

「語りかけ性」を有すると判断される書きことばの表現

保田 祥[†] (国立国語研究所 コーパス開発センター)
柏野 和佳子 (国立国語研究所 言語資源研究系)
立花 幸子 (国立国語研究所 コーパス開発センター)
丸山 岳彦 (国立国語研究所 言語資源研究系)

Addressing Expressions in Written Japanese

Sachi Yasuda (Center for Corpus Development, NINJAL)
Wakako Kashino (Dept. Corpus Studies, NINJAL)
Sachiko Tachibana (Center for Corpus Development, NINJAL)
Takehiko Maruyama (Dept. Corpus Studies, NINJAL)

1. はじめに

我々のプロジェクトでは、『現代日本語書き言葉均衡コーパス』(BCCWJ)に収録されている図書館サブコーパス(10,551 サンプル)の書籍サンプルに、人手で文書分類の観点から情報を付与する作業を行っている(柏野・奥村 2012)。

本稿は、「語りかけ性」に着目する。書きことばではあるが、語りかけている印象を与える表現がある。これまで、どのような観点に基づいて「語りかけ性」の有無が判断されるか、複数作業員の判断が一致した 500 例ほどのサンプルを分析し、高頻度に現れた表現の抽出を試みてきた(保田ほか 2012)。しかし、低頻度であっても、判断の決め手となる表現がある。そこで本稿は、約 4,000 例の観点付与結果より、作業員が「語りかけ性」があると判断したサンプルについて分析を行った。作業員が「語りかけ性」を感じる程度の差を見ることで、「語りかけ性」の判断に関わりの強い表現を明らかにする。「語りかけ性」があると判じられたサンプルのジャンルや販売対象との関連性についても報告する。また、作業員が判断時に記載したコメントの分類を行い、作業員が「語りかけ性」があると判じた根拠となる表現をまとめる。コメントのあったサンプルを確かめることで、表現の量や出現位置に見られる傾向も得られた。

2. 「語りかけ性」について

書籍テキストの中には、著者が読者に対して直接語りかけていると解釈できる文体がある。「あなた」や「みなさん」などの呼びかけ表現や、「でしょう」「ではないでしょうか」といった問いかけや相づちを求めるような文末表現など、「直接的な語り」表現と呼べるような表現が含まれるテキストを、本稿では「語りかけ性を有するテキスト」と呼ぶ(柏野 2010)。以下のようなサンプルが、「語りかけ性」があると考えられる例¹である。

お金を稼ぐために事業を始めるべきでないとしたら、なぜ事業を始めるのでしょうか。答えはあなたの情熱と夢にあります。お気に入りの趣味として事業を始めることを考えることができますか。それはほんの少数の人たちにしか理解できない夢です。なぜかって。まず第1に、たいていの人たちがそんなことが可能とさえ思っていないからです。そんなことは、才能があり、お金持ちで、有名でなければできないと彼らは考えているのです。(LBp3_00158 『世界一わかりやすいほんとうのお金持ちになる法』)

保田ほか(2012)では、書籍サンプルからランダムに選び出した約 500 のサンプルを 1

[†] yasuda_s@ninjal.ac.jp

¹ サンプルの出典は、BCCWJ のサンプル ID と書名とで記す。

セットとし、「語りかけ性あり」として3人の作業者の判断が一致したサンプルの分析を行った。品詞、活用形、語彙素において、観点付与を行う作業者が共通して分類に用いている可能性の高い指標が整理されている(表1)。

また、小磯ほか(2011)は、調査者から得た評定語を指標として分析を行っている。そのとき、「書きことば的一話しことば的」という尺度に、「読み手に語りかける一語りかけの少ない」という尺度や、改まりの程度などの複数の観点が発与する可能性を示している。前掲の保田ほか(2012)でも、作業者が「話しことば的」の観点とは異なる指標で判断を行っている可能性が示唆されている。感動詞、融合(「～じゃない」「～なきゃ」など)、終助詞「よ」のように、「話しことば的」と作業者が判断したサンプルで出現率が高くとも、「語りかけ性あり」と判断されたサンプルでは出現率が低いという要素があるためである。作業者による「語りかけ性」の有無の判断は、「話しことば的」との差異を含め、複雑な条件によって行われているものと考えられる。そこで、本稿は「語りかけ性」を有しているとの判断を行うにあたり、作業者が根拠として用いた表現に注目する。

表1. 「語りかけ性」の有無に関わる表現(保田ほか, 2012)

「語りかけ性あり」	
活用形	: 意志推量形(「～でしょう」「～だろう」など)が多い
語彙素	: 「です(助動詞)」「ます(助動詞)」「事(名詞)」が多い 「あなた(代名詞)」「自分(名詞)」「ね(終助詞)」が多い
「語りかけ性なし」	
品詞・語種	: 固有語が多い
語彙素	: 「た(助動詞)」が多い

3. 調査データ

本稿で扱うデータのセットは表2の通りである。

BCCWJの図書館サブコーパスに含まれる書籍をランダムに並べ替え、6人の作業者がアノテーションを行っている。現在までにアノテーション作業が完了しているサンプル9,000のうち、保田ほか(2012)で行った分析以降にアノテーションが為されたサンプル6,000から、テキスト部分が極端に少ないなど²の理由によりアノテーション対象外とされたサンプルを除く3,814のサンプルを分析対象とした。

「語りかけ性」についてのアノテーションにあたっては、「とても語りかけ性がある」「どちらかといえば語りかけ性がある」「特に語りかけ性はない」の三種類の選択肢から該当すると判断した一つを選択する。作業の結果、「とても(語りかけ性)がある」は334サンプル(本稿で扱うサンプルの8.76%)、「どちらかといえば(語りかけ性)がある」が449サンプル(本稿で扱うサンプルの11.77%)得られた。

また、判断時に用いた表現や印象が備考欄へコメントとして記述されている場合がある。作業者によって記述量にはばらつきがあるが、「語りかけ性」に関しては作業者ごとにそれぞれの作業サンプル数の2%~5%のコメントを得た。

表2. 分析対象データ

語りかけ性の有無	とてもある	どちらかといえばある
サンプル数	334	449
語数	1,009,172	1,350,107

² 対談、Q&A形式、図解、用語解説など形式的に特徴のあるサンプルは、分類対象外(非対象)とされた。作業者は、分類対象としたサンプルのみ観点付与を行っている。

4. 「語りかけ性」を有すると判断される書きことばの表現

4.1 分析手法

まず、作業者が、「語りかけ性」があると判断したサンプルに含まれた表現について、特徴的に見られるはずの表現（保田ほか、2012）を確認し、「とても語りかけ性がある」「どちらかといえば語りかけ性がある」と判断されたそれぞれのサンプル群の差異を確かめる。

形態素解析にあたっては、MeCab 0.98+UniDic 1.3.12 を用いた。品詞・活用形・語彙素のそれぞれの分析については、形態素解析の結果に基づく。

「語りかけ性」が「とてもある」と判断されたサンプル群に多く確認できる表現は、「語りかけ性」を明らかとするものと考えられ、「どちらかといえばある」と判断されたサンプル群にも確認できるはずである。しかし、判断レベルに差が生じる場合には、「語りかけ性」が「どちらかといえばある」と判断されたサンプルでは、判断基準とされた表現の出現頻度が低いことが予測される。

次に、「語りかけ性」を有すると判断されたサンプルの属するジャンルと販売対象の関連性を調べ、ジャンル等による特性を見る。同様に、「とてもある」「どちらかといえばある」の差異も確かめる。

さらに、作業者が自ら判断の根拠とした旨を記載したコメントを分類し、作業者が根拠とした表現が、実際のサンプル群にどのように現れているのかを調べる。

4.2 「語りかけ性」を有すると判断されたサンプルに含まれる表現

品詞と活用形、語彙素について、保田ほか（2012）で抽出された「語りかけ性あり」「語りかけ性なし」の二種類のアノテーション結果に見られた表現との対照を行い、「語りかけ性」が「とてもある」「どちらかといえばある」の差異を見る。

4.2.1 品詞

「語りかけ性あり」と判断されるサンプル群において、「語りかけ性なし」と判断されるサンプル群との有意差（有意水準 0.1%以下）の見られる品詞は、「ね」「よ」などを含む終助詞である。保田ほか（2012）では、「語りかけ性あり」サンプルでは、終助詞がサンプル中に 0.5%程度の出現率であり、「語りかけ性なし」サンプルでは 0.2%に留まる。本稿の調査においても、同程度の出現率が期待される。

本稿の調査では、「語りかけ性」が「とてもある」における終助詞の出現率は 0.69%で、「どちらかといえばある」における出現率は 0.41%であった。「語りかけ性」があると判断されるサンプルにおいては、終助詞の出現頻度が高いことが確かめられたといえる。

また、「とてもある」と「どちらかといえばある」では、終助詞の出現率に 1.7 倍の差異が見られている。そのため、「語りかけ性」があるという判断の根拠として、終助詞が利用されている可能性の高いことが考えられる。

4.2.2 活用形

「語りかけ性あり」と判断されるサンプル群には「でしょう」「だろう」などの意志推量形が 0.46%の出現率で見られ、「語りかけ性なし」とされたサンプル群における 0.24%という出現率とは有意差がある（保田ほか、2012）。

本稿の調査では、「とてもある」群における意志推量形の出現率は 0.41%、「どちらかといえばある」群では 0.31%であった。

但し、「どちらかといえばある」群では、「語りかけ性なし」とされたサンプル群の出現率に近い。意志推量形の出現頻度が低いことが「語りかけ性」があることを妨げず、「語りかけ性」があることに関わっているとは考えられる。しかし、「語りかけ性」があるという判断の根拠として積極的に用いられているとも言い切れないことが明らかとなった。

4.2.3 語彙素

「語りかけ性あり」と判じられるサンプルに有意に出現すると期待される語彙素について、本稿の調査との対照を行った。結果は表3である。

「語りかけ性」の有無に関わると考えられる語彙素は、本稿の調査データにおいても類似した出現率が見られている。また、「とてもある」と「どちらかといえばある」では、「語りかけ性がある」サンプル群に見られる表現の中でも、「とてもある」に顕著である。

但し、「自分」については、語りかけ性が「どちらかといえばある」サンプル群において、「語りかけ性がない」サンプル群とほぼ等しい出現率であり、「語りかけ性」があるという判断を妨げるのではないが、積極的に関係しているとも言い難いことがわかった。

表3. 語りかけ性の有無に関わると考えられる語彙素と本稿の調査における出現率

語彙素	品詞	とてもある	どちらかといえばある	ある(保田ほか, 2012)	ない(保田ほか, 2012)
た	助動詞	2.03%	2.14%	1.65%	3.41%
ます	助動詞	1.69%	1.03%	1.35%	0.09%
です	助動詞	1.15%	0.64%	1.19%	0.05%
事	名詞-普通名詞-一般	0.81%	0.73%	1.05%	0.55%
自分	名詞-普通名詞-一般	0.17%	0.11%	0.25%	0.10%
ね	助詞-終助詞	0.13%	0.06%	0.09%	0.01%
よ	助詞-終助詞	0.13%	0.06%	0.04%	0.01%
貴方	代名詞	0.07%	0.03%	0.04%	0.01%

4.3 「語りかけ性」を有するサンプルと図書分類コード

ジャンル (NDC コード) と販売対象 (C コード) を確認しておく。

本稿の調査対象とした「語りかけ性」を有すると判断されたサンプル群の NDC コードと C コード別の分布は、以下の図1・2の通りである。それぞれ、「語りかけ性」が「とてもある」「どちらかといえばある」とアノテーションされたサンプルをあわせたグラフである。

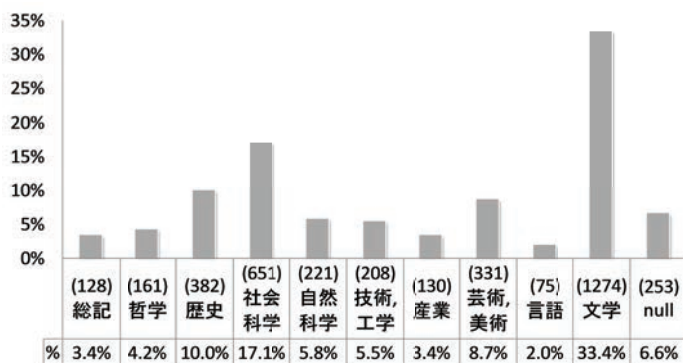


図1. 語りかけ性があると判断されたサンプルのNDCコード別分布

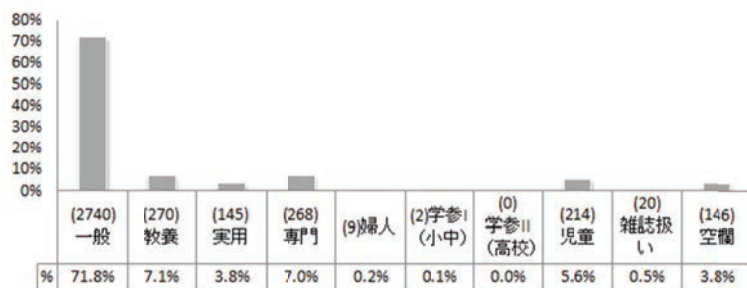


図2. 語りかけ性があると判断されたサンプルのCコード別分布

図1において文学の割合が多いのは、幼児・児童向けのサンプルが含まれるためであり、図2における「児童」を販売対象とした書籍の割合と関係があることがわかる。NDC コードの分布から、全分野に「語りかけ性」が見とめられるのだといえよう。

図2の販売対象分布では、「児童」のほか「実用」が特徴的である。実用書は、いわゆるハウツーものの可能性が高いためである。

次に、「とても語りかけ性がある」と「どちらかといえば語りかけ性がある」の差異を見たい。選択肢毎に、図3はNDCコード、図4はCコードの分布を示している。

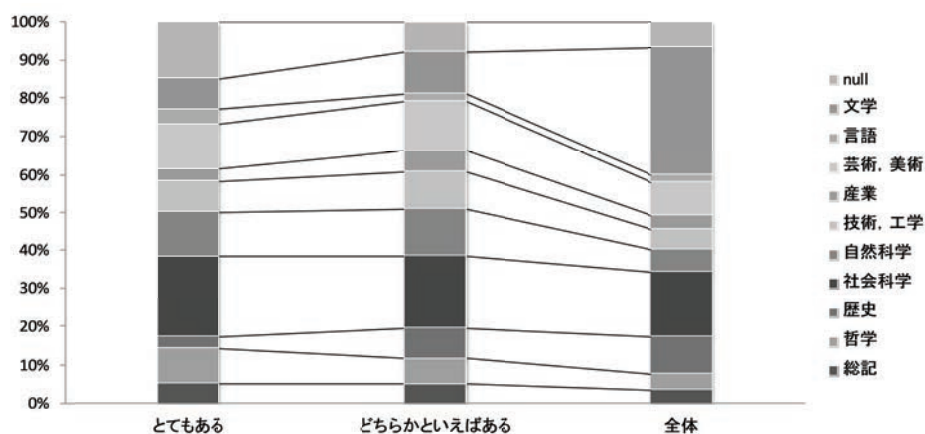


図3. 語りかけ性があると判断されたサンプルのNDCコード別分布（選択肢別）

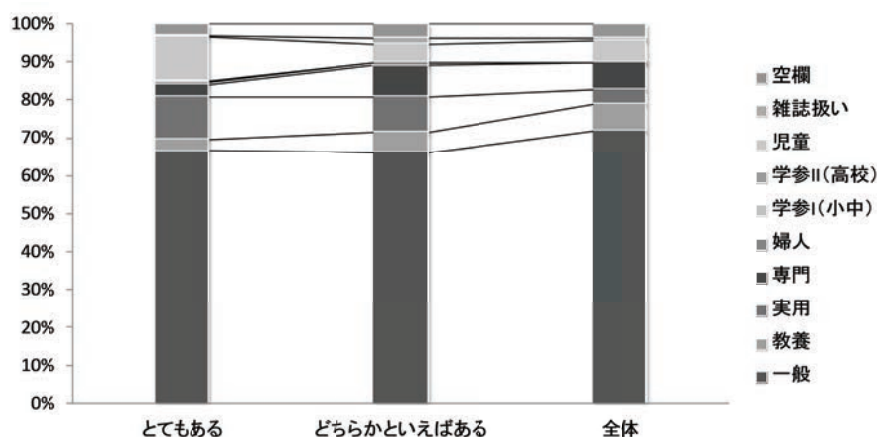


図4. 語りかけ性があると判断されたサンプルのCコード別分布（選択肢別）

図3では、語りかけ性が「とてもある」サンプル群は、哲学でのみ「どちらかといえばある」に比して割合が大きい。哲学分野には啓蒙書が含まれることから、明らかに語りかけている書籍に対して「とてもある」と判断されることが予測される。

また、図4を見ると、語りかけ性が「とてもある」サンプル群は、「どちらかといえばある」に比べ、「実用」と「児童書」の割合が多いことがわかる。従って、「教養」「専門」などでは、「どちらかといえばある」の割合が高い。作業者は語りかけ性があると判断しても、「とてもある」とまでは言い難い場合には、「どちらかといえばある」を選択するため、このような分布になると考えられる。

なお、Cコードで「実用」と分類されていない書籍であっても、「語りかけ性」があると判断される書籍の種類には、いわゆるハウツーものが多いことが明らかである。「語りかけ性」があるとされた書籍は、「とてもある」で、その53%がタイトルに「入門」「～法」「～使い方」「～術」「～ハンドブック」「～がわかる本」「～マニュアル」などが含まれ、「どち

らかといえはる」では同様に 30%、「語りかけ性」があるとされるサンプル群全体で 40% 以上であった。タイトル例は以下のようなものがある。

『気持ちよく家族を介護するための本』『能力と意欲を伸ばす積極育児法』
『基礎から楽しく学べるはじめての陶芸入門』『はがきの書き方』
『資産ゼロから大成功する「魔法の粉」の使い方』『消化器ガン克服マニュアル』
『居住者のためのマンション管理の手引き』『新カラーコーディネート術』
『動物の見つけ方、教えます!』『定年後は山歩きを愉しみなさい』など

NDC コードや C コードの分類には現れないが、啓蒙書や指導書であることが示されているといえる。

4.4 「語りかけ性」に関する作業コメント

作業者は、「語りかけ性」が「どちらかといえはる」という判断を行った際、コメントを記述する傾向がある。「とてもある」と判断されたサンプルにコメントがある場合は、「明らかにあなたに語りかけている体」などにとどまっている。「とてもある」とまでは言い難いが、「語りかけ性」があると感じた場合に「どちらかといえはる」を選択し、その判断根拠を示したものと考えられる。

挙げられた判断根拠は表 4 にまとめた。「昔話」は内容を示すが、コメントのあったサンプルは、「むかしむかし」で始まるいわゆる物語調のテキストである。表現としては、読み手との一体感を生じる「私たち」「われわれ」のような人称のほか、主として文末表現が目されている。また、「多い」と感じたときとされる文末表現以外でも、読み手や読み手の判断を想定した表現が現れると、「語りかけ性」があるとの印象になることがわかる。

表 4. 「語りかけ性」があると作業者が判断した根拠（作業コメント）

- ・ 昔話
- ・ ハウツーものに見られる表現：
 - 希望：「～てください」「～てほしい」「～したい」
 - 注意・禁止：「～ように」「～に注意」「～はいけない」
 - 勧誘：「～てごらん」「～しよう」
 - 可能：「～はず」「～できます」
 - 評価：「～が大切」「～を要する」「～が便利」「～が狙い目」
(その他「よい」「悪くない」などの評価表現もコメントあり)
- ・ 人称：「私たち」「われわれ」が多い
- ・ 文末表現：
 - 「のである」「わけです」「からです」「ものです」などが多い
 - 読み手を想定した表現：
 - 「いただく」「申し上げる」「～の方 (かた)」「ごぞいます」
 - 「あげる」「させてください」「お聞かせすることにしよう」
 - 婉曲表現：
 - 「～と思う (見える/感じ)」「～でしょうか」「～だろう」
 - 読み手の判断を想定した表現：
 - 「思うかもしれません」「おわかりのことでしょう」

前述した通り、「語りかけ性」があると判断されたサンプルは、いわゆるハウツーもの書籍が 4 割を占めるため、ハウツーものに出現する表現に関するコメントが多く見られた。ハウツーもののテキスト例は以下のようなものが典型的であるといえる。

桜の花の塩漬けは、デパートにでもどこにでも売っています。あられは日本橋にぶぶづけ用の、塩だけで味をつけて炒り上げたかわいいあられを売っている店があり、これなら最適ですが、要するにしょうゆを使わない塩味のあられか、かきもちならなんでもいいのです。

ご飯のおこげを形よく包丁して、塩をふっても間に合います。ただ飯粒が乱れてとぶと、いかにも洗い流しのように気持ちが悪いですから、焦げていない反対側のほうもちょっと焦げ目のつく程度に焼いて使しましょう。(LBr5_00039『季節を料理する』)

なお、「多い」「頻出する」という作業者の印象や、作業者が判断根拠とした表現が、実際のサンプルにどのように現れていたのかを見ておく。

表5は、文末に「多い」ことを作業者が判断根拠にしたという表現の、本稿におけるデータ中に現れた数である。また、一文が平均的な文長と仮定したとき(30形態素)の、文末に出現する頻度を算出した。「語りかけ性」が「とてもある」サンプル群では約12%、「どちらかといえばある」サンプル群では約10%の文末に上り、作業者の印象は「多い」に至るものと考えられる。

また、「です」「ます」の出現率が、「とてもある」と「どちらかといえばある」の差異であるため、常体と敬体の出現率が異なっている。

表5. 語りかけ性があるとの判断根拠にされた文末表現の出現率

	とてもある	文末出現率	どちらかといえばある	文末出現率
のだ	148	0.44%	556	1.24%
のである	75	0.22%	635	1.41%
のです	2,183	6.49%	1,485	3.30%
わけだ	16	0.05%	53	0.12%
わけである	2	0.01%	34	0.08%
わけです	246	0.73%	165	0.37%
からだ	23	0.07%	121	0.27%
からである	15	0.04%	93	0.21%
からです	219	0.65%	144	0.32%
ものだ	45	0.13%	93	0.21%
ものである	18	0.05%	113	0.25%
ものです	323	0.96%	240	0.53%
ことだ	37	0.11%	128	0.28%
ことである	27	0.08%	177	0.39%
ことです	460	1.37%	243	0.54%
計	3,837	11.4%	4,280	9.5%

作業者のコメントとして、語りかけている印象のある表現の量がサンプル全体のテキスト量に対して「少ない」か、あるいは「一部分」であったため、サンプル全体としては「語りかけ性」がないという判断をする旨の記載が散見されている。「一部分」には、「冒頭のみ」「文章末のみ」の追記のある場合も見られる。

すなわち、作業者が「語りかけ性」があるとの根拠に用いる表現の位置は、「文章末のみ」である場合に、「どちらかといえばある」を選択した旨が記述された例もあったが、コメントの記載されている場合には、ほぼ「文章末のみ」に語りかけている部分があったものの、全体としては「特になし」を選択したとの内容である。反対に、「冒頭のみ」であると記載されている場合には完全に「特になし」が選択されている。「文章末のみ」がサンプル全体の印象に関わる可能性はある。しかし、判断根拠とする表現が見つかったとしても、サンプル全体のテキスト量に対して「少ない」との印象があれば、「語りかけ性」があるとは判断しない傾向があるということがわかった。

もっとも、アノテーションの行われたサンプルは書籍の全文ではない。サンプル全体における表現の量を判断根拠とする傾向があるのならば、以降に「語りかけ性」のない文が続いている「冒頭のみ」よりも、サンプル内の文字数によっては、「語りかけ性」のない文

の量が予測できない「文章末のみ」に見つかった場合のほうが、「語りかけ性」があるとの判断を行う頻度が上がる可能性がある。

5. まとめ

本稿では、「語りかけ性」についてアノテーションが行われた 3,814 のサンプルのうち、作業によって「語りかけ性」があると判断されたサンプル群の分析を行った。

特に、「語りかけ性」が「とてもある」「どちらかといえばある」の程度の異なる判断が為されたサンプル群の対照を行った。その結果、「語りかけ性」があると判断されたサンプル群に出現率の高い表現が、「とてもある」と積極的に判断根拠とされているのかどうかを確認することができた。「とてもある」に出現率が高い表現でも、「どちらかといえばある」では「語りかけ性なし」サンプルに近い出現率となっていることがある。この場合、「語りかけ性」があるサンプルに特徴的な数値であるとしても、判断根拠とは言い難い。

また、アノテーション作業者が判断の根拠とした旨を記載したコメント欄を分析し、同時に実際のサンプルを確かめることを行った。出現率が低くとも、「語りかけ性」があると判断する根拠となる表現がどのようなものかが明らかとなった。

謝 辞

本研究は、国立国語研究所の共同研究プロジェクト「テキストの多様性を捉える分類指標の策定」に基づくものです。また、BCCWJの構築は、文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築：21世紀の日本語研究の基盤整備」（平成18～22年度、領域代表者：前川喜久雄）による補助を得たものです。

文 献

- 柏野和佳子(2010)「「直接的な語り」という表現スタイルをもつ書籍テキストの人手抽出の試み」『ことば工学研究会』35, pp.63-72.
- 柏野和佳子, 奥村学(2012)「書籍テキストへの分類指標人手付与の試み—『現代日本語書き言葉均衡コーパス』の収録書籍を対象に—」『言語処理学会第18回年次大会』B5-6.
- 小磯花絵, 田中弥生, 小木曾智信, 近藤明日子(2011)「評定実験に基づくテキスト分類尺度の体系化の試み」『現代日本語書き言葉均衡コーパス』完成記念講演会予稿集, pp.47-52.
- 保田祥, 柏野和佳子, 立花幸子, 丸山岳彦(2012)「「語り性」を有する書きことばの典型例の分析」『第1回コーパス日本語学ワークショップ』予稿集, pp.139-146.

関連 URL

国立国語研究所の言語コーパス整備計画 KOTONOHA <http://www.ninjal.ac.jp/kotonoha/>
特定領域研究「日本語コーパス」 <http://www.tokuteicorpus.jp/>