

意味の包含関係に基づく動詞項構造の細分類

竹内 孔一†, 乾健太郎¶, 竹内奈央‡, 藤田篤§
 岡山大学大学院† 奈良先端科学技術大学院大学¶
 言語アナリスト‡ 名古屋大学大学院§

koichi@cl.cs.okayama-u.ac.jp, inui@is.naist.jp, fujita@nuee.nagoya-u.ac.jp

1 はじめに

近年、係り受け解析の精度が向上し文を構成する動詞と名詞との関係、つまり項構造が同定できるようになってきた。項構造レベルでは表現の間で事態の推論を行う基礎となる命題的な意味関係が存在する。例えば「XがYを曲げる」は「Yが曲がる」ことを包含し、「XがYに就職する」ならば「XがYに所属する」ことを包含するであろう。こうした項構造レベルでの意味の包含関係を辞書として構築することが出来れば、言い換えだけでなく、ある行動や動作が他のどのような事態を引き起こすかといった推論の基礎データを提供すると考えられる。

一方、動詞の項構造レベルの意味を分析する手法として語彙概念構造 (lexical conceptual structure (以降 LCS))[4][14][9] が議論されてきた。LCS は動詞の意味を動詞がとる項の間で分解することで多面的な意味記述を行う。例えば、「XがYをZに移転する」の場合、[X CAUSE [BECOME[Y BE [AT Z]]]] と構造化して記述する。[Y BE [AT Z]] は「YがZに存在する」という結果状態を表し、その上位に [BECOME] が BE 以下を包含することで変化を表す。さらに上位に [X CAUSE] が BECOME 以下を包含することで X がその変化を起こすことを表す。しかしながら、上記の意味述語と構造の全体像はまだ議論中 [2][10][12] であるため不明である。また上記の構造化では「移動する」も「就職する」も全く同じとして扱われてしまうことから言語処理の推論として扱うには粒度が粗すぎる。こうした理由から LCS をそのまま辞書構築に利用することは出来ない。

そこで本研究では LCS の持つ包含関係に着目し、より詳細化した分類体系を構築することで動詞項構造の語義を多面的に記述する動詞項構造シソーラスを提案する。大語彙 (7473 語義 (4425 語)) に対する人手による分析から活動、状態変化、状態の分類を行い 940 種類の意味分類を構築した。前回の報告 [11] では単に分類の一部を示したのみであるが本稿では半統制された述語集合によって記述された動詞の意味構造を利用することで LCS の包含関係を保ちつつ詳細化した分類体系を構築できたことを明らかにする。以下では、語彙概念構造からどのように意味特性と語彙化を行

うかについて記述し、他の言語資源である FrameNet, WordNet, VerbNet, EDR と比較を行うことで本言語資源の価値を明らかにする。

2 語の構成的意味分解のコード化

2.1 LCS の利点と問題点

LCS は動詞の語の意味を構造として捉えることで語同士の持つ意味関係を抽象化したフレームで整理する理論である。基本的に動詞の取る項と意味述語の組み合わせで記述し、アスペクト分析 [8] や認知意味論 [6] で議論されてきた動詞の性質を記述する。この構造的な語彙分解によって、上位事象と下位事象、さらに活動性、状態変化性、状態性という動詞の部分的な特性を記述することができるため、それぞれのレベルで特性が似ている動詞の複合的な関係を結びつけることができる。例を図 1 に記述する。

図 1 の左側では上位・下位事象を同時に持つ他動詞「xがyをzに置く」が下位事象だけの自動詞「yがzに移動する」を包含し、さらにその結果状態として「yがzに在る」という状態動詞の意味を包含することを示している。このように動詞の意味を構造として捉えることで物の移動や変化の含意関係を計算するための基礎データとして期待できる。

しかしながら、現段階では BE などの意味述語の粒度が粗く、自然言語処理や知識処理で必要と考えられるより詳細な分類については理論展開されていない。よって事態の推論を行うための基礎データとして LCS の枠組みをそのまま利用することはできない。

2.2 細分類化と包含関係の記述

2.2.1 細分類化

上記で述べた意味構造による包含関係を細分化するために、階層的な分類を構築して対応させる。細分類には大きく分けて 2 種類ある。一つは活動動詞に対して行動の manner の違いによって細分類するものであり、もう一つは状態変化動詞に対して、結果状態を中心に細分類するものである。この様子を図 1 の左と真

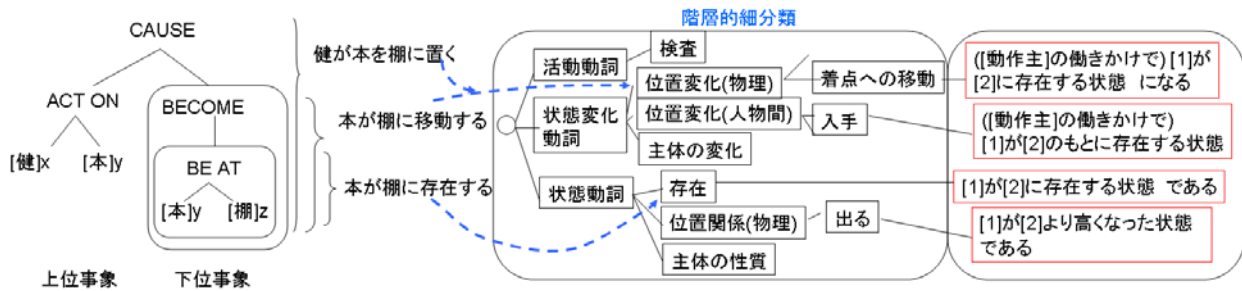


図 1: 詳細化と半統制意味構造記述による辞書構造

ん中の図で説明する。上位事象しか持たない動詞 (活動動詞) は階層のトップレベルにある【活動動詞】¹の細分類に対応させる。一方で、状態変化動詞は上位・下位事象があるものと下位事象のみのものであるがこれらは全て【状態変化動詞】の細分類として記述する。

ここで図1左側のLCSの構造からすると、BECOMEを【状態変化動詞】以下で結果状態で分類することからつまりBE AT (状態動詞)のタイプで分けることになるため、状態動詞の細分類は理想を言えば【状態変化動詞】の細分類における部分木を構成してもおかしくはない。しかしながら、多数の動詞を検査すると状態変化の分類は変化結果だけではなく、変化対象、ある変化に至る manner, 条件、背景が異なると分類が異なるため、それらを考慮した分類体系と状態動詞の分類体系は完全にはマッチしない。そこで、図1のBE ATに対する分類として【状態動詞】をトップレベルの階層に仮定して、状態動詞の細分類を行った。

階層的な分類により動詞間の包含関係がどの程度詳細に分かれたか具体的な事例をあげて説明する。「登場する」「現わす」「出現する」は【状態変化あり:位置変化:位置変化(物理):着点への移動:出現】²と分類する。一方、「通す」「案内する」「導く」は【状態変化あり:位置変化:位置変化(物理):着点への移動:導き】として分類する。これによりこれら2つの動詞グループは状態変化でかつ位置変化が起こり、着点に移動することを示している。異なる点は、移動の背景として『出現』か『導く動作』の違いであることを記述している。

階層を利用して反意語関係も同様に記述する。反意語はある共通する概念が存在するのでそこまでを上位階層でまとめてから細分類する。例えば動詞「曲げる」「丸める」の分類は【状態変化あり:対象の変化:物理的变化:形状(曲げる/伸ばす):曲げる】であり、一方、「伸ばす」「張り切る」は【状態変化あり:対象の変化:物理的变化:形状(曲げる/伸ばす):伸ばす】と分類している。

このように小分類1で反意語関係を並べる分類を用意することで反意語の記述を行うことができる。

2.2.2 半統制意味構造記述

分類体系は木構造であるので図1の階層をたどることで包含関係が得られる。しかしながら、状態動詞と状態変化動詞はもともと包含関係があるにもかかわらず木構造として別の枝となっているため関係を陽に記述することができない。さらに、各分類において項と対応関係が記述できていないため、他の分類との項の対応が明らかではない。この2つの問題を解決するために半統制された述語集合によって記述された動詞の意味構造(以降、半統制意味構造記述)を導入することで分類間の横断的な対応関係を記述する。半統制意味構造記述は状態動詞ならば「～状態である」と記述し、状態変化動詞ならば「～状態になる」としてLCSで見られるBECOMEとBEの包含関係を記述する。例えば図1に示すように「置く」の分類【状態変化あり:位置変化:位置変化(物理):着点への移動】に対して半統制意味構造記述は

([動作主]の働きかけで)
[1]が[2]に存在する状態 になる

と記述する。ここで()は後に処理を容易にするための区切りである。また、[1]と[2]は項であり、次節で説明するように例文の表層格に対応する。これにより[2]は着点であることがわかる。一方で、「存在する」の分類【状態動詞:位置:存在】に対して半統制意味構造記述は

[1]が[2]に存在する状態 である

と記述する。これにより「[]が[]に存在する状態」という文字列の検索を行うことで「存在する」という動詞と「移動する」「置く」との対応関係が項構造を保持したまま関係付けすることが出来る。半統制意味構

¹提案する辞書の分類を【】で表す。

²トップから細分類までの階層を“:”で区切って示す。

造記述を利用することにより、分類の木構造だけではうまく捉えられなかった意味分解の包含関係を多元的に記述することができる。

3 辞書エントリーの具体例

辞書は語義を基本の単位として構成しており、現在は語義の粒度は NTT コミュニケーション科学基礎研究所が開発した Lexeed に対応している。具体的な構成要素は表層、読み、代表例文、表層格、意味役割、格に対して半統制意味構造記述で利用する項番号、意味分類と半統制意味構造記述である。図 2 に記述の具体例を示す。

Lexeed ID	単語	読み
20035550-4	抜ける	ぬける
分類		
【状態変化あり: 主体の変化:属性変化:所属:引退】		
代表例文		
メンバーがグループを抜ける		
表層格	意味役割	表層 項番号
が	動作主	メンバー 1
を・から	起点	グループ 2
半統制意味構造記述		
([動作主] の働きかけで) ([1] が [2] から引退し) [1] が [2] に所属してない状態 になる		

図 2: 「抜ける」の辞書エントリー

図 2 では「メンバーがグループを抜ける」の意味での動詞「抜ける」の辞書エントリーである。分類は【状態変化あり: 主体の変化:属性変化:所属:引退】の 5 階層で記述される。「メンバーが」の意味役割は動作主、「グループを」は起点である。半統制意味構造記述から「ガ格の動作主がヲ格に所属していない状態」であることが示される。

半統制意味構造記述は各最細分類に対して 1 つ付与されるため、語義ごとではなく分類に対して記述している。表層格の付与範囲はガ格、ヲ格、ニ格を中心に、カラ、ヨリ、マデ、ト、ヘ、トシテ、デまで必須の要素と考えられるものについて分析している。表層格の言い換えが単独である場合には図 2 のように「を・から」と併記して記述する。各表層には項番号が付与されており、半統制意味構造記述において番号を参照することで項が表層ではどの助詞に対応するかを記述することができる。

現在付与できている量は 4425 語 (7473 語義) であり、そのうち 271 語 (1762 語義) は基本的な語 (新聞記事中の高頻度語) を含んでいる。分類の種類はで 940

種類あり、暫定的ではあるが意味役割を付与 (全 87 種類) している。

4 他の言語資源との比較

先行研究において FrameNet[1][13], WordNet[3], VerbNet[5], EDR が提案され、利用されている。まず、いくつかの項目について比較した結果を表 1 に示し、各言語資源と比較する。

表 1: 各言語資源の概観を比較

	FrameNet	WordNet	VerbNet	EDR	我々
意味概念の数	884 (フレーム)	N/A (Synset)	237 (クラス)	約 43 万 (クラス)	940 (階層)
	(Ver 1.3)	(Ver 3.0)	(Ver 2.1)	(Ver 3.0)	
	2007 年	2006 年	2006 年	2003 年	2008 年
語数	約 6000	155287	3626	約 41 万	4425
語義数	N/A	117659	4991	約 27 万	7473
付与事例	約 10000	48349	N/A	約 20 万	7473
意味役割	746	N/A	31	28	87
付与対象	動, 名	動, 名	動	全て	動
の品詞	形, 副	形, 副			
対象言語	英他	英他	英	日英	日

FrameNet との比較: FrameNet はある語を想起させる Frame を仮定し、それに対してインスタンスである単語を結びつけることで辞書を構築する。Frame の決定には意味役割 (Frame Elements と呼ばれる) を仮定して意味役割の異なりを Frame の異なりとして記述する。1 つの Frame を 1 つの語義と考えており、各 Frame について意味役割解析をした例文を複数蓄積している。名詞や形容詞について同時に分析をすすめており、日本語では小原ら³が Berkeley と連携をとりつつ開発をスタートしている。

FrameNet と提案する言語資源との比較を行うと、基本的にはどちらも項構造における語義の体系化であり対象とする問題は同じである。我々の意味概念の数も FrameNet と大きく異なる程度である。

WordNet との比較: WordNet は語と語の関係を概念を介さず語のグループ (synset) を利用して記述する。表 1 の語義は synset を表し、動詞に対しては 11529 語で synset は 13767 である。これは FrameNet や我々の辞書と比較して細かい粒度でかつ大規模なデータであると言える。synset では上位・下位関係 (troponymy), 反意関係が記述されている。しかしながら我々の提案する辞書に記述しているような項構造レベルの分解した情報が無く、どの表層語が変化したのかそれとも動作主なのかといった情報が無い。つまり統語構造から意味処理を直接結びつける情報が記録されていない。

VerbNet との比較: VerbNet は Levin[7] の動詞交替

³<http://jfn.st.hc.keio.ac.jp/ja/index.html>

による動詞分類から構築された辞書であり、各動詞の分類に対して動詞の項構造、意味役割、選択制限を記述することで統語構造と語の意味への橋渡しを意識した辞書である。統語との関係を詳しく調べているが、動詞分類間の関係、および動詞間の意味的な関係については Levin の分類以上の情報はない。つまり WordNet や我々の辞書構築で述べた意味の包含関係や反意語については明示的に記録しない構造である。

以上 FrameNet, WordNet, VerbNet との記述の異なりについて述べたが、これらは互いにリンクづけを行っており、英語ではこうした情報を同時に扱うことができる。一方で日本語ではこうしたいくつものタイプの異なる言語資源がフリーで利用できる環境ではない。

EDR との比較: EDR 電子化辞書⁴ は機械処理用に構築された大規模な辞書であり、日本語だけでなく英語も構築している。構成として、単語辞書、対訳辞書、概念辞書、共起辞書、専門用語辞書、EDR コーパスからなり、全ての語義は概念識別子で ID として記録されており辞書とコーパスで統一されている。コーパスは句構造が括弧付けで記述されており、ほぼ全ての語に語義が付与されているだけでなく、意味役割に相当するラベルを句構造に分解した結果に記述している。言語処理の辞書として十分な構成であるといえるであろう。しかしながら我々が提案する動詞間の包含関係を項構造を中心に記述した情報は陽に記録されていない。理由は EDR では各動詞のエントリーに対して項構造の記述が無く、語義の分類も項構造レベルとは異なる観点で行われているためである。

5 まとめ

語彙意味論で議論されてきた語彙分解を言語処理で利用可能にするため、構造的な包含関係を保ちつつ詳細化を行う動詞項構造シソーラスを提案し、動詞 4425 語 (7473 語義) に対してシソーラスを構築した。ここでは従来 BECOME などの意味述語で議論されてきたレベルを詳細化するだけでなく、使役化との関係や状態変化の関係を分類体系 (940 種類) と半統制された述語集合によって記述された動詞の意味構造 (半統制意味構造記述) によって記述した。

本発表の動詞項構造シソーラスは Web サイトで公開する予定である⁵。

謝辞 本研究は、文科省科研費基盤研究 (B) 「語彙意味論に基づく言い換え計算機構の工学的実現と言い換

え知識獲得への応用」(17300047, 代表: 乾健太郎) の支援を受けている。NTT コミュニケーション科学基礎研究所様には電子化辞書 Lexeed の使用を許可いただいた。記して深く感謝する。

参考文献

- [1] Baker, C. F., Fillmore, C. J. and Lowe, J. B.: The Berkeley FrameNet project, *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pp. 86–90 (1998).
- [2] Dorr, B.: Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation, *Machine Translation*, Vol. 12, No. 4, pp. 271–325 (1997).
- [3] Fellbaum, C.: *WordNet an Electronic Lexical Database*, MIT Press (1998).
- [4] Jackendoff, R.: *Semantic Structures*, MIT Press (1990).
- [5] Kipper, K., Dang, H. T. and Palmer, M.: Class-Based Construction of a Verb Lexicon, *Proceedings of the Seventh National Conference of Artificial Intelligence (AAAI-2000)* (2000).
- [6] Langacker, R.: *Foundations of Cognitive Linguistics, Vol.2 :Descriptive Application*, Stanford University Press (1991).
- [7] Levin, B.: *English Verb Classes and Alternation*, University of Chicago Press (1993).
- [8] Venlder, Z.: *Linguistics in Philosophy*, Cornell University Press (1967).
- [9] 影山太郎: 動詞の意味と構文, 大修館書店 (2001).
- [10] 大石亨, 松本裕治: 格パターン分析に基づく動詞の語彙知識の獲得, 情報処理学会論文誌, Vol. 36, No. 11, pp. 2597–2610 (1995).
- [11] 竹内孔一, 乾健太郎, 藤田篤, 竹内奈央: 語彙概念構造に基づく事態上位オントロジーの構築, 言語処理学会第 13 回年次大会, pp. 859–862 (2007).
- [12] 加藤恒昭, 島山真一, 伊藤たかね: 動詞の有限性を訊ねる言語テストに関する考察, 第 12 回言語処理研究会年次大会, pp. 400–403 (2006).
- [13] 肥塚真輔, 岡本紘幸, 斎藤博昭, 小原京子: 日本語フレームネットに基づく意味役割推定, 自然言語処理, Vol. 14, No. 1, pp. 43–66 (2007).
- [14] 影山太郎: 動詞意味論, くろしお出版 (1996).

⁴http://www2.nict.go.jp/r/r312/EDR/J_index.html

⁵<http://cl.cs.okayama-u.ac.jp/rsc/data/index.html>