

COMPOUND VERBS USAGE IN WRITTEN JAPANESE: A CORPUS-BASED ANALYSIS BY *CHAKOSHI*
Mamoru Hatakeyama, Columbia University

キーワード： 複合動詞 茶漉 コーパス 青空文庫

1. はじめに

複合動詞とは、広義ではある形態素に動詞が続き、動詞化した複合語で、名詞（腰かける、目覚める、役立つ）、擬態語の一部（べとつく）、形容詞の語幹（近寄る、若返る）にそれぞれ動詞が続き動詞化したものも含む。また、新美、山浦、宇津野（1987）のように、動詞と動詞の組み合わせで「取っておく」や「書いてみる」、「読んでしまう」、「なっていく」などのように、動詞のテ形に、動詞が続くものも指す場合もあるが、本稿では「話し合う」「落ち込む」などのように前の動詞（以下、前項動詞）の連用形にもう一つの動詞（以下、後項動詞）が結合したものに限定する。

複合動詞は日本語学習において、①結合条件が不透明、②辞書の記述が不十分、③多義性が高い、④単純動詞との意味の区別が困難、⑤教科書で与えられる動詞は単純動詞ばかりで、体系的に複合動詞が提示されることが少ないなどの理由により学習が困難な項目とされている（陳 2004、永井 1996、松田 2002a、2004、森田 1978）。具体例に、①結合条件として、「飲み込む」「食い込む」はいいが、「*かみこむ」「*かじりこむ」「*食べ込む」は非文になり、「（家に）走り込む」「駆け込む」「忍び込む」「上がり込む」「這い込む」「入り込む」はいいが、「*歩き込む」「*行き込む」は非文になることがあげられる。これらの違いを知るために、②辞書の後項動詞「込む」の項を見ても記述が十分にあるとは言えない。さらに、③同形であっても「鍵をかけ忘れる」は「かけるのを忘れる」なのに、「鍵を置き忘れる」は「置くのを忘れる」ではなくて「置いて、そのことを忘れたまま去る」ことを意味し、「呼びかける」の「かける」は相手への働きかけを意味するが、「言いかける」では「かける」は動作の開始直前、不完了を意味する。また、④「話す」と「話しかける」、「座る」と「座り込む」などの違いが分かりにくく、実際に学習者が使う場合には単純動詞を用いても概ね意味が通じてしまうため、回避が進んでしまいやすい。

このように学習が困難だとされる複合動詞について、その効果的な学習方法を検討するための基礎資料として、話し言葉における日本語母語話者、学習者の複

合動詞の使用状況についてコーパスを用いた量的調査がなされている(陳 2007)。これに加えて、本稿では、コロケーション抽出ツール「茶漉」(深田 2007)を使用し、小説などの書き言葉コーパスにおける日本語母語話者の複合動詞の使用状況および共起性の高い複合動詞について調査した。以下に、複合動詞のこれまでの先行研究について触れ、本調査の方法、結果および陳(2007)との比較について述べ、調査方法上の問題点について指摘する。

2. 先行研究

複合動詞についての研究は、これまで前項動詞と後項動詞の結合条件(石井 2001、2007、影山 1993、寺村 1978、1984、長嶋 1976、1997、松本 1998、山本 1984)、意味(姫野 1999)、習得(佐野 2005、寺田 2001、陳 2004、2007、松田 2002b、2004)などがある。またコーパスを用いた研究としては、類義の研究(杉村 2006a、2006b、2006c、2007)、後項動詞の生産性(松村 2007)や実際の複合動詞の使用状況(陳 2007)がある。

陳(2007)では、母語話者(上村コーパス¹(上村 1997))および学習者(KYコーパス²)のOPIコーパスを用いて話し言葉における複合動詞の使用を比較したところ、以下のような結果が得られたとの報告がある。①学習者は母語話者に比べて複合動詞の使用頻度、種類において少なく、習得が困難であることが示唆される。②学習者、母語話者とも複合動詞最頻 15 項目の延べ使用数が総複合動詞使用数の三、四割を占める。③使用頻度の高い上位 15 項目が複合動詞全体使用数の 7、8 割を占める。これらの結果は話し言葉における複合動詞の使用であるが、本稿では書き言葉における使用を調査した。

3. 研究課題

以上のことから、本稿ではコーパスを用いた書き言葉における日本語母語話者の複合動詞の使用について調査を試みるものである。研究課題としては以下の二点である。

研究課題① 書き言葉における使用頻度の高い複合動詞、後項動詞は話し言葉と比べてどのような違いがあるか。

研究課題② 書き言葉における前項と後項の共起性の高い複合動詞は何か。

4. 方法

4-1. 複合動詞の検索のための後項動詞の選定

複合動詞の前項と後項の関係には、並列（泣き叫ぶ、思い煩う、恐れおののく）のものもあるが、日本語の複合語において主要部が右側に来るという「右側主要部の規則 (Righthand Head Rule)」(Williams 1981) のため、後項動詞が複合動詞の意味づけの中心的役割を果たすことが多い(影山 1993、陳 2007)。そこでまず後項動詞を選定し、それと共に起る複合動詞を検索することにした。

後項動詞の選定は以下のように行った。まず、日本語 OPI コーパスを元にした母語話者、学習者の使用した後項動詞のうちそれぞれの上位 15 位に現れたものが 20 語あった(陳 2007:88)。つぎに、現代語辞書 3 種類(学研国語大辞典、岩波国語辞典、新明解国語辞典)及び用例集(国立国語研究所 1971)の見出し語から得られた「既成」の複合動詞 2,494 語(石井 2007:31)の造語成分のうち、前項にも後項にも立つ両項動詞 212 語の上位 15 語、および後項にのみ立つ動詞 272 語の上位 18 語、さらに、姫野(1999)で見出しに取上げられた 23 語、新美、山浦、宇津野(1987)で見出しに取上げられた 40 語の中から、重複を除いた 58 語を選び出した。

辞書および用例集の見出しに現れた複合動詞の後項に現れた動詞だけでも 484 語(石井 2007)もあり、実際に使われているすべての複合動詞のすべての後項動詞について複合動詞を検索するべきであるが、本稿で取上げる 58 語は、結合する前項動詞が多く、使用頻度が高いとされたものであるため、使用実態を概観するための予備調査としては適当であろうと思われる。

4-2. 調査の対象コーパスデータおよび調査に使用したツール

用例調査の対象として青空文庫コーパスを利用した。青空文庫は日本文学 6,125 作品(2008年5月現在)を中心に、主に1890年から1958年³までの間に亡くなった明治、大正、昭和の作家の小説、論文、評論、随筆、エッセイなどの散文および詩、翻訳作品が収められている(夷石、千葉、陳 2006)。青空文庫コーパスの中の用例およびコロケーションの調査のために、茶漣(深田2007)⁴を使用した。茶漣は、奈良先端科学技術大学院大学自然言語処理学講座開発による形態素解析システム『茶筌』を用いてコーパスを検索可能なデータファイルに変換したものを利用し、パーデュー大学深田淳教授により開発されたウェブブラウザ・インターフェースを備えた用例およびコロケーション情報を抽出するシステムである。これまでに茶漣による日常会話における受身形、使役形、程度の表現の使用が発表されている(深田、大曾2007)。

4-3. データ収集方法

4-1 で選定した後項動詞 58 語について、茶漉を用いて用例を検索した。茶漉による検索の手順は以下の通りである。「～あう（合う）」の検索を例として示す。

- ① 検索のコーパスとして青空文庫を指定する。
- ② 検索パターン設定を、スパン前 1 語、後 0 語にする。
- ③ スパン-1 に動詞（品詞）、検索キーワード（kw）の語形に後項動詞の辞書形を指定し、全活用形のチェックボックスを選択する。
- ④ 表記のゆれによる取りこぼしを防ぐために、検索キーワードには、「あう、合う」というふうに漢字、ひらがなの表記をすべて入れ、「、」ではなく、「,」で列記する。また、コロケーション度の低いものも抽出されるように、コロケーション出力設定の t スコア、MI スコア閾値を 0 にする（図 1）。

茶漉 | 日本語用例・コロケーション抽出システム (一般公開版)

[辞書案内 | マニュアル | 問い合わせ]

コーパス指定: 青空文庫(小説集)
 名大会話コーパス【利用上の注意】
 名大会話コーパス(あいつぶ編)【利用上の注意】

検索パターン設定: 青空文庫(小説集)

検索パターン設定: 青空文庫(小説集)

スパン: 前 語 後 語

-3	語形: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外 <input type="checkbox"/> 全活用形
	品詞: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外
-2	語形: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外 <input type="checkbox"/> 全活用形
	品詞: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外
-1	語形: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外 <input type="checkbox"/> 全活用形
	品詞: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外
kw	語形: <input type="text" value="あう, 合う"/>	<input checked="" type="checkbox"/> 全活用形
	品詞: <input type="text"/>	<input type="checkbox"/> 音 <input type="checkbox"/> 除外

tスコア閾値:

MIスコア閾値:

図 1 茶漉 検索条件設定

- ⑤ 検索により、得られたコロケーション情報を、スパン頻度で降順に並べ替える（図 2）。
- ⑥ ここで、「向かい」がスパン頻度 101 となっているが、これは「向かい合う、向かいあう」が 101 例あることを示す。図 2 の範囲外にある「むかい」7 例と合わせ、「むかいあう」が 108 例あったものとして数えた。
- ⑦ 後項動詞 58 語について、それぞれ頻度数上位 15 位までのものについて同様に検索を行った。
- ⑧ 同様に、コロケーションの強さを測るため、t スコア、MI スコアを初期値 (2.0、3.0) にしても、後項動詞 58 語について検索を行った。

●コロケーション情報:

コーパス総語数= 8370720

スパン語数=1108

kwコーパス頻度=19230

形態素	tスコア	Mスコア	Gスコア	コーパス頻度	スパン頻度▲	期待頻度
係り	10.768	7.80	1096.27	227	116	0.0300
向い	10.038	5.56	592.39	931	101	0.1232
付き	7.480	6.85	437.21	211	56	0.0279
見	5.660	0.41	2.65	11427	35	1.5125
押し	5.182	4.38	113.71	565	27	0.0748
云い	4.929	2.02	32.45	2686	25	0.3555
笑い	4.647	2.64	43.75	1539	22	0.2037
握り	4.239	5.94	115.30	128	18	0.0169
覗き	3.997	6.48	115.57	78	16	0.0103
ゆれ	2.720	5.72	95.42	115	14	0.0152

図2 コロケーション情報 検索結果例

また、コーパスを使った研究では、検索結果の中から意図した表現とは異なる表現を排除する作業が必要であるが(杉村 2007: 291)が、その例として、前出の「あう、合う」の場合、動詞でないにもかかわらず「たぐ」が9例抽出された。これは「その隣にはさるしばいの小屋が軒をつらねているといっ たぐ あいでした」のように、形態素解析が正しく行われなかった場合に、このような例を排除することを含む。表記のゆれでは、例えば後項動詞「いる」の検索で、「おちいる」「陥る」「落ち入る」を別の語と数えるかどうかということも問題になる。

5. 結果

本調査において得られた複合動詞は、異なりで791種類、延べ使用数は13,936で、陳(2007)で検出された用例数よりもはるかに多い(表1)。これは、青空文庫コーパス総語数8,370,720語に対して、話し言葉コーパス総文字数は母語話者318,276字、学習者390,907字と、青空文庫コーパスのほうが規模がより大きいことによるが、それぞれの総文字数、総語数(総形態素数)が明らかにならなかったため比率による比較はできなかった。また、本調査では、後項動詞を頻

度の高いとされるもの 58 語に限定したが、それ以外の後項動詞からなる複合動詞も多くあるはずである。さらに、今回は前項動詞と後項動詞が二語からなる複合動詞として形態素解析されたものを数えたが、実際には解析のゆれから、後述する「もうしあげる」のように一語とされるものも多く、それぞれの表記のゆれも含めた検索をする必要があり、実際にはもっと多いものと思われる。

	青空文庫 2008		陳 (2007)	
	58 後項動詞		母語話者	学習者
	全体	各上位 15 位	全複合動詞	
総文字数			318, 276	390, 907
総語数 (総形態素数)	8, 370, 720			
異なり複合動詞使用数	(5, 935)	791	174	105
延べ複合動詞使用数	(24, 073)	13, 936	343	200

表 1 複合動詞使用の概要

一方で、コーパス全体の異なり複合動詞使用数 5, 935、延べ使用数 24, 073 は、前出の「あう、合う」の「たぐ」のような例が排除されていないため、実際にはこれよりも少ないが、参考のため載せた。なお、学習者コーパスの総語数は李、淺尾、濱野、佐野、井佐 (2008)による総形態素数の報告による。

5-1. 複合動詞の使用頻度上位 15 位

本調査において得られた頻度の高い複合動詞は次の通りであった (表 2)。

総合順位	青空文庫2008			陳(2007)学習者			陳(2007)学習者		
	複合動詞	用例数	割合	複合動詞	用例数	割合	複合動詞	用例数	割合
/791	58動詞各15位まで	13936		/174	343		/105	200	
	上位15位合計	3324	23.9%	上位15位合計	122	35.6%	上位15位合計	81	40.5%
1	いいだす	539	3.9%	でかける	25	7.29%	つきあう	16	8.00%
2	わらいだす	275	2.0%	ひっこす	21	6.12%	でかける	8	4.00%
3	なきだす	270	1.9%	はたらきすぎる	9	2.62%	のりかえる	8	4.00%
4	あるきだす	263	1.9%	つきあう	8	2.33%	うけいれる	6	3.00%
5	いいかける	245	1.8%	みつける	7	2.04%	おちつく	5	2.50%
6	みまわす	230	1.7%	もうしあげる	7	2.04%	おもいだす	5	2.50%
7	かけだす	230	1.7%	うちこむ	6	1.75%	とりくむ	5	2.50%
8	ふりかえる	214	1.5%	しあう	6	1.75%	おちこむ	4	2.00%
9	おもいだす	177	1.3%	おそれいる	5	1.46%	とりあげる	4	2.00%
10	しはじめる	161	1.2%	さしあげる	5	1.46%	にあう	4	2.00%
11	かけつける	160	1.1%	はなしかける	5	1.46%	はなしあう	4	2.00%
12	おそれいる	147	1.1%	むすびつく	5	1.46%	おもいきる	3	1.50%
13	のぞきこむ	143	1.0%	もうしこむ	5	1.46%	ききとる	3	1.50%
14	きかえる	135	1.0%	おちつく	4	1.17%	くりかえず	3	1.50%
15	かけこむ	135	1.0%	おもいはじめる	4	1.17%	しりあう	3	1.50%

表 2 複合動詞の使用頻度上位 15 位

これらの上位 15 位までの複合動詞の用例数の合計(3,324)は、全体（正確には、58 後項動詞のそれぞれの動詞の 15 位までの合計=13,936）の 23.9%となった。母語話者書き言葉（23.9%）、母語話者話し言葉（35.6%）、学習者話し言葉（40.5%）の順に、上位 15 位までの用例数が複合動詞全体の用例数に占める割合が低く、より複合動詞の種類が多いことが推測される。 χ^2 検定の結果、書き言葉（3324:13936）は、話し言葉のそれぞれ（122:343、81:200）よりも有意に少ないことがわかった（有意水準 $p < .01$ ）。

また、書き言葉においては、上位 15 位内の上位 4 位までは「いいだす」（1 位）、「わらいだす」（2 位）、「なきだす」（3 位）、「あるきだす」（4 位）で、そのほかに「かけだす」（7 位）、「おもいだす」（9 位）と 15 件のうち 6 件の後項動詞が「だす」であった。件数において 15 位までの半分以上（1754/3324=52.8%）にあたる。これは「～だす」が小説等で登場人物の動作の描写によく使われるためである可能性がある。これに対して、母語話者話し言葉では 15 位以内に一つも後項動詞に「だす」を用いるものはなく、学習者話し言葉に「おもいだす」（6 位）が 1 件あったのみである。茶漉で利用可能なもう一つのコーパス、名大会話コーパスにおいても、後項動詞「だす」の複合動詞のうちで、最も頻度の高かったのは「いいだす」（22 件）で、次に「なきだす」（8 件）となった。話し言葉において、物語り（ストーリーテリング）をする場合に、同じように「だす」の使用頻度が高い可能性もあるが、その点は今後の検証が必要である。

母語話者の書き言葉と話し言葉において複合動詞 15 位内に共通する複合動詞は、「おそれいる」（書き言葉 12 位、話し言葉 9 位）しかなかった。「おそれいる」は、母語話者話し言葉の「もうしあげる」（6 位）、「さしあげる」（10 位）とともに学習者話し言葉にはない待遇表現の複合動詞であった。なお、「あげる」の複合動詞では、第一位は「みあげる」（64 件）で「もうしあげる」は、本調査では後項動詞「あげる」の 15 位以内に入らなかった。しかし、「もうしあげる」を一語として検索すると 700 件以上もあり、後項動詞「あげる」で最も用例数の多い複合動詞となった。

前項を動詞に指定して後項動詞「あげる、上げる、上る」の検索をすると、「申し」で 6 件検索される。二語と分析される「もうしあげる」は、実際にはこのほかに前項動詞の表記が「申」で、「申上げる」「申上る」の活用形が 209 件もあるが、「申」の品詞が動詞と認識されないために一度の検索では探しきれない。さらに、「もうしあげる」およびその表記のバリエーションを一語として検索すると 507 件、合計で 722 件もあった。このように、表記および形態素分析のゆれのために、一語としても検索をする必要がある。同じような例として「にあらう」は、二語の検索では 5 件で「あらう」の上位 15 位に入らないが、一語では 236 件あり、総合一位となる。

5-2. 後項動詞使用頻度上位 15 位

使用頻度の高い後項動詞上位 15 位は次の通りであった（表 3）。後項動詞「だす」は複合動詞の総合順位で 1 位から 4 位および 7 位、9 位と 15 位までの 4 割を独占していただけあり、ここでも書き言葉における「だす」の頻度が突出して多いことがわかる。書き言葉、話し言葉（母語話者）の双方に共通するものがほとんどで、各 15 語のうち 8 割（12 語：あう、あがる、あげる、いる、かける、きる、こむ、だす、つく、つける、つづける、はじめる）は共通していた。

青空文庫 2008				陳(2007)母語話者				陳(2007)学習者			
	後項動詞58種合計	用例数	割合		総合計	用例数	割合		総合計	用例数	割合
58種	上位15位合計	9324	66.9%	68種	上位15位合計	234	68.2%	43種	上位15位合計	156	78.0%
1	だす	2158	15.5%	1	かける	34	9.9%	1	あう	38	19.0%
2	あがる	763	5.5%	2	あう	30	8.7%	2	だす	20	10.0%
3	つける	713	5.1%	3	こむ	23	6.7%	3	こむ	16	8.0%
4	こむ	685	4.9%	4	こす	21	6.1%	4	かえる	14	7.0%
5	かける	634	4.5%	5	あげる	19	5.5%	5	あげる	10	5.0%
6	かえる	628	4.5%	6	すぎる	16	4.7%	6	かける	10	5.0%
7	きる	519	3.7%	7	つく	13	3.8%	7	いれる	8	4.0%
8	あう	490	3.5%	8	つける	13	3.8%	8	すぎる	7	3.5%
9	まわす	478	3.4%	9	だす	12	3.5%	9	つく	7	3.5%
10	はじめる	443	3.2%	10	はじめる	12	3.5%	10	つける	6	3.0%
11	まわる	430	3.1%	11	きる	11	3.2%	11	くむ	5	2.5%
12	つづける	402	2.9%	12	なおす	10	2.9%	12	こす	5	2.5%
13	あげる	364	2.6%	13	いる	8	2.3%	13	あがる	4	2.0%
14	つく	325	2.3%	14	あがる	6	1.7%	14	かえす	3	1.5%
15	いる	292	2.1%	15	つづける	6	1.7%	15	かわる	3	1.5%

表 3 後項動詞の使用頻度上位 15 位

また、それぞれのコーパスで使用頻度上位 15 位の合計は全体の 7、8 割程度（66.9%、68.2%、78%）となっており、これらの後項動詞が頻繁に使われるグループとして類似の傾向を示している。 χ^2 検定の結果も、差は有意でないことがわかった。

5-3. 前項と後項の共起性の高い複合動詞上位 15 位

茶漉のコロケーションスコアを利用し、複合動詞の前項と後項の共起性の高いものを調べたところ次のようになった（表 4）。茶漉では、t スコア、MI スコアの 2 種類のコロケーションスコアが表示される。t スコアとは統計的有意性を測るために、「共起の程度が偶然による確率を超えていると、どのくらいの確かさで言えるかを示す指標」で「広く頻繁に用いられるコロケーション」の判定に適している（齋藤・中村・赤野（編）2005:132）。一方、MI スコアとは、ある語が共起す

る語の情報をどのぐらい持っているかということを示す指標で低頻度でも意味的に興味深いコロケーションを特定するのに有効とされる（齋藤・中村・赤野（編）2005:132）。茶漉の初期閾値（t スコア 2.0、MI スコア 3.0）のまま検索し、MI スコア降順で並べた。スコアは表記と結びついているので、ひらがなにしていない。

	複合動詞	tスコア	コーパス頻度	スパン頻度	複合動詞	MIスコア	コーパス頻度	スパン頻度
1	云いだす	16.891	2686	288	撒きちらす	14.44	30	16
2	笑いだす	16.293	1539	267	困じはてる	13.94	7	6
3	泣きだす	15.719	890	248	頑張りとおす	13.72	8	5
4	見まわす	15.114	11427	230	伐りたおす	13.48	18	6
5	歩きだす	14.843	682	221	蹴ちらす	13.12	89	19
6	駈けだす	12.417	813	155	堰きとめる	13.04	11	11
7	恐いる	12.082	378	146	小突きまわす	12.8	13	11
8	覗きこむ	11.955	235	143	見下げはてる	12.74	16	6
9	駈けつける	11.738	813	138	叱りとばす	12.71	88	12
10	言いだす	11.632	1690	137	ひねくりまわす	12.63	12	9
11	起きあがる	10.946	691	120	引っ掻きまわす	12.63	8	6
12	想いだす	10.803	300	117	喚びおこす	12.59	18	12
13	云いかける	10.78	2686	117	こづきまわす	12.58	11	8
14	係りあう	10.768	227	116	わめきちらす	12.54	56	8
15	出来あがる	10.675	4116	115	なぐりたおす	12.39	32	5

表 4 共起性の高い複合動詞上位 15 位

t スコア降順にした場合、頻度が高いもの（「云いだす」「笑いだす」「泣きだす」など）が多くなった。MI スコア降順にした場合、複合動詞としても、後項動詞としても、頻度は高くないが、共起性高い固有の組み合わせが抽出された。例えば、「ちらす」は後項動詞 58 語のうち頻度数は 35 位であったが、「撒きちらす」、「蹴ちらす」、「わめきちらす」と 3 件も入っている。

5-4. 研究課題 1 に対する結果

研究課題① 書き言葉における使用頻度の高い複合動詞、後項動詞は話し言葉と比べてどのような違いがあるか。

書き言葉において使用頻度の高い 15 位までのうち 6 件が、後項動詞「だす」の複合動詞（いいだす、わらいだす、なきだす、あるきだす、かけだす、おもいだす）で、コーパスが小説中心であるために、登場人物の描写に使われるものが多くなった可能性が考えられる。これに対して、話し言葉では、後項動詞「だす」を使ったものは母語話者コーパスにはなく、学習者コーパスに「おもいだす」があったのみである。また、書き言葉における頻度の高い複合動詞で、話し言葉との共通するものは「おそれいる」1 件のみであったが、「もうしあげる」などのよ

うに、一語として解析されやすいものが他にもある可能性がある。さらに、複合動詞の用例数 15 位までの合計が全体の用例数に占める割合を調べたところ、書き言葉コーパスでは話し言葉コーパスよりも有意に低く、より種類が豊富であることが予測された。頻度の高い後項動詞に注目すると、書き言葉と話し言葉では 15 位中 8 割が共通しており、頻繁に使われるものが多くあることが分かった。また、書き言葉コーパスにおいても、話し言葉コーパス同様に、後項動詞 15 位までの用例数の合計が全体の 7、8 割を占めることが分かった。

5-5. 研究課題 2 に対する結果

研究課題② 書き言葉における前項と後項の共起性の高い複合動詞は何か。

t スコアをもとに、共起性の高い複合動詞を検索すると、頻度数の高いものになる傾向がある。MI スコアをもとに検索すると、それほど頻度の高くない後項動詞の複合動詞も含まれることがわかった。MI スコアによる上位 50 位までは別表の通りである。

5-6. 結果のまとめ

本調査の書き言葉における複合動詞は、小説中心というコーパスのジャンル性のためか、後項動詞の「だす」を用いた複合動詞が多く抽出された。また明治から昭和にかけての時代性を示しうるものとして、「困じはてる」(MI 共起性 2 位)、「うち忘れる」(「わすれる」2 位)、「まかり出る」(「でる」2 位) などの文語的な表現が散見された。

6. 今後の課題

表記、解析のゆれを考慮し、後項動詞 58 語の 5,935 件について複合動詞を二語だけではなく、一語としても検索し、青空文庫だけでなく、均衡コーパスによる調査も正確な使用状況の調査に必要だろう。また、中級段階で体系的に学習する必要がある(森田 1978:74) 複合動詞の中でも、特に前項後項動詞から意味の透明な統語的複合動詞(影山 1993:96) に比べて、より習得の困難な語彙的複合動詞(寺田 2001:26) の効果的な学習のために、習得順序などの研究をさらに進めることが望まれる。

¹ 上村コーパスとは、日本語母語話者、非母語話者に対して行われた 15 分の日本語 OPI の文字化資料である。

陳(2007)ではこのうち、公開されている日本語母語話者 50 名分のデータが使用された。

<http://www.env.kitakyu-u.ac.jp/corpus/>

² KY (鎌田修、山内博之)コーパスとは、90 人分のOPIテープを文字化した言語資料である。90 人の被験者を母語別に見ると、中国語、英語、韓国語がそれぞれ 30 人ずつであり、さらに、その 30 人のOPIの判定結果別の内訳は、それぞれ、初級 5 人、中級 10 人、上級 10 人、超級 5 人ずつとなっている。http://www.opi.jp/shiryo/ky_corp.html

³ 著作権が作家の死後 50 年で切れるため、その年の 50 年前までになくなった作家の作品が公開可能となる。本人の存命中に使用を許諾しているケースもある。<http://aozora.gr.jp/>

⁴ <http://tell.fl.purdue.edu/chakoshi/public.html>

参考文献

- 石井正彦 (2001) 「複合動詞の語構造分類」『国語語彙史の研究 20』 国語語彙史研究会 和泉書院 59-71
- 石井正彦 (2007) 『現代日本語の複合語形成論』 ひつじ書房
- 影山太郎 (1993) 『文法と語形成』 ひつじ書房
- 上村隆一 (1997) 「データベースで調べる」『日本語学』11 明治書院 60-68
- 国立国語研究所 (1971) 『国立国語研究所資料集 7 動詞・形容詞問題語用例集』
- 齋藤俊雄、中村純作、赤野一郎(編) (2005) 『英語コーパス言語学：基礎と実践』 (改訂新版) 研究社出版
- 佐野香織 (2005) 「中・上級学習者の「つける」の典型概念—複合動詞類推研究に向けて—」『アメリカ・カナダ大学連合日本研究センター紀要』28 22-39
- 杉村泰(2006a) 「コーパスを利用した複合動詞「一直す」の意味分析」『言語文化論集』 28(1) 51-66 名古屋大学大学院国際言語文化研究科
- 杉村泰(2006b) 「本動詞「返す」と複合動詞「一返す」の意味の対応について」『ことばの科学』19 157-165 名古屋大学言語文化研究会
- 杉村泰(2006c) 「複合動詞「一忘れる」、「一落とす」、「一漏らす」の用法」『日語学习与研究』127 1-6 对外经济贸易大学
- 杉村泰(2007) 「コーパスを利用した複合動詞の類義分析 - インターネット検索エンジンの利用」『言葉と文化』8 289-304 名古屋大学大学院 国際言語文化研究科 日本言語文化専攻
- 夷石寿賀子, 千葉庄寿, 陳君慧 (2006) 「『青空文庫』を言語コーパスとして使おう—メタデータ構築による歴史的・社会言語学的研究への応用の試み—」『言

- 語処理学会第12回年次大会 (NLP2006) 発表論文集』 915-918
- 陳曦(2004) 「中国人学習者における複合動詞の習得に関する一考察—「～あう」と「～こむ」の理解に基づいて」『ことばの科学』17 59-79 名古屋大学言語文化研究会
- 陳曦(2007) 「学習者と母語話者における日本語複合動詞の使用状況の比較」『日本語科学』22 79-99 国立国語研究所
- 寺田裕子(2001) 「日本語の二類の複合動詞の習得」『日本語教育』109 20-29 日本語教育学会
- 寺村秀夫(1978) 『日本語の文法(上)』日本語教育指導参考書4 国立国語研究所
- 寺村秀夫(1984) 『日本語のシンタクスと意味II』くろしお出版
- 永井鉄郎(1996) 「日本語複合動詞の教育について」『日本語教育』88 140-151
- 長嶋善郎(1976) 「複合動詞の構造」『日本語講座4 日本語の語彙と表現』63-104 大修館書店
- 長嶋善郎(1997) 「複合動詞の構造」『日本語研究資料集 語構成』(齊藤倫明・石井正彦 編) ひつじ書房 213-231
- 新美和昭、山浦洋一、宇津野登久子(1987) 外国人のための日本語例文問題シリーズ4 複合動詞 荒竹出版
- 深田淳(2007) 「日本語用例・コロケーション情報抽出システム『茶漉』」『日本語科学』第22号 「特集: コーパス日本語学の射程」
- 深田淳、大曾美恵子(2007) 『茶漉』で見る日常会話 CASTEL-J in Hawaii 2007 Proceedings 125-128.
- 姫野昌子(1999) 『複合動詞の構造と意味用法』ひつじ書房
- 松田文子(2002a) 「複合動詞研究の概観とその展望 —日本語教育の視点からの考察—」『言語文化と日本語教育』170-184 お茶の水女子大学日本語言語文化学会
- 松田文子(2002b) 「日本語学習者による『～こむ』の習得」『世界の日本語教育』12 43-59 国際交流基金日本語国際センター
- 松田文子(2004) 『日本語複合動詞の習得研究—認知意味論による意味分析を通じて』 ひつじ書房
- 松村一登(2007) 複合動詞の生産性といわゆる「統語的/語彙的」の区別 —コーパスにもとづく考察— 日本言語学会 第134回大会 予稿集 378-383
- 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 日本言語学会
- 森田良行(1978) 「日本語の複合動詞について」『講座日本語教育』69-86 早稲田大学語学教育研究所
- 山本清隆(1984) 「複合動詞の格支配」『都大論究』21 32-49 東京都立大学

李在鎬、淺尾仁彦、濱野寛子、佐野香織、井佐原均 (2008) 「タグ付き日本語学習者コーパスの開発」 言語処理学会 第14回年次大会における発表

Williams, E. (1981). On the notions 'Lexically Related' and 'Head of a Word,' *Linguistic Inquiry* 12, 245-274.

別表1 後項動詞 頻度表

		用例数			
		合計	12174		
		上位15位合計	9324		
順位	後項動詞	用例数	順位	後項動詞	用例数
1	だす	2158	30	ころす	133
2	あがる	763	31	あわせる	124
3	つける	713	32	ぬける	108
4	こむ	685	33	つくす	104
5	かける	634	34	はてる	101
6	かえる	628	35	ちらす	99
7	きる	519	36	たつ	97
8	あう	490	37	ぬく	95
9	まわす	478	38	たおす	94
10	はじめる	443	39	かわる	91
11	まわる	430	40	あわす	89
12	つづける	402	41	うける	79
13	あげる	364	42	いれる	73
14	つく	325	43	はなす	68
15	いる	292	44	とばす	52
16	かかる	287	45	とおす	50
17	すぎる	252	46	なれる	48
18	かねる	242	47	わける	48
19	きれる	229	48	おえる	40
20	でる	225	49	わすれる	38
21	うる	208	50	そこなう	30
22	とめる	202	51	こす	27
23	なおす	191	52	なおる	23
24	かえす	183	53	まくる	23
25	とる	182	54	こめる	22
26	おこす	177	55	とまる	4
27	おとす	175	56	くむ	1
28	たてる	159	57	はいる	1
29	おわる	138	58	あわさる	0

各後項動詞ともそれぞれの上位15位までを集計している。

別表2 複合動詞 頻度表

全体 5935種 24073例

総合順位	複合動詞	用例数	割合	累計	総合順位	複合動詞	用例数	割合	累計
各後項動詞上位15位まで×57動詞									
791		13936	=n/13936						
1	いいだす	539	3.9%	3.9%	51	はいりこむ	46	0.3%	42.8%
2	わらいだす	275	2.0%	5.8%	52	なきつづける	46	0.3%	43.1%
3	なきだす	270	1.9%	7.8%	53	よりかかる	45	0.3%	43.4%
4	あるきだす	263	1.9%	9.7%	54	もえあがる	45	0.3%	43.8%
5	いいかける	245	1.8%	11.4%	55	ひきつける	44	0.3%	44.1%
6	みまわす	230	1.7%	13.1%	56	しまいこむ	44	0.3%	44.4%
7	かけだす	230	1.7%	14.7%	57	しすぎる	43	0.3%	44.7%
8	ふりかえる	214	1.5%	16.3%	58	やりきれる	42	0.3%	45.0%
9	おもいだす	177	1.3%	17.5%	59	よびおこす	41	0.3%	45.3%
10	しはじめる	161	1.2%	18.7%	60	ふりまわす	41	0.3%	45.6%
11	かけつける	160	1.1%	19.8%	61	はりつける	41	0.3%	45.9%
12	おそれいる	147	1.1%	20.9%	62	なりだす	41	0.3%	46.2%
13	のぞきこむ	143	1.0%	21.9%	63	とおりすぎる	41	0.3%	46.5%
14	きかえる	135	1.0%	22.9%	64	つかれきる	41	0.3%	46.8%
15	かけこむ	135	1.0%	23.9%	65	かきあげる	41	0.3%	47.1%
16	できあがる	131	0.9%	24.8%	66	おもいかける	40	0.3%	47.3%
17	おきあがる	121	0.9%	25.7%	67	おさえつける	40	0.3%	47.6%
18	かかわりあう	116	0.8%	26.5%	68	あらわれでる	40	0.3%	47.9%
19	いいきる	111	0.8%	27.3%	69	わきだす	39	0.3%	48.2%
20	むかいあう	108	0.8%	28.1%	70	ひっくりかえる	39	0.3%	48.5%
21	いいつける	105	0.8%	28.8%	71	はいあがる	39	0.3%	48.8%
22	しうる	104	0.7%	29.6%	72	かきまわす	39	0.3%	49.0%
23	しだす	93	0.7%	30.2%	73	おちかかる	39	0.3%	49.3%
24	とびあがる	91	0.7%	30.9%	74	わきあがる	38	0.3%	49.6%
25	しきる	86	0.6%	31.5%	75	やりだす	38	0.3%	49.9%
26	みかえる	84	0.6%	32.1%	76	とりだす	38	0.3%	50.1%
27	あるきまわる	84	0.6%	32.7%	77	とおりぬける	38	0.3%	50.4%
28	おちつく	80	0.6%	33.3%	78	よみおわる	37	0.3%	50.7%
29	わかりきる	78	0.6%	33.8%	79	みまわる	37	0.3%	50.9%
30	しかける	78	0.6%	34.4%	80	はいまわる	37	0.3%	51.2%
31	おおせつける	70	0.5%	34.9%	81	いいつづける	37	0.3%	51.5%
32	まきつける	69	0.5%	35.4%	82	まちうける	36	0.3%	51.7%
33	よびとめる	68	0.5%	35.9%	83	ふりあげる	36	0.3%	52.0%
34	みあげる	64	0.5%	36.3%	84	はなしつづける	36	0.3%	52.2%
35	たちあがる	64	0.5%	36.8%	85	つきおとす	36	0.3%	52.5%
36	さしこむ	60	0.4%	37.2%	86	たちまわる	36	0.3%	52.8%
37	うきあがる	60	0.4%	37.7%	87	すわりなおす	36	0.3%	53.0%
38	かけまわる	60	0.4%	38.1%	88	しかねる	36	0.3%	53.3%
39	しつづける	57	0.4%	38.5%	89	さしだす	36	0.3%	53.5%
40	つきあう	56	0.4%	38.9%	90	こらえきれる、たえきれ	36	0.3%	53.8%
41	こらえかねる	55	0.4%	39.3%	91	かんじだす	36	0.3%	54.1%
42	なりかける	51	0.4%	39.7%	92	ありえる	36	0.3%	54.3%
43	かけあがる	51	0.4%	40.0%	93	みあう	35	0.3%	54.6%
44	すがりつく	50	0.4%	40.4%	94	かたりだす	35	0.3%	54.8%
45	ふみかける	49	0.4%	40.7%	95	ひきおこす	34	0.2%	55.1%
46	ききかえす	49	0.4%	41.1%	96	にげまわる	34	0.2%	55.3%
47	ぬけだす	48	0.3%	41.4%	97	きりこむ	34	0.2%	55.5%
48	なでまわす	48	0.3%	41.8%	98	いあわす	33	0.2%	55.8%
49	しみこむ	47	0.3%	42.1%	99	ぬすみとる	32	0.2%	56.0%
50	まきあげる	46	0.3%	42.5%	100	しめあげる	32	0.2%	56.2%

別表3

	共起性	形態素	tスコア	▲MIスコア	コーパス頻度	スパン頻度
1	撒きちらす	撒き	4	14.44	30	16
2	困じはてる	困じ	2.449	13.94	7	6
3	頑張りとおす	頑張り	2.236	13.72	8	5
4	伐りたおす	伐り	2.449	13.48	18	6
5	蹴ちらす	蹴	4.359	13.12	89	19
6	堰きとめる	堰き	3.317	13.04	11	11
7	小突きまわす	小突き	3.316	12.8	13	11
8	見下げはてる	見下げ	2.449	12.74	16	6
9	叱りとばす	叱り	3.464	12.71	88	12
10	ひねくりまわす	ひねくり	3	12.63	12	9
11	引っ掻きまわす	引っ掻	2.449	12.63	8	6
12	喚びおこす	喚び	3.464	12.59	18	12
13	こづきまわす	こづき	2.828	12.58	11	8
14	わめきちらす	わめき	2.828	12.54	56	8
15	なぐりたおす	なぐり	2.236	12.39	32	5
16	しめしあわせる	しめし	3.464	12.38	25	12
17	いじくりまわす	いじくり	3	12.12	17	9
18	遣りそこなう	遣り	2.236	12.01	203	5
19	しかりとばす	しかり	2.236	11.89	65	5
20	跳ねとばす	跳ね	2.828	11.79	111	8
21	解しかねる	解し	3.741	11.72	49	14
22	ゆりおこす	ゆり	2.236	11.59	15	5
23	剃りおとす	剃り	2.828	11.57	25	8
24	しなだれかかる	しなだれ	3	11.48	11	9
25	反りかえる	反り	2.449	11.37	22	6
26	呑みこめる	呑み	2.236	11.31	89	5
27	抛りこむ	抛り	4.795	11.29	34	23
28	惹きおこす	惹き	3.873	11.25	57	15
29	坐りなおす	坐り	5.656	11.23	183	32
30	口説きおとす	口説き	2.236	11.21	20	5
31	うろつきまわる	うろつき	2.449	11.19	9	6
32	いどみかかる	いどみ	2.449	11.18	9	6
33	はまりこむ	はまり	2.235	11.18	8	5
34	覗きこむ	覗き	11.955	11.14	235	143
35	斬りたおす	斬り	3.162	11.13	153	10
36	のたうちまわる	のたうち	2.645	11.12	11	7
37	塗りこめる	塗り	2.236	11.09	104	5
38	散りつくす	散り	2.828	11.09	72	8
39	揺すりあげる	揺すり	2.236	11.07	6	5
40	潜りぬける	潜り	3	11.07	47	9
41	遮りとめる	遮り	2.449	11.04	24	6
42	揺りおこす	揺り	2.828	10.93	38	8
43	まくりあげる	まくり	3.162	10.84	14	10
44	耐えかねる	耐え	3.161	10.81	66	10
45	断りきれぬ	断り	2.828	10.71	50	8
46	訊きかえす	訊き	6.324	10.65	157	40
47	縫いなおす	縫い	2.449	10.65	51	6
48	守りとおす	守り	2.236	10.63	68	5
49	えぐりぬく	えぐり	2.236	10.63	16	5
50	掻きまわす	掻き	6.162	10.62	204	38