慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	ジャンルによって異なる複合動詞の種類： 小規模コーパスと22後項動詞を用いた計量分析
Sub Title	
Author	村田, 年(Murata, Minori)
Publisher	慶應義塾大学日本語・日本文化教育センター
Publication year	2022
Jtitle	日本語と日本語教育 No.50 (2022. 3) ,p.101- 114
JaLC DOI	
Abstract	
Notes	刊行50周年 特集：修了生の現在 〔調査報告〕1
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00189695-20220300-0101

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

慶應義塾大学　日本語・日本文化教育センター紀要
『日本語と日本語教育』50：101–114（2022.3）

ジャンルによって異なる複合動詞の種類

―小規模コーパスと 22 後項動詞を用いた計量分析ー

村　田　　年

1.　はじめに

筆者は、これまで文章ジャンルの判別の可能性について、いくつかの指標を用いて計量分析を行ってきた。指標としては、助詞相当句・接続語句などの機能語句のほか、慣用句[1]と複合動詞の後項動詞を取り上げた（村田 2011、2012 ほか）。このような研究の目的は、実際の言語使用の状況を統計的手法によって検証し、そこからデータを集積して、大学における専門日本語教育のための教材開発の基礎資料とすることである。

本稿では、村田（2012）の複合動詞の後項動詞を指標とした分析結果[2]を前提に、その後に継続して行ってきた BCCWJ（2009 年版）[3]を対象とした後項動詞の個別調査結果（村田 2016）ならびに個々の複合動詞の用法調査（2020 ほか）を踏まえて、文章のジャンルによって高頻度で出現する複合動詞はある程度絞られると想定し、より少ない後項動詞を指標として、改めて文章のジャンルの判別の可能性について検討した。その結果を報告する。

2.　研究方法

2.1　分析のための文章資料

大学・大学院で研究を行うために日本語学習を続ける留学生、とりわけ人文科学分野で研究を志す留学生にとって、日本語の論文読解は論文作成

のために不可欠だと言える。そこで、本分析では論文を一つのジャンルとして立てることにし、比較対照のために新聞社説と文学作品の二つのジャンルを用いた。利用した文章資料は、村田（2012）で用いた小規模の自作コーパスである。これらの文章資料はいずれも筆者が教育現場で中・上級教材として使用したものを中心に収集したものである。村田（2012）では論述文ジャンルを立てたが、本分析では論文ジャンルに限定するため、資料を再編した[4]。文章資料は全部で 6 種類、合計 379 編で、内訳は以下の通りである。

A. 物理学論文：『日本物理学会誌』（1997）第 52 巻 No.1～12 の各号の「最近の研究から」の掲載論文 2 編ずつ合計 24 編（除　数式・記号）。総文数 2243 文。
B. 工学論文：学術雑誌論文 14 編。専攻分野別に電気工学 6 編、機械工学 4 編、計算機科学 2 編、管理工学 2 編（除　数式・記号）。総文数 1725 文。
C. 文学論文：学術雑誌『中古文学』『中世文学』『近世文藝』『日本近代文学』の四誌各 10 冊の掲載論文から単純無作為抽出による各 6 編ずつ計 24 編（除引用部分）。総文数 4184 文。
D. 経済学論文：『経済研究』一橋大学経済研究所編（2000）VOL.51 No.1～4 の全論文 19 編（除　数式）。総文数 3889 文。
E. 新聞社説：1996 年 12 月 1 日～31 日までの四大紙の朝刊、夕刊の社説（読売新聞 58 編、朝日新聞 55 編、毎日新聞 58 編、日本経済新聞 51 編）の総計 222 編。総文数 6514 文。
F. 文学作品：近代文学作品：12 名の作家の作品計 26 編、総文数 2952 文。
　　現代文学作品：星新一編と泡坂妻夫の短編計 50 編。2403 文。

2.2　指標の選択

姫野（1999）では、国立国語研究所の『複合動詞資料集』（1987）から、異なり語数の多い順に上位 30 位までの後項動詞が挙げられている。姫野は、「数量的には確定的ではないが、全体の傾向は、これでつかめるものと思われる。」と述べ、さらにそれらを外国人児童生徒のための教科書の調査（ぎょうせい 1998）における後項動詞の上位語と比較した上で、そこ

に「見事に重なりが見られる。」と述べて、使用頻度の高い後項動詞はある程度限定されるということを指摘している（姫野 1999, 25–26）。本稿では、姫野（1999）で取り上げられた後項動詞のうち、造語力が強く、重層的に複数の意味を持つとされる 22 の後項動詞を指標として用いた[5]。具体的には、姫野（1999）の第 3 章以下で見出し語となっている 19 の後項動詞に加え、「こむ」の下位見出し語となっている「こめる」、「いる」、「いれる」を含めた合計 22 の後項動詞である。「あう」と「あわせる」の下位項目となっている「あわさる」については、姫野も述べるように、複合動詞の数が少なく、ほとんどの場合に「あわせる」の受身形で代用される可能性が指摘されているため、指標には加えていない。また、姫野（1999）では造語力は強いが、扱われていない十数語があり、その中には、アスペクトを表す「始める」、「続ける」、「終わる／終える」、可能／不可能を表す「得る」、「かねる」、過剰を表す「すぎる」が含まれている。そのため、これらの語も本分析の指標には含まれていない。指標は以下の 22 の後項動詞である。

> あう、あがる、あげる、あわせる、いる、いれる、かかる、かける、きる、こむ、こめる、だす、たつ、たてる、つく、つける、でる、とおす、なおす、なおる、ぬく、まくる

ここで参考までに、村田（2012）の分析[6]で選択された有効な指標と比較したい。村田（2012）では、判別分析に資する後項動詞を探すという目的があり、できるだけ広い範囲から指標を選んだ結果、指標としての後項動詞は 41 語[7]に上った。それらを用いた判別分析の結果、交差検定による正判別率 63.5％で、三つのジャル（論述文、文学作品、新聞社説）の判別が可能なことがわかった。判別に有効な指標として以下の 15 後項動詞が選択されている。

① だす②かける③あがる④かねる⑤いる⑥おわる⑦なおす⑧まくる⑨いれる⑩こむ⑪きる⑫つくす⑬たつ⑭つづける⑮あげる

ここには、姫野（1999）では見出し語とされていない「かねる」、「おわる」、「つづける」の 3 語も含まれているが、それらを外すと、重複していない指標は「つくす」1 語となる。後項動詞「つくす」は、極限までの完了という程度の強さを意味するが、「尽くす」という語自体は複数の意味を持つことはなく、接続する前項動詞も限定的で、造語力は低い語だと言えよう。

後項動詞については、計量分析の指標としては有効でも、複合動詞形成時の意味の理解にさほど困難がなければ、指導の観点からは、重要度ならびに優先度は低いと考えられる。当然ながら、計量的に有効な指標がそのまま日本語教育の指導上、重要な後項動詞となるわけではない。

3. 分析方法

3.1 単変量

個々の変量である 22 後項動詞のジャンルごとの分布の違いを見るため、Kruskal–Wallis 検定（以下、KW 検定）を行って、ジャンルごとに比較対照した。

3.2 多変量

2.1 に挙げた 6 種類の文章資料（A～F）の合計 379 編を以下のように三つのジャンル（論文、新聞社説、文学作品）に分類した。

① 論文ジャンル：A～D ② 新聞社説ジャンル：E ③ 文学作品ジャンル：F

次に、22 の後項動詞の各資料における出現率（各資料における出現回数を一文当たりの出現頻度に換算した相対出現頻度）[8] を求めた。なお、後項動詞の記述に用いられた用字の差異（例：あう／合う／会う）ならびに動詞の活用変化の形は同定して同じ後項動詞として扱った。2.2 で挙げた

22 指標からジャンル判別に有効な指標を選択するため、多変量解析の分類手法の一つである正準判別分析のステップワイズ法を用いて、判別に寄与する後項動詞を選択した[9]。

4. 結果と考察

対象とする文章資料に現れた個別の複合動詞については、既に村田（2008、2012）で全数調査を行っている[10]。生のデータに含まれる情報の概要を把握するため、三つのジャンルごとの 22 指標の各平均出現率を計算し、グラフ化したものを図 1 に示す。図 1 からは各ジャンルによって多用される特徴的な後項動詞が存在することがわかる。論文ジャンルで他のジャンルより多用が目立つのは、「あげる」、「たつ」で、文学作品ジャンルでは「だす」、「あがる」、「かける」、「つける」である。新聞社説ジャンルでは「こむ」、「きる」、「なおす」、「いれる」が目を引くと言えよう。

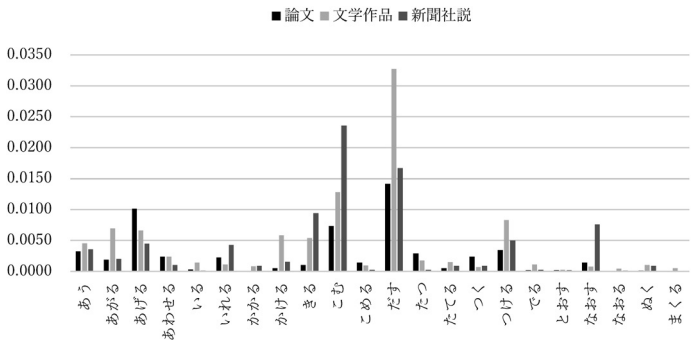

図 1　ジャンルごとの 22 後項動詞の平均出現率

次に単変量と多変量で分析を行う。

4.1　単変量分析結果

KW 検定の結果、22 指標のうち 15 指標が有意であった。その 15 の後項動詞（$p < 0.05$）のジャンル別の平均ランクをまとめたものを表 1 に示す。

平均ランクの違いから、各後項動詞のジャンルごとの出現傾向の差異が見て取れる。論文ジャンルでは、特に「あげる」が際立って高く、それに「たつ」が続く。文学ジャンルでは、特に「だす」が際立ち、「かける」が続く。新聞社説ジャンルはいずれも際立って高い語はなく、中では「こむ」が相対的に高いと言える。

表 1　KW 検定結果

指標	後項動詞	検定統計量	P 値	論文	文学作品	新聞社説
				平均	平均	平均
1	あう	12.035	0.00	206.68	202.95	179.48
2	あがる	22.641	0.00	195.53	217.23	178.66
3	あげる	60.074	0.00	247.13	198.24	166.33
4	あわせる	23.370	0.00	211.78	198.09	179.29
5	いる	12.641	0.00	191.93	200.01	185.87
6	かかる	11.861	0.00	184.00	201.22	188.35
7	かける	32.406	0.00	183.58	222.07	181.36
8	きる	10.087	0.01	167.10	188.72	198.79
9	こむ	13.080	0.00	160.44	178.53	204.71
10	だす	28.843	0.00	184.07	245.20	173.27
11	たつ	36.043	0.00	215.57	191.14	180.28
12	つく	16.308	0.00	207.21	186.89	184.79
13	つける	11.262	0.00	186.86	214.84	182.64
14	でる	10.281	0.01	190.64	198.46	186.87
15	まくる	12.024	0.00	188.50	195.98	188.50

$p < 0.05$

単変量分析の結果は、使用傾向を直接的に捉えることができるので、教育現場では有用な情報になると考えられる。便宜的な基準として、「平均値＋ 0.5SD 相当値」を計算すると 244.7 となる。有意な語のうち、論文ジャンルでは平均ランクが 247.13 の「あげる」、文学作品ジャンルでは 245.20 の「だす」が当該ジャンルでもとりわけ特徴的な語であることがわかる。

ここで、上記の単変量分析の結果をもとに、ジャンルによってどのような複合動詞が多用されているのかを具体的に見てみたい。論文ジャンルと文学作品ジャンルについては、有意な後項動詞のうち上位 2 語（平均ランク 222.00 以上）が形成する複合動詞を、新聞社説については平均ランクが際立って高い語がないので、相対的に高い「こむ（平均ランク 204.71）」について見る。読売新聞を除く対象資料における複合動詞の出現回数は、すでに村田（2008）の表 4「文章資料別複合動詞リスト」で報告しているため、今回、新たに読売新聞社説の資料の調査結果も含めたリストをもとに、出現している複合動詞を確認した（各ジャンルに属す複数資料に共通して出現した語を抜粋）。

1　論文
　あげる：押し上げる、取り上げる。引き上げる、作り上げる、積み上げる
　たつ　：成り立つ
2　文学作品
　だす　：歩き出す、追い出す、送り出す、思い出す、差し出す、飛び出す、取り出す、流れ出す、泣き出す、逃げ出す、抜け出す、掘り出す、見つけ出す、持ち出す、呼び出す
　かける：話しかける、出かける、呼びかける
3　新聞社説
　こむ　：盛り込む、抑え込む、巻き込む、持ち込む、見込む、絞り込む

本分析に用いた文章資料は限定的であるため、一般化はできないが、こ

れらの複合動詞は、指導時の重要度・優先度を決める際にヒントにはなると考えらえる。具体的に個別の複合動詞を見て気づくことは、例えば、「成り立つ」、「思い出す」、「出かける」のように既に一語化している語も、検索結果に含まれてくることである。ちなみに、後項動詞「つける」の用法を詳しく調べた村田（2020）では、どのジャンルでも圧倒的な高頻度語は「見つける」であった。日本語教育の現場では、「見つける」、「出かける」、「思い出す」は、初級レベルで導入され、通常、一語として扱われる。複合動詞の指導の際に、前項動詞と後項動詞に分け、わざわざ語の成り立ちを説明するのは、学習者が後項動詞の意味や造語力の強さを知ることによって、同じ後項動詞を持つ他の未習語についても、その意味を類推できるようにするためである。このように考えると、上記に挙げた語のように、既に一語化している語は、あえて分けて教える必要がない語、むしろ分けない方がいい語だと言えよう。今後、分析の際にそのような語の扱いをどうするか、検討する必要があると考えられる。この点は今後の課題としたい。

4.2　多変量分析結果

次に、判別分析を行い、三つのジャンルを分離するのに有効な指標の有無を検討する。2.1 の文章資料合計 379 編を対象に、22 の後項動詞を指標としてステップワイズ法による判別分析を行った。その結果、三つのジャンルの判別に有効な指標として以下の 11 の後項動詞が選択された。

⑩あがる②あげる③いる④いれる⑤かける⑥きる⑦こむ⑧だす⑨たつ
⑩なおす⑪まくる

ここで選択された 11 後項動詞は、村田（2012）で有効な指標として選択された指標から、「はじめる」、「おわる」、「つづける」、「つくす」の 4 語を除いた指標と同じである。つまり、指標として新たに 22 後項動詞を選び

直し、対象とする文章資料も論文ジャンルに特化したが、三つのジャンル（論文、新聞社説、文学作品）を判別するのに有効な指標は、村田（2012）で選択された指標とほぼ変わらないという結果となった。分析結果の詳細を見ていく。

本分析では文章資料のジャンルは三つなので、表 2 のように二つの判別関数が算出された。表 3 は各ジャンルの判別空間における重心である。選択された 11 後項動詞による判別関数平面での各文章資料の判別得点とジャンルの重心をプロットしたものが図 2 である。これらの結果から、関数 1 が文学作品ジャンルを他の二つのジャンルから大きく分離し、関数 2 が論文ジャンルを新聞社説ジャンルから分離していることがわかる。表 4 の 11 指標の構造係数を見ると、関数 1 では「あがる」「なおす」「いる」「まくる」「いれる」が、文学作品を論文、新聞社説から分離するのに有効となっていて、関数 2 では「だす」「こむ」「きる」「かける」「たつ」「あげる」が、論文を新聞社説から分離している。本分析の判別の可否を評価するために、交差検定を行った。そのクロス集計表を表 5 に示す。交差検定結果から、正判別率は 62.8％となり、選択された 11 の後項動詞によって三つのジャンル（論文、新聞社説、文学作品）の文章資料が判別可能であることが検証された。この結果は、先に挙げた村田（2012）の正判別率 63.5％より若干低いが、ほぼ同様の値となっている。

次に個別のジャンルに特徴的な後項動詞を見ていきたい。

文学作品ジャンルは、「あがる」、「なおす」、「いる」、「まくる」、「いれる」の 5 指標によって、論文ならびに新聞社説ジャンルから 71.3％の寄与率で分離されている。複合動詞の使用は三つのジャンルの中で一番多いことがわかる。文学作品ジャンルで特徴的に高い頻度で用いられている語は「あがる」、「いる」、「まくる」で、逆に頻度が低いのは「なおす」と「いれる」である。具体的な複合語としては、以下のような語が挙げられる（文学作品 2 資料に共通して出現した語を抜粋）。

表 2　判別関数の固有値等

判別関数	固有値	寄与率	p 値	Λ	χ 2 乗
関数 1	0.373	71.3	0.00	0.633	169.541
関数 2	0.150	28.7	0.00	0.870	51.879

表 3　判別空間における各ジャンルの重心

ジャンル	関数 1	関数 2
論文	0.268	−0.720
文学作品	1.069	0.366
新聞社説	−0.464	0.137

表 4　選択された 11 後項動詞の構造係数

後項動詞		関数 1	関数 2
1	あがる	*0.304	0.260
2	なおす	*−0.295	0.172
3	いる	*0.268	0.136
4	まくる	*0.246	0.209
5	いれる	*−0.236	0.074
6	だす	0.352	*0.419
7	こむ	−0.313	*0.414
8	きる	−0.237	*0.408
9	かける	0.296	*0.388
10	たつ	0.229	*−0.381
11	あげる	0.143	*−0.336

*有意な係数

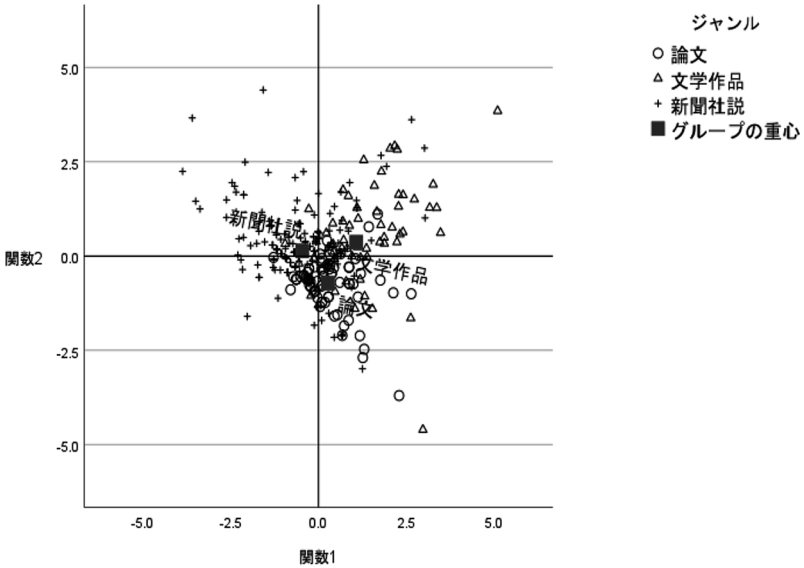

図 2　文章資料とジャンルごとの重心のプロット

表 5　判別関数による予測グループと実際のジャンルのクロス集計（交差検定後）

ジャンル	判別分析に基づく予測グループ			
	論文	文学作品	新聞社説	合計
論文	13	7	61	81
文学作品	6	30	40	76
新聞社説	14	13	195	222
合計	33	50	296	379

文学作品
あがる：起き上がる、立ち上がる、でき上がる、飛び上がる、舞い上がる
いる　：寝入る、踏み入る、見入る、聞き入る、染み入る
まくる：追いまくる、撃ちまくる、うんざりしまくる

論文ジャンルと新聞社説ジャンルを分離するのは、「だす」、「こむ」、「きる」、「かける」、「たつ」、「あげる」の 6 指標である。論文ジャンルで特徴的な語は「あげる」と「たつ」である。新聞社説ジャンルでは「だす」、「こむ」、「きる」、「かける」である。ちなみに、この 6 指標は単変量分析の結果でも有意である。

論文ジャンルの具体的な複合動詞は、すでに 4.1 で見た通りである（4.1 参照）。

新聞社説ジャンルの具体的な複合動詞は、4.1 で見た「こむ」の複合動詞以外に、以下のような語が挙げられる（新聞社説 4 資料に共通して出現した語を抜粋）。

新聞社説
だす　：打ち出す、引き出す、噴き出す、踏み出す、持ち出す、乗り出す、飛び出す、生み出す、見い出す、作り出す、抜け出す
きる　：打ち切る、踏み切る、断ち切る、押し切る、乗り切る、思い切る
かける：呼びかける、投げかける、持ちかける、働きかける

以上、単変量と多変量を用いて、論文、文学作品、新聞社説の三つのジャンルの分析を行った。小規模なコーパスを対象とした分析であり、結果を安易に一般化すべきではないが、判別分析の結果、有効な 11 の後項動詞が選択でき、これらの後項動詞がジャンルによって出現頻度に特徴的な差を持つことが検証された。また、具体的に有効な指標となった後項動詞を含む複合動詞にはどのような語があるかをジャンル別に確認することができた。同じ後項動詞「だす」を持つ複合動詞が使用されていても、文

学作品ジャンルと新聞社説ジャンルの語を比較するとわかるように、使用頻度の高い語の種類は大きく異なる。このような知見は教育現場では有用な情報となろう。

5. おわりに

対象資料は限定されたものではあるが、計量分析を用いることによって、専門日本語教育で重要な論文ジャンルについて、多用される複合動詞の存在を確認し、それをある程度特定することができた。また同時に、対照的なジャンルである文学作品ならびに新聞社説についても、特徴的な複合動詞を抽出できた。今後は、大規模コーパスを対象とした更なる検証を進める必要があると考えられる。また、これらの知見を生かすための教材の開発にも取り組んでいきたい。

注

1 村田・山崎（2011）に詳しい。
2 村田（2012）では紙幅の都合で、判別分析結果のみを報告し、詳しい考察は報告していない。
3 山崎 (2009) を参照されたい。
4 村田（2012）で用いた資料から、専門分野の入門書である経済学入門教科書の資料を除いた。
5 この 22 指標は、村田（2013）以降、BCCWJ を対象とした複合動詞調査で用いているが、統計解析で用いるのは初めてである。
6 村田（2012）では解析用プログラムとして SPSS Statistics17 を用いた。分析方法については柳井・高木編『多変量解析ハンドブック』に詳しい。
7 41 指標（後項動詞）：あう、あがる、あきる、あぐねる、あげる、あやまる、あわせる、いる、いれる、おえる、おくれる、おわる、かかる、かける、かねる、きる、こむ、こめる、すぎる、そこなう、そこねる、そびれる、そんじる、だす、たつ、たてる、つくす、つく、つける、つづける、でる、とおす、なおす、なおる、なれる、ぬく、のこす、はじめる、はてる、まくる、わすれる
8 村田（2012）のデータを利用した。
9 解析用プログラムとして、SPSS Statistic27 を用いた。
10 村田（2008）では、本分析に利用した文章資料（但し、2008 年時点では読売新聞の資料は未収集につき、含まれていない）と 22 指標を含む 26 の後項動詞の出現回数について全数調査を報告し、個別の複合動詞について考察を行った。

参考文献

柳井晴夫・高木廣文編（1986）『多変量解析ハンドブック』現代数学社
野村雅昭・石井正彦（1987）『複合動詞資料集』国立国語研究所報告』
村上征勝・金明哲共著（1998）『講座　人文科学研究のための情報処理　第 5 巻　数量的分析編』尚学社
ぎょうせい［26］（1998）『外国人児童生徒のための日本語指導　第 2 分冊　算数（数学）・理科の教科書―語彙と漢字―』東京外国語大学留学生センター
姫野昌子（1999）『複合動詞の構造と意味用法』ひつじ書房
村田　年（2008）「文章と複合動詞―論述的な文章ジャンルを特徴づける新たな指標を探して―」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター 36, 1–33.
山崎　誠 (2009)「代表性を有する現代日本語書籍コーパスの構築」『人口知能学会』 24(5), 623–631.
村田　年・山崎　誠（2011）「「 手 」の慣用句を指標とした文章のジャンル判別―現代日本語書き言葉均衡コーパスを用いて」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター 39, 75–88.
村田　年（2012）「文章のジャンル判別に寄与する指標の研究―専門日本語教育への応用―」『コーパスとテキストマイニング』共立出版，166–180.
村田　年（2016）「BCCWJ を用いた複合動詞使用頻度調査表の改訂―22 後項動詞を指標として―」『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター 44, 115–131.
村田　年（2020）「BCCWJ に現れた複合動詞「押しつける」」―自然科学系（含技術・工学）ジャンルと社会学ジャンル―『日本語と日本語教育』慶應義塾大学日本語・日本文化教育センター 48, 31–54.