

BCCWJにおける複合動詞後項の表記の実態

小椋秀樹 (立命館大学文学部)

Orthographic Variation of the Letter Form of Compound Verbs as Seen in the BCCWJ

Hideki Ogura (College of Letters, Ritsumeikan University)

要旨

本稿の目的は、統語的複合動詞等の表記のうち、特に後項動詞の表記に着目し、そのゆれの実態を明らかにすることにある。

具体的には、『現代日本語書き言葉均衡コーパス』の図書館サブコーパス・書籍を調査資料とし、影山(1993)、姫野(1999)で統語的複合動詞を構成する後項動詞として挙げられている動詞を中心に30語を取り上げた。調査では、後項動詞の度数、出版年別、著者の生年代別といった観点から、漢字表記と平仮名表記の割合を明らかにした。

1. はじめに

現代日本語における語表記のゆれについては、小椋(2012)で、『現代日本語書き言葉均衡コーパス』(以下、BCCWJとする。)のコアデータ⁽¹⁾を資料として実態調査を行った。小椋(2012)では、和語に語表記のゆれが多く見られること、和語の中でも動詞が語表記のゆれの割合の高い語群であることを明らかにした。さらに動詞を対象に、どのようなゆれの類型があるのかも調査し、漢字と平仮名の対立によるゆれ(例:「合うーあう」)が最も多いことを指摘した。漢字と平仮名の対立によるゆれは、語表記にゆれのある単純動詞の約7割に、同じく複合動詞の約8割に見られる。複合動詞では、「とりくむー取り組む」「くりかえすーくり返すー繰り返す」などのように、語全体を平仮名表記にしたものや、前項動詞又は後項動詞を平仮名表記にしたものがあり、高い割合になっているものと考えられる。

小椋(2012)は、短単位を用いた調査であるため、上に述べた複合動詞とは語彙的複合動詞を指す。短単位では、最小単位二つから成る語彙的複合動詞は「／食べ歩く／」「／走り回る／」のように1短単位とするが、統語的複合動詞は「／降り／出す／」「／調べ／尽くす／」のように前項動詞、後項動詞をそれぞれ1短単位とする⁽²⁾。そのため小椋(2012)では、統語的複合動詞を構成する前項動詞、後項動詞を単純動詞として集計している。統語的複合動詞についても、語彙的複合動詞と同様に、漢字と平仮名の対立による語表記のゆれが多く見られるのか実態を調査する必要がある。

ところで、統語的複合動詞の表記については、表記の基準の面で興味深い点がある。それは、統語的複合動詞の後項動詞について、常用漢字で表記できても、一般的な表記として平仮名表記を挙げるものが見られることである。例えば、白石、野元、高田(2010)では、

(1) BCCWJ の設計等については、前川(2008)、山崎(2011)を参照。

(2) 短単位の認定基準のうち和語の動詞については、小椋、小磯、富士池(2011:29)を参照。

次のような例がある。

かねる **かねる**、兼ねる [例] 申し上げかねる、母の帰りを待ちかねる
 すぎる ……**すぎる**、……過ぎる [例] 考えすぎる、食べすぎる、飲みすぎる、
 気にしすぎる
 だす ……**だす**、……出す [例] 雪が降りだす、勉強をやりだす、笑いだす

白石、野元、高田(2010:2)は、語の書き表し方について、「常用漢字・現代仮名遣い・送り仮名の付け方を原則とする、一般的と考えられる書き表し方を、ゴシック体で示し」ている。上の例では、常用漢字「兼」「過」「出」による表記ではなく、平仮名表記を「一般的と考えられる書き表し方」として示していることになる。このような基準の妥当性を検証するという意味でも、統語的複合動詞の表記の実態、特に後項動詞を漢字表記にするか、仮名表記にするかについて調査を行う必要がある。

なお近年、情報機器の普及に伴う漢字の多用化傾向ということがよく言われる。語表記のゆれに経年変化が見られるのかといったことを調査することも重要である。

以上のことから、本稿では、BCCWJ を資料として、複合動詞のうち統語的複合動詞を取り上げ、特にその後項動詞に着目して、計量的な面から表記の実態を調査することとする。なお本稿では、小椋(2012)の結果を踏まえ、漢字と平仮名の対立によるゆれを取り上げることとする。したがって、送り仮名の対立によるゆれや仮名遣いの対立によるゆれなどは、取り上げない。

以下、2 節で今回の調査対象とする統語的複合動詞の後項動詞の範囲、調査資料とするレジスター、調査方法について述べた後、3 節で調査結果を報告する。最後に 4 節で本稿をまとめる。なお本稿では、語の表記を示す際には「出す」「だす」のように鍵括弧を付け、語を示す際には《ダス》のように二重山括弧を付ける。

2. 調査対象・資料・方法

2. 1 調査対象

語彙的複合動詞、統語的複合動詞という区分は、影山(1993)に述べられたものである。統語的複合動詞の特徴の一つとして、後項動詞が限定されるということが挙げられる。統語的複合動詞の後項動詞として、影山(1993:96)では、《カケル》《ダス》《ハジメル》など 27 語を挙げており、姫野(1999:19)では《カカル》《ハテル》《ソコネル》の 3 語を加えている。

短単位では、原則として、統語的複合動詞の後項動詞を付属要素(接尾的要素)として扱うこととし、「/降り/出す/」「/調べ/尽くす/」のように前項動詞と結合させずに単独で 1 短単位としている⁽³⁾。

ただし例外もあり、影山(1993)、姫野(1999)で示された 30 語全てを付属要素(接尾的要素)としているわけではない。BCCWJ における出現状況から造語力が高いと判定しなかったもの、単位認定が難しいと判断したものについては、付属要素(接尾的要素)として扱わなかった。これに該当するのは、《アウ》《アキル》《カケル》《カカル》《ノコス》《アヤマル》《ナオス》の 7 語である。

一方、影山(1993)、姫野(1999)で統語的複合動詞の後項動詞として挙げられていない語

(3) 付属要素、及び短単位の認定における付属要素の扱いについては、小椋、小磯、富士池(2011:33、(38)-(55))を参照。

であっても、BCCWJ における出現状況、単位の統一性などから付属要素(接尾的要素)に加えたものもある。例えば、《サス》《ハテル》など6語である。

以上のようなことから、本稿で調査対象とする語は、影山(1993)、姫野(1999)と若干異なるところがある。本稿で調査対象としたのは、以下の30語である。

A: 影山(1993)、姫野(1999)に挙げられている語

アグネル(あぐねる)	エル(得る)	オエル(終える)	オクレル(遅れる)
オワル(終わる)	カネル(兼ねる)	カワス(交わす)	キル(切る)
コナス(熟す)	スギル(過ぎる)	ソコナウ(損なう)	ソコネル(損ねる)
ソビレル(そびれる)	ソズル(損ずる)	ダス(出す)	ツクス(尽くす)
ツケル(付ける)	ツツケル(続ける)	トオス(通す)	ナレル(慣れる)
ヌク(抜く)	ハジメル(始める)	マクル(捲る)	ワスレル(忘れる)

B: 影山(1993)、姫野(1999)に挙げられていない語

オオセル(果せる) サス(止す) ツツク(続く) ハタス(果たす)
ハテル(果てる) ワタル(渡る)

本稿では、影山(1993)、姫野(1999)で統語的複合動詞の後項動詞として挙げられていない語も合わせて調査対象とすることから、以下、上記30語を後項動詞とする複合動詞を、統語的複合動詞等と呼ぶ。

2. 2 資料

本稿では、統語的複合動詞等の後項動詞の表記について、経年変化の有無も調査する。

BCCWJ で経年変化を見ることのできるレジスターは、図書館・書籍(1986-2005年)、特定目的・ベストセラー(1971-2005年)、特定目的・白書、特定目的・国会会議録(以上、1976-2005年)である。このうち、特定目的・白書、特定目的・国会会議録は、公文書であり、公文書の表記の基準に基づいて書かれている。そのため、一般社会における表記とずれが見られる可能性がある。また経年変化が見られたとしても、基準の改定によるものという可能性もある。したがって、今回の調査に適したレジスターとは言えない。特定目的・ベストセラーは、多くの人々に読まれたものとして、重要なレジスターではあるが、延べ語数が約370万語と、データ規模の面で問題がある。

以上のようなことから、本稿では、資料として図書館・書籍を用いることとした。図書館・書籍は、全体で延べ約3000万語と大規模なデータである。5年ごとに区切った場合、最も語数の少ない1986-1990年で延べ約480万語、最も語数の多い2001-2005年で延べ約880万語と、経年変化を見るのにも十分な規模のデータである。

2. 3 方法

用例の収集に当たっては、短単位データ1.0.0を対象に、『中納言』1.1.0により、次の検索条件式で検索した(例として、《ハテル》の検索条件式を示す。)

```
キー: 語彙素 = "果てる" AND 前方共起: 品詞 LIKE "動詞%" ON 1 WORDS FROM キー IN (registerName="図書館・書籍" AND core="false") WITH OPTIONS unit="1" AND tglWords="20" AND limitToSelfSentence="0" AND endOfLine="CRLF" AND tglKugiri="|" AND encoding="UTF-8" AND tglFixVariable="2"
```

本稿は、漢字と平仮名の対立による語表記のゆれの実態を把握することを目的としてい

る。その際、注意しなければならないのは、児童向けの書籍の用例である。当然のことながら、児童向けの書籍の表記には、平仮名が多用されており、児童向けの書籍に多く使われている語は、平仮名表記の割合が高くなる可能性がある。本稿では社会一般の表記の実態を把握したいと考えており、児童向けの書籍の用例は除外するのが望ましい。

そこで、児童向けの書籍の用例を除外するために、BCCWJ の DVD に格納されている書誌情報データベースを利用した⁽⁴⁾。書誌情報データベースでは、書籍のジャンル情報の一つとして C コード(図書分類コード)が記載されている。C コードは 4 桁の数値で、左から 1 桁目が対象読者を示す「販売対象コード」である。児童向けの書籍は、この 1 桁目が「8」となっている。この C コードの情報を使って児童向けの書籍の用例を除外した。

具体的には、まず各語の検索結果を統合した後、関係データベースにインポートした。次にサンプル ID ベース書誌情報データ(Joinud_info.txt)をインポートし、サンプル ID で検索結果と関係付けて、検索結果に C コードの情報を付与し、児童向けの書籍の用例を除外できるようにした。

3. 調査結果

3. 1 後項動詞の表記のゆれ

本節では、後項動詞の表記の実態について、漢字表記、平仮名表記がそれぞれどの程度用いられているのか見ていくこととする。なお、先にも述べたように図書館・書籍は、出版年で 20 年の幅を持つレジスターであるが、ここでは一つの共時態と見なして(年代幅を無視して)調査する。

30 語の後項動詞について、漢字表記、平仮名表記がそれぞれどの程度用いられているかを、表 1 にまとめた。表 1 では、漢字表記、平仮名表記の度数と、それぞれの表記が語の度数全体に占める割合とを示した(「漢字」「平仮名」の各欄)。

表 1 : 後項動詞の表記

	漢字		平仮名		種別		漢字		平仮名		種別
	度数	割合	度数	割合			度数	割合	度数	割合	
アグネル	0	0.0%	40	100.0%	固定	ソズル	23	100.0%	0	0.0%	固定
エル	1769	37.4%	2961	62.6%	ゆれ	ダス	1660	54.1%	1410	45.9%	ゆれ
オエル	404	87.4%	58	12.6%	独占	ツクス	365	54.4%	306	45.6%	ゆれ
オオセル	0	0.0%	47	100.0%	固定	ツケル	0	0.0%	13	100.0%	固定
オクレル	96	85.0%	17	15.0%	独占	ツヅク	55	65.5%	29	34.5%	ゆれ
オワル	568	89.6%	66	10.4%	独占	ツヅケル	2925	60.6%	1899	39.4%	ゆれ
カネル	36	2.9%	1207	97.1%	独占	トオス	168	73.0%	62	27.0%	ゆれ
カワス	73	64.6%	40	35.4%	ゆれ	ナレル	297	83.7%	58	16.3%	独占
キル	778	27.4%	2065	72.6%	ゆれ	ヌク	272	69.7%	118	30.3%	ゆれ
コナス	1	0.6%	180	99.4%	独占	ハジメル	3352	48.1%	3615	51.9%	ゆれ
サス	0	0.0%	12	100.0%	固定	ハタス	76	84.4%	14	15.6%	独占
スギル	463	23.0%	1552	77.0%	ゆれ	ハテル	301	71.2%	122	28.8%	ゆれ
ソコナウ	36	36.0%	64	64.0%	ゆれ	マクル	4	1.0%	380	99.0%	独占
ソコネル	21	44.7%	26	55.3%	ゆれ	ワスレル	150	98.0%	3	2.0%	独占
ソビレル	0	0.0%	39	100.0%	固定	ワタル	272	47.1%	306	52.9%	ゆれ

(4) 書誌情報データベースについては、丸山、中村(2011)を参照。

表1を見ると、《アグネル》《オオセル》など語表記にゆれの見られない語が6語あるが、それ以外は全て漢字と平仮名の対立による語表記のゆれが見られる。

ゆれの見られる24語を見ると、漢字表記、平仮名表記のいずれかに集中しているものがある。例えば、《カネル》《コナス》は平仮名表記が9割を超えている。そこで、ゆれの程度に応じた分類を試みることにする。まず、ゆれの見られない語を「固定」、一方の表記が8割以上を占めている語を「独占」、それ以外を「ゆれ」と呼ぶこととし、表1の「種別」欄に記載した⁵⁾。

「固定」「独占」「ゆれ」について、語の度数との関連を見ていくことにする。表2は、語の度数別に、「固定」「独占」「ゆれ」がどの程度、出現するかをまとめたものである。この表では、度数を100以下、101-200というふうに六つに区分し、それぞれの区分における「固定」「独占」「ゆれ」の語数を示した。

表2: 「固定」「独占」「ゆれ」と語の度数

	100以下	101-200	201-300	301-400	401-500	501以上
固定	6	0	0	0	0	0
独占	1	3	0	2	1	2
ゆれ	2	2	1	1	1	8

表2を見ると、「固定」は、度数100以下の低頻度層にのみ見られることが分かる。《アグネル》《オオセル》《サス》《ソビエル》《ツケル》は平仮名表記のみが、《ソズル》は漢字表記のみが用いられている。《アグネル》《オオセル》《サス》《ソビエル》で平仮名表記のみが用いられるのは、これらの語が常用漢字(又は表内訓)で表記できないことによると考えられる。《ソズル》で漢字表記のみが用いられるのは、この語が漢語サ変動詞だからであろう。ただ表1から分かるように、「固定」に属する語は度数50以下であり、BCCWJでは度数が低いため、たまたま一方の表記しか出現しなかった可能性もある。

「独占」は、度数100以下の低頻度層、度数501以上の高頻度層にも見られるが、度数101-500の中頻度層に6語あり、中頻度層を中心に分布している。「ゆれ」も低頻度層から高頻度層まで出現しているが、度数501以上が8語と最も多く、高頻度層を中心に分布している。よく使われる語ほど、表記にゆれが生じるということができよう。これは「固定」が低頻度層にのみ見られることと、ちょうど逆のことといえる。なお宮島(1997:100)でも、「度数順にみると、当然度数1のものにはゆれがなく、度数のたかいものほど、これがおおい」述べられている。本稿は、統語的複合動詞等の後項動詞に限定した調査ではあるが、宮島(1997)と同様の傾向が確認できた。

「独占」に属する語のうち、漢字表記に集中する語は《オエル》《オクレル》《オワル》《ナレル》《ハタス》《ワスレル》の6語、平仮名表記に集中する語は《カネル》《コナス》《マクル》3語で、漢字表記に集中する語の方が多い。

また「ゆれ」に属する語のうち、漢字表記の割合が高い語は《カラス》《ダス》《ツク

(5) この3区分は、1956年発行の雑誌90種を対象に、語表記のゆれを調査した宮島(1997)を参考にしたものである。ただし宮島(1997)は、「独占」を「特定の形式が9割以上をしめているもの」(p.103)としており、本稿と異なる。

ス》《ツヅク》《ツヅケル》《トオス》《ヌク》《ハテル》の 8 語、平仮名表記の割合が高い語は《エル》《キル》《スギル》《ソコナウ》《ソコネル》《ハジメル》《ワタル》の 7 語である。平仮名表記の割合が高い語のうち、《エル》《キル》《スギル》《ハジメル》《ワタル》の 5 語は、度数 500 以上の高頻度語である。ゆれの見られる高頻度の語は、平仮名で表記される傾向があるといえよう。

3. 2 後項動詞の表記の変化

次に、後項動詞の表記の経年変化について見ていくこととする。ここでは、漢字表記、平仮名表記の割合に変化があるのか見ていく。

改定常用漢字表(2010 年 2 月、文化審議会答申)において、「情報機器による漢字使用が一般化し、社会生活で目にする漢字の量が確実に増えていると認められる」(p.(3))と述べられているように、近年、情報機器の普及という書記環境の変化に伴って、漢字が多用される傾向にあるといわれる。図書館・書籍は、1981 年の常用漢字表(内閣告示第 1 号・同訓令第 1 号)の実施から 2010 年の改定までの 29 年間のうち、20 年をカバーするレジスターである。本稿は、統語的複合動詞の後項動詞に限定した表記の実態調査ではあるが、情報機器が一般化し、書記環境が変化していく中で、表記がどのように変化していったのか(あるいは、しなかったのか)を明らかにするための調査としても位置付けられる。

まずは、出版年代別の経年変化を見ていく。図 1 に、出版年代別に後項動詞の漢字表記、仮名表記の割合を示した。この図では、出版年が 1986-1989 年のものを 80 年代後半、1990-1994 年のものを 90 年代前半、1995-1999 年のものを 90 年代後半、2000-2005 年のものを 2000 年代と、四つの年代に区分した。語別に漢字表記、平仮名表記の割合を算出するのではなく、各年代ごとに全ての語の漢字表記、平仮名表記の度数を合計した上で、漢字表記、平仮名表記の割合を算出している。なお、その際、度数 100 以下の低頻度の語群(ゆれの見られなかった《アグネル》《オオセル》《サス》《ソビエル》《ツケル》《ソソズル》と、《ソコネル》《ハタス》)を除外して集計した。

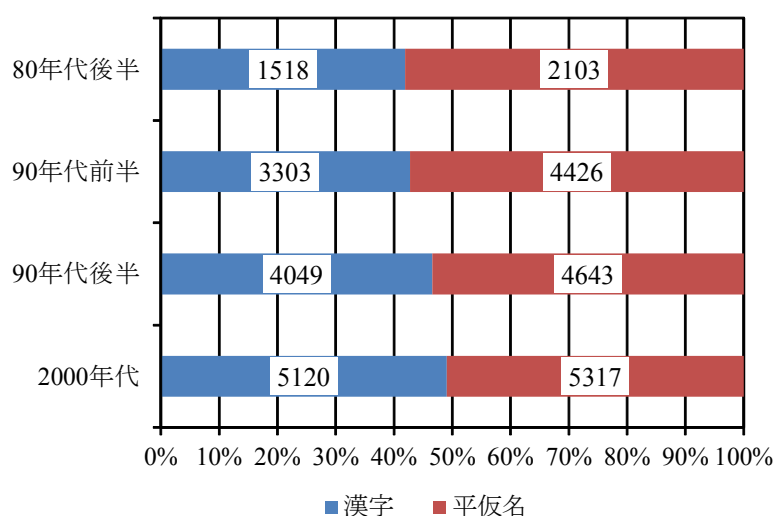


図 1 : 後項動詞の表記の変化(出版年代)

図1を見ると、年代が下るにしたがって、徐々に漢字表記の割合が高くなっていることが分かる。図1の用例数を基に漢字表記の割合を求めると、80年代後半は41.9%であったが、90年代前半は42.7%、90年代後半は46.6%、2000年代は49.1%となる。2000年代には、漢字表記が平仮名表記と拮抗するまでになっているのである。

次に、著者の生年代別の経年変化を見ていく。図2に、著者の生年代別に後項動詞の漢字表記、仮名表記の割合を示した。図1と同様に、各年代ごとに全ての語の度数を合計した上で、漢字表記、平仮名表記の割合を算出しているが、その際、著者が複数のサンプルの用例は集計の対象外としている。また、語の度数の合計が300以上の年代を図に示した。

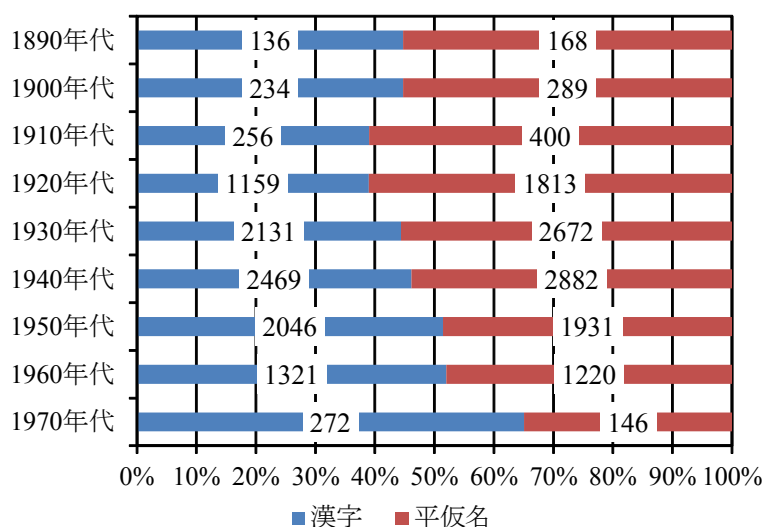


図2：後項動詞の表記の変化(著者生年代)

図2を見ると、1890年代生まれのグループから1910年代生まれのグループまで、漢字表記の割合が低下するが、その後は、1970年代生まれのグループまで漢字表記の割合が増加傾向にある。1950年代以降のグループは、1950年代生まれのグループが51.4%、1960年代生まれのグループが52.0%、1970年代生まれのグループは65.1%と、いずれも漢字表記が過半数を占めている。ただし1970年代生まれのグループは、全体の度数が418と少ないため、慎重に見る必要はあろう。とは言え、度数が2000を超える1950年代、1960年代生まれのグループで漢字表記の割合が5割を超えており、増加傾向にあることは指摘できよう。

4. 終わりに

本稿では、BCCWJの図書館・書籍を資料として、統語的複合動詞等の後項動詞の語表記のゆれ、特に漢字と仮名の対立による語表記のゆれについて実態調査を行った。その結果、次のことが明らかとなった。

- (1) ゆれの程度に応じて「固定」「独占」「ゆれ」の3区分で見た場合、「固定」は低頻度層に、「独占」は中頻度層を中心に、「ゆれ」は高頻度層を中心に分布している。高頻度語ほど、表記にゆれが生じるといえる。
- (2) 「独占」に属する語は、漢字表記に集中する語が多い(9語中6語)。「ゆれ」に属する語(15語)で、平仮名表記の割合が高い語は7語あるが、そのうち5語は度数

500以上の高頻度語である。

- (3) 出版年別の経年調査、著者の生年代別の経年調査のいずれにおいても、年代が下るにしたがって、漢字表記の割合が高くなる。

本稿で取り上げた30語のうち、「ゆれ」に属する語は16語ある。今回は、後項動詞の表記のみの調査ではあったが、統語的複合動詞は、語彙的複合動詞と同様、語表記にゆれが多いと考えられる。

また30語中24語は常用漢字で表記できる語であるが、そのうちの23語に漢字と平仮名の対立によるゆれが見られる。姫野(1999:28)では統語的複合動詞の後項動詞を「接尾辞的複合動詞」と呼ぶが、接尾辞的な性格から、平仮名表記が選択されやすくなっていると考えられる。

しかしながら、出版年代別、著者生年代別の調査結果を見ると、漢字表記が増加傾向にある。今後、語別に経年変化を見ていくことも必要であろう。また、漢字表記の増加傾向が情報機器の普及によるものかを検証するために、特定目的・知恵袋、特定目的・ブログを対象とした調査も必要となろう。今後の課題としたい。

なお、1節で取り上げた《カネル》《スギル》《ダス》の表記の基準については、本稿の結果からいえば、《カネル》《スギル》は平仮名表記を一般的な表記とするのは妥当である。しかし《ダス》については、漢字表記が約54%、平仮名表記が約46%でゆれており、本稿の結果からは、どちらが一般的か決めることは難しい。表記の基準を考えるためにも、本稿のような実態調査を行っていくことが必要である。

謝 辞

本研究は、国立国語研究所共同研究プロジェクト(基幹型)「コーパス日本語学の創成」(リーダー：前川喜久雄)、同「多角的アプローチによる現代日本語の動態の解明」(リーダー：相澤正夫)、JSPS 科研費「大規模コーパスに基づく現代語表記のゆれの実態解明」(代表者：小椋秀樹)による補助を得た。

参 考 文 献

- 小椋秀樹(2012)「コーパスに基づく現代語表記のゆれの調査 — BCCWJ コアデータを資料として —」、『第1回コーパス日本語学ワークショップ予稿集』、pp.321-328.
- 小椋秀樹、小磯花絵、富士池優美、宮内佐夜香、小西光、原裕(2011)『『現代日本語書き言葉均衡コーパス』形態論情報規程集 第4版(上・下)』(国立国語研究所内部報告書).
- 影山太郎(1993)『文法と語形成』、ひつじ書房.
- 白石大二編、野元菊雄新版監修、高田智和改訂新版監修(2010)『例解辞典 改訂新版』、ぎょうせい.
- 姫野昌子(1999)『複合動詞の構造と意味用法』、ひつじ書房.
- 前川喜久雄(2008)「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発」、『日本語の研究』4-1、pp.82-95.
- 丸山岳彦、中村武範(2011)「第8章 書誌情報データベース」、国立国語研究所コーパス開発センター『『現代日本語書き言葉均衡コーパス』利用の手引き』第1.0版、pp.117-152.
- 宮島達夫(1997)「雑誌九十種表記表の統計」、『日本語科学』1、pp.92-103.
- 山崎誠(2011)「第2章 『現代日本語書き言葉均衡コーパス』の設計」、国立国語研究所コーパス開発センター『『現代日本語書き言葉均衡コーパス』利用の手引き』第1.0版、pp.113-20.