

語彙概念構造に基づく事態上位オントロジーの構築

竹内孔一[†]，乾健太郎[†]，藤田篤^{††}，竹内奈央^{†††}

岡山大学大学院，[†] 奈良先端科学技術大学院大学，^{††} 名古屋大学大学院，^{†††} 言語アナリスト
koichi@it.okayama-u.ac.jp，[†]inui@is.naist.jp，^{††}fujita@nuee.nagoya-u.ac.jp

1 はじめに

日本語の動詞の含意関係を語彙概念構造 (LCS) の意味特性をベースに整理して事態の上位オントロジーを構築することを目指している。事態のオントロジーとはここでは動詞間の関係を語の含意関係を取り上げてまとめあげる体系をさしている。例えば LCS では「X が Y を Z に与える/渡す」などの客体変化動詞は [x DO [x ACT ON (y)] CAUSE [BECOME [y BE AT z/C]] と記述できるが、これは「X が意志的な行為によって Y が Z にある状態にする」という意味を記述しており、変化の結果「Y が Z にある状態」という意味、さらに「Y が Z に渡る」という変化を含意することを示している。こうした動詞間の動作主性の有無や位置変化といった LCS で議論されてきた抽象的な意味関係をガイドとして事態の上位オントロジーを構築するのが本研究の目的である。

LCS の付与 [2][8][1][4] [7] で問題となる課題は語の多義性と派生の扱い、ならびに統語的振る舞いと意味との対応である。我々はこの問題に対して (1) 国語辞典の語義ごとに意味記述を行う、(2) 意味記述 (意志性や完了性など) を複数の意味特性に分解し属性と属性値の組で意味を表す、(3) 意味特性に統語的振る舞いの検査を用意しデータに記述する、という方針で付与を行う。

この方針のもとに、動作主性とアスペクト性に関する意味特性について NTT コミュニケーション科学基礎研究所で開発された電子化辞書 Lexeed のデータ (動詞 5711 件 (語義単位)/4153 件 (単語)) を対象に人手で行った。アスペクト性は完了性の分析を行うことで状態変化動詞かどうか判定できる手法として従来語彙意味論では重視されてきた。しかし、Lexeed に対する付与の結果、分類した動詞と意味特性を比較して完了性に対する意味属性と、含意関係の整理に重要な状態変化性が同一ではないということが明らかになった。つまり状態変化性と完了性はそれぞれ独立の意味特性として記述する必要があることが明らかになった。

そこで本稿ではまず実データに対する付与結果を整理し、アスペクト分析 (特に完了性) と状態変化性についての相関関係を明らかにする。これにより完了性と状態変化性を軸として動詞を分類することで動詞の振る舞いを整理できることを示す。次に含意関係を整理するために状態変化の細分類が必要であることを具体的に示し、Lexeed に対して状態変化の細分類、52 種類の動詞のカテゴリーを仮定する。この動詞カテゴリー

と既に行っている格パターンを利用することで動詞の含意関係を整理できる辞書データの形式になることを明らかにする。

2 アスペクト分析に関する再考察

付与した意味特性は動作主性に関する意志性、経験主、causer の 3 種類とアスペクト分析 (時間幅/瞬間性、完了性) に状態性、活動性、状態変化性の 3 種類である。(付与結果における統計的な値などは文献 [6] を参照) 付与を行って明らかになったのは、動作主性に関する意味特性は検査も含め問題なく分類できた一方で、アスペクト分析は状態性、活動性、状態変化性の判別がうまく行かない例が現れてきたことである。例えば

- 彼は池袋に出没する。(活動?変化?)
- 彼はこまを回す。(完了?)

「出没する」は「池袋に」と位置を指定しているものの、現在池袋にいる訳ではなく結局日常的な活動を述べている。同様な語としては「往復する」や「勤務する」などもあげられる。こうした語は特別な語ではなく日常的に良く使われる語であるにもかかわらず分析例として検討されてきていなかったのが問題である。また後者の「回す」の例では、従来の判定では完了性 (telicity) があるので状態変化と判断しているが、よく考察すると下位事象である「回る」は未完了 (atelic) の動作であり矛盾している。これは 2 段階の事象、つまり、上位事象である「彼がまわすという動作」が瞬間に完了して、下位事象である「回る」が未完のまま続いているというのが現象の意味するところである (図 1 参照)。これは「投げる」や「飛ばす」も同様である¹。

上位事象: 回す動作 (瞬間に完了)
下位事象: 回る -----> (未完)

図 1: 「回す」の上位事象と下位事象の時間関係

さらに、完了性があると必ず状態変化であると語彙意味論的には扱われているが、含意関係の計算という観点から見た場合の状態変化として欲しい情報と異なることが明らかになってきた。例をあげると

- 彼はマラソンを完走する。(完了)(変化?)
- 彼は渋谷に到着する。(完了)(変化)
- 彼はウィルスに感染する。(完了)(変化)

「完走する」は動作の終点がはっきりした完了性のある動詞である。語彙意味論の視点からは、走っている (活

¹ 言語学の分析にはこうした上位下位それぞれに対してアスペクト分析を行っている文献 [5] もある。

動)を終るので状態変化と考えられている。しかしながら動作の終了を変化として扱うことが言語処理における含意関係記述としてどれだけ有効であろうか? さらに動作の完了は状態が変化したかどうかは明らかではない。一方「到着する」や「感染する」は「彼」の位置変化、属性変化であり、何が変化しているかがはっきりしている。こうした考察から、もともと決まった動作が終了したことをのみをあらわす動詞の場合は変化と考えるのではなく、ただ動作が完了したと分析すべきである。つまり、完了性と変化性は異なり、状態変化性は独立の意味特性として分類軸とする必要がある。

3 完了性と変化に基づく動詞の分類

完了性と状態変化性という意味特性を元に動詞全体の分類を行った。その結果、2つの見方による組み合わせで4通りの分類ができる。さらに未完了 (atelic) で状態変化無しの一部の動詞が格パターンや動詞がとる名詞によって完了性・状態変化性が変わるグループがあるため²別分類として5つの動詞分類となる。この様子を図2に示す。図2では垂直方向に状態変化性:(状態

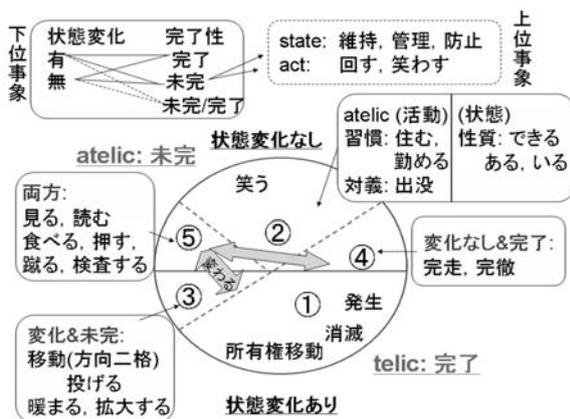


図 2: 完了性と状態変化性による分類

変化ありとなし)で分類し、少し左下がりに傾いて(点線で示してある境界)縦方向に完了性 (telicity)(完了と未完)で分類している。完了性と状態変化性を斜めに交差させたのは多くのものは完了性と状態変化性は同じものが多いためである。分割した領域は動詞の分類を示す。各領域について説明する。

1. 状態変化有・完了 「発生」「消滅」といった動詞や「移動」に関する動詞がここに分類される。いわゆる典型的な状態変化動詞である。完了性があるため「T 時間で V する」といった検査も行うことができる。

²これは領域 5 のみ完了性の付加が起こることを述べてるのでなく領域 5 以外でも起こりえる。例えば領域 3 の動詞は「太平洋側の海岸線まで拡大する」とすると完了性が付加される。

2. 状態変化無・未完 従来、活動動詞 (e.g. 「笑う」)、状態動詞 (e.g. 「できる」「ある」「いる」) と分けてきた動詞がここに分類される。さらに習慣的な活動も含まれる (e.g. 「住む」「勤める」「出沒する」)。この分類ではもう少し積極的に従来のいわゆる活動動詞を分割している。例えば人間の動作である「見る」「食べる」など基本的には同じ活動動詞である「笑う」と同じグループとして扱われてきたが後の例に示すように、動詞がとる名詞との解釈によって状態変化動詞に変わることが指摘されている。こうした文脈による変化が起こるものは後に示すように別カテゴリーとして分類することで完了性・変化性に沿って動詞を正しく分類することができる。

3. 状態変化有・未完 未完了であっても状態変化があることは既に文献 [8] などで指摘されてきた。例えば「領土が 10 年間拡大する」のように時間幅があるにもかかわらずその時間の中で少しづつ「領土」が大きくなる様子を示しており完了点が存在しない。ここには従来語彙意味論で問題になっている「彼にボールを投げる」[3] という未完了の状態変化も分類できる。これは上位事象の投げる行動は完了しているが投げたボールが目標である「彼」に到達したかどうかははっきりしないというものである。

こうした語彙意味論で取り上げられている現象をここに分類できる一方で、言語処理として考えた場合こうした繊細な未完という情報がどれだけ重要かは定かではない。例えば通常のテキスト中に「彼に投げた」といえば届いていることが推測される。またもし仮に届かなかった場合は文書中に「届かなかった」という文が来るはずである。よって処理の面から重要な情報は未完ではなくこの動詞も状態変化 (ヲ格の位置移動) というのが重要であり、我々の今回の分類基準もこれに合致する。

4. 状態変化無・完了 完了性は認められるが状態変化が無い動詞である「完走する」や「完徹する」がここに分類される。これらの他には「決められたコースを消化する」や「見届ける」といったもともと決まっていた動作を終了する意味のある動詞が分けられる。

5. 状態変化性・完了性が変わる 状態変化・完了性のどちらも格パターンや動詞がとる名詞によって変わる性質をもつ動詞である。基本的には未完で状態変化無しの動詞であるが文脈によって完了性や変化が付加される。まだ多くは発見できていないが、以下がその例である。

- 彼が哲学書を読む/10 分で彼が哲学書を読む。(変化無, 完了)
- 彼がボールを蹴る/彼が友達にボールを蹴る。(変化, 未完)

上例の「読む」は基本的には未完了の動詞であるが副詞との共起によって完了性を持つ例である。下例の「蹴

る」は二格がない場合はただ単に方向も無くボールを蹴る状態変化無しの場合であるのに対して、「友達に」という二格が付与されることでボールの位置移動が含意される。こうした特性は従来のいわゆる活動動詞である「笑う」「勤める」には無く、異なるタイプの動詞カテゴリーであることがわかる。

よって下位事象をまとめると状態変化性は有無の2つの属性値、完了性は完了、未完、2種類のあわせて4種類、さらに状態変化と完了性が交替する動詞のタイプがあり合計で5種類のカテゴリーが存在する。さらに上位事象と関連した動詞の分類を考察する。上位事象が下位事象と異なるアスペクトを持つと考えることによって従来分割が困難であったものもうまくまとめることができる。例えば使役の主語が付いた使役動詞である

- 彼がコマを回す。
- 彼が客を笑わす。

なども下位事象(笑う/回す)が未完の状態変化無の動詞でそれに使役が付いて他動詞になっていることがうまく扱える。また継続のための動作や事を起こさないための動作も上位・下位事象に分割して整理することができる。

- 火事を防止する。
- ビルを管理する。

これらは下位事象が未完の状態変化無(状態動詞)であり、そうなるように使役の主語が活動するという様子を捉えることができる。

4 含意関係を捉えるための動詞の細分類

完了性と状態変化の軸によって動詞分類の全体像を掴むことができたが、含意関係を捉えるためには状態変化性と完了性の記述だけでは不十分である。つまり動詞の取るどの格がどう変化するかという変化の種類に対する分析が必要である。具体例で示そう。

- 客が東京に到着する。
- 客が東京に着く。
- 客がホテルに感謝する。
- 客がホテルに満足する。

現在の情報ではガ格の状態変化という分類しか行わないため、これらは全て同じ状態変化動詞という分類になる。しかしながら含意関係を考慮したLCS辞書ならば「XがYに到着する/着く」は「Xが移動してYにいる状態になる」と記述されるべきであるし、「XがYに感謝する/満足する」は「Yに対してXがある心理的な状態になる」であろう。つまり(1)どの格が変化するのか、(2)変化のタイプは何かを記述する必要がある。これを具現化するには「着く/到着する」は位置変化の動詞、「感謝する/満足する」は心理的变化の動詞として細分類を行う必要がある。

そこで、NTTコミュニケーション科学基礎研究所で開発された電子化辞書Lexeedの状態変化の動詞(6455

件(語義単位)を対象に細分類化を人手で行った。手法としては典型的な動詞を探し、似ている動詞をボトムアップにくみ上げてカテゴリーを作成した。その結果、現段階では状態変化に対して52種類のカテゴリーを仮定できた。以下分類の全体像について記述した後、いくつか典型的な動詞カテゴリーについて説明し、格情報との組み合わせにより含意関係が抽出できることを明らかにする。

4.1 分類の全体像

現在分類している状態変化動詞のカテゴリーを明らかにする。カテゴリーはおおむね次の5つの系列に分類できる: 位置変化系、属性変化系、判断系、物理的变化系、解決系、使役系、その他。よってこれらの系列に整理して表1に示す。

表1: 状態変化に対する動詞の細分類

系列	動詞分類
位置変化系	位置変化(物理),(所有),(情報) 位置関係の変化(物理),交換 関係の変化(人),(対象),勝敗 取得(所有),(情報) 消費・喪失(所有),(情報) 生成(物理),(情報),(物理)
属性変化系	身体的変化,心理的变化,意識の変化 習得,属性変化,入賞・落選 状況変化
判断系	判断(決心),(認識),(認定),(許可) 分類,証明
物理的变化系	物理的变化,程度の変化,制限 質の変化,利用の権利の変化
解決系	対処・解決,補足,改良・修正
使役系	ACTの引き起こし
その他	開始,停止・終了,変更,回復 進歩,完成・達成,突破,時間的变化 状態変化,準備,影響・被害,時の推移 最終的結果,的中,状態不変化 その他の変化

表1中の()は一つのカテゴリーを省略して記述している。例えば位置変化(物理),(所有),(情報)は位置変化(物理),位置変化(所有),位置変化(情報)の3つのカテゴリーを表している。これらのカテゴリー,ならびに系列によるまとめは第一段階の提案でさらなる洗練が必要であるが,今まで状態変化として扱ってきたものがどれだけその種類に幅があるのかを掴むことに役立っている。

さらに格パターンが整理されているため格パターンの選択によって同じ動詞の分類内で格パターンが異なる含意関係のある動詞を取り出すことができる。また同じ格パターンで取り出せば統語的振る舞いと意味が似た動詞の集合を取り出すことができる。次節ではこれらの具体例を示す。

4.2 動詞分類カテゴリーの例

「位置関係の変化(物理)」(218件)を例に挙げて説明する。まず「位置変化(物理)」(532件)と名称が似て

いるが「位置変化」はある対象が移動してある着点に至るといふ移動 (e.g. 移動する, 急ぐ) であるのに対して「位置関係の変化」は幅広い位置の変化を含む。例えばある対象が対象の中に入ったり (注射する, 投入する, 給水する, 給油する, 充電する, インプットする, 注入する), あるところから出たり (排水する, 引き抜く, 取り出す), あるところに合体したり (重なる, 積み重ねる, 装着する) する動詞である。これらは1つのグループとしてまとめられているが, 本LCS辞書データはテーブル上の属性と属性値の組み合わせですべて整理されており, 格パターンが同じものについて簡易に取り出すことができる。例えば「Xガ (動作主)Yヲ (対象)Zニ (着点/方向/場所) Vする」という格パターン³で検索すると

- 注射する, 注入する, インプットする, 収納する, 投入する, 給油する

といった事例が取り出される (90件)。これらは「XがYをZにあるという変化を意志的に行う」という事例群⁴であり, 含意関係の同様なものを取り出すことができる。少し変えて「Xガ (対象)Zニ (着点)」で検索すると

- 伝染する, 埋没する, 混入する

など「YがZにあるという状態に至る変化」という上例の下位事象の関係に当たる事例集合を集めることができる。もちろんこうしたカテゴリー内にはまだ検討が必要な動詞が存在し個別に整理していく必要があるが, 格パターンと変化の整理によって含意関係が似ている動詞を整理できる枠組みが提案できたことが大きな利点である。

5 他の言語資源との比較

現在, 言語処理で利用できるシソーラスとして国立国語研究所の分類語彙表 [10] とEDR電子化辞書 [9] が挙げられるが, こうした動詞間の含意関係に注意を払って構築されていないため部分的には参考にはできないが直接参考にすることができない。例えば我々の分類では「卒業する」はガ格の属性変化として分類し「辞職する」も同じグループである。一方, 分類語彙表では「卒業する」は「活動-生活-学事・兵事」というクラスに分けられているが「辞職する」は「活動-生活-処世・出处進退」という分類になり「幅が利く」「一花咲かせる」などまた少し異なる分類にカテゴリーに属している。また大きな違いとして分類語彙表は語義や例文が無いので, どの語義について記述しているかが言語処理で直接参照できない。

また参考にはできるシソーラスとしてNTTコミュニケーション科学基礎研究所の日本語語彙大系 [11] があるが名詞に対する分類の粒度が細かく整備されている一方で用言に対しては含意関係を整理できるほど細かくないように見受けられる。我々のデータは, 格パター

ンと例文まで含めて公開を予定しているなのでその部分においても異なっている。

6 まとめ

LCSの意味特性を利用して事態上位オントロジーの構築を行い具体的な分類を提示した。従来, 語彙意味論で重要視されてきた完了性の分析と, 言語処理として必要な状態変化性に差異がある事を明らかにし, 全動詞の分類を完了性と状態変化性でまとめることを提案した。これにより, 従来分類がうまく行かなかった動詞に対して整理することができた。さらに状態変化動詞の細分類を行うことで含意関係を捉えた動詞辞書が構築できることを示した。

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殿には電子化辞書 Lexeed の使用を許可いただいた。東京大学の伊藤たかね氏, 加藤恒昭氏, 畠山真一氏, 甲南大学の中谷健太郎氏には有益なコメントをいただいた。奈良先端科学技術大学院大学の阿部修也氏には種々のデータ作成に多大な協力をいただいた。記して深く感謝する。

参考文献

- [1] Dorr, B.: Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation, *Machine Translation*, Vol. 12, No. 4, pp. 271–325 (1997).
- [2] Jackendoff, R.: *Semantic Structures*, MIT Press (1990).
- [3] Levin, B. and Hovav, M. R.: *Argument Realization*, Cambridge (2005).
- [4] 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).
- [5] 金子亨: 言語の時間表現, ひつじ書房 (1995).
- [6] 乾健太郎, 藤田篤, 竹内孔一: 含意関係計算のための事態オントロジーの開発に向けて, 信学技法, *NLC2006-89*, pp. 13–18 (2007).
- [7] 加藤恒昭, 畠山真一, 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).
- [8] 影山太郎: 動詞意味論, くろしお出版 (1996).
- [9] 日本電子化辞書研究所: EDR 電子化辞書使用説明書 (第2版) (1995).
- [10] 国立国語研究所編: 分類語彙表, 大日本図書 (2004).
- [11] 池原 悟他編: 日本語語彙大系, 岩波出版 (1997).

³LCS辞書には格パターンに「対象」「着点」など深層格も試験的にはあるが付与している

⁴1件の事例には動詞とその語義, さらに代表例文が付いている。