

令和 2 年 6 月 19 日現在

機関番号：62618

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K02800

研究課題名(和文) 学術的文書作成のための文体差のある語の計量的分析

研究課題名(英文) Quantitative analysis of stylistic differences between words for academic writing

研究代表者

柏野 和佳子 (KASHINO, Wakako)

大学共同利用機関法人人間文化研究機構国立国語研究所・音声言語研究領域・准教授

研究者番号：50311147

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：「書き言葉的」な語で記述すべき学術的文章(レポート, 論文等)に「話し言葉的」な語が混じるという問題を解決するために、「書き言葉的」「話し言葉的」といった注釈のある語を作文技術に関する文献等から2,141語抽出した。それらに、品詞とコーパスの頻度情報を新たに付与した。また、コーパスに基づき、代表的な書き言葉的および話し言葉的な語の対に対して、場面別での使用頻度を詳細に明らかにした。

研究成果の学術的意義や社会的意義

「書き言葉的な語」と、「話し言葉的な語」との文体差をコーパスに基づいて計量的に明らかにする点が、学術的な特色である。学術的文章作成という目的に沿う「書き言葉」の、可能な限り体系的、網羅的な収集を目指す点が、これまででない独創的な点である。

文体差のある語のデータベースを整備することにより、関連する語全体を見渡しながら学術的文章作成時に留意すべき語を十分に理解することが可能になる。本研究成果は作文教育や日本語教育、また、自然言語処理における文章生成という応用場面においての活用が期待できる。

研究成果の概要(英文)：To solve the problem of "spontaneous" words being mixed in academic texts (reports, articles, etc.) that should be described in writing-style words, the words with annotations such as "writing style" or "spontaneous" were extracted from literature on composition techniques. For each of those 2,141 words, we newly annotated part-of-speech and word frequency information appearing in corpora. In addition, frequency of typical writing style and spontaneous words was clarified in various scenes based on the corpus.

研究分野：言語学・日本語学

キーワード：文体 書き言葉 話し言葉 コーパス 位相 学術的文章 文章作成 日本語教育

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年、「話し言葉的な語」を多用する SNS 等の影響もあり、「書き言葉的な語」で記述すべき学術的文章(レポート, 論文等)に、「話し言葉的な語」が混じるという問題が大学教育や日本語教育の現場でしばしば指摘されている。しかしながら, 学術的文章で用いるとよい語を網羅的に調査する文献や先行研究は残念ながら見当たらなかった。そこで, 日本語の作文技術に関する文献及び, 書き言葉と話し言葉の相互関係に関する文献を広く調査した(柏野ほか 2016)。書籍約 170 冊, 論文約 290 本を収集し, 書籍 32 冊, 論文 31 本より「書き言葉的」と「話し言葉的」として示される語・表現を抽出した。その結果, 抽出した語は, ペア 1,172 語, 単独:書き言葉的 581 語, 単独:話し言葉的 364 語であった。文献により着目する段階によって, 同じ語が「書き言葉的」とされたり, 「話し言葉的」とされたりしていた。それらの重複を除いた異なり語は約 1,900 語であった。

得られた語に使用の目安を客観的に付与するためには, コーパスに基づく計量的分析が欠かせない。また, 文献から抽出した語の品詞情報にゆれがあったため, その修正も課題として残っていた。

柏野和佳子・田嶋明日香・平本智弥・木田真理(2016)「学術的文章作成時に留意すべき「書き言葉的」「話し言葉的」な語の文献調査」『言語処理学会第 22 回年次大会発表論文集』pp.1041-1044.

### 2. 研究の目的

「書き言葉的な語」と「話し言葉的な語」とされる語を対象に, 大学生や日本語学習者の作文, 論文データベース, 各種会話コーパスを用いてその文体差を計量的に分析し, 学術的文書作成のための語の使用指針を策定することが目的である。従来, 主に文体差のあるものとして議論されてきた接続・副詞・文末表現以外に, 文体差が問題となるような形容詞, 名詞等も収集し, 分析する。

### 3. 研究の方法

#### (1) コーパスに基づく頻度情報の付与

これまでに抽出したデータから見出し語および言い換えのペアを抽出し, 表記のゆれ, 単位のゆれを出来る限り統一する。

以下に挙げるコーパスから頻度情報を抽出し, 付与する。

- ・現代日本語書き言葉均衡コーパス (BCCWJ)
- ・科学技術論文

#### (2) 追加の文献調査と品詞情報の付与

追加の文献調査を行い, 2,141 語のリストを得た。

2,141 語の品詞情報の見直しと付与を行った。語ではなく表現の場合は, 「品詞 + 品詞 + 品詞」や「連語 (品詞 & 品詞 & 品詞)」といった形で記述した。また, 特に長いものについては細かな分解を避けて, 単に「名詞句」「動詞句」などとした。

### (3) コーパスに基づく計量的分析

文体差のある語を対象に、『現代日本語書き言葉均衡コーパス(BCCWJ)』ほか、中学校・高校の教科書類、日本語教育の教科書、ベトナム人学習者の論文、日本語学習者コーパス等を用いた計量的分析を行った。

## 4. 研究成果

### (1) コーパスに基づく頻度情報の付与

『現代日本語書き言葉均衡コーパス(BCCWJ)』のレジスター別頻度および、科学技術論文データを用いて、図1のような頻度表を作成した。これにより、文体差を計量的に把握することができる。

表記	B	C	D	E	F	G	H	I	J	K	L	M	N
	Yahoo ブログ	Yahoo 知恵袋	国会会 議録	広報紙	新聞	書籍	検定教 科書	法律	白書	科学技 術論文	雑誌	韻文	
～いるんだ	356	241	230	32	9	2892	28	0	0	9	148	2	
～上で	520	626	995	104	71	5088	52	21	608	2312	252	13	
～おり	4164	2415	26225	832	306	19557	236	14	7249	8025	907	78	
～が	267242	345213	122766	48332	23350	1639169	22387	12618	86553	357874	109321	3353	
～か否か	45	73	52	4	4	1180	3	16	101	1358	19	0	
～化する	161	171	275	88	70	2999	45	5	519	2990	144	1	
～かどうか	812	1708	2151	98	93	7879	85	55	112	3227	381	9	
～かもしれない	1460	1200	273	27	99	15510	39	0	16	582	854	14	
～から	67012	62488	37187	16735	4704	376852	4927	2428	16726	69004	24797	737	
～くせに	160	174	1	1	2	888	2	0	0	14	39	2	
～くて	37	13	0	0	0	10	0	0	0	12	0	1	
～けど	536	241	0	0	0	616	0	0	0	57	32	1	
～けれども	120	99	45	3	6	3237	20	0	2	106	142	25	
～けれども	53	21	45	1	4	1896	16	0	2	65	60	15	
～させられた	48	32	11	1	8	613	6	0	4	8	25	0	
～されてしまった	43	44	12	3	3	481	2	0	0	75	17	0	
～し	136078	170828	70370	47904	12544	895578	13826	10842	75172	290237	55689	1817	
～しか	3782	5657	764	110	182	17772	95	0	79	2414	1098	60	
～したって	77	98	34	0	0	541	0	0	0	2	24	0	
～したのだが	35	2	1	0	2	338	1	0	0	5	20	3	
～したんだが	5	0	2	0	0	28	0	0	0	0	3	0	
～しちゃった	255	109	5	2	1	308	0	0	0	0	40	2	
～していない	585	1087	415	329	76	4499	34	90	344	2079	282	0	
～している	7952	10072	5273	2440	1246	76641	1012	386	12529	28346	4897	91	
～してく	16	9	0	1	0	15	0	0	0	8	3	0	
～してしまった	399	326	63	17	13	2582	4	0	5	141	150	6	

図1 文体差のある語の頻度調査結果(一部)

### (2) 追加の文献調査と品詞情報の付与

- ・書籍12冊、論文59件、ネット記事3件の調査結果を追加した。
- ・品詞情報を図2のように付与した。

B	C	D	E	F	G	K	L	M	N	O	P	Q
書名	著者名	出版社	出版年	章	ページ	品詞・言葉の種類						
							読	ペア:話し言葉的	品詞	読	ペア:書き言葉的	品詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	接続助詞	から	から	接続助詞	ので	ので	接続助詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	接続助詞	して	して	動詞(連用形)+接続助詞	し	し(連用中止法)	動詞(連用形)
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	接続助詞	しないで	しないで	動詞(連用形)+助動詞+接続助詞	せずに	せずに	動詞(未然形)+助動詞+接続助詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	接続助詞	したら	したら	動詞(連用形)+助動詞(仮定形)	すれば	すれば	動詞(仮定形)+接続助詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	接続助詞	のに	のに	接続助詞	にもかかわらず	にもかかわらず	接続助詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	接続助詞	けど	けど	接続助詞	が	が	接続助詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	副詞	ぜんぜん	全然	副詞	まったく	まったく	副詞
この1冊できちんと書ける! 論文・レポートの基本	石黒圭	日本実業出版社	2012	論文の文体	127	副詞	たぶん	多分	副詞	おそらく	おそらく	副詞

図2 文体差のある語の品詞情報付与結果(一部)

(3) コーパスに基づく計量的分析

『現代日本語書き言葉均衡コーパス(BCCWJ)』には、「国会会議録」という話し言葉を転記した資料や、話し言葉的な語が多く用いられる「yahoo!知恵袋」「yahoo!ブログ」の用例が収録されている。たとえば、書き言葉的な「次第に」と、話し言葉的な「どんどん」の使用を比較すると、それら三つや雑誌において「どんどん」の使用率が高いことが図3からわかる。しかしながら、図4に示すとおり、書き言葉的な「重要」と話し言葉的な「大事」の比較では、国会会議録では「重要」の使用率が高いという結果が出ている。つまり、公的な場面の話し言葉では、書き言葉的な語の使用率が高い場合もあることを明らかにした。(柏野 2018)

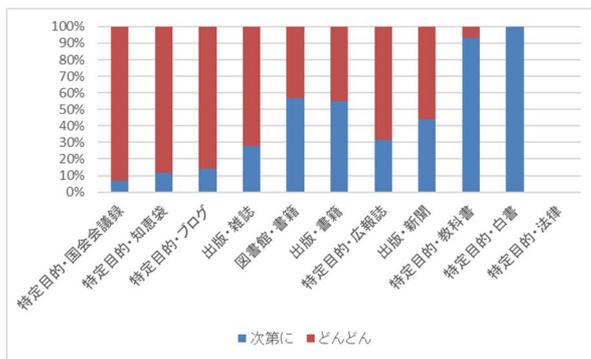


図3 「次第に」「どんどん」(ppmの比率)

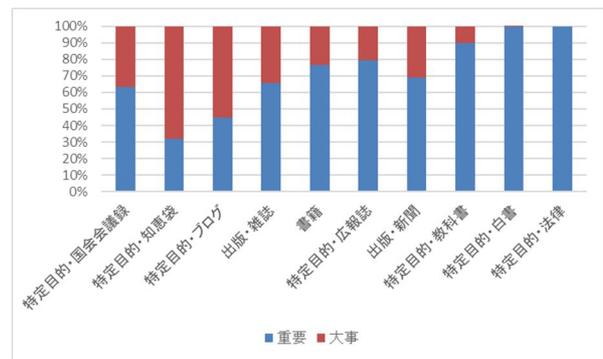


図4 「重要」「大事」(ppmの比率)

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>「日本語教育における音声言語と文字言語」国際シンポジウム『ビジネス日本語教育とグローバル人材育成』星野和子，招待，ハノイ貿易大学，2017.                  「ベトナム人日本語学習者の卒業論文・修士論文に見られる話し言葉的な副詞」国際シンポジウム『ビジネス日本語教育とグローバル人材育成』NGUYEN THI BICH HA，招待，ハノイ貿易大学，2017.                  「フォーマルな話し言葉に現れやすい書き言葉的な語」『シンポジウム「日常会話コーパス」III』柏野和佳子，査読無，国立国語研究所，2018.                  「学術的文章作成時に留意すべき「話し言葉的な語」とは」『第7回 早稲田大学 ライティング・フォーラム』柏野和佳子，招待，早稲田大学，2019.                  「日本語の書き言葉と話し言葉コーパスの構築と活用 学習者コーパスとの比較利用のために」国際シンポジウム『日本語学習者向けコーパスの構築と応用研究』柏野和佳子，招待，北京外国語大学，2019.                  「Making of 岩国第八版」『岩波国語辞典 第八版』刊行記念イベント』柏野和佳子，招待，紀伊國屋書店新宿本店，2019.                  「大学生の「話し言葉・書き言葉」についての認識」『日本語教育セミナー 日本語作文で気をつけたい話し言葉的な言葉とは』佐渡島紗織，主催，ハノイ大学，2020.                  「卒業論文に現われた話し言葉的な語 副詞を中心に」『日本語教育セミナー 日本語作文で気をつけたい話し言葉的な言葉とは』星野和子，主催，ハノイ大学，2020.                  「複合格助詞（「において」「によって」など）について」『日本語教育セミナー 日本語作文で気をつけたい話し言葉的な言葉とは』丸山直子，主催，ハノイ大学，2020.                  「作文教育のための日本語コーパスの活用」『日本語教育セミナー 日本語作文で気をつけたい話し言葉的な言葉とは』木田真理，主催，ハノイ大学，2020.                  「書き言葉的な語と話し言葉的な語のリスト作成について」『日本語教育セミナー 日本語作文で気をつけたい話し言葉的な言葉とは』柏野和佳子，主催，ハノイ大学，2020.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	丸山 直子 (MARUYAMA Naoko)  (00199936)	東京女子大学・現代教養学部・教授  (32652)	
研究分担者	佐渡島 紗織 (SADOSHIMA Saori)  (20350423)	早稲田大学・国際学術院・教授  (32689)	
研究協力者	星野 和子 (HOSHINO Kazuko)		
研究協力者	グエンビック ハーティ (NGUYEN BICH HA THI)		

## 6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	前坊 香菜子  (MAEBO Kanako)		
連携研究者	木田 真理  (KIDA Mari)  (80401727)	政策研究大学院大学・政策研究科・講師   (12703)	
連携研究者	仁科 喜久子  (NISHINA Kikuko)  (40198479)	東京工業大学・留学生センター・名誉教授   (12608)	