Investigating Japanese FrameNet Data with FrameSQL

Kyoko Hirose Ohara & Hiroaki Sato
Keio University & Senshu University, Japan
ohara@hc.cc.keio.ac.jp, thc0408@gmail.com
5th September, 2010
Charles University, Prague, Czech Republic

Outline
1. Introduction
2. FrameNet (BFN) and Japanese FrameNet (JFN)
3. FrameSQL
4. Analyses of surprise.v & odorokasu.v
5. Conclusions
6. Future Work
References

1. Introduction

• Existing lexical resources
  – Still a widespread assumption that a small number of universally-defined semantic roles are sufficient and more preferable than frame elements (FE’s) defined in terms of specific semantic frames
  • E.g. “Verb Argument Structure Thesaurus” (VAST) (Takeuchi et al. 2010)
  • But cf. Matsubayashi et al. (2010)

1.1. Claims

• JFN is more suitable to analyze actual uses of words in corpora than other existing lexical resources, because in JFN annotations,
  1) not only core frame elements but also non-core frame elements are used;
  2) the information about which frame element is paired with which grammatical functions (GF’s)/phrase types (PT’s) is recorded as valence;
  3) frame elements are linked to each other by frame-to-frame relations.

2. FrameNet and Japanese FrameNet

2.1. FrameNet

• “The general purposes of the [Berkeley FrameNet] project are both to provide reliable descriptions of the syntactic and semantic combinatorial properties of each word in the lexicon, and to assemble information about alternative ways of expressing concepts in the same domain.” --- Fillmore & Baker (2010:321)

2.2. Japanese FrameNet

• Creating a prototype of an online Japanese lexical resource in the FrameNet style:
  • By describing the sense of each lexical unit with respect to the semantic frame it evokes;
  • By annotating corpus examples of each word with frame-semantic tags (frame elements).
• Research questions
  • To what extent is the Frame-semantic approach suitable for analyzing the Japanese lexicon?
  • To what extent are the existing English-based semantic frames applicable to characterizing Japanese lexical units?
• Currently annotating Balanced Corpus of Contemporary Written Japanese (BCCWJ)
3. FrameSQL

- a web-based application created originally for the Berkeley FN.
- puts together JFN and BFN lexical databases, and the user can access them seamlessly, as if they were a unified database.
- useful for comparing related frames and lexical units.

3.1. BFN search menu

- hyper link to JFN

3.2. JFN search results

4. Analyses of *surprise.v* & *odorokasu.v*

The Experiencer_obj frame in BFN & JFN:

Some phenomenon (the STIMULUS) provokes a particular emotion in an EXPERIENCER.

Core FE’s:
- EXPERIENCER: The EXPERIENCER reacts emotionally or psychologically to the STIMULUS.
- STIMULUS: The STIMULUS is the event or entity which brings about the emotional or psychological state of the EXPERIENCER.

Non-core FE’s:
- MEANS: The MEANS by which the STIMULUS affects the EXPERIENCER.
- TIME: The TIME when the EXPERIENCER has an emotion as caused by the STIMULUS.

4.1. FN: Phrase Types of STIMULUS X

1. NP: X surprised Y
2. PP: Y was surprised by X (Passive)
3. Sfin: It surprised Y [STIMULUS that S] (Extraposition)
4. VP to: It surprised Y [STIMULUS to VP] (Extraposition)
5. when: ... it always surprises me [STIMULUS when people turn out to be such bad listeners]
6. if: ... it wouldn’t surprise me in the least [STIMULUS if, if you analyzed this dream you discovered it had nothing to do with being at school and nothing to do with taking exams]

4.2. FN: when clauses with devastate.v in the Experiencer_obj frame

1. [EXPERIENCER A family I knew] were devastated [STIMULUS TIME when the thirty-five-year-old mother died after a short illness].
4.3. Frame-to-Frame Relations

- “every FE is defined in terms of specific frames”
- Problem: too many FEs
  - “At the moment (2008.10.1) there are slightly more than 9,000 distinct types of Frame Element in FrameNet.”
- Solution: Frame-to-Frame Relations
  - “But don’t panic! They are related to each other via [several kinds of] relations, and in most cases can be recognized as a subtype of a small set of FEs that resemble thematic roles.”

4.4. Relations of STIMULUS

- Emotions. STIMULUS shares some properties with other related FEs.

4.5. [STIMULUS when ...] in the Emotions_by_stimulus frame

1. LUs: glad, joyful
2. Sfin: X was glad [STIMULUS that S]
3. VPto: X was glad [STIMULUS to VP]
4. when: We were as glad as anything [STIMULUS TIME when we saw this rain coming down].

4.6. seeing with STIMULUS

1. [STIMULUS Seeing Sadako’s dance] (Sadako no odori o mi-TE) GEN dance ACC see-TE
   Toyama was much surprised.
2. [Means Starting the game] (geemu o hajime-TE) game ACC start-TE
   Deumi was surprised.
- Question: Is there any data to support this distinction?
- Answer: There are currently only three LUs in the Experiencer_obj frame in JFN, but Frame-to-Frame relations give us more data in related frames.

4.7. seeing with CONTENT

1. Experiencer_obj <=> Experiencer_focus
   - odorokasu (surprise.v) <=> odoroku (be_surprised.adj)
2. Experiencer_obj, STIMULUS <=> Experiencer_focus .CONTENT
3. [CONTENT] Seeing his hands] (Jibun no te o mi-TE) Mr. Maeda was surprised.
   - JFN tries to annotate the data consistently.

4.8. Annotating dictionary example phrases

   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’
4.9. 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.

4.10. Annotations of (1) in VAST and JFN

(2') VAST annotations
a. [<Person > seken o] odorokaseta [<Causer > ziken] public ACC surprised incident
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

4.11. Summary: Treatment of ‘peripheral’ phrases
• Our analysis of odorokasu (to surprise) shows that 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 describe them properly, just as BFN does.
  – JFN: [STIMULUS Seeing Sadako’s dance] (Sadako no odori o mitte), Tooyama was much surprised.
  – BFN: ... It always surprises me [STIMULUS when people turn out to be such bad listeners]

4.12. Annotating Corpus Sentences

(4) Sentence from the BCCWJ corpus
Sadako ga dansu o suru siin o soozosita koto mo nakatta tame, Sadako no odori o mite, did.not.exist SUB GEN 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.’

4.13. 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.

5. Conclusions
• FrameSQL highlights interesting and important valence patterns hidden in corpus data, showing that the FrameNet/Frame Semantics approach is suitable for analyzing Japanese texts.
6. Future Work

• Constr. Name: When- \textit{STIMULUS} Construction
  – Form: \([\text{STIMULUS} \ when \ S] \ ... \ [<\text{Emotion-related frame}> \ LU] \)
  – Meaning: EXPERIENCER experiences an emotion when S \textit{STIMULUS}.
• Constr. Name: \textit{STIMULUS} –TE Construction
  – Form: \([\text{STIMULUS} \ ... \ \text{VP-TE}] \ ... \ [<\text{Emotion-related frame}> \ LU] \)
  – Meaning: EXPERIENCER experiences an emotion when VP \textit{STIMULUS}.

References


URLs

• Balanced Corpus of Contemporary Written Japanese (BCCWJ) website
• FrameNet website
  http://framenet.icsi.berkeley.edu/
• FrameSQL website
• Japanese FrameNet website
  http://jfn.st.hc.keio.ac.jp/
• Japanese FrameNet on YouTube
  http://www.youtube.com/watch?v=kfqR9aUcp1c