

複合動詞データベース構築のための付与情報

著者	神崎 享子
雑誌名	国立国語研究所論集
号	3
ページ	1-18
発行年	2012-05
URL	http://doi.org/10.15084/00000487

複合動詞データベース構築のための付与情報

神崎 享子

国立国語研究所 理論・構造研究系 プロジェクト研究員

要旨

「動詞＋動詞」型の複合動詞は、使用頻度の面でも表現力の面でも、日本語に特徴的な語彙であるが、統語的、意味的情報を付与してデータベース化している研究はまだ少ない。そこで、本稿では、語彙的複合動詞の形態的、統語的、意味的情報にとって何が必要かを検討する。まず、研究書や辞書などから収集した約 2500 語の複合動詞について量的観点から構成をとらえる。次に、情報付与の検討にあたって、既存のデータベースの現状を調査し、どのような情報が不足しているかを探る。そして、現在の言語学の複合動詞研究と、既存の基本動詞辞書の両方の観点から、必要な情報をまとめ整理し、それらの情報を実際に付与するにあたり、どのような基準あるいは知見を参考にするかを述べる。最後に、第一段階で構築中のデータベースの一部を掲載する。

キーワード：和語動詞、動詞＋動詞型複合動詞、付与情報、データベース

1. はじめに

本稿では、日本語の「動詞＋動詞」型複合動詞のデータベースを構築する際に、どのような情報を付与することが必要かを述べる。日本語の「動詞＋動詞」型の複合動詞は、東アジアの他の言語と比べても特徴的であり、とりわけバリエーションの豊かさは他に類を見ない。また、野村・石井（1987）の報告にもあるように日常的に使用頻度が高く、このタイプの複合動詞の役割は非常に大きい。このため、言語学や日本語学でも、理論的、記述的、計量的観点から「動詞＋動詞」型複合動詞は研究が進められている（以下「動詞＋動詞」型複合動詞を、「複合動詞」と呼ぶ）。たとえば、理論言語学では、影山（1993）や由本（2005）などがその語形成メカニズムを解明しようとし、日本語学の分野では姫野（1999）などが語構成を分類整理し、また、野村・石井（1987）や石川（2010）、Tamaoka, Lim and Sakai（2004）などは、計量的に語構成や用法を調査・分析している。

一方、日本語教育では、森田（1978）などが指摘しているように、複合動詞の学習・習得は必要性があるにもかかわらず難しいと言われている。日本語学習者に役立つような情報を付与した辞書が、Tagashira and Hoff（1986）などに一部見られるものの、まだ規模も小さく少数である。規模の大きい日本語教育用のデータベースとしては山下（2007）があるが、量的な面が充実している一方で、日本語学習で使用する場合には、質的な面でさらなる情報を追加する必要がある。自然言語処理分野でも、複合動詞の登録語数は単純語に比べてまだ少ない。

このように、「動詞＋動詞」型複合動詞の研究は、語彙論、語形成論、語構成論などで多角的に進んできたが、実際に言語研究や日本語学習、自然言語処理などに利用できるようなまとまったデータベースは存在しないのが、現状である。そこで筆者は、先行研究の知見を踏まえながら、

汎用的に役立つ複合動詞データベースを構築する作業を進めている。これによって、今まで扱えなかった統語情報や意味情報からの複合動詞検索や、動詞と動詞の組み合わせパターンからの検索、また、前項動詞と後項動詞のどちらの統語・意味情報が複合動詞に継承されているかなども調べられるようにし、システムティックな複合動詞学習にも貢献できるようにする。

本論文では、複合動詞データベースの一般的な性質を俯瞰した上で、「動詞＋動詞」型複合動詞のデータベースをどのように設計し、どのような情報を付与するのが適切であるのかを考察する。

2. 対象とする複合動詞の範囲

現在構築中の複合動詞のデータベースでは、和語動詞からなる「動詞＋動詞」型複合動詞を扱う。影山（1993）ではこのタイプの複合動詞を、後項動詞（後ろに位置する動詞）が直接、前項動詞（前に位置する動詞）をとるか、あるいは動詞句などをとるかという観点から、2種類に分けている。一つは、語彙的複合動詞と呼ばれ、一語としてのまとまりが強い複合動詞である。たとえば「聞き漏らす」はよいが「見漏らす」は使われないことから、「漏らす」がどのような動詞にも結合するわけではないことがわかる。もう一つは、統語的複合動詞と呼ばれるもので、統語的な派生から形成され、比較的制限を受けずに自由に結合することができるものである。たとえば、「始める」などは、「歌い始める」「勉強し始める」「通い始める」「作り始める」などどんな動詞にでも結合することができ、補助動詞の役割をもつ。意味の面でも、二つの単純語の動詞の語義が変化を受けずに合成されている。また、統語的複合動詞は後項動詞の数が少なく、影山（1993）では27語、姫野（2001）では、3語追加して30語としている。これらの動詞は、前項動詞と自由に結合することから、生産性が高い。

本データベースでは、辞書や研究書の用例などから、まず語彙的複合動詞を抽出する。この語彙的複合動詞と統語的複合動詞を区別する基準については、影山（1993）にいくつか上がっており、松本（1998）などでも検証を行っている。最も判別しやすい基準の一つは「そうする」を代用する方法である。語として不可分な語彙的複合動詞の場合は、語の一部が文中を照応することができないという特徴を用いた方法である。たとえば、「歌い始める」は「そうし始める」と言うことができるが、「取り戻す」は「そうし戻す」と言えない。これは、前者が統語的複合動詞で、後者が語彙的複合動詞となる。姫野（2001）では、「燃え出す」は「そうなり出す」のほうが自然な代用なので、「そうなる」の代用も許すほうがよい、と提案している。このような区別を行い、本稿では語彙的複合動詞のみを取り出して論じる。

3. 収集した複合動詞の構成

本データベースでは、複合動詞の研究書の用例や辞書などから、語彙的複合動詞を2585語収集した。複合動詞を収集した資料を下記に示す。

・影山太郎（1993）『文法と語形成』東京：ひつじ書房

- ・田中茂範・松本曜（1997）『空間と移動の表現』東京：研究社
- ・姫野昌子（1999）『複合動詞の構造と意味用法』東京：ひつじ書房
- ・姫野昌子（2004）『日本語表現活用辞典』東京：研究社
- ・山下喜代（2007）『日本語教育のための合成語のデータベース構築とその分析』平成 17 年度（2005）～平成 19 年度（2007）科学研究費補助金 基盤研究（c）研究成果報告書

収集した複合動詞の量的構成を調べるために、異なり語の複合動詞のリストから、前項動詞と後項動詞の出現頻度を計算した。表 1 が前項動詞の出現度数分布で、表 2 が後項動詞の出現度数分布である。それぞれの語彙リストは本論文の最後に付録で示す。

集計データでは、前項動詞の合計語数が 789 語、後項動詞の合計語数が 452 語となり、前項に出現する動詞は後項に出現する動詞と比べて種類が豊富であることがわかる。一方、一つの動詞がいくつの動詞と結合するかを見ると、後項動詞の方が、一つの語に対して結合する動詞が多いことがわかる。「込む」「出す」「つける」「上がる」などは、100 種類以上の前項動詞と結合していることが指摘できる。また、両方に共通する特徴としては、頻度 1 の語数が多い。表 1 では、前項動詞の頻度が 1 の語数は 372 語あり、全体の 47% (372/789) を占める。表 2 では、後項動詞の頻度が 1 の語数は 211 語あり全体の 46% (211/452) を占める。一方、組み合わせでみると頻度 1 どちらの組み合わせは、「ついばむ」「裂け広がる」「凝り固まる」「忌み嫌う」「慌てふためく」とごく少数であることがわかった。組み合わせの表の一部は表 9 にも掲載している。

表 1 前項動詞出現度数分布

出現度数	要素数
70-79	1
60-69	0
50-59	2
40-49	2
30-39	3
20-29	8
10-19	24
5-9	93
2-4	284
1	372
合計	789

表 2 後項動詞出現度数分布

出現度数	要素数
200-300	1
100-199	5
90-99	0
80-89	3
70-79	1
60-69	0
50-59	3
40-49	10
30-39	4
20-29	9
10-19	22
5-9	36
2-4	147
1	211
合計	452

このうち、前項にも後項にも出現する動詞は、217語であった。また、多くの動詞がどちらか一方に出現することから、前項か後項かによって動詞の意味に一定の傾向があることが示唆される。両方に出現する動詞、前項に出現する動詞、後項に出現する動詞のそれぞれの表の一部を表3、表4、表5に掲載する。

表3 前項にも後項にも出現する動詞リストの一部

よみ	前項・後項	例（よみ）	例（表記）
あおぎ	前項	あおぎみる	仰ぎ見る
あおぐ	後項	ふりあおぐ	振り仰ぐ
あがり	前項	あがりこむ	上がり込む
あがる	後項	おびえあがる	おびえ上がる
あきれ	前項	あきれいる	あきれ入る
あきれる	後項	おどろきあきれる	驚きあきれる
あけ	前項	あけはらう	開け払う
あける	後項	おしあける	押し開ける
あたり	前項	あたりちらす	当たり散らす
あたる	後項	いきあたる	行き当たる
あて	前項	あてつける	当て付ける
あてる	後項	ひきあてる	引き当てる
あふれ	前項	あふれおちる	あふれ落ちる
あふれる	後項	みちあふれる	満ちあふれる

表4 前項のみに出現する動詞リストの一部

前項よみ	例（よみ）	例（表記）
にぎり	にぎりしめる	握りしめる
かこい	かこいこむ	囲い込む
いばり	いばりちらす	威張り散らす
うつし	うつしうえる	移し植える
むせび	むせびなく	咽び泣く
ひき	ひきこもる	引きこもる
おし	おしもどす	押し戻す
おっ	おっばなす	押っ放す
なぐり	なぐりかかる	殴りかかる
くわえ	くわえだす	加え出す
うたい	うたいかける	歌いかける
かせぎ	かせぎだす	稼ぎ出す
すぎ	すぎさる	過ぎ去る
とき	ときほぐす	解きほぐす

表5 後項のみに出現する動詞リストの一部

後項よみ	例（よみ）	例（表記）
あがる	まきあがる	巻き上がる
あげる	かりあげる	刈り上げる
あつめる	よせあつめる	寄せ集める
いれる	かりいれる	刈り入れる
おこす	まきおこす	巻き起こす
おこる	まきおこる	巻き起こる
かえす	まきかえす	巻き返す
かかる	あまえかかる	甘えかかる
きらう	いみきらう	忌み嫌う
くる	よせくる	寄せ来る
こむ	かりこむ	刈り込む
つける	まきつける	巻き付ける
なおる	おきなおる	起きなおる
よせる	すいよせる	吸い寄せる

4. 複合動詞のデータベースに必要な情報

4.1 節では、まず、既存の複合動詞データベースの付与情報を概観して、4.2 節で現状で何が不足しているかを検討する。次に、4.3 節で複合動詞研究で着目される特徴と動詞データベースで付与されている情報を俯瞰し、4.4 節で複合動詞データベースを構築する際に、どのような情報が必要かを検討する。

4.1 既存の複合動詞データベースの情報

まず既存の複合動詞データベースの付与情報を概観しながら、言語学研究や日本語学習などへの活用のためには、どのような情報が不足しているかについて述べる。

(1) 野村雅昭・石井正彦 (1987) 『複合動詞資料集』

野村・石井 (1987) のデータベースには複合動詞が 7432 語収録されており、文学作品・論説文・雑誌・国語辞典など幅広い資料から複合動詞が収集されている。掲載されている表は、文学作品や辞典などの出典情報がついた複合動詞五十音順リストや、複合動詞連発表、五十音順構成要素表、度数順構成要素表、頻度 10 以上の前接率構成要素表、後接率構成要素表などである。

これらのリストに対して付与された情報は、複合動詞の出現度数、ジャンル別（文学作品等か辞書か）における前項要素および後項要素としての出現度数、全体における前項要素および後項要素としての出現度数、全体における前接率、後接率である。接合率や度数分布の表を見ることで、前に結合しやすい動詞、後ろに結合しやすい動詞など、複合動詞の構成を量的な観点からとらえることができる。

野村・石井 (1987) の複合動詞リストは、複合動詞の異なり語を幅広く集め、量的に十分なリストを作成しているところに利点がある。しかし、その一方で統語情報や意味情報などは付与されていない。そのため、統語や意味を扱う言語学研究や日本語学習などでは、本資料集をそのまま活用することが難しい。

(2) Yoshiko Tagashira and Jean Hoff (1986), *Handbook of Japanese Compound Verbs*

本書は、日本語学習者や日本語を母語としない話者が、複合動詞の意味や用法を正しく理解することをサポートする目的で編纂されたものである。研究社の日英辞書の見出し語から取り出した 1000 語の複合動詞のリストを作成し、そのうち、日本語学習にとって重要な複合動詞 200 語を選定して、情報を付与したものである。

付与情報は、①前項動詞と後項動詞、複合動詞自体の日本語読みと英語訳、並びに前項動詞と後項動詞の意味関係、②複合動詞の定義文、③文型と文型パターンに入る名詞、④例文（日本語文と英訳）から構成されている。たとえば、「握りしめる」は、次のように記載されている。

(例) NIGIRISHIMERU

- ① V1=nigiru=grasp

V2=-shimeru=(intensification)

CV="activity-manner"=grip tightly

- ② Hold something with one or both hands as one's own possession with no intention of letting it go;
- ③ [people] ga [thing/body part] o nigirishimeru
- ④ 三つか四つかの例文 (ローマ字表記の日本語文, 日本語表記の日本語文, 英訳)

本書は、日本語学習にも、また統語や意味を研究する研究者にも有益な情報を付与しているが、その反面、200語程度の複合動詞しか含まないので、量的に少なすぎるのが問題である。複合動詞の数は非常に多いので、やはり実際に利用する場合は、ある程度の数の複合動詞に対してこのような情報をつける必要があるだろう。

加えて、二つの動詞が複合動詞の構成要素になった場合の、意味・文型パターン・用法の変化が情報として付与されているとさらに有効であろう。これらの変化を示す情報があれば、言語学研究においても語形成のメカニズムを解明する手掛かりになり、また、日本語教育においても、複合動詞の丸暗記ではなく、複合動詞のシステムティックな合成や未知語の複合動詞における意味の推測などを可能にする一助になると予想される。

(3) 山下喜代 (2007) 『日本語教育のための合成語のデータベース構築とその分析』

山下のデータベースも日本語教育における複合動詞学習を念頭に置いて作成されたものである。国語辞典7種類と『分類語彙表増補改訂版』、『日本語能力試験出題基準語彙表』から、複合動詞だけでなく派生語も含む全ての合成語を抽出し、語種も和語や漢語など幅広くとられている。延べ16040語、異なり4828語が収録されており、合成語と造語成分のデータベースとなっている。造語成分とは、語の構成要素として造語力をもつ形態素で、山下(2007)では、造語成分を日本語学習に利用することで、既習の語彙と結びつけて新たな語が作られることがわかるので、データベース化することは有用であろうと述べている。

付与されている情報は、造語成分では、①見出し語、②表記、③前項と後項のどちらの成分になるかを示す接続、④語種、⑤語例、⑥資料出現頻度、⑦資料採録状況である。一方、合成語データベースでは、⑧見出し語、⑨よみ、⑩品詞、⑪語種、⑫『日本語能力試験出題基準語彙表』にあるか無いか、⑬造語成分が前項と後項のどちらにあるかを示した接続、⑭結合成分(造語成分ではない方の語)、⑮新聞の頻度や⑯webの頻度である。表6、表7に例をあげる。

このリストは、新聞やwebなどで複合動詞の出現頻度を計算することで、前項動詞や後項動詞が造語成分としてどれくらい生産性があるか把握できるようになっている。日本語学習において、複合動詞の生産性が高い構成要素(前項や後項にたつ動詞)を体系的に習得させ、複合動詞を一つ一つ覚えなくてもシステムティックに造語できるようにすることを目的としている。その点では、今までのデータベースにはない動的な観点であるといえる。しかし、それは量的な側面からのアプローチであるため、野村・石井(1987)の場合と同様に、統語的・意味的情報は付与されていない。そのため、言語研究や日本語学習にはそのまま活用ができないことが問題である。

表6 山下 (2007) の造語成分データベース

一かか る	掛か る・掛る	後接	和語	もたれかかる, 攻めかかる, 倒 れ掛かる	3				新明	学研			分語
①	②	③	④	⑤	⑥	⑦							

表7 山下 (2007) の合成語データベース

攻め掛かる	せめかかる	動詞	和	C	後接	掛か る	一かかる	攻め	35	74
⑧	⑨	⑩	⑪	⑫	⑬			⑭	⑮	⑯

これ以外にも加藤・林・伊藤 (2011) では、「動詞+動詞」型の複合動詞を、その構成要素である動詞と特定の意味関係に関連づけたり、複合動詞の語彙概念構造をその構成要素の語彙概念構造と結びつけて関係づけて表現したりすることを目的として、複合動詞に統語的・意味的情報の付与を行っている。

4.2 付与信息のまとめ

ここで、既存の複合動詞データベースにどのような情報が付与されてきたかを、大きくまとめると次のようになる。

- ・複合動詞の量的構成：複合動詞の出現頻度、前項動詞と後項動詞の出現頻度
- ・複合動詞の定義文、前項動詞と後項動詞の定義文
- ・前項動詞と後項動詞の意味的關係
- ・複合動詞の文型パターンとその文型に代入される名詞（名詞はカテゴリ化されている）
- ・複合動詞が収録されている辞書
- ・造語力のある形態素かどうか
- ・用例
- ・類例（類義表現）

これらの情報は複合動詞の基本的な特徴を示す重要な情報であるが、逆に、従来のデータベースで不足している情報は、大きく二つ考えられる。

一つ目は、統語的・意味的情報が付与された複合動詞の数が少ない。野村・石井 (1987) の報告では、雑誌や文学作品などから抽出した2万語の異なり語の中で35%が「動詞+動詞」型複合動詞であると述べている。日常的に使用頻度の高いこのタイプの複合動詞をデータベース化する際には、ある程度の規模が必要であろうと考える。

二つ目は、二語の動詞が結合して複合動詞になる際の、動詞どうしの有機的な関係がわからないため、体系的に複合動詞をとらえることができない。日本語学習においても、正しい組み合わせで複合動詞を生成したり、知らない複合動詞の用法や意味を推測したりするなど、動的な複合動詞学習にも役立つのではないかと考える。

このような問題点を踏まえて、次に、複合動詞の特徴を示す情報付与のためには何が必要かを検討する。

4.3 言語学研究と動詞辞書で着目される情報

これまで言語学的な研究では、複合動詞の語形成過程や語構成などの研究が進められており、複合動詞のいろいろな特徴が明らかになっている。このような言語学的研究から得た知見を取り入れて情報を付与することが重要である。一方、単純語の動詞辞書は、言語学研究や、言語処理、日本語学習など幅広い分野で有用なこともあり、従来から統語的・意味的情報が吟味されてきた。既存の動詞辞書では、どのような情報に着目しているかを調べることは有効であろう。以下、この二つの観点から付与情報を検討する。

4.3.1 言語学研究で分析される観点

複合動詞の語形成や語構成などの研究では、次のような複合動詞の特徴に着目している。

(1) 複合動詞、前項動詞、後項動詞のそれぞれの文型と名詞の選択制限

たとえば、「雪が降り積もる」は「N1が(N2に)降り積もる」(Nは名詞)という文型をとり、「降る」も「積もる」も「N1が(N2に)V」(Vは動詞)という文型をとる。しかし共起する名詞をみると「雪が積もる」とは言えても、「雨が積もる」とはいえない。「雨が」が主語にたたないという状況から考えると「降り積もる」は「積もる」の方の特徴を持っていると考えられる。このように文型に違いはないが、そこにどんな名詞がたつかによって、複合動詞の特徴は前項動詞に似ているのか後項動詞に似ているのかが判断できる。文型がどんな名詞をとるかの情報は Tagashira and Hoff (1986) にも見られたが、複合動詞研究においても日本語学習においても非常に重要な情報である。

(2) 主語の意志性の有無

主語の意志性の有無は、動詞結合の際の制約条件になることがある。たとえば、影山(1993)では「他動性調和の原則」を提案しており、動詞結合の際の制約として、基本的に意志性のある主語をとる動詞は、意志性のない主語をとる動詞とは結合できないことを述べている(非対格自動詞は、非対格自動詞と結合するが他動詞や非能格自動詞とは結合しない)。たとえば、「転がり落ちる」とは言えるが、意志性をもつ主語をとる「(私はボールを)転がす」を「落ちる」と結合させて「転がし落ちる」とは言わない。主語の意志性は、複合動詞形成の制約にかかわる重要な要素となる。

(3) 複合動詞の主要部の位置

「主要部」の考え方にはいくつかの立場があるが、基本的に複合動詞がもつ統語的・意味的情報が前項と後項のどちらの動詞から継承されるかで主要部が決まる。継承されている情報をもつ動詞の方を「主要部」と考える。通常は、後項の動詞の統語構造が継承されるのが一般的であり、それを「右側主要部」と呼ぶ。しかし、必ずしも全ての統語的・意味的情報がいつも右側、つまり後項動詞から受け継がれているとは限らず、前項と後項の両方の動詞から継承しているものや、前項動詞から意味の特徴を継承しているものなど、複合動詞によって異なる。主要部がどちらの動詞にあるかがわかることで、複合動詞の統語的・意味的情報が推測できるとともに、複合動詞形成のしくみを解明するのに重要な手掛かりとなる。

(4) 前項動詞と後項動詞の意味的關係

複合動詞を構成する前項動詞と後項動詞の関係をとらえるものである。たとえば、「恋い慕う」なら並列の關係、「語り明かす」なら語りながら夜を明かすことから、付帯状況という關係になる。日本語を学習する際にも、動詞どうしの關係がわかることにより、複合動詞習得の手助けになる。また言語分析においても、前項と後項動詞の意味關係は、複合動詞の組み合わせによってどのような傾向を示すのかなど複合動詞研究においても重要な情報となる。

(5) 複合動詞のアスペクト

状態、動作、変化、などのアスペクトについても、その複合動詞にたつ名詞や副詞などの制限がかかわるため、重要な情報である。

4.3.2 既存の動詞データベースで分析される観点

動詞については従来から研究が進んでいることもあり、得られた知見を用いて情報が付与されているデータベースが多くみられる。複合動詞に対しても利用できる情報が多いと考えられる。

(1) 活用、文型パターン、共起名詞、受身、使役、アスペクトなどの情報

情報処理振興事業協会（1997）で作成した IPAL 基本動詞辞書は、使用頻度上重要と思われる基本動詞 861 語を取り上げ、詳細に統語的・意味的情報を記述している。

(2) 類義語、上位語

『日本語表記活用辞典』（姫野 2004）、『分類語彙表』（国立国語研究所 1964, 2004）、『日本語ワードネット』（情報通信研究機構 2009）、『EDR 電子化辞書』（情報通信研究機構 2002）、などにも付与されている。複合動詞の類義語や上位語、あるいは上位概念をとらえておくことで、前項動詞と後項動詞から推測できない意味をとらえることができる。たとえば「寝返る」という語は、「敵に寝返った」という表現で使われるが、前項動詞と後項動詞の「寝る」と「返る」からは想定できない意味である。これは、「裏切る」「そむく」などの動詞と関連づけることで日本語学習や情

報検索の助けとなる。

(3) 意味役割の付与

意味役割は、文型で出現する「Nガ」, 「Nヲ」, 「Nへ」などが述語動詞に対してどのような役割をもつかを表すものである。意味役割が重要なのは、これによって動詞がある程度分類できるからである。たとえば「[N1へ]行く」, 「[N2へ]出発する」, 「[N3へ]送る」, 「[N4へ]帰る」などでは「Nへ」は方向を表し、上記四つの動詞は、モノや人の移動を表す動詞のクラスとなっている。この格のパターンによって、大まかではあっても類似した動詞のクラスを特定できる。また、日本語学習の面でも、意味役割の付与によって文型とそこにたつ名詞の羅列だけよりもさらに効率よく動詞をとらえることができる。たとえば「[海外へ]行く。」「[母親へ]荷物を送る。」の場合、「行く」にも「送る」にも[方向]のへ格をとるという意味役割が付与されていれば、助詞「へ」の前にたつ名詞が何であっても方向の解釈をとることができる。

意味役割付与の試みは、アメリカでは Fillmore の格理論を応用した Baker, Fillmore and Cronin (2003) の FrameNet や、日本では、小原 (2011) や黒田ほか (2008) などが FrameNet に基づいたデータ構築を試みている。FrameNet とは別に、Kipper, Hoa and Palmer (2000) では、Levin (1993) の統語的情報で分類した動詞クラスに対して意味役割を付与した VerbNet などを構築している。

このように、意味役割は国内外でデータベースが作られるほど有用な情報である反面、未だに明確な判断基準がないことから、意味役割の粒度や付与の方針がデータベースによって異なり、標準的な意味役割の付与が実際難しい。

4.4 従来研究や既存動詞データベースからとらえた付与情報のまとめ

前節のまとめとして、従来の言語研究や既存の動詞データベースから得られた重要な情報を①～⑧までの項目に整理する。

- ① 複合動詞, 前項動詞, 後項動詞の自他 (主語の意志性の有無も含む) と定義文
- ② 類義語, 上位語
- ③ 文型パターンと用例
- ④ 文型に代入される名詞
- ⑤ 意味役割の付与
- ⑥ 主要部の判定
- ⑦ 前項動詞と後項動詞の意味的關係
- ⑧ 複合動詞のアスペクト

①～④までの項目は、実際に複合動詞データに付与する際に、既存のデータベースの情報が部分的にでも活用できるものである。⑤～⑧までは、情報を付与したデータベースが少なかったり一定の基準が未だ固まっていなかったり、既存の情報を直接利用できないものである。言い換

えれば、⑤～⑧の情報を広範囲の複合動詞に統一的に付与したデータベースは、これまでにない。上記①～⑧の項目の中で、構築中の複合動詞のデータベースに実際どのような情報を付与するかについて、次の第5節で述べる。

5. 複合動詞の語形成データベースへの付与情報

現在構築中のデータベースでは、これまでの辞書やデータベースに不足していた、複合動詞が合成される際の変化がわかるようにしていく点が特徴である。以下 4.4 節の項目に従って、実際に付与する情報について述べる。

(1) 複合動詞, 前項動詞, 後項動詞の自他 (主語の意志性の有無も含む) と定義文 (①), 文型パターンと用例 (③), 文型に代入される名詞 (④)

現在、複合動詞を対象にして、このような情報を載せているデータベースはまだ見られていない。したがって、動詞辞書の情報が手掛かりになる。情報処理振興事業協会 (1997) 『CD-ROM 版 計算機用日本語基本辞書 IPAL—動詞・形容詞・名詞—』には、主に単純語の動詞に関して①, ③, ④のような情報が付与されている。

(2) 類義語, 上位語・上位概念 (②)

類義語や上位語・上位概念は、シソーラスや類義語辞典などを利用することになる。辞書では『日本語表現活用辞典』(姫野 2004) に、動詞表現とともに類義語の複合動詞が掲載されている。シソーラスには、たとえば『分類語彙表』(国立国語研究所 1964, 2004) などがあるが、複合動詞の掲載は多くないので、やはり資料を組み合わせるとらえていく必要がある。また『日本語ワードネット』(情報通信研究機構 2009) や『EDR 電子化辞書』(情報通信研究機構 2002) では、類義語や上位語, 上位概念などを扱っており、そのような情報も利用できると考える。工程としては、それぞれの複合動詞に基本的な統語的・意味的情報を付与した後、付与作業を行う予定である。

(3) 意味役割の付与 (⑤)

意味役割 (または深層格) の設定には従来からいろいろな説があるが、国立国語研究所 (1997), 石綿 (1999) などが、これまでの意味役割の諸説をまとめている。益岡・田窪 (1992) で提案している格の設定は比較的広く普及しているが、判断する際の標準的な基準や用語の統一などが明確ではないので、現段階では意味役割を付与しない。

(4) 主要部の判定 (⑥)

複合動詞の主要部のとらえ方については、影山 (2009) の冒頭でまとめられている。

a. 範疇による定義 (Williams 1981)

語構造の主要部とは、その語全体の範疇を決定する要素である。

b. 包摂関係による意味的な定義 (Haspelmath 2002)

語構造の主要部とは、その語全体の上位概念 (hypernym) を表す要素である。

c. 範疇と意味を混合させた定義 (Arnoff and Fundeman 2005 など)

語構造の主要部とは、全体の範疇と意味を決定する要素である。

しかし、実際に複合動詞を観察してみると、範疇による主要部と意味による主要部は必ずしも一致しないことがわかる。たとえば「思い当たる」を例に考えると、「N1 は N2 に思い当たる」(例:[元木は][あることに]思い当たり, 部屋を飛び出した。)という文型パターンをとるが、二格をとるのは「思う」ではなく「当たる」の方の統語的特徴である。「[あることを]思う」は言えても「[あることに]思う」とは言わない。一方、「[壁に]当たる」という形で「当たる」に対して二格がたつのは自然である。このことから「思い当たる」の二格は「当たる」から継承されていることがわかる。しかし、今度は「思い当たる」と共起する名詞をみると、「あることに(思い当たる)」のように、思う内容を表す事柄が二格にたっている。「当たる」には、壁や電柱などのような具体的な名詞がたつので、思う内容を表す事柄は「当たる」と共起する名詞ではない。このことから、共起する名詞の意味的性質(選択制限)は「思う」から継承されていることがわかる。言い換えると、「思い当たる」は「思う」の意味の影響を帯びていることになる。そしてこれは、bの包摂関係(上位下位関係)による主要部の判定方法の一つにもなる。

このように、前項動詞と後項動詞のどちらに主要部があるかを定める場合、統語的主要部の決定は格の継承からとらえることができ、意味的主要部の決定は共起名詞の選択制限の継承からとらえることができる。

以上から、統語的主要部と意味的主要部を分けて、下記のように、それらが重なる場合と重ならない場合がわかるように記述することが必要になる。

A. 前項と後項のどちらかが主要部の場合

「統語的主要部」: 統語的情報(格の情報)を継承する動詞の方

「意味的主要部」: 共起名詞の選択制限を継承する動詞の方

B. 両方の動詞から情報が同程度に継承され、一方に決められない場合

C. どちらの動詞からも継承されていない場合

(5) 複合において生じる意味関係の情報(前項動詞と後項動詞の意味的關係 (⑦), 複合動詞のAspect (⑧))

⑦ 前項動詞と後項動詞の意味的關係と時間的順序

前項動詞と後項動詞の時間的順序については、影山(1993)と松本(1998)でも述べられている。二つの動詞が表す事象の時間的順序は、同時か、ずれるか、関係ないかに大別される。さらに細かく見ていくと、下記のような関係に分けられる。(前項動詞をV1, 後項動詞をV2とする。)

・時間的に同時の場合

a) V1とV2が同時進行(V1 and V2)(請い願う, 恋い慕う)

b) 付帯状況(V2 while V1)(飲み明かす, 攻め上る)

・時間的にずれる(関係ない)場合

- c) 手段・様態・程度などの意味関係 (V2 by V1 など) (流れ落ちる)
 - d) 前要素の意味が希薄化している場合 (ついでに)
 - e) 前要素に意味的主要部がある場合 (V1 as if V2) (咲き狂う, 晴れ渡る)
- a) ~ e) の区別をつけることで, まずは意味関係を大きく分類し, その後さらに精緻化していく予定である。

⑧ 複合動詞のアスペクト

Vendler (1967) などが activity (活動), achievement (達成), accomplishment (到達), state (状態) の四つのアスペクトのカテゴリーを提唱しているが, それ以外にも Pustejovsky (1995) の state, process, transition の三分類もある。動詞の時間相は一つの動詞に一つの側面だけとは限らない。同じ動詞でも共起する名詞によって, アスペクトの別の側面が現れる。たとえば, 「焼く」を例にとると, 「オープンで5分魚を焼いた」の場合は魚を焼いているプロセスに着目しており, 「オープンでケーキを焼いた」の場合は焼いてケーキが作り出されたことに着目している。

このように動詞は, 共起する名詞や副詞などによって, 多面的にアスペクトを表すこともあるので, 一つの動詞に対して一つのカテゴリーの情報と限定せず, 場合によっては複数を組み合わせて記述する。

6. 最後に——現段階の複合動詞データベース——

本データベースの特徴は, 前述したように, 複合動詞の語形成に係る情報も付与する点である。具体的には, 複合動詞全体だけではなく, 前項動詞や後項動詞についても, 単純語の動詞としての文型パターンや文型パターンに代入される共起名詞, アスペクトなども調べられるようにする。これによって, 複合動詞が前項動詞と後項動詞のどのような特徴を受け継いでいるかを見ることができる。たとえば主要部には, 統語的, 意味的にどちらの動詞の情報を引き継いでいるかを情報として付与している。さらに具体的にどのような変化が起きているかについて調べる場合は, 単純語の動詞としての情報と複合動詞の情報を見比べることができるようにする。

複合動詞のデータベースの情報としては, 前節の①~⑧の情報が必要であるが, まずは意味役割 (⑤) 以外の①~⑥の情報を付与する作業を進めている。表8に, それらの情報を付与したデータベースの一部を掲載する。左から, 「V1 + V2 (複合動詞) のよみ」「V1 + V2 の表記」「前項 (V1) よみ」「前項表記」「前項頻度」「後項 (V2) よみ」「後項表記」「後項頻度」「活用」「自他」「定義文」「文型」「例文」「類義語」「統語的主要部」「意味的主要部」「V1 と V2 の意味関係」の順になっている。

表 8 複合動詞データベースの一部

V1+V2 よみ	V1+V2 表記	前項 よみ	前項 表記	前項 頻度	後項 よみ	後項 表記	後項 頻度	活用	自他	定義文	文型	例文	類義語	統語的 主要部	意味的 主要部	V1 と V2 の 意味関係
あおぎみる	仰ぎ見る	あおぐ	仰ぐ	1	みる	見る	3	上一	他	上を向いて高いところをみる。	N1 ガ N2 ヲ	彼女は 空を 仰ぎ見た。 彼は 彼女を 仰ぎ見た。	見上げる	見る	見る	V1 は V2 の 付帯状況 (V2 while V1)
あおぎみる	仰ぎ見る	あおぐ	仰ぐ	1	みる	見る	3	上一	他	尊敬する。	N1 ガ N2 ヲ N3 ト	彼は 先輩を 師と 仰ぎ見た。	見上げる／尊敬する	見る	仰ぐ	V1 は V2 の 付帯状況 (V2 while V1)
あおぎみる	仰ぎ見る	あおぐ	仰ぐ	1	みる	見る	3	上一	他	尊敬する。	N1 ガ N2 ヲ S ト	彼は その友達を 戦後最大の芸術家だと仰ぎ見た。	見上げる／尊敬する	見る	仰ぐ	V1 は V2 の 付帯状況 (V2 while V1)
あおりたてる	煽り立てる	あおる	煽る	1	たてる	立てる	48	下一	他	他人を刺激してある行動に煽り立てる。	N1 ガ N2 ヲ	あおる, 挑発する, けしかける, 煽り立てる, 扇動する, 奮い起こす, 掻き立てる	立てる	煽る	V2 は V1 の 様態・程度	
あおりたてる	煽り立てる	あおる	煽る	1	たてる	立てる	48	下一	他	他人を刺激してある行動に煽り立てる。	N1 ガ N2 へ	(排日移民法の可決が, … 日本の上論を反米感情へと煽りたてることになる。)	あおる, 挑発する, けしかける, 煽り立てる, 扇動する, 奮い起こす, 掻き立てる	煽る	煽る	V2 は V1 の 様態・程度
あがりこむ	上がり込む	あがる	上がる	1	こむ	込む	279	五段	自 (意志)	他人の場所に遠慮なく入っていく。	N1 ガ N2 ニ	男は部屋に上がり込んで,	入る	込む	上がる	V1 は V2 の 付帯状況 (V2 while V1)
あがりこむ	上がり込む	あがる	上がる	1	こむ	込む	279	五段	自 (意志)	他人の場所に遠慮なく入っていく。	N1 ガ N2 カラ N3 へ	彼は縁側から座敷へ上がり込んで,	入る	込む	上がる	V1 は V2 の 付帯状況 (V2 while V1)
あがりこむ	上がり込む	あがる	上がる	1	こむ	込む	279	五段	自 (意志)	他人の場所に遠慮なく入っていく。	N1 ガ N2 へ	彼女は部屋へあがりこんだ。	入る	込む	上がる	V1 は V2 の 付帯状況 (V2 while V1)

このデータベースでは、後項動詞または前項動詞を中心にソートしたり、文型を中心にソートしたりして、それぞれの特徴をみることが出来る。前項動詞や後項動詞の情報も同様に付与している。さらに、複合動詞の前項と後項の動詞がどのような出現頻度の組み合わせかについてもリストを提供する。前項と後項がそれぞれ頻度1ずつの組み合わせであれば、前項と後項の動詞が固定化した組み合わせであり、出現頻度が高いものどうしの組み合わせであれば、前項と後項の動詞が多義的であったり、意味が希薄になり接頭辞化していたりなどの可能性が考えられる。前項動詞と後項動詞の出現頻度からみた組み合わせリストの一部を表9に示す。

表9 複合動詞の前項動詞と後項動詞の出現頻度からみた組み合わせリストの一部

VV 表記	VV よみ	前項表記	前項よみ	前項度数	後項表記	後項よみ	後項度数
慌てふためく	あわてふためく	慌て	あわて	1	ふためく	ふためく	1
忌み嫌う	いみきらう	忌み	いみ	1	嫌う	きらう	1
返り咲く	かえりざく	返り	かえり	1	咲く	ざく	2
ささくれ立つ	ささくれだつ	ささくれ	ささくれ	1	立つ	だつ	2
去りゆく	さりゆく	去り	さり	1	ゆく	ゆく	2
光り輝く	ひかりかがやく	光り	ひかり	1	輝く	かがやく	2
もがき苦しむ	もがきくるしむ	もがき	もがき	1	苦しむ	くるしむ	2
養い育てる	やしないそだてる	養い	やしない	1	育てる	そだてる	2
仰ぎ見る	あおぎみる	仰ぎ	あおぎ	1	見る	みる	3
怒り狂う	いかりくるう	怒り	いかり	1	狂う	くるう	3
飼いならす	かいならす	飼い	かい	1	ならす	ならす	3
伝え聞く	つたえきく	伝え	つたえ	1	聞く	きく	3
出っ張る	でっばる	出っ	でっ	1	張る	ばる	3
うかがい知る	うかがいしる	うかがい	うかがい	1	知る	しる	4
変わり果てる	かわりはてる	変わり	かわり	1	果てる	はてる	4
計り知る	はかりしる	計り	はかり	1	知る	しる	4
売れ残る	うれのこる	売れ	うれ	1	残る	のこる	5
行い澄ます	おこないすます	行い	おこない	1	澄ます	すます	5
すれ違う	すれちがう	すれ	すれ	1	違う	ちがう	5
開き直る	ひらきなおる	開き	ひらき	1	直る	なおる	5
乱れ飛ぶ	みだれとぶ	乱れ	みだれ	1	飛ぶ	とぶ	5
もらい受ける	もらいうける	もらい	もらい	1	受ける	うける	5

参考文献

- Baker, Collin F., Charles J. Fillmore and Beau Cronin (2003) The structure of the FrameNet Database. *International Journal of Lexicography* 16: 281-296.
- 姫野昌子 (1999) 『複合動詞の構造と意味用法』東京：ひつじ書房。
- 姫野昌子 (2001) 「複合動詞の性質」『日本語学』20(8): 6-15.
- 姫野昌子 (2004) 『日本語表現活用辞典』東京：研究社。
- 石川慎一郎 (2010) 「現代日本語書き言葉均衡コーパス (BCCWJ) における複合動詞「～出す」の量的分析」『統計数理研究所研究レポート』238: 15-34.
- 石綿敏雄 (1999) 『現代言語理論と格』東京：ひつじ書房。
- 情報処理振興事業協会 (1997) 『CD-ROM 版 計算機用日本語基本辞書 IPAL—動詞・形容詞・名詞—』
- 情報通信研究機構 (2002) 『EDR 電子化辞書』http://www2.nict.go.jp/tr/r312/EDR/J_index.html
- 情報通信研究機構 (2009) 『日本語ワードネット』<http://nlpwww.nict.go.jp/wn-ja/>
- 影山太郎 (1993) 『文法と語形成』東京：ひつじ書房。
- 影山太郎 (2009) 「外心構造における意味と形態のミスマッチー「太っ腹」タイプの形容名詞—」由本陽子・岸本秀樹 (編) 『語彙の意味と文法』25-45. 東京：くろしお出版。
- 加藤恒明・林良彦・伊藤たかね (2011) 「複合動詞を用いた複合動詞の特徴分類」『言語処理学会第 17 回年次大会発表論文集』568-571.
- Kipper, Karin, Dang Trang Hoa and Martha Palmer (2000) Class-based construction of a verb lexicon. In: *The Proceedings of Seventeenth National Conference on Artificial Intelligence*, Austin, TX.
- 国立国語研究所 (1964) 『分類語彙表』東京：秀英出版。
- 国立国語研究所 (1997) 『日本語における表層格と深層格の対応関係』, 国立国語研究所報告 113, 東京：三省堂。
- 国立国語研究所 (2004) 『分類語彙表増補改訂版』東京：大日本図書。
- 黒田航・李在鎬・渋谷良方・井佐原均 (2008) 「複層意味フレーム分析 (の簡略版) を使った意味役割タグづけの現状: タグづけデータから派生する言語資源の紹介を中心に」『言語処理学会 14 回年次大会予稿集』PB2-5.
- Levin, Beth (1993) *English verb classes and alternations*. Chicago: University of Chicago Press.
- 益岡隆志・田窪行則 (1992) 『改訂基礎日本語文法』東京：くろしお出版。
- 松本曜 (1998) 「日本語の語彙的複合動詞における動詞の組み合わせ」『言語研究』114: 37-83.
- 森田良行 (1978) 「日本語の複合動詞について」『講座日本語教育』14: 69-86. 早稲田大学語学教育研究所。
- 野村雅昭・石井正彦 (1987) 『複合動詞資料集』(文部省科学研究費補助金特定研究 (1) 言語データの収集と処理の研究)。
- 小原京子 (2011) 「日本語フレームネットの全文テキストアノテーション: BCCWJ への意味フレーム付与の試み」『言語処理学会第 17 回年次大会予稿集』703-704.
- Tagashira, Yoshiko and Jean Hoff (1986) *Handbook of Japanese compound verbs*. Tokyo: Hokuseido.
- Tamaoka, Katsuo, Hyunjung Lim and Hiromu Sakai (2004) Entropy and redundancy of Japanese lexical and syntactic compound verbs. *Journal of Quantitative Linguistics* 11(3): 233-250.
- 田中茂範・松本曜 (1997) 『空間と移動の表現』東京：研究社。
- Vendler, Zeno (1967) *Linguistics in philosophy*. Ithaca: Cornell University Press.
- 山下喜代 (2007) 『日本語教育のための合成語のデータベース構築とその分析』平成 17 年度 (2005) ~平成 19 年度 (2007) 科学研究費補助金 基盤研究 (c) 研究成果報告書。
- 由本陽子 (2005) 『複合動詞・派生動詞の意味と統語—モジュール形態論から見た日英語の動詞形成』東京：ひつじ書房。

付録

付録1 前項動詞の頻度 10 以上の語彙リスト

前項	度数	前項	度数
取り	71	駆け	15
引き	55	持ち	14
打ち	51	呼び	13
押し	43	降り	13
見	43	乗り	13
突き	35	食い	13
思い	30	読み	13
切り	30	泣き	12
立ち	29	攻め	12
差し	26	売り	12
振り	22	行き	11
吹き	22	走り	11
言い	21	塗り	11
書き	21	投げ	11
掻き	21	引っ	10
踏み	21	巻き	10
飛び	19	組み	10
聞き	18	待ち	10
追い	16	流れ	10
買い	16	連れ	10

付録2 後項動詞の頻度 10 以上の語彙リスト

前項	度数	前項	度数
込む	279	あわせる	24
出す	150	まとめる	24
あげる	136	回す	24
上げる	136	合わせる	24
揚げる	136	抜く	24
付ける	103	切る	21
あがる	80	返す	20
おろす	80	去る	17
上がる	80	散らす	16
入れる	78	飛ばす	16
出る	56	込める	15
付く	53	倒す	15
入る	50	返る	15
おとす	49	払う	14
落とす	49	寄る	13
たつ	49	歩く	13
立つ	49	替える	12
たてる	48	起こす	12
立てる	48	起す	12
合う	46	下げる	11
かける	44	詰める	11
掛ける	44	殺す	11
取る	40	さがる	10
かかる	38	下がる	10
掛かる	38	止める	10
下ろす	30	直す	10
降ろす	30	渡す	10
回る	28	渡る	10
落ちる	28	戻す	10

Toward the Construction of a Database of V + V Compound Verbs

KANZAKI Kyoko

Postdoctoral Research Fellow, Department of Linguistic Theory and Structure,
National Institute for Japanese Language and Linguistics

Abstract

Compound verbs of the form “Verb+Verb” are known as a productive word formation type in Japanese. There are, however, few versatile databases of V+V compounds available. This paper investigates what kinds of linguistic information are expected in an adequate database of compound verbs. First, a list of compound verbs that are gathered from several dictionaries and other resources is explained. Then the kinds of information included in the existing databases of compound verbs are discussed with a view to clarifying the essential semantic, syntactic, and morphological information from two viewpoints: one is the viewpoint of linguistic research on the word-formation and lexical semantics of V+V compounds, and the other is the viewpoint of the information provided by Japanese dictionaries for simplex verbs. Finally, part of the database of compound verbs under construction is shown by way of illustration.

Key words: Verb+Verb compound verbs, database, semantic and syntactic information