

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 4 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25370532

研究課題名(和文)大規模コーパスに基づく現代語表記のゆれの実態解明

研究課題名(英文)Corpus-based Survey of the Orthographic Variation in Contemporary Japanese

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) 3,100,000円

研究成果の概要(和文)：本研究は、現代語表記のゆれの実態について、『現代日本語書き言葉均衡コーパス』に収録された複数のレジスターを対象にした調査から明らかにするものである。現代語表記のゆれの実態として、次のようなことが明らかとなった。

- (1) 統語的複合動詞の後項動詞の表記の変遷を調査したところ、漢字表記が増加する傾向が観察された。
- (2) 外来語については、長音の表記に関するゆれが多く見られた。具体的には、語末長音のゆれと、原語の二重母音[e i]の表記のゆれが観察された。

研究成果の概要(英文)：This study clarifies orthographic variation in contemporary Japanese by survey of multiple registers recorded Balanced Corpus of Contemporary Written Japanese. As the actual situation of orthographic variation, the following thing became clear.

- (1) Results of changes in the latter form of syntactic compound verbs, tendency of kanji is to increase has been observed.
- (2) For loan words, it was seen orthographic variation of long vowels. In particular, orthographic variation of word-final long vowels and hiatus [ei] has been observed.

研究分野：日本語学

キーワード：文字 コーパス

1. 研究開始当初の背景

(1) 語表記のゆれ

音節、語など、種々の言語単位において、形式が一つに定まらず、複数の形式が許容されることがある。この複数の形式が共時的に存在する現象を「ゆれ」と呼ぶ。

語のレベルにおけるゆれには、語形やアクセントのゆれのほか、日本語においては表記のゆれが多く見られる。例えば、「俺 - おれ」「さくら - サクラ」のような異なる文字体系間の対立によるゆれのほか、「付属 - 附属」のような異なる漢字の対立によるゆれ、「行 - 行なう」のような送り仮名の違いによるゆれ等がある。また、「上げる - 挙げる - 揚げる」のような異字同訓も、それぞれを別語とせず同一の語と見なした場合、《アゲル》に「上げる」「挙げる」「揚げる」という複数の表記が共時的に存在すると捉えられ、《アゲル》の表記のゆれとして扱うことができる。

(2) 語表記のゆれに関する調査

語表記のゆれについては、これまでに次の三つの調査が行われている(調査対象年順に示す。)

宮島達夫(1997)「雑誌九十種表記表の統計」『日本語科学』1(1956年発行雑誌90種の調査)

国立国語研究所(1983)『現代表記のゆれ』(1966年発行朝日・毎日・読売3紙の調査)

国立国語研究所(2006)『現代雑誌の表記』(1994年発行雑誌70誌の調査)

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点がある。

1点目は、いずれも調査対象が単一のレジスターという点である。書き言葉は、新聞・雑誌にとどまるものではなく、書籍・教科書・公文書・法律など様々なレジスターがある。また、1966年当時には存在しなかったレジスターとしてWebが挙げられる。現代における語表記のゆれの実態解明という面からは、複数の多様な媒体を対象に調査を行い、レジスター差の有無等を明らかにする必要がある。

2点目は、国立国語研究所(2006)の調査対象年(1994年)から既に18年が経過しているという点である。1990年代から、情報機器の急速な普及に伴って書記環境が大きく変化するとともに、それに伴う漢字使用の増加が指摘されている。その結果、表記のゆれの実態にも変化が生じていることが予想されることから、より現在に近い時期における語表記のゆれを調査する必要がある。

2012年に、本研究代表者は語表記のゆれについて『現代日本語書き言葉均衡コーパス』の一部(コアデータ:白書・新聞・雑誌・書

籍・Yahoo!知恵袋・Yahoo!ブログから成る延べ100万語のデータ)を利用した予備調査を行った。この予備調査において、現代語表記のゆれの傾向を概観するとともに、コーパスを活用した語表記のゆれの調査手法等について検討するなど、本研究を開始するための準備を行った。

2. 研究の目的

本研究は、『現代日本語書き言葉均衡コーパス』に収録された2001年~2005年発行の新聞・雑誌・書籍、2008年4月~2009年4月に投稿されたブログという四つの媒体を対象にして、より現在に近い時期における語表記のゆれの実態を明らかにすることを目的とする。

3. 研究の方法

(1) 調査対象・データ規模

本研究では、『現代日本語書き言葉均衡コーパス』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。また、いわゆる一般語を対象とすることとし、固有名詞・数詞・感動詞・助詞・助動詞・記号・補助記号は対象から除く。各レジスターのデータの異なり語数・延べ語数を次に示す(「/」の左が異なり語数、同じく右が延べ語数である。)

新聞	約3万語	/	約74万語
雑誌	約4.7万語	/	約232万語
書籍	約8.6万語	/	約1520万語
ブログ	約5.5万語	/	約520万語

(2) 「表記一覧表」の作成

『現代日本語書き言葉均衡コーパス』に収録した新聞・雑誌・書籍・ブログの各レジスターを対象に、調査・分析の基本資料となる「表記一覧表」を作成する。「表記一覧表」は、各レジスターに出現した語をその出現表記形、出現表記形の頻度とともに一覧にしたもので、新聞・雑誌・書籍・ブログのレジスター別に作成し、それぞれ次に挙げる情報を付与する。

《「表記一覧表」の付与情報》

見出し語、見出し語の表記、品詞、語種、見出し語の頻度、出現表記形、出現表記形の頻度

「表記一覧表」を基に、レジスターごとに語表記にゆれの多い語群、少ない語群やゆれない語群を把握する。

さらに、語種・品詞・語構成・頻度の面からも分析を行う。具体的には、語種・品詞等の別に分析を行い、どの語種・品詞等の語に語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。

レジスター別の語表記のゆれに関する調査結果を比較し、語表記のゆれが媒体によってどのように異なるのか、あるいはどのような点について共通性があるのかといったことを明らかにする。

語表記のゆれが生じる要因について、以下のような言語内的観点、言語外的観点から考察を行う。

言語内的観点：品詞や語種・語構成のほか、どのような意味・用法のときに、どのような表記が用いられているか。

言語外的観点：新聞漢字表等、そのレジスターにおける表記の基準の有無や国語施策の影響。著者に関する情報(性、年齢など)の別。

4. 研究成果

『現代日本語書き言葉均衡コーパス』のコーデータを資料として和語・漢語・外来語・混種語の全語種を対象に行った予備調査である小椋秀樹(2012)「コーパスに基づく現代語表記のゆれの調査 BCCWJ コアデータを資料として」(『第1回コーパス日本語学ワークショップ予稿集』, pp.321-328.)及び、本研究において作成を進めた「表記一覧表」の分析を基に、語表記のゆれが多く見られる外来語と複合動詞とを取り上げ、より詳細な分析を行い、語表記のゆれの実態を明らかにした。

(1) 外来語表記のゆれ

『現代日本語書き言葉均衡コーパス』のコーデータを資料として、そこに収録された新聞・雑誌・書籍・白書・Webの五つのレジスターを対象に外来語表記のゆれの実態を計量的な手法によって明らかにした。

具体的には、外来語の表記がどの程度ゆれているのかレジスターごとに調査し、レジスターによる差異を明らかにした上で、各レジスターにおける外来語表記のゆれの類型についても明らかにした。

ゆれのレジスター差

各レジスターにおいて、どの程度、外来語表記のゆれが見られるのか、度数2以上の語を対象に調査を行った。

その結果、異なり語数で見ると、雑誌・Webではゆれが見られる語の割合は20代であった。一方、新聞、書籍、白書は10%未満であった。このように、外来語表記のゆれにはレジスターによる差異が見られることが明らかとなった。

外来語表記のゆれの類型

各レジスターで表記がゆれている語を対

象に、どのような表記のゆれが、どのくらいあるのかについても調査を行った。この調査では、a. 外来語表記法の対立(例：バイオリン - ヴァイオリン) b. 異なる文字体系間の対立(例：ページ - 頁) c. 符号等による表記と他の表記との対立(例：パーセント - %)の3類型に分類した。

その結果、表記のゆれの類型についてもレジスター差が見られた。雑誌・Web・新聞は、他のレジスターに比べて類型bの割合が高く、書籍・白書は他のレジスターに比べて類型aの割合が高かった。

外来語表記法の対立

類型aに属する語を対象に、どのような外来語表記法の対立(ゆれ)が見られるのか調査した。その結果、長音に関する表記のゆれ(語末長音、語中長音)が全てのレジスターにわたって多く見られた。

(2) 外来語語末長音のゆれ

上記(1)の調査を踏まえ、『現代日本語書き言葉均衡コーパス』の新聞・雑誌・書籍・Webの四つのレジスターを対象に、外来語語末長音のゆれ(例：コンピューター - コンピュータ)の実態について調査を行った。

この調査では、(1)の調査において原語の語末が-er, -or, -ar, -gy, -dy, -ry, -tyの語に表記のゆれが多く見られたことから、これらの語を対象とした。

ゆれのレジスター差

この調査では、対象を語末長音のゆれにしぼったが、ゆれの割合にレジスター差が見られた。ゆれの見られる語の割合は、Web・書籍が20%台、雑誌が20%弱であり、これらはゆれの割合の高いレジスターといえる。一方、新聞は約5%とゆれが極端に低く、『外来語の表記』、新聞各社の表記の基準に忠実に従い、表記をかなり統一していることが明らかとなった。

原語の語末による差

語末長音の表記のゆれには、原語の語末による差異も見られた。語末が-rの語よりも語末が-yの語の方がゆれの割合が高い。

また語末が-yの語の中でも、語末が-dy, -tyの語のゆれの割合が高い。

原語の語末が-yの語に、ゆれが多く見られる要因として、原語の発音の影響を指摘した。語末yの発音は、弱母音/i/であり、明確な長音でない。そのため原語の発音を意識した場合、長音符号を省く傾向が生じている可能性がある。専門用語の場合は、特にその傾向が強くなるのではないかとした。

(3) 外来語における二重母音[eɪ]のゆれ

上記(1)の調査において、ゆれが多く見られた語中長音のゆれとは、原語の二重母音

を長音符号で書くか連母音で書くというゆれで、その中でも原語の二重母音[eɪ]の表記のゆれ(例: プレーヤー - プレイヤー)が最も多かった。

そこで、原語の二重母音[eɪ]の表記を取り上げ、『外来語の表記』の原則に当たる長音表記と原則に当たらない連母音表記とがそれぞれの程度用いられているのか実態調査を行った。

レジスター差

全レジスターで長母音表記の割合が80%以上を占めており、『外来語の表記』の原則に当たる長音表記が定着していることが確認された。しかし、レジスター別に見ると、差異も見られた。連母音表記の割合は、雑誌が13.7%で最も高く、新聞が5.6%で最も低い。書籍、Webは約10%である。なお上記(1)(2)と比べた場合、レジスター差は小さいといえる。

語別の傾向

語別に見ると、連母音表記に偏るものが見られる。連母音表記の割合が90%以上の語としては、「メード」「ネール」「ブレーク」「ネーティブ」などが挙げられる。これらは、いずれも比較的新しい語と位置付けられる。比較的新しい語では、原語の発音を重視して、連母音表記が定着していると考えられる。

意味・用法と表記

長音表記か連母音表記かについては意味用法との関係が見られる。例えば、「ディスプレイ」は、「展示」の意味よりも「モニター」の意味の方が連母音表記の割合が高い。「デイ」は「Xデー」のような「~日」という意味よりも、「昼間」の意味の方が連母音表記の割合が高くなっている。

(4) 複合動詞のゆれ

予備的調査においては、語彙的複合動詞の語表記のゆれが大きいことを指摘した。それを踏まえて、統語的複合動詞、特にその後項動詞の表記について、『現代日本語書き言葉均衡コーパス』の図書館サブコーパスを資料として、出版年代、著者の生年代といった観点から、漢字表記と平仮名表記の割合がどのように変化するのか明らかにした。

出版年代

年代が下るにしたがって、徐々に漢字表記の割合が高くなっている。漢字表記の割合は、80年代後半は41.9%であったが、90年代前半は42.7%、90年代後半は46.6%、2000年代は49.1%となる。このように2000年代には、漢字表記が平仮名表記と拮抗するまでになる。

著者の生年代

1890年代生まれのグループから1910年代生まれのグループまで、漢字表記の割合が低下するが、その後は、1970年代生まれのグループまで漢字表記の割合が増加傾向にある。1950年代以降のグループからは、漢字表記が過半数を占めている。

出版年代においても、著者の生年代においても年代が下るにしたがって、漢字表記の割合が高くなることが確認された。この要因については、漢字政策の改定(漢字表への訓の追加)と情報機器による漢字の多用化という二つの言語外的要因を指摘した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計6件)

小椋秀樹、現代における外来語表記のゆれ 語末長音と[eɪ]の表記、日本語学、査読無、35(8)、2016、印刷中

小椋秀樹、コーパス活用の勘所 第19回【現代語】表記(2)『現代日本語書き言葉均衡コーパス』による表記の研究、日本語学、査読無、34(12)、2015、78-83

小椋秀樹、複合動詞後項の表記の経年変化 BCCWJを資料として、論究日本文学、査読無、102号、2015、1-12

小椋秀樹、コーパス活用の勘所 第8回【現代語】表記(1)『現代日本語書き言葉均衡コーパス』による表記の研究、日本語学、査読無、33(13)、2014、78-83

小椋秀樹、外来語語末長音の表記のゆれについて、論究日本文学、査読無、100、2014、195-208

[学会発表](計3件)

小椋秀樹、外来語における[eɪ]の表記のゆれ、第8回コーパス日本語学ワークショップ、2015年9月1日、国立国語研究所(東京都立川市)

小椋秀樹、BCCWJにおける複合動詞後項の表記の実態、第6回コーパス日本語学ワークショップ、2014年9月10日、国立国語研究所(東京都立川市)

小椋秀樹、外来語語末長音の表記のゆれについて、第4回コーパス日本語学ワークショップ、2013年9月6日、国立国語研究所(東京都立川市)

[図書](計1件)

相澤正夫(編)、金愛蘭・新野直哉・松田謙次郎・金澤裕之・尾崎喜光・石井正彦・

小椋秀樹 他、おうふう、現代日本語の動態研究、2013、262 (151-171)

6 . 研究組織

(1)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