『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーションの分析 (2)

丸山 岳彦　（国立国語研究所 言語資源研究系）†

Analyzing Variation of Sentence Final Expressions in the BCCWJ (2)

Takehiko Maruyama　(Dept. Corpus Studies, NINJAL)

1　はじめに

本稿では、『現代日本語書き言葉均衡コーパス（以下、BCCWJ と記す）』を利用した現代日本語文法研究および社会言語学的な研究の試みとして、レジスターごとに特徴的に観察される文末表現のバリエーションについて分析を行なう。ここでの分析の前提となる丸山 (2012) では、BCCWJ に含まれる 12 種類のレジスターのテキストを分析し、各レジスターに見られる典型的な文末表現を抽出した。その続編となる本稿では、これらの文末表現が実際にどのような形で用いられているのかについて、単語（短単位）N-gram を用いた集計結果をもとに分析を行なう。

2　BCCWJ の各レジスターに頻出する文末表現

本稿での議論の前提として、丸山 (2012) で示した分析結果を示す。丸山 (2012) では、BCCWJ に含まれる 12 種類のレジスターから約 20,000 文ずつを取り出し、文末位置から逆向きに文字単位の N-gram を抽出して、そこに頻出する文末表現を分析した。その上で、レジスターごとに異なる文末表現が特徴的に観察されることを指摘し、テキストが持つ機能や定形性という特徴が、頻出する文末表現に影響を与えていることを指摘した。表 1 に、5gram 以上に現れた文末表現を集計し、レジスターごとに出現率の高い文末表現をリスト化した結果を示す。なお、文末の「。」は省略してある。ゴシック体は、表 1 中、そのレジスターにしか現れていない文末表現を表す。

表 1: BCCWJ の各レジスターに頻出する文末表現（5gram 以上、上位 10 位）

出版書籍		図書館書籍		雑誌		新聞		白書		教科書	
のである	1.85%	のである	2.11%	ています	1.27%	している	2.32%	っている	9.91%	ましょう	4.08%
ています	1.73%	なかった	1.62%	している	1.00%	っている	1.43%	している	9.08%	**てみよう**	2.54%
っている	1.41%	っている	1.48%	っている	0.94%	れている	1.04%	**なっている**	6.28%	ています	2.09%
している	1.40%	している	1.18%	れている	0.90%	なかった	0.92%	れている	5.41%	れている	2.04%
なかった	1.34%	であった	1.18%	なかった	0.65%	**になった**	0.74%	**となっている**	5.09%	っている	1.57%
れている	1.21%	ています	1.17%	のである	0.65%	していた	0.67%	されている	2.73%	している	1.50%
あります	1.00%	っていた	0.97%	あります	0.61%	**たという**	0.58%	となった	2.29%	**みましょう**	1.37%
であった	0.96%	れている	0.95%	されている	0.46%	ています	0.54%	のである	2.03%	**てみましょう**	1.27%
りません	0.83%	たのである	0.78%	**でしょう**	0.46%	となった	0.49%	**めている**	1.99%	あります	1.25%
なります	0.73%	あります	0.74%	ください	0.45%	**るという**	0.47%	**ものである**	1.90%	りました	0.98%

広報紙		Yahoo!知恵袋		Yahoo!ブログ		法律		国会会議録	
ください	9.07%	**しょうか？**	5.46%	ています	1.72%	**ならない**	21.55%	**ございます**	14.08%
ています	8.46%	**でしょうか？**	5.44%	**いました**	1.48%	**なければならない**	18.79%	**でございます**	12.35%
しています	3.59%	**のでしょうか？**	2.79%	りました	1.21%	**ができる**	13.79%	**おります**	10.98%
てください	2.76%	思います	2.77%	**きました**	1.15%	**ことができる**	13.78%	**ております**	10.84%
あります	2.66%	ています	2.77%	しました	1.11%	**ることができる**	11.95%	思います	8.47%
しました	2.49%	ください	2.75%	思います	0.82%	**しなければならない**	10.62%	あります	7.76%
れました	2.30%	と思います	2.58%	と思います	0.73%	**ものとする**	8.12%	と思います	7.16%
なります	2.17%	てください	2.44%	あります	0.64%	**することができる**	6.88%	**であります**	6.87%
してください	2.05%	りません	2.17%	**ていました**	0.61%	**るものとする**	6.36%	**いと思います**	4.18%
ましょう	1.94%	あります	1.60%	**っています**	0.57%	**準用する**	4.60%	**けでございます**	4.16%

† maruyama@ninjal.ac.jp

3　本稿の目的、分析の対象および方法

丸山 (2012) での分析結果を受け、本稿では、各レジスターで特徴的に見られる文末表現について、単語（短単位）単位による N-gram を抽出し、各文末表現が実際にどのような形で用いられているかを分析する。分析対象とする文末表現は、表 1 の結果を参考にして、図 1 に示す 13 種類を選び、4 つのグループに分類した。

デス形：です。 / でした。 / でしょう。
マス形：ます。 / ました。 / ましょう。 / ません。
質問・依頼形：か。 / か？ / ください。
その他：ている。 / できる。 / ならない。

図 1: 分析対象とする文末表現（13 種類）

BCCWJ 検索サイト「中納言[1]」を用いて、対象とする文末表現を全てのレジスタを対象に検索し、結果をダウンロードして、13 個の KWIC データを得た。検索式の例を図 2 に示す。

```
キー: （書字形出現形 = "です" AND 品詞 LIKE "助動詞%"） AND 後方共起: （書字形出現形 = "。"
AND 品詞 LIKE "補助記号-句点%"） ON 1 WORDS FROM キー IN (subcorpusName="出版・書籍" AND
core="true") OR (subcorpusName="出 版・書 籍" AND core="false") WITH OPTIONS unit="1"
AND tglWords="20" AND tglKugiri="|" AND tglFixVariable="2"
```

図 2: 中納言の検索式の例

次に、各 KWIC データを 11 種類のレジスターごとに分割し、合計 143 個の分析用データを得た。ここで分析する 11 種類のレジスターを、図 3 に挙げる。BCCWJ に格納されている全データのうち、「ベストセラー（OB）」「韻文（OV）」は、分析上の都合から、除外した。

出版 SC：書籍（PB）、雑誌（PM）、新聞（PN）
図書館 SC：書籍（LB）
特定目的 SC：Yahoo!知恵袋（OC）、Yahoo!ブログ（OY）、白書（OW）、
国会会議録（OM）、広報紙（OP）、教科書（OT）、法律（OL）

図 3: 分析対象とするレジスター

さらに、143 個の分析用データから、文末表現から逆向きに単語（短単位）N-gram を抽出した。N-gram の抽出には、田野村忠温氏（大阪大学）が開発・公開している「BNAnalyzer [2]」を利用した。

「BNAnalyzer」は、「中納言」の検索結果（KWIC データ）を入力として、検索キーの前後文脈の N-gram の一覧（1gram〜8gram）を Excel ファイルとして出力するツールである（田野村, 2012）。処理結果として出力される Excel ファイルの例を、図 4 に示す。

[1] https://chunagon.ninjal.ac.jp/
[2] http://www.tanomura.com/research/BNAnalyzer/

	A	B	C	D	E	F
1	1-gram	2-gram	3-gram	4-gram	5-gram	6-gram
2	て(2162)	して(676)	てみて(221)	してみて(79)	ないようにして(28)	ないように注意して(11)
3	で(195)	みて(243)	にして(141)	ようにして(77)	ので、注意して(16)	ので、注意して(7)
4	ご覧(26)	ないで(121)	注意して(85)	を参照して(62)	を参考にして(14)	／ふたつ／３つ(6)
5	_(23)	おいて(63)	参照して(63)	考えてみて(29)	ように注意して(12)	ますので注意して(6)
6	を(16)	考えて(59)	ておいて(60)	に注意して(26)	想像してみて(10)	を想像してみて(6)
7	許し(13)	あげて(58)	てあげて(52)	参考にして(25)	にしてみて(9)	足して４で割って(6)
8	待ち(12)	せて(55)	しないで(41)	で注意して(16)	の見積書をご覧(9)	。_「ちょっと待って(5)
9	注意(11)	見て(54)	をして(35)	覚えておいて(16)	を考えてみて(9)	いうことを忘れないで(5)
10	安心(10)	教えて(47)	と考えて(32)	気をつけて(15)	_「ちょっと待って(8)	五十九を参照して(5)
11	ごらん(9)	やって(30)	させて(30)	試してみて(14)	ことを忘れないで(8)	参考にしてみて(5)
12	せ(9)	いて(29)	を見て(28)	をクリックして(13)	するようにして(8)	くらいお灸をして(4)
13	参照(7)	行って(27)	を教えて(25)	を選択して(13)	ボタンをクリックして(8)	ことを知っておいて(4)
14	聞き(7)	待って(27)	ていて(24)	してあげて(11)	で、注意して(7)	医師の診察を受けて(4)
15	つ(6)	言って(24)	になって(23)	を忘れないで(11)	ページを参照して(7)	楽しみにしていて(4)
16	入り(6)	なって(23)	確認して(22)	見積書をご覧(10)	、気をつけて(6)	気を悪くしないで(4)
17	休み(5)	つけて(22)	ちょっと待って(20)	「ちょっと待って(9)	」を参照して(6)	細心の注意を払って(4)

図 4: BNAnalyzer の処理結果の例（LB 「ください。」に前接する要素）

143 個の分析用データを BNAnalyzer で処理し、Excel ファイルを出力させた。そのうち、前文脈の N-gram を示すシート「前文脈の N-gram」のみを取り出し、1gram から 8gram までの各結果をそれぞれ抽出した。N-gram の表現と出現頻度をタブ区切りで配列し直し、全体で合計 282,171 行の頻度付き N-gram の一覧を得た。集計結果の例を図 5 に示す。以下では、この N-gram データを用いて分析を行なう。

	media	EOS	gram	string	freq
1	LB	kudasai	2gram	して	676
2	LB	kudasai	2gram	みて	243
3	LB	kudasai	3gram	てみて	221
4	LB	kudasai	3gram	にして	141
5	LB	kudasai	2gram	ないで	121
6	LB	kudasai	4gram	してみて	79
7	LB	kudasai	4gram	ようにして	77
8	LB	kudasai	3gram	注意して	65
9	LB	kudasai	3gram	参照して	63
10	LB	kudasai	2gram	おいて	63
11	LB	kudasai	4gram	を参照して	62
12	LB	kudasai	3gram	ておいて	60

図 5: N-gram の集計結果の例

4　分析 1：総文数に占める各文末表現の比率

分析の 1 点目として、分析対象とする 13 種類の文末表現が、各レジスターにおける総文数に占める割合について見る。BCCWJ に格納されている「文」の数をどう捉えるかについては複数の解釈の可能性があり得るが、ここでは、BCCWJ-DVD 版に格納されている「文書構造タグ」において、属性なしの <sentence> タグで囲まれている範囲（すなわち、句点類で終わる「文」に相当する範囲）を収集し、その数を総文数とした。各レジスターごとの総文数を表 2 に、13 種類の文末表現が各レジスターの総文数に占める割合を図 6 に、それぞれ示す。

表 2: 各レジスターの総文数

SC	レジスター	総文数
出版 SC	書籍 (PB)	1,087,715
	雑誌 (PM)	197,069
	新聞 (PN)	54,932
図書館 SC	書籍 (LB)	1,276,651
特定目的 SC	白書 (OW)	95,267
	教科書 (OT)	39,966
	広報紙 (OP)	97,454
	Yahoo!知恵袋 (OC)	582,862
	Yahoo!ブログ (OY)	487,167
	法律 (OL)	17,637
	国会会議録 (OM)	116,022
合計		4,223,682

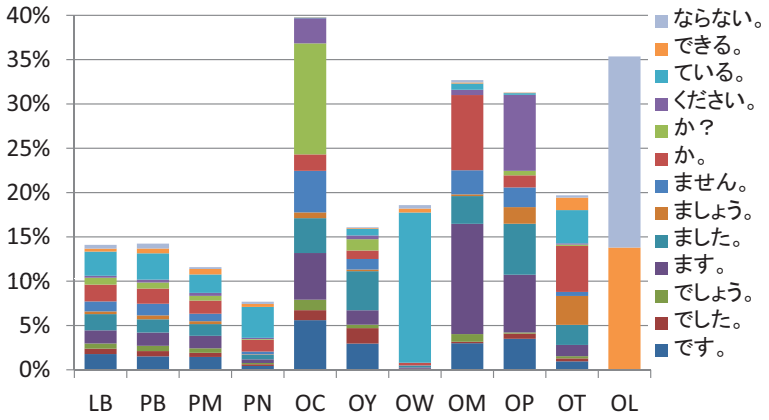

図 6: 各レジスターの総文数に占める各文末表現の比率

以下では、図 6 から読み取れるいくつかの特徴について論じる。

13 種類の文末表現が総文数に占める割合　13 種類の文末表現の合計が総文数に占める割合を各レジスターごとに見てみると、Yahoo!知恵袋 (OC)、法律 (OL)、国会会議録 (OM)、広報紙 (OP) が 30%を超えているのに対して、新聞 (PN)、雑誌 (PM)、書籍 (LB、PB) は 15%を下回っていることが分かる。丸山 (2012) で示した結果と同様、前者のレジスターに含まれるテキストが比較的少ない種類の文末表現によって構成されているのに対して、後者のレジスターに含まれるテキストには多様な種類の文末表現が現れていると見ることができる。

図書館 SC、出版 SC における文末表現　図書館 SC の書籍 (LB)、出版 SC の書籍 (PB)、雑誌 (PM)、新聞 (PN) の結果を見ると、どのレジスターもほぼ同じような分布を示していることが分かる。大半が普通体で書かれる新聞のみ、デス形・マス形の比率が低く、「ている。」の比率が高い、という違いが観察されるものの、出版される書き言葉の文末表現は、ほぼこのような分布を示すのであろう。

特的目的 SC における文末表現　一方、特定目的 SC では、レジスターごとに分布が大きく異なっている様子が見て取れる。特徴的な個所を挙げてみると、以下のようになるだろう。

- Yahoo!知恵袋 (OC) では、「か？」の比率が顕著に高い。これは、質問と回答の組で構成される Yahoo!知恵袋において、疑問符「？」を付けて質問をする表現が多用されるためと考えられる。
- 白書 (OW) では、「ている。」の比率が顕著に高い。これは、国内外の現況について客観的に記述・報告を行なう文体の中で、「ている。」が好まれていることに起因すると考えられる。
- 国会会議録 (OM) では、「ます。」「か。」の比率が顕著に高い。前者は、会議中「ございます。」という文末が多用されるため、後者は、「その理由はなぜですか。」「可能な条件があるのかどうか。」など質問をしたり疑義を呈したりする場合に「か。」が多用されるためと考えられる。
- 広報紙 (OP) では、「ください。」の比率が顕著に高い。「郵便往復はがきでお申し込みください。」「納税通知書をご覧ください。」のように、市民に対する広報記事の中で「ください。」が多用されているためと考えられる。
- 教科書 (OT) では、「ましょう。」「か。」の比率が顕著に高い。それぞれ、「あてはまる数を書きましょう。」「それぞれ何個含まれているか。」のような問題文の中で、多用されていると考えられる。

- 法律 (OL) では、「ならない。」「できる。」のみで、総文数の 35%を占める。丸山 (2012) でも述べたように、法律はある行為を命令・禁止したり、ある権利を保障したりすることを明示的かつ曖昧性のないように述べるためのテキストであり、そのことを示すための文末表現が特徴的に表れているものと考えることができる。

5　分析 2：各文末表現の N-gram と出現率

分析の 2 点目として、13 種類の文末表現が実際にどのような形で用いられているのかについて、N-gram の集計結果を用いて分析する。ここでは、図 5 に示したような N-gram の一覧の中から、3 グラム（3 短単位）以上の連接を分析の対象とする。各 N-gram の出現数をそのメディアの総文数で割って出現率を求め、出現率で降順ソートすることにより、出現しやすい文末表現の用いられ方を集計した。以下、4 グループ・13 種類の文末表現ごとに、出現率で上位 5 位までの結果を示す。なお、例えば「(OM 3)」は、「国会会議録で現れた 3gram の例」であることを表わす。

デス形

表 3: 「です。」の前接要素

と思うの (OM 3)	293 (0.25%)
と思うん (OM 3)	258 (0.22%)
ているわけ (OM 3)	220 (0.19%)
次のとおり (OP 3)	63 (0.06%)
は次のとおり (OP 4)	49 (0.05%)

表 4: 「でした。」の前接要素

ありません (OC 3)	440 (0.08%)
ありません (OY 3)	245 (0.05%)
ありません (LB 3)	519 (0.04%)
ありません (PB 3)	435 (0.04%)
いません (OC 3)	209 (0.04%)

表 5: 「でしょう。」の前接要素

ているわけ (OM 3)	43 (0.04%)
ているの (OT 3)	10 (0.03%)
が、いかが (OM 3)	27 (0.02%)
ているの (OC 3)	124 (0.02%)
ですが、いかが (OM 4)	20 (0.02%)

「です。」の結果の上位 5 位は国会会議録 (OM) と広報紙 (OP) が占めており、特に国会会議録の「思う (の | ん) です。」という形の比率が高い。一方「でした。」の結果を見ると、Yahoo!知恵袋 (OC)、Yahoo!ブログ (OY)、図書館書籍（LB)、出版書籍（PB) という 4 つのレジスターで「ありませんでした。」という形が上位 4 位を占めている。つまり、「でした。」という文末表現は、幅広いレジスターにおいて「ありませんでした。」という形で多用されていると言える。また、「でしょう。」の結果を見ると、国会会議録に現れた「ているわけでしょう。」「いかがでしょう。」、教科書 (OP)・Yahoo!知恵袋に現れた「ているのでしょう。」が上位を占めている。このうち「ているわけでしょう。」以外は、基本的に問いかけとして用いられる文末表現であると言える。

マス形

表 6: 「ます。」の前接要素

わけでござい (OM 3)	1268 (1.09%)
たいと思い (OM 3)	1132 (0.98%)
してい (OP 3)	645 (0.66%)
考えており (OM 3)	632 (0.54%)
お願いし (OC 3)	1428 (0.24%)

表 7: 「ました。」の前接要素

開催され (OP 3)	172 (0.18%)
が行われ (OP 3)	129 (0.13%)
ことにし (OT 3)	46 (0.12%)
してい (OP 3)	94 (0.10%)
と言われ (OC 3)	553 (0.09%)

「ます。」の結果では、国会会議録が多く現れている点が目立つ。「わけでございます。」「たいと思います。」「考えております。」などは、デスマス体で話される国会答弁の中で、多用される文末表現

表 8: 「ましょう。」の前接要素

してみ (OT 3)	73 (0.18%)
考えてみ (OT 3)	69 (0.17%)
調べてみ (OT 3)	60 (0.15%)
ようにし (OP 3)	120 (0.12%)
ないようにし (OP 4)	67 (0.07%)

表 9: 「ません。」の前接要素

かもしれ (OC 3)	1594 (0.27%)
ではござい (OM 3)	297 (0.26%)
ではあり (OC 3)	1321 (0.23%)
ではあり (PB 3)	1717 (0.16%)
なければなり (OP 3)	151 (0.15%)

であると言える。一方、「ました。」の結果では、広報紙が多く現れている。「開催されました。」「が行われました。」のように、区市町村で開催されたイベントを報告するような記事が多く含まれていることを示唆する。また、「ましょう。」の結果では、教科書 (OT)、広報紙が上位 5 位を占めている。教科書では「してみましょう。」「考えてみましょう。」「調べてみましょう。」という形で、読み手である児童・生徒に対する指示が表わされている。一方、広報紙では「(ない) ようにしましょう。」という形で、読み手である区市町村民への呼びかけが行なわれている。

依頼・質問形

表 10: 「か。」の前接要素

みません (OP 3)	463 (0.48%)
てみません (OP 4)	421 (0.43%)
ではない (OM 3)	454 (0.39%)
じゃないです (OM 3)	403 (0.35%)
ありません (OM 3)	379 (0.33%)

表 11: 「か？」の前接要素

なのでしょう (OC 3)	2055 (0.35%)
ないのでしょう (OC 3)	1290 (0.22%)
なんです (OC 3)	1253 (0.21%)
みません (OP 3)	91 (0.09%)
てみません (OP 4)	80 (0.08%)

表 12: 「ください。」の前接要素

を教えて (OC 3)	2533 (0.43%)
提出して (OP 3)	268 (0.28%)
てみて (OC 3)	1252 (0.21%)
はお問い合わせ (OP 3)	200 (0.21%)
をして (OP 3)	196 (0.20%)

「か。」の上位 5 位は、広報紙と国会会議録が占めている。一方、「か？」の上位 5 位は、すべて Yahoo!知恵袋が占めている。広報紙では、区市町村民に呼び掛ける表現として「みませんか。」「みませんか？」の両方が用いられているが、数の上では疑問符よりも句点の方が好まれているようである。一方、Yahoo!知恵袋では質問を投稿する際に「なのでしょうか？」「ないのでしょうか？」「なんですか？」のような形が好んで用いられていると言える。また、「ください。」の上位 5 位は、Yahoo!知恵袋および広報紙が占めている。「教えてください。」「提出してください。」のように、読者や区市町村民に依頼をする表現として用いられている。

その他

「ている。」の上位 3 位は白書 (OW) が占めている。「こととしている。」「となっている。」という文末表現による客観的な描写が、白書で多用されていると言える。一方、「できる。」の上位 5 位は、すべて「〜することができる。」という形が占めている。「動詞+ことができる。」という文末表現が、コロケーションとして用いられていることが示唆される。「できる。」と「ならない。」の上位 3 位は、いずれも法律 (OL) が占めている。分析 1 でも述べたように、ある行為を命令・禁止したり、ある権利を保障したりすることを述べるために、これらの文末表現が多用されていると考えることができる。

表 13: 「ている。」の前接要素

こととし (OW 3)	642 (0.66%)
）となっ (OW 3)	378 (0.39%)
%となっ (OW 3)	344 (0.35%)
」と話し (PN 3)	56 (0.10%)
と考えられ (OT 3)	30 (0.08%)

表 14: 「できる。」の前接要素

することが (OL 3)	1175 (6.66%)
命ずることが (OL 3)	279 (1.58%)
を命ずることが (OL 4)	277 (1.57%)
することが (OT 3)	105 (0.26%)
することが (OW 3)	88 (0.09%)

表 15: 「ならない。」の前接要素

しなければ (OL 3)	1850 (10.49%)
しては (OL 3)	255 (1.45%)
届け出なければ (OL 3)	242 (1.37%)
しなければ (PB 3)	1296 (0.12%)
しなければ (OW 3)	66 (0.07%)

6 分析 3：長い一致を持つ文末表現

最後に、BNAnalyzer が出力する最長の 8gram の例について見てみよう。各レジスターにおいて、8gram の文末表現が総文数に占める比率の高い順に上位 3 位までを抽出した。結果を表 16 に示す。

表 16: 8gram の文末表現（各レジスター上位 3 位まで）

図書館書籍	ていたのではないだろうか。	39 (0.03 ‰)
	しているのではないだろうか。	13 (0.01 ‰)
	について次のように述べている。	12 (0.01 ‰)
出版書籍	といっても過言ではありません。	26 (0.02 ‰)
	ていたのではないだろうか。	24 (0.02 ‰)
	しているのではないだろうか。	18 (0.02 ‰)
雑誌	中から１つ選んでお答えください。	7 (0.04 ‰)
	に満足した？何歳ですか？	6 (0.03 ‰)
	条件をＣＨＥＣＫ！確定申告は必要か？	4 (0.02 ‰)
新聞	郵送かファクスで資料をお寄せください。	4 (0.07 ‰)
	の連絡、資料の返却はしません。	3 (0.05 ‰)
	。落選の方へは連絡しません。	2 (0.04 ‰)
Yahoo!知恵袋	教えてください。よろしくお願いします。	76 (0.13 ‰)
	にはどうしたらいいのでしょうか？	25 (0.04 ‰)
	するにはどうしたらいいですか？	23 (0.04 ‰)
Yahoo!ブログ	ブログは重たいのでこちらからご覧ください。	80 (0.16 ‰)
	心ある対応をよろしくお願い致します。	38 (0.08 ‰)
	と作品の内容は一切関係ありません。	23 (0.05 ‰)
白書	五入のため，合計は百にならない。	21 (0.22 ‰)
	四捨五入のため合計は百にならない。	12 (0.12 ‰)
	．５%）の順となっている。	12 (0.12 ‰)
国会会議録	ますが、御異議ございませんか。	87 (0.75 ‰)
	ますが、御異議ありませんか。	68 (0.59 ‰)
	たいというふうに考えております。	53 (0.46 ‰)
広報紙	※当日、直接会場へお越しください。	49 (0.50 ‰)
	について考えてみませんか。	21 (0.22 ‰)
	ています。詳しくはお問い合わせください。	19 (0.19 ‰)
教科書	１人ぶんは何まいになりますか。	7 (0.18 ‰)
	と，何人に分けられますか。	7 (0.18 ‰)
	どのようにかかわっているのだろうか。	5 (0.13 ‰)
法律	をとるべきことを命ずることができる。	54 (3.06 ‰)
	の物件を検査させることができる。	41 (2.32 ‰)
	、その旨を公示しなければならない。	33 (1.87 ‰)

以下、表 16 に見られる特徴的な点を指摘する。まず、出現率が群を抜いて高いのが法律である。「をとるべきことを命ずることができる。」が 54 回、「の物件を検査させることができる。」が 41 回出現している。これらは、法律文における定形的な言い回しとして用いられている表現と考えられる。同様に、国会会議録の「ますが、御異議 (あり | ござい) ませんか。」「たいというふうに考えております。」も比較的高い出現率を示している。これらも、国会における発言での定形的な言い回し（特に前者は議長や委員長による発言）であると考えてよい。

広報紙では「※当日、直接会場へお越しください。」「について考えてみませんか。」が頻出している文末表現となっている。このうち前者について調べてみると、49 例中 46 例が『広報はままつ』からの例であった。すなわち、特定の広報紙において定形的に用いられている表現であると言える。一方、後者については、さまざまな広報紙が出典となっていた。また、白書の「四捨五入のため合計は百にならない。」という表現は、すべて『中小企業白書』からの例であった。これも、特定の白書において定形的に用いられている表現と言える。

Yahoo!知恵袋の「教えてください。よろしくお願いします。」「にはどうしたらいいのでしょうか?」「するにはどうしたらいいですか？」なども比較的高い出現率を示している。これらは質問を投稿する際の「型」のようなものがユーザー間で共有されているのだろう。

最後に、Yahoo!ブログについて見ると、「ブログは重たいのでこちらからご覧ください。」が 80 回、「心ある対応をよろしくお願い致します。」が 38 回などとなっている。これは、田野村 (2012) が指摘する、同一データが重複して出現している例であると思われる。前者は、宣伝用ブログ記事に含まれるリンクへ誘導するために貼り付けられる文言である。また、後者について調べてみると、「☆乱筆乱文は文学的素養不足の為お許し下さい。今後とも心ある対応をよろしくお願い致します。」という 2 文を、記事の末尾に常に記している書き手によるものであった。

このように見てくると、書き言葉には、（典型的には小説のように）書き手が新たな文章表現を創出する場合だけでなく、ある固定的な文章表現が繰り返し利用される場合も少なからずあることが分かる。後者には、あるレジスターにおいて定形的な表現が好まれて多用されているという場合と、機械的に複製された文章表現が繰り返し出現しているという場合とがあることになる。しかしながら、両者の違いを厳然と区別することは簡単ではない。

そのような重複の問題をノイズとして回避したいという立場がある一方で、例えば、機械的に複製された文章表現が繰り返し出現するのがブログという書き言葉の実態である、という見方もできる。このような問題をどのように捉えるかは、分析を実施する個人に任せられていると言ってよい。

7　まとめ

BCCWJ の 12 種類のレジスターに含まれるテキストを N-gram によって分析し、各レジスターに見られる典型的な文末表現を抽出した。この結果、特に特定目的 SC において、いくつかの文末表現が選考的に多用されていることを明らかにした。

参考文献

丸山岳彦 (2012). 「『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーションの分析」. 『言語処理学会 第 18 回年次大会 発表論文集』, pp. 591–594. 言語処理学会.

田野村忠温 (2012).「BCCWJ に含まれるウェブデータの特性について—データ重複の諸相と BCCWJ 使用上の注意点—」. 『第 2 回 コーパス日本語学ワークショップ 予稿集』.