

大規模コーパスを活用した外来語表記のゆれの調査

小 椋 秀 樹

一 はじめに

日本語における言語問題として外来語の氾濫ということが繰り返し言われる。特に近年では、意味の分かりにくい外来語が多用されていることが問題となることが多い。この問題に、日本語学の立場から取り組んだのが、小泉純一郎首相（当時）の肝いりで始まった国立国語研究所の「外来語」言い換え提案（二〇〇二年から二〇〇六年）であった。「外来語」言い換え提案では、公文書や新聞・放送等において使用されるものの、一般の国民には意味の分かりにくい外来語に焦点を絞り、別の語に言い換えたり、注釈を付けたりして分かりやすくする工夫が具体的に提案された。

一方、外来語には、意味の分かりにくさという問題のほかに、表記のゆれという問題もある。国が示した外来語表記の基準として『外来語の表記』（二九九年、内閣告示・内閣訓令）がある。これは、「法令、公用文書、新聞、雑誌、放送など、一般の社会生活において、現代の国語を書き表すための「外来語の表記」のよりどころを示したものである。この『外来語の表記』は、原則を示しつつも、慣用として定着している表記を認めるなど、一語に複数の表記を認め得るような緩やかな基準である。そのため、「バイオリンーヴァイオリン」「コンピューターーコンピュータ」「メールーメール」のように一語に複数の表記が見られることがある。

それでは、現代日本語において外来語表記のゆれは、実際にどの程度見られるのであろうか。外来語表記のゆれについては、これまでに次の二つの調査が行われている。

宮島達夫・高木翠（一九八四）…一九六四年発行雑誌九十種の調査

国立国語研究所（一九八三）…一九六六年発行朝日・毎日・読売三

紙の調査

しかし、これらの調査については、二つの問題点がある。一点目は、いずれも調査対象が単一の媒体という点である。書き言葉には、新聞・雑誌・書籍・公文書等、様々な媒体がある。また、一九六四年当時には存在しなかった新しい媒体としてウェブがある。このようなことを考えると、雑誌のみ、又は新聞のみを対象とした調査では、外来語表記のゆれの実態解明には不十分である。

二点目は、調査対象となった一九六四年、一九六六年から既に四十年以上が経過しているという点である。この間、外来語の多用化が進むとともに、使われる外来語も変化しており、外来語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。そこで、より現在に近い時期における外来語表記のゆれを調査する必要がある。

このように現代日本語における外来語表記のゆれの実態を把握するためのデータは欠けており、新たに調査を行うことが必要である。その際には、より現在に近い時期における外来語表記のゆれの調査を行うと

もに、複数の媒体を対象として、外来語表記のゆれに媒体差があるのか明らかにすることも求められる。

そこで、本稿では、『現代日本語書と言葉均衡コーパス』（以下、BCGWSJ）のコーデータ^②を資料として、そこに収録された新聞・雑誌・書籍・白書・ウェブの五つの媒体を対象に、より現在に近い時期における外来語表記のゆれの実態を明らかにする。

二 調査対象

本稿の目的は、複数の媒体を対象として、より現在に近い時期における外来語表記のゆれの実態を明らかにすることにある。そこで、複数の媒体のデータを収録したBCGWSJを調査対象とした。ただしBCGWSJ全体を対象とするのではなく、今回はコーデータのみを対象とした。

コーデータは、次の媒体から成り、延べ語数は、短単位で約一一〇万語、長単位で約八四万語（短単位、長単位とも記号、空白、補助記号を除く）である。

出版サブコーパス …二〇〇一～二〇〇五年発行の新聞、雑誌、

書籍

特定目的サブコーパス…二〇〇一～二〇〇五年発行の白書

二〇〇四年一月～二〇〇五年一月投稿

のヤフー知恵袋

二〇〇八年四月～二〇〇九年四月投稿のヤ

フーブログ

コーデータは、自動形態素解析後に、全体に対して人手による確認、修正を行ったデータで、解析精度は長単位・短単位とも約九九%以上である。

BCGWSJを対象とした語表記のゆれの調査では、BCGWSJ全体を対象

とすることも考えられるが、コーデータ以外のデータの精度が約九八%と少し低いことから、調査に当たって誤解析がどの程度影響するかが気になるところである。そこで、今回はBCGWSJ全体を対象とした語表記のゆれの調査を行う前の予備調査として、より精度の高いコーデータのみを対象とすることとした。

本稿では、固定長・可変長サンプルの両方を対象とした。また、長短二種類のデータのうち短単位を用い、人名・地名等の固有名詞を除く、いわゆる一般語を対象とした。

三 外来語表記のゆれの認定

語表記のゆれとは、一つの語形（発音）に対して二種類以上の表記が共時的に存在する現象を指す。この定義に厳密に従えば、外来語表記のゆれとは、「バイオリンーヴァイオリン」「ビーナスーヴィーナス」のようなゆれを指すことになる。

しかし、一般には「コンピューターーコンピュータ」「マネージャーーマネジャー」「メールーメール」なども外来語表記のゆれとして問題にされる。しかし「コンピューターーコンピュータ」等は、本来発音の違いに基づく表記の違いである。例えば、「コンピューター」は、語末を長音化した発音を写したものの、「コンピュータ」は語末を短呼した発音を写したものであって、厳密な意味での語表記のゆれには当たらない。それにもかかわらず、これらが外来語表記のゆれとして扱われるのは、語表記のゆれか発音のゆれかの判定が難しいことによる。つまり「コンピューター」と書かれていても「コンピュータ」と、「コンピュータ」と書かれていても「コンピューター」と発音する人がいる可能性は高く、語末の長音符号の有無を発音の違いによるものと言い切れないのである。

表 1 : UniDic の階層構造

語彙素	語形	書字形
ヤハリ【矢張り】	ヤハリ (副詞)	やはり
		ヤハリ
		矢張り
	ヤッパリ (副詞)	やっぱり
		ヤッパリ
		矢っ張り

宮島達夫・高木翠（一九八四：四四）でも「ピアノくピアノ」は、発音がちがうはずだが、その差は微妙である。「ア」または「ヤ」がかいてあるからといって、そのとおり発音しているとみるのは、あぶない。」と述べ、発音のゆれと表記のゆれとの区別の難しさを指摘している。その上で、宮島達夫・高木翠（一九八四）では、「ピアノーピアノ」のようなゆれも外来語表記のゆれとして扱っている^③。本稿でも、宮島達夫・高木翠（二九八四）と同様の立場を取り、「コンピューターコンピュータ」等のゆれも外来語表記のゆれとして扱うこととした。

BCWJ の形態素解析には、形態素解析辞書 UniDic^④を用いた。UniDic では、表記や語形の違いにかかわらず、同じ語であれば、同一の見出しを与えるという方針を取り、語を階層化した形で登録している。この階層の最上位を語彙素（国語辞典の見出しに相当）と呼んでおり、この語彙

素の下に語形（語形の違いを区別する層）、更に語形の下に書字形（表記の違いを区別する層）という階層を設けている（表 1 参照）。

この階層構造は、BCWJ の形態論情報にも反映しており、コーパス中の全ての短単位に対して語彙素・語形・書字形という階層的な見出しや品詞情報が付与されている。本稿では、この階層的な見出しを利用して外来語表記のゆれを認定することとしたが、その際、一部に修正を加えた。その修正とは、次に説明する「集計用語形」の追加である。

UniDic では、表 2 に示したように「コンピューター」「コンピュータ」は別語形として扱っている。しかし、これらを別語形として扱

大規模コーパスを活用した外来語表記のゆれの調査

表 2 : UniDic における外来語の辞書登録

語彙素	語形	書字形
コンピューター	コンピューター (名詞)	コンピューター
	コンピュータ (名詞)	コンピュータ

表 3 : 集計用語形の例

語彙素	集計用語形	語形	書字形
コンピューター	コンピューター (名詞)	コンピューター (名詞)	コンピューター
		コンピュータ (名詞)	コンピュータ

表 4 : 媒体別語数（異なり・延べ）

媒体	異なり	延べ
ウェブ	2,055	7,431
書籍	1,071	3,960
雑誌	2,175	10,090
新聞	2,023	9,508
白書	663	4,550

わず、同一語形における表記のゆれとして扱おうというのが本稿の立場である。そこで、表 3 のように新たに「集計用語形」という層を設けた。「集計用語形」は、語彙素と同じ語形で、語形と同じ品詞を持つこととした。この語彙素、集計用語形、書字形という階層を利用して、表記のゆれを認定した。具体的には、「任意の異なる二つの書字形が、同じ語彙素・集計用語形・品詞を持つ場合、同じ語の表記のゆれと認める。」とした^⑤。集計用語形レベルで集計を行うため、以下に示す語数は、全て集計用語形の数である。なお、表 4 に媒体別に外来語の異なり語数・延べ語数

表5：表記にゆれのある語の割合（媒体別、異なり）

媒体	異なり	ゆれ	%
ウェブ	2,055	234	11.4%
書籍	1,071	31	2.9%
雑誌	2,175	277	12.7%
新聞	2,023	92	4.5%
白書	663	16	2.4%

媒体別に、どの程度の外来語に表記のゆれが見られるのか見ていく。語表記にゆれのある外来語の異なり語数と割合を表5に示した。表5を見ると、外来語表記のゆれの割合は、雑誌が一二・七%で最も高く、ウェブが一一・四%でそれに次ぐ。この二つの媒体のゆれの割合が一割を超えている。書籍・新聞・白書のゆれの割合は、五%以下で、いずれも一桁台である。この中では、新聞が四・五%で最も高く、書籍が二・九%、白書が最も低く二・四%である。

雑誌とウェブのゆれの割合が一割を超え、他の媒体に比べて、かなり高くなっている点が注目される。この要因は、雑誌・ウェブにおいて、英字で表記された外来語が多く用いられていることにある^⑥。そこで、英字表記の外来語を除外して、語表記のゆれの割合を算出し直した結果を表6にまとめた。

全ての媒体でゆれの割合が低下しているが、雑

四 媒体別の外来語表記のゆれ

を示した（集計用語形の数で、固有名詞・数詞・感動詞・助詞・助動詞・記号・補助記号を除く）。本研究では、ヤフー知恵袋とヤフーブログとを併せてウェブとして集計することとした。

なお、本稿では、語の表記を示す際には「コンピューター」「コンピュータ」のようにカギ括弧を付けて示し、語を示す際には《コンピューター》のように二重山括弧を付けて表記する。

表6：表記にゆれのある語の割合（媒体別、異なり、英字除く）

媒体	異なり	ゆれ	%
ウェブ	2,010	94	4.7%
書籍	1,067	20	1.9%
雑誌	2,082	68	3.3%
新聞	2,007	45	2.2%
白書	659	10	1.5%

誌・ウェブの低下は他の媒体に比べて著しい。雑誌で約九%、ウェブで約七%低下している。一方、新聞・書籍・白書は一%から二%程度の低下にとどまっている。表5と表6との比較からも雑誌・ウェブに英字で表記された外来語が多く用いられており、それがゆれの割合を押し上げた要因になっていることが分かる。

表6に示した外来語表記のゆれの割合は、いずれの媒体においても五%以下と低い。媒体差は指摘できる。ゆれの割合が三%以上のウェブ・雑誌と二%前後の新聞・書籍・白書とに大きく二分される。このように五つの媒体が大きく二分される要因の一つとして、その媒体に共通の表記の基準があるか否かということが考えられる。白書では、全省庁とも国が定めた外来語表記の目安である『外来語の表記』の原則に当たる規定に基づいて、外来語を表記している。新聞は、新聞社ごとに用字に関する基準を定めているが、外来語については、やはり『外来語の表記』に基づいており、その中でも原則に当たる規定が共通の基盤となっている。そのため、新聞社による違いは余り大きくないと考えられる。

一方、雑誌については『外来語の表記』をよりどころとしているであろうが、白書・新聞とは異なり、出版社や雑誌ごとに自由度があると思われる。ウェブは、どのような表記を用いるかは全く個人の自由である。そのため、ウェブ・雑誌のゆれの割合が新聞・白書よりも高くなっていると考えられる。

なお、書籍についても雑誌と同様、著者個人の自由度が高いと思われる

るが、今回の調査ではゆれの割合が一・九%と新聞を下回っている。したがって、その媒体に共通の表記の基準の有無は、要因の一つではあるが、それだけでは、表6に示した結果について十分に説明することはできない。媒体に共通の表記の基準の有無以外に、どのような要因が考えられるか、今後の検討課題としたい。

五 外来語表記のゆれの類型

本節では、外来語表記のゆれに、どのようなゆれの類型があるのかを見ていく。外来語表記のゆれの類型について、次の五類型に整理する。

a 外来語表記法の対立

例…コンピュータ—コンピュータ、バイオリン—ヴァイオリン

b 片仮名と英字の対立

例…ニュース—NEWS、ページ—page

c 片仮名と平仮名の対立

例…アップ—あっぷ、クレヨン—くれよん

d 平仮名と英字の対立

例…あつと—at、ファン—ふあん

e 漢字・符号等による表記と他の表記との対立

例…アット—@、グラム—g、コーヒー—珈琲

語表記にゆれのある外来語について、右の類型に分類した結果を表7として示した。表7では媒体別に分けず、コアデータ全体をまとめて集計した。なお、ゆれのある外来語の中には、一つの類型に分類されるものだけでなく、複数の類型に分類されるものがある。例えば、《アーティスト》の表記は、コアデータ中に「アーティスト」「アーティスト」「Artist」の三種類が出現する。このうち、「アーティスト—アーティスト」のゆれは

表7：外来語表記のゆれの類型（異なり、延べ）

類型	異なり	%	延べ	%
a	198	26.4%	3,599	19.9%
b	459	61.3%	10,671	59.0%
c	37	4.9%	869	4.8%
d	9	1.2%	426	2.4%
e	46	6.1%	2,511	13.9%

「a. 外来語表記法の対立」に分類され、「アーティスト—アーティスト」のゆれは「b. 片仮名と英字の対立」に分類される。したがって《アーティスト》は類型aとbとの二つに分類されることになる。

表7を見ると、「b. 片仮名と英字の対立」に属する語が最も多く、異なりで六一・三%、延べで五九・〇%と、異なり・延べ共に約六割を占めている。次いで、「a. 外来語表記法の対立」が多く、異なりで二六・四%、延べで一九・九%となっている。一般に外来語表記のゆれとして問題となるのが、この類型aである。外来語表記のゆれの実態を明らかにするためには、具体的にどのような外来語表記の対立が見られるのかを明らかにする必要がある。例えば、先に挙げた《アーティスト》では、類型aに属するゆれとして「アーティスト—アーティスト」がある。これは外来音「アーティスト」と書くか、「チ」と書くかというゆれであり、外来語表記法の対立の一つとして「ティ—チ」を挙げる

ことができる。

このように類型aを、具体的にどのようなゆれなのかという観点から分類し、各分類に属する語の異なり語数を表8としてまとめた。なお、表7に示した類型と同様、語によっては複数の種類に分類されるものがある。例えば、《ウィンドウ》の表記は、コアデータ中に「ウィンドウ」「ウィンドー」の二種類がある。ここには、外来音「ウイ」と書くか「ウイ」と書くかというゆれと、語末の才段長音を長音符号で書くか「ウ」と書くかというゆれがある。したがって、《ウィンドウ》は「ウイ—ウイ」「オー—オウ」の二つに分類されることになる。また、表8ではコアデータ全体を

表8：外来語表記法の対立（異なり、延べ）

分類	異なり	語 例
語末長音の有無	51	カテゴリーーカテゴリー
エーーエイ	23	メールーメール
清音ー濁音	14	スムーズースムーズ
語中長音の有無	10	マシンーマシーン
ウイーウィ	8	ウイークーウィーク
ウエーウェ	8	ウェブーウェブ
イアーイヤ	6	ダイヤモンドーダイヤモンド
バーヴァ	6	キャラバンーキャラヴァン
チーティ	5	アーティストーアーティスト
ホーフォ	5	テレホンーテレフォン

まとめて集計し、異なり語数が五以上のゆれのみを掲げた。ここに掲げた一〇種類のゆれで全体の約六割を占める。

表8を見ると、語末長音の有無が五一語で最も多く、「エーーエイ」の対立に属するものが二二語でそれに次ぐ。この二種類で全体の約三割を占める。以下、異なり語数の最も多い「語末長音の有無」について詳しく見ていく。

「語末長音」の有無の対立を持つ語（五一語）のうち頻度一〇以上の語を表9に示した。表9では、各語について語末長音を長音符号で示

した表記（以下、長音あり）の用例数とその割合、語末長音の長音符号を略した表記（以下、長音なし）の用例数とその割合も示した。その上で、略した表記の用例数の割合が高いものから低いものへと順に配列した。長音ありの用例数の割合が高いものから低いものへと順に配列した。

『外来語の表記』では、語末長音のうち、英語の語末-er、-or、-arなどに当たるものの表記について、次のように定めている。

英語の語末-er、-or、-arなどに当たるものは、原則としてア列の長音とし長音符号「ー」を用いて書き表す。ただし、慣用に応じて「ー」を省くことができる。

〔例〕 エレベーター ギター コンピューター マフラー
 エレベータ コンピュータ スリッパ

語末ア段長音は長音符号で表記することを原則としており、語によって

表9：語末長音の有無の対立（頻度10以上）

語彙素	長音あり	%	長音なし	%
モニター	17	94%	1	6%
ユーザー	38	93%	3	7%
ドライバー	25	93%	2	7%
サポーター	14	88%	2	13%
カウンター	23	85%	4	15%
マネージャー	17	85%	3	15%
ファイバー	11	85%	2	15%
コーディネーター	10	83%	2	17%
テクノロジー	10	77%	3	23%
レーダー	8	73%	3	27%
ボディー	31	72%	12	28%
パーティー	12	71%	5	29%
メモリー	19	66%	10	34%
コンピューター	25	49%	26	51%
コメディ	4	40%	6	60%
エア	5	33%	10	67%
シティー	5	33%	10	67%
セキュリティ	2	10%	19	90%
ブラウザ	1	8%	12	92%
ヘア	1	7%	13	93%

は慣用に基づく長音符号を省いた表記も認めるといふものである。

表9には、語末がア段長音の語が《モニター》《ユーザー》《ドライバー》《サポーター》《カウンター》《マネージャー》《ファイバー》《コーディネーター》《エアー》《ブラウザ》《ディーネーター》《レーダー》《コンピューター》《エア》《ブラウザ》《ヘアー》の一二語ある。これらは、『外来語の表記』の原則と慣用との間で表記がゆれている語ということになる。これら一二語のうち《モニター》《ユーザー》《ドライバー》は九割以上が、《サポーター》《カウンター》《マネージャー》《ファイバー》《コーディネーター》は八割以上が『外来語の表記』の原則に当たる長音ありの表記となっている。慣用に当たる表記が見られるものの、それはわずかであり、『外来語の表記』の原則に当たる表記にほぼ固定していると言える。一方、《ブラウザ》《ヘアー》は九割以上が長音なしの表記となっており、『外来語の表記』の慣用に当たる表記にほぼ固定していると言える。

長音あり、又は長音なしで、表記がほぼ固定している語があるのに対

して、《レーダー》《コンピューター》《エア》のように表記が固定していない語もある。特に《コンピューター》は長音ありと長音なしとが拮抗している。

語末がア段長音以外の語は六語ある。これらは、英語の語末が *-dy* *-gy* *-ry* *-ty* で、イ段長音の語である。これらのうち、《セキュリティ》は九〇%が長音なしの表記となっており、この表記でほぼ固定していると考えられる。一方、長音ありの表記で固定している語はなく、《テクノロジ》《ボディ》《パーティー》《メモリー》《コメディ》は表記が固定せず、ゆれている。

英語の語末が *-dy* *-gy* *-ry* *-ty* の語をどのように表記するかについて『外来語の表記』に規定はないが、一般の国語辞典等では、イ段長音として語末に長音符号を用いて書き表されている。今回の調査結果では、一般の国語辞典等における表記とは異なる表記、つまり語末長音を略した表記が多く用いられていることが明らかとなった。

以上、外来語表記法の対立のうち、「語末長音の有無」の対立についてコアデータ全体の状況を把握した。次に表9に示した「語末長音の有無」の対立がある外来語の表記のうち、長音なしの表記に注目し、その出現状況を媒体別に見ていくこととする。長音なしの表記の出現状況を媒体別に整理して、表10に示した。

まず語末がア段長音の語を見ていく。長音なしの表記が定着している《ブラウザ》《ヘア》については、その長音なしの表記「ブラウザ」「ヘア」は、ウェブ・雑誌のみに出現しており、使用される媒体に偏りが見られる。ウェブ・雑誌で多用されていることよって、全体でまとめて集計した際に、長音なしの比率が八割を超えたのである。一方、長音ありとなしとが拮抗している《コンピューター》については、長音なしの表記「コンピューター」は、新聞を除く全ての媒体に出現しており、媒

大規模コーパスを活用した外来語表記のゆれの調査

表 10：語末長音なしの表記（頻度 10 以上、媒体別）

媒体	用 例（語末長音なし）
ウェブ	エア(2), カウンタ(4), コンピュータ(1), サポータ(2), ドライバ(2), ブラウザ(1), ヘア(5), ユーザ(3)
	コメディ(2), セキュリティ(7), パディ(1), ボディ(2), メモリ(10)
書籍	コンピューター(10)
	シティ(5), セキュリティ(1), テクノロジ(3), ボディ(1)
雑誌	エア(8), コンピュータ(6), ブラウザ(11), ヘア(8), マネージャ(3), モニタ(1)
	コメディ(4), シティ(2), セキュリティ(6), パーティ(5), ボディ(8)
新聞	シティ(3)
白書	コーディネータ(2), コンピュータ(9), ファイバ(2), レーダ(3)
	セキュリティ(5)

体の偏りは見られない。まさに表記がゆれている語と言える。その他の語については、ウェブ・雑誌で多く用いられており、使用される媒体に偏りが指摘できる。

次に語末がイ段長音の語を見ていく。新聞に「シティ」、白書に「セキュリティ」が見られるものの、基本的に語末長音なしの語はウェブ・書籍・雑誌に多く出現している。先に見た「ブラウザ」「ヘア」と同様、長音なしが出現する媒体に偏りがある。なお、《セキュリティ》は長音なしの表記で固定している語で、新聞を除く全ての媒体に出現している。今回の調査結果からは、長音なしの「セキュリティ」で固定していると言える。

ここで注目した長音なしの表記は、ア段長音については『外来語の表記』で慣用とされる表記であり、イ段長音については一般の辞書類とは異なる表記である。つまり、いずれも原則とは異なる表記である。コアデータ全体で見えた場合、《ブラウザ》《ヘア》は、長音なしの比率が八割を超えており、長音なしでほぼ固定していると考えられた。しかし媒体別に見ると、長音なしの使用には媒体に偏りがあることが分か

る。また、「コンピュータ」「セキュリティ」を除き、ア段長音・イ段長音ともに長音なしの表記は、やはりウェブ・雑誌で多く用いられている。表5・6で見たようにウェブ・雑誌は、外来語表記のゆれの割合が高い媒体である。本節で見た語末長音の表記についても、慣用に当たる表記（原則とは異なる表記）が多く用いられており、このことからウェブ・雑誌は表記の自由度が比較的高い媒体ということが確認される。

一方、新聞で長音なしは「シティ」のみである。新聞における外来語の表記が基準に従った、統一の図られたものとなっていると言える。書籍・白書は、表記のゆれのある語の割合（表5・6）は新聞よりも低い。長音なしの表記は、新聞よりも多く見られた。ただ、書籍について見ると、「コンピュータ」は十例のうち七例が森川和行『医療法人・医療生協の会計改革』に集中している。ほかの複数例ある表記「シティ」「テクノロジー」も、それぞれ特定の著者が用いたものである。つまり、特定の著者の使用に偏ったものである。白書についても、長音なしの表記は、情報通信白書（「コンピュータ」五例、「セキュリティ」二例、「ファイバ」二例）に偏っている。このように特定の著者、特定の省庁で用いられているものであり、媒体全体にわたる傾向とは言えない。

六 終わりに

本稿では、BCCWJコアデータを資料として、白書・新聞・雑誌・書籍・ウェブの五つの媒体を対象に外来語表記のゆれの実態調査を行った。その結果をまとめると、次のようになる。

- (1) 英字による表記を除いて集計すると、外来語表記のゆれの割合は、各媒体とも5%以下と低い。しかし媒体差は指摘でき、ゆれの割合の高い媒体と低い媒体と、大きく二つに分けることができる。前者

はウェブ・雑誌であり、後者は白書・新聞・書籍である。

- (2) 五つの媒体がゆれの割合の高い媒体と低い媒体とに二分される要因の一つとして、その媒体に共通の表記の基準があるか否かということが考えられる。
- (3) 外来語表記のゆれの類型については、「片仮名と英字の対立」に分類されるものが最も多く、「外来語表記法の対立」に分類されるものがそれに次ぐ。
- (4) 「外来語表記法の対立」としては「語末長音の有無」による対立が最も多い。また長音なしの表記は、ウェブ・雑誌に多く用いられる傾向があり、媒体による偏りを指摘できる。また、このことからウェブ・雑誌は表記の自由度が比較的高い媒体と行うことができる。

今回の調査では、主として媒体差という観点から使用実態を見ていき、外来語表記のゆれの割合に媒体差が見られる要因について、表記の基準という観点から考察を加えた。しかし語表記のゆれの要因としては、他の要因も考えられる。例えば、著者、分野、語義、語構成などである。今後は、こうした説明変数を導入し、より詳細な分析を行って、外来語表記のゆれの実態、ゆれが生ずる要因について詳細に明らかにする必要がある。また、BCCWJのコアデータのみではなく、全体を対象とした調査を行う必要がある。今後の課題としたい。

注

- ① 「外来語」言い換え提案については、国立国語研究所「外来語」委員会（二〇〇六）を参照。
- ② BCCWJの設計等については、山崎誠（二〇〇七）、前川喜久雄（二〇〇八）を参照。コアデータの設計・構成等については、小椋秀樹・小木曾智信・小磯花絵ほか（二〇〇九）を参照。
- ③ 国立国語研究所（一九八三）でも語形のゆれと見るか表記のゆれと見る

か判定の難しいものがあることが指摘されている。国立国語研究所(一九八三)では厳密に「バイオリナーヴァイオリン」のようなゆれのみを語表記のゆれと考えている。ただし語形のゆれか表記のゆれか判定の難しいものも付表に一覧している。

④ Undicの概要については伝康晴・小木曾智信・小椋秀樹ほか(二〇〇七)を参照。

⑤ 小椋秀樹(二〇一一)ではBCCWJコアデータを資料として、語種別に語表記のゆれの割合を調査した。この調査では、語表記のゆれの認定について「任意の二つの書字形が、同じ語彙素・語形・品詞を持つ場合、同じ語の表記のゆれと認める」とした。集計用語形を設けていなかったため、「コンピューターコンピュータ」のような一般に外来語表記のゆれとされているものが除外されているという問題があった。また、表記のゆれの類型等の分析では、和語動詞のみを取り上げており、外来語については分析を行っていない。

⑥ BCCWJでは、英文は形態素解析(単語分割、品詞・読み・語種情報の付与等)の対象から除外したが、日本語文中に出現した英単語や商品名等については形態素解析の対象とし、品詞・読み・語種等の情報を付与している(外来語として扱った)。そのため、表5の集計では英字で表記された外来語も対象となっている。

参考文献

小椋秀樹・小木曾智信・小磯花絵・富士池優美・宮内佐夜香・渡部涼子・竹内ゆかり・小川志乃・小西光・原裕・中村壮範(二〇〇九)『現代日本語書き言葉均衡コーパス』における形態論情報付与作業の進捗状況』『特定領域「日本語コーパス」平成20年度公開ワークショップ(研究成果報告会)予稿集』、五七～六四頁

小椋秀樹(二〇一一)「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—」『第一回コーパス日本語学ワークショップ予稿集』、三二一～三二八頁

国立国語研究所(一九八三)国立国語研究所報告七五『現代表記のゆれ』

国立国語研究所「外来語」委員会(二〇〇八)『分かりやすく伝える外来語言い換え手引き』ぎょうせい

伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵(二〇〇七)「コーパス日本語学のための言語資源—形態素解析用電子化辞書の開発とその応用—」『日本語科学』二二、一〇一～一二三頁、国書刊行会

前川喜久雄(二〇〇八)「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発」『日本語の研究』四一、八二～九五頁

宮島達夫・高木翠(一九八四)「雑誌九十種資料の外来語表記」国立国語研究所報告七九『研究報告集』五、四三～七六頁

山崎誠(二〇〇七)『現代日本語書き言葉均衡コーパス』の基本設計について』『特定領域「日本語コーパス」平成一八年度公開ワークショップ(研究成果報告会)予稿集』、一二七～一三六頁

付記

本稿は、国立国語研究所基幹型共同研究「多角的アプローチによる現代日本語の動態の解明」(プロジェクトリーダー…相澤正夫)及び同「コーパス日本語学の創成」(プロジェクトリーダー…前川喜久雄)による研究成果の一部である。

(本学文学部准教授)