

## 『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—

間淵 洋子 (国立国語研究所 コーパス開発センター) †

### Notational Diversity of Kanji Compounds in the Taiyo Corpus: An Approach Using the XML Tags

MABUCHI, Yoko (Center for Corpus Development, NINJAL)

#### 1. はじめに

現代語ではほぼ統一的に表記されるものの、『太陽コーパス』において複数の表記形で出現するような漢語がある。

- (1) 而して大氣が之れに入るときは腹部は膀大し之れを出づるときは【縮小】す、  
(1895年8号 石川千代松「蝶の話」P141B20)
- (2) 現今支那人が女人の足を【縮少】ならしめんとして鐵骨履を穿たしめて  
(1895年9号 福羽美静「教育に就て」P149A26)

上例「縮小」「縮少」は、いずれも物理的な対象が縮まって小さくなる(対象を縮めて小さくする)意で用いられており、同義の語として機能していると考えられる。この二つの表記で出現する「シュクショウ」という漢語は、現代語においてほぼ統一的に「縮小」と表記されるが<sup>1</sup>、『太陽コーパス』に例を求めると、双方の出現数が殆ど拮抗しており、どちらかをどちらかの誤用・誤字と簡単に言い切ることができない。ここでは、このような関係性にある語を「同義異表記語」<sup>2</sup>と呼ぼう。

このような「同義異表記語」には、「縮小」に対する「縮少」のように同音・同義(類義)字形による表記の他、以下に見られる「喝采」に対する「喝采」のように、異音・異義字形による表記のバリエーションを持つものも少なくない。

- (3) 之を小にしては數年前最も多く江戸ツ子の【喝采】を博した藏原惟郭は  
(1917年6号 浮田和民「総選挙の回顧的批評(立憲政治と群衆心理)」P011A16)
- (4) 其の方に完成せる著書を公衆の前に朗誦して亦異常の【喝采】を博せしが  
(1895年2号 森田思軒「紀元前の著名なる航海者(承前)」P047B22)

『太陽コーパス』では、このような表記のバリエーションを、言語コーパスとしての検索性を考慮した上で、現代語における規範的表記に集約・校訂してテキスト化を行なって

† mabuchi@ninjal.ac.jp

<sup>1</sup> コーパス検索アプリケーション中納言を用いて『現代日本語書き言葉均衡コーパス』における語彙素「縮小」を検索したところ、その原文表記形は「縮小」2002例に対して「縮少」7例(うち1例は国会会議録の用例、3例は同一著者による用例)であった。

<sup>2</sup> 田島(1998)は漢字表記語を「音」「義」「表記」の三つの構成要素により①同音・同義・同表記②同音・同義・異表記③同音・異義・同表記④同音・異義・異表記⑤異音・同義・同表記⑥異音・同義・異表記⑦異音・異義・同表記に分類する。これに倣う。

いる。そのため、上の例(2)は「縮小」に、例(4)は「喝采」にコーパス本文が改められているのだが、その校訂・原文情報は XML タグとして保持されている。

本発表では、この校訂情報を元に、『太陽コーパス』における「同義異表記語」を調査し、その一部について、出現の実態と経年変化を報告する。コーパスを利用した言語研究の方法論として、文字列の情報に加えて、コーパスの XML タグを積極的に用いて情報を抽出する試みの一端を示してみたい。

## 2. 『太陽コーパス』

本発表で調査対象とする『太陽コーパス』は、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作られたものであり、月刊総合雑誌『太陽』（博文館）の明治 28（1895）年、明治 34（1901）年、明治 42（1909）年、大正 6（1917）年、大正 14（1925）年について、著作権処理ができなかった記事を除くほぼ全文を対象にしたものである。

雑誌『太陽』は、分量の多さ、ジャンルの広さ、執筆陣の多彩さ、読者層の厚さなどの点で、当時の文献資料として格別の価値を持っていることが指摘されており（田中 2005）、それを裏付けるように、2005 年の『太陽コーパス』公開以来、口語化、濁点や仮名遣い等の表記法の確立・統一化、新漢語の定着など、当時の語用の実態や現代語に連なる変化の過程を明らかにした論考が多く発表された。

『太陽コーパス』の収録記事が発行された明治～大正期は、現在のように表記の規範意識が高くなく、同語に対し多様な表記が存在していたことが知られている（今野 2012）。例えば、明治前期の文献に見られる同義の異表記関係にあった表記には、「華麗」と「麗華」、「遊戯」と「戲遊」のように字順が入れ替わっているペアが多く存在することが報告されており（田島 1998）、実際に『太陽コーパス』においても同様の字順の入れ替わった漢語対についての報告がある（吉川 2005）。本研究で取り上げる「同義異表記語」も、そのような明治～大正期の表記の多様性を映す現象の一つであり、『太陽コーパス』はこの現象を観察するのに有用な調査対象資料である。

## 3. 同義異表記語

次に、本研究で扱う「同義異表記語」について確認しておきたい。

「同義異表記語」を捉えるにあたって、極めて近接的な関係にある「同音異表記語」について、意味論上の意味の関係と使用される文字の関係に基づき整理すると、以下のよう

- に示すことができる。
- ① 異義語：意義 異義
  - ② 類義語：回答 解答
  - ③ 同義語：
    - A) 異体字：国語 國語
    - B) 表意宛字：本当 真実（ルビに「ほんとう」）
    - C) 別字：縮小 縮少
- ①、②は、文字の異なりが直接意味の異なりに繋がるもので、語の本来の意味から相互に区別して用いられるべき関係にある。これらの語を混用することは表記の規範という観点からは「誤用」とされるべきものだが、この混用が個人的・一時的の使用ではなく、ある程度社会的に認められ一般的に広く、また意図的に使用されるという状況を伴う場合、

これを「通用」と言って差し支えないだろう。『太陽コーパス』においても、このタイプの表記通用は存在するが、現代語においても同様の事象が多く見られ、現代語とは異なる明治期の特徴的な言語実態を捉えるという今回の目的と異なるため、分析対象とはしない。

③-A)は、「字体」の選択に関わるものであり、問題の所在が異なるため、やはり分析対象としない。

③-B)は、ルビを考慮すれば「同音同義」と言うべきものだが、当該漢語が通常の音で読まれる際に想定される、別語としての性質をも内包する表記であるため分析対象から外す。

③-C)は先に前掲例(1)(2)のようなものだが、このうち現代語における規範的な表記に対して、別字による異表記が特異発生的に出現しているとは考えにくいような例について、今回調査・分析対象とするものである。

なお、③-C)の「同義-別字」の異表記には、「同音」ではない文字によるものがあり、前掲例(3)(4)がこれに相当する。これは、本来異義の別字形として文字レベルで相互に区別して使用されるべきものであり、その混用は「誤字」と判断されるものだが、表記バリエーション発生のメカニズムや、当時の雑誌編集・組版の態度、活字の受容等を勘案する時、「同音-同義-別字」の異表記と連続的な事象として見えてくる。詳しく分析することは別の機会に譲るが、今回の調査手法によって副次的にデータを得ることができるため、その出現実態について同時に報告する。

#### 4. 調査概要

##### 4. 1 調査対象コーパス

調査には、現在構築中の『太陽コーパス』増補改訂版データを用いた。これは、2005年の『太陽コーパス』公開時に著作権処理の関係で収録対象とならなかった記事から保護期間終了のものを追加収録したもので、今後形態論情報を付与した上で再データ化される予定のものである。今回の分析では、より多くの用例を収集する目的で、データ量の多い構築中データを用いた。

##### 4. 2 調査対象語

本研究で対象とする「同義異表記」の語を抽出するために、『太陽コーパス』のXMLタグを用いた。『太陽コーパス』は、XML形式により様々な情報が付加されているが、その中に、雑誌『太陽』の原文を校訂しテキスト化したことを示す要素「注」がある。この「注」要素は、言語研究用コーパスとしての検索性を考慮し設けられた要素で、現代語の規範と異なるような語用・表記に対して、以下の校訂種別を「分類」属性として示した上で、原文の情報（「原文」属性）と共に情報付けを行なっている（田中 2005）。分類属性に示される校訂の種別は表1の通りである。

このうち、「A 誤字通用」については、挙げられた例に、今回分析の対象とすべき「同義異表記語」が含まれる（下線引用者）。

<注 原文="近頃" 分類="A 誤字通字">近頃</注>ノースロツブ氏より直接概畧の話を  
(1895年7号「樹栽日に就て」牧野伸頓 P150A08)

随分御<注 原文="氣嫌" 分類="A 誤字通字">機嫌</注>宜しく

(1895年8号「妄語戒即ち真語律に就て」渡辺龍聖 P163B27)

そこで、今回はXML文書から、直接このタグの情報を利用して「同義異表記語」の抽出を試みた。具体的な抽出手法を次節に示す。

表1 『太陽コーパス』XML「注」要素「分類」属性

分類	意味
A 誤字通字	誤植, 誤用, 漢語の漢字表記における通用。誤植・誤用のうち, 以下のB~Eについては別の分類を立てるので, ここには含めない。
B 衍字	なくてもよい文字が紛れ込んだと思われるもの。
C 脱字	あるべき文字が脱落していると考えられるもの。
D 転倒	二つの文字の配列順が, 転倒していると考えられるもの。
E 欠損	活字不良などのため, 文字が欠けたり, つぶれたりしているもの。
F 濁点脱落	濁音表記が期待される箇所に濁音表記がないもの。
G 仮名遣	仮名遣いが歴史的仮名遣いの規範と異なる使われ方をしているもの。
H 正誤表	次号等の『太陽』誌上で, 誤りを告知して訂正している場合は, その告知にしたがって, 該当号の該当部分を修正した。

#### 4. 3 調査手順

##### 《手順1》

XML 文書から XPath を用いて, 「注」要素「分類」属性の値として「A 誤字通用」を持つものを全て抽出した。XPath は, XML 文書内で特定のノードの位置を指定することを目的とした構文である。コンピュータにおけるファイル・システムパスと同様に, スラッシュ区切りで階層を明示した式 (ロケーション・パス) を使ってノードを参照することができ, 比較的容易に XML 文書内の特定の要素や属性を抽出することができる。今回は, XPath 式「//注[@分類="A 誤字通用"]」を用いて情報抽出を行なった<sup>3</sup>。この際, 「注」要素「原文」属性 (校訂前の雑誌原文を格納するための属性) の情報を同時に取得した。

これによって抽出されるのは, 今回調査対象として扱う「同義異表記語」の他, 「。」－「、」, 「ぬ」－「ね」といった類形の文字間の誤植, 濁点の打ち誤り, 「ある」－「あり」のような活用形の誤り等, 多種多様な誤用 (通用) である。本研究で対象とするのは, 一時的な誤用とは言いがたく, 規範表記と等価の表記として用いられる漢語表記バリエーションであるため, 出現頻度を手掛かりに一時的な誤用を排除することとし, 出現度数 10 以上の語を選定した。更に, 固有名詞を除く 2 字以上の漢語を今回の調査対象語 (規範表記と通用表記の対) と定めた。

なお, 分析にあたって, 通用表記・規範表記相互の関係性を見る際に, 規範表記は『太陽コーパス』の校訂本文によったが, 以下の 2 語に対しては独自に規範表記を当てた。

(ア) 「シュドウ」: 通用表記(原文)「主働」/コーパス校訂本文「主導」→規範表記「主動」

『太陽コーパス』原文においては, 「主導」が 1895 年の号にいずれも「主導者」の形で 4 例現れるが, その後は「シュドウ者」の接続も含めて「主動」の表記で 32 例が出現する。

(5) 彼は、實に討幕派の【主導】者にして、

(1895年3月 落合直文「しら雪物語(承前)」 P099B14)

(6) 彼等猶世界歴史の【主動】者はアリアン人種にありといふを得る乎。

(1895年9号 田岡嶺雲「十三世紀に於ける蒙古民族の雄図」 P073B03)

(7) 我帝國が最大有力なる【主働】者の地歩を占めざるべからざるや、

(1895年12号 川崎三郎「東邦革新」 P018A01)

<sup>3</sup> 作業環境: Windows8, Cygwin(1.7.17), perl(5.14.2)+XPath モジュール

このように「主導」と「主動」が混用されているのであれば、字形の近さと頻度から「主動」の規範表記を「主動」とするのが妥当と判断した。

(イ)「ケンジツ」: 通用表記(原文)「**健實**」/コーパス校訂本文「**賢實**」→規範表記「**堅實**」

「**賢實**」の表記は、『太陽コーパス』原文に1例も出現せず、辞書の表記としても見られない。「**健實**」の意に隣接する「**堅實**」を規範表記とするのが妥当と判断した。

(8) わが國の職人にも以上の如き【**健實**】なる美風があつたのである

(1925年2号 藤原銀次郎「労働問題の解決策として工場の官僚化を排す」P034C12)

(9) 保守的にして【**堅實**】なる氣風を有する

(1917年1号 安井正太郎「列強海軍と下級艦」P112A16)

以上によって定めた具体的な調査対象語は、表2に示す88語である。

表2 調査対象「同義異表記語」リスト

アイソウ(愛想/愛相), イコウ(意向/意嚮), イチズ(一途/一圖), オウタイ(應對/應待), オクビョウ(臆病/憶病), カイチョク(戒飭/戒飾), カシヤク(呵責/苛責), カッサイ(喝采/喝采), カンケイ(關係/干係), カンケイ(關係/關繫), カンタン(簡單/簡短), カンニン(堪忍/勘忍), カンネン(觀念/感念), カンパ(看破/觀破), カンマン(緩慢/緩漫), キオク(記憶/記臆), キガイ(氣概/氣慨), キセツ(季節/期節), キソウ(皮相/皮想), キネン(記念/紀念), キュウシュウ(吸收/吸集), キョウショウ(狹小/狹少), クフウ(工夫/工風), ゲキレツ(激烈/劇烈), ケネン(懸念/掛念), ゲンイン(原因/源因), ケンキュウ(研究/研窮), ケンジツ(堅實/健實), コウカ(効[效]果/功果), コウジョウ(向上/昂上), コウセキ(功績/効[效]績), コウテツ(更迭/交迭), コウテン(公轉/行轉), ゴウモン(拷問/拷問), コジン(個人/箇人), サイツツ(採掘/採掘), ザンシン(斬新/嶄新), ジコ(自己/自個), シハイ(支配/司配), シマツ(始末/仕末), シャカイ(社會/社界), シュウカク(收穫/收獲), シュウキ(周期/週期), シュクショウ(縮小/縮少), シュッパン(出版/出板), シュドウ(主動/主動), ショウバイ(商賣/商買), ショクタク(囑託/囑托), ショチ(處置/所置), シンコク(深刻/深酷), シンセツ(親切/信切), セイコウ(成功/成效[效]), セイセキ(成績/成績[跡]), セイチュウ(掣肘/制肘), センエツ(僭越/潛越), センコウ(銓衡/詮衡), ゼンゼン(全然/全々), センレン(洗練/洗鍊), ソウゴン(莊嚴/壯嚴), ソウホウ(双方/相方), ソガイ(阻害/沮害), ソッセン(率先/卒先), タイテイ(大抵/大低), タイハイ(頽廢/頽敗), タイヘイ(太平/太平), テイケツ(締結/訂結), ドウサ(動作/働作), トウタ(淘汰/淘汰), ドウダン(道斷/同斷), ハクセキ(白晳/白晳), ヒエキ(裨益/裨益), ヒジュン(批准/批准), ヒッキョウ(畢竟/必竟), フウサイ(風采/風采), フクザツ(複雜/復雜), フクシャ(複寫/復寫), フシギ(不思議/不思議), ヘキエキ(辟易/避易), ボウガイ(妨害/妨害), ボウシ(防止/妨止), ホウドウ(報道/報導), ホントウ(本當/本統), マスイ(麻醉/魔睡), ムゾウサ(無造作/無雜作), ヨソウ(予想/預想), ヨロン(輿論/輿論), ラチ(拉致/羅致), リュウギ(流儀/流義)
---

## 《手順2》

次に、ここで得られた調査対象語リストにおける規範表記を、コーパス本文全体から抽出した。この中には、実際には対になる原文の通用表記（やその他の異表記）が「注」タグの付いた形で含まれるから、先に《手順1》において抽出した注タグ付き出現例を除いて規範表記出現例とした。

## 5. 調査結果と分析

### 5.1 表記バリエーションの分類

抽出した同義異表記語（表記対）は、規範表記と通用表記の関係性について、以下の二つの観点で分類を行なった。

A) 語義...辞書の記述に基づく語の意味関係 (『日本国語大辞典第二版』小学館<sup>4</sup>による)

**同義**: 下記「異義」「類義」に当てはまらないもの。①規範表記により表記される辞書項目において、通用表記が見出し語漢字欄、辞書欄 (主要な古辞書, 明治期辞書類での記載を示す欄), 用例内に現れているもの。②別項目であっても語釈の記述が【「○○」に同じ。】のみのもの。③通用表記が辞書に全く現れないもの。

例: 「記念」と「紀念」

き-ねん【記念・紀念】

**類義**: 通用表記と規範表記が異なる辞書項目で、両者に意味の重なりがあるもの (どちらかがより限定された意味のみを持つ場合も含む)。

例: 「観念」と「感念」

かん-ねん [クワン:] 【観念】 [名] (3)ある (抽象的な) 物事に対する考え、意識。

かん-ねん 【感念】 [名] 物事についての感じ方、考え方。

**異義**: 通用表記と規範表記が異なる辞書項目で、両者に意味の重なりがないもの。

例: 「妨害」と「防害」

ぼう-がい [パウ:] 【妨害・妨碍・妨礙】 [名] さまたげること。邪魔すること。ぼうげ。

ぼう-がい [パウ:] 【防害】 [名] 害するものから身を守ること。

## B) 字義...表記対における交代字同士の意味関係 (白川静『字通』平凡社による)

**同義**: 異体字関係にある文字 (今回は対象としない)

**類義**: 部首が交代している文字, 字義に重なりや隣接が見られる文字

例: 「更迭/交迭」における「更」と「交」

「更」①かえる、あらためる。④こもごも、いれかわる。

「交」②たがいに、こもごも、かわるがわる、入りみだれる。③とりかわす、とりかえる。

**異義**: 上記「同義」「類義」に当てはまらない文字

例: 「應對/應待」における「対」「待」

「対」①うつ、あげる、土をうつ。②むかう、あたる、あう、あいて、つい。③こたえる、あわせる、たぐえる、むかえる。

「待」①まつ、まちうける。②ふせぐ、そなえる、もてなす、あてにする。

語義と字義, 双方の関係によって同義異表記語を分類した結果を表3に示す (表内下線は辞書に一切掲載のない表記)。

表3から, 同義異表記の多くは, 漢字列のうちの一文字を類義同音の別字 (多くは部首の置換や追加・削除による極めて近接的な別字) に置き換えることにより生じていることが分かる<sup>5</sup>。屋名池(2004)に「「読みさえ定まれば (読めさえすれば)、どのような書き方をしてもよい」 (表記と言語は多対一に対応する) という「一意表記」の原則がおこなわれていたものらしく、実際、この原則に違反するものはごく少数にとどまる。」と指摘される通りの実態が映しだされている。また、笹原 (2005)は, 字体に生じる「同化」「衝突」<sup>6</sup>と

<sup>4</sup> Web上で利用できる JapanKnowledge 版 (<http://www.jkn21.com/stdsearch/displaymain>) を用いた。字義確認に用いた白川静『字通』も同様。

<sup>5</sup> ここでは規範表記→通用表記という発生順や起源関係を意図しない。

<sup>6</sup> 同化とは字体が近接する他の字体の干渉を受けて同じ構成要素を持つように変化する現象 (例: 「模糊」... 「模」が「糊」の米偏に干渉を受け「模糊」と表記される)。干渉を受けた結果, 既に存在する他の字体と一致するケースを衝突, さらに音義が一致するケースを暗号と呼ぶ (笹原 2005)。

いう現象が『太陽コーパス』に少なからず見られること、またその中で、部首が置換された文字に同化しそれが継続的に使用される「定着」状態にあるものが散見されることを報告している。ここから垣間見えるのは、雑誌『太陽』の著者・編集者・読者にとって、文字の一部の構成要素が置換された字体を、置換前の字体と同義のものとして認識・受容する言語環境があったということである。このような環境下、音の一致によって語としての同一性が確保されれば、なおさら異表記の共存状態は容易に成り立ち得たものと思われる。

表3 同義異表記語の語義・字義による分類

語義\字義	類義	異義
同義	愛相, 意嚮, 憶病, 苛責, 戒飾, 掛念, 關繫, 勘忍, 簡短, 觀破, 緩漫, 記臆, 氣慨, 紀念, 狹少, 劇烈, 研窮, 源因, 簡人, 交迭, 功果, 効績, 拷問, 採掘, 嶄新, 自個, 收獲, 週期, 縮少, 出板, 主働, 所置, 囑托, 深酷, 制肘, 成効, 成蹟, 洗鍊, 潛越, 詮衡, 壯嚴, 沮害, 相方, 卒先, 大低, 大平, 訂結, 淘汰, 働作, 批準, 皮想, 避易, 稗益, 必竟, 不思議, 復寫, 復雜, 報導, 妨止, 魔睡, 預想, 流義 (62)	【同音】 應待, 工風, 昂上, 行轉, 仕末, 司配, 社界, 商買, 信切, 全々, 頽敗, 無雜作, 輿論 (13) 【異音】 喝采, 白晳, 風采 (3)
類義	干係, 感念, 期節, 吸集, 健實 (5)	本統, 羅致 (2)
異義		一圖, 同斷, 防害 (3)

一方、異義の文字での置き換えによって生じる異表記も少なからず見られる。これらの内訳をみると、語義からの類推によって異義の文字を当てたと思われるもの(例10)が多く、また隣接する別語からの干渉による混淆形と思われるもの(例11)、その他、極めて字形の近い異音の別字による置き換え(「采/采」「皙/皙」)によるものがある。

(10) 「向上」→語義から上昇のイメージ喚起→「昂上」

(11) 「應對」→隣接する「接待」等の語による干渉→「應待」

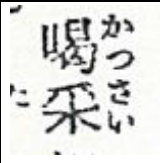

		幕のあとで観客の喝采(図1)に應へる時にはベルナルは立つてはゐるが、左手を軽く他の役者の肩において、それに凭りかかつてゐた。四たび五つたびと幕をあげさせる満場の喝采(図2)は光榮ある彼女の過去の歴史に對する敬意を示すもので、 (1917年6月号 厨川白村「老女優サラ・ベルナル」P167A25-28)
図1	図2	

図1, 図2は、同一記事の同頁同段の3行の範囲に規範表記「喝采」と通用表記「喝采」が出現している例である。これだけ隣接した箇所同語の表記に別字「采」「采」が用いられていることから、これらは別字として強く意識されていない可能性が考えられる。

さて、表3を見ると、今回調査対象として取り出した語には、語義において類義の語(同音類義語)、異義の語(同音異義語)が含まれていることが分かる。

このうち「語義」を「類義」としたのものには、(a)完全に規範表記の語義に内包されるもの(「干係」「吸集」)、(b)規範表記の語義と隣接するもしくは規範表記の語義と一部が重なるもの(「感念」「期節」「健実」)、(c)規範表記の語義と重なる部分と規範表記の語義には

含まれない異義の両方を持つもの(「本統」「羅致」)があり、(a)は「同義」としても差し支えない。(b)は、通用表記が現代語で殆ど用いられないことがないという点で校訂の対象となったもので<sup>7</sup>、本来は類義語の混用と同様の事象であり、厳密には今回取り上げる「同義異表記」とは性質を異にする。(c)も(b)と同様通用表記が現代語で殆ど用いられないという点で校訂の対象となったものだが、弁別される語義において通用表記が用いられている場合は、むしろこれは通用ではなく正用であって、仮に現代語において当該の表記が使用されないとしても、現在では消滅してしまった語の出現と見なすべき事象である。

例えば「ラチ」では、辞書記述を見ると、「羅致」が語義の一部に「拉致」の語義を完全に内包しており、より広い意味で用いられる語であることが分かる。

ら-ち【拉致】〔名〕無理に連れて行くこと。捕えて連れて行くこと。

ら-ち【羅致】〔名〕(1)鳥を網で捕えるように、広く人材を集めること。招きよせること。(2)「らち(拉致)」に同じ。

(12) 海内の學者を闕下に【羅致】したるを以て、濟々たる多士は翰林に充滿せり、

(1895年6号 中西牛郎「清朝全盛の時代」P030A13)

(13) 天心の衣鉢を襲ふの大觀も亦新たに結社するに方つて己れを虚うして逸才を【羅致】した。

(1917年13号 内田魯庵「案頭三尺」P045A10)

一方、2例のみ出現する規範表記「拉致」についてその用例を確認すると、以下のとおり「羅致」(1)の意で用いられていると思われることから、『太陽コーパス』における「ラチ」は専ら「羅致」(1)の意で用いられ、「拉致」の表記形がむしろ通用している状況と見ることが出来る。

(14) 一進會を作るや李容九一派の天道教徒を巧みに【拉致】して全國を風靡したる怪手腕を以て、

(190906号 浅田江村「政治、外交 統監政治の失敗」P028A07)

また、同様に「語義」を「異義」としたもの(「一圖」「同断」「防害」)についても、本来は同音異義語の誤用・混用の現象であり、「同義異表記」とは性格が異なる。これらの現象は、今回の調査において全容を把握することはできないが、調査対象になった一部の語については、異表記通用の実態を報告することを目的として、この先の分析対象から省くことはしない。

## 5. 2 出現率からみた通用表記

次に、それぞれの表記の出現実態について見ていく。表4は、同義異表記語の出現度数に占める通用表記形の割合(=通用表記率)を示したものである。

通用表記率は、極端に高いもの低いものなどさまざまではあるが、今回調査対象とした通用表記が出現度数10以上の語の場合、通用表記率は比較的高く、表記の通用という言葉運用が広く行われ定着していたことを伺わせる。通用表記の構成要素や辞書への記載の有無による分布の偏りはあまり見られないが、異音異義の文字による通用表記は総じて通用表記率が高い点が特徴的である。これは、先に示した通り字形認識において異なりがさほど意識されていないことに起因するものと考えられる。また、極めて通用表記率の高い「ラチ」は、その用例を精査すると、専ら前掲例(7)(8)に見た「羅致」(1)の意で用いられていることが分かる。この場合、「羅致」は「拉致」が本来持たない語義で用いられているため、「羅致」の表記で現れているものであり、通用表記独自の原義に起因するものである。

<sup>7</sup>『太陽コーパス』では『広辞苑』等中型国語辞典を目安に校訂対象を定めている(田中2005)。



表4 同義異表記語の通用表記率(使用率の高い順。下線は別義字, 網掛けは異音字による異表記)

通用表記率*	辞書記述あり	辞書記述なし
~100%	預想, <u>羅致</u> (2)	<u>行轉</u> , <u>詮衡</u> , <u>喝采</u> , <u>潜越</u> (4)
~68%	苛責, 魔睡, 紀念, 緩漫, 氣慨, 縮少, <u>無雜</u> <u>作</u> , <u>一圖</u> , 流義, <u>頹敗</u> , 洗鍊, <u>同斷</u> , 記臆, 主 働 (14)	復寫, <u>風采</u> , 拷問, <u>白晝</u> , 避易, 戒飾, 勘 忍, 週期, 囑托, 壯嚴, 制肘 (11)
~26%	健實, 嶄新, 成績, 交迭, 意嚮, <u>本統</u> , 狹少, 掛念, 觀破, 收獲, 期節, <u>工風</u> , 愛相, 劇烈, 憶病, 出板, 商賈, 効績 (18)	卒先, <u>應待</u> , 稗益, 深酷, 採掘, 淘汰, 沮 害, 働作, <u>仕末</u> , 報導, 皮想, <u>信切</u> , 所置, 批准, 訂結, 功果 (16)
~7%	成效, 必竟, 感念, 簡短, <u>防害</u> , 箇人, 吸集, 源因, 大平, 不思儀, 研窮, 干係, 關繫 (13)	相方, 妨止, <u>昂上</u> , 復雜, <u>輿論</u> , <u>全々</u> , <u>司</u> <u>配</u> , 大低, 自個, <u>社界</u> (10)

\* 分布に開きのあるところで適宜区分して示した。

### 5. 3 経年変化

次に, 通用表記の出現頻度が高い上位5語「紀念/記念」「成績/成績」「記臆/記憶」「成效/成功」「本統/本當」について, それぞれの表記の出現状況が年を追ってどのように変化するかを確認してみたい。図3に通用表記率の推移を示す。

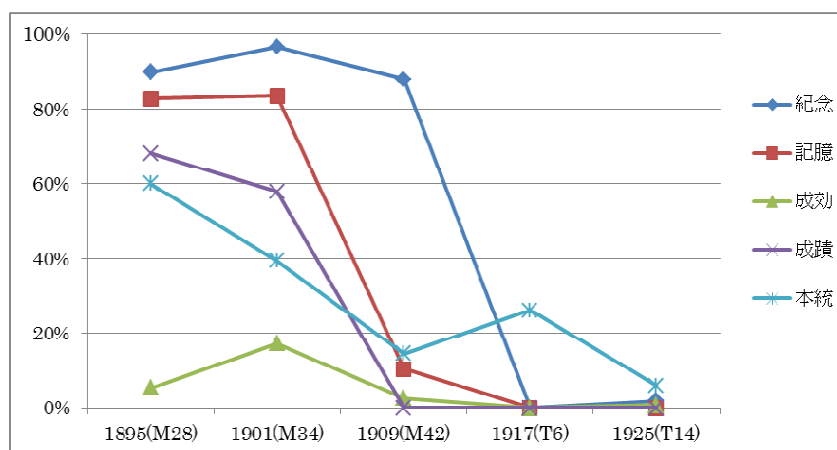


図3 通用表記率の推移 (出現度数上位5位まで)

図3より, 「紀念」を除く語で1901(明治34)年から1909(明治42)年で通用表記率が劇的に減少していることが分かる。1917(大正7年)には唯一勢力を保っていた「紀念」も例が見られず, 1925(大正14)年には1917年に若干勢力を戻した「本統」を含め, 全ての語において, 通用表記が規範表記に駆逐されている様子が見て取れる。

この背景には, 明治期の国語施策が関与しているものと思われる。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, 次第に規範表記への統一化が実現されてきたことの現れと見ることができる。

## 6. まとめ

本発表では、『太陽コーパス』の XML タグを直接用いて、同義異表記語の抽出を試み、以下の分析・考察を行なった。

- ・ 同義異表記語の語・字構成要素による分類と、異表記発生・受容の背景の考察
- ・ 出現度数による異表記通用状況の実態把握
- ・ 通用表記率の経年変化と国語施策との関連性の指摘

今回の調査・分析は、コーパスの研究用付加情報を生かした研究手法のトライアルケースとして行なったものであり、明治期における漢語表記の多様性についてほんの一端を捉えたものにすぎない。今回の手法の問題点や限界、あるいは時間的制約によって課題として残されたものの幾つかを以下に示す。

- ・ 著者による使用実態・個人差の把握

表記問題は、個人に帰結する部分も大きく、著者ごとの使用実態把握が欠かせない。今回の調査では、著者別の集計を行なったものの、時間的・能力的制約によって分析に至らなかった。今後の課題としたい。

- ・ 『太陽コーパス』の校訂態度（「A 誤字通用」の対象選定基準）に依拠したことにより抽出できなかった同義異表記語の把握（例：「障碍」「障害」「障礙」<sup>8</sup>）。
- ・ 通用表記の議論として保留した、類義語・異義語の誤用・通用の実態把握（例：「異義」と「異議」<sup>9</sup>）。

これらについては、別途、辞書や同音異表記のリストなどを用いて、調査すべき語を選定する必要がある。いずれも今後の課題としたい。

## 参考文献

- 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
- 武部良明(1981)『日本語表記法の課題』三省堂
- 国立国語研究所(2005a)『太陽コーパス—雑誌『太陽』日本語データベース—』(CD-ROM) 博文館新社
- 国立国語研究所(2005b)『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
- 笹原宏之(2005)「漢字文字列における字体の同化と衝突」国立国語研究所(2005b), pp.293-312
- 田島優(1998)『近代漢字表記語の研究』和泉書院
- 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(2005b), pp.1-48
- 文化庁[編](2006)『国語施策百年史』ぎょうせい
- 吉川明日香(2005)「字順の相反する2字漢語—「掠奪—奪掠」「現出—出現」について—」国立国語研究所(2005b), pp.143-156
- 屋名池誠(2004)「明治語の表記」『日本語学』23-12, pp.65-72

<sup>8</sup> 『太陽コーパス』では校訂対象となっていない。

<sup>9</sup> 同音異義語とされる表記対だが、中納言による BCCWJ の検索結果では、「イギ申し立て」「イギあり/なし」等「異議」が期待される接続において「異義」が見られる（異義 40 例／異議 1574 例）。これらは語彙素がそれぞれ「異義」「異議」となっているが、同一語の異表記として扱うべきものと思われる。