

Japanese FrameNet: Lexicon and Construction Building for Japanese



JAPANESE FRAMENET

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

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

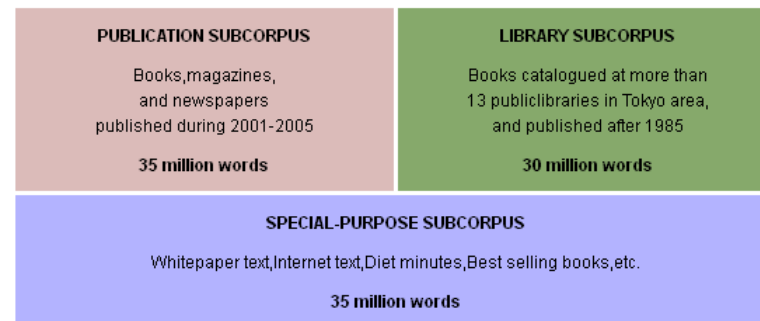


What is Japanese FrameNet



Japanese FrameNet (JFN): Input

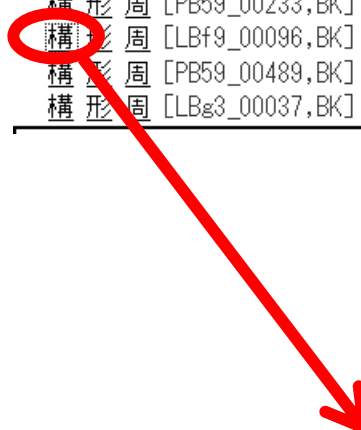
- Balanced Corpus of Contemporary Written Japanese (BCCWJ) by National Institute for Japanese Language and Linguistics (NINJAL)
 - the first available balanced and representative corpus of Modern Written Japanese (2011-)
 - Copyright-free
 - Contains 143-million words of texts taken from:
 - Magazines,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



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] 「ほとんどか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 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

Annotation Tool

Frame

Lexical Unit

Frame Elements

Goal <G>		Source <S>	
Theme <H>	DNI	Time <F2>	
Manner <M>		Cotheme <K>	
Means <F12>		Depictive <D>	
Mode_of_transportation ...		Goal_conditions <F8>	
Path <F5>			

Windows taskbar: スタート | 10May2007-2 | コマンド プロンプト - J... | Lexical Entry Report ... | Japanese FrameNet ... | 98% | 4:34

Japanese FrameNet (JFN): Output Lexical Entry Report

[Arriving](#)
[たどりつく.v](#)
[つく.v](#)
[入る.v](#)
[来る.v](#)
[着く.v](#)
[至る.v](#)
[Arson](#)
[Artifact](#)
[Assessing](#)
[Assistance](#)
[Atonement](#)
[Attaching](#)
[Attack](#)
[Attempt](#)
[Attempt suasion](#)
[Attention](#)
[Attention getting](#)
[Attributes](#)
[Avoiding](#)
[Awareness](#)
[Bail setting](#)
[Bearing arms](#)
[Become silent](#)
[Becoming](#)
[Becoming aware](#)
[Behind the scenes](#)
[Being attached](#)
[Being born](#)
[Being dry](#)
[Being employed](#)

Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns			
(1) TOTAL	Depictive	Goal	Theme	
(1)	AVP Dep --	NP Dep =	NP Ext ガ	
(1) TOTAL	Depictive	Theme	Time	
(1)	Sub Dep --	NP Ext ハ	AVP Dep --	
(1) TOTAL	Goal			

ColorsOff

- 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、**ヘンリー卿の着**たときは駅へも行ったし、ノーサンバランド・ホテルへ尾行したりして出歩いている間、細君はホテルの居間に監禁しておいたのだ。**DNI**
- やっと**着**いたか。**DNI DNI**
- 相互タクシー（大阪市城東区）でも、依頼場所に到着した際に「車が着きました」という連絡が客にできず、客から「**車は着**いたのか」といった苦情が数件寄せられたという。**DNI**
- 美智子がタクシーに飛び乗ってくれたんですが、なかなか**着**かないから僕は覚悟を決めた。**DNI**
- **携帯のメールが着**いたら、相手に自動連絡するような新手のサービスだろう。**DNI**
- **ひつぎが着**いてから、骨壺（つぼ）を持って遺族が退場するまで通常約1時間半と見込まれて

Japanese FrameNet (JFN): Output Full Text Annotation

- A semantically annotated corpus

全文テキストアノテーション

[PB56_00002.txt]

6. パイロット People,by,vocation 最たる Being,obligated 仕事 Documents は 資格 Documents 維持 Activity,ongoing 「その2」 Relative,time に 操縦 Operate,vehicle 関係 Relation の ライセンス Documents だが、これも 管制 Control 機関 Organization と 文信 Communication する 航空 Operate,vehicle 網 Network 無線 Communication,means 通信 People,by,vocation 士の ライセンス Documents と、操縦 Operate,vehicle そのものの ライセンス Documents とに 分かれる Being,separated
- ここで 操縦 Operate,vehicle そのものの ライセンス Documents に 話 Communication を 絞つ Place,weight,on て 進める Activity,ongoing と、パイロット People,by,vocation は 訓練 Education,teaching 生 Education,teaching となっ Becoming たその 日 Calendaric,unit から、約 Relational,quantity 三十 時間 Measure,duration の 飛行 Operate,vehicle 時間 Measure,duration で ようやく Place,weight,on (単独、) 飛行 Operate,vehicle が できる Capability ようになる。
- その後、自家用 Operate,vehicle 操縦 People,by,vocation 士 免許 Documents、事業 Businesses 用 操縦 People,by,vocation 士 免許 Documents、一切 Completeness 外 Part,inner,outer を 見る Perception,active ことなく 計器 Gizmo の みで Operate,vehicle 飛行 Operate,vehicle 計器 Gizmo 飛行 Operate,vehicle 証明 Documents、双発以上の エンジン Vehicle,subpart を 持つ Possession 飛行機を 操縦 Operate,vehicle できる Capability 陸上 Operate,vehicle 多発機などの 基本 Being,necessary 的な 免許 Documents を 約 Relational,quantity 四 年間 Measure,duration 飛行 Operate,vehicle 時間 Measure,duration に して 約 Relational,quantity 三百 数十時間 Measure,duration を かけて 取得 Getting する。
- ここで 言う Communication 「陸上 Vehicle 多発機」とは、海上 Relational,natural,features や 湖国 Relational,natural,features で 離着陸 Motion できる Capability 水上 Vehicle 機に対して、通常 Typicality の 滑走 Motion、Locale,by,use 路 Roadways を 使用 Using して 離着陸 Motion する 飛行機を 指す Referring,by,name 用語 Simple,name と 理解 Grasp していただきたい。
- しかし 航空 Operate,vehicle 法 Law では、これら 自家用 Operate,vehicle 操縦 People,by,vocation 士、事業 Businesses 用 操縦 People,by,vocation 士、計器 Gizmo 飛行 Operate,vehicle 証明 Documents、陸上多発の ライセンス Documents だけでは、まだ 最大 Extreme,value 離陸 Motion 重量 Dimension が五千七百キログラム(五、七トン、)未滿の 飛行機 Operate,vehicle し 操縦 Operate,vehicle でき Capability ない。
- 「最大 Extreme,value 離陸 Motion 重量 Dimension が五千七百キログラム」の 飛行機 Operate,vehicle とは…？ それは具体的に、十数人乗りの 飛行機 Operate,vehicle を イメージ Awareness していただくといいたい。
- 一方、定期 Frequency 航空 Motion で 使用 Using している 機材 Supply は、数十名乗りの 飛行機 Operate,vehicle から 最大 Extreme,value 五百六十八席の ジャンボ Vehicle ジェット (最大 Extreme,value 離陸 Motion 重量 Dimension は三十八万キログラム/三百八十 トン 前後 Relational,quantity) まであり、最大 Extreme,value 離陸 Motion 重量 Dimension が五、七トン以上の 大きさの 飛行機 Operate,vehicle を 操縦 Operate,vehicle するには、これらの ライセンス Documents では まだまだ 不足 Sufficiency で、それらの 飛行機 Operate,vehicle に 独自の ライセンス Documents を 取得 Getting しなければならない。
- 最大 Extreme,value 離陸 Motion 重量 Dimension を 基準 Being,necessary に 個々の 飛行機 Operate,vehicle に 個々の ライセンス Documents が 要求 Being,necessary される 理由 Reason は、飛行機 Operate,vehicle によって、それぞれの 設計 Config,with、思想 Thought や 飛行 Motion、特性 Degree が 大きく Degree 異なる Similarity ことによる。

その後、自家用操縦士免許、事業用操縦士免許、**一切の飛行機を** **計器のみで** **飛行** **できる** **計器飛行証明**、**双発以上のエンジンを持つ** **飛行機を** **操縦** **できる** **陸上多発機などの** **基本的な** **免許** **を** **約** **四** **年間**、**飛行時間** **にして** **約** **三百** **数十時間** **を** **かけて** **取得** **する**、**陸上** **多発機**



Japanese FrameNet (JFN): Output FrameSQL

Sato (2012)

• Thesaurus

Commerce_pay
Contingency
Departing
Desirability
Existence
Expectation
Experiencer_focus
Experiencer_obj
Explaining_the_facts
Fluidic_motion
Grant_permission
Motion
Motion_directional

Search Reset Gloss

POS
Sort&display
Phrase Type
Gram. Func.
Other Tags
Display limit 9999

[BFN]

Lexical Units Frame Elements

あきる.V
あきれる.V
おそれる.V
おどろく.V
ひるむ.V
悔る.V
喜ぶ.V
怯える.V
恐れる.V
恥じる.V
悔いる.V
愛す.V

Beneficiary
Circumstances
Content
Degree
Event
Event+Expressor+State+Topic
Experiencer
Experiencer+Content
Explanation
Expressor
Manner
Parameter

Experiencer_focus

Definition:

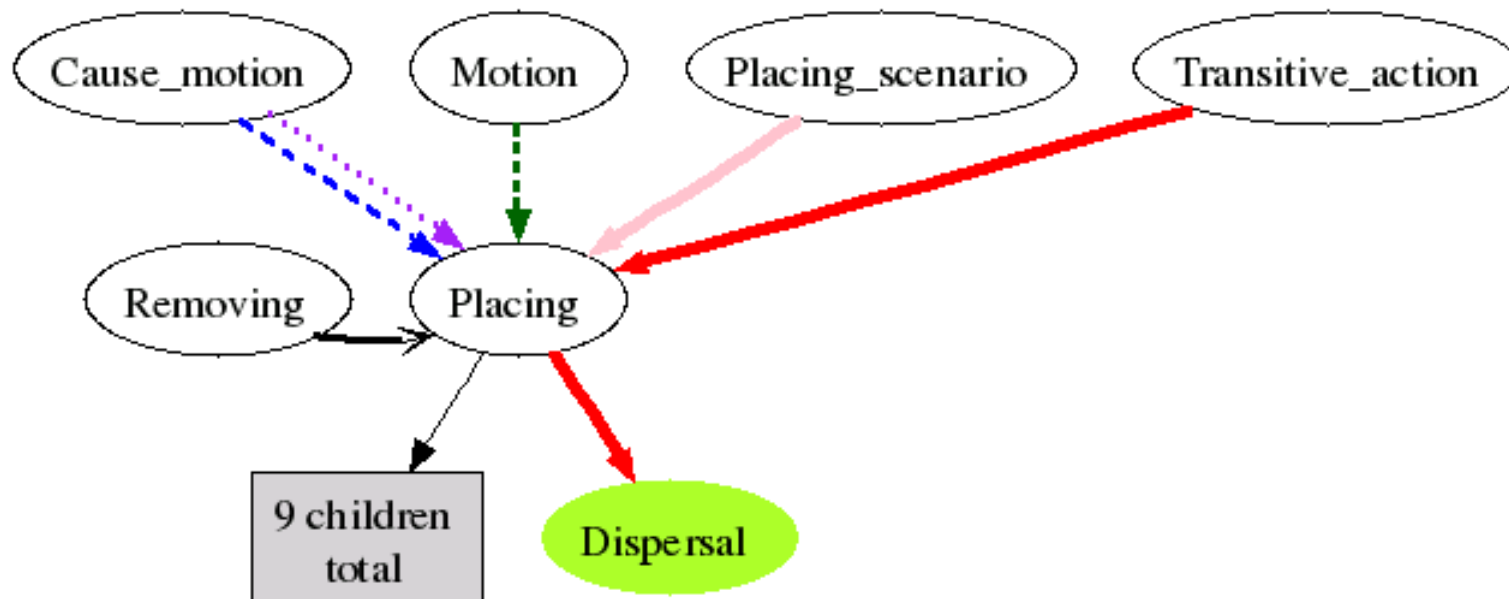
The words in this frame describe an **Experiencer**'s emotions with respect to some **Content**. A **Explanation** for the emotion in current state of affairs, quite often it refers to a general situation which causes the emotion.

- 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 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 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.”



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**.

Annotating Corpus Sentences

(4) Sentence from the BCCWJ corpus

Sadako ga dansu o suru siin o sozosita 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 sozosita 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 sozosita 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

- Baker, Collin. (2006). "Frame Semantics in Operation: The FrameNet Lexicon as an Implementation of Frame Semantics" ICCG4.
- Fillmore, Charles J. (2006). "The Articulation of Lexicon and Construction" 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 Lingüística 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 Construction" 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.), *Multilingual 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>

