

## 現代日本語書き言葉における非外来語のカタカナ表記事情

柏野 和佳子\* (国立国語研究所 言語資源研究系)  
中村 壮範 (国立国語研究所 コーパス開発センター)

### Frequency of Katakana Representation for Japanese Non-loan Words as Observed in the BCCWJ Corpus

Wakako Kashino (Dept. Corpus Studies, NINJAL)  
Takenori Nakamura (Center for Corpus Development, NINJAL)

#### 1. はじめに

「カタカナ語」といえば、主として外来語をさす。しかしながら、非外来語である、和語や漢語にカタカナ表記が用いられることは、少ないことではない。

これまで、新聞や若者雑誌を中心に、非外来語である、和語や漢語のカタカナ表記が増加する傾向にあることが指摘されてきた。佐藤(1991)は若者を対象にカタカナ使用についてのアンケート調査を行った。中山(1998)と成田・榊原(2004)は新聞記事の表記を調査した。堀尾・則松(2005)と則松・堀尾(2006)は、若者雑誌の表記を調査した。堀江(2001)は、少し広範囲に、小説、マンガ、雑誌、新聞記事の表記を調査した。

また、柏野・奥村(2012)は、『現代日本語書き言葉均衡コーパス(Balanced Corpus of Contemporary Written Japanese; 略称 BCCWJ)』(前川 2013)に収録される書籍テキストデータを用いて、書籍における非外来語のカタカナ表記の実態を調査した。その結果、雑誌や新聞同様にカタカナ表記率の高い語がある一方、ひらがな表記率や漢字表記率が高い語もあり、表記の実態は、新聞、雑誌、書籍といった媒体で差があることを明らかにした。

非外来語のカタカナ表記については、教育的な関心も高く、高原(2012)、茂木(2012)、黎(2012)らに取り上げ、調査や分析を行っている。

このように、非外来語のカタカナ表記に関する先行研究はいくつもある。しかしながら、そのほとんどは調査対象語や着目語の選定が主観的に行われている。そこで、本研究は、調査対象語をコーパス分析によって客観的に選ぶことで、現代日本語書き言葉全般における、非外来語のカタカナ表記の実態を明らかにすることを目的にする。その方法は、現代語の均衡コーパスである、BCCWJ の全収録語の自動解析から得られる、同一見出し語の表記別の頻度情報を用いるというものである。本研究の一つ目の成果として、本発表では、カタカナ表記頻度の多い非外来語の語彙表を報告する。さらに、カタカナ頻度の多い語を、カタカナ表記率、ひらがな表記率、漢字表記率、の三つの観点により捉え直して得られた語彙表を報告する。

#### 2. 非外来語のカタカナ表記の実態調査

##### 2.1 カタカナ表記頻度の多い非外来語の語彙表の作成

『現代日本語書き言葉均衡コーパス(BCCWJ)』の DVD 版(国立国語研究所 2012 年)の短単位データを用いた。短単位の全収録語数は、約 1 億 5 百万語である。サブコーパス別の収録語数は、次の表 1 のとおりである。

BCCWJ の短単位データの作成は自動形態素解析で行われている(「コア」と呼ばれる一部には人手修正あり)。形態素解析器に「MeCab<sup>1</sup>」、解析用辞書に「UniDic<sup>2</sup>」が使用されている。この BCCWJ の形態論情報付与に使用されている UniDic では、表記が異なっても同じ語であれば一つの見出し語にまとめられている。よって、解析結果より、見出し語単

\* waka@ninjal.ac.jp

<sup>1</sup> <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

<sup>2</sup> <http://sourceforge.jp/projects/unidic/>

表1 サブコーパス別収録語数 (短単位)

レジスター名	略表記	概要	語数	割合
出版・書籍	PB	出版実態サブコーパス・書籍	28552283	27.22%
出版・雑誌	PM	出版実態サブコーパス・雑誌	4444492	4.24%
出版・新聞	PN	出版実態サブコーパス・新聞	1370233	1.31%
図書館・書籍	LB	図書館実態サブコーパス・書籍	30377866	28.96%
特定目的・ベストセラー	OB	特定目的サブコーパス・ベストセラー	3742261	3.57%
特定目的・白書	OW	特定目的サブコーパス・白書	4882812	4.65%
特定目的・広報紙	OP	特定目的サブコーパス・広報紙	3755161	3.58%
特定目的・法律	OL	特定目的サブコーパス・法律	1079146	1.03%
特定目的・国会会議録	OM	特定目的サブコーパス・国会会議録	5102469	4.86%
特定目的・教科書	OT	特定目的サブコーパス・教科書	928448	0.88%
特定目的・韻文	OV	特定目的サブコーパス・韻文	225273	0.21%
特定目的・知恵袋	OC	特定目的サブコーパス・Yahoo!知恵袋	10256877	9.78%
特定目的・ブログ	OY	特定目的サブコーパス・Yahoo!ブログ	10194143	9.72%
<b>合計</b>			<b>104911464</b>	<b>100.00%</b>

表2 カタカナ表記頻度上位 100 語の語彙表

番号	見出し	読み	カタカナ		頻度総数	番号	見出し	読み	カタカナ		頻度総数
			頻度	率					頻度	率	
1	物	モノ	3798	1.3%	285804	51	蝦	エビ	784	47.9%	1636
2	俺	オレ	3455	11.4%	30265	52	兎	ウサギ	779	51.4%	1517
3	馬鹿	バカ	3179	40.3%	7890	53	喧嘩	ケンカ	772	23.9%	3229
4	本当	ホントウ	2383	6.8%	35043	54	ばちんこ	バチンコ	771	97.3%	792
5	僕	ボク	2282	4.8%	47657	55	散らし	チラン	769	91.0%	845
6	良い	ヨイ	2198	1.1%	198994	56	親父	オヤジ	746	23.6%	3155
7	癌	ガン	2188	32.9%	6652	57	熊	クマ	746	40.9%	1824
8	わた	ネタ	2037	97.5%	2089	58	はつきり	ハッキリ	744	7.1%	10452
9	嫌	イヤ	1991	15.6%	12741	59	餓鬼	ガキ	742	71.8%	1033
10	塵	ゴミ	1918	40.9%	4690	60	ばらばら	バラバラ	738	57.2%	1291
11	車	クルマ	1856	8.6%	21623	61	怪我	ケガ	731	25.5%	2864
12	人	ヒト	1828	1.2%	156890	62	眼鏡	メガネ	730	30.9%	2363
13	子	コ	1719	5.8%	29535	63	茸	キノコ	692	55.8%	1241
14	こつ	コツ	1643	91.1%	1803	64	びん	ビン	685	72.2%	949
15	勤める	ススメル	1582	13.5%	11683	65	ぱつ	パツ	682	43.8%	1557
16	此れ	コレ	1571	0.8%	199500	66	壺	ツボ	675	43.5%	1550
17	薔薇	バラ	1534	72.5%	2116	67	ちび	チビ	670	68.0%	986
18	蛋白	タンパク	1528	48.3%	3164	68	鸚鵡	オウム	668	86.9%	769
19	奴	ヤツ	1356	11.1%	12186	69	貴方	アナタ	667	1.4%	48127
20	格好	カッコウ	1345	20.9%	6432	70	伯父	オジ	666	10.8%	6191
21	漫画	マンガ	1329	37.9%	3510	71	齒	ハ	657	11.5%	5706
22	猫	ネコ	1301	18.5%	7037	72	ぎりぎり	ギリギリ	653	45.9%	1422
23	丸	マル	1287	18.6%	6919	73	訳	ワケ	650	0.9%	73097
24	綺麗	キレイ	1262	10.8%	11715	74	草	ソウ	646	41.1%	1572
25	林檎	リンゴ	1186	49.3%	2408	75	茄子	ナス	640	48.6%	1317
26	苛々	イライラ	1125	68.6%	1639	76	蛸	タコ	630	58.4%	1079
27	私	ワタシ	1108	2.5%	45106	77	蟹	カニ	629	56.3%	1117
28	びっくり	ビックリ	1106	23.1%	4790	78	葱	ネギ	626	45.1%	1389
29	葉書	ハガキ	1097	36.3%	3025	79	にこにこ	ニコニコ	622	64.0%	972
30	猿	サル	1071	46.8%	2288	80	虎	トラ	622	38.4%	1621
31	事	コト	996	0.1%	740923	81	蟬	セミ	621	65.3%	951
32	嘘	ウソ	993	17.9%	5537	82	鍵	カギ	620	16.1%	3846
33	金	カネ	961	4.9%	19742	83	癖	クセ	619	15.8%	3913
34	どきどき	ドキドキ	942	79.7%	1182	84	杉	スギ	614	44.5%	1380
35	雌	メス	936	47.2%	1985	85	内	ウチ	611	1.4%	45175
36	まじ	マジ	925	65.5%	1412	86	犬	イヌ	594	7.0%	8495
37	阿呆	アホウ	901	65.0%	1386	87	彼れ	アレ	594	4.8%	12404
38	鼠	ネズミ	888	55.9%	1588	88	桜	サクラ	589	13.1%	4506
39	台詞	セリフ	866	46.5%	1861	89	すっきり	スッキリ	585	30.4%	1922
40	巢	ス	856	34.8%	2461	90	惚け	ボケ	585	74.0%	791
41	携帯	ケイタイ	852	10.1%	8408	91	蛙	カエル	584	79.2%	737
42	家	ウチ	851	59.6%	1429	92	大蒜	ニンニク	578	39.1%	1479
43	木	キ	826	5.8%	14356	93	有り	アリ	572	11.0%	5177
44	葡萄	ブドウ	824	44.4%	1855	94	乗り	ノリ	571	43.6%	1311
45	餌	エサ	816	32.2%	2531	95	手	テ	558	1.2%	47213
46	はい	ハイ	809	10.5%	7697	96	蝶	チョウ	555	41.0%	1353
47	烏	カラス	802	59.5%	1348	97	此処	ココ	553	1.0%	56797
48	烏賊	イカ	801	63.5%	1261	98	行く	イク	549	0.2%	220638
49	雄	オス	792	41.4%	1913	99	段	ダン	548	11.4%	4804
50	徴	カビ	792	87.0%	910	100	勤	カン	546	44.9%	1216

位の表記別の頻度情報を自動的に得ることができる。その頻度情報をもとに、カタカナ表記頻度の多い語を抽出した。その際、数詞、助詞、接頭辞、接尾辞は対象外とした。もともと頻度の多かったのは「モノ(物)」で、その頻度は3,798である。298位は4語あり、その頻度は246である。頻度246~3,798の301語が調査範囲として扱いやすい適当なサイズであると判断し、それら301語を今回の調査対象語として選定することとした。そのうち、上位100語で作成した語彙表を表2に示す。

表2を見ると、頻度総数に対する、カタカナ表記率が語によって大きく異なることが目につく。たとえば、1位の「モノ(物)」は、ひらがな表記、漢字表記とを合わせた頻度総数が285,804と多く、そのカタカナ表記率はわずか1.3%に過ぎない。

## 2.2 表記の比率別にみる語彙表の作成

調査対象の301語を、カタカナ表記率、ひらがな表記率、漢字表記率の高い順に並び替え、それぞれの上位50語で作成した各語彙表を順に、表3~5に示す。

表3 カタカナ表記率上位50語の語彙表

番号	見出し	読み	カタカナ		ひらがな		漢字		頻度総数
			頻度	率	頻度	率	頻度	率	
1	もてる	モデル	285	99.7%	1	0.3%	0	0.0%	286
2	女	メ	419	99.1%	0	0.0%	4	0.9%	423
3	ねたばれ	ネタバレ	263	98.5%	4	1.5%	0	0.0%	267
4	蛇	ジャ	320	98.2%	0	0.0%	6	1.8%	326
5	壁蝨	ダニ	301	97.7%	6	1.9%	1	0.3%	308
6	ねた	ネタ	2037	97.5%	52	2.5%	0	0.0%	2089
7	海豚	イルカ	430	97.5%	5	1.1%	6	1.4%	441
8	ばちんこ	バチンコ	771	97.3%	21	2.7%	0	0.0%	792
9	ぶす	ブス	349	91.6%	32	8.4%	0	0.0%	381
10	こつ	コツ	1643	91.1%	160	8.9%	0	0.0%	1803
11	散らし	チラシ	769	91.0%	70	8.3%	6	0.7%	845
12	軟派	ナンパ	250	90.9%	0	0.0%	25	9.1%	275
13	びら	ピラ	285	90.5%	30	9.5%	0	0.0%	315
14	ぱりぱり	バリバリ	380	88.6%	49	11.4%	0	0.0%	429
15	こく	コク	401	87.4%	58	12.6%	0	0.0%	459
16	劫	コウ	406	87.3%	0	0.0%	59	12.7%	465
17	かりかり	カリカリ	252	87.2%	37	12.8%	0	0.0%	289
18	徴	カビ	792	87.0%	55	6.0%	63	6.9%	910
19	鸚鵡	オウム	668	86.9%	46	6.0%	55	7.2%	769
20	鸚哥	インコ	265	86.6%	19	6.2%	22	7.2%	306
21	焔炉	コンロ	290	85.0%	46	13.5%	5	1.5%	341
22	でぶ	デブ	394	82.8%	82	17.2%	0	0.0%	476
23	ぼち	ボチ	498	81.8%	111	18.2%	0	0.0%	609
24	蛾	ガ	536	81.0%	10	1.5%	116	17.5%	662
25	がらがん	ガンガン	365	80.8%	87	19.2%	0	0.0%	452
26	がらがら	ガラガラ	282	80.6%	68	19.4%	0	0.0%	350
27	鍍金	メッキ	252	80.5%	61	19.5%	0	0.0%	313
28	鱈	コマ	426	80.2%	65	12.2%	40	7.5%	531
29	どきどき	ドキドキ	942	79.7%	240	20.3%	0	0.0%	1182
30	どん	ドン	517	79.5%	133	20.5%	0	0.0%	650
31	蛙	カエル	584	79.2%	99	13.4%	54	7.3%	737
32	ぶな	ブナ	290	78.6%	71	19.2%	8	2.2%	369
33	片仮名	カタカナ	445	78.2%	9	1.6%	115	20.2%	569
34	家鴨	アヒル	268	77.5%	52	15.0%	26	7.5%	346
35	鮑	アワビ	318	77.2%	34	8.3%	60	14.6%	412
36	ちん	チン	276	77.1%	82	22.9%	0	0.0%	358
37	蜻蛉	トンボ	397	76.8%	104	20.1%	16	3.1%	517
38	麒麟	キリン	400	75.5%	26	4.9%	104	19.6%	530
39	とん	トン	246	75.5%	80	24.5%	0	0.0%	326
40	とんとん	トントン	246	74.5%	84	25.5%	0	0.0%	330
41	わん	ワン	311	74.2%	108	25.8%	0	0.0%	419
42	栗鼠	リス	416	74.0%	106	18.9%	40	7.1%	562
43	惚け	ボケ	585	74.0%	158	20.0%	48	6.1%	791
44	上	カミ	452	73.0%	51	8.2%	116	18.7%	619
45	けち	ケチ	397	72.7%	138	25.3%	11	2.0%	546
46	薔薇	バラ	1534	72.5%	88	4.2%	494	23.3%	2116
47	びん	ビン	685	72.2%	264	27.8%	0	0.0%	949
48	蜥蜴	トカゲ	250	72.0%	29	8.4%	68	19.6%	347
49	餓鬼	ガキ	742	71.8%	82	7.9%	209	20.2%	1033
50	括弧	カッコ	372	71.0%	15	2.9%	137	26.1%	524

表3に示すカタカナ表記率上位50語は、その比率が71.0%以上といずれも高く、主にカタカナ表記される語であることが確認できよう。ただし、たとえば、16位の「コウ(劫)」のように、あまり見慣れない語も入っている。実際の用例を確認してみると、囲碁用語や、仏教で長い時間を表す語としての「劫」のカタカナ表記の使用例のほか、音を表すだけの「コウ」が「劫」と誤解析された結果もまじっているため、頻度や比率が高くなってしまっていた。44位の「カミ(上)」もわかりにくい例であるが、これは、「カミサン(上様)」(＝主婦を呼ぶのに用いる)のカタカナ表記の使用例であった。ただし、その中に、「神」のカタカナ表記を「上」と誤解析したものも含まれていた。このように、自動解析にはある程度の誤解析は含まれはするが、表記実態の傾向は自動解析結果より十分にみてとれる。

表4は、ひらがな表記率上位50語の表である。上位を占める半分以上の語が、ひらがな表記が主な語である。それらは、カタカナ表記率は低いが、使用総数が多いためにカタカナ表記の頻度が多くなっている語であることがわかる。しかし、カタカナ表記率が20%を超える語は、ひらがな使用が主であるといっても、カタカナ使用の目立つ語の一群であるとの印象を受けるものになっている。

表4 ひらがな表記率上位50語の語彙表

番号	見出し	読み	カタカナ		ひらがな		漢字		頻度総数
			頻度	率	頻度	率	頻度	率	
1	為る	スル	395	0.0%	2563160	100.0%	305	0.0%	2563860
2	此れ	コレ	1571	0.8%	195971	98.2%	1958	1.0%	199500
3	一寸	チョット	344	1.2%	29147	98.1%	206	0.7%	29697
4	此処	ココ	553	1.0%	55682	98.0%	562	1.0%	56797
5	此方	コチラ	322	1.8%	17508	97.5%	126	0.7%	17956
6	私	ワタシ	1108	2.5%	43951	97.4%	47	0.1%	45106
7	あつ	アツ	406	4.2%	9256	95.8%	1	0.0%	9663
8	彼れ	アレ	594	4.8%	11779	95.0%	31	0.2%	12404
9	事	コト	996	0.1%	702885	94.9%	37042	5.0%	740923
10	うん	ウン	273	5.2%	4995	94.8%	0	0.0%	5268
11	物	モノ	3798	1.3%	268202	93.8%	13804	4.8%	285804
12	訳	ワケ	650	0.9%	68478	93.7%	3969	5.4%	73097
13	筈	ハズ	271	1.1%	23796	93.4%	1408	5.5%	25475
14	貴方	アナタ	667	1.4%	44900	93.3%	2560	5.3%	48127
15	はつきり	ハツキリ	744	7.1%	9708	92.9%	0	0.0%	10452
16	所	トコロ	427	0.4%	100940	90.2%	10506	9.4%	111873
17	えっ	エツ	448	9.8%	4116	90.2%	0	0.0%	4564
18	内	ウチ	611	1.4%	40576	89.8%	3988	8.8%	45175
19	はい	ハイ	809	10.5%	6888	89.5%	0	0.0%	7697
20	わし	ワシ	290	6.0%	4196	87.4%	317	6.6%	4803
21	へえ	ヘエ	301	15.3%	1664	84.7%	0	0.0%	1965
22	只	ただ	485	11.0%	3694	83.7%	235	5.3%	4414
23	きつい	キツイ	373	16.7%	1854	83.3%	0	0.0%	2227
24	有り	アリ	572	11.0%	4307	83.2%	298	5.8%	5177
25	御免	ゴメン	436	10.0%	3555	81.9%	349	8.0%	4340
26	でかい	デカイ	247	18.7%	1068	80.9%	5	0.4%	1320
27	良い	ヨイ	2198	1.1%	160909	80.9%	35887	18.0%	198994
28	凄い	スゴイ	454	3.1%	11569	78.4%	2724	18.5%	14747
29	さっ	サツ	405	21.6%	1445	77.1%	24	1.3%	1874
30	わあ	ワア	401	24.0%	1272	76.0%	0	0.0%	1673
31	ぴったり	ピッタリ	541	24.0%	1716	76.0%	0	0.0%	2257
32	びっくり	ビックリ	1106	23.1%	3600	75.2%	84	1.8%	4790
33	ぐっ	グッ	306	27.7%	796	72.2%	1	0.1%	1103
34	すっきり	スッキリ	585	30.4%	1337	69.6%	0	0.0%	1922
35	ちら	チラ	340	33.5%	675	66.5%	0	0.0%	1015
36	御洒落	オンシャレ	322	17.3%	1216	65.2%	327	17.5%	1865
37	菠蘿	ホウレン	255	35.0%	465	63.9%	8	1.1%	728
38	やばい	ヤバイ	437	36.1%	772	63.9%	0	0.0%	1209
39	胡麻	ゴマ	425	19.9%	1358	63.6%	353	16.5%	2136
40	餡	アン	311	25.3%	779	63.3%	140	11.4%	1230
41	行く	イク	549	0.2%	134413	60.9%	85676	38.8%	220638
42	増し	マシ	496	32.2%	910	59.0%	136	8.8%	1542
43	ずれ	ズレ	288	41.1%	412	58.9%	0	0.0%	700
44	奇麗	キレイ	1262	10.8%	6895	58.9%	3558	30.4%	11715
45	伯母	オバ	542	11.8%	2682	58.5%	1358	29.6%	4582
46	大蒜	ニンニク	578	39.1%	854	57.7%	47	3.2%	1479
47	己等	オイラ	477	42.6%	641	57.2%	2	0.2%	1120
48	ばれる	バレル	516	42.9%	682	56.7%	5	0.4%	1203
49	ぱっ	パッ	682	43.8%	875	56.2%	0	0.0%	1557
50	にやり	ニヤリ	283	44.3%	356	55.7%	0	0.0%	639

表5は、漢字表記比率上位50語の表である。上位50語のうち、カタカナ表記率が20%を超える語は少ない。しかし、20%以下でも、カタカナ表記頻度の多い語を中心に、カタカナ表記が多いと感じる語がこの表にも多くあがっている。たとえば、22位の「ケイタイ(携帯)」は、そのカタカナ表記の内訳が、「ケータイ(797)/ケイタイ(54)/ケータイ(1)」であるのだが、「ケータイ」という使用例がもはや標準であるかのような印象を与える語である。しかし、それはあくまでも、「携帯電話」をさす「ケータイ」であり、「携帯電話」以外にも、「携帯する」「携帯手荷物」といった用法をもつ「ケイタイ(携帯)」という語は、漢字表記が圧倒的に多いということである。また、この表の中では、従来の先行研究ではあがってこなかった34位の「ソン(損)」に注目したい。漢字表記率が高い一方、カタカナ表記率も20.5%あり、カタカナ表記頻度が382ある。よって、この語も、カタカナ使用の目立つ実態をもつ語の一つであると言えるだろう。

表5 漢字表記率上位50語の語彙表

番号	見出し	読み	カタカナ		ひらがな		漢字		頻度総数
			頻度	率	頻度	率	頻度	率	
1	手	テ	558	1.2%	32	0.1%	46623	98.8%	47213
2	日	ヒ	388	0.7%	589	1.0%	58221	98.3%	59198
3	機	キ	421	1.9%	190	0.9%	21124	97.2%	21735
4	山	ヤマ	277	1.5%	254	1.4%	18094	97.1%	18625
5	人	ヒト	1828	1.2%	3932	2.5%	151130	96.3%	156890
6	村	ムラ	360	2.8%	152	1.2%	12398	96.0%	12910
7	奥	オク	252	2.2%	219	1.9%	11159	96.0%	11630
8	彼	カレ	290	0.3%	4254	4.6%	87495	95.1%	92039
9	金	カネ	961	4.9%	15	0.1%	18766	95.1%	19742
10	茶	チャ	262	3.9%	94	1.4%	6279	94.6%	6635
11	薬	クスリ	247	4.1%	98	1.6%	5671	94.3%	6016
12	首	クビ	457	4.6%	119	1.2%	9412	94.2%	9988
13	子	コ	1719	5.8%	233	0.8%	27583	93.4%	29535
14	個	コ	529	4.4%	265	2.2%	11203	93.4%	11997
15	黒	クロ	470	5.0%	174	1.8%	8802	93.2%	9446
16	無理	ムリ	400	3.9%	365	3.6%	9511	92.6%	10276
17	裏	ウラ	417	7.0%	33	0.6%	5529	92.5%	5979
18	犬	イヌ	594	7.0%	86	1.0%	7815	92.0%	8495
19	鳥	トリ	377	8.5%	2	0.0%	4047	91.4%	4426
20	車	クルマ	1856	8.6%	71	0.3%	19696	91.1%	21623
21	切れる	キレル	276	8.9%	6	0.2%	2822	90.9%	3104
22	携帯	ケイタイ	852	10.1%	2	0.0%	7554	89.8%	8408
23	傷	キズ	258	7.8%	93	2.8%	2945	89.4%	3296
24	木	キ	826	5.8%	753	5.2%	12777	89.0%	14356
25	案	ラク	419	8.5%	207	4.2%	4277	87.2%	4903
26	段	ダン	548	11.4%	72	1.5%	4184	87.1%	4804
27	歯	ハ	657	11.5%	86	1.5%	4963	87.0%	5706
28	体	カラダ	381	1.6%	2690	11.5%	20379	86.9%	23450
29	松	マツ	252	11.3%	84	3.8%	1897	85.0%	2233
30	虫	ムシ	429	12.8%	94	2.8%	2829	84.4%	3352
31	大根	ダイコン	271	12.6%	104	4.8%	1776	82.6%	2151
32	麦	ヘン	504	8.7%	547	9.4%	4741	81.9%	5792
33	皿	サラ	502	15.3%	107	3.3%	2669	81.4%	3278
34	損	ソン	382	20.5%	0	0.0%	1480	79.5%	1862
35	鍵	カギ	620	16.1%	203	5.3%	3023	78.6%	3846
36	椅子	イス	539	8.8%	866	14.1%	4749	77.2%	6154
37	誰	ダレ	375	1.0%	8112	22.2%	28105	76.8%	36592
38	猫	ネコ	1301	18.5%	380	5.4%	5356	76.1%	7037
39	鶏	ニワトリ	381	17.8%	146	6.8%	1616	75.4%	2143
40	暇	ヒマ	373	12.5%	362	12.2%	2242	75.3%	2977
41	無駄	ムダ	480	12.9%	472	12.7%	2762	74.4%	3714
42	下手	ヘタ	288	12.1%	330	13.9%	1758	74.0%	2376
43	桜	サクラ	589	13.1%	608	13.5%	3309	73.4%	4506
44	豚	ブタ	426	22.8%	75	4.0%	1365	73.2%	1866
45	米	コメ	534	25.7%	62	3.0%	1483	71.3%	2079
46	本当	ホントウ	2383	6.8%	7724	22.0%	24936	71.2%	35043
47	麵	メン	310	15.8%	302	15.4%	1348	68.8%	1960
48	嘘	ウソ	993	17.9%	799	14.4%	3745	67.6%	5537
49	蛇	ヘビ	392	27.2%	82	5.7%	969	67.2%	1443
50	稲	イネ	257	31.3%	21	2.6%	544	66.2%	822

### 3. 今後の課題

BCCWJ 全体の自動解析結果を用いた非外来語のカタカナ表記の調査分析は着手した

かりである。先に、柏野・奥村(2012)においては、その時の調査対象語に対し、カタカナ表記になりやすい語のタイプ(「表外漢字」「表外音訓」「表外熟字訓」などを含む語、動植物名の語、擬音語・擬態語、第一義でない意味の語、特別な意味を加味する用法の語)やケース(強調用法のある場合、同字語を避けたい場合、音を明示したい場合)別に使用実態を分析した。BCCWJで得られる今回の調査対象語に対しても、そうしたタイプやケース別の分析を進める予定でいる。また、表1に示したとおり、BCCWJは複数のレジスタから構成されており、レジスタ別の分析が可能である。試行的に行っている予備調査では、たとえば、教科書や白書ではカタカナ表記が避けられ、逆に、新聞、雑誌、知恵袋、ブログでは、カタカナ表記が好まれる、という傾向がみてとれている。今後、レジスタ別の使用傾向の差異についても明らかにしていきたい。

### 謝 辞

本研究は、国立国語研究所の共同研究プロジェクト基幹型「コーパス日本語学の創成」による補助、並びに、文部科学省科学研究費補助金基盤研究(C)「コーパス分析に基づく辞書の位相情報の精緻化」(課題番号: 23520572)の助成を受けたものです。

### 文 献

- 柏野和佳子・奥村学(2012)「和語や漢語のカタカナ表記:『現代日本語書き言葉均衡コーパス』の書籍における使用実態」『計量国語学』28(4), pp.153-161.
- 佐藤栄作(1991)「若者のカタカナ使用と外来語表記—語種意識から—」『日本語学』10-7, pp.76-88, 明治書院.
- 高原真理(2012)「日本語学習辞書における表記情報の課題と今後の取り組み」(2012年度マレーシア国際研究集会「日本語学習辞書開発の支援を考える」口頭発表資料 <http://jishokaken.sakura.ne.jp/doc/Malaysia/M3.pdf>) .
- 中山恵利子(1998)「非外来語の片仮名表記」『日本語教育』96, pp.61-72, 日本語教育学会.
- 中山恵利子・桐生りか・山口昌也(2007) 第3部第3章「新聞に見る基幹外来語」『国立国語研究所報告126 公共媒体の外来語—「外来語」言い換え提案を支える調査研究—』.
- 成田徹男・榊原浩之(2004)「現代日本語の表記体系と表記戦略—カタカナの使い方の変化—」『人間文化研究』2, pp.41-55, 名古屋市立大学.
- 則松智子・堀尾香代子(2006)「若者雑誌における常用漢字のカタカナ表記化—意味分析の観点から—」『北九州市立大学文学部紀要』72, pp.19-32.
- 堀江紫野(2001)「カタカナ表記の研究—非外来語系を中心に—」『国文目白』40, pp.16-24, 日本女子大学.
- 堀尾香代子・則松智子(2005)「若者雑誌におけるカタカナ表記とその慣用化をめぐって」『北九州市立大学文学部紀要』69, pp.35-44.
- 前川喜久雄(編)(2013)『コーパス入門』(講座 日本語コーパス) 朝倉書店.
- 茂木俊伸「第5課「チョー恥ずかしかったヨ!」なカタカナの不思議」, 定延利之(編著)『私たちの日本語』朝倉書店, pp.47-57.
- 黎婉珊(2012)「非外来語のカタカナ表記—「中納言」検索および日本語教科書についての調査から—」(日本語教育国際研究大会名古屋2012年ポスター発表) A494.

### 関連 URL

国立国語研究所コーパス開発センター [http://www.ninjal.ac.jp/corpus\\_center/](http://www.ninjal.ac.jp/corpus_center/)