

語釈文を用いた複合動詞の特徴分類

加藤 恒昭¹ 林 良彦² 伊藤 たかね¹

東京大学 大学院総合文化研究科¹ 大阪大学 大学院言語文化研究科²

1 はじめに

複数の語基が結び付いて、動詞として振る舞うひとつの語をなしているものを複合動詞と呼ぶ[6]。日本語動詞において複合動詞が占める割合は大きく、接辞と語基の結びつきである派生動詞を含めると国語辞典の見出し語の半数を超えると報告されている。そして、このうちの7割以上が「動詞(連用形)+動詞」の形をした複合動詞(動詞的な造語成分を持つものを含む)だという[5]。

このタイプの複合動詞は辞書の見出し語として数が多いだけでなく、生産的に形成され、英語において結果構文や句動詞で表現されるような動詞概念がこれらによって表されることが知られている[4]。

我々は、この「動詞(連用形)+動詞」の複合動詞の意味的統語的特徴を、その構成要素である動詞のそれと関係づけることを目指している¹。語彙的オントロジの中に複合動詞を位置づけ、その構成要素である動詞と特定の意味関係で関連づけることや、複合動詞の語彙概念構造をその構成要素のそれと関係づけて表現することが目的である。

言語理解において、語の意味はそのままでは他と関係づけられていない原子であり、その関係づけるためには、個別の語彙的知識が必要となる。複合動詞もそれを語として見ている限りでは、それらに関する知識を個別に記述しなければならないが、その生産性から、少なくとも一部においては、その構成要素と構成的な関係が成立すると推測される。複合動詞の意味的統語的特徴を、その構成要素のそれと系統的に関係づけられれば、例えば、ある複合動詞が表わす状況がその構成要素である動詞が表わす状況を含意する等の推論を可能にするような知識を効率的に構築できると期待できる。

本稿では、そのための第一歩として行った複合動詞抽出の手法についての検討と、計算機可読辞書の語釈文を利用して行った複合動詞とその構成要素の意味的関係の初期調査について、報告する。

2 位置づけ

複合動詞をその構成要素との意味関係に着目して分類する試みは既に多くの研究者によって行われている。寺村は、構成要素である動詞²の意味のどちらが複合動詞の中

に保持されているかに着目して、複合動詞を4種類に分類している[8]。森田は、v1とv2の意味が並列関係になっているもの、v1とv2が主述・補足関係になっているもの、v2が抽象的な意味に変化したもの等々、5つに分類している。由本は、複合動詞が意味の透明性と生産性が高い統語的なものと慣習化・語彙化が進んでいる語彙的なものに分けられるという分析[3]をふまえ、統語的複合動詞の意味関係を補文関係とし、語彙的複合動詞の意味関係を、並列関係、付帯状況、手段、原因、そして補文関係に分類している[9]。

これらの中でも由本の分析は語彙概念構造の枠組みを用いた形式性の高いもので、それぞれの意味関係を持つ複合動詞を構成できるための構成要素間の制約も示されている。一方で、森田の分類に比べると構成要素の意味への関心が薄いように思われ、例えば「書き落とす」は「書くことに失敗する」と分析されるが、ここでは「落とす」の意味が「失敗する」であることが前提とされる。森田の分析であればこの「落とす」は抽象的な意味に変化していて、語としての「落とす」の意味全体とは区別されなければならない。このような異なる研究関心からの知見を統合していくことが必要である。またいずれの研究も言語学的な観点に立つもので、網羅的な言語資源を構築するための指針や、処理アルゴリズムの設計に用いるためには、それらをふまえつつ異なる観点からの検討を追加する必要もある。

本稿では、計算機可読辞書の語釈文を利用して、複合動詞とその構成要素の意味的関係の分析を試みる。計算機可読辞書を用いること理由のひとつに、語と語義を区別して扱うことが可能となり、語と語の関係だけでなく、語義と語義の関係として問題をとらえられることがある。例えば「書き残す」には「書いて後に残す」「書くべきことを書かないでおく。書かずに残す」という2つの語義がある³。この2つの語義では「書く」ことへの含意が正反対で、語ではなく、それぞれの語義に関して議論する必要がある。また「洗い出す」には「洗い始める」「隠れていた事実を調べ上げる」の語義があり、前者は「洗う」の語義全体と関連するが、後者はそのひとつである「隠れているものを調べ出す」とのみ関連を持つ。「洗い出す」と「洗う」の関係として捉えていてはこのことは見えてこない。これらの問題に向かい合うために、語に対して可能な複数の語義が関係づけられている辞書の語釈文を活用する。

¹以下、誤解のない範囲でこの「動詞(連用形)+動詞」のものを単に複合動詞と呼ぶ。

²以下、前方の動詞をv1、後方の動詞をv2と表わす。

³語釈文は「基本語データベース」[7]から引用し、必要に応じて見出し語とで繋ぐ。語釈文の一部のみを取り出している場合もある。

表 1: 毎日新聞記事からの複合動詞の抽出

	抽出数	適切な抽出	誤抽出	解析失敗	抽出例
分類 (A)	787	787	-	-	取り組む, 繰り返す, 盛り込む, 振り返る
分類 (B)	198	166(42)	9	23	勝ち上がる, 引きこもる, 食べ残す, 使い回す
分類 (C)	15	2	11	2	貸し渋る, 生まれ育つ

*分類 (B) の括弧内は v2 が動詞-自立であるものの内数, 解析失敗は文字化けを含む形態素解析の誤りによるものを示す.

3 複合動詞の抽出

複合動詞の分析のためには, まずそれを文章中から抽出する必要がある. 形態素解析⁴を行った場合, 複合動詞の解析結果は以下の3種類のいずれかとなる.

- (A) 解析用辞書に語として登録されており, 1語として分析される.
- (B) 動詞連用形 + 動詞の2語(もしくはそれ以上の連鎖)として分析される. 後方の動詞は「回る」「始める」「続ける」のように動詞-非自立の場合と, 「勝ち上がる」の「上がる」や「食べ残す」の「残す」等, 動詞-自立の場合がある. 前者は統語的複合動詞に対応すると考えられる.
- (C) 名詞 + 動詞の2語として分析される. 前方の名詞は「踊り」「貸し」等, 動詞の連用形と同じ表記を持つ名詞(名詞-一般)で, 「踊り狂う」「貸し渋る」はこの分類となる.

分類 (A) の場合, 解析用辞書に複合動詞であるかの情報は含まれないため, 解析用辞書の項目に人手で, 複合動詞であるかの判断と, 複合動詞である場合はその構成要素を註釈づけ, それを参照して複合動詞であるかを判断する. ちなみに, 送り仮名の違い等の漢字平仮名の異表記をまとめあげた解析用辞書項目(動詞-自立)7361語について作業した結果, 4142語(56%)が「動詞(連用形) + 動詞」の形のものを含む複合動詞全般であり, そのうちの2996語(72%)が「動詞(連用形) + 動詞」の形をしていた⁵. この結果は [5] での報告と一致する.

分類 (C) については, 名詞のうち, 動詞の連用形と同じ表記を持つものを機械的に抽出しリストを作成した. 具体的には解析用辞書に含まれる名詞(名詞-一般のみ, 名詞-サ変接続は含まない)について「ます」を後続させて, 形態素解析を行い, 当該部分が動詞連用形と解析されるものを選び出した. またこの時, 解析結果となった動詞の基本形等の情報を註釈づけた. 1472語がえられている.

これらを用いて複合動詞を抽出する. 抽出の傾向をえるために, 毎日新聞記事 2008年1年分から抽出を行った.

⁴本稿では形態素解析システムとして茶筌 (version 2.2.4) をその解析用辞書に IPADIC (version 2.70) [2] を用いている. 品詞分類等もそれに従う.

⁵現在より厳密な判断基準で見直しを行っている. それを受けて数字の細かい部分が変わってくる可能性がある. 本稿全般にわたって, 報告している数値が初期分析の結果であることをご容赦願いたい.

また「基本語データベース」[7]の見出し語を解析し, 複合動詞を抽出した.

毎日新聞から抽出したもののうち, 頻度の多い上位1000語の内訳を表1に示す. 8割近くが辞書に登録されており, 分類 (B) として抽出されるものも3/4はv2が動詞-非自立である. なお, この中には「見てる」「生きてる」等複合動詞に分類すべきではないもの, 「し続ける」「し始める」のようにv1が「する」で直前に名詞-サ変接続をとっているであろうものも含んでいる. ここからは, 複合動詞の大多数が解析用辞書に登録されているか, 特定の動詞-非自立をv2とする統語的複合動詞相当のものであることがわかる. しかしそれでも44語の「新しい」複合動詞が高い頻度で用いられており, その割合は頻度が低くなるにつれて増加することが推測される.

誤抽出には以下のような場合がある. v2を「する」として, 「(キャンプ) 入りする」や「(お) 届けする」を複合動詞としている. 「割」が「割る」の連用形であるので「(1) 割増える」等も複合動詞としている. また, 頻度は高くなく上位1000語には含まれないが, 記事見出しに現れるような「組み合わせ決まる」を分類 (C) として誤抽出することや, 複合動詞が複合動詞の構成要素となることを無条件に許しているために「繰り返し求める」等を全体で分類 (B) と誤抽出するケースも見られる.

「基本語データベース」の動詞見出し語3057語からは, 1317語の複合動詞を抽出した. 語義数では, 7529語義から2286語義となる. そのうち, 分類 (B) は「言い遅れる」「生き長らえる」「抱きかかえる」等, 149語, 分類 (C) は「遊び暮らす」「押し殺す」「踊り狂う」等, 13語であった. 解析用辞書に含まれる複合動詞の数は本辞書の倍以上であるが, それでも多くが語として登録されていないことがわかる.

以上から, 複合動詞の一覧を辞書等に求めることは難しく, 分析のためにはそれを文章中から抽出する仕組みが必要であることがわかる. 現在の方式は精度の点でまだ不十分であるので, 関連研究 [1] 等を参考にして, 再現率を下げることなく精度を上げることを検討する必要がある.

4 語釈文を用いた複合動詞の分析

複合動詞とその構成要素の意味的關係を明らかにするために, 計算機可読辞書の語釈文の分析を行った. 対象は「基本語データベース」から抽出した1317語・2286語義

言いそびれる：言う機会を失って、 <u>言わず</u> に終える。	(v1 否定形式)
売れ残る：商品が <u>売れない</u> で残る。	(v1 否定形式, v2 終止形式)
あきれ返る：完全に <u>あきれ</u> る。	(v1 終止形式)
言い表わす：思っていることを言葉で <u>表す</u> 。言葉で表現する。	(v2 終止形式)
植え込む：草や木を土の中にしっかりと <u>植えて</u> 入れる。	(v1 連用形式)
打ち上げる：打って高く <u>上げ</u> る。	(v1 連用形式, v2 終止形式)
撃ち殺す：鉄砲などで <u>撃って</u> 殺す。	(v1 連用形式, v2 終止形式)
書き添える：文章や絵の側に <u>添えて</u> 書く。	(v1 終止形式, v2 連用形式)
送り出す： <u>送って</u> 外へ <u>出す</u> 。出かける人を <u>送る</u> 。	(v1 連用形式, v1 終止形式, v2 終止形式)

図 1: v1 や v2 を含む語釈文の例

の語釈文である「書き残す：書いて後に残す。後に伝えるために書いて残す。」等、語釈文は語の意味を簡潔な文もしくは文章で説明したものである。これについて、「書き残す」の構成要素である「書く」「残す」が語釈文に含まれていることから、「x が y を書き残した」ことが「x が y を書いた」とことと「x が y を残した」ことを含意するという情報が獲得できるのではないかというのが本調査の着眼である。

なお、語釈文には、このような説明に加え、「開け放つ：開け放す」のように動詞のみからなり、言い換えを示すもの、「打ち返す：「返す」を強めた言い方」のようなメタな記述からなるものがある。スペースの関係でこの種類の語釈文の分析については割愛する。

前節で述べた抽出方式により、見出し語や語釈文中に現れる語が複合動詞であるかの判断が行え、複合動詞の場合はその構成要素を明らかにできる。また、語によって1語2語等、異なる形に分析されるというような揺れを吸収し、複合動詞に共通した形態素解析結果を与えられるため、分析のためのパタン記述も容易になっている。

語釈文の形態素解析結果に以下のようなパタンを適用することで、見出し語の構成要素である v1 と v2 がどのような形式で語釈文に現れているかを調査した。終止形式での出現はその語（の語義）がその複合語の語義の主たる部分、意味の主辞であることを示唆し、連用形式での出現はその語が付帯状況や様態を表わし、意味の主辞を修飾する修飾要素の役割を担っていることを示唆すると考えている。否定形式での出現は、その語が表す行為を行わないことがいずれかの役割を果たしていると期待される。

終止形式	「v[基本形]」, 「v[連用形]たり v[連用形]たりする」
連用形式	「v[連用形]」, 「v[連用形]て」
否定形式	「v[未然形]ず」, 「v[未然形]ないで」, 「v[未然形]ない」

ここで、v は単独で語をなしているもので、複合動詞の構成要素 (v2) として現れたものは含まない。

表 2: 語釈文中での v1 と v2 の出現

	否定	連用	終止	連用 U 終止
v1	10	359	327	662
v2	0	51	280	326
v1∩v2	-	-	-	95

表 3: v1 と v2 を含む語釈文での v1 と v2 の出現形式

	v2 連用	v2 終止	v2 連用 U 終止
v1 連用	4	74	78
v1 終止	12	7	18
v1 連用 U 終止	16	80	95

結果を表 2 に示す。あわせて幾つかの例を図 1 に示す。2286 語義中の 793 語義、約 1/3 の語釈文に v1 もしくは v2 が含まれていることがわかる。v1 は意味の主辞と修飾要素のいずれの役割も果たすことがあること、v2 の出現は v1 に比べて少なく、その役割は意味的主辞が大半であることがわかる。また、否定形式で現れたのは v1 のみである。v1 と v2 が共に現れたものの内訳を表 3 に示すが、v1 が修飾要素となり v2 が意味の主辞となっていると思われるものが大半を占めている。

語義数で数えて、多くの複合動詞を形成する v1 と v2 の上位 6 位（形成される語義数が 50 以上のものにあたる）について、それを含む複合動詞の語義毎の語釈文が v1 や v2 を含む割合を表 4, 5 に示す。語義全体の平均では、v1 で 29%、v2 で 14% であり、v1 や v2 に依存してその割合が大きく異なるのがわかる。複合動詞と構成要素との意味関係を推定するベースラインとして用いることもできそうである。

5 考察

2/3の語釈文にv1もv2も含まれなかったことについて考えてみる。第一の理由は言うまでもなく表現の多様性である。例えば、上例に挙げた「売れ残る」と同じような構成を持つと思われる「焼け残る」の語釈文は「同じ火事で他が焼けたのに、それだけが無事に残る。」でありv1の否定形式は含まれない。語もv1やv2そのものではなく類義語となっていることが多い。「寝静まる：人々が皆眠って辺りが静かになる。」「寝入る：深くよく眠る。」や「言い寄る：話し掛けながら近づく。」「走り寄る：走って近づく。」等が見られる。また、語釈文の中に複合動詞が含まれる場合もある。「編み出す：新しい方法や策略などを考え出す。」「言い伝える：後世に語り伝える。」等多くの例がある。これらの複合動詞は見出し語とv1やv2を同じくするものも多いが、現時点ではその意味づけが明らかではなく、今後分析を進めていきたい。

第二には、「洗い出す：隠れていた事実を調べ上げる。」「上がり込む：勧められてもいないのに、人の家に入る。」のように、v1やv2の特定の語義だけを引き継いで、意味的な構成要素としていると考えられる場合である。この例では、それぞれ「洗う：隠れているものを調べ出す。すっかり出す。」「上がる：座敷や部屋の中へ入る。」が引き継がれているものと思われる。それらを抽出する方法として、v1やv2そのものではなく、その語釈文に現れる動詞等とマッチングを行うことが考えられるが、この例からもわかるように表現の多様性の問題から直接的な方法でよい結果を得ることは難しそうである。

第三が、慣習化・語彙化が進んで、もしくは構成要素の意味が抽象化あるいは造語成分となることで、複合動詞の語釈文にその構成要素であるv1やv2が現れない場合である。「居直る：急に態度を変える。」「入れ込む：夢中になる。ひどくのぼせる。」等、この例はもちろん多い。

一方で、語釈文にv1やv2が含まれ、含意関係が示唆されるものにも注意が必要である。「歌い上げる：声を上げて最後まで歌う。」「泣き腫らす：激しく長く泣いてまぶたを腫らす。」はv1とv2を語釈文に含むが、「アリアを歌い上げた」時に「アリアは上がら」ないし、「泣き腫らした顔」をしていても「顔が泣いた」わけではない。このような項関係までを語釈文から抽出するためには更に深い解析が必要なことは言うまでもない。

6 おわりに

複合動詞とその構成要素との意味的关系を計算機可読辞書の語釈文から取得することを試み、その初期調査の結果を報告した。語釈文は簡潔ではあるが多様な表現を含んでおり、単純なマッチングでは充分でないが、既存の言語知識を活用したある程度の解析により、多くの知識を取得する可能性はあるという感触をえた。今後これらについて検討を進めていく。また、複合動詞を高い精度で抽出する

表 4: 生産性の高いv1を持つ複合動詞の語義

	語義数	v1	v2	v1∩v2
引く	126	10 (8%)	23 (18%)	3
見る	105	48 (46%)	5 (5%)	4
取る	102	17 (17%)	20 (20%)	1
言う	57	24 (42%)	7 (12%)	1
突く	51	15 (29%)	13 (25%)	6
打つ	51	8 (16%)	9 (18%)	1

表 5: 生産性の高いv2を持つ複合動詞の語義

	語義数	v1	v2	v1∩v2
込む	224	76 (34%)	0 (0%)	0
出す	157	38 (24%)	37 (24%)	17
上げる	102	36 (35%)	11 (11%)	4
合う	79	22 (28%)	2 (3%)	0
付ける	88	31 (35%)	6 (7%)	3
上がる	55	12 (22%)	13 (24%)	2

ことにも問題は残っており、この問題も引き続き検討していく。

謝辞

本研究は、挑戦的萌芽研究「大規模語彙知識を融合した語彙概念構造体系の構築」および財団法人電気通信普及財団より助成いただいている「日本語意味理解のための動詞語彙概念構造の研究」の一部として進められている。ご理解とご支援をここに深く感謝する。

参考文献

- [1] J. Breen and T. Baldwin. Corpus-based Extraction of Japanese Compound Verbs, In Procs of the 2009 Australasian Language Technology Workshop (ALTW 2009), pp. 35-43, 2009.
- [2] <http://chasen-legacy.sourceforge.jp/>.
- [3] 影山太郎. 文法と語形成, ひつじ書房, 1993.
- [4] 影山太郎. 形態論と意味, くろしお出版, 1999.
- [5] 森田良行. 日本語学と日本語教育, 凡人社, 1990.
- [6] 森田良行. 動詞・形容詞・副詞の事典, 東京堂出版, 2008.
- [7] NTTコミュニケーション科学基礎研究所監修. 基本語データベース, 学習研究社, 2008.
- [8] 寺村秀夫. 日本語のシンタクスと意味 II, くろしお出版, 1984.
- [9] 由本陽子. 複合動詞・派生動詞の意味と統語, ひつじ書房, 2005.