

『計量国語学』アーカイブ

ID	KK300501
種別	論文A
タイトル	近代二字漢語における同語異表記の実態と変化 — 形態論情報付きコーパスを用いて —
Title	Diachronic Variation in Orthography of Two-Character Sino-Japanese Words in Modern Japanese: A Corpus-based Study
著者	間淵 洋子
Author	MABUCHI Yoko
掲載号	30巻5号
発行日	2016年6月20日
開始ページ	257
終了ページ	274
著作権者	計量国語学会

論文 A

近代二字漢語における同語異表記の実態と変化

—形態論情報付きコーパスを用いて—

間淵 洋子 (明治大学大学院生・日本学術振興会特別研究員 DC)

要旨

近代における漢語表記の多様性と、現代にかけての表記の淘汰・統一の様相を明らかにすることを目的として、形態論情報付き近代語・現代語コーパスを用い、漢語の異表記の実態調査・分析を行った。近現代両用の約 2,000 語について、異表記の数や個々の表記の使用率等を調査し、以下 3 点を明らかにした。1) 一語が有する表記数は、近代より現代で少ない。2) 現代で消失・衰退した表記が、近代では高頻度で使用されていた事例が多く見られる。これは、近代において一般的だった表記が衰退し淘汰されたことや、語の標準的な表記が変化・交替したことを表している。3) 表記の淘汰や交替は、近代で既に生じている語とそうでない語がある。近代に生じた表記の統一化は、近代学校制度の確立や活版印刷による出版文化の興隆などを背景とした、言語使用の実用性・合理性重視が起因していると考ええる。一方、昭和以降に生じたと思われる表記の淘汰・交替は、当用漢字等の国語政策の影響が大きいものと推測される。

キーワード：漢語、近代語、表記の多様性、通時的変化、形態論情報付きコーパス

1. 研究の背景

日本史における近代は、政治、国際環境、科学、哲学、文化、全てにおいて最も激しく揺れ動き、劇的な変化がもたらされた重大な転換期であった (坂野 2012)。日本語史における近代もまた、それまでの漢文・文語中心のスタイルから今日の書き言葉に繋がる口語文へと移り変わる最大の変革期であったが、この変革期の日本語 (= 近代語) は、模索的・流動的であり、現代語の確立に至るまでに多くの「ゆれ」や「変化」が見られたことが、これまで多く指摘されてきた (今野 2012)。特に、近代化に伴う新しい物質・事象・概念の大量な流入により、飛躍的に語彙を増やした漢語 (宮島 1977) については、表記や語法の周辺的なバリエーションとその消長が、現代語との差異を中心にこれまで多く取り上げられてきた。

例えば、同音の語に対して全く同じ意味で複数の表記が用いられるといった事象がある。以下にコーパスからの例を挙げよう。なお、用例の引用においては、問題とする漢語に【 】を付し、原文の傍点・傍線・ルビ等は省略する。また、用例本文に続けて“(『雑誌名』年号、著者「タイトル」)”の形式で出典情報を示す。

- (1) 本館は【政治】、經濟、科學、哲學、宗教、文學、美術、娛樂等に關する全國無數の出版物中より (『太陽』1909年5号, 大橋新太郎「在外日本人諸君に告ぐ」)
- (2) 現今同志社に在る神學科の外に於て、【政事】、經濟、哲學、文學、法學等に在り、(『国民之友』1888年34号, 新島襄「同志社大學設立の旨意」)

例 (1), (2) に見るように, 近代において「セイジ」という漢語は, 同様の文脈で「政治」「政事」という異なる表記が用いられているが, これらは, 現代においてほぼ統一的に「政治」と表記される. ここでは, このように, 同一の語に対して複数の表記が存在する場合, それらの表記を「同語異表記」と呼ぶ. 「同語異表記」は上掲した漢語「セイジ」に固有に見られる現象ではなく, 多くの漢語に同様に見られることがこれまで数多く報告されてきたが, その実態は, 必ずしも明らかになっているとはいえない.

そこで, 本稿では二字漢語¹を対象として, 「同語異表記」について以下の点を明らかにすべく, コーパスを対象とした調査・考察を行う.

- 1) 近代語・現代語を通じて, 同語異表記の実態はどのようなものであるか.
- 2) 近代語と現代語の間で語の表記が変わったものはどの程度あるか, また, 具体的にどのような語であるか.
- 3) 語の表記が変わった時期はいつか, また, その要因は何か.

2. 先行研究の整理

漢語の同語に対する表記バリエーションについては, これまで主に, 漱石を中心とした明治期の文学作品における特異な用字として注目され, これに端を発する今野の一連の研究や (今野 2008, 2012, 2014), 「セイジ (政治/政事)」や「キシヨウ (気象/気性)」等の個々の語について近世から近代にかけて詳細に表記と語義の変遷を追う田島の一連の研究 (田島 1998, 2009), 国語辞書から「異語の同語」を抽出し, 類義語との関連を視野に語義や字音について整理・検討し, 国語施策や新聞・法律・速記等における用字用語の検討に発展させた武部の一連の研究 (1981) など, 多くの先行研究により実に様々な資料から多くの語について報告がなされてきた. これらはいずれも知見に富み, 漢語や日本人の言語感覚の本質に迫る研究であるが, 同語異表記という事象の総体やその各時代における実態については明らかにされてこなかった.

この問題を解決する模索的研究として, 筆者は間淵 (2013) において, 2005 年公開された『太陽コーパス』の XML ファイルに格納された校訂情報を用いて一部の異表記対を抽出し, 約 80 語について表記の使用状況や, 異表記の構成字義・字音による分類等を試みた. その結果, 『太陽コーパス』においては, 今日一般的ではない表記の勢力が高いも

1 本稿では, 「漢語」を, 成立や出自を問わず, 日本漢字音で読まれる漢字表記語全ての意で用いる. また, 分析対象「二字漢語」に限定するのは, ①二字漢語が, 日本語の漢語語彙の種類や頻度において中心的な位置を占めていること, ②三字以上の漢語の殆どは, 二字漢語が語基になって成り立っていること, ③一字漢語は, 和語の複合要素との結合が固く, 他の二字以上の漢語と文法的振る舞いが大きく異なること, などによる. 一字漢語や三字以上の漢語は, 二字漢語の様相を確認した後に別途検討する必要があると考えるため, 今回の分析対象からは外し今後の課題としたい.

のも多く見られること、それらの一部は『太陽コーパス』の収録期間 30 年間で衰退し、今日一般的な表記に統一されたことなどを示した。しかし、この調査はコーパスの校訂情報から得られた異表記対のみが分析対象であり、元となる校訂情報が、明確な校訂基準や整合的で妥当な同語判定基準に基づくものとは言いきれないために、網羅的・体系的な表記バリエーションの取得が不可能であった。

そこで、本稿では、上記問題点を解消する方策として、現在公開されている（または開発中の）近代語と現代語の形態論情報付きコーパスを用い、同語異表記の網羅的・体系的実態把握を目指すこととする。

3. 研究方法

3.1 調査対象コーパス

近現代における漢語表記の実態を把握するために、現在公開されている近代語と現代語のコーパスを用いた調査を行う。調査対象とするコーパスの概要を表 1 に示す。

表 1: 調査対象コーパス

時代	コーパス	発行年	語彙量（自立語）
近代	明六雑誌	1874-1875	11 万語
	国民之友	1887-1888	57 万語
	太陽（増補）	1898-1925	555 万語
現代	BCCWJ 出版 SC	2001-2005	3450 万語

現代語のコーパスとしては、2011 年公開の『現代日本語書き言葉均衡コーパス（以下、BCCWJ と表記）』の出版サブコーパス（以下、SC と表記）を用いる。

BCCWJ は、日本初となる、また唯一の、大規模バランストコーパスであるが、その中でも出版 SC は、2001-2005 年に出版された書籍・雑誌・新聞について、極めて精密な文字量調査に基づき推定された母集団に対して、厳密・精緻なランダムサンプリングにより、その縮図となるよう設計されたコーパスであり（前川 2008）、日本語の一般的な書き言葉のサンプルとしてまさに好適なデータである。

近代語コーパスとしては、2013 年公開の『明六雑誌コーパス』、2014 年公開の『国民之友コーパス』、更に、2005 年公開の『太陽コーパス』と 2006 年公開『近代女性雑誌コーパス』を作業用に形態素解析したデータを用いる。

『明六雑誌コーパス』は、明治初頭の 1884（明治 7）年から 1885（明治 8）年にかけて刊行された、日本の学術雑誌の先駆けとされる『明六雑誌』の全 43 号全文を対象とするコーパスである（近藤ほか 2012）。『国民之友コーパス』は、徳富蘇峰の設立した民友社により 1887（明治 20）年から 1898（明治 31）年にかけて刊行され、当時の著名な有識者が多く名を連ね、雑誌発行部数は第 1 位で全国の知識層に広く普及し彼らの言論活動に大きな影響を与えたとされる雑誌『国民之友』の 1887（明治 20）～1888（明治 21）年刊行分である 1～36 号の全文をコーパス化したものである（近藤ほか 2014）。『太陽コーパス』は、博文館から刊行され、分量の多さ・ジャンルの広さ・執筆者の多彩さ・読者層の厚さ、いずれをとっても当時類を見ない資料とされる総合雑誌『太陽』（1895～1928 年）

の、1895（明治28）年、1901（明治34）年、1909（明治42）年、1917（大正6）年、1925（大正14）年の5年分の全文を格納し、言文一致を経て口語体による書き言葉が安定し普及する時期（明治時代後期～大正時代）の書き言葉を代表できるコーパスとして開発・公開された（田中 2005）。これら近代語コーパスは、BCCWJ と異なりバランストコーパスではなく、またサンプルコーパスではないが、いずれも当時の日本語の書き言葉を知る上で欠くことのできない代表的な重要資料をコーパス化したものであり、研究価値の高いデータである。

調査対象としたこれらのコーパスの最も特徴的な点は、形態素解析が施されている点である。さらに、解析に用いた解析辞書 UniDic は、以下のような特長を有しており、この情報を用いることで、本研究で対象とする同義異表記を持つ漢語を容易に抽出することができる。

- 「短単位」と呼ばれる斉一な単位により単語の区切り目が認定される。特に漢語については、2 文字 1 単位とする単純なルールを用い、単位認定に揺れが生じない。
- 語彙素＞語形＞書字形・発音形といった階層構造を持ち、表記の揺れや語形の変異にかかわらず同一の見出しを与える。
- 品詞、語種等の言語研究に有用な情報を有する。

例えば、「詩歌」という漢語を例に UniDic によってどのような形態論情報が付与されるかを示すと図 1 のようになる。

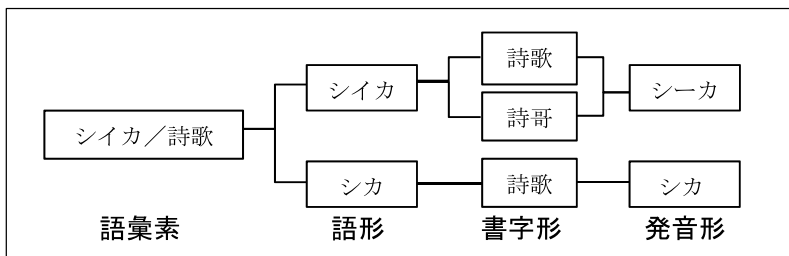


図 1: UniDic の階層構造

「語彙素」とは、辞書見出しに相当するものであり、語の様々なバリエーションを統合的に位置づけるために設定されている最上位の階層である。標準的な読みと標準的な表記によって規定される。語彙素は、その下位に「語形」という階層を持つ。ここには、読み方・発音のバリエーションが含まれている。図 1 に挙げた例「詩歌」については、一般的な読み「シイカ」の他に、「シカ」という読みが含まれる。形態は異なるが、同語と判別されたものが兄弟関係として位置づけられる。それぞれの語形は「書字形」という下位階層を持つ。これは、実際に現れる表記形であり、文字種・字体・用字などのバリエーションが兄弟関係として位置づけられる。つまり、この「語彙素」の情報を用いることで、語の様々なバリエーションを一括して検索することが可能となっている。

これらの形態論情報は全て、国立国語研究所の形態論情報データベース（小木曾・中村 2011）に格納されており、本研究では 2015 年 6 月時点の短単位情報データを用いた。

3.2 調査対象語の選定

次に、設定した調査対象コーパスから、調査・分析対象とする語を抽出・選定する。本稿で調査・分析の対象とする語は「現代語とは異なる二字漢語の同義異表記関係にある表記対」である。そこでまず、現代と近代の漢語表記の全体像を把握し近代の特異性を確認するため、「現代と近代の両方に用いられている二字漢語」を抽出することとし、短単位情報データベースより、以下の条件に当てはまる用例を抽出した。

「語種」が「漢」（漢語）であり、「語彙素」の表記が2字であるもののうち、調査対象の近代語コーパスにも現代語コーパスにも用例のある「語彙素」

その結果、異なり語彙素数約 28,300 語、異なり書字形数 36,000 語、近代語コーパス延べ 182 万用例、現代語コーパス延べ約 529 万用例が抽出された。その中から、数詞、誤解析、誤字や当て字と思われるもの、孤例などを除き、更に両時代のコーパスにおいて、語彙素の自立語 100 万語あたりの調整頻度が 10 以上の語²を分析対象として、2,075 語彙素、3,053 書字形（仮名表記のものを除く³）を抽出した。

4. 調査結果

4.1 近現代における漢語の表記事情

抽出した 2,075 語彙素の同語異表記の実態を明らかにするため、近代・現代それぞれの時代において、実際にどのような表記のバリエーションが見られるのかを確認しよう。

まず、1 語が持つ表記バリエーションの数について検討する。同一の語彙素に対して、幾つの書字形が用いられたかを、横の列に近代語コーパスにおける書字形数、縦の行に現代語コーパスにおける書字形数を配してクロス集計したものが表 2 である。

表 2: 語彙素に対する書字形数

近代 \ 現代	1	2	3	4	5	6	総計
1	1618	195	33	6			1852
2	93	83	19	10		2	207
3	3	2	7	1	1		14
4				1	1		2
総計	1714	280	59	18	2	2	2075

表 2 より、両時代を通じて一つの表記のみ用いられている語彙素が 1,618 語（全体の約

2 調整頻度 10 の語は、現代語コーパスの粗頻度で 350 語、近代語コーパスの粗頻度で約 60 語であり、例えば、現代語では「委任」「故障」「正体」「遭遇」「排気」などにあたる。計量的な分析が可能となる程度の用例数を確保し、かつ基本的・一般的な語彙が抽出されることを意図して定めた基準である。

3 本調査では、仮名表記の書字形を調査分析の対象外としたが、近代から現代への漢語表記の変遷においては、漢字表記から仮名表記へという変化も見逃せない。仮名表記される理由も個々の語によって様々であり、後述する表記割合や異表記の消失・衰退などと関連する可能性も高く、明らかにすべき大きな課題であるが、これについては、今後の課題としたい。

78%)と大勢であり、これを含めて近代・現代での書字形数が一致する語彙素(表中背景色付きセル)は1,709語(約82%)と8割以上を占めていることが分かる。一方、書字形数に差のある語彙素では、近代から現代で書字形が増えた語彙素(表中太字下線付き)が98語(全体の約5%)、減った語彙素(表中太字斜体)が268語(全体の約13%)であり、現代では近代に比べて表記のバリエーションが減少している傾向が見て取れる。

次に、実際の出現形である書字形と、語の一般的・標準的な表記である語彙素の表記との差異がある場合に、その差異がどのようなものであるかを検討する。横の列に各時代での書字形数の単複を、縦の行に書字形と語彙素表記の関係性を配してクロス集計したものが表3である。書字形数の単複は、時代ごとに、ある語彙素の実際に出現する際の書字形が単一であるか、複数あるかを表し、近現代を通じて単一であるものを「通時単数」、近現代を通じて複数であるものを「通時複数」、近代でのみ複数であるものを「近代複数」、現代でのみ複数であるもの「現代複数」として示す。また、書字形と語彙素表記の関係性は、差異の見られる文字同士の関係性を表し、以下の通り分類する⁴。

一致 書字形と語彙素表記が一致しているもの

(例:「自分」など)

新旧 新字体・旧字体の差異のみがあるもの

(例:「亜鉛↔亞鉛」における「亜」と「亞」)

異体 新旧に含まれない異体字関係の差異があるもの(新旧を含む場合あり)

(例:「寄付↔寄附」における「付」と「附」)

別字 異体字に含まれない別字関係の差異があるもの(新旧、異体字を含む場合あり)

(例:「一緒↔一所」における「諸」と「所」)

表3から、現代のみで複数表記を持つ場合、97例中85例と9割程度が字体の新旧の差異に集約される異表記であるが、近代のみで複数表記を持つ場合は、234例中137例と6割程度が新旧以外の異体字・別字を含む異表記であることが分かる。

表3: 書字形数と字体関係

書字形数 字体関係	通時 単数	通時 複数	近代 複数	現代 複数	総計	書字形例
一致	1309				1309	自分, 必要, 問題, 時間, 世界
新旧	309	26	97	85	517	社會(会), 學(学)校, 經(経)濟(済)
異体字		15	18	1	34	箇(個)人, 最后(後), 規模(模)
別字		85	119	11	215	已(以)上, 情(状)態, 家屬(族)
総計	1618	126	234	97	2075	

以上の観察から、近代と現代における漢語表記の実態として、以下を確認することができた。

- 1) 近現代に共通して用いられる漢語は、両時代を通じて語と表記が1対1で対応するものが8割近くを占めている。

4 字体の新旧については常用漢字表を、異体字の判断については『角川新字源』を用いた。

- 2) 近代は、現代と比較して、1 語に対応する表記数が多い。近代から現代にかけて、表記バリエーションは減少している。
- 3) 現代で表記にバリエーションが見られる場合、そのほとんど（9 割近く）は字体の新旧の差異によるものである。字体の新旧は国語施策的な背景が明確であり、実質的な表記バリエーションとは言いがたい。
- 4) 近代で表記にバリエーションが見られる場合は、字体の新旧以外の差異によるものが半数以上（6 割程度）を占めており、現代における表記バリエーションとは性質が大きく異なる。

4.2 近代に見られる異表記とその比率

次に、近現代間で表記に差異が認められる語彙素について、差異の実態を確認する。

なお、ここでは、国語施策的な背景が明確である字体の新旧に関わる差異は分析対象から除外する。そのため、実際に出現した書字形の字体の新旧を統合（新字体に統一）した上で表記を比較し、標準表記と異なる表記バリエーション（以下「異表記」）として 288 書字形を抽出した。これらについて、近現代それぞれにおける出現頻度と使用割合（当該語彙素の全用例数に占める異表記の割合）を求め、双方の使用割合に 2 倍以上の開きがある 279 書字形を対象に、変化の実態を明らかにすることとした。このうち、10 書字形は近代で見られず、現代で新たに発生した異表記⁵であるが、本稿では近代の表記の多様性と現代にかけての表記バリエーションの減少を主に扱うため分析対象から除外する。こうして残った 269 書字形について、近代で有していた表記を現代で失ったもの（＝消失：現代での用例粗頻度が 0 のもの）と、近代から現代にかけて、語の出現全体に対する割合が著しく減っているもの（＝衰退：現代での用例粗頻度が 1 以上で、使用割合が近代での使用割合の半分以下のもの）とに分類し、近代での使用割合と共に示したものが表 4 である。書字形例には近代に見られる異表記を、現代の標準的な表記（＝語彙素表記）と対にして、異表記の粗頻度が多い順に 10 例までを示す。

表 4 に見るように、現代で消失・衰退している異表記には、近代において 50% 以上の割合で用いられていたものが 33 書字形（約 12%）あり、10% 以上の割合で用いられていたものになると 115 書字形（43%）にも及ぶ。

5 意図的な新規表記「豆腐（豆腐）」「処長（所長）」「公苑（公園）」「聚会（集会）」、当用漢字表による用字制限による代用表記「車両（車輛）」「年令（年齢）」、誤用の定着「撮映（撮影）」「同士（同志）」「固有（個有）」「合州（合衆）」等の例が見られた。

表 4: 異表記の分類と使用割合 (括弧内は標準表記, ゴシック体は異字)

異表記割合	消失 (計 166 語)		衰退 (計 103 語)	
	数	書字形例	数	書字形例
50% 以上	2	関聯(関連), 縮少(縮小)	31	附近(付近), 殖民(植民), 恢復(回復), 編輯(編集), 紀念(記念), 刺戟(刺激), 寄附(寄付), 根柢(根底), 活潑(活発), 附属(付属)
10-50 %	38	撰挙(選挙), 記臆(記憶), 権理(権利), 成蹟(成績), 成効(成功), 構造(構造), 本統(本当), 結構(結構), 模様(模様), 規模(規模)	44	充分(十分), 位地(位置), 蒸汽(蒸気), 書翰(書簡), 營養(栄養), 恰度(丁度), 情況(状況), 批難(非難), 撰択(選択), 差違(差異)
1-10%	81	箇人(個人), 干繫(関係), 冀望(希望), 一処(一緒), 計劃(計画), 太陽(太陽), 所置(処置), 個月(箇月), 簡短(簡単), 仕末(始末)	23	源因(原因), 容子(様子), 莫迦(馬鹿), 個処(箇所), 深切(親切), 領解(了解), 明了(明瞭), 功果(効果), 撰出(選出), 侵掠(侵略)
1% 未満	45	干係(関係), 関繫(関係), 法方(方法), 自個(自己), 已来(以来), 汽関(機関), 已下(以下), 反体(反对), 事状(事情), 完然(完全)	5	勞動(労働), 已上(以上), 已後(以後), 安神(安心), 編輯(編集)

このように近代での使用率が高いにも関わらず消失・衰退する表記がある一方で, 表 4 の異表記割合と消失, 衰退の語数分布からは, 近代での使用率が高いものほど現代でも消失せず残り, 使用率の低いものは消失する傾向があるように見える。

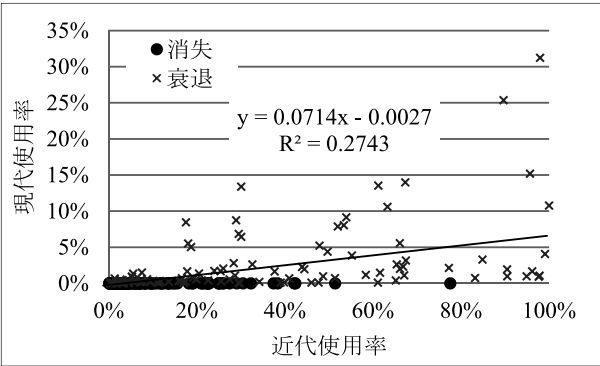


図 2: 近代・現代における異表記割合

そこで、近代での使用率と現代での使用率の間に相関があるかを確認するため、両者を散布図に示したものが図 2 である。図 2 より、近代表記率と現代表記率の相関係数 (R^2) は 0.274、消失、衰退の層別による近代表記率との相関比 (η^2) は 0.303 であり、近代での表記率が高いほど現代での表記率が高く異表記が残存する傾向はあるものの、強い相関があるとは言えなかった。

以上の観察から、近代と現代とで表記に差が見られる語の表記実態について、以下を確認することができた。

- 1) 字体の新旧を統合した上で比較した近現代間の表記の差異は、現代で新たに異表記が発生・増加したものが若干数ある以外、大多数において、近代に使用が見られた異表記が、消失または衰退（表記割合が著しく減少）するという方向性を持っている。
- 2) 現代において消失または衰退する異表記は、近代で元々使用率の低いものも多いが、10% 程度の語においては、半数以上という極めて高い割合で現代の標準的な表記（語彙素の表記）を凌駕して用いられており、一定の勢力を有していた。

4.3 異表記使用の経年変化

4.3.1 標準表記定着時期の推定

続いて、4.2 節に見た異表記が、消失・衰退し標準的な表記（以下「標準表記」とする）に統一された、または、優勢となったのはいつ頃であったかを確認する。

ここでは、標準表記の定着時期を特定するために、収録テキストの成立年（雑誌発行年）を手掛かりに、異表記使用の経年的な変化を調査することとし、変化の観察・分析が可能となる用例数を確保するため、近代コーパスでの粗頻度が 100 例以上の 74 語を対象を絞った。雑誌の発行年を目安とした a)1874 年、b)1887 年、c)1895 年、d)1901 年、e)1909 年、f)1917 年、g)1925 年の 7 カ年⁶での異表記と標準表記の使用割合を求め、(1) 標準表記の使用割合が 7 割を超え、かつ、(2) 続く年次で使用割合がそれ以下に下降しないことをもって、標準表記が定着したものとみなした。表 5 に移行時期の判定例を示す。表中の数字は実数が粗頻度、% 表示は標準表記の使用割合で、網掛けは 7 割を超えていることを表す。

語例①「工夫」の表記は、全ての年次において標準表記（上段）が 7 割以上の使用率を占めていることから、既に 1874 年の段階で標準表記が優勢であり、その勢力は変わっていない。これらの語は、元々標準表記が優勢であった可能性もあるし、それ以前に表記の勢力交替が完了しているとも考えられるが、ひとまず標準表記の定着時期を「1874 以前 (A)」として分類する。続いて語例②「希望」を見てみると、1887 年以降で標準表記の勢力が安定している。よって、年次と前の年次の間を標準表記の定着時期とみなし、「1874-1887 年 (B)」と分類する。一方、語例⑦「十分」の表記は、1901 年頃から概ね標準表記が優勢と思われるものの、1925 年では 50% に下がり、標準表記と異表記の出現頻度に差が見られない。このような語は、標準表記に統合されたとはみなし難いため「未統合拮抗（未：拮抗）」として分類する。また、語例⑧「区画」は、1887 年まで標準表記が優勢だが、1895 年以降異表記の勢力が増し、1909 年以降標準表記は出現しない。これらは、調

6 a)1874 年には 1874-5 年が、b)1887 年には 1887-8 年が含まれる。各年の総自立語数は、a)1874 年 11 万語、b)1887 年 57 万語、c)1895 年 122 万語、d)1901 年 114 万語、e)1909 年 106 万語、f)1917 年 101 万語、g)1925 年 112 万語である。

査した近代語コーパスの範囲では標準表記の定着期が特定できず、また明らかに異表記が優勢であることから、「未統合異表記優勢（未：異表記）」と分類する。

表 5: 標準表記の定着時期判定例（ゴシック体は標準表記）

語例	表記	1874	1887	1895	1901	1909	1917	1925	時期
①工夫	工夫	6	26	69	57	72	32	77	A
		75%	90%	75%	81%	91%	73%	96%	
	工風	2	3	23	13	7	12	3	
②希望	希望	2	205	196	257	369	303	237	B
		20%	95%	96%	89%	100%	99%	100%	
	冀望	8	11	9	31		2	1	
③権利	権利	63	54	259	230	187	183	125	C
		40%	45%	99%	98%	99%	99%	99%	
	権理	95	65	2	4	2	1	1	
④構造	構造	1	19	42	131	68	26	79	D
		33%	58%	35%	81%	100%	100%	100%	
	構造	2	14	77	30				
⑤記憶	記憶	1	10	14	13	86	102	105	E
		20%	14%	17%	18%	90%	100%	100%	
	記臆	4	62	70	60	10			
⑥記念	記念		4	10	3	16	74	106	F
			29%	10%	3%	12%	100%	98%	
	紀念		10	88	84	115		2	
⑦十分	十分	21	74	196	341	325	356	219	未 (拮抗)
		95%	35%	49%	54%	65%	67%	50%	
	充分	1	137	206	293	175	172	217	
⑧区画	区画	3	20	8	13				未 (異表記)
		100%	83%	32%	45%	0%	0%	0%	
	区劃		4	17	16	21	17	11	

このような基準で標準表記の定着期（移行期）を推定した結果を表 6 に示す。

表 6 より、調査対象とした 74 語中、(a) コーパス収録範囲以前から標準表記が優勢だった語が 10 語、(b) 近代語コーパスの収録範囲内で標準表記に移行されている語が 26 語、(c) 移行せず勢力が拮抗している語が 21 語、(d) 移行せず異表記が優勢な語が 17 語あることが分かった。

以上の観察から、近代と現代とで表記に差が見られる語の表記について、近代に見られる異表記が現代で用いられる標準表記に統合される時期について、以下の点を確認することができた。

1) 既に、近代において標準表記への移行が進んでいるものは半数程度存在する。残りの半数は、昭和期に統合された可能性が高い。

2) 近代語コーパスの収録範囲では、期間 E (1901-1909 年) に移行が進んだ語が多い。

表 6: 標準表記への移行時期

期間区分	語数	語彙素 (括弧内異表記. 五十音順, 旧字は新字に統合, ゴシック体は異字)
A. 1874 以前	10	関係(干係 / 干繫 / 関繫), 個人(箇人), 工夫(工風), 広大(宏大), 自動(自働), 成功(成効), 丁度(恰度), 馬鹿(莫迦), 非難(批難), 様子(容子)
B. 1874-1887	2	希望(冀望), 原因(源因 / 元因)
C. 1887-1895	5	位置(位地), 記事(紀事), 権利(権理), 収穫(収獲), 選挙(撰挙)
D. 1895-1901	4	構造(構造), 出版(出板), 状況(情況), 模様(模様)
E. 1901-1909	9	一緒(一所 / 一処), 規模(規模), 記憶(記憶), 結構(結構), 構成(構成), 差異(差違), 成績(成績), 本当(本統), 曖昧(曖昧)
F. 1909-1917	4	栄養(營養), 記念(紀念), 形態(形体), 選択(撰択)
G. 1917-1925	2	書簡(書翰), 蒸気(蒸汽)
H. 1926 以降 (拮抗)	21	回復(恢復 / 快復), 肝心(肝腎), 起因(基因 / 歸因), 起源(起原), 交付(交附), 作戦(策戦), 砂漠(沙漠), 支度(仕度), 実情(実状), 十分(充分), 縮小(縮少), 障害(障碍 / 障礙), 情勢(状勢), 植民(殖民), 総合(綜合 / 湊合), 知恵(智慧 / 智恵), 賃金(賃銀), 丁寧(丁寧), 布団(蒲団), 放棄(抛棄), 了解(諒解 / 領解 / 領会)
I. 1926 以降 (異表記優勢)	17	回転(廻転), 格好(恰好), 活発(活潑), 関連(関聯), 寄付(寄附), 区画(区劃), 経路(徑路), 根底(根柢 / 根抵), 刺激(刺戟 / 刺撃), 収集(蒐集 / 蒐輯 / 聚集), 付加(附加), 付近(附近), 付属(附属), 付着(附着), 付与(附与), 編集(編輯), 理屈(理窟)

4.3.2 表記割合の変化

前節において、それぞれの語の統合がいつ頃起こったのかを確認したが、その統合の過程を明らかにするために、本節では、上記 (a) から (d) のそれぞれのカテゴリーにおいて、異表記と標準表記の分布がどのように変化したかを詳細に見てみたい。年次毎に求めた標準

表記割合の推移を確認することで、変化の様相を明らかにする。

まず、(a) コーパス収録範囲以前から標準表記が優勢だった語（既統合語）について、標準表記割合の推移を見ると、図 3 の「丁度」「広大」のように、8 割程度の使用率で推移する語もあるが、それ以外のほとんどは「関係」「個人」のように、標準表記が極めて高い割合で安定している。

次に、(b) 近代語コーパスの収録範囲内で標準表記に移行した語（統合過程の語）については、図

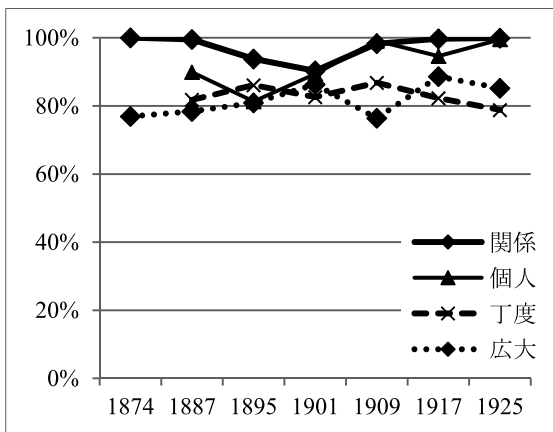


図 3: 既統合語の標準表記割合

4「原因」「状況」「本当」のように、比較的緩やかに増加するものがある一方で、図5「希望」「選挙」「記憶」「記念」のように、標準表記が急激に増加し一気に統合が進む語が多く見られる。統合過程の語 26 語について、前期から表記割合が 40 ポイント以上上昇して統合されるものを「急激な統合」、それ以外を「漸次的な統合」として変化のパターンを分類すると、概ね以下ようになる。

急激な統合 17 語（希望、選挙、権利、記事、収穫、出版、構造、成績、記憶、結構、規模、構成、記念、栄養、形態、蒸気、書簡）

漸次的な統合 9 語（原因、位置、状況、模様、一緒、本当、差異、曖昧、選択）

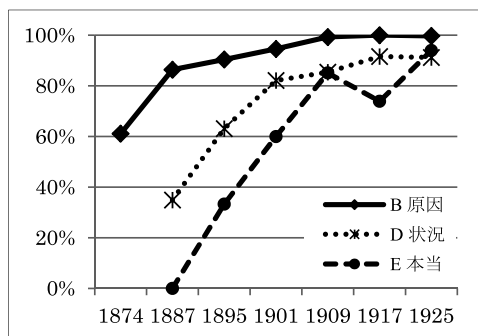


図 4：統合過程の語の推移①

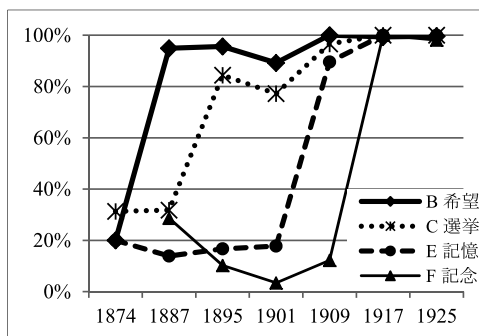


図 5：統合過程の語の推移②

また、(c) 近代語コーパスの範囲では統合が進まず勢力が拮抗している語は、図6「縮小」のように均衡状態にある語のほか、「賃金」のように緩やかに標準表記割合が上昇しており 1926 年移行に同様に増加し移行が進むと思われる語や、「了解」「作戦」のように、逆に標準表記割合が下降していくことが予想される語もあり、はっきりした統合の方向性は確認できない。

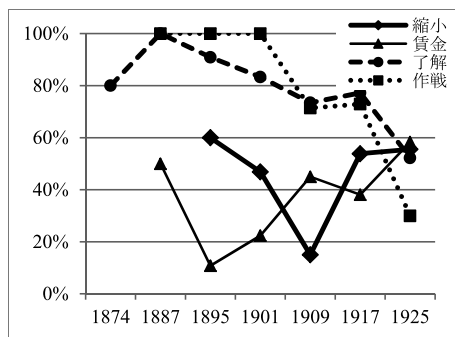


図 6：未統合拮抗の語の推移

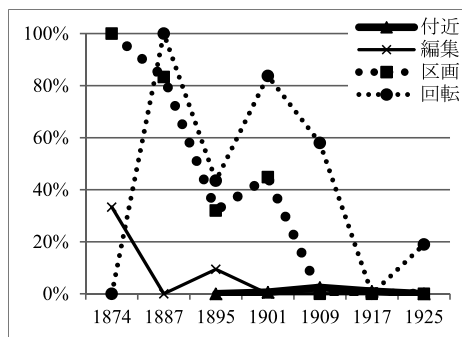


図 7：未統合異表記優位の語の推移

更に、(d) 標準表記に移行せず、むしろ異表記が優勢な語については、図7「付近」「編集」のように収録範囲内で一貫して標準表記割合が低い語（＝異表記が圧倒的に優勢な語）、「区画」「回転」のように一時期用いられていた標準表記が異表記に駆逐され淘汰さ

れたものが見られる点は注目される。異表記の優勢な 17 語のうち、「付近」「編集」と同様に、この期間で一貫して異表記が優勢な語について、「表記の統合」という点に着目して統合時期を確認すると、以下のものであった。

A.1874 以前 10 語（寄付, 活発, 寄与, 経路, 収集, 付加, 付近, 付属, 付着, 付与）

D.1895-1901 1 語（格好）

E. 1901-1909 3 語（関連, 区画, 理屈）

F. 1909-1917 3 語（回転, 根底, 刺激）

これを、先の表 6 に配置し、「（標準表記に限らず）表記の統合時期」を今一度まとめ直すと以下になるだろう。

A.1874 以前 20 語, B.1874-1887 2 語, C.1887-1895 5 語,

D.1895-1901 5 語, E. 1901-1909 12 語, F. 1909-1917 7 語

G. 1917-1925 2 語, H. 1926 以降 21 語

以上の観察から、近代における表記割合の推移について、以下の点を確認することができた。

- 1) コーパス収録範囲における表記割合の推移から、表記の統合過程は、緩やかに移行する語より、一気に統合される語が多い。
- 2) 標準表記への統合が進まず異表記が優勢である語の中には、収録範囲において一貫して異表記が統一的に用いられる語や、ある期間に異表記へと統合される語が少なくない。これは、標準表記への移行とは別に、「表記の統合」という流れがこの時期に強く働いていた可能性を示唆している。
- 3) 「表記の統合」という点で、表記割合をまとめ直すと、収録範囲以前に統合されていたものが 20 語、統合過程にあった語が 33 語、統合されず拮抗し併用されていた語が 21 語となり、調査対象の 74 語中 53 語、7 割以上の語で、近代において表記の統合がほぼ完了していたことになる。
- 4) 表記の統合時期は、収録範囲以前のものが 20 語と多いが、範囲内では 1901 年から 1909 年の間に見られるものが多い。

5. 考察—同語異表記における表記統合の背景

5.1 標準表記への移行

4 節に示した調査結果から、近代語コーパスの収録範囲における同語異表記の勢力分布の変化には、二つの観点が必要であることが明らかになった。一つは 4.3.1 節で見た (a) 「標準表記への移行」が進んだかという点、もう一つは、4.3.2 節で見た (b) 標準表記かどうかを問わず、「表記の統合」が進んだかという点である。

このうち、本節では「標準表記への移行」という点について、移行する語とその時期が意味するものは何かという点について検討したい。

4.3.1 節より、本稿で同語異表記の勢力分布の経年変化を明らかにするために調査対象とした 74 語のうち、36 語が近代において標準表記への統合が進み、残りの 38 語が未統合であった。このうち、近代において標準表記への統合が進む 36 語に特徴的なのは、その語自体の使用頻度が高いことであるように思われる。以下に、統合の期間区分と共に、語の粗頻度を掲載する。

表 7: 移行の有無と語の使用頻度

移行	語例 (括弧内は粗頻度)
済 (36)	関係(4610), 成功(1238), 個人(1196), 丁度(777), 非難(698), 馬鹿(686), 様子(683), 自動(528), 工夫(402), 広大(331), 原因(2048), 希望(1631), 選挙(3161), 位置(1307), 権利(1271), 記事(485), 収穫(337), 出版(664), 状況(566), 模様(521), 構造(489), 一緒(670), 成績(659), 本当(572), 記憶(537), 結構(434), 規模(370), 差異(325), 構成(207), 曖昧(157), 記念(512), 栄養(302), 選択(274), 形態(189), 蒸気(434), 書簡(355)
未 (38)	経路(119), 付近(654), 編集(342), 寄付(272), 活発(260), 収集(245), 付属(245), 付加(139), 付着(137), 付与(137), 格好(148), 関連(147), 区画(130), 理屈(359), 刺激(478), 根底(372), 回転(228), 十分(2733), 植民(757), 回復(718), 障害(266), 放棄(223), 知恵(210), 情勢(198), 丁寧(196), 起源(177), 実情(165), 支度(163), 総合(159), 縮小(144), 布団(144), 交付(139), 砂漠(109), 肝心(102), 了解(417), 起因(176), 作戦(164), 賃金(288)

移行の進んだ 36 語の頻度平均は 823, 未移行の 38 語の頻度平均は 317 であり, 両者は t 検定により 5% 水準で有意差のあることが確認された。

語の頻度は, 語の定着度に関与する。近代書き言葉としてこれらの漢語が定着するにつれ, 用法や表記の偏りが意識されやすくなり, 表記の固定化・統一化がより進む可能性がある。表記の統合と, 語の出現頻度に相関があるとすれば, 漢語の定着という観点から表記の統合・安定を位置づけることができるのではないかと。今後も検討を行いたい。

なお, 今回の調査で, 標準表記への統合が見られなかった語のうち, 特に近代で異表記が優勢である語については, その語構成漢字を調査した結果, 殆どの語が 1946 年 (昭和 21 年) に告示された「当用漢字表」や 1954 年 (同 29 年) の補正資料における代用字・代用表記の対象となっていることが分かった。よって, 恐らく, これらの異表記は, その時期を境に表記の統合が生じた結果, 現在のように消失・衰退したものと推測される。今後, 同時期の資料が調査可能になれば, 実態を確認できるものと考ええる。

5.2 表記の統合

4.3.2 節より, 近代において表記が集中して統合される時期があった。1901 年から 1909 年にかけての期間である。なぜ, この時期表記の統合が集中的に起こりやすかったのか, その時代背景を考えてみたい。

明治前期は, 国語・国字政策, 国語・国字論が盛んであった。この背景には, 1894 年の日清戦争頃から始まる, 社会の近代化・国家の近代化に伴う, 実用主義的・教育的国語国字政策があったものと思われる。また同時に, 1901 年から 1909 年にかけては, 口語化が特に進んだ時期である可能性が高い。『太陽コーパス』の年次別文体別記事数を調査した田中 (2005) によると, 以下の通り, 1901 年から 1909 年を境に口語記事の割合が文語記事を凌駕し, 以降も直線的に口語記事が増え 1925 年には口語化がほぼ完了する。

このような, 近代口語文の定着に伴い, 言語使用自体も合理性重視の気運が高まり, 複雑で合理的とは言えない同語異表記のような事象が淘汰されていくことになったものと思われる。明治 30 年代から 40 年代にかけては, 送り仮名や仮名遣いなど様々な表記についての統一施策が展開された。1916 (大正 5) 年には臨時国語調査会による「字体整理案」「漢語整理案」等が発表され, 漢字表記に関しても統制を指向した言語政策が敷かれた。これらの流れは, 上記のような合理性重視による表記統一の流れに影響を受けたものであろうし, 更にこれらの影響を受け揺れや通用といった表記のバリエーションを許容してき

た言語的環境が崩れ、より加速的に表記の統一化が実現されてきたのではないか。

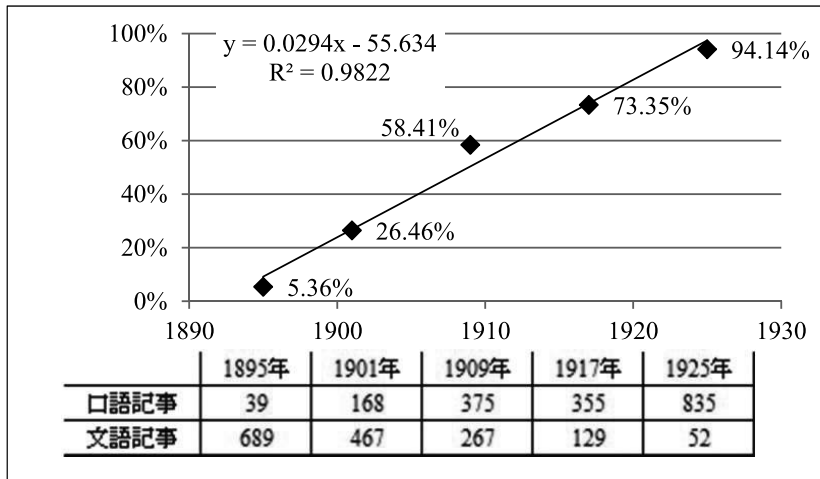


図 8:『太陽コーパス』の年次別口語記事率（田中 2005 を元に作成）

6. まとめ

本研究では、形態論情報付き近代語・現代語コーパスを用いて、漢語表記のバリエーションの網羅的な抽出を試み、近現代間に見られる表記の変化について、以下の結論を得た。

- 近代は現代よりも多くの表記バリエーションが存在する。字体の新旧を同一視した場合でも、異体字や別字による異表記が多くあり、現代で全く消失した表記や極めて低い頻度でしか用いられない表記が、近代では高頻度で用いられていた。
- これらの表記は、近代において既に表記の衰退が生じたものと、近代では勢力を保っていたものとがある。近代で表記の衰退が生じたものは、変化の時期より、近代化による実用主義的国語国字政策や、近代口語文の定着に伴う言語使用の合理性重視の気運があったものと考えられるほか、漢語そのものの使用頻度より、多用され近代書き言葉に定着することによって、当時より一般的であった表記が安定して用いられていた可能性がある。
- 一方、近代で現代の標準表記に移行されなかった表記は、その構成字から、1946 年（昭和 21 年）に告示された「当用漢字表」とその周辺の字体表の登場により、代用字・代用表記としての標準表記に移行されたと推測される。

最後に、本稿で扱うことができなかった点を、今後の課題として掲げる。いずれも、本研究が目標とする、近代から現代にかけての漢語の表記バリエーションに関する実態と変化の把握には重要な問題であり、引き続き調査・分析を進めることとする。

a. 表記の位相差

現在利用可能な近代語コーパスは、雑誌という限られたメディアのテキストであるため、本稿では“雑誌における漢語表記の実態”を明らかにしたに過ぎない。語法や語彙に比して自由度が高く、より書き手の志向性が反映されやすい“表記”という事

象を対象とした分析においては、位相の異なるテキスト間での対照が欠かせない。特に、公的な性質の高さなどから統一性を重視する傾向にある教科書や新聞といった媒体と、作家による個性的な表記（当て字や文学的作為の高い表記など）が許容・重視される小説・韻文等の文学的作品とでは、その表記のあり方に大きな差が存在することは明らかである。現在、公的な性質の高い媒体として、教科書（国定読本）のコーパス構築を進めており、今後、新聞・小説等のデータも追補していく計画である。これらとの比較対照により、近代の総合的な漢語表記の実態・変化を明らかにしたい。

b. 同語異語判別の問題

本稿では、「同語異表記」の認定に際して、形態素解析辞書 UniDic の構造に依存した語の同定を行ったため、実際の語の使用実態にそぐわない語の認定がなされているケースが少なからず存在する。語史の観点からは、近代語において同語異表記に相当すると思われるものの、他の時代における同語認定の規定等により「別語」（別語彙素）とされ、本稿で分析対象とならなかった語、例えば、古く「別義による別表記」だったものの、近代以降意味の区別がなくなり「同語の表記揺れ（＝同語異表記）」として用いられたもの（前掲例 (1)(2) における「政事」「政治」）、また、「同語の表記揺れ（＝同語異表記）」から、「別義の同表記」化を経て、現代語では「別語の別表記」となっているもの（例 (3)(4) における「気象」「気性」）などである。これらについては、形態素解析辞書の整備やコーパスの情報整備に対して問題点をフィードバックすることで、適切な情報付けや情報抽出を可能にする枠組みを検討したい。

- (3) あの快活な【気象】は何處に行いたのか、（『太陽』1909 年 1 号、斎藤野の人訳；トルストイ作「流刑者」）
- (4) 明るく快活な【気性】はどこへやら、（『太陽』1925 年 3 号、三上於菟吉「長篇小説 蛇人（第三回）」）

c. コーパス収録範囲外の実態把握

本稿では、1925 年以降に表記が統合されたことが明らかになった語については、当用漢字等の国語施策の影響を想定したが、実際の統合時期を確定し仮説を実証するためには、現在データ不在の期間についての調査・検討が不可欠である。データの追補を検討しつつ、今後のコーパス開発にも期待したい。

付記

本稿は、科学研究費基盤研究 (c)「近代語コーパスの計量分析に基づく同語異表記の通時的研究」（代表：問淵洋子）による成果の一部である。なお、本稿は、計量国語学会第 59 回大会（神戸大学）で口頭発表した内容に加筆・修正したものである。席上で多くの貴重な御意見・御助言を賜った。また本稿の執筆に際して、査読者の方々に非常に有益な御助言を数多くいただいた。ここに記してお礼申し上げる。

文献

- 池上禎造（1984）『漢語研究の構想』岩波書店
- 井之口有一（1982）『明治以降の漢字政策』日本学術振興会

- 小川環樹・西田太一郎ほか編（1994）『新字源 改訂版』角川学芸出版
- 小木曾智信・中村壮範（2011）『『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装 改訂版』（特定領域研究「日本語コーパス」平成 22 年度研究成果報告書）
- 小椋秀樹ほか（2011）「特定領域研究「日本語コーパス」平成 22 年度研究成果報告書『現代日本語書き言葉均衡コーパス』形態論情報規程集第 4 版（下）科研費特定領域研究「日本語コーパス」データ班
- 国立国語研究所（2005）『雑誌『太陽』による確立期現代語の研究—『太陽コーパス』研究論文集—』博文館新社
- 近藤明日子（2014）『『国民之友コーパス』解説書 第 1.1 版』（http://www.ninjal.ac.jp/corpus_center/cmj/doc/kokumin_manual_v1_1.pdf）
- 近藤明日子・小木曾智信ほか（2012）『『明六雑誌コーパス』の開発—近代語コーパスのモデルとして—』『第 2 回コーパス日本語学ワークショップ予稿集』国立国語研究所, pp.329-334
- 今野真二（2008）『消された漱石 明治の日本語の探し方』笠間書院
- 今野真二（2012）『百年前の日本語——書きことばが揺れた時代（岩波新書）』岩波書店
- 今野真二（2014）『日本語の近代——はずされた漢語（ちくま新書）』筑摩書房
- 須永哲矢・近藤明日子（2012）「近代語コーパスのための形態論情報付与規程の整備」田中牧郎ほか『近代語コーパス設計のための文献言語研究 成果報告書（国立国語研究所共同研究報告 12-03）』国立国語研究所, pp.93-117
- 武部良明（1981）『日本語表記法の課題』三省堂
- 田島優（1998）『近代漢字表記語の研究』和泉書院
- 田島優（2009）『漱石と近代日本語』翰林書房
- 田中牧郎（2005）「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」国立国語研究所（2005）
- 田中牧郎（2006）『『近代女性雑誌コーパス』の概要』『日本学術振興会科学研究費補助金研究成果報告書 基盤研究 (B)「20 世紀初期総合雑誌コーパス」の構築による確立期現代語の高精度な記述』, pp.55-62
- 文化庁〔編〕（2006）『国語施策百年史』ぎょうせい
- 坂野潤治（2012）『日本近代史（ちくま新書）』ちくま書房
- 前川喜久雄（2008）「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発（＜特集＞資料研究の現在）」『日本語の研究』4-1, pp.82-95
- 間淵洋子（2013）『『太陽コーパス』における漢語表記の多様性—コーパスの XML タグを利用した研究手法の試み—』（「第 4 回コーパス日本語学ワークショップ予稿集」）
- 宮島達夫（1977）「近代日本語における漢語の位置」鈴木康之編『国語国字問題の理論』麦書房, pp.135-186

（2015 年 11 月 30 日受付・2016 年 4 月 8 日再受付）

Paper A

Diachronic Variation in Orthography of Two-Character Sino-Japanese Words in Modern Japanese: A Corpus-based Study

MABUCHI Yoko (Graduate student at Meiji University,
Research Fellow of Japan Society for the Promotion of Science)

Abstract:

This study reveals the varieties of Sino-Japanese orthography as well as the elimination and consolidation of certain orthographies in contemporary usage of Japanese. Using morphologically annotated corpora of modern and contemporary Japanese, I conducted a survey and analysis of different orthographies of Sino-Japanese words. By examining the number of different orthographies and the usage rates of individual ones, the following three points became clear regarding 2,000 words used in the modern and contemporary eras: 1) There are fewer orthographies used to write a word in the contemporary era than the modern era; 2) Orthographies that disappeared in the contemporary era were used frequently in the modern era. This not only means that the general orthographies used in the modern era were eliminated, but also that the orthographic standards have changed; and 3) The elimination and alteration of orthographies occurred with certain words and not others in the modern era because of an emphasis placed upon practicality and reasonability of language in the wake of modern school system and the rise of typography in publishing. In addition, the eliminations and shifts that occurred after the Showa era were greatly influenced by national language policies.

Keywords: Sino-Japanese words, modern Japanese, variation in orthography, diachronic changes, morphologically annotated corpora