

日本語意味理解のための動詞語彙概念構造の研究

代表研究者 加藤 恒 昭 東京大学大学院総合文化研究科 教授

1 はじめに

テキストを単なる文字の並びとして扱うのではなく、それが伝える意味を処理できると、様々な通信サービスが可能となる。例えば、電子カルテに記載された情報を意味に基づいて検索し類似した症例を検討したり、ウェブ等に溢れている多量の情報から必要なものを（現在の文字列の一致ではなく）意味的な判断で選択し提供したりすることが可能となる。本研究「日本語意味理解のための動詞語彙概念構造の研究」は、このようなテキストの意味に基づく高度な通信サービスを可能とすることを目指したものである。

テキストの意味理解のためには、文の意味的な構造を明らかにすることが重要である。この構造は項構造と呼ばれ、動詞とその項（格要素）との関係が表現され、これにより動詞が意味する出来事を行った主体が何でその対象が何か等が明らかとなる。近年、このような項構造解析の技術は進歩が目覚ましいが、大きな問題となるのは、動詞が意味する出来事そのものの意味記述が不十分なことである。そのため、例えば、「流す」と「流れる」、「流れる」と「流れ始める」や「流れ込む」との意味関係が理解されず、全く異なる出来事であると判断されることになる。本研究では、このような動詞の間の意味関係を明らかにすることを通じて、動詞の意味、動詞が意味する出来事を明らかにすることを試みた。ここでのアプローチは、動詞の意味をそれが関係する動詞との関係で記述しようというものである。現在の解析では「流す」「流れる」「流れ始める」「流れ込む」はお互い無関係なものとして扱われる。また、「流す」や「流れる」がどんな意味を持つかを形式性の高い形で定義することも容易ではない。そこで、「流す」の意味を「流れる」の意味との関係で記述する、「流れ始める」の意味を「流れる」と「始める」の意味との関係で記述することを考える。そのために、まず様々な動詞の関係を整理していくことから研究を進めた。

現時点で得られた成果は以下の通りである。

- (1) 日本語テキストに対して一般的な形態素解析（単語分割）を行った結果として得られる動詞群に対し、それらの意味的形態的關係を記述したデータベース（動詞関係データベース）を作成した。対象とした関係は、自動詞と他動詞の対応（「流れる」「流す」）、動詞とその可能形との対応（「流す」「流せる」）、動詞とその受身形との関係（「流す」「流される」）、動詞とその使役形（「流す」「流させる」）との対応である。
- (2) 上記の動詞群のうち、複合動詞（「流れ始める」や「流れ込む」のように複数の動詞もしくは動詞的造語成分から構成される動詞）を明らかにし、それとその構成要素との関係をデータベース化（複合動詞データベース）した。
- (3) 日本語テキストから複合動詞を抽出する方式について検討し、基本的な手法を評価した。複合動詞はそのすべてが辞書に記述されているものではなく、生産的に創り出される。そのため、どのような複合動詞が存在し、それらがテキスト中でどう振る舞っているかを知るためには、それをテキスト中で認定する仕組みが必要である。このような仕組みとして、複合動詞の抽出を実装した。
- (4) 複合動詞とその構成要素の意味的な関係を明らかにするために、計算機可読辞書の語釈文からそれを抽出する実験を行い、一定の質と量の意味関係が得られることを確認した。更なる検討の余地は大きいものの、ある程度の知識を抽出できるという感触を得た。

これらは、いずれも検討のための環境整備もしくは初期的な検討にとどまるものであるが、動詞の意味を中心としたテキストの意味理解を進めるにおいて、貴重な一歩であると考えている。また、動詞関係データベースと複合動詞データベースは公開準備中である。以下、これらの成果について、順に説明する。

2 動詞関係データベースの構築

日本語テキストの一般的な形態素解析に用いられる形態素解析用辞書 IPADIC (version 2.70) に含まれる自立動詞について、それらの間の意味的・形態的關係づけを行った。本作業は次の2つの目的を持っている。

- (1) 動詞の間の意味的な関係を理解するための足掛かりとなる情報を作成する。
- (2) 工学的な理由で語として辞書登録されているものに、その後の意味処理の観点から望ましい解析結果を与えるための情報を作成する。

この作業で付与した関係は、次の6種類である。

- ・ 自他動詞（他動詞を自動詞に関係づける）
- ・ 受動能動関係1（形態的に判断できる受動の動詞を、能動の動詞に関係づける）
- ・ 受動能動関係2（形態では判断できない受動の動詞を、能動の動詞に関係づける）
- ・ 可能関係（形態的に判断できる可能動詞を、もとの動詞に関係づける）
- ・ 使役関係1（～ス/サスの形で形態的に判断できる使役動詞を、もとの動詞に関係づける）
- ・ 使役関係2（～セル/サセルの形で形態的に判断できる使役動詞を、もとの動詞に関係づける）

このうち、自他動詞の対応づけ、受動能動関係2は、(1)の目的で行い、意味的な判断に基づいて行った。一方、可能関係、使役関係(1,2)、受動能動関係1は、(2)の目的で行い、形態的な情報で判断を行った。ただし、意味的な関係が実際にあるかどうかの判断を補助的に付与している。

作業では、辞書項目に対し、同じ読み、同じ品詞情報（活用分類等）を持ちながら、漢字表記と平仮名表記の間の表記揺れによって異なる項目されているものをまとめて語として認定しIDを与え、それらそれぞれに代表表記を設定した。その後の動詞関係の認定はこの単位で行っている。

関係認定の基準は以下の通りである。

自他動詞 「AがBをV1する」と「BがV2する（Aは二格等に現れない）」という意味関係にあるV1とV2を対応づける。例えば、「エアコン[A]が部屋[B]を温める[V1]」と「部屋[B]が温まる[V2]」のような、温める(V1)⇒温まる(V2)を関係づける。なお、自他動詞関係の本質は、「AがBをV1する」と「BがV2する」という意味関係にある他動詞V1と自動詞V2であるが、この基本的な定義は可能、受動能動1,2、使役1,2の関係にある動詞の対が持つ意味関係と重複するため、それを判断するための詳細な意味判断基準を設けて関係記述を行っている。

受動能動関係1 動詞V2の未然形+レル/ラレルの形をしていながら、辞書中に登録されている動詞V1を、V2と関係づける。大抵は「AがBをV2する」と「BがAにV1」の関係にある。ただし、「～に憑く」⇒「憑かれる」の例もある。例えば「地上[B]が雪[A]におおわれる[V1]」と「雪[A]が地上[B]をおおう[V2]」のような、おおわれる(V1)⇒おおう(v2)を関係づける。

受動能動関係2 「AがBにCをV2する」と「BがCをV1する」の意味関係あるいは、「AがBをV2する」と「BがAにV1する」の意味関係にあるが、動詞V2の未然形+レル/ラレルの形でない動詞V1を、V2と関係づける。例えば、「生徒[B]が英語[C]を教わる[V1]」と「先生[A]が生徒[B]に英語[C]を教える[V2]」や、「ニュース[B]が人々[A]に知れる[V1]」と「人々[A]がニュース[B]を知る[V2]」のような、教わる(V1)⇒教える(V2)、知れる(V1)⇒知る(V2)を関係づける。

可能関係 「V2（五段活用）仮定形・命令形+ル」がV1になるようなV2、「V2（五段活用以外）未然形+レル」がV1になるようなV2（いわゆる「ら抜き」）、「V2（五段活用以外）未然形+ラレル」がV1になるようなV2でありながら、辞書中に登録されているV1を、V2と関係づける。例えば、上がれる(V1)⇒上がる(V2)、見れる(V1)⇒見る(V2)、見られる(V1)⇒見る(V2)等を関連づける。

使役関係1 動詞V2の未然形+ス/サスの形をしていながら、辞書中に登録されているV1を、V2と関係づける。例えば、読ます(V1)⇒読む(V2)が関係づけられる。

使役関係2 動詞V2の未然形+セル/サセルの形をしていながら、辞書中に登録されているV1を、V2と関係づける。例えば、読ませる(V1)⇒読む(V2)が関係づけられる。

これらの作業結果を関係データベース(SQLiteを利用)、及びpythonのオブジェクトシステムとしてまとめ、利用可能とした。

3 複合動詞データベースの構築

複数の語基（語や造語成分となるもの）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[森田08]。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[森田

90]. ここでは、複合動詞と判断される動詞の抽出と、そのそれらと構成要素となる語（や造語成分）との関係づけを行った。

作業は、前述の動詞関係データベースと同様に、表記の揺れを吸収して ID と代表表記を与えた語について行った。以下の分類と関係づけを行った後、関係データベース、及び python のオブジェクトシステムとしてまとめ、利用可能とした。

- ・ 基本 そもそも動詞（ひとつの形態素であり他の語と関係づけられないと思われるもの）
- ・ 動詞＋ 「立ち上がる（立つ，上がる）」「売れ残る（売れる，残る）」等，動詞の連用形と動詞の繋がりによる複合動詞。
- ・ 名詞 「愛す（愛）」のような名詞と関連しているもの。
- ・ 名詞＋ 「色あせる（色，あせる）」等，名詞と動詞の繋がりによる複合動詞。
- ・ 形容 「暖める（暖かい）」のような形容詞と関連していると思われるもの。
- ・ 形容＋ 「近づく（近い，つく）」等，形容詞と動詞の繋がりによる複合動詞。
- ・ 形動 「静まる（静か）」「馬鹿げる（馬鹿）」のような形容詞と関連していると思われるもの。形容動詞は、「～な」で名詞を修飾できることで，形容詞や名詞と区別する。
- ・ 形動＋ 形容動詞と動詞の繋がりによる複合動詞（存在しない可能性あり）
- ・ 接辞＋ 「相継ぐ（相，継ぐ）」等，接辞（単独で語にならないもの程度の意味）と動詞の繋がりによる複合動詞。「べとつく（べと，つく）」等，擬態語＋「つく」はここに含める。
- ・ サ動＋ 「理解し合う」「和解し合う」とその可能形（サ変動詞の連用形と動詞が繋がりによる複合動詞）。一般には語と認定されるものではないので，辞書設計の特殊な事情で含まれているものと思われる。
- ・ 句 「悦に入る」のような助詞を含む慣用句，「打って出る」のようなテ形（「打って」）を含んだ複合動詞等，語と考えづらいもの
- ・ 古語 現代仮名遣いでないもの
- ・ 不明 作業者にとって分類が不明で，広辞苑に載っていないもの

ちなみに、この作業の対象となった語は 7361 語であったが、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。この結果は本節先頭で述べた[森田 90]での報告と一致している。

4 複合動詞の抽出

複合動詞の 7 割以上を占める「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）であるが、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[影山 99]。前述したように、本研究の目的は、動詞の意味をそれが関係する動詞との関係で記述しようということである。そのため、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く。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。なお以下では、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び、前方の動詞を v1、後方の動詞を v2 と表わす。

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。一般的な形態素解析（形態素解析システム茶釜(version 2.2.4) [chasen]を解析用辞書 IPADIC (version 2.70) で用いたもの、以下の記述での品詞分類等もそれによって得られるものを用いる)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。

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。また「基本語データベース」[NTTCS08]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。表において、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している。表から、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。

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

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、今後、関連研究[Breen09]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していく。

5 語釈文からの複合動詞意味関係の抽出

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、その関係づけのためには、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。

本研究では、そのための第一歩とし、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

的關係抽出についての検討を行った。

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[寺村 84]。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[影山 93]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[由本 05]。

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す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、それらをふまえて異なる観点からの検討を追加する必要もある。

本研究では、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た。計算機可読辞書を用いることの理由のひとつに、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置く。書かずに残す」という2つの語義がある(語釈文は「基本語データベース」[NTTCS08]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)。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。

複合動詞とその構成要素の意味的関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、分析のためのパター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ー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、意味的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、意味的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。

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 は意味的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 が意味的主辞となっていると思われるものが大半を占めている。

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

- 言いそびれる：言う機会を失って、言わずに終える。(v1 否定形式)
 売れ残る：商品が売れないで残る。(v1 否定形式, v2 終止形式)
 あきれ返る：完全にあきれる。(v1 終止形式)
 言い表わす：思っていることを言葉で表す。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植えて入れる。(v1 連用形式)
 打ち上げる：打って高く上げる。(v1 連用形式, v2 終止形式)
 撃ち殺す：鉄砲などで撃って殺す。(v1 連用形式, v2 終止形式)
 書き添える：文章や絵の側に添えて書く。(v1 終止形式, v2 連用形式)
 送り出す：送って外へ出す。出かける人を送る。(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

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、表 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。

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

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。そ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

本研究では、テキスト意味理解を高度化することを目標に、動詞の意味をそれが関係する動詞との関係を用いて記述するための基盤作りと検討を行った。様々な意味的形態的關係にある動詞を整理すると共に、複合動詞とその構成要素との関係を明らかにした。これらで得られたデータベースは現在、公開準備中である。また、特に複合動詞の意味についての検討を進めるため、その抽出方式を提案した。そして、複合動詞とその構成要素との意味的關係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。

【参考文献】

- [Breen09] 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.
[chasen] <http://chasen-legacy.sourceforge.jp/>.

- [影山 93] 影山太郎. 文法と語形成, ひつじ書房, 1993.
 [影山 99] 影山太郎. 形態論と意味, くろしお出版, 1999.
 [森田 90] 森田良行. 日本語学と日本語教育, 凡人社, 1990.
 [森田 08] 森田良行. 動詞・形容詞・副詞の事典, 東京堂出版, 2008.
 [NTTCS08] 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.
 [寺村 84] 寺村秀夫. 日本語のシンタクスと意味 II, くろしお出版, 1984.
 [由本 05] 由本陽子. 複合動詞・派生動詞の意味と統語, ひつじ書房, 2005.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
語釈文を用いた複合動詞の特徴分類	言語処理学会第17回年次大会	2011年3月