

令和 2 年 6 月 11 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K02794

研究課題名(和文) 書き言葉コーパスと話し言葉コーパスとを活用した外来語表記のゆれの研究

研究課題名(英文) Study of Orthographic Variation in the Japanese Loanwords using Written and Spoken Japanese Corpus

研究代表者

小椋 秀樹 (OGURA, Hideki)

立命館大学・文学部・教授

研究者番号：00321547

交付決定額(研究期間全体)：(直接経費) 1,600,000円

研究成果の概要(和文)：本研究は、『外来語の表記』(1991年、内閣告示・内閣訓令)で複数の表記が許容されている外来音等を取り上げ、『現代日本語書き言葉均衡コーパス』を資料として表記のゆれの実態調査を行うとともに、『現代日本語話し言葉コーパス』を資料として外来語の発音のゆれを調査し、その結果を表記のゆれに関する調査結果と突き合わせることで、表記のゆれと撥音のゆれとの関係を明らかにしようとするものである。調査の結果、外来語語末長音については、表記と発音との間にずれのあることが明らかとなった。

研究成果の学術的意義や社会的意義

大規模な書き言葉コーパス、話し言葉コーパスを活用して表記と発音のゆれの実態を解明する研究はこれまで行われていない。本研究は、外来語表記のゆれと発音のゆれとについて、最新のデータを提供するとともに、表記のゆれと発音のゆれとの関係に関するデータも提供するものである。

このような本研究の成果は、表記辞典の編纂や新聞・放送をはじめとする様々な分野における外来語表記の基準作成にとって重要な基礎データとなる。このことから、本研究は学術の社会的貢献として高い意義を有するといえる。

研究成果の概要(英文)：The present study is based on the fact that the "Spelling of Foreign Words" (1991, Cabinet Notice and Cabinet Instruction) allows for multiple spellings, and the The reality of orthographic distortions in "Modern Japanese Written Language Equilibrium Corpus". In addition to the survey, we investigated the pronunciation of foreign words using the "Corpus of Contemporary Japanese" as a reference The results were cross-checked with the results of the investigation of the wavering of the notation and of the repetition. The purpose of this study is to clarify the relationship between the two.

As a result of the survey, it was found that there was a gap between the notation and the pronunciation of the terminal long vowels of foreign words.

研究分野：日本語学

キーワード：コーパス 外来語 表記 撥音

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

#### (1) 外来語表記のゆれ

日本語における言語問題として外来語が取り上げられる場合、外来語の多用やその意味の分かりにくさが問題となることが多い。しかし問題はそれらにとどまるものではない。外来語には表記のゆれという問題もある。例えば、報道機関においては外来語の表記に関する基準をどうするかということが繰り返し議論されている。

外来語の表記にゆれが見られる要因として、国が定めた外来語表記の基準である『外来語の表記』(1991年、内閣告示・同訓令)の性格が挙げられる。『外来語の表記』は「グローブ・クラブ」のように語形にゆれのあるものについて語形の統一を図ったり、語形選択の基準を示したりするものではない。また表記の統一を図ろうとするものでもなく、各分野における慣用を認める立場を取っている。このような緩やかな性格が外来語表記のゆれを生む要因の一つとなっている。

#### (2) 外来語表記のゆれに関する調査

外来語の表記に関する実態調査としては、NHK放送文化研究所がウェブ検索やコーパスを使った調査を行っている。また、荻野綱男もウェブ検索による調査を行っている。しかしこれらの調査については、以下のような二つの問題がある。

1 点目は、一部の事象に関する調査となっていないという点である。NHK放送文化研究所の調査は、放送における発音・表記の基準の改定を目指したものであるため、放送で問題となる語が調査の中心となっており、調査の範囲が狭いといえる。荻野綱男の調査も、ゆれの見られる一部の外来音が対象となっており、調査の範囲は狭い。

2 点目は、ウェブ検索という手法に関わる問題である。例えば、『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)には研究用付加情報として形態論情報や著者情報、書誌情報等が付与されている。また検索ツールによってコーパス中から語連鎖を抽出することもできる。そのためコーパスを活用することによって、付加情報や前後の語、文脈を活用した詳細な分析を行うことができる。一方、ウェブ検索で得られた用例には、付加情報が付与されておらず、語連鎖の抽出もコーパスほど容易ではない。またそもそも検索した日時によって検索結果が変わるといって、不安定性が既に指摘されているところである。

以上のことから、現代における外来語表記のゆれの実態を解明するためには、まず『外来語の表記』で複数の表記が許容されている外来音を含む外来語を取り上げ、コーパスの研究用付加情報等を活用した詳細な分析を行うことが必要である。

### 2. 研究の目的

現代における外来語表記のゆれについては、ウェブ検索などを用いた実態調査が行われている。しかしこれらの調査の対象は、表記にゆれの見られる語の一部にとどまっているため、外来語表記のゆれの実態を体系的に明らかにするには至っていない。

本研究は、このような研究の現状を踏まえて、『外来語の表記』(内閣告示・同訓令)で複数の表記が許容されている外来音を取り上げ、BCCWJを対象に外来語表記のゆれの実態調査を行う。さらに『日本語話し言葉コーパス』(以下、CSJ)を対象に、外来語の表記のゆれに関連を持つと考えられる発音のゆれを調査し、その結果を表記のゆれに関する調査と突き合わせることで、外来語表記のゆれの実態を多角的に明らかにするものである。

### 3. 研究の方法

#### (1) 調査対象

本研究では、外来語表記のゆれの実態と外来語の発音のゆれの実態とを大規模コーパスを用いて明らかにする。前者については、BCCWJに収録したレジスターのうち、2001年~2005年に発行された新聞・雑誌・書籍、2008年4月26日~2009年4月25日に投稿されたブログを調査対象とする。後者についてはCSJを調査対象とする。

対象とする外来語は、『外来語の表記』で複数の表記が許容されている外来音を含む語である。

#### (2) 分析の観点・方法

以下に挙げる調査・分析を行い、現代における外来語表記のゆれの実態、外来語の発音のゆれの実態、表記(発音)のゆれが生じる要因を明らかにする。さらにBCCWJ、CSJの調査結果を突き合わせることで、表記のゆれと発音のゆれとの関係についても明らかにしていく。

どのような外来語に表記(発音)のゆれが生じるのか

BCCWJ、CSJそれぞれのデータベースを基に、どのような外来語に、どのような外来音の表記のゆれが見られるのかを明らかにする。

さらに、品詞別、語義別(多義語か単義語か)、語構成別、頻度別、単語のなじみ度別に表記のゆれを持つ語の割合などを分析し、どの品詞(語義、語構成、頻度、単語のなじみ度)の語に表記のゆれが多く見られるのか(あるいは語表記のゆれが少ないのか)を明らかにする。

表記のゆれと発音のゆれとはどのような関係にあるのか

上記の調査・分析を通して、外来語表記のゆれの実態と外来語の発音のゆれの実態とを明らかにした上で、両者を突き合わせて表記のゆれと発音のゆれとの関係について明らかにする。

この調査・分析は、外来音ごとに突き合わせを行う。例えば、語末長音であれば、長音符号を付ける表記と付けない表記とでゆれが見られる場合、発音でも長音での発音と短呼化した発音

とでゆれが見られるのか明らかにする。これによって、表記のゆれが発音のゆれを反映したもののなか、あるいは表記と発音とに関連性はなく、表記のみ（又は発音のみ）にゆれが見られるのかを明らかにしていく。

#### 4. 研究成果

外来語語末長音を対象に、BCCWJ と CSJ とを用いて外来語語末長音の表記と発音のゆれの実態調査を行い、外来語語末長音の表記と発音との関係について明らかにした。

##### (1) 語末母音と表記・発音のゆれ

BCCWJ(出版・書籍)における外来語語末長音の表記(長音符号による表記(長音符号表記)か長音符号を省く表記(符号無表記)か)の調査結果を表1に、CSJにおける外来語語末長音の発音(長音か短音か)の調査結果を表2にまとめた。

表1,表2の「全体」列を見ると、BCCWJ(出版・書籍)では長音符号表記が83.0%,符号無表記が17.0%,CSJでは長音が92.3%,短音が7.7%となっている。BCCWJでは長音符号表記が、CSJでは長音による発音が定着していることが分かる。

表1：外来語語末長音の表記のゆれ(延べ)

	全体	ア段	イ段	ウ段	エ段	オ段
符号	80986	47509	21968	6414	1953	3142
	83.0%	83.9%	74.9%	99.2%	98.8%	99.9%
符号無表記	16597	9145	7373	53	23	3
	17.0%	16.1%	25.1%	0.8%	1.2%	0.1%

表2：外来語語末長音の発音のゆれ(延べ)

	全体	ア段	イ段	ウ段	エ段	オ段
長音	19578	11556	5405	1533	610	474
	92.3%	90.6%	93.1%	99.3%	98.9%	98.5%
短音	1630	1204	401	11	7	7
	7.7%	9.4%	6.9%	0.7%	1.1%	1.5%

語末母音別に見ると、BCCWJ(出版・書籍)、CSJとも語末の母音によって表記・発音のゆれに違いがある。符号無表記率を見ていくと、BCCWJ(出版・書籍)ではイ段音の符号無表記率が最も高く25.1%で、ア段音が16.1%でそれに次ぐのに対し、ウ段音、エ段音、オ段音は1%前後と非常に低い割合となっている。CSJにおいてもBCCWJ(出版・書籍)と同様の違いがある。ア段音は短音率が最も高く9.3%で、イ段音が7.2%でそれに次ぐのに対し、ウ段音、エ段音、オ段音はやはり1%前後となっている。ただしBCCWJ(出版・書籍)に比べると、短音率の高いア段音、イ段音でもその割合は10%以下であり、短音率の低い他の母音との差は小さくなっている。

##### (2) 語末音と表記・発音のゆれ

語末がア段音、イ段音の語を対象として、より細かく語末音別に長音符号表記と符号無表記、長音と短音の割合を調査した。図1,2はBCCWJ(出版・書籍)の、図3,4はCSJの調査結果である。図1,2には度数500以上の語末音を、図3,4には度数200以上の語末音を示した。

図1から図4を見ると、語末音によって表記・発音のゆれに違いのあることが分かる。

ここでも符号無表記率を見ていくと、BCCWJのア段音(図1)では、「ファ」「ダ」「タ」の3音の符号無表記率が高い。「ファ」は59.7%で最も高く、「ダ」(25.7%),「タ」(23.7%)がそれに続く。一方、他の語末音はいずれも10%以下で低い割合となっている。イ段音(図2)でも同様の傾向が見られる。「ティ」(78.3%),「ディ」(74.5%)では符号無表記率が7割を超えており、「リ」(23.2%)も2割台である。一方、他の語末音はいずれも5%以下で低い割合となっている。

CSJのア段音(図3)では「ア」の短音率が38.4%で最も高く、「タ」が11.8%でそれに次ぐ。しかしその他の語末音の短音率は1割を下回っている。イ段音(図4)については、全体的に短音率が低く、最も短音率が高い「リ」でも10.9%で、「リ」以外の音は全て10%を下回っている。

3.1節で述べたように、語末の母音別に見た場合、BCCWJ,CSJのいずれにおいてもア段音、イ段音の符号無表記率、短音率が他の母音よりも高く、符号無表記、短音化の生じやすさに語末の母音による偏りが見られた。本節では語末のア段音、イ段音を対象に、更に語末音別に表記・発音の実態を調査した。その結果、上に述べたとおりア段音、イ段音とも全ての語末音において符号無表記や短音化が一樣に見られるのではなく、BCCWJ(出版・書籍)では「ファ」「ダ」「タ」「ティ」「ディ」「リ」の符号無表記が、CSJでは「ア」「タ」「リ」の短音化が他の語末音に比べて多いことが分かった。つまり、語末音によって符号無表記、短音化の生じやすさに偏りが見られるのである。

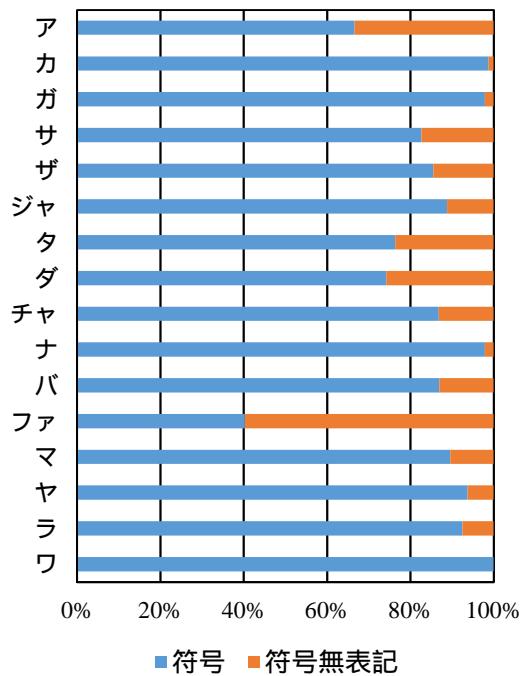


図 1: 表記のゆれ (ア段音, 語末音別)

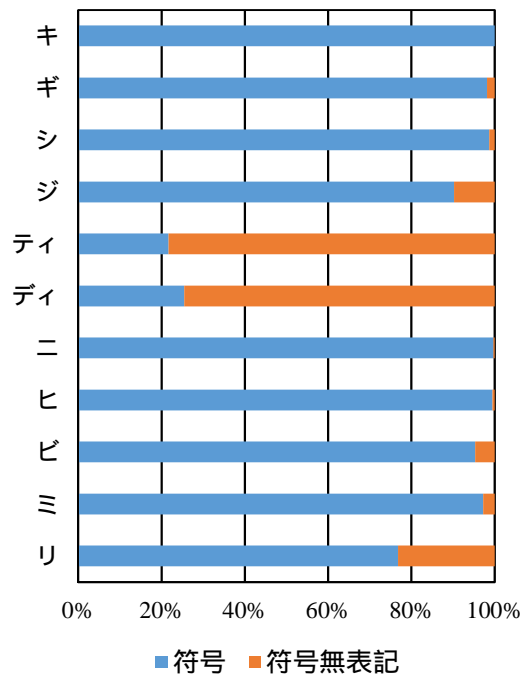


図 2: 表記のゆれ (イ段音, 語末音別)

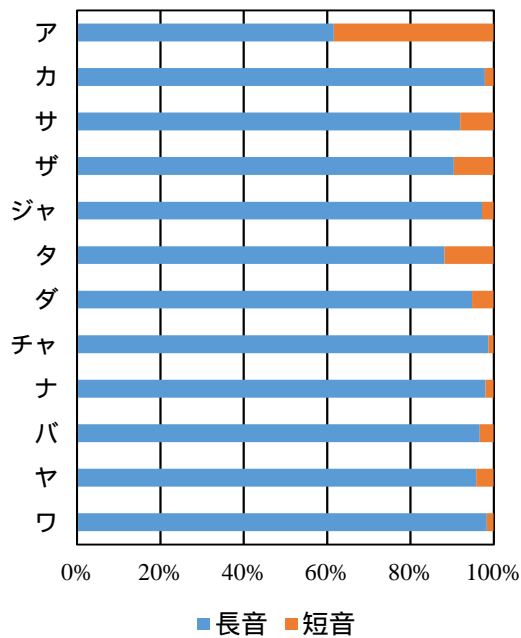


図 3: 発音のゆれ (ア段音, 語末音別)

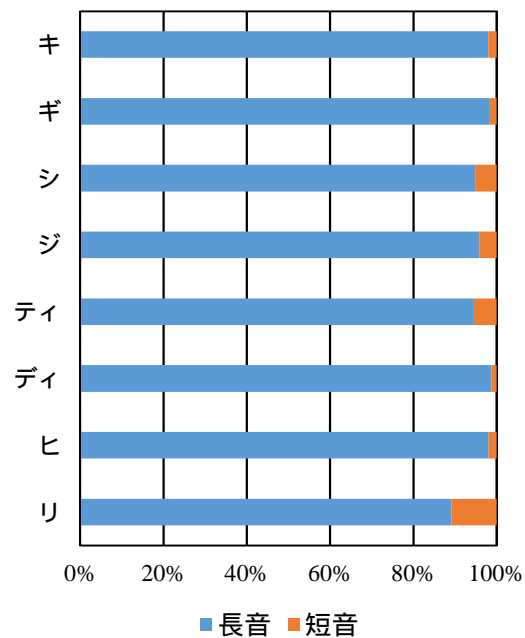


図 4: 発音のゆれ (イ段音, 語末音別)

### (3) 語と表記・発音のゆれ

BCCWJ (出版・書籍) において語末音「ア」を持つ語を示した表 3 を見ていく。表 3 の下にあるとおり、語末音「ア」を持つ語は異なり 33、延べ 1124 である。延べ 1124 のうち 369 が符号無表記で用いられたものである。この表 3 から、《エアー》の符号無表記の度数が非常に高いことが分かる。《エアー》だけで語末音「ア」の符号無表記の 64.5% を占める。また表 3 に挙げた 5 語で、語末音「ア」の符号無表記の 96.7% を占める。

表 4、表 6、表 7 についても、表 3 と同様の傾向が認められる。表 4 を見ると、《ソファー》が語末音「ファ」の符号無表記の 64.5% を占めている。また語末音「ファ」の符号無表記は、表 4 に挙げた 4 語のみで生じている。表 6 では、《ボディー》が語末音「ディ」の符号無表記の 41.5% を占めており、表 6 の 5 語で、語末音「ディ」の符号無表記の 82.0% を占める。表 7 では、《ドア》が語末音「ア」の短音の 55.3% を占め、表 7 の 5 語で語末音「ア」の短音の 86.3% を占める。

以上のことから、BCCWJ（出版・書籍）の語末音「ア」「ファ」「ディ」においては、一部の語に符号無表記が生じており、しかもそれらの語の符号無表記の度数や符号無表記率が高いため、これら三つの語末音の符号無表記率が高くなっているといえる。CSJの語末音「ア」についても同様である。

しかし、BCCWJ（出版・書籍）の語末音「ティ」については異なる傾向が見られる。表5の下に示したように、語末音「ティ」を持つ語は異なり138、延べ5393、符号無表記の延べ数は4223である。表5に示した語の符号無表記の度数は合計2197であるが、全体に占める割合は52.0%で、他の語末音に比べると低い割合となっている。

語末音「ティ」を持つ語は異なりで138あるが、このうち118語（85.5%）で符号無表記が生じている。したがって、語末音「ティ」は他の語末音とは異なり、特定の語に符号無表記が生じているのではなく、語末音「ティ」を持つ広範囲の語に符号無表記が生じている。また表5を見ると、符号無表記の度数が高く、符号無表記率も高い。広範囲の語に符号無表記が生じ、しかも度数、符号無表記率が高いということが語末音「ティ」全体の符号無表記率を上げていると考えられる。これは語の問題というより、「ティ」という語末音の問題としてとらえることができる。

表3：符号無表記度数上位語（語末「ア」）

語彙素	度数	無表記 度数	%
エアー	283	238	84.1%
ピアー	33	33	100.0%
オンエアー	32	26	81.3%
ビューアー	30	24	80.0%
ドラッグストアー	23	21	91.3%

異なり33、延べ1124、符号無表記369

表4：符号無表記度数上位語（語末「ファ」）

語彙素	度数	無表記 度数	%
ソファー	518	388	74.9%
バッファー	154	146	94.8%
サイファー	11	11	100.0%
メタファー	55	2	3.6%

異なり16、延べ、916、符号無表記547

表5：符号無表記度数上位語（語末「ティ」）

語彙素	度数	無表記 度数	%
コミュニティー	608	548	90.1%
パーティー	829	467	56.3%
プロパティー	464	456	98.3%
セキュリティー	408	392	96.1%
アイデンティティー	368	332	90.2%

異なり138、延べ5363、符号無表記4199

表6：符号無表記度数上位語（語末「ディ」）

語彙素	度数	無表記 度数	%
ボディー	473	431	91.1%
レディー	206	168	81.6%
メロディー	228	118	51.8%
スタディー	85	74	87.1%
パロディー	69	61	88.4%

異なり31、延べ1394、符号無表記1039

表7：短音度数上位語（語末「ア」）

語彙素	度数	短音 度数	%
ドアー	160	141	88.1%
クラシファイアー	37	23	62.2%
バリアー	22	20	90.9%
アウトドアー	26	18	69.2%
ストアー	47	18	38.3%

異なり35、延べ665、短音255

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 小椋秀樹	4. 巻 109
2. 論文標題 書き言葉と話し言葉における外来語語末長音のゆれ	5. 発行年 2018年
3. 雑誌名 論究日本文学	6. 最初と最後の頁 1-12
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 小椋秀樹
2. 発表標題 書き言葉と話し言葉における外来語語末長音のゆれ
3. 学会等名 言語資源活用ワークショップ2017
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 小椋秀樹・富士池優美・宮内佐夜香・金愛蘭・柏野和佳子	4. 発行年 2020年
2. 出版社 朝倉書店	5. 総ページ数 160
3. 書名 コーパスで学ぶ日本語学 日本語の語彙・表記	

〔産業財産権〕

〔その他〕

-

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
---------------------------	-----------------------	----