

日本語フレームネット：意味タグ付きコーパスの試み

小原 京子 (慶大), 大堀 壽夫 (東大), 鈴木 亮子 (慶大),
藤井 聖子 (東大), 斎藤 博昭 (慶大), 石崎 俊 (慶大)

1. はじめに

本論文では, 語彙情報資源構築プロジェクト日本語フレームネット(JFN)¹の意味タグ体系について述べる(藤井・小原 2003, Ohara et al. 2003, 2004, 石崎 2004). 第2節ではJFNの背景と概要を示す. 第3節ではJFNのコンテンツについて説明する. 第4節ではJFNを意味体系の観点から他の電子化資源と比較し, 第5節ではそれらに記されていない新たな知見を報告する.

2. JFNの概要

JFNは, 2002年から始めた日本語語彙情報資源構築プロジェクトであり, カリフォルニア大学バークレー校において開発中のフレームネットプロジェクト(The FrameNet Project, 以下FN)との連携のもとに進められている(Fontenelle ed. 2003, Fillmore and Baker 2004)². その目的は, コーパスデータを基に精度の高い語の意味・用法の分析を行い, 電子化語彙情報を蓄積することである. 具体的には, 次のような特徴をもつ. (i) フレーム意味論を意味記述の枠組みとして採用している(Fillmore 1982, 1992). これによって語と語の間の意味関係がより適切に捉えられる. (ii) コーパスデータにFEの情報, および構文環境の情報をタグの形で付与することで解析を進める. 結果として得られるのは, 語彙情報資源と共に, 意味タグの付いた解析済みコーパスである.

フレーム意味論の優れた点は, 日常世界についての慣習化された知識のモデル, すなわち意味フレームを参照しつつ語彙を記述する点である. 語や構文のもつ意味は, 日常世界のモデルとしての意味フレームを背景として, 一部を焦点化することで得られる. フレーム意味論における「意味フレーム」の概念は従来の「格フレーム」よりもずっと粒度が高く, 認知的妥当性をもったものである(cf. Fillmore 1968, 1977および第4節参照).

例えば, “risk”という語のコーパス分析において, Fillmore and Atkins (1994)は Protagonist, Bad, Decision, Goal, Setting, Possession, Sourceのフレーム要素(Frame Element, 以下FE)によってRISKフレームを規定し, “run a risk”と“take a risk”の相違を, BadとDecisionのどちらのFEが焦点化されるかという違いによって捉えた:

(1) Newborn babies run the risk of hypothermia [=Bad].

[TAKEによる代替は不可]

(2) He took the risk of jumping the cliff [=Decision].
[RUNによる代替は不可]

JFNも同様に, 意味フレームを規定してFEのタグ付けを行うという方法論によって日本語の語彙を分析する. と同時に, 英語の記述のために作られたFNの枠組みが日本語データにどこまで適用できるかを検討・吟味している.

3. JFNのコンテンツ

JFNのコンテンツは, 以下の5種類から成る. 1) フレーム辞書; 2) コーパスデータに基づく語義; 3) 意味タグ付き例文; 4) 結合価パターン情報; 5) 検索のためのインターフェースである(cf. Ruppenhofer et al. 2002). 1)のフレーム辞書とは, 話者がある語彙項目(lexical unit, 以下LU)を用いる際に想起する意味フレームの定義(どのようなFEが関与しているかを含む)を集め, 相互の意味フレーム間の関係を記述したものである. JFNでは, 英語の語義記述のためにFNで想定されている意味フレームやそれらのFEが, 日本語の語義記述のためにも必要であると考えられる場合には, 可能な限りFNと同じ意味フレームとFEを想定して日本語語彙の語義記述を行っている³.

2)のコーパスデータに基づく語義に関しては, 従来国広 1997, 國廣編 1982(2003), 柴田編 1976(2002), 1979(2003), 森田 1989などの日本語基本語彙の意味に関する記述に, コーパスデータを詳細に検討することで新たに知見を付け加えられるかどうか焦点を当てて分析している.

3)の意味タグ付き例文とは, 各LUについて, 該当意味フレームに典型的な統語的文脈やコロケーションを網羅するよう多様な例文をコーパスデータから選び, 各文において具体的に, i) FEがどの句で実現されているか; ii) その句の句タイプは何か(名詞句, 副詞句など); iii) その句が名詞句の場合, どの助詞をとるかなどの情報をタグ付けしたものである. JFNおよびFNにおけるタグ付け例を(3), (4)に示す.

(3) JFNにおけるタグ付け例

台湾人が 香港経由で	中国に	<u>渡る</u> .
<Theme> <Path>	<Goal>	
名詞句 副詞句	名詞句	
が 経由で	に	

¹ <http://www.nak.ics.keio.ac.jp/jfn/ja/index.html>

² <http://www.icsi.berkeley.edu/~framenet/>

³ ただし, FNで定義されたFEには日本語の語義記述のためには不十分と思われるものもある(cf. Ohara et al. 2003, 2004).

小学生が 青信号で 横断歩道を 渡る。
 <Theme> <Path> <Goal>
 名詞句 名詞句 前置詞句
 が を へ

(4) FN におけるタグ付け例

We crossed the border into Italy.
 <Theme> <Path> <Goal>
 名詞句 名詞句 前置詞句
 into

I crossed the desert on foot.
 <Theme> <Path> <Goal>
 名詞句 名詞句

ここでは、Motion (移動) フレームに關与する日本語動詞「わたる」と英語動詞“cross”を含む文に、この意味フレームを構成する FE 名をタグ付けした例を示している⁴。Motion フレームとは、「ある主題 Theme が、ある場所 Source から、他の場所 Goal へ、その 2 点をつなぐスペース Path を経由して動く」事象に關する意味フレームである。

4) の結合価パターン情報とは、JFN では、単に述語が表層的にどのような助詞をとるかだけでなく、その述語に係る句の句タイプは何か、またその句は該当する意味フレームにおいてどの FE を具現化しているかについて情報を組み合わせたものを指す。このような LU ごとの結合価パターンに關する情報をアノテーション済みの用例から Lexical Entry Report として抽出できるようにしている。図 1 に、その一例を示す。

Number Annotated	Path	Theme	Source	Goal	Vehicle
5.000 TOTAL	Path	Theme			
2.000	NP	NP			
	を	が			
1.000	NP	NP			
	を	DNI			
2.000	NP	NP			
	を	は			
1.000 TOTAL	Path	Theme	Manner		
1.000	NP	NP	NP		
	を	DNI	で		
2.000 TOTAL	Path	Theme	Source		
2.000	NP	NP	NP		
	を	DNI	から		
1.000 TOTAL	Path	Theme	Source	Vehicle	
1.000	NP	NP	NP	NP	
	を	DNI	から	で	

図 1 結合価パターン情報

図 1 の右側は、コーパスから抽出した Path_shape (経路形態) フレームに關与する日本語動詞「越える」の例文に

⁴ 厳密には、「わたる」と“cross”は、Motion フレームより具体的な別フレーム Path_shape (経路形態) フレームに關与していると考えられる。これらの動詞は、ある主題が移動する際にできる経路の形態との関連で具体的な移動事象を表現するからである。

意味タグ付けを施したものである。左側の表は、既に意味タグ付けされたこれらの例文から結合価パターン情報を抽出した Lexical Entry Report の一部である。DNI とは、Definite Null Instantiation の略で、限定指示的ゼロ表示を示す。ある FE が、該当文中には現われないが具体的に何を指しているかを先行文脈より復元できる際に付与すると言うように定義されている⁵。

最後に、5) の意味タグ付きコーパスデータ検索のインターフェースについては、FrameSQL⁶ と呼ばれるフレームネットデータベース検索用インターフェースを用いて FN の英語語彙や SFN のスペイン語語彙に關する情報同様検索できるよう、意味タグ付きコーパスを構築中である。

4. 他資源の意味体系との比較

JFN 以外の、日本語の結合価情報を記述している言語資源としては、『計算機用日本語基本動詞辞書 IPAL (Basic Verbs)』(以下 IPAL)(情報処理振興事業協会技術センター1987)、『日本語語彙体系』(池原他編 1999)、『日本語動詞の結合価』(荻野他 2003) などがある。JFN の結合価パターン情報がこれらの言語資源の結合価情報と根本的に異なるのは、フレーム意味論の枠組みを前提としていること、すなわち意味フレームの概念を用いていることである。上記の言語資源の背後にある意味の体系と JFN が前提としているフレーム意味論の違いは、フレーム意味論では、1) 各 LU が抽象的・先験的に規定された概念に対応づけられるのではなく、日常活動を通じて形成された典型的場面についての構造化された知識である意味フレームに關連付けされること⁷; 2) これらの意味フレーム同士が Inheritance (継承) などの関係によって体系付けられていること; 3) 動詞の項に、あらかじめ設定された少数の、粒度の低い意味役割 (Agent, Patient など) のどれかを割り当てるのではなく、該当する意味フレームに即して精密に定義される粒度の高い FE を割り当てていくことである。この意味の体系に關する違いが、すなわち上記言語資源と JFN との違いである。

ここではいずれかの優劣を議論するつもりはないが、JFN の利点の一つとして考えられるのは、他言語の言語資源とのリンクの可能性である (第 3 節参照)。既に述べたとおり、同じフレーム意味論の枠組みで、英語・スペイン語・ドイツ語を記述する試みである FN・SFN・ドイツ語フレームネットプロジェクトが既に開始されている。JFN では他言語における語義記述のために定義されたフレームや FE が日本語語義記述に適用できるのかを詳細に検討しつつも、適用できる場合にはなるべく適用していく方針であり、従って将来的にはこれらの他言語フレームネットとリンクさせることが可能と考えられる。

以上見たように、JFN と先行する結合価情報を含む他

⁵ 詳しくは以下を参照のこと。
<http://www.icsi.berkeley.edu/~framenet/book/book.html#id19>
 ただし日本語においてDNIの概念がそのまま通用するかどうかは検討の価値がある。
⁶ FrameSQL は、佐藤弘明氏 (専修大学) によって開発され、現在 FN および SFN データが検索可能である (Subirats and Sato 2004)。
<http://sato.fm.senshu-u.ac.jp/fn22/notes/index2.html>
⁷ 詳しくは以下を参照のこと。
<http://www.icsi.berkeley.edu/~framenet/book/book.html#frameRels>

の言語資源は、動詞の項に付与する意味情報に明らかな違いがある。しかしながら、これらの言語資源を JFN と相補的に使用することは可能であるし、使用者にとっても有益であると考えられる。

5. JFN の分析例: 評価を伴う伝達動詞について

次に、評価を伴う伝達動詞「おこる」を題材に、コーパスデータに基づく JFN の分析例を紹介する。JFN では、目下『CD-毎日新聞(データ集)』11年分を中心に書き言葉データを分析用コーパスとして使っている。JFN 分析用コーパスと検索用 JFN kwic ツールの結合価値情報抽出機能を用いることにより、既存の電子化資源には含まれていない結合価値をシステムティックに見出せる。たとえば、『日本語動詞の結合価値』では、「腹を立てる」と言う語義の「おこる」は、「に」格名詞句を伴うとされているが、「を」格名詞句については記述がない：

- (5) 左翼は、今でもサンチェス解任に怒っている。
(『日本語動詞の結合価値』より)

JFN コーパスでも「に」格名詞句を伴う「おこる」の用例は多く見られるが、「を」格名詞句を伴う例も見つかった⁸：

- (6) 恩納(おんな)村の比嘉昇一さん(78)は、米軍基地が減らない現状を怒る。

JFN では、目下これを Experiencer_obj フレーム(Experiencer に対して、ある事柄(stimulus 刺激)がある感情を引き起こす)に關与する「おこる」であると分析している。「を」格名詞句はこの意味フレームにおける FE Stimulus を具現化しているとし、FE をタグ付けする：

- (6') Experiencer_obj フレーム：

恩納(おんな)村の比嘉昇一さん(78)は、
<Experiencer>

米軍基地が減らない現状
<Stimulus>

怒る。

次に、『日本語動詞の結合価値』には「強くしかりつける」という語義の「おこる」の例文として、「お袋はそれを知ってえらく怒った」などがあり、「~と」は「おこる」の結合価値として記述されていない。ところが、JFN コーパスでは、「時々学校を休む女の子に『なぜ休むんだ』と皆の目の前で厳しく怒った」など、「~と」と共起する「おこる」の用例が多く見られる。JFN では、これを Judgment_direct_address フレーム(Communicator が Addressee に関する判断を直接 Addressee に伝える。その判断とは、ある Reason か Topic に関するものである)

⁸ 以下の JFN コーパスデータの原典は『CD-毎日新聞 2002 年版』である。

に關与する「おこる」の用法として分析している。「~と」は、Communicator が Addressee に判断を伝える際の Message を具現化していると考えられる：

- (7) Judgment_direct_address フレーム：

[DNI] 時々学校を休む女の子に
<Communicator> <Addressee>

「なぜ休むんだ」と 皆の目の前で
<Message> <Place>

厳しく怒った。

さらに、「おこる」を含む用例文における「に」格名詞句の検討を行った。「に」格名詞句は、(8)においては怒りの感情を引き起こす Stimulus、(7)では非難という価値判断を伝える相手 Addressee を指している⁹：

- (8) 2月、成田空港の出国手続きの非効率ぶりに怒った。(7)

時々学校を休む女の子に『なぜ休むんだ』と皆の目の前で厳しく怒った。

従って、(8)は Experiencer_obj フレームに、(7)は Judgment_direct_address フレームに關与していると分析される¹⁰：

- (8') Experiencer_obj フレーム：

[DNI] 2月、
<Experiencer> <Time>

成田空港の出国手続きの非効率ぶりに 怒った。
<Stimulus>

- (7'') Judgment_direct_address フレーム：

[DNI] 時々学校を休む女の子に
<Communicator> <Addressee>

「なぜ休むんだ」と 皆の目の前で
<Message> <Place>

厳しく怒った。

(7'')の「おこる」の用法は、Judgment_direct_address フレームに關与しているという点で、「母は『そんなことで泣くことはおまへんで』と叱った」などにおける「しかる」の用法に似通っている。

以上をまとめると、JFN kwic ツールの結合価値情報抽出

⁹ と同時に、(7)の「に」格名詞句内の修飾節「時々学校を休む」は非難という価値判断を下すに至った Reason を具現化していると考えられる。

¹⁰ ただし、「寝間着や浴衣を左前、おまけに帯も縦結びにされたお年寄りが若いヘルパーに怒る」のように両方の意味フレームに關与していると分析されるケースもある。

機能を用いることにより、既存の電子化資源には含まれていない結合価をシステムティックに発見することができた。特に、「～と」のように、構文上の役割において「が」「を」のような中心的な助詞ではないと考えられるものでも、コーパスにおける用例の豊富さからみればより細密な分析を要すると考えられるものがあることが明らかとなった。また、コーパスの膨大な用例を意味フレームを設定して分析すると、表層格が同じ場合にも、意味的な相違をFEに照らして分析できる。さらに、一つの動詞の意味の違い(「おこる」)や、同様の用法を持つ動詞間の類似点や相違点(「しかる」と「おこる」)なども分析できる。

6. おわりに

本論文では、フレーム意味論を意味分析の枠組みとし、その分析結果を意味役割タグ付きコーパスとして蓄積しようとしている語彙情報資源構築プロジェクトJFNについて述べた。JFNでは、このような方法論により既存の資源に記されていない事実を発見することができた。今後、FNで定義された意味フレームやFEがどこまで日本語語彙の意味記述に適しているかを検討していく予定である。

謝辞

本研究は、文部科学省学術フロンティア推進事業・慶應義塾大学「超表象デジタル研究センター」プロジェクト助成金により行われた。

主要参考文献

- Fillmore, Charles J. (1968). "The case for case." Bach, E. and Harms, R. (Eds.). *Universals in Linguistic Theory*. Holt, Rinehart, and Winston, New York. 1-88.
- Fillmore, Charles J. (1977). "The case for case reopened." Cole, Peter and Sadock, Jerrold (Eds.). *Syntax and Semantics*. Vol.8: Grammatical Relations. Academic Press, New York. 59-82.
- Fillmore, Charles J. (1982). "Frame semantics." *The Linguistic Society of Korea* (Ed.). *Linguistics in the Morning Calm*. Seoul: Hanshin. 111-137.
- Fillmore, Charles J. (1992). "'Corpus linguistics" or "Computer-aided armchair linguistics".' *Directions in Corpus Linguistics: Proceedings from a 1991 Nobel Symposium on Corpus Linguistics*. Mouton de Gruyter, Stockholm, 35-60.
- Fillmore, Charles J. and Atkins, B.T.S. (1994) "Starting where the dictionaries stop: The challenge for computational lexicography." Atkins, B.T.S. and Zampolli, A. (Eds.). *Computational Approaches to the Lexicon*. Oxford University Press, Oxford, 349-393.
- Fillmore, Charles J. and Baker, Collin. (2004). "The Evolution of FrameNet Annotation Practices." Fourth international conference on Language Resources and Evaluation (LREC 2004). Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora", 1-8.
- Fontenelle, Thierry. (Ed.). (2003). Special Issue: FrameNet and Frame Semantics. *International Journal of Lexicography*. Vol.16, Special Issue 3, Oxford, Oxford University Press.
- 藤井聖子, 小原京子. (2003). 「フレーム意味論とフレームネット」, *英語青年*, 149(6), 373-376.
- 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦 (編). (1999). *日本語語彙体系 CD-ROM 版*. 岩波書店.
- 石崎俊(2004). 「コンピュータ時代を支える言語学: コンピュータによる言語理解へ向けて」, *言語*, 33(12), 42-49.
- 情報処理振興事業協会技術センター(1987). *計算機用日本語基本動詞辞書 IPAL (Basic Verbs)*.
- 国広哲弥. (1997). *理想の国語辞典*. 大修館書店.
- 國廣哲彌, 柴田武, 長島善郎, 山田進, 浅野百合子. (1982(2003)) *ことばの意味 3: 辞書に書いてないこと*. 平凡社.
- 森田良行. (1989). *基礎日本語辞典*. 角川書店.
- 荻野孝野, 小林正博, 伊佐原均. (2003). *日本語動詞の結合価*. 三省堂.
- Ohara, Kyoko Hirose, Seiko Fujii, Hiroaki Saito, Shun Ishizaki, Toshio Ohori, and Ryoko Suzuki. (2003). "The Japanese FrameNet Project: A Preliminary Report." *Proceedings of Pacific Association for Computational Linguistics (PACLING'03)*, 249-254.
- Ohara, Kyoko Hirose, Seiko Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki. (2004). "The Japanese FrameNet Project: An introduction." *Fourth international conference on Language Resources and Evaluation (LREC 2004). Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora"*, 9-11.
- Ruppenhofer, Josef, Baker, Collin F., and Fillmore, Charles J. (2002). "Collocational Information in the FrameNet Database." Braasch, Anna and Povlsen, Claus (Eds.). *Proceedings of the Tenth Euralex International Congress*. Copenhagen, Denmark. Vol.I: 359-369.
- 柴田武, 國廣哲彌, 長島善郎, 山田進. (1976(2002)). *ことばの意味: 辞書に書いてないこと*. 平凡社.
- 柴田武, 國廣哲彌, 長島善郎, 山田進, 浅野百合子. (1979(2003)). *ことばの意味 2: 辞書に書いてないこと*. 平凡社.
- Subirats, Carlos, and Sato, Hiroaki. (2004). "Spanish FrameNet and FrameSQL." *Fourth international conference on Language Resources and Evaluation (LREC 2004). Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora"*, 13-16.

データ

CD-毎日新聞(データ集)92年版~2002年版 毎日新聞社.