

『計量国語学』アーカイブ

ID	KK300301
種別	論文B
タイトル	コーパスにおける格助詞の使用実態 —BCCWJ・CSJにみる分布—
Title	Usage of Case Particles in Corpus: BCCWJ and CSJ
著者	丸山 直子
Author	MARUYAMA Naoko
掲載号	30巻3号
発行日	2015年12月20日
開始ページ	127
終了ページ	145
著作権者	計量国語学会

論文 B

コーパスにおける格助詞の使用実態

—BCCWJ・CSJ にみる分布—

丸山 直子 (東京女子大学)

要旨

現代日本語の格助詞について、『現代日本語書き言葉均衡コーパス』(BCCWJ)¹及び『日本語話し言葉コーパス』(CSJ)²における用いられ方を観察し、書き言葉と話し言葉の違い、及びそれぞれのレジスターごとの違いを明らかにした。BCCWJ はコアのみ(新聞、雑誌、書籍、白書、知恵袋、ブログ)を調査対象とし、CSJ は、同一話者による独話(学会講演)と対話(自由会話)4件ずつを対象として調査を行った。

BCCWJ も CSJ も、全語数の約 30% が助詞であり、助詞の中では格助詞が最も多い。BCCWJ において、白書と知恵袋は対極にある。白書はデの代わりにニオイテを用いるなど、かたい書き言葉としての性質を持っている。新聞は多少白書に似た性質を持つ。知恵袋とブログは、ともに話し言葉的な性質を帯びているが、知恵袋の方がより独特である。CSJ においては、学会講演と自由会話で性質が異なる。CSJ の学会講演と、BCCWJ のかたいタイプの書き言葉(白書・新聞)には類似点が見られる。

キーワード: 格助詞, 書き言葉と話し言葉, BCCWJ, CSJ, コレスポネンス分析

1. はじめに

コーパスにおける助詞の使われ方に関しては、大曾美恵子(2007)・宮内佐夜香(2012)・松田真希子(2014)(2015)のような調査がある。しかし、書き言葉・話し言葉における助詞の分布を全体として把握した調査は見当たらない。筆者は、まず助詞全体の分布について調査をし、丸山(2015)で報告した³。本稿では、格助詞に絞って、より詳しく調査を行い、報告する。

調査対象とするコーパスは、『現代日本語書き言葉均衡コーパス』(BCCWJ)及び『日本語話し言葉コーパス』(CSJ)で、書き言葉と話し言葉の違い、及びそれぞれのレジスターごとの違いを明らかにする。BCCWJ はコアのみ(新聞、雑誌、書籍、白書、知恵袋、ブログ)を調査対象とし、CSJ は、同一話者による独話(学会講演)と対話(自由会話)

1 http://pj.ninjal.ac.jp/corpus_center/bccwj/

2 http://pj.ninjal.ac.jp/corpus_center/csj/

3 助詞全体の分布としては、BCCWJ も CSJ も、全語数の約 30% が助詞であり、助詞の中では格助詞が最も多い。BCCWJ においては、白書と知恵袋・ブログは、助詞の使用法に関して、様々な点で対極にある。白書はかなり特殊で、格助詞相当の複合辞が多く、短単位と長単位で大きく分布が異なる。新聞は多少白書に似た性質を持つ。知恵袋とブログは、格助詞が少なく終助詞が多いという共通の性質を帯びているが、内訳を見ると、知恵袋の方が特殊性がある。CSJ は、講演の方が格助詞が多く、会話には副助詞・終助詞が多い。講演の方が書き言葉に近いと言える。

4 件ずつを対象とする。

2. 調査対象

BCCWJ・CSJ の、調査対象としたものを表 1・表 2 に記す。BCCWJ は、コアすべてで、短単位で約 100 万語、長単位で約 80 万語である。CSJ は、4 名の学会講演・自由会話 1 件ずつで、計 8 件である。こちらは短単位で計約 3 万語という小さなサンプルである。

表 1: BCCWJ の調査対象

	短単位総数	長単位総数
出版・新聞コア	308,504	224,140
出版・雑誌コア	202,268	159,883
出版・書籍コア	204,050	169,730
特定目的・白書コア	197,011	129,646
特定目的・知恵袋コア	93,932	78,770
特定目的・ブログコア	92,746	75,242
計	1,098,511	837,411

表 2: CSJ の調査対象

講演者 ID	性別	生年代	基にした学会講演 ID	短単位数	長単位数	自由会話 ID ⁴	短単位数	長単位数
1185	女	70to74	A11F0703	5,634	4,697	D03F0034	3,021	2,699
19	女	65to69	A05F0043	3,512	2,655	D03F0058	2,330	2,039
471	男	75to79	A11M0369	3,119	2,246	D03M0004	2,491	2,178
373	男	45to49	A11M0469	6,763	5,379	D03M0038	3,638	3,278
計				19,028	14,977		11,480	10,194

3. 助詞の分類及び格助詞の認定

本稿では、BCCWJ は中納言オンライン版の短単位・長単位分割及び品詞分類に基づき、CSJ は、DVD に収められている短単位・長単位データに基づき集計した。BCCWJ・CSJ とも、格助詞・副助詞・係助詞・接続助詞・終助詞・準体助詞の六分類である。判定基準は両者間で異なっている。今回は、異なった状態のまま集計し、比較した⁵。

4. 調査で得られた格助詞

以下に、それぞれのコーパスに含まれていた格助詞の一覧を表にして示す。長単位の欄は、短単位にない形のを載せている。どれも、出現形が異なるものを含む。ガ・ヲ・

4 CSJ 自由会話は、インタビューアーの発話を含むものである。同一話者のレジスター別比較を厳密に行うには、インタビューアーの発話を除いて検討する必要があるが、今回は対話全体（一つのファイル全体）の数値を示すこととした。

5 CSJ は現在、中納言での検索が可能になるように調整中である（渡部他：2015）。来年度以降、両者同じ基準での比較が可能になる。

ニ・ト・デ・ヘ・ヨリ・カラ・ノが格助詞として認められているというのは標準的な捉え方である。しばしば格助詞とされるマデは副助詞になっている。ノは連体助詞も含む。ガからカラまでの順番は、概ね中核的な格から周辺の格への順番とした。長単位は、それらの助詞を含む形になっているので、その順に示した。BCCWJ, CSJ 共通のものに下線を引いた。

表 3: コーパス中の助詞一覧

	BCCWJ	CSJ
格助詞 (短単位)	<u>ガ</u> , <u>ヲ</u> , <u>ニ</u> , <u>ト</u> , <u>デ</u> , <u>ヘ</u> , <u>ヨリ</u> , <u>カラ</u> , <u>ノ</u> , <u>トテ</u> , <u>ニテ</u> , <u>サ</u>	<u>ガ</u> , <u>ヲ</u> , <u>ニ</u> , <u>ト</u> , <u>デ</u> , <u>ヘ</u> , <u>ヨリ</u> , <u>カラ</u> , <u>ノ</u> , <u>デハ</u> (じゃ)
格助詞 (長単位)	ヲ+ α : <u>ヲ通ジテ</u> , <u>ヲハジメ</u> , <u>ヲメグル</u> , <u>ヲモツテ</u> , ニ+ α : <u>ニアタツテ</u> , <u>ニアタリ</u> , <u>ニイタルマデ</u> , <u>ニオイテ</u> , <u>ニオケル</u> , <u>ニ関シテ</u> , <u>ニ関スル</u> , <u>ニ際シ</u> , <u>ニ際シテ</u> , <u>ニシテ</u> , <u>ニ対シ</u> , <u>ニ対シテ</u> , <u>ニ対スル</u> , <u>ニツイテ</u> , <u>ニツギ</u> , <u>ニトツテ</u> , <u>ニヨツテ</u> , <u>ニヨリ</u> , <u>ニヨル</u> , <u>ニヨルト</u> , <u>ニヨレバ</u> , <u>ニワタツテ</u> , <u>ニワタリ</u> , <u>ニワタル</u> , <u>際ニ</u> , ト+ α : <u>トイウ</u> , <u>トイッタ</u> , <u>トシテ</u> , カラ+ α : <u>カラシテ</u> , <u>カラスルト</u> , <u>カラスレバ</u> , ノ+ α : <u>タメノ</u>	ヲ+ α : <u>ヲモトニシタ</u> , <u>ヲモトニシテ</u> , ニ+ α : <u>ニオイテ</u> , <u>ニオケル</u> , <u>ニ関シテ</u> , <u>ニ関シマシテ</u> , <u>ニ関スル</u> , <u>ニ比ベテ</u> , <u>ニ従ツテ</u> , <u>ニ対シテ</u> , <u>ニ対シマシテ</u> , <u>ニ対スル</u> , <u>ニツイテ</u> , <u>ニツキマシテ</u> , <u>ニトツテ</u> , <u>ニ伴ウ</u> , <u>ニ基ヅイタ</u> , <u>ニ基ヅイテ</u> , <u>ニ基ヅク</u> , <u>ニヨツテ</u> , <u>ニヨル</u> , <u>ニヨリマス</u> , <u>ニヨリマス</u> , ト+ α : <u>トイウ</u> , <u>トイッタ</u> , <u>トシテ</u> , <u>トイタシマシテ</u>

※漢字仮名交じりの表記にしたのは、見やすさのためである。実際には「語彙素読み」は片仮名、「語彙素」は漢字仮名交じりの代表的な表記で示されており、個々のデータにおいてはそれぞれ多様な表記が見られる。

5. BCCWJ における格助詞

5.1 格助詞全体の分布

BCCWJ コアにおける格助詞の内訳は表 4 の通りである。

表 4: BCCWJ 格助詞の数 (短単位)

	ガ	ヲ	ニ	ト	デ	ヘ	ヨリ	カラ	ノ	その他	計
新聞コア	7235	10294	9826	6165	5102	522	141	1362	16390	15	57052
雑誌コア	5000	6192	6645	4168	2728	228	137	900	9714	10	35722
書籍コア	5372	6576	7361	5079	2530	277	140	861	9880	7	38083
白書コア	3545	6265	7948	3615	1456	341	129	675	12636	9	36619
知恵袋コア	2375	2029	2641	1931	1552	57	97	326	3261	5	14274
ブログコア	2018	1965	2643	1710	1371	118	82	410	3773	14	14104
計	25545	33321	37064	22668	14739	1543	726	4534	55654	60	195854

サンプルの大きさが異なるので、1万語あたりに換算すると、表 5 のようになる。

表 5: BCCWJ 格助詞の数 (短単位) (1 万語当たりの数)

	ガ	ヲ	ニ	ト	デ	ヘ	ヨリ	カラ	ノ	格助詞全体
新聞コア	235	334	319	200	165	17	5	44	531	1849
雑誌コア	247	306	329	206	135	11	7	44	480	1766
書籍コア	263	322	361	249	124	14	7	42	484	1866
白書コア	180	318	403	183	74	17	7	34	641	1859
知恵袋コア	253	216	281	206	165	6	10	35	347	1520
ブログコア	218	212	285	184	148	13	9	44	407	1521

これを棒グラフで示すと、図 1 のようになる。

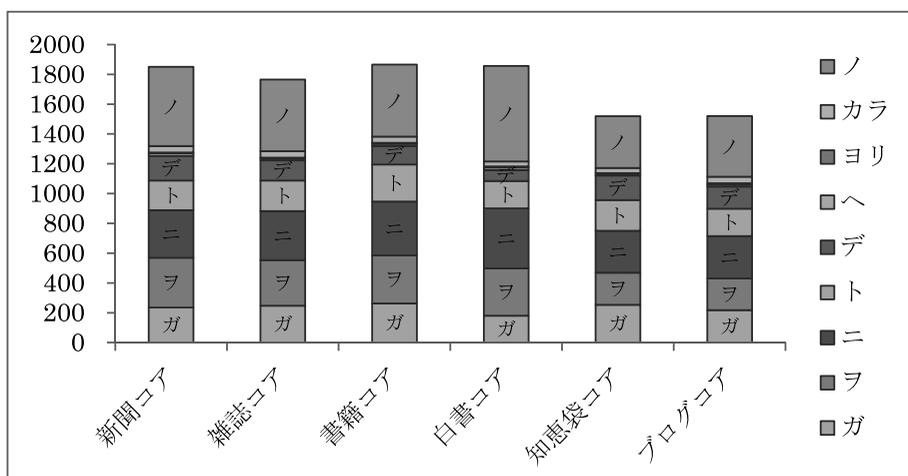


図 1: BCCWJ 格助詞の数 (短単位) (1 万語当たりの数)

新聞はヲが多い。白書はノ・ニが多く、ト・ガ・デは少ない。知恵袋はガ・デが多く、ヲ・ヘは少ない。ブログもヲが少ない。

表 5 のクロス表の内容をもとに、数の少ないヘ・ヨリを除いて、レジスターと助詞の関係をより詳細に把握するため、各助詞を第 1 アイテム、コーパス種別を第 2 アイテムとしてコレスポネンス分析を行った。その結果、下記の散布図を得た (図 2)。なお、第 1 次元の寄与率は 81.88%、第 2 次元の寄与率は 9.9%、2 つの次元による累計寄与率は 91.78% であるため、2 つの次元に基づく解釈に一定の妥当性があると判断した。第 1 次元の寄与率が圧倒的である。軸解釈を行うと、第 1 次元はブログや知恵袋などのくだけた話し言葉的ジャンル (+) と、白書のようなかたい書き言葉的ジャンル (-) を区分している軸と考えられる。新聞・雑誌・書籍は 0 のところに一直線に並んでいる。また、第 2 次元は新聞・ブログが +、書籍・知恵袋が - となっている。

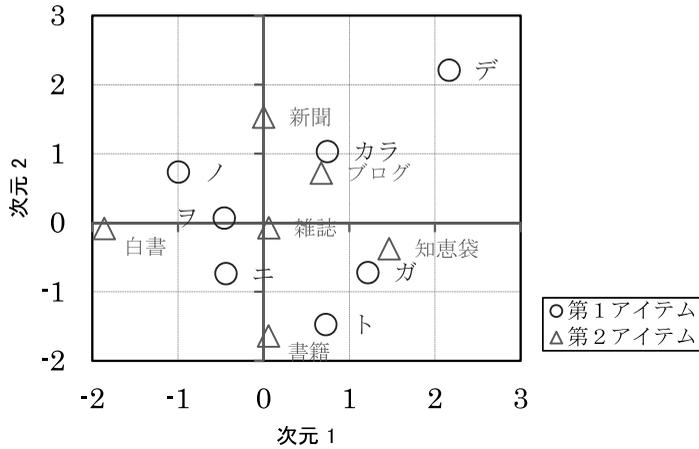


図2: コレスポネンス分析の散布図 (短単位)

長単位で調査すると、だいたい値が異なる。調査結果を、表6・図3⁶に示す。ニを伴う複合辞、トを伴う複合辞の数が多くなる。特に白書には、ニツイテ・ニオイテ・ニヨル等、ニを伴う複合辞が多い。短単位で白書にニが多いのはニを伴う複合辞が多いからである。白書は、単独のニは、むしろ他に比べて少ない。複合辞に関わる格助詞はニとトが主である。

長単位についても、1万語換算をしてから、数の少ないへ・ヨリを除いて、各助詞を第1アイテム、コーパス種別を第2アイテムとしてコレスポネンス分析を行った。その結果、図4のような散布図を得た。なお、第1次元の寄与率は90.48%、第2次元の寄与率は5.34%、2つの次元による累計寄与率は95.82%であるため、2つの次元に基づく解釈に一定の妥当性があると判断した。白書・知恵袋、及びニ + a (ニを伴う複合辞)・ト + a (トを担う複合辞) の位置が特徴的となっている。

表6: BCCWJ 格助詞の数 (長単位)

	ガ	ヲ	ニ	ニを伴う複合辞	ト	トを伴う複合辞	デ	へ	ヨリ	カラ	ノ	その他	計
新聞コア	7168	10177	8402	1180	5160	910	4998	522	133	1356	16290	146	56442
雑誌コア	4895	6156	5860	479	3126	882	2579	228	128	896	9633	81	34943
書籍コア	5155	6534	6297	658	3651	1187	2375	277	126	818	9826	73	36977
白書コア	3473	6148	3843	3676	2688	636	1417	341	128	675	12450	262	35737
知恵袋コア	2310	2027	2390	130	1586	261	1450	57	93	321	3247	16	13888
プログコア	1967	1957	2386	136	1375	257	1197	118	77	399	3756	28	13653
計	24968	32999	29178	6259	17586	4133	14016	1543	685	4465	55202	606	191640

6 図3は1万語換算したものである。

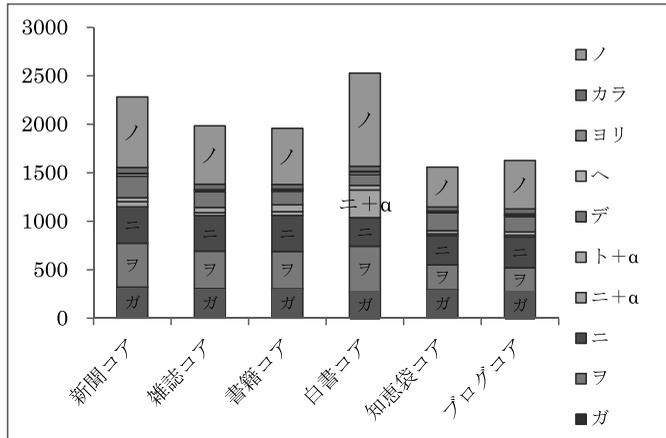


図 3: BCCWJ 格助詞の数 (長単位) (1 万語当たりの数)

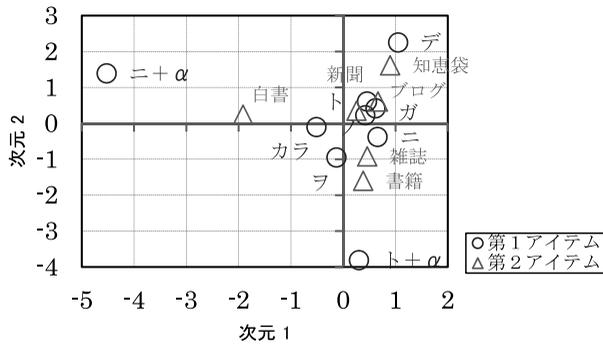


図 4: コレスポネンス分析の散布図 (長単位)

5.2 各格助詞の使用実態

5.2.1 ガ

格助詞ガの使用数, 割合は表 7 の通りである.

表 7: BCCWJ 格助詞ガの数 (短単位)

ガ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ガの数	7235	5000	5372	3545	2375	2018	25545
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	12.68	14.00	14.11	9.68	16.64	14.31	13.04
助詞中の割合 %	8.64	8.35	8.07	7.21	7.81	7.45	8.06
総語数中の割合 %	2.35	2.47	2.63	1.80	2.53	2.18	2.33

(網掛け部分は, 他と比較して多い箇所 (太字) と少ない箇所. 以下同様.)

格助詞中の割合も、助詞中の割合も、総語数中の割合も、すべて白書はガが少ない。(白書は、ニ・ノが多い。)格助詞中の割合を見ると、知恵袋はガが多い。どの割合を見ても、知恵袋はブログよりガが多い。知恵袋とブログの差が表れている。助詞中の割合を見ると、新聞のガが多い。新聞は、助詞中の割合は大きい、格助詞中の割合は小さい。(新聞は格助詞の中ではヲが多い。)総語数中の割合を見ると、書籍のガが多い。

5.2.2 ヲ

格助詞ヲの使用数、割合は表 8 の通りである。

表 8: BCCWJ 格助詞ヲの数 (短単位)

ヲ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ヲの数	10294	6192	6576	6265	2029	1965	33321
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	18.04	17.33	17.27	17.11	14.21	13.93	17.01
助詞中の割合 %	12.30	10.34	9.87	12.74	6.68	7.25	10.52
総語数中の割合 %	3.34	3.06	3.22	3.18	2.16	2.12	3.03

格助詞中の割合も、助詞中の割合も、総語数中の割合も、すべて知恵袋・ブログはヲが少ない。(知恵袋・ブログはデが多い。)新聞は、格助詞中の割合・助詞中の割合・総語数中の割合ともヲが多い。白書もヲが多い。

5.2.3 ニ

格助詞ニの使用数、割合は表 9 の通りである。

表 9: BCCWJ 格助詞ニの数 (短単位)

ニ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ニの数	9826	6645	7361	7948	2641	2643	37064
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	17.22	18.60	19.33	21.70	18.50	18.74	18.92
助詞中の割合 %	11.74	11.10	11.05	16.17	8.69	9.75	11.70
総語数中の割合 %	3.19	3.29	3.61	4.03	2.81	2.85	3.37

ニは、格助詞中・助詞中・総語数中とも、白書に多い。知恵袋・ブログは、助詞中・総語数中においてニが少ない。格助詞中の割合を見ると、新聞はニが少ない。(ヲが多い。)

長単位で調べると、かなり違う数値となる。表 10 と表 11 に示す。白書に多いのは、ニそのものではなく、ニを伴う複合辞であることがわかる。

表 10: BCCWJ 格助詞ニの数 (長単位)

ニ 長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ニの数	8402	5860	6297	3843	2390	2386	29178
格助詞の数	56442	34943	36977	35737	13888	13653	191640
助詞の数	77398	53516	57785	43809	25947	23735	282190
総語数	224140	159883	169730	129646	78770	75242	837411
格助詞中の割合 %	14.89	16.77	17.03	10.75	17.21	17.48	15.23
助詞中の割合 %	10.86	10.95	10.90	8.77	9.21	10.05	10.34
総語数中の割合 %	3.75	3.67	3.71	2.96	3.03	3.17	3.48

表 11: BCCWJ 格助詞ニを伴う複合辞の数 (長単位)

ニを伴う複合辞 長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
ニを伴う複合辞の数	1180	479	658	3676	130	136	6259
格助詞の数	56442	34943	36977	35737	13888	13653	191640
助詞の数	77398	53516	57785	43809	25947	23735	282190
総語数	224140	159883	169730	129646	78770	75242	837411
格助詞中の割合 %	2.09	1.37	1.78	10.29	0.94	1.00	3.27
助詞中の割合 %	1.52	0.90	1.14	8.39	0.50	0.57	2.22
総語数中の割合 %	0.53	0.30	0.39	2.84	0.17	0.18	0.75

長単位で見ると、白書は、ニそのものは、格助詞中の割合も助詞中の割合も総語数中の割合も、全て他に比べて少ない。ニを伴う複合辞は、白書に圧倒的に多く、知恵袋・ブログに少ない。白書は、格助詞ニの 46.25%⁷が複合辞である。新聞が 12%、あとのレジスターは 1 ケタである。白書には、ニオイテが圧倒的に多い。ニオイテは、白書以外は少ない。ニヨリ・ニヨッテは、白書以外はニヨリよりニヨッテが多い。ニ対シ・ニ対シテは、新聞のみニ対シが多い。

5.2.4 ト

格助詞トの使用数、割合は表 12 の通りである。

表 12: BCCWJ 格助詞トの数 (短単位)

ト 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞トの数	6165	4168	5079	3615	1931	1710	22668
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	10.81	11.67	13.34	9.87	13.53	12.12	11.57
助詞中の割合 %	7.36	6.96	7.63	7.35	6.35	6.31	7.15
総語数中の割合 %	2.00	2.06	2.49	1.83	2.06	1.84	2.06

⁷ 表 11 によると、白書におけるニを伴う複合辞が 3676 個、表 9 によると、白書における短単位としてのニの数は 7948 個。割り算すると 46.25% となる。

トは全体にあまり差がない。書籍に若干多く、白書・ブログは若干少ない。
長単位で調べると、表 13、14 のようになる。

表 13: BCCWJ 格助詞トの数 (長単位)

ト 長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞トの数	5160	3126	3651	2688	1586	1375	17586
格助詞の数	56442	34943	36977	35737	13888	13653	191640
助詞の数	77398	53516	57785	43809	25947	23735	282190
総語数	224140	159883	169730	129646	78770	75242	837411
格助詞中の割合 %	9.14	8.95	9.87	7.52	11.42	10.07	9.18
助詞中の割合 %	6.67	5.84	6.32	6.14	6.11	5.79	6.23
総語数中の割合 %	2.30	1.96	2.15	2.07	2.01	1.83	2.10

表 14: BCCWJ 格助詞トを伴う複合辞の数 (長単位)

トを伴う複合辞 長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
トを伴う複合辞の数	910	882	1187	636	261	257	4133
格助詞の数	56442	34943	36977	35737	13888	13653	191640
助詞の数	77398	53516	57785	43809	25947	23735	282190
総語数	224140	159883	169730	129646	78770	75242	837411
格助詞中の割合 %	1.61	2.52	3.21	1.78	1.88	1.88	2.16
助詞中の割合 %	1.18	1.65	2.05	1.45	1.01	1.08	1.46
総語数中の割合 %	0.41	0.55	0.70	0.49	0.33	0.34	0.49

トを伴う複合辞は、書籍に多い。トイウが多い。白書にはトシテが多いが、トイウが少ないために全体としては少なくなっている。

5.2.5 デ

格助詞デの使用数、割合は表 15 の通りである。

表 15: BCCWJ 格助詞デの数 (短単位)

デ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞デの数	5102	2728	2530	1456	1552	1371	14739
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	8.94	7.64	6.64	3.98	10.87	9.72	7.53
助詞中の割合 %	6.09	4.56	3.80	2.96	5.11	5.06	4.65
総語数中の割合 %	1.65	1.35	1.24	0.74	1.65	1.48	1.34

デは、格助詞中の割合・助詞中の割合・総語数中の割合とも、白書には少ない。多いの

は、格助詞中の割合では、知恵袋・ブログ、助詞中の割合では新聞、総語数中の割合では、新聞・知恵袋となっている。

デの代わりに用い得る複合辞ニオイテ・ニヨッテ・トシテと比較してみると、表 16 のような結果を得る。

表 16: BCCWJ 格助詞デと類似の意味の複合辞の比較

長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞デの数	4998	2579	2375	1417	1450	1197	14016
総語数	224140	159883	169730	129646	78770	75242	837411
1 万語当たりの数	223	161	140	109	184	159	167
長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ニオイテの数	12	20	41	556	3	5	637
総語数	224140	159883	169730	129646	78770	75242	837411
1 万語当たりの数	1	1	2	43	0	1	8
長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞トシテの数	365	246	285	466	54	92	1508
総語数	224140	159883	169730	129646	78770	75242	837411
1 万語当たりの数	16	15	17	36	7	12	18
長単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ニヨッテの数	86	61	125	69	35	19	395
総語数	224140	159883	169730	129646	78770	75242	837411
1 万語当たりの数	4	4	7	5	4	3	5

白書はニオイテ、トシテが多い。ニオイテは、白書以外は少ない。トシテは、知恵袋に少ない。

5.2.6 ヘ

格助詞への使用数、割合は表 17 の通りである。

表 17: BCCWJ 格助詞への数 (短単位)

ヘ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞への数	522	228	277	341	57	118	1543
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	0.91	0.64	0.73	0.93	0.40	0.84	0.79
助詞中の割合 %	0.62	0.38	0.42	0.69	0.19	0.44	0.49
総語数中の割合 %	0.17	0.11	0.14	0.17	0.06	0.13	0.14

へは全体に少ない。知恵袋が特に少ない。多いのは白書・新聞である。

5.2.7 ヨリ

格助詞ヨリの使用数、割合は表 18 の通りである。

表 18: BCCWJ 格助詞ヨリの数 (短単位)

ヨリ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ヨリの数	141	137	140	129	97	82	726
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	0.25	0.38	0.37	0.35	0.68	0.58	0.37
助詞中の割合 %	0.17	0.23	0.21	0.26	0.32	0.30	0.23
総語数中の割合 %	0.05	0.07	0.07	0.07	0.10	0.09	0.07

ヨリは全体に少ない。他に比べると、知恵袋、ブログに若干多い。

5.2.8 カラ

助詞カラの使用数、割合は表 19 の通りである。

表 19: BCCWJ 格助詞カラの数 (短単位)

カラ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞カラの数	1362	900	861	675	326	410	4534
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	2.39	2.52	2.26	1.84	2.28	2.91	2.31
助詞中の割合 %	1.63	1.50	1.29	1.37	1.07	1.51	1.43
総語数中の割合 %	0.44	0.44	0.42	0.34	0.35	0.44	0.41

カラは大差ない。白書と知恵袋が若干少ない。

5.2.9 ノ

格助詞ノの使用数、割合は表 20 の通りである。

表 20: BCCWJ 格助詞ノの数 (短単位)

ノ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
格助詞ノの数	16390	9714	9880	12636	3261	3773	55654
格助詞の数	57052	35722	38083	36619	14274	14104	195854
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
格助詞中の割合 %	28.73	27.19	25.94	34.51	22.85	26.75	28.42
助詞中の割合 %	19.58	16.23	14.83	25.71	10.73	13.93	17.57
総語数中の割合 %	5.31	4.80	4.84	6.41	3.47	4.07	5.07

白書はノが多く、知恵袋は少ない。

5.2.10 BCCWJ における格助詞のまとめ

これまでの調査結果をまとめると、以下のようになる。

- ・白書と知恵袋は対照的である。
- ・白書は、ガ・デが少なく、ニ・ノが多い。
- ・白書にニが多いのは、ニを伴う複合辞が多いからである。長単位検索を行うと、単独のニは、むしろ少ないことがわかる。
- ・知恵袋・ブログはデが多い。白書はデが少なく、それは他の複合辞（ニオイテ等）に置き換えられているからであると思われる。デは話し言葉的であり、白書では、デの代わりに複合辞を用いる。
- ・新聞は、ヲが多い。白書もヲが多い。

レジスターによって上記のような違いが現れるのはなぜか。他の項目も調べてみた。ガが少ないということはどういうことかを見るために、似たような環境に現れるハはどうか調べた結果を、表 21 に示す。また、名詞の数、文長を調べた結果は、表 22 に示す。

表 21: BCCWJ 係助詞ハの数 (短単位)

ハ 短単位	新聞	雑誌	書籍	白書	知恵袋	ブログ	計
係助詞ハの数	8956	6350	7316	3577	2780	2655	31634
係助詞の数	11838	9108	10426	4290	4215	4137	44014
助詞の数	83713	59866	66606	49157	30392	27095	316829
総語数	308504	202268	204050	197011	93932	92746	1098511
係助詞中の割合 %	75.65	69.72	70.17	83.38	65.95	64.18	71.87
助詞中の割合 %	10.70	10.61	10.98	7.28	9.15	9.80	9.98
総語数中の割合 %	2.90	3.14	3.59	1.82	2.96	2.86	2.88

表 22: BCCWJ 名詞・助詞・句点の数

短単位	名詞	%	助詞	%	句点 (中分類)	%	総語数
新聞コア	142679	46.25	83713	27.14	12042	3.90	308504
雑誌コア	76812	37.98	59866	29.60	9832	4.86	202268
書籍コア	59822	29.32	66606	32.64	8915	4.37	204050
白書コア	96297	48.88	49157	24.95	4038	2.05	197011
知恵袋コア	26352	28.05	30392	32.36	6953	7.40	93932
ブログコア	31965	34.47	27095	29.21	5536	5.97	92746
計	433927	39.50	316829	28.84	47316	4.31	1098511

白書はハも少ない。白書は、名詞が多く、助詞が少ない。句点の数が少ないことから、文が長いことがわかる。名詞が多いことからノが多くなり、文が長いということから、同じ文字数でも主語に当たるものは少なくなり、ハヤガが少なくなるのだと思われる。知恵袋は句点の数が多く、文が短い。但し、新聞も白書の次に文が長く、名詞も多い。にもかかわらず、新聞においては、ハもガも多い。ここに、新聞と白書の文体の違いがあると考えられる。この究明は、今後の課題としたい。

6. CSJ における格助詞

CSJ における格助詞の内訳は表 23 の通りである。ともに、話者は女性二人、男性二人であり、A11F0703 と D03F0034 は同一話者、以下同様、4 名分のデータである。

表 23: CSJ 格助詞の数 (短単位)

		ガ	ヲ	ニ	ト	デ	ヘ	ヨリ	カラ	ノ	その他	計
学会講演	A11F0703	114	135	157	196	75	0	0	42	205	0	924
	A05F0043	110	60	94	92	20	0	3	7	180	1	567
	A11M0369	88	82	59	93	31	5	0	20	176	2	556
	A11M0469	144	191	154	210	60	2	0	79	271	0	1111
	A グループ計	456	468	464	591	186	7	3	148	832	3	3158
自由会話	D03F0034	47	4	34	71	50	0	1	6	53	0	266
	D03F0058	36	19	28	60	38	0	5	1	30	0	217
	D03M0004	51	20	51	41	35	0	0	7	61	0	266
	D03M0038	100	60	63	87	50	3	0	15	43	1	422
	D グループ計	234	103	176	259	173	3	6	29	187	1	1171
総計	690	571	640	850	359	10	9	177	1019	4	4329	

1 万語に換算すると、表 24 のようになる。更にそれを図にしたものが図 5 である。学会講演と自由会話で、かなり傾向が異なることがわかる。

表 24: CSJ 格助詞の数 (短単位) (1 万語当たりの数)

短単位 (1 万語中)	ガ	ヲ	ニ	ト	デ	ヘ	ヨリ	カラ	ノ	その他	格助詞 全体	
学会講演	A11F0703	202	240	279	348	133	0	0	75	364	0	1640
	A05F0043	313	171	268	262	57	0	9	20	513	3	1614
	A11M0369	282	263	189	298	99	16	0	64	564	6	1783
	A11M0469	213	282	228	311	89	3	0	117	401	0	1643
	A グループ計	240	246	244	311	98	4	2	78	437	2	1660
自由会話	D03F0034	156	13	113	235	166	0	3	20	175	0	881
	D03F0058	155	82	120	258	163	0	21	4	129	0	931
	D03M0004	205	80	205	165	141	0	0	28	245	0	1068
	D03M0038	275	165	173	239	137	8	0	41	118	3	1160
	D グループ計	204	90	153	226	151	3	5	25	163	1	1020
総計	226	187	210	279	118	3	3	58	334	1	1419	

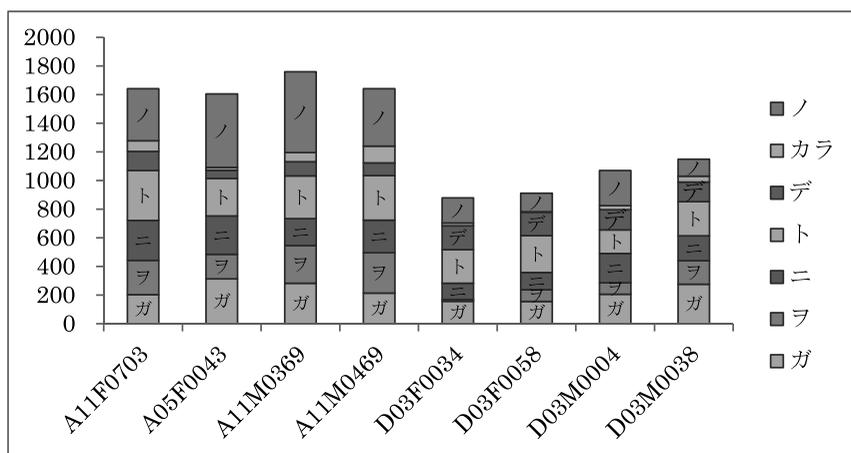


図 5: CSJ 講演・会話の格助詞の数 (短単位) (1 万語当たり)

表 24 の内容をもとに、数の少ないへ・ヨリを除き、ジャンルと助詞の関係をより詳細に把握するため、各助詞を第 1 アイテム、コーパス番号を第 2 アイテムとしてコレスポネンス分析を行った。その結果、下記の散布図を得た (図 6)。なお、第 1 次元の寄与率は 58.52%，第 2 次元の寄与率は 21.81%，2 つの次元による累計寄与率は 80.33% であるため、2 つの次元に基づく解釈に一定の妥当性があると判断した。軸解釈を行うと、第 1 次元は A (学会講演) が -、D (自由会話) が + となって、かたい話し言葉かくだけた話し言葉かできれいに区分されている。第 2 次元は男女で分かれるということもなく、あまり意味がないように思われる。しかし、男性二人については、D03M0038 と A11M0469、D03M0004 と A11M0369 はそれぞれ同一人物であり、縦軸上の同じような位置に存在している。ジャンルを超えた個人差が若干現れているのかもしれない。しかし、個人差より

もジャンル差の方が大きいことは明らかである。自由会話はデが多く、ヲが少ない。学会講演には、ヲとノが多く、デが少ない。

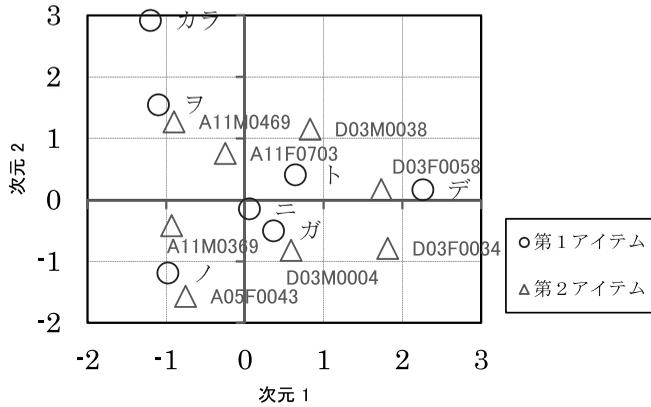


図 6: コレスポネンス分析の散布図 (短単位)

複合辞について調べるため、長単位検索をしてみると、表 25 のような結果を得る。

表 25: CSJ 格助詞の数 (長単位)

長単位		ガ	ヲ	ニ	ニを伴う複合辞	ト	トを伴う複合辞	デ	ヘ	ヨリ	カラ	ノ	その他	計
学会講演	A11F0703	114	134	129	26	66	143	69	0	0	17	205	1	904
	A05F0043	110	60	70	24	58	34	20	0	3	7	180	1	567
	A11M0369	88	79	29	27	74	19	31	5	0	20	176	5	553
	A11M0469	144	189	113	39	70	129	54	2	0	28	271	2	1041
	Aグループ計	456	462	341	116	268	325	174	7	3	72	832	9	3065
自由会話	D03F0034	45	4	31	2	34	12	35	0	1	6	53	0	223
	D03F0058	36	19	27	0	23	17	32	0	5	1	30	0	190
	D03M0004	51	20	51	0	24	23	31	0	0	7	61	0	268
	D03M0038	100	60	57	4	54	37	37	3	0	15	43	1	411
	Dグループ計	232	103	166	6	135	89	135	3	6	29	187	1	1092
総計	688	565	507	122	403	414	309	10	9	101	1019	10	4157	

1万語当りに換算すると、表 26・図 7 のようになる。

表 26: CSJ 格助詞の数 (長単位) (1 万語当たりの数)

長単位 (1 万語中)	ガ	ヲ	ニ	ニを伴 う複合 辞	ト	トを伴 う複合 辞	デ	ヘ	ヨリ	カラ	ノ	その他	格助詞 全体	
学会講演	A11F0703	243	285	275	55	141	304	147	0	0	36	436	2	1925
	A05F0043	414	226	264	90	218	128	75	0	11	26	678	4	2136
	A11M0369	392	352	129	120	329	85	138	22	0	89	784	22	2462
	A11M0469	268	351	210	73	130	240	100	4	0	52	504	4	1935
	Aグループ計	304	308	228	77	179	217	116	5	2	48	556	6	2046
自由会話	D03F0034	167	15	115	7	126	44	130	0	4	22	196	0	826
	D03F0058	177	93	132	0	113	83	157	0	25	5	147	0	932
	D03M0004	234	92	234	0	110	106	142	0	0	32	280	0	1230
	D03M0038	305	183	174	12	165	113	113	9	0	46	131	3	1254
	Dグループ計	228	101	163	6	132	87	132	3	6	28	183	1	1071
総計	273	224	201	48	160	164	123	4	4	40	405	4	1652	

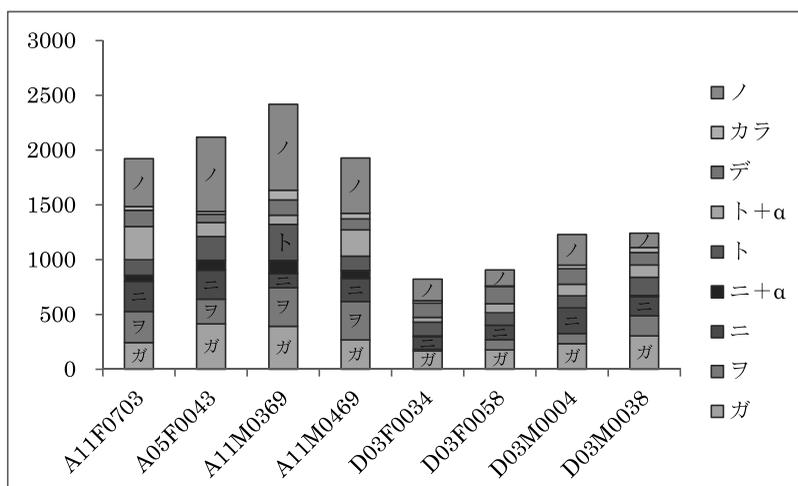


図 7: CSJ 講演・会話の格助詞の数 (長単位) (1 万語当たり)

表 26 の内容をもとに、数の少ないへ・ヨリを除きジャンルと助詞の関係をより詳細に把握するため、各助詞を第 1 アイテム、コーパス番号を第 2 アイテムとしてコレスポネンス分析を行った。その結果、下記の散布図を得た (図 8)。なお、第 1 次元の寄与率は 48.54%、第 2 次元の寄与率は 30.34%、2 つの次元による累計寄与率は 78.88% であるため、2 つの次元に基づく解釈に一定の妥当性があると判断した。軸解釈を行うと、第 1 次元は A (学会講演) が-, D (自由会話) が+にほぼなっており、かたい話し言葉かくだけた話し言葉かで区分されている。ニを伴う複合辞は、学会講演に特異である。

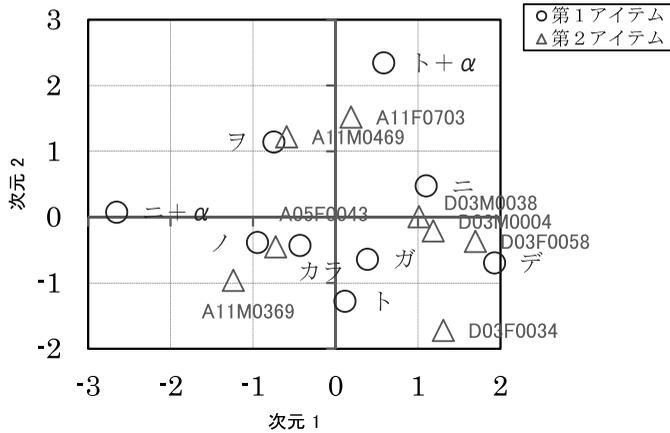


図 8: コレスポネンス分析の散布図 (長単位)

講演は、ニを伴う複合辞とトを伴う複合辞が多い。特に多いのは、トシテ・ニツイテ・ニ関シテである。

7. BCCWJ と CSJ を比較して

書き言葉 (BCCWJ) と話し言葉 (CSJ) を比較すると、ノが多いことや、複合辞が多い等から、学会講演と、かたいタイプの書き言葉 (白書・新聞) が似ていることがわかる。

例えば、ノについては、学会講演は 1 万語中平均 437 語、自由会話は 1 万語中平均 163 語、それに対して、BCCWJ の方は、知恵袋が最も低くて 347 語、白書が最も高く 641 語、新聞が 531 語である。白書 > 新聞 > 書籍 > 雑誌 > 学会講演 > ブログ > 知恵袋 > 自由会話の順である。複合辞については、ニを伴うものの場合、白書は 1 万語中 284 語、学会講演は平均 77 語、自由会話は平均 6 語である。白書 > 学会講演 > 新聞 > 書籍 > 雑誌 > ブログ > 知恵袋 > 自由会話の順となる。

書き言葉 > 話し言葉ではなく、その中のレジスターごとの性質の違いが存在する。

8. 複合辞について

複合辞については、長単位として扱う場合と、長単位にしない場合がある。例えば「の」「のに」は、BCCWJ では、短単位分割でも長単位分割でも準体助詞「の」+ 助動詞「だ」、準体助詞「の」+ 格助詞「に」として扱うが、CSJ では、長単位分割では接続助詞として扱っている。複合辞の扱いは今後の課題である。

9. 複数の分類にまたがるものの扱いについて

例えば「って」は、少なくとも三種に分けられる。「手術って聞いてびっくりした」は格助詞、「～なんですって。」は終助詞、「人生って楽しいことばかりじゃないよ」は係助詞である。しかし、BCCWJ 及び CSJ においては、すべて副助詞として扱っている。形態素解析としては副助詞として扱うとしても、その働きの違いを明らかにする必要がある。

10. まとめと今後の課題

全語数の約 30% が助詞である。助詞の中では格助詞が最も多い。

BCCWJ において、白書と知恵袋は対照的である。白書は、ガ・デが少なく、ニ・ノが多い。白書にニが多いのは、ニを伴う複合辞が多いからである。長単位検索を行うと、単独のニは、むしろ少ないことがわかる。知恵袋・ブログはデが多い。白書はデが少なく、それは他の複合辞（ニオイテ等）に置き換えられているからであると思われる。デは話し言葉的であり、白書では、デの代わりに複合辞を用いる。新聞は、ヲが多い。白書もヲが多い。新聞は多少白書に似た性質を持つ。しかし、新聞は、ガ（係助詞ハも）が多いのに対して、白書はガもハも少ない。この理由の究明は今後の課題としたい。

CSJ においては、話者による違いよりも、学会講演か自由会話かの違いの方が顕著である。講演は、ニを伴う複合辞とトを伴う複合辞が多い。

CSJ の学会講演と、BCCWJ のかたいタイプの書き言葉（白書・新聞）には類似点が見られる。かたい書き言葉 > かたい話し言葉 > くだけた書き言葉 > くだけた話し言葉という段階がありそうである。

助詞の分類の仕方、認定の仕方には課題も残る。さらに、格助詞の担う意味役割（深層格）について、現在分析を進めている。レジスターによって、偏りがありそうである。

文献

- 大曾美恵子（2007）「コーパスに見られる格助詞の運用実態」『姫路獨協大学外国語学部紀要』20, 121-134.
- 松田真希子（2014）「ニ格深層格の定量的分析」『言語処理学会第 20 回年次大会発表論文集』516-519.
- 松田真希子（2015）「Web 日本語 N グラムコーパス分析に基づく深層格の偏りの検証」『計量国語学会第 59 回大会予稿集』43-48.
- 丸山直子（2015）「助詞の使用実態—BCCWJ・CSJ にみる分布—」『第 8 回コーパス日本語学ワークショップ予稿集』179-188.
- 宮内佐夜香（2012）「接続助詞とジャンル別文体的特徴の関連について—『現代日本語書き言葉均衡コーパス』を資料として—」『国立国語研究所論集』3, 39-52.
- 渡部涼子・田中弥生・小磯花絵（2015）「『日本語話し言葉コーパス』UniDic 版形態論情報の構築」『第 8 回コーパス日本語学ワークショップ予稿集』279-288.

（2015 年 8 月 15 日受付・2015 年 9 月 30 日再受付）

Paper B

Usage of Case Particles in Corpus: BCCWJ and CSJ

MARUYAMA Naoko (Tokyo Woman's Christian University)

Abstract:

We studied the use of case particles in modern Japanese, focusing on the difference between written and spoken languages as well as the difference among sub corpus (registers). As the data source, we used the core part (newspapers, magazines, books, white papers, *Chiebukuro* (Wisdom Q&A), blog) of Balanced Corpus of Contemporary Written Japanese (BCCWJ) and four monologues and four dialogues in the Corpus of Spoken Japanese (CSJ). In both BCCWJ and CSJ, 30% of all the words are particles, among which the case particles are the majority. In BCCWJ, white papers and *Chiebukuro* exhibit completely opposite results. White papers and newspapers have more written-language-style, formal expressions such as "ni-oite" instead of more colloquial "de." *Chiebukuro* and blogs are more towards spoken language. Furthermore, *Chiebukuro* has its own unique style. In CSJ corpus, monologues and dialogues showed different characteristics. Finally, we found that some similarity between monologues in CSJ and formal documents (such as white papers) in BCCWJ.

Keywords: case particles, written Japanese, spoken Japanese, BCCWJ, CSJ, correspondence analysis