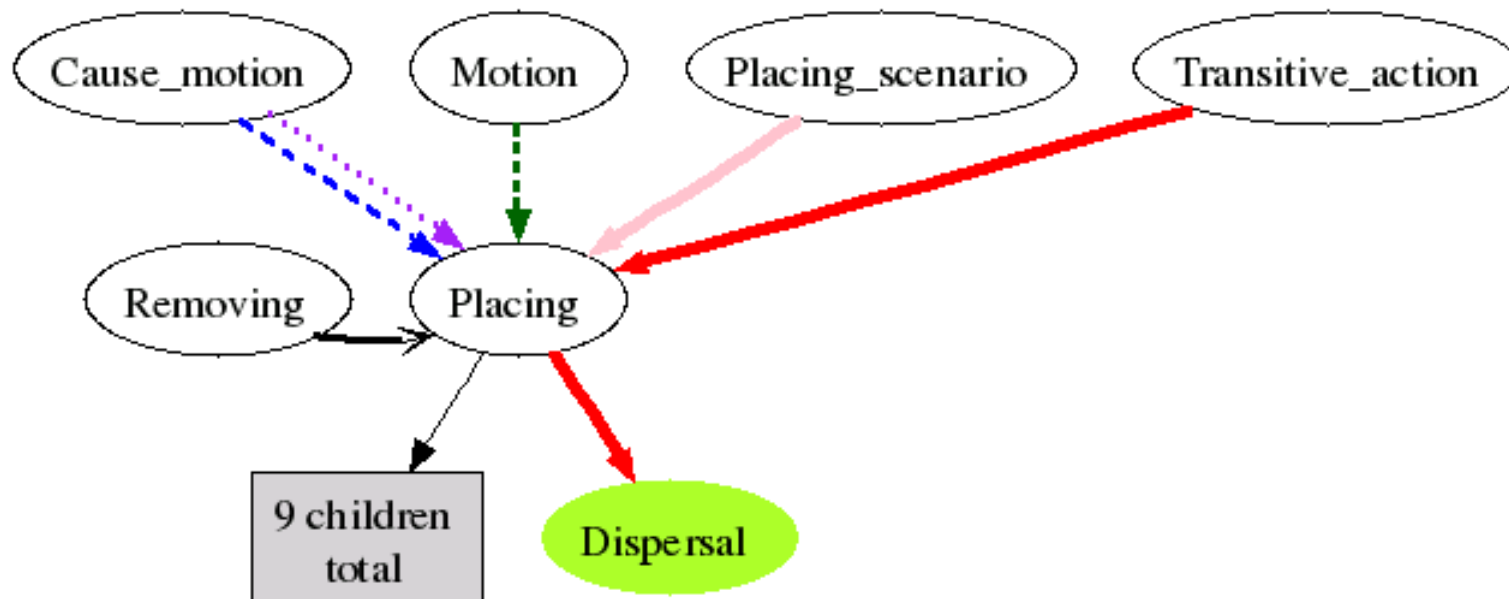
What are the theories
behind Japanese FrameNet

Frame Elements *vs.* Semantic Roles

Frame Elements (FEs) are <u>relativized to frames</u> and much <u>more fine-grained</u> 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

(1) *odorokasu.v* (to surprise) in *Daijirin: The Second Edition* (2006)

a. *seken o   odorokaseta ziken*

   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. <**Agent**> *ga*   <**Person**> *o*   *odorokasu*

               NOM   ACC surprise

    b. <**Causer**> *ga*   <**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. [< **Person** > *seken o*]   *odorokaseta* [< **Causer** > *ziken*]

        public ACC surprised   incident

  'the incident which surprised the public'

  b.   [< **Person** > *zimoku*   *o*]   *odorokasu*

      many_people's_attention ACC surprise

       'to surprise people'

(3') JFN annotations

  a. [EXPERIENCER *seken o*]   *odorokaseta* [STIMULUS   *ziken*]

  b. [EXPERIENCER *zimoku*   *o*]   *odorokasu*

25

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

30

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

- Baker, Collin. (2006). "Frame Semantics in Operation: The FrameNet Lexicon as an Implementation of Frame Semantics" ICCG4.

- Fillmore, Chalres J. (2006). "The Articulation of Lexicon and Constructicon" ICCG4.

- Fillmore, Charles J. (2008). "Border Conflicts: FrameNet meets construction grammar." In E. Bernal & J. DeCesaris (Eds.), *Proceedings of the XIII Euralex International Congress* (49-68). Barcelona: Institut Universitari de Lingüistica Aplicada.

- Fillmore, Charles J. and Collin F. Baker (2010) "A Frames Approach to Semantic Analysis," *The Oxford Handbook of Linguistic Analysis*, Heine, Bernd and Heiko Narrog (eds.), OUP.

- Fillmore, Charles J., Russell R. Lee-Goldman, and Russell Rhodes. (To Appear). "The FrameNet Constructicon" Boas, H.C. and Sag, I.A. (Eds.) *Sign-based Construction Grammar.*

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

- Hasegawa, Yoko, Russell Lee-Goldman, Kyoko Hirose Ohara, Michael Ellsworth, and Charles J. Fillmore. (2012). The Frames-and-Constructions Approach to Paraphrase. ICCG7.

- Lee-Goldman , Russell & Russell Rhodes. (2009). "Corpus-based analysis and annotation of constructions" Paper presented at Frames and Constructions Conference, 31 July-2 August 2009.

- Lonneker-Rodman, Birte. (2007). "Multilinguality and FrameNet" International Computer Science Institute Technical Report.

- Ohara, Kyoko Hirose. (2008). "Lexicon, Grammar, and Multiinguality in Japanese FrameNet". LREC2008.

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

- Ohara, Kyoko Hirose, Hiroaki Saito, Seiko Fujii, and Hiroaki Sato. (2011). "Japanese FrameNet: Building a lexical and constructional resource based on BCCWJ and Semantic Frames." (In Japanese).

- Ohara, Kyoko Hirose (2012). "Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English ". LREC2012.

- Ruppenhofer, Joseph, et al. (2010). "FrameNet II: Extended Theory and Practice."

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