

国立国語研究所学術情報リポジトリ

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

メタデータ	言語: Japanese 出版者: 公開日: 2016-06-15 キーワード (Ja): キーワード (En): 作成者: 国立国語研究所, The National Institute for Japanese Language メールアドレス: 所属:
URL	https://doi.org/10.15084/00001355

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—

国立国語研究所編



博文館新刊

刊行のことば

本書は、同時に刊行した国立国語研究所編『太陽コーパス—雑誌『太陽』日本語データベース—』（CD-ROM）について、その設計と活用に関わる研究論文を集成したものである。『太陽コーパス』は、国立国語研究所が構築を目指している日本語コーパス（大規模言語データベース）の一環として、現代語の書き言葉が確立する20世紀初期にもっともよく読まれた総合雑誌『太陽』（博文館刊）を対象として作成したものである。

本書に収めた論文は、いずれも『太陽コーパス』の作成に参画した研究者の手になるものであり、11回にわたって開催した「『太陽』研究会」における討議の成果をふまえて執筆されたものである。本書の編集は、主として田中牧郎（研究開発部門第一領域主任研究員）が担当し、吉田谷幸宏（研究開発部門第一領域研究補佐員）がこれを助けた。また、各論文の執筆者、研究会への参加者をはじめ、多くの研究員、非常勤研究員、所外協力者などの尽力があった。

本書が、日本語に関心をもつ広い範囲に読まれることを期待するものである。

平成17年3月

独立行政法人国立国語研究所長
甲斐 睦朗

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——

目次

刊行のことば	国立国語研究所長 甲斐睦朗
研究の目的と本書の構成	田中牧郎 i

第1部 設計

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	田中牧郎 1
構造化テキストに対応した全文検索システム『ひまわり』	山口昌也 49
構造化テキストを直接利用するアプリケーション—『ブリズム』と『たんぼぼ』—	小木曾智信 83

第2部 活用

I 語彙

漢語「優秀」の定着と語彙形成—主体を表す語の分析を通して—	田中牧郎 115
字順の相反する二字漢語—「掠奪—奪掠」「現出—出現」について—	吉川明日香 143
外国地名表記について—漢字表記からカタカナ表記へ—	井手順子 157
逆接の接続詞・接続語句	馬場俊臣 173

II 文法

「そして」の用法について—用例に基づく類型の分類と分析—	島田泰子 193
副詞「とても」について—陳述副詞から程度副詞への変遷—	中尾比早子 213
尊敬待遇表現—動作性の名詞や動詞連用形に付く形式について—	近藤明日子 227
漢語サ変動詞の可能的形—「～できる」の展開—	小木曾智信 251

III 文字・表記

漢字の実態と処理の方法	田中牧郎 271
漢字文字列における字体の同化と衝突	笹原宏之 293
異体仮名について	中川美和 313
濁点文字使用率から見る濁音表記	近藤明日子 331
仮名遣いについて	小木曾智信 351

英文題目	377
あとがき	379
執筆者一覧	383
『太陽コーパス』に含めなかった記事の一覧	384

研究の目的と本書の構成

——田中 牧郎

1

待望される日本語コーパス言語学

言語学では、1960年代に始まった「コーパス言語学」(corpus linguistics)と呼ばれる領域が、この20年ほどの間のコンピューター技術の普及にともなって飛躍的に進展してきた。コーパス言語学とは、一定の方針で大量に集められ、コンピューターで管理された生の言葉の資料集(コーパス)を用いて、さまざまな言語事象につき、実証的な研究を行う領域である。英語研究が導いてきた領域であるが、現在では各個別言語の研究においても、コーパスを用いた研究が隆盛に向かっている。日本におけるコーパス言語学も、英語研究から進展を見せ、10年ほど前から日本語の研究においても試みが始まるようになった(注1)。

コーパスというとき、広義に電子化資料全般を指すこともあるが、狭義には、言語の実態を代表するように一定の方針でデータの質と量を設計した大量の電子化資料を指す。英語を対象としたコーパス言語学が、数多くの狭義の良質なコーパスに基盤を置いて多方面に展開してきているのに対して、日本語においては狭義のコーパスはまだ少なく、日本語研究としての展開は不十分である。日本語の研究にコーパスを利用する場合、現状では、新聞社や出版社、電子図書館などが提供する電子資料を活用することが一般的である。しかし、それらはデータの質や量、付与される情報を考慮して設計されたものではなく、言語研究資料としては偏りの大きいものである。コーパス言語学で重視されている、バランスのとれた言語資料による記述研究に直接活用できるようなコーパスを構築して研究を進めることが、日本語においても求められる。

また、英語研究の分野ではコーパスを用いて言語研究を行うための便利なソフトウェアが数多く開発され、高度なコンピューター技術をもたない言語研究者でもコーパスを活用できる環境が整えられ、研究の裾野を広げている。日本語コーパスに関しても、

日本語の研究に適したソフトウェアの開発と普及が望まれる。

以上のように、日本語を対象としたコーパス言語学の本格的な展開が望まれているが、その実現のためには、良質の（狭義の）コーパスを作成すること、コーパスを利用するソフトウェアを開発すること、そしてそれらを用いて実践的な研究を多角的に進めることが待望されている。

2

国立国語研究所による現代語の実態調査

2.1 創立以来の実態調査

国立国語研究所は、創立以来の任務のひとつとして、現代日本語の実態を客観的にとらえることのできる資料を整備し、その資料に基づいた調査研究を行ってきた。この任務に基づく研究の成果として、さまざまな媒体の日本語の実態を調査し、データ集や研究報告書を公開してきたが、そこには次の三つの柱があった（注2）。

○計量的な調査研究

『現代雑誌九十種の用語用字』（1962-1964）

『高校教科書の語彙調査』（1983-1984）

『テレビ放送の語彙調査』（1995-1999）など

○体系的な調査研究

『話し言葉の文型』（1960-1963）

『分類語彙表』（1964）、『同一増補改訂版一』（2004）

『動詞の意味用法の記述的研究』（1972）など

○歴史的な調査研究

『明治初期新聞の用語』（1959）

『雑誌用語の変遷』（1987）など

これらは、それぞれの調査目的に応じて、現代日本語の実態を記述するために必要な資料を選定し、記述方法を学術的見地から開拓しつつ研究を進めてきたものである。調査研究の成果が国語施策の基礎資料として役立てられるとともに、日本語研究の進展にも貢献してきた。こうした国立国語研究所の調査研究の基本的な役割は、今後も不変であると考えられる。

2.2 コーパス構築の事業へ

国立国語研究所は、上に見た基本的役割に加えて、新しい活動

も展開してきた。1980年代から本格的に着手した事業に、国語辞典編集のための用例採集事業がある。これは、「日本大語誌」と呼ばれる、大規模国語辞典編集のための用例採集を進める構想に基づくもので（注3）、最初の成果物として『国定読本用語総覧』全12巻（1985～1997）、同CD-ROM版（1997）を刊行した。この事業は文脈付き語彙用例集の編集刊行という形で成果を蓄積することから始まったが、コンピューター技術の急速な発展にともない、コンピューターを用いた手法を順次取り入れていった。手法の再検討の中で、用例集の編集よりもコーパスの構築を通して構想を実現することが、コストの削減と成果の拡充を期待できると考え、用例採集事業からコーパス構築事業に移行した（注4）。

この移行の時期と前後して、科学技術振興調整費による事業として、理工系の研究機関と連携した大規模音声コーパス作成に着手した。これは、音声認識や自動要約などへの応用を直接の目的とするコーパス作成事業であったが、完成した『日本語話し言葉コーパス』は、言語学的に高度な分析を行うことも可能な形に作成された（注5）。こうして、用例採集事業からの移行と音声コーパス作成との二つの流れにより、国立国語研究所にコーパス構築事業という新しい柱が作られることになった。

コーパス構築の事業は、国立国語研究所の従来の調査研究と比較して、対象にする資料の範囲が膨大であり、報告書を成果物とするという性質の調査研究ではなく、コーパスを作成すること自体が目的とされ、完成したコーパスは汎用的な研究目的に用いることが想定されている点に特色がある。この汎用性は、国立国語研究所の基本的任務である現代日本語の計量的・体系的・歴史的な研究をも包含しうるものである。国立国語研究所によるコーパスの構築は、従来の実態調査型の調査研究の流れと合流して、より太い流れを作っていくことが見込まれる。

3

汎用コーパス構築のために

3.1 求められる研究の段階

現代日本語の研究において多目的に使うことのできる汎用コーパスの具体像はまだ明らかではない。考えられるさまざまな可能性を検討しつつ、本格的なコーパスを構築する準備を進めていく

必要がある。タイプの異なるコーパスをいくつか作りつつ、コーパスの設計と活用についての研究を進めていくことが現実的であろう。その際、次の3点について研究することが重要だと考えられる。

○コーパスの設計（対象）

現代日本語を代表できる資料を、どのように選び、どの程度の量をコーパス化すればよいか。

○コーパスの設計（方法）

汎用的な目的で活用できる形にするためには、どのような形式を用意し、どのような情報を付与すればよいか。

○コーパスの活用

コーパスを利用して研究することで、日本語研究としてどのような知見がもたらされ、どのような方向に研究が進展できると期待されるか。

○コーパスの設計（対象）

まず、現代日本語を代表できる資料については、なるべく多様な資料を扱いながら、現代日本語の多様性をとらえるための資料の類型を見出していくことが必要であろう。一度に多くの資料を扱うことは現実には難しいので、限られた範囲で多様性をとらえることができる資料をまずは対象にし、多様性をとらえるための視点や方法を研究することが望まれる。こうした観点からこれまでの国立国語研究所の調査研究をふりかえると、ひとつの媒体から多様性をとらえるのにもっとも適していると考えられる媒体として、雑誌があげられよう（注6）。

○コーパスの設計（方法）

次に、形式や付加情報に関しては、その種類からコーパスを分類すると、次の四段階が考えられる。段階が進むほど複雑で高度なコーパスということになる。

（1）プレーンテキストコーパス

テキストのみからなり、付加情報を何ももたないものの

（2）構造化テキストタグ付きコーパス

文章や文のレベルまで構造化し、ジャンル、文体、著者、引用箇所、話者などの情報をタグ付けしたものの

（3）形態素解析タグ付きコーパス

単語や形態素のレベルまで構造化し、見出し語や品詞などの情報をタグ付けしたもの

(4) 構文解析タグ付きコーパス

文における単語や形態素の係り受けまで構造化し、構文機能に関する情報をタグ付けしたもの

高度な段階である(3)(4)についても、自然言語処理の研究分野ではすでに作られているコーパスがあるが、日本語研究への活用は十分でない。その理由はいろいろ考えられるが、対象とする資料についての吟味が不十分であること、形態素や構文の認定基準や付与されている情報が、言語研究者にとって必ずしも満足できるものではないことなども想定される。言語研究を目的としてコーパスを利用する場合、現状では、(1)を対象にして、研究者が自らの研究目的に応じて、データを加工して活用している傾向が強い。しかし、この段階にとどまる限り、研究者間の言語情報の共有化は望みにくく、コーパスによる日本語研究の普及も難しいのではないと思われる。研究の活性化のためには、(2)(3)(4)の段階への展開が望まれるが、そのためには段階を踏んだ実践的な研究を重ねていくことが不可欠であろう。一気に高い水準を目指すのではなく、(3)(4)への展開は次の研究段階に送り、まずは(2)の段階のコーパスとしてある程度完成度の高いものを確実に作成し、それを用いて具体的な研究を行うことを優先させようと考えた(注7)。

コーパスの分類の観点を資料の配し方に移せば、次のような分類ができる。

・共時バランストコーパス

ある一つの時代について、その言語を代表する言語層のバランスをとって資料を配したもの

・通時コーパス

ある言語の歴史的な変化をとらえることができるように資料を配したもの

現代日本語の実態を反映した汎用コーパスとしては、現代という時代の共時バランストコーパスがまず必要であろう。そして、現代日本語の諸問題を考えるためには、それを歴史的所産として見る視点も不可欠であり、現代からさかのぼれる形での通時コーパスも整備されることが望まれよう。しかし、いずれのコーパスも日本語を代表させる規模をそろえるにはきわめて大がかりな作業を要する。まずは、原型となるコーパスを作ることで、資料の選び方や配し方について研究を行うことが求められよう。このよう

な考え方から、共時バランストコーパスと通時コーパスの両方に
通じる問題点が整理でき、できればそれらの原型となるようなコー
パスを作って研究することを目指すことにした。

○コーパスの活用

コーパスを利用することによって開拓できる新しい研究領域
は、英語の例を見れば、言語の記述研究のあらゆる領域について
精度を上げ見通しをよくすること、辞書編集・言語教育など応用
言語学的な展開など、広範囲に及んでおり、日本語についても、
同様な展開が期待できよう。ただ、さまざまに活用可能であるとい
っても、研究の目的が広すぎると問題点が拡散して、まとまり
のよい確実なコーパスが作りにくいという問題も出てくる。汎用
性は意識しつつも具体的な目的のもとにコーパスを活用し、研究
を進めるなかで実り多い議論が可能な方向をとるのが得策だと考
えられる。

3.2 『太陽コーパス』の設計と活用

3.1に述べたような汎用コーパス構築に向かうために現在求め
られている研究段階を実現すべく取り組んだものが『太陽コーパ
ス』である。『太陽コーパス』とは、1895（明治28）年から
1928（昭和3）年まで博文館から刊行された月刊の総合雑誌
『太陽』を資料とするコーパスである。『太陽』を電子化する企画
は、2.2に述べた「日本大語誌」構想において、『国定読本用語
総覧』の次の成果物を作成していくなかで生まれたものである。
この作業は、国語辞典編集のための用例採集として、雑誌『太陽』
から任意で採択する方式（スカウト式）で集められた用例に対し
て文脈をつける目的で、『太陽』の本文を入力するところから始
まった。ところが、先に述べたような経緯から国語辞典編集のた
めの用例採集事業はコーパス構築の事業に移行し、『太陽』の入
力作業は『太陽コーパス』作成の作業に衣替えをし、事業の位置
づけも軌道修正が図られたのである。

汎用コーパスを目指す最初の段階で作るコーパスの対象として
雑誌『太陽』が適切だと考えられる理由には、次のようなことが
あげられる。

- ・『太陽』は、著者、ジャンル、文体といった文章の種類か
らみて広い範囲をカバーしており、文章に対する構造化と
タグ付けを行う構造化テキストタグ付きコーパスの対象事
例として適切である。

- ・『太陽』が刊行されていた19世紀末期～20世紀初期は現代日本語の書き言葉が確立する時代である。この時代をあつかうことで、現代日本語の広がりや確立への過程とを知ることができる。また、今後他の資料や時代をも対象に加えて、現代語のバランストコーパス、現代語につながる通時コーパスを作成する足がかりとなる研究対象として適切である。
- ・『太陽』は単体の資料でありながら、多様な言語の層を反映しており、現代語の確立期は言語の変化も激しい。コーパスを用いて多様な日本語の姿を記述する対象として適切である。

このような考え方から、『太陽コーパス』は、次のようなコーパスを目指して設計することにした。

- ・構造化テキストタグ付きコーパスの雛型となるコーパス。
- ・共時バランストコーパス・通時コーパスのいずれの方向にも発展可能な、原型となるコーパス。
- ・コーパスを活用した日本語の記述研究を多角的に試みることができるコーパス。

『太陽』を対象としたコーパスを作ることを通して、確立期現代語の記述研究を行うことを直接の目的とするが、その先には、日本語コーパスの作成とそれを用いた研究についての一般的な問題を展望していくことを目指したわけである。

4

本書の構成

4.1 全体の構成

本書は、研究論文16編を中心とする。その執筆者はすべて『太陽コーパス』の作成に関わった研究者で、2000年度～2003年度にかけて11回にわたって開催した「『太陽』研究会」における討議をふまえた成果を編集した。『太陽』研究会では、言語資料としての雑誌『太陽』を多角的に分析するとともに、コーパス化の作業で生じた諸問題を解決する方法を検討し、『太陽コーパス』を作成しながらそれを用いてさまざまな角度から探索的な研究を行った。基本的には、参加した研究者の問題意識に基づく研究成果を持ち寄る形で進めたが、コーパスによらない従来型の研究とは違う、新しい視点や方法を得ることを目指して討議を重ね

た。その成果の一部はすでに他の媒体に発表したものもあるが(注8)、本書におさめた論文は、そこでの討議を経て参加者の問題意識を熟成させ、現段階で到達できた水準で書かれたものである。

4.2 「第1部 設計」について

本書は、『太陽コーパス』の設計に関わる論文を掲載した第1部と、『太陽コーパス』を活用した確立期現代語の記述研究の論文を掲載した第2部との二部構成とした。

第1部には3編の論文を掲載した。「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」(田中牧郎)は、雑誌『太陽』をコーパス化することの意義と方法を論じ、『太陽コーパス』の仕様を総合的に記述したものである。『太陽コーパス』の全体像を理解するために一読してほしいものであり、『太陽コーパス』を使いながら疑問に思うところが出てきたら、その都度参照してほしい論文である。文献資料をもとにコーパスを作成する際の一般的な問題点を考える際に参考になる事例を多く含んでいると思う。

「構造化テキストに対応した全文検索システム『ひまわり』」(山口昌也)、「構造化テキストを直接利用するアプリケーション―『プリズム』と『たんぽぽ』―」(小木曾智信)の2編は、『太陽コーパス』を利用するために開発されたソフトウェアの仕様と利用法を記述したものである。利用マニュアルとして参照すべきことはもちろん、コーパス活用のための本格的ソフトウェアの解説論文になっているので、日本語コーパス言語学のためのソフトウェア研究の基本論文としても参照されるべきものである。どのソフトウェアも『太陽コーパス』以外のコーパスに適用する方向も考えられている。

なお、第1部の3編をもとに、要点を平易にまとめ直した解説を、小冊子『『太陽コーパス』解説書』として、『太陽コーパス』CD-ROMに添付した。コーパスの操作方法などについて簡潔な解説が必要な場合は、その解説書を参照してほしい。

4.3 「第2部 活用」について

4.3.1 「I 語彙」の論文

第2部は、『太陽コーパス』を用いた記述研究の論文を、「I 語彙」「II 文法」「III 文字・表記」の順で、扱う言語現象の分野別に並べた。「I 語彙」には、4編の論文を掲げた。「漢語『優秀』

の定着と語彙形成—主体を表す語の分析を通して—」（田中牧郎）は、漢語「優秀」の定着が、和語「すぐれる」との間に緊密な対語関係を形成していく過程とともに進んだことを記述したものである。コーパスから得られる用例から帰納的に語の意味を分析する際に、共起する語との統語的な関係に着眼する方法の有効性を示すものにもなっている。「字順の相反する二字漢語—「掠奪—奪掠」「現出—出現」について—」（吉川明日香）は、明治期に顕著に見られる字順が反対になる二字漢語が併存する現象について、時代を経て一方のみが残存する場合と、時代を経ても両方が共存していく場合とを比較したものである。一方のみが残存する場合は同義であり、共存していく場合は意味による使い分けがあることを明らかにしている。用例の分析において、「格成分」として現れる語句を検討する方法をとっており、意味分析の方法として、統語的な整理が有効であることを示している点は田中論文と共通している。この二つの論文は、コーパスの豊富な用例からどのようにして語の意味研究に向かうのかについて、有効な方法のひとつを示したものであることができる。

「外国地名表記について—漢字表記からカタカナ表記へ—」（井手順子）は、外国地名の表記が漢字表記からカタカナ表記に移行する目立った現象を全体的に記述し、移行の要因を探る論文である。著者の生年と表記選択との相関が強いこと、1925年になると、それ以前であれば漢字表記を選択した世代も、カタカナ表記を選択する傾向があることを明らかにし、世代と時代とが交錯する言語変化の具体相を浮かび上がらせている。「逆接の接続詞・接続語句」（馬場俊臣）は、先行研究などから接続詞・接続語句を網羅的に採取し、それらが『太陽コーパス』にどれぐらいの頻度で、どの年次・文体に出現しているかについて、詳細に調査したものである。接続詞・接続語句を例として、『太陽コーパス』の語彙の広がりや経年的な語彙変化の程度を具体的に示している。井手論文と馬場論文は、あるまとまった語彙を網羅的に収集し、用例の出方を徹底的に調査する方法をとることによって、現代語の確立過程における語彙の大きな変容を実証的に明らかにしたものである。コーパスを活用することで、言語の変容を広範囲にわたってとらえることができるようになることがわかる。

4.3.2 「Ⅱ文法」の論文

「Ⅱ文法」にも、4編の論文を掲げた。「『そして』の用法について—用例に基づく類型の分類と分析—」（島田泰子）は、接続

詞「そして」によってつながられる語句の品詞性を指標に用法の分類を行うことによって、先行研究の分類では未解決であった「そして」の文法機能の問題に、解決を与えることができることを見通している。「副詞『とても』について―陳述副詞から程度副詞への変遷―」（中尾比早子）も、先行研究で混乱していた「とても」の機能について、かかり先の語の品詞性を指標として類型化することを通して、陳述副詞から程度副詞への変遷の過程を記述したものである。島田論文と中尾論文は、語の文法機能の記述において、コーパスにおける大量の用例に対して有効な指標をあてがって帰納する方法をとることで、明晰な記述の方向が見えてくることを示唆している。

「尊敬待遇表現―動作性の名詞や動詞連用形に付く形式について―」（近藤明日子）は、「御～なさる」から「御～になる」への推移を中心とした、尊敬待遇の表現形式9種の変化の様相を、「～」の部分に入る語、使われる文章の種類、話し手の性別という、三つの視点から詳細に記述したものである。「特化係数」という数値的な指標を用いることで、各形式の微細な差異までもくつきりと浮かび上がらせており、文法形式の推移が、位相や表現価値による使い分けと連動しながら進んでいくさまが描かれている。「漢語サ変動詞の可能の形―『～できる』の展開―」（小木曾智信）は、漢語サ変動詞に「できる」を付けて可能を表す諸形式の形成と展開について考察し、「～できる」がもつとも基本的な形式となっていく潮流を発見している。用いられる文章の性質や上接するサ変動詞の文体的特徴など、文体との相関についても明快に記述されており、一見して雑然としている用例も視点を定めて整理していくことで、言語変化の確かな流れが見えてくることかわかる。近藤論文と小木曾論文は、文法形式の推移を、位相や文体など推移の要因と関連づけることで豊かな記述ができることを示した形になっている。

4.3.3 「Ⅲ文字・表記」の論文

「Ⅲ文字・表記」には、漢字に関わる2編と、仮名に関わる3編の論文を掲げた。まず、「漢字の実態と処理の方法」（田中牧郎）は、『太陽コーパス』の設計における漢字処理の方法について考察したものである。用いられる漢字の種類と字体のゆれ幅が非常に大きい『太陽』の漢字の現象を、電子化において生じる問題を軸に整理し、包摂・代用の処理や字体変換辞書の整備などの方策によって、コーパスとして扱いやすい電子テキストを作成する方

法を提案している。「漢字文字列における字体の同化と衝突」(笹原宏之)は、ある字体が文中で近くにある別の字体に影響を与えて字体を変えてしまう「同化」の現象と、発生した字体がそれまでに存在していた別の字体とたまたま一致してしまう「衝突」の現象が、『太陽』の活字字体にも広く見られることを詳しく報告している。これは、『太陽』において、著者の手書き原稿がそのまま活字化された場合があったことを示すものであり、そうした漢字を電子化する際は、校訂対象とするのが現実的であることを述べている。田中論文・笹原論文は、『太陽』における漢字や異体字の実態を記述する側面と、異体字の多い文献資料を電子化する場合の一般的な方法論について考察した側面とをあわせもつものである。

「異体仮名について」(中川美和)は、『太陽』に残存している異体仮名の実態を詳しく記述したものである。活字印刷が一般化していくのとともに、異体仮名は衰退し一つの仮名に収斂していくように見えるが、なかには機能的な書き分けが行われているものなどもあり、異体仮名の衰退過程には、活字印刷の普及だけには帰せられない、日本語を表記する文字の歴史として考察すべき問題があることを示している。『太陽コーパス』には異体仮名の情報は収録できなかったが、『太陽』以前の資料を文字資料としても使える形に電子化する際などには、異体仮名の処理の方法は検討を要する課題となることを教えている。

「濁点文字使用率から見る濁音表記」(近藤明日子)は、『太陽コーパス』に付けられた校訂注記に濁点脱落の情報があるものを網羅的に分析し、年次別、文体別、記事別、著者別に実態を整理している。濁点が表記される表記法は、口語文の普及とともに整備されていくものであり、濁点表記法の定着は口語文の定着と密接に関わる問題であることを明らかにしている。「仮名遣いについて」(小木曾智信)は、校訂注記に仮名遣いの情報があるものを網羅的に分析し、経年的推移にも語別の現象としても複雑な様相にあるさまを丁寧に記述している。仮名遣いの現象の推移は、異体仮名や濁点表記のように原理的な説明を簡潔に行うことは困難であり、語彙や語法による個別の事情が深く関わっている状況が、多面的に示されている。近藤論文と小木曾論文は、コーパスに付けられた校訂注記を用いてコンピューターによる一括処理でデータを得ることで、人手による研究では展望しにくい全体的な傾向を見出すことに成功している。タグ付きコーパスを作成して研究することの成果をわかりやすく示したのものにもなっている。

4で、本書に収めた論文一つ一つの内容を簡単に紹介し、コーパスを用いた日本語研究としての一般的な観点からみたときに、それぞれの論文のもつ意義についても言及した。それぞれ、雑誌『太陽』という資料からコーパスを設計するための研究、また『太陽コーパス』を使って確立期現代語における特定の言語現象を記述するための研究であり、その目的を実現したこと自体で十分意義のある論文である。一方、コーパスによる日本語研究を進展させるための一段階を示すことを目指した本書の観点からは、コーパス言語学の方法によってひらけてきそうな日本語研究の新しい展開について、16編の論文はいくつかの方向性を示している。具体的個別的な問題に新しい知見を加えつつ、一般的な方法論としても問題意識が高められる方向で、本書が読まれることを期待したい。

『太陽コーパス』の価値を十分に生かし切るには、本書の論文で扱ったテーマは、まだまだ部分的なものに止まり、今後さまざまな角度から活用が進められるべきである。また、コーパスを用いた日本語研究として想定できそうな課題のうち、『太陽コーパス』のできる範囲は限定的でもある。しかし、コーパスを作り、それを用いて研究することが、言語の記述研究に広がりや深みをもたせていく可能性をはらんでいることは、具体的に示せたのではないかと思う。『太陽コーパス』がきっかけとなって、日本語コーパス言語学が活性化することを期待したい。

注

- (1) コーパス言語学の現状に関してはMcEnery (2001)が参考になる。日本で書かれたものでは、英語研究では齊藤・中村・赤野 (2005)、日本語研究では宮地・甲斐監修 (2003) などが、扱う範囲が広く参照価値が高い。
- (2) 国立国語研究所の調査研究は他に社会言語学や言語教育など多様に展開してきたが、書かれたり話されたりした生の媒体を対象にした実態調査型の調査研究は、この三つの柱にまとめられる。
- (3) 国語辞典編集のための用例採集事業の概要は、木村・加

- 藤・田中(1999)にまとめられている。用例採集の対象資料の目録や採集方法など、この事業についての詳細な情報は、国立国語研究所国語辞典編集準備室・同国語辞典編集室(1980-1995)。
- (4) 国語辞典編集のための用例採集事業からコーパス構築事業への移行については、木村・加藤・田中(1999)に言及がある。
- (5) 『日本語話し言葉コーパス』については、前川(2004)に概説がある。
- (6) 国立国語研究所(1953)(1957・1958)(1962・1964)(1987)(2002)のように、雑誌を対象にした実態調査は繰り返し行われ、雑誌資料の有効性は確認されている。
- (7) 話し言葉を対象とした『日本語話し言葉コーパス』は(3)の形態素解析タグ付きコーパスとして本格的なものであり、部分的には(4)の構文解析タグ付きコーパスにあたる情報も付与されており、書き言葉のコーパスを設計する際にも参考になるものである。しかし、多様な書き言葉を対象とした本格的なコーパスは、(2)の構造化テキストタグ付きコーパスの段階もまだ未整備である。まずは(2)の段階を踏まえてから(3)(4)の段階に進むべきだと考えた。
- (8) 本書以前の『太陽コーパス』を用いた研究成果については、本書「あとがき」を参照。

参考文献

- 木村睦子・加藤安彦・田中牧郎(1999)「国語辞典編集のための用例データベース」(国立国語研究所編『日本語科学』5,109-128頁,国書刊行会)
- 国立国語研究所(1953)『婦人雑誌の用語』(国立国語研究所報告4,秀英出版)
- 国立国語研究所(1957・1958)『総合雑誌の用語 前編・後編』(国立国語研究所報告12・13,秀英出版)
- 国立国語研究所(1962・1964)『現代雑誌九十種の用語用字 第1〜3分冊』(国立国語研究所報告21・22・25,秀英出版)
- 国立国語研究所(1987)『雑誌用語の変遷』(国立国語研究所報告89,秀英出版)
- 国立国語研究所(1985-1997)『国定読本用語総覧1-12』(三省堂)

-
- 国立国語研究所 (1997) 『国定読本用語総覧CD-ROM版』 (三省堂)
- 国立国語研究所 (2002) 『現代雑誌の漢字調査』 (国立国語研究所報告119, 国立国語研究所)
- 国立国語研究所国語辞典編集準備室・同国語辞典編集室 (1980-1995) 『国語辞典編集準備資料』 1-11 (国立国語研究所)
- 齊藤俊雄・中村純作・赤野一郎 (2005) 『改訂新版 英語コーパス言語学—基礎と実践—』 (研究社)
- 前川喜久雄 (2004) 「『日本語話し言葉コーパス』の概要」 (国立国語研究所編『日本語科学』 15, 111-133頁, 国書刊行会)
- 宮地裕・甲斐睦朗監修 (2003) 『日本語学 臨時増刊号 (22巻5号) コーパス言語学』 (明治書院)
- McEnery, T. & A. Wilson (2001) *Corpus Linguistics* 2nd Edition, Edinburgh University Press.

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

——田中牧郎

1

はじめに

本格的な日本語コーパス構築と日本語コーパス言語学の進展を目指した最初の研究段階として、『太陽コーパス』の作成と研究を進めた。『太陽コーパス』は、次のようなコーパスとして設計することにした（本書「研究の目的と本書の構成」（田中牧郎）vii頁参照）。

- 構造化テキストタグ付きコーパスの雛型となるコーパス。
- 共時バランストコーパス・通時コーパスのいずれの方向にも発展可能な、原型となるコーパス。
- コーパスを活用した日本語の記述研究を多角的に試みることができるコーパス。

『太陽コーパス』の対象とした雑誌『太陽』は、こうしたコーパスを作るのに適した資料だと考えられる。本論文は、雑誌『太陽』が言語資料としてどのような位置にあり、どのような特徴をもっているのか、そうした『太陽』の特徴を生かしてどのようなコーパスを設計したのかについて総合的に述べるものである（注1）。

2

現代語確立期の資料としての雑誌『太陽』

2.1 現代語確立期の資料としての総合雑誌

現代日本語の書き言葉は、社会の近代化とともに言う言語の変革とともに形成され、近代化の完成とともにほぼ確立した。その形成と確立が、言語現象としてもっとも目立った形に現れたのは、漢語を中心とする新しい語彙の創造と定着、言文一致による口語文の創成と普及、の二つのできごとであったと見られる。新語の増大がもっとも顕著であり、言文一致運動が最盛期に達したのは、

19世紀後半（明治前期）である。そして、新しい語彙が定着し、口語文が普及し、語彙と文体が安定に向かうのは20世紀初期（明治後期から大正時代）である。この、書き言葉が安定に向かう20世紀初期を、現代語の確立期と見ることができる。

現代語確立期の書き言葉の実態を調査研究するために活用することが期待される資料は多岐にわたるが、その全体像を把握することは容易でない（注2）。多種多様な資料を一度に扱うことは現実には困難であるので、範囲を限った扱いやすい資料でありながら、それなりに十分な規模と多様性をもち、当時の日本語を代表できるものがあれば、そうした資料を優先的に研究対象としていく工夫が望まれよう。そのような条件を満たす可能性をもつ資料として、新聞や雑誌があげられるが（注3）、単体での内容のまとまり、分量の多さ、多様性という点では、新聞よりも雑誌の方がまさっている。とりわけ、総合雑誌と呼ばれるメディアは、単一の資料でありながら、当時の日本語をかなりの程度まで代表させることができる資料として、高い価値をもっているのではないかと考えられる。

社会の近代化にともない、19世紀後期に雑誌というメディアが誕生するが、当初は『明六雑誌』（1874年創刊）、『東京経済雑誌』（1879年創刊）などの啓蒙雑誌が中心であった。以後、ジャンルや読者層ごとに、さまざまな雑誌が創刊されていくが、1880年代後半には、広範なジャンルをおさめ読者層を拡大させた、『国民之友』（1886年創刊）、『日本人』（1887年創刊）などの総合雑誌が誕生した。総合雑誌の登場によって、当時の日本人の読む生活における雑誌メディアの影響力は飛躍的に大きくなったと考えられる。こうした時代の流れをとらえて、文字通り画期的な総合雑誌として1895年に創刊されたのが、博文館の月刊誌『太陽』であった。

2.2 言語資料としての『太陽』の位置

『太陽』は、記事の分量、ジャンルの広さ、執筆陣や読者層の厚さの点で、画期的な総合雑誌であった。『太陽』創刊号の、四六倍版200余ページという体裁は、先行の総合雑誌『国民之友』の四六版約50ページという体裁に比較して、数倍以上の分量を誇る。博文館は、1894年までに刊行していた『日本商業雑誌』『日本農業雑誌』『日本大家論集』『日本之法律』『婦女雑誌』を統合して『太陽』を創刊し、それら前誌群がカバーしていたジャン

ル、執筆陣、読者層を、『太陽』に継承し拡大させた。分量の多さと、ジャンル、著者、読者の観点から見た多様性という点で、この時期の日本語を代表する資料として、研究対象にすべき優先度は非常に高い資料であると言えよう（注4）。

ジャンルについて、『太陽』創刊年（1895年）12冊における欄（記事を配列するカテゴリー）の名称を列挙すると次の通りである。

論説、講演、史伝、地理、小説、雑録、文苑、芸苑、家庭、政治、法律、軍事、文学、科学、美術、教育、宗教、医事、商業、農業、工業、社会、海外思想、与論一斑、社交案内、新刊案内、実業案内、海外彙報、海内彙報

学術や産業の各分野を網羅し、啓蒙的かつ実用的なジャンルの文章が広く収められていることが分かる。鈴木貞美（2001）は、こうした『太陽』の性格を、「国民のための知識と趣味をひとつの器に盛るという意味での「総合雑誌」だったのである」とまとめている。

次に、著者については、『太陽』創刊号の大橋新太郎「太陽の発刊」の記事中に、

今『太陽』の期する處は普く専門諸大家の力を集め、廣く中外諸人に紹介して以て相互の智見を交換せしめんとするに在り。是我が『太陽』が當代第一流の諸名家にのみ執筆寄稿の勞を請ひ、成るべく平易に成るべく趣味多からしめんと力むる所以なりとす。

とあるように、当時一流の各界の専門家が名を連ねている。創刊号の著者を列挙すると、次の通りで（括弧内は専門分野など、著者の属性）、当時を代表するさまざまな分野から著者が選ばれている。

大橋新太郎（博文館創立者）、久米邦武（歴史学者）、千頭清臣（貴族院議員）、井上哲次郎（哲学者）、坪内逍遙（小説家）、三宅雪嶺（ジャーナリスト）、上田万年（国語学者）、坪井正五郎（人類学者）、井上辰九郎（日本興業銀行理事）、横井時敬（農学者）、尾崎行雄（政治家）、中西牛郎（宗教思想家）、森田思軒（ジャーナリスト）、戸川残花（評論家）、福地桜痴（小説家）、中川四明（俳人）、渡辺千吉郎（未詳）、尾崎紅葉（小説家）、饒庭篁村（小説家）、幸田露伴（小説家）、志賀重昂（地理学者）、石橋忍月（文芸評論家）、飯田武郷（国学者）、大和田建樹（国文学者）、羽南外史（未詳）、捻華主人（未詳）、

佐々木指月（彫刻家）、幸堂得知（小説家）、三島通良（医学士）、寒沢振作（未詳）、大橋乙羽（小説家）、大隈重信（政治家）、品川弥二郎（未詳）、板垣退助（政治家）、市村塘（理学者）、前田香雪（鑑識家）、矢部規矩治（日本醸友会会長）

さらに、読者層については、永嶺（1997）に詳しい研究があり、「中学生から壮年層にわたる全国的な中産層読者を獲得していた」とし、『太陽』の意義を、「和漢洋・古今東西の多様な文化的要素を融合し、それを全国的な配付網を通じて地方の末端にまで普及させ、国民的文化へと練り上げていく溶融炉としての側面にあった」とまとめている。

現代語の確立期における目立った言語現象のうち、口語文の普及に関しては『太陽』を用いた調査報告が出されている（注5）。1897～1907年の『太陽』について口語文の比率を調査した見坊（1957）によると、当初は文語文が大部分であったのが、年次を追うごとに次第に口語文が増えていき、文語から口語への変化の過程をとらえることができるという。また、「研究の目的と本書の構成」（田中牧郎、iii頁）で述べた、国立国語研究所の「日本大語誌」構想を進める際の用例採集対象文献を選定する作業において、有識者10人の評定委員全委員が推薦した雑誌のひとつが『太陽』であった（国立国語研究所国語辞典編集準備室1983）。

以上のように、雑誌『太陽』は、20世紀初期の現代語確立期を代表する資料のひとつであり、コーパスの対象とする価値の高い資料だと言える。

3

雑誌『太陽』の本文の様態

3.1 『太陽』本文の特徴

言語資料を組織化してコーパスにまとめるためには、資料の特質を引き出しやすい形に本文等を構造化し、必要な情報をあらかじめタグ付けしたデータベース等の形式で整備することが望まれる。そのためには、文献資料の本文をどのように構造化し、原資料のもつ言語情報をすくい上げることのできるタグ付けをどのように行うかをよく検討しなければならず、原資料の特徴をよく把握しておく必要がある（注6）。『太陽』の本文は、現代の総合雑

誌の本文とはかなり異なる特徴をもっており、これについては、土屋（1966）（1967）に、句読法・補助符号などについて報告がある。土屋に言及がある特徴は、それを踏まえ、それ以外の現象についても、広く見わたしていくと、『太陽』の本文の様態として、次の8点を特徴として指摘することができる。

- (1) 引用表示法が多様
- (2) 句読法が多様
- (3) 振り仮名が豊富
- (4) 漢字字体が多様
- (5) 誤植・誤用が多い
- (6) 濁音無表記例が多い
- (7) 仮名遣い規範が未整備
- (8) 特殊な表記法が多様

以下、この8点について、具体的な特徴を概観し、構造化とタグ付けにあたって考慮すべき点を整理する。

3.2 引用表示法

文章中に、他の典拠から文章を引用したり、小説等で会話部分を引用したりする場合、改行・字下げ・鉤括弧あるいは助詞「と」等で、引用部分を明示する方式が、『太陽』では確立していない。『太陽』でよく見かける引用表示法をあげると、次の通りである。下線部が引用文、傍点部が引用表示形式である。例文の後の括弧内には、年号、記事名、著者名、頁番号（Pの後の3桁の数字）、段記号（1段組の場合はA、2段組の場合はAB、3段組の場合はABCという具合）、行番号を示す。

- (1) 或るものは雨霽れて後ち出立すべしと言ひしも（1895年1号「利根水源探検紀行」渡辺千吉郎P197A21）
- (2) 又ベルナルも馳せ來りて曰く。敵既に我が右翼に接したりと（1895年1号「ヲートルロー合戦の記」戸川残花P062B06）
- (3) ゲーテの所謂、一國語のみを知るものは、畢竟其國語を知らざるものなりといふのは、此事であります。（1895年1号「国語研究に就て」上田万年P029B08）
- (4) 案内者として連れる山賊の友人云ふ様、之れより先きは彼等の住所なれば吾等を普通旅客と見違ひ、何時長銃を向けるやらも知れざれば、其節は驚くことなく靜かに馬車を留む可しとのことにて、（1901年2号「特別通信

欧米奇聞」鈴木東馬P211B15)

- (5) 古書にて二三を例すれば(源平盛衰記) 鷲尾三郎が一谷にて義經を案内する條「みなくれなゐに日出だしたる扇を以て鷲尾に賜ひ云云(1895年1号「日章旗」 旭日生P114B17)
- (6) 和尚は晝寢でも仕て居たのであらう青膨れのした氣倦るさうな顔をして「庫から出すのが大ソウでなどきも大儀ソウに無機嫌氣にあつた。(1925年12号「蕪村寺」 橋本閑雪 P038A14)
- (7) 支考が辻談義説に昔祖翁の此ものゝ名を稱して、露の一字には新古の別あらんと可笑がり給へり」と見え、(1895年2号「昔の落語家」 二橋生P137A19)
- (8) 「こりや奈何も厄介だねえ。」 観音丸の船員は累々しき盲翁の手を執りて、舳より本船に扶乗する時、恁は呟きぬ。(1895年1号「取舵」 尾崎紅葉P083B11)

(1) のような「と言う」の形式は、現代の標準的な引用法であるが、『太陽』では、この例のように鉤括弧を付けない方式も一般的である。(2)「曰く〜と」、(3)「所謂〜と」、(4)「云ふ様〜と」、(5)「云云」などは、古来の文語の形式であり、『太陽』にはこれらが普通に見られる。また、鉤括弧を用いる場合、(8)のように始点・終点が呼応する形も創刊年から見られるものの、(5)(6)(7)のように、始点と終点とが呼応しないことも多い。全体的に見れば年次を追うごとに現代語の書き言葉の引用表示法と同じものに近付いていくが、後の時代になっても(6)の例に見るように、現代語では普通は用いない引用法が残っている。

本文の構造化として、引用された部分を他の部分から区別しておくことが望まれるが、現代語と異なる多様な引用表示法をもつ『太陽』の場合、特定の形式に着眼して機械的に引用部分を抽出することは困難である。引用部分にあらかじめタグ付けを行っておく必要性は高いと思われる。

3.3 句読法

『太陽』の句読法のありようについては、土屋(1966)に、やや詳しい言及があり、『太陽』創刊号(1895年1号)には次の4種類があることが報告されている。

- (1) 句点「。」と読点「、」を用いるもの
(2) 読点「、」のみのもの

(3) 句点「。」のみのもの

(4) 読点「、」と無表記のもの

そして、『太陽』全体を見ると(3)(4)は早くに消滅し、(2)は長く続いたが、やがて(1)に移っていくと述べられている。『太陽』創刊号をさらに詳しく観察すると、句読法のありようはもっと複雑であり、句点と読点は「、」「。」で示し、段落の切れ目を「。」で示すものや、句読点を一切用いないものなどがある。

このように、『太陽』において、「、」「。」は句・文・段落など、さまざまな文章の区切りを示す記号として、いくつかの使われ方があったと見られる。段落の切れ目に「、」「。」が用いられることがあるのは、段落冒頭の一字下げの表示方法が確立していないこととも関わっている。こうした「、」「。」の用い方は、記事や年次によって、そのありようを変えるだけでなく、同一の記事のなかでもゆれている場合もある。現代語では自明である文章や文の切れ目の位置が明示的でない『太陽』の本文には、現代語の場合とは異なった構造化の方法が求められる。

3.4 振り仮名

『太陽』の時代は、振り仮名(ルビ)が活発に使用されていた時代である。創刊年(1895年)『太陽』の記事は、(1)総ルビ(ほとんどすべての漢字に振り仮名があるもの)、(2)パラルビ(一部の漢字に振り仮名があるもの)、(3)無ルビ(特別な場合を除き振り仮名がないもの)の三種にほぼ分かれる。そして、小説は総ルビ、論説や政治・文学・教育・宗教・商業・工業・農業・彙報などの記事では無ルビ、史伝・地理や家庭の記事ではパラルビという傾向が顕著である。1910年頃までの『太陽』は、このように記事の属する欄によって、振り仮名のありようがほぼ決まる。ところが、1910年代から後は、新刊紹介や彙報などごく少数の記事が無ルビのほかは、ほとんどすべての記事で総ルビとなっている。

欄によって、あるいは年によって振り仮名のありようが変わることから、振り仮名を付けるか否かは著者の判断だけでなく編集部の判断も大きかったのではないかと推測される。パラルビの場合、読みにくい漢字列、あるいは誤読のおそれがある漢字列に、振り仮名が付されているかといえ、必ずしもそうではない。パラルビ記事の一例として、創刊号(1895年1号)の「紀元前の著名なる航海者」(森田思軒)の冒頭から、ルビ付きの語とルビ無

しの語を抜き書きすると、次のようになる。

ルビ付きの語 ^{せいやうしか}西洋史家 ^よ據る ^{ざいせき}載籍
^{とく}特に ^{つた}傳ふ ^{ふる}古き ^{かうかいしや}航海者

ルビ無しの語 考る 所 余輩 其の 名 最も

ルビ無しの語は、振り仮名がなくとも読めるものが多いと見られるが、ルビ付きの語の多くも、容易に読むことのできる語である。読みにくい語、読み誤られる可能性の高い語に振り仮名が付けられているわけではない。このように、振り仮名には言語情報として必ずしも有用性の高くない場合も含まれているのである。コーパスに含める電子テキスト作成にあたり振り仮名をどの程度生かすのかは、検討を要するところである。

3.5 漢字字体

『太陽』では、漢字字体のゆれが大きく、安定していない。また、現在では普通用いられない漢字も多く見られる。

- | | | |
|-----------|-----------|-----------|
| (1) 羽 [羽] | (2) 青 [青] | (3) 鷗 [鷗] |
| (4) 沿 [沿] | (5) 膈 [膈] | (6) 捷 [捷] |
| (7) 煮 [煮] | (8) 憑 [憑] | (9) 杓 [杓] |
| (10) 筭 | | |

(1) から (9) は、それぞれ、そのままの字体をコンピューターで表示させることはできない。現在もっとも一般的なコンピューターで表示される文字の中で、もっとも近いものを引き当てると、[] 内の文字がそれにあたると見られる。また、(10) はそれに近い文字をコンピューターで扱える漢字のなかに見出すことができないものである。

電子テキストで採用する日本語文字集合には、汎用性のあるものとして、JISとユニコードとが候補になろう。上の10種の文字について、JIS X0208 1997 (第1・第2水準)、JISX0213 2000・2004 (第3・第4水準)、ユニコード (CJK統合漢字) での扱いをまとめると、表1のようになる。

表1 JISとユニコードの字体の扱い例

	JIS X0208	JIS X0213	ユニコード
(1)	包摂	包摂	包摂
(2)	包摂	包摂	別字
(3)	包摂	別字	別字
(4)	外字	包摂	包摂
(5)	外字	外字	包摂
(6)	外字	外字	別字
(7)	外字	外字	外字
(8)	外字	外字	外字
(9)	外字	外字	外字
(10)	外字	外字	外字

ユニコードのCJK統合漢字については、例えば(1)は、「羽」「羽」両方を用意しているが、「羽」は台湾・韓国向けに用意されているもので、日本語のコンピューターでは、通常「羽」しか出ない。これに対して(2)は、日本語用にも「青」と「青」とは別字としてともに使えるようになっている。表1における「ユニコード」の欄は、日本語用のコンピューターでユニコードが使えるものが通常登載している文字セットに関してまとめたものである。

『太陽』のように使用文字の種類が多く、そのゆれも大きい資料を電子テキスト化する場合、用いる文字集合について、次の二つの要件が満たされることが望まれる。

- ・収録文字が豊富であること
- ・字体のゆれ幅の認定規準（包摂規準）が規定してあること

JISとユニコードの収録文字数と包摂規準のありようを対比的にまとめると、表2の通りである。

文字数の観点からは、文字数の多いユニコードが勝るが、包摂規準の観点からは、規準の全体が明示的に定義されているJISが

表2 JISとユニコードの収録文字数と包摂規準

	JIS	ユニコード
文字数	6879 (第1・第2水準)	70195 (ユニコード4.0)
	4344 (第3・第4水準)	
包摂規準	全体明示	一部明示

すぐれている。使用文字の種類の多い『太陽』では、文字数の多いユニコードを用いれば、外字の数を減らすことができる利点があるが、(7)(8)(9)(10)のように、ユニコードによっても外字として残るものは多い。また、字体のゆれが大きい『太陽』では、包摂できる字体の範囲が明確なJISの包摂規準は有益であるが、その規準では対応できない字体のゆれも大きく、JISの包摂規準では不十分な場合も多い。

このように、JISによるにしてもユニコードによるにしても、字体処理の困難な問題は解決しない。また、JISもユニコードも、見直しや改訂が繰り返されており、いずれかに合わせて電子テキストを作成したとしても将来にわたってそれがそのまま安定して使える保証はない。固定した文字集合や固定した包摂規準は、現状では望めないのである。なるべく問題の少ない字体処理の方針を独自に決め、今後の文字集合の改訂にも対応できる形にしておくことが必要である。

なお、『太陽』における漢字字体の実態とそれに応じた電子テキスト作成における処理の方法については、本書に別に掲載した「漢字の実態と処理の方法」(田中牧郎, 271頁から)に詳しく述べたので、参照してほしい。

3.6 誤植・誤用・通用

『太陽』には、現代の総合雑誌と比べて誤植が多い。

- (1) 歴史編纂の(1895年2号「史料の編纂は目下の急務たるを論ず」坪井九馬三P009A14)
- (2) 合衆國にに於ては(1895年1号「檳榔子」市村塘P174B14)
- (3) 之を退治たる程(1895年3号「加藤清正(承前)」小倉秀貫P038B15)
- (4) 達るすものにして(1895年4号「海外彙報」P195A16)
- (5) 二雜二を發行して(1895年1号「漢字の利害」三宅雪嶺P023A04)

『太陽』原文に見られるこれらの例は、(1)「吏」は「史」の誤字、(2)「に」が衍字、(3)「治」と「た」の間に「し」が脱字、(4)「るす」は「する」の転倒、(5)「誌」の欠損、の例だと、それぞれ判断される。

また、誤用ではないかと疑われるものの、『太陽』のなかに類例がいくつも見つかる場合もある。

(6) 心持を察しるとね (1901年1号「楹紅葉」広津柳浪P105B05)

(7) 氣嫌を取るを (1901年8号「文芸時評」高山樗牛P038A19)

(6) は、「察する」の誤りではないかと疑われるが、「衣服の風から察すると」(1901年2号「難破船」中内蝶二P095B01)、「視線の達しと」(1901年3号「海賊村」江見水蔭P027B12)などのように、サ変動詞の連体形が「しる」となる箇所は『太陽』において他にも例は少なくない。また、(7) は現代語では誤用とされる表記だが、「御氣嫌宜しく」(1895年8号「妄語戒即ち真語律に就て」渡辺龍聖P163B28)、「一杯氣嫌で」(1925年11号「浅草放浪記」記者P226A21)のように、例は少なくない。これらは、当時通用していた語法や表記であったとも考えられ、誤用と通用の線引きは容易でない。全般に語法や表記の規範は現代語の公的な書き言葉よりもゆるかったと見られる。

こうした誤植・誤用や通用には手を加えず、原文のままの形で電子テキストを作成するという考え方もあるが、活字媒体でのテキスト作成の場合に行われる本文批判の方法を、電子テキスト作成にも適用し、一貫した方針で誤植・誤用や通用について判断し、安心して使える校訂本文を確定することが望まれよう。また、文字列検索による利用形態が中心となるコーパスとして、利用者が検索しやすい本文を作成する配慮も必要であろう。

3.7 濁音無表記

濁音の表記に濁点を付ける規範は『太陽』においては確立しておらず、濁音が期待される箇所に濁点がない場合が多く見られる。例えば、次の例の下線部は、(1)「ぶ」、(2)「だ」、(3)「づ」、(4)「ず」、(5)「ぐ」の濁点表記が期待されるが、清音表記になっている。

(1) 佛語を學ぶこと (1895年1号「科学」P173C27)

(2) 甚た惜むべき (1895年2号「広島形勢」野口勝一P071A25)

(3) 先づ母親の顔を見知りて (1895年1号「婦人の令名」寒沢振作P150A06)

(4) 論するまでもなく、(1895年1号「国語研究に就て」上田万年P029A18)

(5) 妙作を掲ぐ (1895年1号「小説の記事説明」記者P08

3A02)

これらは、語形が清濁にゆれていたのではなく、濁音の正書法が未確立であることによるものだと考えられる。その実態についての詳細は、本書に収録した「濁点文字使用率から見る濁音表記」（近藤明日子、331頁から）を参照してほしい。表記がゆれていることが検索性を損ねないように、電子テキスト作成において何らかの対処を行うことが望まれる。

3.8 仮名遣い

『太陽』においては、仮名遣いの規範が確立しておらず、語によってゆれが大きいものが目立つ。歴史的仮名遣いとされている表記と異なる例は、多い。

- (1) 子の愛にも（1895年7号「子煩悩」大橋乙羽P107B01）
- (2) 奇峯を超へて（1895年1号「利根水源探検紀行」渡辺千吉郎P080A22）
- (3) 教えられて（1895年2号「階級制と君子の道」久米邦武P035A09）
- (4) 蒸汽を用いぬ（1895年2号「工業教育」手島精一P041B21）
- (5) 近衛の中將（1895年1号「正月」大橋乙羽P155A01）

上の語は、歴史的仮名遣いでは、(1) 愛、(2) 超え、(3) 教へ、(4) 用ゐ、(5) 中^{ちゆう}であることが期待されるが、それと異なった表記になっている。仮名遣いの実態については、本書収録の「仮名遣いについて」（小木曾智信、351頁から）を参照してほしい。電子テキストでは、仮名遣いの不統一が検索もれを起こす可能性は高いので、濁音無表記の場合と同じく、電子テキスト作成において対処することが望まれる。

3.9 特殊な表記法

上にあげたもののほか、現代日本語の書き言葉には通常見られないか、きわめて限定的になっている表記法が、『太陽』においてはかなり広く使われている例として、踊り字、割書、敬意欠字、合字、片仮名小書なども目立つ。これらの中には、電子テキスト作成において、対処が求められるものも多い。

4

コーパスの対象とする範囲

『太陽コーパス』は、現代日本語の確立期のコーパスという位置づけから、まず20世紀の始点として1901年を対象としてとった（注7）。言語の通時的な変化を段階的にとらえることができるようにし、またその中の一時点における共時的な記述にも堪えられる分量をそろえることを考えた。こうした考えから1901年を起点に8年刻みで、1909年、1917年、1925年をとり、さらに創刊年にさかのぼり1895年をとった。各年は臨時増刊号を除く12冊全号を対象にし、全体で60冊とした（注8）。

原則として雑誌各冊の全記事・全文を対象とするが、次のような記事は、対象外とした。

- （1）表紙。目次。
- （2）口絵や図表を中心とする記事。口絵や図表に付随する文章記事。
- （3）広告。博文館からの社告や読者アンケートの呼びかけなどのうち、自社広告と判断される記事。
- （4）漢文や欧文が主体の記事。
- （5）前号などの訂正記事。ただし、訂正記事での、誤植の指摘などは、これを反映させて電子テキストを修正した。
- （6）連載記事や前号等で予告していた記事について、不掲載を詫げる記事。
- （7）著者に関わる著作権処理ができなかった記事。

（1）から（6）は、『太陽コーパス』の作成目的に照らして、対象に含める必要性が低いと判断したものであるが、（7）は、目的に照らせば対象に含めたいものであったが、法律上あるいは倫理上の問題を解決することができなかったために、やむなく除外したものである。著作権処理ができずにコーパスに含めなかった記事の一覧は、本書の末尾に掲げた（注9）。

また、対象にした記事の中でも、次の部分はコーパス化の対象にはしなかった。

- （8）図表。
- （9）漢文や欧文の長文の引用。
- （10）挿絵や写真のキャプション。

このうち、図表、漢文、欧文は、本文の途中で割り込んで現れる

ことがあり、その存在が電子テキスト上から完全に消えてしまうと本文の読解ができなくなる場合がある。そうした場合は、その位置に図表等の存在があることを、示すようにした。

5

構造化の基本方針

欧米などの本格的なコーパスにおける構造化テキストの記述には、構造化マークアップ言語が用いられることが一般的である。最近では、W3C (World Wide Web Consortium, インターネット関連技術の標準化組織) によって勧告された国際規格、XML (eXtensible Markup Language) を用いることが主流になりつつあり、今後もその流れは進展するものと考えられる(注10)。『太陽コーパス』の記述にもこのXMLを採用したが、コーパスの記述にXMLを用いる利点として次のようなことがあげられる。

- (1) 各種の情報を一つのファイルに書き込むことが可能であり、それらを構造化して管理することができる。
- (2) データの妥当性の検証機能によって、データの保守管理が容易になる。
- (3) データ形式を変換したり必要な情報を抽出したりすることが容易である。

構造化の単位は、言語の階層構造に即せば、文章→文→句→語→形態素といった階層的なものが期待される。しかし、『太陽』では文章や文の単位が明示的でないため、階層的な構造化を厳密に行うことは技術的に困難である。また『太陽コーパス』は形態素解析や構文解析は行わないので、文・語・形態素などの認定は、仕様上必須ではない。こうした事情から、『太陽コーパス』における構造化は、言語の階層構造と厳密に対応させたものとはせず、ゆるやかな単位によるものとした。

一方、多彩な著者・ジャンル・文体等の情報は、検索や抽出が可能な形に生かすことが望まれ、これらをタグに書き入れることで、情報を容易に取得できるように工夫した。さらに、ゆれが大きかったり規範が未確立だったりする表記法については、規範形や仮に設定した規範形に修正したテキストをつくり、原文の形はタグに書き入れておき、必要に応じて参照できるようにした。こ

これらのタグによって付加する情報が整合するような構造化を行うことで、整った構造化タグ付きコーパスを実現させようとした。

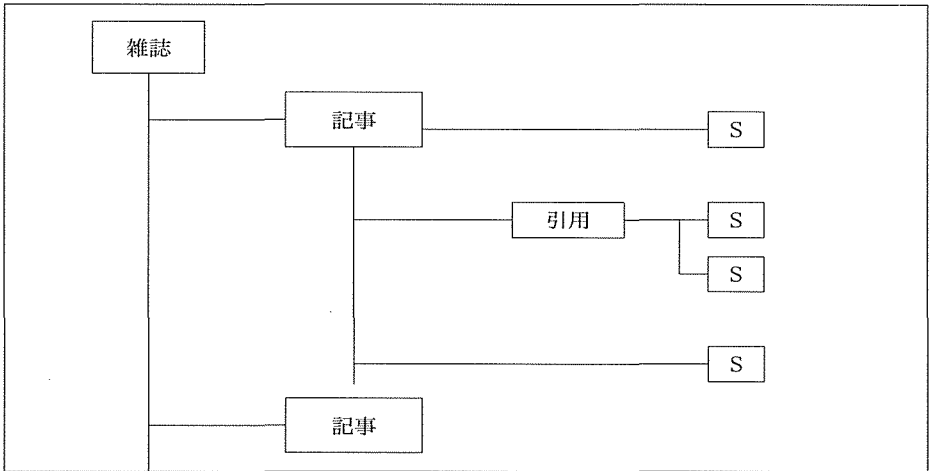


図1 『太陽コーパス』の構造化

まず、図1のように「雑誌」（各号1冊全体）、「記事」、「引用」、「s」の四つの箱形（ブロック）要素を設定し、基本構造を構築した。現代語の書き言葉では目立たないが、『太陽』によく出現する特有の表記法については、行内（インライン）要素としてタグ付けを行い、必要に応じてブラウザ表示や印刷、情報抽出等に用いることができるようにした。著者、ジャンル、文体等の情報は、上記の箱形要素や行内要素のタグ属性に書き込み、検索結果に表示したり、検索条件に指定できるようにした。

このようにして、XMLによって構造化してタグを付けたコーパスも、変換・抽出・検索などを効率よく行うことができなければ、その価値を調査研究に生かすことは難しい。そこで、コーパスを用いた言語研究に役立つソフトウェアをいくつか開発しコーパスに実装した。『太陽コーパス』を利用するためのソフトウェアについては、本書の「構造化テキストに対応した全文検索システム『ひまわり』」（山口昌也、49頁から）、「構造化テキストを直接利用するアプリケーション『プリズム』と『たんぽぽ』」（小木曾智信、83頁から）の二つの論文を参照してほしい。以下では、『太陽コーパス』の構造化とタグ付けの方法について、具

体的に述べる。

6

タグセット

表3 『太陽コーパス』のタグセット

	タグ(要素)	説明	属性
箱形要素	雑誌	『太陽』1冊全体	雑誌名, 年, 号, version
	記事	各記事全体	題名, 著者, 欄名, 文体, ジャンル
	引用	他の典拠・話者からの引用部分	種別, 話者, 文体
	s	句読点等によって区切られた, 擬似的な文	
行内要素	br	強制改行, 空要素	
	段落記号	段落の切れ目を示す鉤括弧, 空要素	原文
	l	『太陽』原誌上の頁行番号, 空要素	位置, 元位置
	r	振り仮名(ルビ)	rt
	外字	JIS外字	文字番号, 文字情報
	注	校訂部分への注記	分類, 原文
	踊字	踊り字	種類, 値
	割書	割り書き	
	割書改行	割り書き中の改行, 空要素	
	敬意欠字	皇室への敬意を示す欠字	
	合字	合字	
	小書	片仮名小書き	
	上付	数式等における上付き文字	
	下付	数式等における下付き文字	
	非入力対象	コーパスに含めなかった本文, 空要素	種類, 表見出し

『太陽コーパス』の構造化と情報付与に用いたタグは、表3の通りである。タグ付けの書式は、タグ名を<>の中に入れて、例えばsタグであれば、

<s>一陽復歸して萬象維れ新に、</s>

のように、開始タグ<s>と終了タグ</s>とで本文を囲む形をとる。囲まれる本文をもたないタグは「空要素タグ」と呼ばれ(表3で「説明」の欄に「空要素」とあるもの)、例えばbrタグであれば、

<s>(中略)江湖諸君の賛成を仰がんとす。
</s>

のように、
の形で記し、開始タグと終了タグの区別はない。

「属性」（タグで囲まれた本文が持っている性質のうち特に記しておくべきことからや、タグによって埋め込みたい情報）をもつ場合には、開始タグの中書き込む。例えばrタグの場合であれば、

```
<r rt="ふり">降</r>
```

のように、属性名「rt」の後に「=」を付し、"の中"に「値」（属性として記される内容）を書き入れる。これらは、XMLの記法にのっとって、独自に定義したものである。

以下、一つ一つのタグによってどのような言語情報が取得できるようにしたかについて、詳しく述べていきたい。

7

雑誌タグ

雑誌『太陽』の、各1冊分を構造化の最上位の階層に置き、ルートタグ（おおもとの構造化単位）とした。コーパスを構成する雑誌本文の電子テキストは、雑誌タグとしてまとめられる要素一つを1ファイルとした。属性として、「雑誌名」「年」「号」「version」をもつ。「雑誌名」の属性を設けたのは、将来『太陽』以外の雑誌にも拡張したコーパスとして発展させる可能性を想定したからである。「年」「号」のほか、ルートタグになるのでXML定義のversionも属性に加えた。下に示すものは、1895年1号の雑誌タグ（開始タグ）である。四つの属性の後には、『太陽コーパス』のXML定義に関わる情報も書き入れてある。

```
<雑誌 雑誌名="太陽" 年="1895" 号="01" Version="1.0"
xsi:schemaLocation="http://www.kokken.go.jp/taiyo
zassi.xsd" XMLNs:xsi="http://www.w3.org/2001/XM
LSchema-instance" XMLNs="http://www.kokken.go.
jp/taiyo">
```

『太陽コーパス』が含む電子テキスト60ファイルは、『太陽コーパス』に含まれる雑誌60冊の電子テキストに相当するが、その概要を示すと表4のようになる。

表4 『太陽コーパス』に含まれる雑誌一覧

巻号	発行年月	記事数	文字数	ファイル名	特集等
第1巻第 1号	1895年 1月	71	288512	t189501.XML	
第1巻第 2号	1895年 2月	62	285349	t189502.XML	
第1巻第 3号	1895年 3月	57	259004	t189503.XML	
第1巻第 4号	1895年 4月	66	263722	t189504.XML	
第1巻第 5号	1895年 5月	59	266692	t189505.XML	
第1巻第 6号	1895年 6月	57	263254	t189506.XML	
第1巻第 7号	1895年 7月	62	273043	t189507.XML	
第1巻第 8号	1895年 8月	65	291180	t189508.XML	
第1巻第 9号	1895年 9月	62	281857	t189509.XML	
第1巻第10号	1895年10月	58	286722	t189510.XML	
第1巻第11号	1895年11月	55	283574	t189511.XML	
第1巻第12号	1895年12月	55	292457	t189512.XML	
第7巻第 1号	1901 年 1月	58	267395	t190101.XML	
第7巻第 2号	1901 年 2月	59	265322	t190102.XML	
第7巻第 3号	1901 年 3月	61	266444	t190103.XML	
第7巻第 4号	1901 年 4月	49	271316	t190104.XML	
第7巻第 5号	1901 年 5月	56	270162	t190105.XML	
第7巻第 6号	1901 年 6月	52	268062	t190107.XML	
第7巻第 8号	1901 年 7月	55	271824	t190108.XML	
第7巻第 9号	1901 年 8月	57	259745	t190109.XML	
第7巻第10号	1901 年 9月	54	264239	t190110.XML	
第7巻第12号	1901 年10月	52	251111	t190112.XML	
第7巻第13号	1901 年11月	41	244547	t190113.XML	
第7巻第14号	1901 年12月	41	254387	t190114.XML	
第15 巻 1号	1909年 1月	59	254978	t190901.XML	新年大附録(小説2編)
第15 巻 2号	1909年 2月	60	228858	t190902.XML	
第15 巻 4号	1909年 3月	57	235719	t190904.XML	
第15 巻 5号	1909年 4月	52	255686	t190905.XML	
第15 巻 6号	1909年 5月	56	253037	t190906.XML	
第15 巻 8号	1909年 6月	55	202109	t190908.XML	
第15 巻10号	1909年 7月	53	240886	t190910.XML	
第15 巻11号	1909年 8月	53	234143	t190911.XML	
第15 巻12号	1909年 9月	56	229355	t190912.XML	
第15 巻13号	1909年10月	53	238068	t190913.XML	

巻号	発行年月	記事数	文字数	ファイル名	特集等
第15巻14号	1909年11月	48	244070	t190914.XML	
第15巻16号	1909年12月	50	243443	t190916.XML	
第23巻1号	1917年1月	43	295021	t191701.XML	新年大附録(小説6編)
第23巻2号	1917年2月	41	228082	t191702.XML	
第23巻3号	1917年3月	41	230337	t191703.XML	
第23巻4号	1917年4月	42	312399	t191704.XML	春期大附録(小説4編)
第23巻5号	1917年5月	26	165115	t191705.XML	
第23巻6号	1917年6月	34	182244	t191706.XML	
第23巻8号	1917年7月	40	199065	t191708.XML	
第23巻9号	1917年8月	50	190045	t191709.XML	
第23巻10号	1917年9月	46	290328	t191710.XML	秋期大付録(小説6編)
第23巻12号	1917年10月	45	186535	t191712.XML	
第23巻13号	1917年11月	59	207866	t191713.XML	
第23巻14号	1917年12月	37	171135	t191714.XML	
第31巻1号	1925年1月	123	212180	t192501.XML	特別記事:最近科学発明界の進歩
第31巻2号	1925年2月	76	207831	t192502.XML	特別記事:少壮新人の天下
第31巻3号	1925年3月	65	178783	t192503.XML	
第31巻4号	1925年4月	79	182103	t192504.XML	特別記事:無線電話号
第31巻5号	1925年5月	61	166168	t192505.XML	
第31巻7号	1925年6月	65	198206	t192507.XML	
第31巻9号	1925年7月	68	205474	t192509.XML	
第31巻10号	1925年8月	65	194802	t192510.XML	
第31巻11号	1925年9月	83	252287	t192511.XML	
第31巻12号	1925年10月	72	207309	t192512.XML	
第31巻13号	1925年11月	70	212102	t192513.XML	労働問題研究号
第31巻14号	1925年12月	62	229851	t192514.XML	農村問題号

1909年(第15巻)までは特集も少なく、各巻同じような形式で編集されている。1917年(第23巻)以後は付録や特別記事あるいは特集を立てる号が目立ってくるが、そうした号についても原則として1冊全文をコーパスの対象としたので、1冊あたりの分量(文字数)が他の号より多くなっている場合がある。

表5 『太陽コーパス』（公刊版）の年次別記事数・文字数

年	記事数	文字数
1895年	729	3335366
1901年	635	3154554
1909年	652	2860352
1917年	504	2658172
1925年	889	2447096
計	3409	14455540

年次単位で記事数・文字数をまとめると、表5のようになる。1895年と1901年は300万字を超えているが、この2年分は各冊のほとんど全文をコーパスに含めている。1909年以後、徐々に減少していくのは、記事著者の著作権の問題により公刊するコーパスからは除外した記事が、年次を追うごとに増えていくためである（除外した記事の一覧は本書の終わりに掲げた）。経年的な比較を厳密な統計処理によって行う場合など、この分量の出入りに留意する必要がある。なお、記事除外前の『太陽コーパス』（評価版・内部資料）の年次別の記事数・文字数を表6に掲げた。本書第2部の研究論文のほとんどは、『太陽コーパス』を作成しながら研究を進めたものであるので、この表6に示す範囲のデータを対象に調査・研究を行ったものである。

表6 『太陽コーパス』（評価版・内部資料）の年次別記事数・文字数

年	記事数	文字数
1895年	732	3340782
1901年	643	3239029
1909年	715	3102795
1917年	550	2960783
1925年	1081	3423588
計	3721	16066977

8

記事タグ

8.1 属性の種類

各雑誌に含まれる記事を、「雑誌」要素につぐ二番目の階層の単位とし、記事一つ一つを「記事」要素として設定した。原誌において、記事の切れ目か記事内部の節や項の切れ目かが不明瞭な場合もあるが、原則として、著者が変わることが明示されている位置を、記事の切れ目と認定した。ただし、無署名記事が続く場合など、欄の切れ目を記事の切れ目に認定した場合もある。

記事タグには、記事の「題名」のほか、「著者」「欄名」「ジャンル」「文体」といった、言語研究にとって基本的な情報を、属性として設けた。各属性の内容は次の通りである。著者以下の属性については、次項以下で詳しく述べる。

題 名：記事の題名。本文に表示される題名を基本とするが、連載記事などで号によって表記が変わるものなど、必要に応じて表記を統一した場合がある。

著 者：著者の氏名。ペンネーム等を用いて、複数の名前前で記事を書いている著者や、一般的に知られた名前と異なる名前前で記事を書いている著者は、現在最も知られている一般的な名前に統一した場合がある。無署名記事の場合、「*」を記入した。

欄 名：記事のいくつかをひとまとめにして「欄」を構成している場合、その欄の名称を記入した。欄のない記事の場合、「**」を記入した。

ジャンル：日本十進分類法（NDC）により、記事をジャンルに分類し、分類番号を記入した。扉など分類できない記事の場合、「***」を記入した。

文 体：文末辞を指標として、口語・文語の別を記入した。奥付など、名詞の列挙の形をとる記事で口語・文語の別のないものは、「項目」という属性値を記入した。

記事タグ（開始タグ）は、次のような形で記述した。

<記事 題名="戦勝後の教育" 著者="千頭清臣" 欄名="論説" 文体="文語" ジャンル="NDC371">

8.2 「著者」属性

「著者」については、著者名だけでなく、その著者の生年や所属・分野といった属性情報を整理しておき、その情報がコーパスを利用する際に取得できる形にしておけば、言語の研究において有益である。そうした著者情報は、別の一覧（著者データベース）を作成し、コーパスを使って変換や検索を行う際に、その一覧から情報を取得し、変換結果や検索結果として表示できるようにした。著者データベースの利用方法、著者を一覧する方法については、本書の「構造化テキストに対応した全文検索システム『ひまわり』」（山口昌也）の「6.1.5 著者データベース機能」（75頁）を参照してほしい。

著者データベースに含まれる著者の総数は約1000人で、ここには、短歌・俳句等、ごく短文で、独立の記事には扱わなかった著者は含んでいない。著者データベースから、変換・検索時に引き出せるように整えた情報は、次のものである。

氏 名：各記事の著者名。複数名を使い分けたり、表記にゆれがあったりする場合など、同一人物でありながら氏名が二通り以上考えられる場合は、一つにまとめた。また、現代もっとも通用している氏名やその表記にまとめた場合もある。

所 属：『太陽』原本に、勤務先や居住地などの記載がある場合はそれを生かし、『著作権台帳』（日本著作権協議会）、『現代日本人名録物故者編1901～2000』（日外アソシエーツ）、『日本人名大辞典』（講談社）などに情報のあるものはそれを生かした。所属情報が不明のもの、次の「分野」に書き入れた情報だけで十分と思われるものは、空欄にした。所属が多数にわたる著者の場合は、原則として二つまで（最大で三つまで）記した。

分 野：『太陽』原本に、職業や専門分野に関する記載がある場合はそれを生かし、『著作権台帳』や、『現代日本人名録物故者編1901～2000』『日本人名大辞典』などに情報のあるものは、それを生かした。分野情報が不明のもの、上の「所属」に書き入れた情報だけで十分と思われるものは空欄にした。職業や専門分野が多数にわたる著者の場合は、原則として二つまで（最大で三つまで）記した。

生 年：著者の生年。『著作権台帳』や、『現代日本人名録物故者編1901～2000』『日本人名大辞典』などを参照して記した。生年が不明の場合は空欄とした。

没 年：著者の没年。『著作権台帳』や、『現代日本人名録物故者編1901～2000』『日本人名大辞典』などを参照して記した。没年が不明の場合は空欄とした。

著者データベースの具体的な記述は、次のような形になっている。

<氏名>千頭清臣</氏名>
 <所属>貴族院議員</所属>
 <分野>官僚 政治家</分野>
 <生年>1856</生年>
 <没年>1916</没年>

この著者データベースを参照すれば、『太陽コーパス』に含まれる著者の特徴を概観できる。まず、生年は、1806年から1908年の約100年の範囲をカバーしている。その分布を整理すると、表7のようになる。1851年から1890年までの40年間に生まれた人が、約8割を占める。19世紀生まれの人々の言葉の実態を知ることができ、特に19世紀後半生まれの人々の言葉の実態をもっともよく反映しているコーパスであると考えることができる。この生年情報を有効に活用することで、著者の生年から見た言葉の新古や変化、世代による言葉の違いなどの記述が可能になると考えられる。

表7 『太陽コーパス』における著者の生年

生年範囲	人数
1801～1810年	1
1811～1820年	3
1821～1830年	13
1831～1840年	30
1841～1850年	58
1851～1860年	135
1861～1870年	197
1871～1880年	167
1881～1890年	88
1891～1900年	25
1901～1910年	5

著者の所属、専門分野の情報を見ると、政治家、実業家、作家、翻訳家、芸術家、評論家、宗教家、教育家、法律家、学者、技術者、軍人、詩人、ジャーナリスト、官僚など広範囲に及んでいることがわかる。

8.3 「欄名」属性

「欄名」は、記事のまとまりやジャンルを知るのに役立つものであり、『太陽』自体の内発的な記事分類枠と見なすことができ、コーパスにおいても活用することが望まれる。欄の機能には、1冊のなかで関連あるテーマの記事を束ねる働きと、複数号を通して同一シリーズとして関連づける働きと、二つがあると考えられる。『太陽』では、1895年、1901年、1909年は、この両方の機能を持たせた欄が整備され、ほとんどすべての記事がいずれかの欄に配される形をとっている。ところが、1917年と1925年では、欄は1冊のなかで記事を束ねる機能に限られるようになり、いずれの欄にも属さない記事が多くなる。

下に、各年の欄名の一覧を掲げる。

1895年

論説、講演、史伝、地理、小説、雑録、文苑、芸苑、家庭、政治、法律、軍事、文学、科学、美術、教育、宗教、医事、商業、農業、工業、社会、海外思想、輿論一斑、社交案内、新刊案内、実業案内、海外彙報、海内彙報

1901年

太陽、論説、人物月旦、文芸時評、政治時評、法律時評、教育時評、宗教時評、経済時評、名家談叢、小説雑俎、歴史地理、家庭談叢、文苑、農業世界、工業世界、商業世界、科学世界、社会事情、海外事情、輿論一斑、寄書、海内彙報

1909年

太陽、時事評論、人物月旦、論説、文芸、学芸、雑纂、新年大附録

1917年

○複数の号にわたって設定されている欄名

時局の印象、無名隠士夜話、案頭三尺、婦人界評論、経済財政時論、教育時言

○特定の1号のみに設定されている欄名

財時界事、旅と酒と、文部省展覧会評、文部省展覧会日本画家の作意と苦心、初めて議会に列して、故奥田義人男、

新年大附録，春期大附録，秋期大附録

1925年

○複数の号にわたって設定されている欄名

財界時事小話，文芸時評，凡人非凡人，同人会談叢，現代怪傑伝，茶話

○特定の1号のみに設定されている欄名

欧米印象記三篇，科学新話，驚嘆すべき最近科学界の進歩，困難に打克つた経験，失敗から得た教訓，最近奇聞画報，産児制限か産児保護か，思想悪化の傾向と批判，秋の美術界評論，出世の早い社員の特長，勝負事の趣味，新国難襲来の真相と国民の覚悟，世界のラヂオ，大衆文学の傾向を如何に見るか，大町桂月翁の面影，第六回美術展覧会批評，年のくれの印象，能率増進と生産費節減，発明界珍事，抜擢した人と抜擢された人々，不景気は何時までつづくか，無線電話欄，予が最も感化をうけたる人々三篇，予の最も敬服したる書物と人物

1895年から1901年への流れを見ると，廃止されたり（「雑録」「芸苑」「軍事」「美術」など），新設されたり（「太陽」「人物月旦」「寄書」），統合されたり（「史伝」「地理」→「歴史地理」）するものが多少あるが，大部分は名称を変えて継続するものである（「文学」→「文芸時評」，「政治」→「政治時評」，「法律」→「法律時評」，「教育」→「教育時評」，「宗教」→「宗教時評」，「小説」→「小説雑俎」，「家庭」→「家庭談叢」，「農業」→「農業世界」，「工業」→「工業世界」，「商業」→「商業世界」，「科学」→「科学世界」など）。1895年から1901年にかけて，欄構成に大きな変化はないと考えてよい。

ところが1909年には，欄の数が大幅に減少している。記事の内容を見ると，1909年の「時事評論」には，1901年の「政治時評」「法律時評」「教育時評」「宗教時評」「経済時評」などが引き継がれ，同じく「文芸」には「文芸時評」のほか「小説雑俎」が継承されているようである。さらに，「学芸」には「農業世界」「工業世界」「科学世界」などが，「雑纂」には「家庭談叢」「海外事情」「寄書」「海内事情」などが，継承されていると見られる。このように，1909年では掲載されている記事のジャンルには大きな変化はないが，欄の区分けが大きくくりになっているのである。

1917年・1925年は，上で複数の号にわたって設定されている欄名か，特定の一号に限って設定されている欄名かに分けて示し

た通り、特定号のみで設定されている欄の方が多くなっている。これは、特集や各号のテーマ設定に基づく編集であると考えられる。創刊以来の基本的な欄構成は、この段階では解体されており、欄名から記事のジャンルを探ることはできなくなっている。欄名は『太陽コーパス』全体を通じてジャンルの指標とすることはできず、欄名以外にも、ジャンルを分類する指標を求める必要がある。

8.4 「ジャンル」属性

総合雑誌が扱うかなり広範なジャンルの記事全体を分類できる枠組みとして、『太陽コーパス』では、図書館での書籍の分類に用いられる「日本十進分類法」(NDC)を採用した。NDCは、内容によってほとんどすべての書籍を分類できる枠組みとして長年にわたる実績がある。第1次から第3次までの3層構造を000から999までの3桁の数字で示すものである(注11)。『太陽』の各記事に、NDCにしたがって3桁の番号を付与した。

『太陽コーパス』の年次別のジャンル別記事数と比率をまとめると、表8ようになる。「***」は、扉、奥付、予告等、分類が不可能なものに付けている。全体に「3. 社会科学」がもっとも多く、『太陽』が社会的な素材をもっともよく扱っていることを示している。

どの年次も、ほぼ全ジャンルをカバーしており、『太陽』のジャンルの広さを確認することができる。年次による比率の相違を見ると、「0 総記」が、1925年、1917年で高くなっている。これは、1925年では、いくつかの号で雑学的な短文が特集されており(約120件)、1917年では、新刊紹介が短い記事に分かれている(約80件)ことによるものである。また、「7 芸術」が1917年でやや多くなっているが、2回にわたって美術展に関する諸家の文章を特集しているためである。その他は、年次による出入りはそれほど大きくなく、『太陽コーパス』は、全体を通じて各ジャンルにわたる文章を広く収めていると言うことができる。

なお、表8で全体に多かった「3 社会科学」の記事の内訳をNDCの第2次区分で整理すると、表9ようになる。

表8 『太陽コーパス』の年次別ジャンル一覧

NDC	1895年		1901年		1909年		1917年		1925年	
	記事数	比率	記事数	比率	記事数	比率	記事数	比率	記事数	比率
0 総記	22	3%	9	2%	46	7%	93	18%	218	21%
1 哲学	45	6%	46	8%	13	2%	8	2%	12	1%
2 歴史	78	11%	60	10%	34	6%	10	2%	83	9%
3 社会科学	201	29%	248	42%	288	47%	195	42%	191	23%
4 自然科学	36	5%	28	5%	34	5%	14	3%	33	5%
5 技術	41	6%	35	7%	18	3%	16	4%	70	9%
6 産業	51	7%	51	9%	20	3%	6	2%	38	5%
7 芸術	60	9%	14	2%	28	4%	46	11%	74	8%
8 言語	12	2%	4	1%	3	0%	2	0%	2	0%
9 文学	153	22%	97	16%	133	22%	77	17%	147	18%
計	699	100%	592	100%	617	100%	467	100%	868	100%
***	30	—	43	—	35	—	37	—	21	—
全記事計	729	—	635	—	652	—	504	—	889	—

表9 『太陽コーパス』における「3社会科学」の下位分類

NDC	1895年		1901年		1909年		1917年		1925年	
	記事数	比率	記事数	比率	記事数	比率	記事数	比率	記事数	比率
30 社会科学	54	27%	61	25%	95	33%	24	12%	37	19%
31 政治	32	16%	50	20%	108	38%	84	43%	64	34%
32 法律	13	6%	16	6%	8	3%	17	9%	8	4%
33 経済	41	20%	51	21%	18	6%	22	11%	47	25%
34 財政	7	3%	13	5%	11	4%	6	3%	1	1%
35 統計	0	0%	0	0%	0	0%	0	0%	0	0%
36 社会	14	7%	16	6%	15	5%	4	2%	27	14%
37 教育	13	6%	24	10%	17	6%	19	10%	4	2%
38 風俗習慣, 民俗学,民族学	15	7%	11	4%	3	1%	4	2%	0	0%
39国防, 軍事	12	6%	6	2%	13	5%	15	8%	3	2%
計	201	100%	248	100%	288	100%	195	100%	191	100%

全般に、「31政治」「33経済」のジャンルが多く、「35統計」の記事が皆無であるなど、ジャンルによる記事の多寡は見られるが、

年次による極端な変化はない。

ジャンルに関する情報は、目的に応じて、「欄名」属性と「ジャンル」属性とを補い合わせながら活用すると、いっそう有効度が上がると考えられる。

8.5 「文体」属性

「文体」については、言文一致に向かうこの時期の顕著な潮流の中で言語現象をとらえることができるように、文語を基調とする記事か、口語を基調とする記事かの区別を、属性として記入した。文語か口語かの認定は、原則として次に掲げる主要な文末辞を指標に行った。なお、韻文の場合も、口語詩か文語詩（和歌・俳句を含む）かの別にしたがって、文語か口語かの情報を与えた。

文語の指標とした文末辞 なり たり あり り つ ぬ
 き べし

口語の指標とした文末辞 です ます ござる である
 だ た

口語と文語が混在しているかに見える記事が若干あるが、どちらがより中心的な文体（基調文体）であるかという観点から、口語か文語かのいずれかに判別した（注12）。基調文体から外れる箇所が引用部分である場合は、引用タグ（後述）でタグ付けしたところに「文体」属性を書き入れるようにした。文章の形をとらない奥付や一覧表等の記事には、「項目」という属性値を与えた。表10のように、1895年では大半が文語記事であるが、年次を追うごとに、次第に口語の比率があがっていき、1925年では大半が口語記事になっている。文体の経年変化は、通時的観点から『太陽コーパス』を見たときに最も際立つ特徴である。

表10 『太陽コーパス』の年次別文体別記事数

年	総記事数	口語記事	口語記事率	文語記事	文語記事率	項目記事
1895年	729	39	5.3%	689	94.5%	1
1901年	635	168	26.5%	467	73.5%	0
1909年	652	376	57.7%	267	41.0%	9
1917年	504	355	70.4%	129	25.6%	20
1925年	889	835	93.9%	52	5.8%	2
計	3409	1773	52.0%	1604	47.1%	32

9

引用タグ

「記事」は、著者が読者に向けて書いたものであるが、記事の中に、著者以外の人が用いた言葉を引用した部分が含まれることがある。その部分の言葉は、引用された書き手または話し手（これをまとめて「話者」とする）の言葉と考えることもできる。また、フィクションの場合は、登場人物の言葉と扱うこともできる。そうした引用された言葉の場合、誰に向かってどんな状況で発せられた言葉であるかというような、場面に関する情報が扱いやすい形に整理してあれば言語の分析にとって有益である。つまり、引用された部分は話者や場面という観点から、地の部分と区別して扱えるようにタグ付けを行っておくことが望まれるのである。特に、『太陽』では、3.2で見たように、引用の表示形式が十分確立しておらず、コーパス利用者が引用箇所を認定することが容易でない場合があるため、引用箇所のタグ付けをあらかじめ行っておくと、利用の際の利便性は格段に高まると思われる。

『太陽コーパス』では、引用部分に引用タグを付し、文章の種別と話者（場合によって文体も）を識別できるように、「種別」「話者」（場合によって「文体」も）の属性を書き入れた。引用と認定するのは、次の四つのいずれかに該当するものである。例示では、下線部が引用部分であることを示す。

(1) 当該記事の著者以外の言葉の部分

例：ウォルツウオスはコレツヂが山水を摸寫するを見て、曰く、之を措めよ之を措めよ、而して寧ろ活きたる天地に親炙せよ、紫山緑水自ら口を衝て湧き出でんと。是れ亦た心靈を説きたるものに非ずや。(1895年5号「明治の心靈界」松村介石P029B12)

(2) 当該記事の著者の言葉であっても、別の場で述べたことを引用している場合

例：余は長詩形に關しては、「今の世の中には、まさか和歌のみを唯一の詩形とすべしと主張するが如き頑冥の徒はあらざるべし」と云ひ、(1901年1号「文芸時評」大町桂月P041A03)

(3) 当該記事の著者の言葉であっても、鉤括弧等で、他の部分とは区別して特に強調してある場合

例：是に於てか余は大膽にも論斷せり。曰く「支那は遠からずして滅亡すべし。其人民は侵略者の爲めに逐はれて流離四散し、第二のジウ人と爲るべし」と。(1901年1号「対清政策」尾崎行雄P041B17)

(4) フィクションにおける登場人物の発話の部分

例：「こりや奈何も厄介だねえ。」観音丸の船員は累々しき盲翁の手を執りて、舳より本船に扶乗する時、恁は呟きぬ。(1895年1号「取舵」尾崎紅葉P083B10)

「種別」属性には、引用部分を次の6種に分類し、それぞれの種別を記入した。

会話：会話のほか、演説や独話の引用部分。フィクションにおける登場人物の発話部分。

心話：声や文字の形では発話されない心内で発せられた言葉の部分。フィクションにおける登場人物の心話部分。

典拠：他の文献などからの引用部分。手紙や法令・メディアなどからの引用、学説や声明などの引用も含む。

韻文：散文記事における、韻文の部分。全体が韻文の記事にはこの属性は付けない。

漢文：漢文の部分。全体が漢文の記事は、コーパスの対象外。

記事説明：記事に対する編集者などによる説明の部分。

「話者」属性には、発話者名または典拠になった文献名などを記入した。同一記事や連載記事のなかで、呼び名が変わる話者は、原則として一つの呼び名に統一して記した。また、記事本文に表示される名前では分かりにくい場合などは、分かりやすい名前に修正したり、肩書きを付与した。

「文体」属性は、記事タグに書き入れた「文体」と異なる文体となる引用部分の場合にのみ記入する。文体の認定方法は、記事タグの「文体」の場合と同じである。

引用タグの記述例を、上記(1)の例によって示せば次の通りである。

<s>ウオルヅウオスはコレツヂが山水を摸寫するを見て、</s>

<s>曰く、</s>

<引用 種別="典拠" 話者="ウオルヅウオス">

<s>之を措めよ之を措めよ、</s>

<s>而して寧ろ活きたる天地に親炙せよ、</s>

<s>紫山緑水自ら口を衝て湧き出でん</s></引用>

<s>と。</s>

<s>是れ亦た心靈を説きたるものに非ずや。
</s>

なお、引用部分を一覧して見る方法については、本書の「構造化テキストを直接移用するアプリケーションー『プリズム』と『たんぽぽ』ー」（小木曾智信）の「3.1.2情報抽出」（90頁から）を参照してほしい。

10

s（文）タグ

5で述べた通り、『太陽コーパス』では、言語の階層構造に即した厳密な構造化は行わず、文の認定も行わないが、文章に適度な長さで切れ目を入れておくことは、コンピューター処理や人間による読解のためには不可欠なことである。そこで、段落や文の認定に代わる便宜的な切れ目を設けて構造化を行った。

現代の書き言葉であれば、句点「。」で形式的な文の区切りと認めるが、『太陽』では、文の切れ目が「。」だけでなく「、」で表される場合も多いので、形式的な区切りとして「。」または「、」のいずれかで区切られた範囲を、擬似的な文「s」とした。一律に句読点を指標としたために、「s」の単位で区切られる範囲は、名称の並記の部分など非常に短い場合から、段落の切れ目まで区切りのない非常に長い場合まで、さまざまな長さになっている。ただし、句読点の極端に少ない記事や、句読点が脱落していると考えられる箇所など、原文に「。」や「、」がなくても適宜「s」の区切りを入れた場合もある。なお、sタグには属性は書き入れない。

11

br(強制改行)タグ、段落記号タグ、l(原文位置)タグ

段落に近似する単位の切れ目を表示させるために、強制改行を示す行内要素brタグを設けている。誌面上で、段落末に強制改行がある位置や、箇条書き、列挙等の場合、行末の切れ目の位置に、この要素を空要素タグとして挿入した。

また、『太陽』では段落の切れ目を表す記号として鉤括弧が用

いられることがあるが、コーパス利用者にとっては、引用を表示する鉤括弧と紛らわしい。この種の鉤括弧は本文からは削除し、その位置に、次のような段落記号タグを空要素として挿入したことがある。

<段落記号 原文="』"/> <段落記号 原文="『"/>

『太陽』原文での位置を、電子テキスト中に埋め込み、「位置」属性に、頁・段・行を組み合わせた番号を書き入れたものが、l（原文位置）タグである。たとえば、2頁の1段目の3行目であれば、「位置」属性にP002A03のように書き入れる。一つの記事の文章が、頁を隔てて連続するような場合は、原文の頁・段・行とは異なる番号を入れ、記事内の文章が電子テキスト内で連続するようにした。その場合、原文での頁・段・行を示す記号番号は、「元位置」属性に記入した。このlタグは、コーパスデータを原文と照合する必要がある場合や、検索や変換のソフトウェアで、別々に索引化された情報を関連づける場合などに有用になる。

sタグ、brタグ、lタグが付与された電子テキスト本文は、次の例のようになる。

<s>今後の同胞四千餘萬は復た深窓に眠るの日本<l 位置
="P002A03" />人に非ずして、</s>

<s>五大洲中に濶歩するの大日本人と爲れり。</s>

<s>豈我邦第二の維新を爲す時ならずと謂はんや。<l 位置
="P002A04" />
</s>

<s>此時に當りて大に智識を世界に求め、</s>

<s>我邦文明の真相を發揮して之を宇内に宏にせんこと、
</s>

<s>蓋し國民の任務たり。</s>

(1895年1号「太陽の発刊」大橋新太郎P002A03)

12

r（振り仮名）タグ

『太陽』誌面上に豊富に付けられている振り仮名は、『太陽』の本文の一部であり、電子テキストに反映させることは望まれよう。熟字訓や宛字、複数通りの読みが想定される場合などは、振り仮名がなければ正しく読めない漢字も多く、その有用性は大きい。一方3.4で見た通り、『太陽』における振り仮名の様相は、

記事が属する欄や年によって異なっており、雑誌の編集方針に左右される面も強い。著者の書いた言語や表記の問題ではないとも言えそうな部分もあり、そうした場合の振り仮名は冗長なデータに過ぎないと考えることもできる。また、単線的な配列を基本とする電子テキストを複線的な配列にする振り仮名を含めるのは、機械処理上煩雑でもある。そうした冗長さや煩雑さは、むしろ避けるべきだという考え方もある。

こうした長所・短所を踏まえると、振り仮名が有用な部分はその価値を生かし、冗長な部分は煩雑さを避けるような方式をとるのが現実的であろう。『太陽』の全巻を通して常に総ルビであるのは、小説・戯曲のジャンルの記事である。これらのジャンルの記事は、熟字訓や宛字が多く、振り仮名がなければ読めなかったり、誤読のおそれのある漢字表記が特に多い。また詩歌のジャンルに属する記事のなかにも、熟字訓や宛字表記の多いものがある。史伝、地理や家庭の記事など、パラルビによるものは、3.4に述べたように、振り仮名がなくても読める漢字列にルビが付されている場合が多い。

以上のことから、『太陽コーパス』全体に対する振り仮名の扱いの方針を次のように定めた。

- (1) 小説・戯曲・詩歌の類の記事については、原文にある振り仮名を入力する。
- (2) 小説・戯曲・詩歌の類以外の記事については、原文に振り仮名があっても入力しない。

小説・戯曲・詩歌の認定は、記事タグ中のジャンル属性の認定基準として用いたNDCによった。具体的には、NDCの第一区分で「9 文学」となっているもののうち、第2次区分に各国別の文学を立てる「91日本文学」から「98ロシア・ソヴィエト文学」までの範囲で、第3次区分が、1（詩歌など）、2（戯曲）、3（小説・物語）のいずれかの記事が、振り仮名を入力する対象となる。もちろん、これらのジャンル以外でも、振り仮名の有用性の高い漢字表記が用いられている場合もあるが、明示的な採否の規準は立て難いため、NDC番号のみの規準によって判別を行った。

r タグの書式は、次の通りである。

- (3) <r rt="やく">厄</r><r rt="かい">介</r>
- (4) <r rt="どう">奈何</r>

原則として(3)のように、漢字一字ごとに r タグを付し、「rt」

属性の属性値に振り仮名を書き入れていく形をとっているが、複数の漢字の連続にまとめて対応している熟字訓の振り仮名の場合は、(4)のようにその漢字連続に対してrタグを付して振り仮名を書き入れている。

熟字訓か否かの判別は、単漢字に分割すると漢字と語形の対応の根拠が失われるものを熟字訓と扱う原則のもとで行った。

(5) 熟字訓とするもの あす 明日 あした 明日 あさつて 明後日 アトモスフェア 雰圍氣
アフリカ 阿弗利加

(6) 熟字訓としないもの あくるひ 明日 あひて 對手 すごろく 双六 たぐる 手繰

こうして入力した振り仮名は、テキスト本文ではなくrタグの属性値となるため、検索のためには特別の方策をとることになる。その方法については、本書の「構造化テキストに対応した全文検索システム『ひまわり』」(山口昌也)の「4. 検索エンジン」(59頁から)および「5. 検索条件入力インターフェイス」(68頁から)、「構造化テキストを利用するアプリケーション『プリズム』と『たんぽぽ』」(小木曾智信)の「4. 検索用アプリケーション『たんぽぽ』」(103頁から)に説明があるので、参照してほしい。

13 外字タグ

13.1 包摂

3.5で述べたような『太陽』の字体の多様性と日本語文字集合の現状をふまえ、『太陽コーパス』として用いる文字集合と包摂規準を明確にした上で、電子テキストを作成することが必要である。『太陽コーパス』における漢字の実態と処理の方法については、別に「漢字の実態と処理の方法」(田中牧郎)として、本書(271頁から)に詳述したので、あわせて参照してほしい。

『太陽コーパス』では、JISによって入力を行い、JIS包摂規準に独自の規準を加えて包摂を行うことを基本方針とすることとした。JISのうち、一般的なコンピューターで普通に利用できるJISX0208 1997 (第1・第2水準)を用い、包摂規準は、JISX0213 2000・2004 (第3・第4水準)で追加された規準も加えたものを用い、さらに独自に規準を補足し、検索性を高める工夫を施した。

○JIS包摂規準によるもの

3.5で示した(1)から(4)の例のような、JISで包摂されるもの、すなわち、JISX0208 1997 (第1・第2水準)、JISX0213 2000・2004 (第3・第4水準)で示される包摂規準によって包摂できるものは、包摂する。上の例を、適用される包摂規準の連番とともに示すと表11の通りである。

表11 JIS包摂規準による包摂例

番号	『太陽』原文	JIS漢字	包摂規準の連番	備考
(1)	羽	羽	018	
(2)	青	青	146	
(3)	鷗	鷗	6.6.4	過去の規格との互換性を維持するための包摂規準
(4)	沿	沿	189	JIS X0213 2000で追加された包摂規準

このうち、(4)の包摂規準(連番189)はJISX0208 1997にはなかったが、JISX0213 2000で追加されたものである。また、(3)の包摂規準はJISX0208 1997で「過去の規格との互換性を維持するための包摂規準」とされたもので、JISX0213 2000では包摂規準から外され二つの字体は別字扱いになっているが、現状の一般的なコンピュータで「鷗」を用いるには、ある水準以上の知識が必要である。(3)のタイプの漢字の包摂規準は、JISX0208 1997によった。

○包摂規準を加えたもの

JIS包摂規準にはないが、『太陽コーパス』で独自に加えた規準によって包摂することにした字体も多い。例えば、3.5にあげた(5)「膈」の字体は、JIS漢字に掲げる「膈」と字形差は軽微である。そして次に示すように、「膈」字が用いられる用例と比較しても、読み・意味の両面から見て等価であると考えられる。

然而偶々肝膈を披きて赤誠を吐き (1895年5号「明治の心霊界」松村介石P029A20)

自ら其眞肝膈を吐露せしむるに (1909年16号「元老の現在及び未来」山路愛山P040A11)

また、「鴉」の字体は、芝野耕司編『増補改訂JIS漢字字典』(日本規格協会)において、「字形例」としてあげている康熙字典内府本のものにもっとも近く、次のように、「鴉」との間に、読

み・意味に相違は見られない。

初冬や人を嘲る群鴉（1909年4号「冬六十句」八界坊 P151C06）

群鴉騒ぎ鼓聲天に震ふの様は（1895年2号「初午」大橋乙羽 P154A15）

さらに、「𪛗」の字体は、冠の部分の内部が「ハ」のような形になり、「民」との間に空間がない。現代では見かけない字体であり、JISにはこれを包摂する規準はないが、下の例のように、「𪛗」と読み・意味は等しいと考えられる。

或は陷穽があり𪛗があり、（1925年10号「書翰礼讃（上）書翰の蒐集と研究並に大隈家書簡の涉獵」市島春城（談） P153A26）

養父の𪛗にかゝる身の（1895年7号「子煩悩」大橋乙羽 P110B15）

このように、読み・意味ともに等価であり、字体差が軽微なもので、JIS包摂規準に示す諸規準と比較しても、包摂とすることには問題がないと判断されるものは、独自に規準を立てて包摂した。

13.2 代用

読み・意味の一致する字体がJIS漢字内にあるが、JISの包摂規準に照らし、包摂の域を超える字体差があると認められる場合、包摂はできないが、「代用」の扱いによってJIS漢字を入力した。3.5に見た（6）から（9）の例などが、「代用」の処理としたものにあたる。用例によれば、下のように、それぞれ「捷」「煮」「憑」「杓」と、読み・意味が等しいことが確認できる。しかし、字体差はJIS包摂規準に照らして軽微なものとはいえない。

（6）敏捷なる天資（1909年11号「個人としての犬養木堂君」川尻琴湖 P205A01）

敏捷で、利に敏い（1917年14号「治作と米造」福永渙 P213A06）

（7）じりじりと煮え立つ蠟燭に（1909年6号「貸家一覧」泉鏡花 P110B19）

血は心臓に集りて、煮え立ち湧き立ち、（1895年12号「銀釵」渡部霞亭 P110A23）

（8）机に凭かゝつて（1901年14号「セバストウボルの落城」嵯峨の屋おむろ（訳）；トルストイ（作） P084B20）

机に悪懸りながら (1901年9号「セバストウポルの火花 (承前)」嵯峨の屋おむろ (訳) ; トルストイ (作) P096A17)

- (9) 柄杓にて三度汲みて (1895年1号「新年の礼式」有住斉PI53A10)

桶の水を柄杓で掬って (1901年3号「富籤」遅塚麗水P095B02)

「代用」とは、読み・意味の等価なJIS漢字で代用してテキスト化を行うものであり、包摂のように同じ漢字と認めるものではない。包摂と一線を画するために、原文での字体を別の文字集合から画像等によって参照させるしくみを留意した。具体的には、(6)の例であれば、当該の漢字に次のような外字タグを付与した。

敏<外字 文字番号="012445">捷</外字>なる天資

「文字番号」属性は、JISとは別の文字集合における番号を記入するためのものである。別の文字集合として用いたものは、文字鏡研究会による『今昔文字鏡』(注13)と、国立国語研究所の『太陽コーパス外字集』である。当該の外字が『今昔文字鏡』に収録されている場合、その番号を書き入れ、『今昔文字鏡』に収録されていない場合は、『太陽』原誌から取り込んだ文字画像を集めた『太陽コーパス外字集』の番号を書き入れた。後者の例は次のようになる。

<外字 文字番号="T003">巷</外字>

『太陽コーパス』の本文検索などにおいては代用したJIS漢字を用い、ブラウザで閲覧する際などには、この文字番号をもとに、『今昔文字鏡』もしくは『太陽コーパス外字集』の画像ファイルを、本文にはめ込んで閲覧できるようにした。

13.3 外字

代用は外字の一種だが、タグを無視すればJISでテキスト入力のできたものでもあり、外字画像として扱うかテキストとして扱うかは利用目的に応じて変えることができる。一方、『太陽』の外字の中には、代用できる字体がJIS漢字の中にない場合も多く、この場合はテキスト化は不可能で、まさしく外字となる。このような外字は「𪛗」を外字タグで囲む形で表すことにした。例えば3.5にあげた(10)「簞」の場合は、『今昔文字鏡』にもなく、次のようになる。

<外字 文字番号=" T108">■</外字>

また、『今昔文字鏡』に登録されている場合は、次の「磚」の例のようになる。

<外字 文字番号="024597">■</外字>

代用も含んだ広い意味での「外字」を、『太陽コーパス』利用時に閲覧したり一覧を作成したりする方法については、本書の「構造化テキストを直接利用するアプリケーションー『プリズム』と『たんぽぽ』ー」（小木曾智信、83頁から）を参照してほしい。

14 注タグ

14.1 注タグの種類

3.6で見た、誤植・誤用・通用に関しては、一定の方針で校訂を行うことにした。校訂を行った場合、正しい形に修正したテキストに注タグを付し、原文の形を「原文」属性に記入した。明らかな誤植・誤用のほか、誤用とは言い切れないが、表記規範が確立していないことなどのために表記のゆれが著しく、原文のままテキスト化を行うと、検索性を損ねると考えられるものも校訂の対象とした。具体的には、濁点無表記、歴史的仮名遣いに反する仮名遣い、漢語の漢字表記で現在の規範と異なるものについては、現代語の規範を背景にして検索文字列を指定する利用者にとって、原文の表記法を想定することが難しい場合が多いと考えられるため、できるだけ校訂の対象にした。

注タグをつけた場合には、以下に示すA～Hの8種類があり、それぞれを「分類」属性に記入した。A～Gについては、その例を図2～図9に原文画像で示し、各図の該当する箇所のXMLテキストの部分で、注タグを付けた形でそのまま抜き書きすると、下の通りである。

A誤字通用（図2・図3）

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。

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

國民も亦貧窮に陥あり、

必すしも細語を要せず、

渡航した御朱印船なものは、慶長の九年より元和の三年迄

之が下に立ち這つくばはざるを得ず、學者此有様を稱して、優劣
勝敗劣弱肉強食の数とはいふなり、されば氣の利た雀は鷹

嵐山にものして

從二位 東久世通禧

わが爲になくかと思ふ年をへてどひし嵐のやまほゝきす

字會も亦意見の衝突ありて、新舊の二派となり。而して最後に
に皆諸共に泣き寝入りとなりぬ。主唱者や熱心ならざりし、會

隨分御氣嫌宜しくとか

近頃ノースロップ氏より直接概略の話を聞

図2

図3

図4

図5

図6

図7

図8

図9

隨分御<注 原文="氣嫌" 分類="A誤字通用">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)

B衍字 (図4)

なくてもよい文字が紛れ込んだと思われるもの。

而して最後に<l 位置="P035A20"/><注 原文="に" 分類="B
衍字"/>皆諸共に泣き寝入りとなりぬ。(1895年7号「欧州諸
国に於ける綴字改良論」上田万年P035A20)

C脱字 (図5)

あるべき文字が脱落していると考えられるもの。

<注 原文="やまほゝきす" 分類="F濁点脱落">やまほ<注 分類
="C脱字"></注>ゝぎす</注> (1895年7号〈短歌〉
P173A05)

D転倒 (図6)

二つの文字の配列順が、転倒していると考えられるもの。

優<l 位置="P180A05"/>勝<注 分類="D転倒">劣敗</注>弱肉
強食の

(1895年8号「強者と弱い者」関旭巖P180A06)

E 欠損 (図7)

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。

御朱印舩な<注 分類="E欠損">る</注>ものは、(1895年10号「海国日本に於る海事思想」久松義典P052A25)

F 濁点脱落 (図8)

濁音表記が期待される箇所に濁音表記がないもの。

<注 原文="必ず" 分類="F濁点脱落">必ず</注>しも細語を要せず、(1895年7号「北海道拓殖論」近衛篤磨P001E10)

G 仮名遣 (図9)

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。

國民も亦貧窮に<注 原文="陥り" 分類="G仮名遣">陥り</注>、(1895年7号「朝鮮問題」川崎三郎P008A24)

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。コーパス作成者の判断による修正ではないことを示すために、「H正誤表」の分類枠を立て、別扱いとした。

自から奉ずる所は<注 原文="八百金" 分類="H正誤表">八圓金</注>を超えず、(1895年3号「教事些語(下)」巖本善治P174B29)

なお、注タグで囲む単位は、原則として単語とし、振り仮名の部分を修正して注記を付与した場合には、振り仮名をつける文字列の単位(通常は単字、熟字訓等の場合は複数文字列)としている。これらの注記を一覧する方法は、本書に掲載した「構造化テキストを直接利用するアプリケーションー『プリズム』と『たんぽぽ』」(小木曾智信)の、「3. 形式変換と情報抽出」(84頁から)を参照してほしい。

15

踊り字タグ

「踊り字」と呼ばれる繰り返しの符号の用い方に関して、『太陽』は、現代に比べてかなり複雑な用法をもっている。現代の表記法

と異なる種類の踊り字の場合、それをどうテキスト化するかということ、踊り字によって埋もれてしまう語形をどのようにして検索できるようにするかということの二つが問題になる。

まず、『太陽』に見られる踊り字の種類には、次の4種があり、(3)「二字点」と(4)「くの字点」とが、JISコードで表現できないものである。



図10

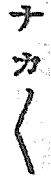


図11

- (1) 一字点 > ズ、ヅ
- (2) 同の字点 々
- (3) 二字点 (図10)
- (4) くの字点 (図11)

『太陽』でも、(1) 一字点、(2) 同の字点は、現代と同様、通常は(1) 仮名の繰り返し、(2) 漢字の繰り返しに、それぞれ用いられる。

(3) 二字点は、『太陽』では仮名の繰り返しにも漢字の繰り返しにも用いられるが、仮名の繰り返しとして用いられている場合は(1) 一字点(> ズ、ヅ)で、漢字の繰り返しで用いられている場合は(2) 同の字点(々)で、それぞれ代用することとした。そして、原文は二字点であることを、「種類」属性を立てて書き入れた。図10の例の場合、次のようなタグ記述になる。

各<踊り字 種類="二字点">々</踊り字>

(4) くの字点は、二字以上の繰り返しに用いられ、(1) 一字点または(2) 同の字点で代用することは難しい。そこで便宜的に「～」(清音表記の場合)、「～°」(濁音表記の場合)という記号で代用することとし、図11の例の場合であれば、次のように記した。

ナカ<踊り字 種類="くの字点">～～</踊り字>

踊り字によって埋もれてしまう語形を検索できるようにするために、踊り字タグに、「値」属性を立て、必要に応じて踊り字を開いた表記を書き入れた。(1) 一字点や(2) 同の字点が現代

の標準的な用法で用いられている例（「ア、」「人々」など）については、検索者が検索文字列を指定する際に、指定もれが起こる可能性は低いので「値」属性は設定しない。しかし、下に示したような場合は、現代の標準的な用法とは異なっており、電子テキストの側で対応をしておくことが望まれる。（３）二字点についても、代用した「々」の通常の用法から外れる二字以上の連続で用いられる場合は、同様の対応が必要である。また、（４）くの字点も、何字分繰り返されるかは場合によりさまざまであるので、同様の対応が望まれる。「値」属性を記述した具体例は、次の通りである。

おほ<踊字 値="ほほほ">>>></踊字>

神社<踊字 値="神社">々々</踊字>

一日<踊字 種類="二字点" 値="一日">々々</踊字>

ナカ<踊字 種類="くの字点" 値="ナカ">～～</踊字>

16 その他のタグ

伊達蜂須賀諸氏と私に婚を結び、
たてはちすかしよし
此事は前號加藤清正の
 事蹟中に記述し置けり

図12

大詔已に下ると雖ども、
 至尊尙ほ聖慮を

図13

圖るとを要す

図14

グスタフ、クーツ
てんきくわいしゃ
 電氣會社の

図15

$x^2 - a^2,$

図16

CaCl_2

図17

16.1 割書タグ

これまで述べてきたタグ以外のタグについて、最後にまとめて説明する。まず、本文に注釈などを付す場合で、行中に一行を二行に割って記す、「割書」と呼ばれる表記法がある（図12）。この場合、「割書」のタグで囲み、割書内の改行位置に<割書改行/>の空要素タグを入れた。

伊達蜂須賀諸氏と私に婚を結び、</s><s><割書>此事は前號
加藤清正の<割書改行/>事蹟中に記述し置けり</割書></s>
(1895年12号「石田三成（承前）」小倉秀貫P073A20)

なお、角書き、スペース節約等のための二行書等は、割書に認定しない。

16.2 敬意欠字タグ

天皇などに敬意を表する意図で、天皇関連の語の直前を、一字または二字分空白にする表記法が行われる場合がある（図13）。「欠字」と言われることのある書法だが、印刷不鮮明等による欠字の場合と区別して、「敬意欠字」と呼ぶことにし、<敬意欠字></敬意欠字>のように、「_」をタグで囲む方式で記入した。

下ると雖ども、<敬意欠字>_</敬意欠字>至尊尚ほ（1895年
9号「台湾論」島田三郎P002B07）

16.3 合字タグ

ある種の語を表す文字連続で、二つ以上の文字を一字にあわせた「合字」と呼ばれる表記法が行われる場合があり（図14）、次のようにタグ付けした。

圖る<合字>こと</合字>を要す（1895年9号「戦後財政案」
河島醇 P006B23）

合字表記されることのある語には、「こと」「まいらせ候」「ヨリ」「トモ」などがある。

16.4 小書タグ

図15の「フ」のような、カタカナ表記で小書きされるもののうち、JISにないものにこのタグを付している。

グスタ<小書>フ</小書>へ、クーツ電気會社の（1901年7号
「工業世界」金子篤寿P169A02）

小書タグが付けられることのある片仮名は、次のものである。

キ コ ネ ハ フ ヘ ホ ム ル ワ キ エ ヲ

16.5 上付タグ、下付タグ

数式や化学式など、上付（図16）または下付（図17）の表記法をとる場合があるが、次のようにタグ付けした。

$x^{<\text{上付}>2^{</\text{上付}>-a^{<\text{上付}>2^{</\text{上付}>}$ （1901年7号「実業教育に於ける数学」数藤斧三郎P199B08）

$\text{CaCl}^{<\text{下付}>2^{</\text{下付}>}$ （1895年8号「石灰岩は如何にして生ぜしや」佐藤伝蔵P145A17）

16.6 非入力対象タグ

記事内にある図表や漢文・欧文の長文にわたる引用などはコーパスの対象外としているが、本文から当該部分が削除されることで、本文の読み取りに支障をきたす場合もある。図表や漢文・欧文自体はコーパスとして不要であっても、その位置に図表等のあったことは記されている方が都合がよい。非入力対象タグを空要素タグとして立て、「種類」属性には、図表・漢文・欧文等の種類を、「表見出し」属性には、原文で付けられている図表等の名称を、それぞれ記入した。

$<\text{非入力対象 種類}=" \text{《図表》} " \text{表見出し}=" \text{獨逸國（千八百八十二年調査）} ">/>$ （1895年1号「経済的闘争」井上辰九郎P035A24）

17

おわりに

本論文ではまず、雑誌『太陽』が現代語の確立期を代表する資料として価値が高いことを確認した上で、その本文の様態を観察して、資料の価値をコーパスとしてどのようにして生かすべきかを考察した。それをふまえてどのように『太陽コーパス』を設計し仕様を策定したかについて、全体的な記述を行った。こうして作成されたコーパスを使ってどのような研究成果が生まれるかによって『太陽コーパス』の価値は評価されるべきものである。本書における個々の論文が扱う個別的具体的な問題についての研究成果は、その判断材料になるであろうし、これらの論文を機縁としてさまざまな観点から『太陽コーパス』が活用されることで、コーパスについての議論が盛んになることを期待したい。

活用されるうちに『太陽コーパス』の設計や仕様自体に改善すべき点は多々見つかるであろうし、このコーパスでは研究できない限界も明らかになるであろう。雑誌『太陽』以外の資料をもとに構造化テキストタグ付きコーパスを作成する場合には、本論文で述べた設計や仕様では不十分なところも多いと思われる。『太陽コーパス』を踏み台として、コーパスの設計や作成が広い範囲で盛んになることも望みたい。また、われわれ『太陽コーパス』の作成に携わった者も、このコーパスの保守・普及と活用を図りながら、次の段階のコーパスを作成するための歩みを進めていきたい。

注

- (1) 本論文に述べる内容の一部は、田中・小木曾（2000）、田中（2004a）などで述べたことがあるが、『太陽コーパス』の設計に関わる基本的ことがらであるので、本論文に取り込んだ。
- (2) 湯浅（2000）は、明治期の資料が膨大で多様であるため、かえって当時の日本語の全体像を把握することが困難になっていることを指摘している。
- (3) 明治期の言語資料を分類して示した飛田（1973）では、新聞・雑誌を「総合資料」と呼んで、他の種類の資料の「性格を総合した資料」という位置づけを与えている。
- (4) 『太陽』の書誌、社会的・歴史的位置については、鈴木正節（1979）、鈴木貞美編（2001）に詳細な研究があり、言語資料としての『太陽』の位置づけを考える上でも参考になる。『太陽』の復刻版と総目次は、日本近代文学館編（1999）にあるが、原誌を所蔵する大学図書館や公共図書館も多い。『太陽コーパス』作成においては、国立国語研究所蔵の原誌を底本とした。
- (5) 『太陽』における口語文の増加のありようについては、『太陽コーパス』作成の過程で行った二つの予備的な研究でも扱った。小椋・小木曾・近藤（2002）は、『太陽コーパス』全体を概観し、田中（2004b）は創刊年（1895年）『太陽』における詳細な状況を報告した。これらにおいても『太陽』が口語文の普及に関する調査研究に有効であることが示されている。
- (6) 日本語学の領域で『太陽』を資料とした先行研究には、見坊（1957）、土屋（1966）（1967）（2004）、深澤（2001）（2003）などがあるが、ここで述べるような『太陽』の言語

- 資料としての特徴を十分に生かした研究は未開拓である。
- (7) 本書の「研究の目的と本書の構成」(田中牧郎, iii頁)でも記した, 国立国語研究所の「日本大語誌」構想において, 現代を3期に分け, その第2期の始点を1901年とした。『太陽コーパス』の対象年を決める起点を1901年としたのは直接的にはここに始まる。
- (8) 『太陽コーパス』に着手した当初は『太陽』での完結性に配慮し, 『太陽』終刊年の1928年(2号で終刊)の2冊も対象に加えていたが(田中・小木曾2000参照), 年次間隔も冊数も他とバランスを欠くため, 最終的には対象外とした。
- (9) 著作権処理ができなかった場合には, ①著者の没年が未詳であったり著作権者の連絡先が不明であったりして, 著作権処理のための情報収集ができなかったもの, ②コーパス収録について著作権者に許諾を依頼したが許諾が得られなかったもの, の二通りがある。未処理に終わった著者は, ①の場合が大部分を占め, 依頼できた場合の多くの著作権者からは許諾が得られた(依頼104人, 許諾97人, 不許諾7人, 許諾率93.3%)。①の未詳者についての調査方法については課題を残した。著作権処理の方策は継続して検討し, 次の公刊の機会にはコーパスに含められる記事を増やしていけるようにしたい。
- (10) 電子テキストの構造化記述に用いられるマークアップ言語の歴史についてまとめた近藤(2003)は, 今後の日本語コーパスのマークアップ方式はXMLが主流になるだろうと見通している。XMLが登場する以前に, 文献資料の特徴に応じた構造化とタグ付けの方法について, 安永(1998)が国文学の見地から詳細に研究している。この研究は, 日本語学の見地から研究資料を構造化テキストにする方法を考える際にも参考になるところが多い。
- (11) NDCの詳細は, 日本図書館協会分類委員会編(1995)を参照。
- (12) 連載記事で文語と口語の混在の度合いが, 号によって変化するものが若干あるが, その場合は連載全体で基調をなす文体を属性値に記入した。
- (13) 『今昔文字鏡』は, 『大漢和辞典』の全文字5万字強を含んだ10万字にのぼる文字を収めた文字集合であり, 漢字を多く扱う資料の電子的流通や印刷等にある程度普及している(<http://www.mojikyo.org/>)。

参考文献

- 小椋秀樹・小木曾智信・近藤明日子（2002）『『太陽コーパス』
を使った近代語表現の通時的研究—口語文体・可能表現・待遇
表現について—』（『国語学会2002年度春季大会要旨集』，
177-184頁，国語学会）
- 見坊豪紀（1957）「明治時代の文語文」（『言語生活』74，32-
41頁，筑摩書房）
- 国立国語研究所（1987）『雑誌用語の変遷』（国立国語研究所報
告87，秀英出版）
- 国立国語研究所国語辞典編集準備室（1983）『用例採集のための
主要雑誌目録』（国語辞典編集準備資料3，国立国語研究所）
- 近藤泰弘（2003）「古典語のコーパス」（『日本語学』22-5，
62-81頁，明治書院）
- 鈴木貞美（2001）「明治期『太陽』の沿革，および位置」（鈴木
貞美編『雑誌「太陽」と国民文化の形成』3-39頁，思文閣出
版）
- 鈴木貞美編（2001）『雑誌「太陽」と国民文化の形成』（思文閣
出版）
- 鈴木正節（1979）『博文館『太陽』の研究』（アジア経済出版会）
- 田中牧郎（2004a）「近代語研究資料と研究 雑誌『太陽』（『日
本語学』23-12，208-220頁，明治書院）
- 田中牧郎（2004b）「雑誌『太陽』創刊年（1895年）における
口語文—敬体を中心に—」（飛田良文編『国語論究11言文一致
運動』78-107頁，明治書院）
- 田中牧郎・小木曾智信（2000）「総合雑誌『太陽』の本文の様
態と電子化テキスト」（国立国語研究所編『日本語科学』8，
141-152頁，国書刊行会）
- 土屋信一（1966）「雑誌『太陽』（明治28-昭和3）に見る表記の
変遷」（『言語生活』182，36-42頁，筑摩書房）
- 土屋信一（1967）「雑誌『太陽』の用字の変遷」（『言語生活』
193，34-43・87頁，筑摩書房）
- 土屋信一（2004）「雑誌記事の言文一致—明治二十八年『太陽』
の場合—」（飛田良文編『国語論究11言文一致運動』321-348
頁，明治書院）
- 永嶺重敏（1997）『雑誌と読者の近代』（日本エディタースクエ
ル出版部）
- 日本近代文学館編（1999）『CD-ROM版近代文学館⑥「太陽」』

-
- (八木書店)
- 日本図書館協会分類委員会編 (1995)『日本十進分類法 新訂9版』
(社団法人日本図書館協会)
- 飛田良文 (1973)「近代語研究の資料」(『文学・語学』66, 45-60頁, 三省堂)
- 深澤愛 (2001)「雑誌『太陽』創刊年における外国地名片仮名表記」(『国語文字史の研究』6, 187-218頁, 和泉書院)
- 深澤愛 (2003)「漢字片仮名交じり文中における表記の選択—博文館『太陽』における外国地名とカタカナ表記—」(国立国語研究所編『日本語科学』14, 29-52頁, 国書刊行会)
- 安永尚志 (1998)『国文学研究とコンピュータ』(勉誠社)
- 湯浅茂雄 (2000)「近代語研究の要点と課題」(『日本語学』19-11, 138-148頁, 明治書院)

構造化テキストに対応した 全文検索システム『ひまわり』——山口 昌也

1

はじめに

『太陽コーパス』のように大規模な構造化テキストを研究に利用するには、構造化テキストから文字列を高速に全文検索したり、付与されている情報を抽出するツールが必要である。

本稿で説明する『ひまわり』は、言語研究用に設計された全文検索システムである(図1)。本来、『ひまわり』は、『太陽コーパス』にだけでなく、広範な構造化テキストを検索対象とすることができるが、ここでは、『太陽コーパス』に適用する場合を中心に述べる。

『ひまわり』の主な機能を次に示す。

- ・構造化されたテキストに対する高速な全文検索機能
 - 索引(indexing)方式に Suffix Array 方式(山下 2000)を採用して検索の高速化を実現
 - 前後文脈や著者、ジャンルなどXMLの要素属性を、検索条件として指定することが可能
- ・言語研究支援機能
 - KWIC 形式での検索結果の表示
 - 書誌情報、引用情報、著者情報など、『太陽コーパス』にタグ付けされた、さまざまな言語情報の閲覧(図1を参照)

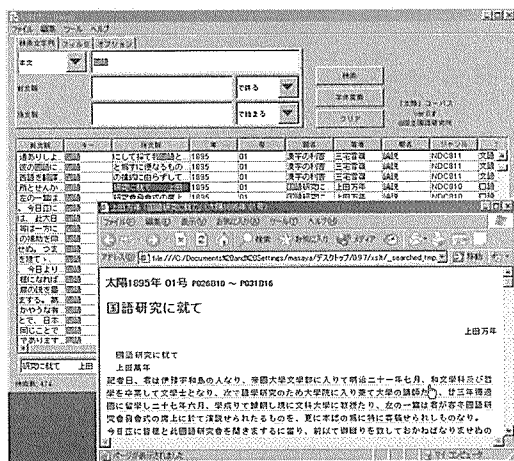


図1 全文検索システム『ひまわり』

2 『ひまわり』の設計

2.1 想定する利用形態

『ひまわり』の設計を行う前に、想定する利用形態を定めておく。ここでは、

- ・どのような利用者が
- ・どのような資料を
- ・どのように全文検索するのか

という観点から考えていく。

2.1.1 利用者と計算機環境

『ひまわり』が想定する利用者は、主として言語研究者である。より具体的に言うと、構造化テキストに付与された言語情報を閲覧し、分析する利用者を想定する。また、利用者は構造化テキスト自体に対する専門知識を持たないものとする。

一方、利用者の計算機環境については、より多くの計算機環境で動作することが理想的である。ただし、任意の環境で動作させることは、難しい。そこで、上で述べた利用者に対する想定から、最も多くの利用者が存在する Windows で動作することを必要条

件として設定する。

2.1.2 検索対象の資料

『ひまわり』は、『太陽コーパス』のようにさまざまな言語情報が付与された構造化電子テキストを検索対象として想定する。構造化の形式は、XML とする。

対象とする構造化テキストの規模は、150MB 程度を想定する。これは、現時点で 約100MB の規模を持つ『太陽コーパス』を格納するのに十分な量である。なお、個人で入手可能な電子テキストのうち、規模の大きいと思われる毎日新聞 CD-ROM の記事本文の規模は、1 年分で 90～150MB である。

次に、構造化テキストと検索システムの設置形態について述べる。構造化テキスト、および、検索システムの設置形態としては、(1) 両者とも利用者の計算機に置く、(2) 検索用サーバを構築し、利用者はクライアントとして検索用サーバに接続して検索を行う、などが考えられるが、『ひまわり』では (1) を採用する。この設置形態は、『太陽コーパス』を CD-ROM などで配布することも考慮しているが、言語研究者自身がコーパスを作成して、それを『ひまわり』で検索する、という利用形態も勘案している。

2.1.3 全文検索の利用方法

2.1.1 の想定する利用者のところでも述べたとおり、全文検索の利用方法としては、構造化テキストに付与された言語情報をもとに、さまざまな角度から閲覧を繰り返し、分析を進めていくということを想定する。例えば、分析対象の語が、どの年代に多く出現するか、また、用法、出現する文脈はどのようなものかを把握するために、さまざまな条件で検索を繰り返し、分析に生かすという使い方である。

2.2 設計方針

2.1 節で示した利用形態を考慮して、次のような設計方針を立てた。

- ・高速、かつ、言語情報を検索条件として利用可能な検索
- ・容易な検索条件の指定
- ・資料に適した言語情報の表示

2.2.1 高速、かつ、言語情報を検索条件として利用可能な検索

コーパスを利用する場合、すばやく目的の文字列が検索され、思考の流れを妨げないことが望ましい。特に、言語研究者は、試行錯誤をしつつ、さまざまな条件で検索することが考えられるために、検索の高速性は重要になる。

また、検索の高速性を保持しつつ、構造化テキストに付与されている言語情報、例えば、著者や文体情報を検索条件として利用することも必要である。

これらのことを実現するために、まず、既存の各種検索システムを調査し、既存のシステムを『ひまわり』の検索エンジンとして設計に組み込めないかを探った。ここでは、大きく分けて、次の3タイプの検索システムを調査した。

- XML データベース

『太陽』の構造化テキストは、XML で記述されており、XML データベースとの親和性はよい。さらに、XML の要素や要素属性を考慮しつつ、全文検索をすることもできる製品も存在する（例：NetCore 社の NeoCore XMS (<http://www.neocore.jp/>)、（株）メディアフュージョン社の EsTerra XSS (<http://www.mediafusion.co.jp/>) など）。しかし、現状では、XML データベースは未だ発展段階のシステムであり、高価でもある（例えば、EsTerra XSS は、2004-10-25 現在、Windows 版が250万円である）。したがって、一般の利用者が気軽に利用できるとはいいがたい。

- リレーショナルデータベース

リレーショナルデータベースは、XML データベースと異なり、確立された技術となっており、信頼性の高いシステムが利用できる。また、無料で利用できるフリーソフトウェアのシステム（MySQL や PostgreSQL など）も存在する。しかし、データベースに格納されるデータは、表形式で管理されており、XML のように記述形式の自由度が高いテキストに対して、高速に全文検索を行うには適していない。

- 文字列検索プログラム、ライブラリ

grep などに代表される文字列検索プログラムは、構造化されたテキストを検索対象とすることができるものを含めて、数多く

開発されている。例えば、XML 文書を検索できるライブラリとしては、Xerces (<http://xml.apache.org/>) が挙げられる。ただし、索引づけ (indexing) を行っていないため、高速な検索は難しい。

一方、sufary (<http://cactus.aist-nara.ac.jp/lab/nlt/ss/>) や sary (<http://sary.namazu.org/>) のように高速な全文検索を目的として設計されたライブラリも存在する。これらのライブラリは、Suffix Array を使って索引づけを行い、高速な検索を実現している。ただし、XML の要素や属性を考慮して検索するなど、構造化されたテキストの検索を行うには、さらなる拡張が必要である。また、Unix、Linux でしか動作せず、動作環境が限定される。

以上のことから、既存の検索システムをそのまま組み込んだシステムでは、『ひまわり』が想定する利用形態に適合させることや高速、かつ、言語情報を検索条件として利用可能な検索を実現するのは困難であると考える。

そこで、『ひまわり』では、Suffix Array 方式で索引づけを行い、かつ、XML 文書を検索できる検索エンジンを新たに実装する。Suffix Array 方式を採用したのは、(1)アルゴリズムが単純であることから、構造化テキストへ対応するための拡張やプログラムの実装が容易であること、(2)sufary などで高速に検索できることが実証されていることによる。

2.2.2 容易な検索条件の指定

2.1.1 で示したように、『ひまわり』では、構造化テキスト自体に対する知識を持たない利用者を想定している。このような利用者が『太陽コーパス』に付与されている情報を利用して検索を行うためには、容易に検索条件を指定できる仕組みが必要である。『ひまわり』では、GUI (Graphical User Interface : 視覚的なユーザインターフェイス) を採用して、容易に検索条件を指定できるようにする。

また、『太陽コーパス』のように明治、大正期の資料には、現在では通常用いられない字体が使用されている。したがって、検索文字列を指定する際には、字体の知識が必須である。『ひまわり』では、利用者が入力した字体を『太陽コーパス』で使われている字体に変換する機能を持たせることにより、利用者が容易に

検索文字列を指定できるようにする。

2.2.3 資料に適した言語情報の表示

『太陽コーパス』には、さまざまな言語情報が付与されている。それらの言語情報を言語研究に役立つよう、適切、かつ、効率的に表示することが必要である。そこで、『ひまわり』では、次のことを実現することとした。

- ・ KWIC (KeyWord In Context) 形式での表示

KWIC は、従来から言語研究で利用されてきた方式であり、KWIC を利用してさまざまな分析ができることがわかっている。『ひまわり』では、KWIC に加えて、検索対象語に付与されている言語情報（例：著者や記事名など）を併せて表示できるようにする。

- ・ 記事全体での表示

KWIC での表示は、表示範囲が限られるため、フォントサイズなどの表示スタイルや、付与されている言語情報を資料に即して表示することは難しい。特に、段落や文中での位置といった情報は、記事全体を閲覧できることが好ましい。そこで、『ひまわり』では、KWIC で表示される文脈だけでなく、記事全体を閲覧できるようにする。

3

『ひまわり』の機能と構成

本節では、『ひまわり』の機能とシステム構成について解説する。すでに述べたように、『ひまわり』は『太陽コーパス』以外の構造化テキストも検索対象とすることができるが、ここでは、『太陽コーパス』への適用例という形で『ひまわり』の機能と構成を説明することにする。

3.1 『ひまわり』の機能

『太陽コーパス』に同梱される『ひまわり』の機能を次に列挙する。

- ・ 全文検索機能

- －検索対象
 - + 本文（全要素の要素内容）
 - + 「r」要素の rt 属性（ルビ自体）
- －絞込み条件
 - + 検索対象の文字列に対する前文脈，後文脈
 - + 「雑誌」要素の「年」，「号」
 - + 「記事」要素の「題名」，「著者」，「欄名」，「文体」，「ジャンル」属性
 - + 「l」要素の「位置」属性
 - + 「引用」要素の「種別」，「話者」属性
 - + 「注」要素の「原文」属性
- ・文字列照合条件
 - －検索対象の文字列の場合：完全一致
 - －絞込み条件の場合：完全一致，前方一致，後方一致，正規表現
- ・一覧表示機能
 - －雑誌一覧表示：コーパスに登録されている雑誌一覧を表示する。
 - －記事一覧表示：コーパスに登録されている記事一覧を表示する。
 - －著者一覧表示：コーパスに登録されている記事の著者一覧を表示する。
- ・検索結果閲覧機能
 - －KWIC 形式閲覧機能：検索対象文字列を KWIC 形式で表示する。
 - + 並び替え機能：検索結果を並べ替える。
 - + 絞込み機能：検索した結果を絞り込む。
 - + 著者検索：検索結果の著者データを検索する。
 - + 頻度計測機能：年ごとなどの一定の基準で，検索対象文字列の出現頻度を計測して表示する。
 - －記事閲覧機能：記事全体を整形し，Web ブラウザで表示する。
- ・字体変換機能：検索文字列に対する異体字の候補を提示する。
- ・その他の機能
 - －検索結果のテキストファイルへの出力
 - －検索結果のクリップボードへのコピー

3.2 『ひまわり』のシステム構成

3.2.1 システム構成図

図2に『ひまわり』のシステム構成図を示す。図中の実線はデータの流れを、点線は、データの参照を表す。

システムは、大きく分けて、次の四つのモジュールからなる。個々のモジュールの説明は、この後の章で順次詳細に解説する。

- ・検索条件入力インターフェイス

利用者が検索条件を入力するのを支援する。GUI (Graphical User Interface : 視覚的なユーザインターフェイス) を持ち、『太陽』構造化テキストの構造や付与されている言語情報に対する知識がなくても検索条件を指定できるようになっている。具体的には、『太陽』構造化テキスト中のどの部分を検索対象とするか、どの付与情報を検索時の制約条件として用いるのか、ということメニュー形式で選択することができる。

また、検索条件入力インターフェイスは、異体字の入力支援用に、字体変換辞書(後述)に基づいた字体変換機能を持つ。

- ・検索エンジン

検索条件入力インターフェイスで指定された検索条件でコーパスファイルを検索し、検索結果を検索結果ブラウザに出力する。検索の高速化のため、Suffix Array方式の索引ファイルを利用している。なお、コーパスファイルは、『太陽』構造化テキストに含まれるすべての「雑誌」要素を一つのファイルに結合したものである。

- ・検索結果ブラウザ (KWIC 形式)

検索エンジンで検索された結果を KWIC 形式で表示する。また、KWIC に基づいた並べ替え、検索結果の絞込みなど、分析を支援する役割も持つ。KWIC 形式の表示では、KWIC だけでなく、検索文字列に対する記事のタイトル、著者、文体情報など、『太陽』構造化テキストに付与されている各種の言語情報を表示する。言語情報のうち、著者情報は、著者情報ファイル(後述)から取得する。検索結果ブラウザに表示される結果は、テキストファイル、および、クリップボードを介して、外部プログラムに

渡すことができる。

・検索結果ブラウザ（記事閲覧）

検索結果の記事全体を閲覧する。記事閲覧用の検索結果ブラウザ自体は実装せず、一般的な Web ブラウザで代用する。『ひまわり』側は、検索エンジンにより閲覧対象の記事をコーパスから抽出し、Web ブラウザに渡す。その際、XSL スタイルシート、および、CSS ファイルを指定する。XSL スタイルシートは HTML 形式への変換方法を、また、CSS ファイルは、Web ブラウザでの表示形式を規定するものである。

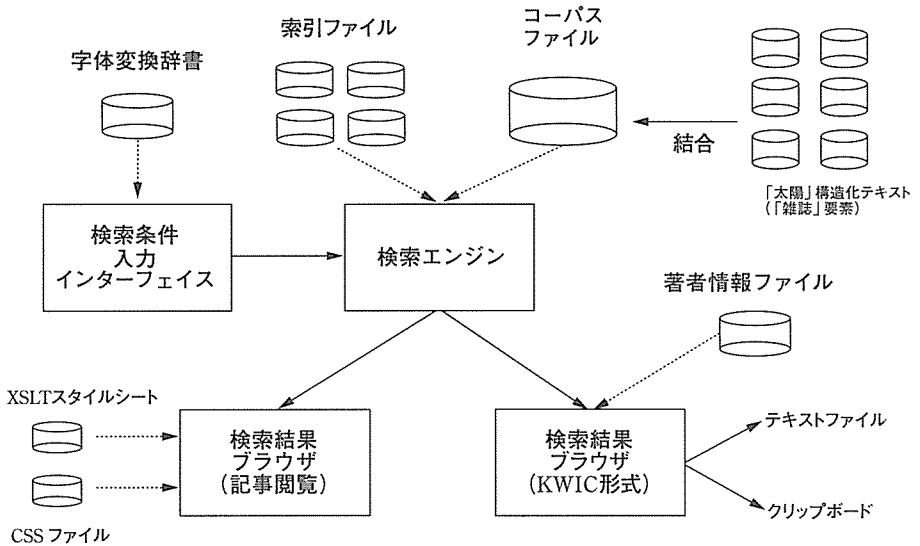


図2 『ひまわり』のシステム構成図

3.2.2 『ひまわり』関連ファイル

『ひまわり』に関連するファイルを次に示す。これらは、『太陽コーパス』CD-ROM の himwari フォルダに収録されているファイルである。詳細は、関連する節を参照のこと。

・コーパスファイル[corpus.xml]（4.1 節参照）

ー形式：XML スキーマ <http://www.kokken.go.jp/taiyo/corpus.xsd>、および、zassi.xsd で検証済みの XML 文書であること。ただし、実体参照は不可。

- －文字符号化方式：UTF-16（Little Endian, Byte Order Mark 付き）
- －改行コード：LF
- ・要素内容へのインデックス[corpus.cix]（4.2.2 節参照）
- ・要素へのインデックス[*eix]（4.2.1 節参照）
 - －corpus.zassi.eix ... 「雑誌」要素
 - －corpus.article.eix ... 「記事」要素
 - －corpus.ref.eix ... 「l」要素
 - －corpus.quote.eix ... 「引用」要素
 - －corpus.tyu.eix ... 「注」要素
- ・要素属性へのインデックス[*aix]（4.2.1 節参照）
 - －corpus.r.rt.aix ... 「r」要素の rt 属性
- ・著者情報ファイル[authors.xml]（6.1.5 節参照）
 - －形式：authordb.dtd で検証済みの XML 文書であること
- ・字体変換辞書[jitaidic.xml]（5.3 節参照）
- ・記事閲覧用 XSL スタイルシート [xsl/*xsl]（6.2 節参照）
 - －zassiHL.xsl（横書き・行番号表示）
 - －zassiHP.xsl（横書き・段落表示）
 - －zassiVL.xsl（縦書き・行番号表示）
 - －zassiVP.xsl（縦書き・段落表示）
 - －zassi_common.xsl（上記 xsl ファイルの共用部分）
- ・記事閲覧用 CSS ファイル [xslt/zassi.css]（6.2 節参照）
- ・記事閲覧用テンポラリファイル[xslt/__searched_tmp.xml]（4.4.1 節参照）

3.3 システムの実装とシステム要件

『ひまわり』は、Java 言語で記述した。Java 言語を選択したのは、Java 言語が基本的にプラットフォーム非依存の言語であり、Java 言語で記述したプログラムが、多くの OS で動作するためである。

『ひまわり』のシステム要件は、次のとおりである。

OS：

Java Runtime Environment (JRE) 1.4.1_05, もしくは、
1.4.2_04 が動作する環境

メモリ：

512MB 以上を推奨

ハードディスク：

最低 500MB以上の残り容量を推奨

なお、『ひまわり』は、Debian GNU/Linux 3.0r2 (kernel 2.4.20) 上で開発した。動作は、次の環境において確認している。

- ・ JRE 1.4.2_04
 - －Windows XP Professional
 - －Windows 2000 (SP2)
- ・ JRE 1.4.1_05
 - －Debian GNU/Linux 3.0r2 (kernel 2.4.20)
 - －Windows XP Professional
 - －Windows 2000 (SP2)

4 検索エンジン

本節では、『ひまわり』の検索エンジンについて述べる。検索エンジンは、『ひまわり』の中で、全文検索、記事閲覧時における記事検索などの検索処理一般を担当する。

この後の節では、まず、検索エンジンが想定するコーパスの形式を示す。次に、検索エンジンを構成する三つのモジュールの詳細について解説する。その後、これらのモジュールを組み合わせ、『ひまわり』における検索機能を実現する方法を説明する。

4.1 コーパスファイルの形式

『ひまわり』の全文検索エンジンは、検証済み XML 文書 (valid XML Document) を検索対象とする。文字符号化方式は、UTF-16 とする。ただし、実体参照はサポートしておらず、あらかじめ展開しておく必要がある。

コーパスの文書型は、このあと説明するモジュールの入力パラメータを変更することにより設定できる。『太陽コーパス』を利用する場合は、「corpus.xsd」に則った設定になっている。

corpus.xsd は、次に示すように、「雑誌」要素を複数包含する「雑誌」要素を定義している。「雑誌」要素は、個々の雑誌全体の文書構造を定義するための要素であり、『太陽コーパス』の文書型定義ファイル zassi.xsd により定義されている。corpus.xsd 中では、zassi.xsd を include する形で定義される(5行目参照)。

```
<?xml version="1.0" encoding="utf-16" ?>
<xsd : schema xmlns :
xsd="http://www.w3.org/2001/ XMLSchema">
  <xsd : element name="雑誌" >
    <xsd : include schemaLocation="zassi.xsd" />
    <xsd : complexType>
      <xsd : choice maxOccurs="unbounded">
        <xsd : element ref="雑誌"/>
      </xsd : choice>
      <xsd : attribute name="コーパス名"
        type="xsd : string" use="optional"/>
    </xsd : complexType>
  </xsd : element>
</xsd : schema>
```

『ひまわり』では、このように複数の構造化テキストを単一のコーパスファイルとして検索対象とするように設計されている。これは、小規模な複数の構造化テキストを検索対象とすると、管理方法が複雑になったり、索引ファイルが分散することにより、検索速度が低下するといった問題が起りうるためである（もちろん、コーパス全体の規模が大きくなれば、分散させて管理することも必要である）。

今回検索対象とする『太陽』の構造化テキストの場合、『太陽』1号分が一つのファイルに対応し、『太陽コーパス』全体では、60個ものファイルから構成されている。一つのファイルは、おおむね 1～2Mバイトである。『ひまわり』で検索対象とするデータでは、これら60個のファイルを一つのファイルに連結して、『太陽コーパス』全体として、一つのファイル (corpus.xml) になるようにしている。なお、『太陽』の構造化テキストでは、文字符号化方式として、Shift JIS を採用しているが、

corpus.xml としてまとめる際には、『ひまわり』の仕様どおり、UTF-16 に変換している。

4.2 検索モジュール

検索エンジンは、次の三つの検索モジュールから構成され、これらを組み合わせて各種の検索処理を行う。

- ・要素検索モジュール
- ・要素内容検索モジュール
- ・要素属性検索モジュール

名前からわかるとおり、XML 文書の要素 (element)、要素内容 (element content)、要素属性 (attribute) を検索するモジュール群であり、一般的な整形形式 XML 文書 (well-formed XML Document) を検索対象とするよう設計した。この設計により、異なる文書型のコーパスに対応することが容易にできるようになる。

これらのモジュールは、個々のモジュールが個別に管理する索引を使って、コーパスファイルに直接アクセスする。また、索引の構築は、個々のモジュールが内部の機能として持っている。

4.2.1 要素検索モジュール

■機能、入出力

要素検索モジュールは、XML 文書中の指定された位置を含んでいる要素の範囲を検索する機能を持つ。入力と出力は、次のとおりである。

- ・入力
 - －検索対象の要素名
 - －XML 文書中における位置
- ・出力
 - －要素が要素内容を持つ場合：
 - + 指定された位置を含む検索対象要素の開始タグの開始位置
 - + 指定された位置を含む検索対象要素の終了タグの終了位置
 - + なお、検索対象の要素が入れ子になっている場合は、

最も内部の要素の開始タグと終了タグの位置を返す。

－要素が要素内容を持たない（空要素の）場合：

- + 指定された位置に対して、直前に出現する検索対象要素タグの終了位置、および、
 - + 次に出現する検索対象要素タグの開始位置
- （なお、直前に出現する検索対象要素タグが存在しない場合は、XML 文書の先頭位置となる）

例えば、次の XML 文書の s 要素に対する検索を考えてみる。

```
<s> 太陽 <l 位置="P001C01"/><br/></s>
<s> 明治三十四年 <l 位置="P001C02"/><br/></s>
<s> 人生には限りありて冀望には際涯なし、</s>
```

要素検索モジュールへの入力として、「太陽」の「太」の位置を指定して、s 要素に対する検索を実行した場合、次の範囲の先頭と末尾の位置情報を返す。

```
<s> 太陽 <l 位置="P001C01"/><br/></s>
```

次に、空要素の場合について説明する。「明治」の「明」の位置を指定して、l 要素を検索した場合、一つ目の l 要素タグ（「位置」属性値がP001C01）の末尾の位置情報と、その次の l 要素タグ（「位置」属性値がP001C02）の末尾の位置情報を返す。

```
<br/></s>
<s> 明治三十四年 <l 位置="P001C02"/>
```

■検索アルゴリズム、および、索引づけ

索引は、作成対象の要素の範囲（開始位置と終了位置）を開始位置で並べ替えたものである。索引ファイルは、要素ごとに作成する。検索は、索引の開始位置を二分木検索（binary search）（石畑1989）方式を採用している。計算量は、要素数を n とすると、 $O(\log n)$ となる。索引を格納しておくための領域は、 $2 \times 4n$ バイト必要となる。これは、一つの索引に対して、開始位置と終了位置の二つ位置情報を格納すること、また、一つの位

置情報は 4 バイトで表現されることから導かれる。

4.2.2 要素内容検索モジュール

■機能，入出力

要素内容検索モジュールは，文字列をキーとして，要素内容の文字に対して検索を実行する。要素内容に対する検索の入力と出力は，次のとおりである。

- ・入力
 - －検索対象とする要素名，検索対象の文字列
 - －検索対象の文字列の前文脈，後文脈，文脈長（任意）
- ・出力
 - －指定された文字列の XML 文書中での位置

■検索アルゴリズム，および，索引づけ

索引には，検索対象とする要素中の文字に対する Suffix Array (山下2000) を作成し，それを昇順に並べ替えたものを用いている。索引は，要素ごとに作成する。検索は，Suffix Array を二分木検索することにより実現している。計算量は，索引づけした文字数を n とすると， $O(\log n)$ である。なお，検索結果が複数ある場合は，最初に検索された結果の Suffix Array のインデックス値を前後させることにより検索できる。したがって，その際の計算量は $O(1)$ である。索引を格納しておくための領域は， $4n$ バイト必要となる。また，索引づけ可能な文字数は，Java 言語の仕様上，最大 $2^{31}-1$ 個である。

検索時のマッチングは，完全一致である。ただし，あくまでも要素内容に対して，マッチングのテストを行い，要素タグは，すべて無視される。例えば，文字列「制度」のマッチングを行う場合について見てみる。

`<s>単に制<l 位置="P002A05"/>度の變化を以てするも，</s>`

この例では，「制」と「度」の間に 1 要素タグ「`<l 位置="P002A05"/>`」が入っているが，要素タグは無視するので，マッチングに成功する。

要素内容検索モジュールは，検索文字列に対する前文脈と後文脈，文脈長を指定できる。これらの指定は，任意である。指定されると，文脈長の範囲の前後文脈を制約条件として，検索対象文

字列とのマッチングを試みる。前後文脈は、正規表現によって指定することができる。正規表現とのマッチングの際も、要素タグは無視される。

なお、正規表現によるマッチングには、Java 言語の標準ライブラリに含まれる `java.util.regex` を使っている。したがって、正規表現の文法、マッチングに関する規定は、`java.util.regex` に依存する。詳細は、
<http://java.sun.com/j2se/1.4/ja/docs/ja/api/java/util/regex/package-summary.html> を参照されたい。

4.2.3 要素属性検索モジュール

■機能，入出力

要素属性検索モジュールは、指定された要素属性の属性値をキーとして検索を行い、その属性値の XML 文書中での位置情報を返す。要素属性検索モジュールの入出力は、次のとおりである。

- ・入力
 - －検索対象の要素名，属性名，属性値
 - －検索対象の文字列の前文脈，後文脈，文脈長（任意）
- ・出力
 - －指定された属性値の XML 文書中での位置

■検索アルゴリズム，および，索引づけ

索引づけ，検索アルゴリズムともに，要素内容検索モジュールとほぼ同様である。異なるのは，検索と索引づけの対象が，要素属性値であるという点である。検索時のマッチングは，属性全体に対する完全一致である。したがって，索引づけは，属性値の先頭の文字だけを対象にしている。

索引づけした文字数を n とすると，検索の計算量は $O(\log n)$ ，索引を格納する領域も， $4n$ バイトである。この場合の n は，付与する属性数と同数である。索引ファイルは，要素属性ごとに作成する。

4.3 『ひまわり』の全文検索機能

ここでは，三つの検索モジュールを使って，どのように『ひまわり』の全文検索機能を実現しているかを説明する。

全文検索機能の流れを次に示す。各処理の詳細は，この後の節

で順次解説していく。

1. 検索条件の指定
2. 検索キーの検索
3. 要素属性の取得と検索結果の絞込み

4.3.1 検索条件の指定

『ひまわり』における検索条件は、検索キーと絞込み条件からなる。検索キーは、必ず指定しなければならない条件である。一方、絞込み条件は、検索キーにマッチした結果を絞り込むための条件で、指定は任意である。『ひまわり』で使用できる検索キーと絞込み条件を次に示す。なお、検索エンジンの仕様の上では、ここに示した以外の要素や属性を検索キーや絞込み条件にすることも可能であるが、『太陽コーパス』を利用する上で必要性の高いものを優先して設定してある。

- ・検索キー
 - －すべての要素の要素内容（つまり、本文）
 - －r 要素の rt 属性値
- ・絞込み条件
 - －検索キーに対する前文脈、後文脈
 - －「雑誌」要素の「年」、「号」
 - －「記事」要素の「題名」、「著者」、「欄名」、「文体」、「ジャンル」属性
 - －「l」要素の「位置」属性
 - －「引用」要素の「種別」、「話者」属性
 - －「注」要素の「原文」属性

検索キーとして指定できるのは、基本的に単純な文字列である。一方、絞込み条件には、正規表現を使うことができる。

4.3.2 検索キーの検索

設定された検索キーを入力として、検索を実行する。検索キーとして、要素内容が指定された場合は、要素検索モジュールを使い、r 要素の rt 属性などの要素属性値が指定された場合は、要素属性検索モジュールを使って検索する。要素属性検索モジュールの出力は、属性値の位置情報なので、この位置情報を要素検索モジュールの入力として、再度検索を実行し、その要素内容を取得する。

検索結果は、「検索結果集合」として主記憶内に保持される。

結果として保持されるのは、検索対象文字列と XML 文書中における検索対象文字列の位置情報である。「検索結果集合」には、複数の検索結果を格納しておくことができるとともに、年や号など検索文字列に付随する要素属性を関連づけて格納しておくことができる。

4.3.3 要素属性の取得と検索結果の絞込み

絞込み条件に適合しているか調べるために、検索キーの検索により得られた検索結果集合の個々の結果に対して、各種の要素属性を取得する。

要素属性の取得には、要素検索モジュールを用いる。要素検索モジュールへの入力、検索結果集合に格納されている検索対象文字列の位置情報であり、その出力として、対象とする要素の範囲を取得する。そこから得られる開始位置に基づいて、XML 文書から要素の開始タグ、ひいては、各種属性を取得する。

以上の処理により取得された属性値を絞込み条件と比較し、条件に適合しない場合は、検索結果集合から削除する。適合した場合は、検索した属性値を検索結果集合に追加していく。そして、4.3.1 で示したすべての絞込み条件に適合した検索結果だけが、検索結果集合に残ることになる。

例えば、検索キーが「太陽」とし、これから「記事」要素の「文体」属性を取得する場合を考える。要素内容検索モジュールの結果により、「太陽」の位置情報が得られているので、その位置情報を入力として、「記事」要素に対して、要素検索を行う。その結果、「記事」要素の範囲が得られ、そこから、「記事」タグの開始位置が取得できる。得られた開始位置に基づいて、「文体」属性を取得し、絞込み条件と適合するかチェックする。

```
<記事 題名="展覧会是非" 著者="内田魯庵" 欄名="案頭三尺"
  文体="口語" ジャンル="NDC706">
<s> 案頭三尺 内田魯庵 <l 位置="P059A01"/><br/></s>
<s> 展覧会是非 <l 位置="P059A02"/><br/></s>
<s> (一) 美術を禍ひする人氣 <l 位置="P059A03"/>
<br/></s>
```


4.4 その他の検索処理

4.4.1 記事閲覧機能

記事閲覧機能は、記事全体を Web ブラウザで表示する機能である。機能の詳細については、6.2 節で詳しく述べることにし、ここでは、検索モジュールを使って、記事閲覧用のデータを構築する手順を説明する。

記事を閲覧するためには、「記事」要素を抽出し、Web ブラウザで閲覧できる HTML 形式に変換する必要がある。その手順は、次のとおりである。

- (1)要素検索モジュールにより、「記事」要素の範囲情報（開始位置と終了位置）を検索
- (2)検索された範囲情報を元に、XML 文書から当該の「記事」要素を抽出
- (3)当該の記事を含む「雑誌」要素を、要素検索モジュールにより検索
- (4)検索された「雑誌」要素の全属性を持ち、当該の「記事」要素だけを含む「雑誌」要素を生成
- (5)生成された「雑誌」要素からなる XML 文書を生成する。
生成される XML 文書の例を次に示す。生成する際には、スタイルシート処理命令 (xml-stylesheet) を付加し、XSL スタイルシートを指定する。下の例では、2 行目で XSL スタイルシート zassi.xsl を指定している。

```
<?xml version="1.0" encoding="utf-16"?>
<?xml-stylesheet href="zassi.xsl"
type="text/xsl" ?>
<雑誌 題名="文学と英国の書肆" ... (その他の属性) ... >
  : (中略)
</雑誌>
```

- (6)生成された XML 文書を記事閲覧用テンポラリファイル (3.2.2 参照) を介して、Web ブラウザに渡し、HTML 形式に変換
- (7)HTML 形式に変換された結果を Web ブラウザで表示

この流れを見ると分かるとおり、(6) の段階で、Web ブラウザに記事表示用の XML 文書が渡され、HTML 形式への変換は、Web ブラウザに内蔵されている XSLT (XSL Transformations) プロセッサが行う。Web ブラウザで表示する際の文字色などの物理的構造を設定する CSS (Cascading Style Sheet) ファイルは、XSL スタイルシートの中で指定しておく。

このように、表示方法の設定は、『ひまわり』と独立しているため、利用者側で柔軟に設定することも可能である。実際の表示例は、6.2 節の図13、図14を参照していただきたい。

4.4.2 雑誌一覧、記事一覧表示

雑誌一覧表示機能、記事一覧表示機能は、それぞれ、コーパスファイル (corpus.xml) に登録されている雑誌一覧、記事一覧を表示する機能である。この機能の実現には、要素検索モジュールが使用される。

雑誌一覧表示機能の場合は、「雑誌」要素に対して作成された索引を元にすべての「雑誌」要素を検索し、その開始タグから、必要な属性を抽出し、一覧にして表示している。

記事一覧機能も、「記事」要素に対して、同様の処理を行っている。なお、記事一覧表示の際に表示される属性には、「年」、「号」など「雑誌」要素の属性も含まれるが、これらの属性は、当該「記事」要素が含まれる「雑誌」を再度要素検索することにより実現している。

5

検索条件入力インターフェイス

検索条件入力インターフェイスは、XML 文書や『太陽コーパス』自体に知識を持たない利用者が容易に検索条件を指定できるよう、次の仕組みを用意している。

- ・検索キー入力インターフェイス
- ・絞り込み条件入力インターフェイス
- ・字体変換辞書

5.1 検索キー入力インターフェイス

検索キー入力インターフェイスは、検索条件のうち、検索キーの入力を支援する。検索キー入力インターフェイスの例を図3に示す。

図3 検索キー入力インターフェイス

図3のように、構造化テキストのどの部分を検索対象とすることをメニュー形式で選択することができる。現在のインターフェイスでは、本文（すべての要素の要素内容）、「r」要素の rt 属性がメニューに登録されている。検索キー自体は、このメニューの右側の欄に入力する（図3では、「理想的」が入力されている）。

5.2 絞込み条件入力インターフェイス

絞込み条件のうち、前文脈、後文脈については、検索キーの入力欄の下に指定欄を設けている（図4）。複数の絞込み条件を指

前文脈	キー	後文脈	絞り込み条件
私立学校に至つては其	理想的	なるもの...	で終る
の統治する國家を以て	理想的	なりと爲...	を含む
淨められたればこそ其	理想的	な協同生...	正規表現
つてしまふ。自分は	理想的	な社會が...	と一致しない
化場を設け、完全な	理想的	な養魚を...	で始まる
人は一面に於て科學的	理想的	なる新建...	で始まる

図4 絞込み条件の入力インターフェイス（前文脈、後文脈）

定すると、and 条件として解釈される。

絞込み条件には、図4のように、メニュー形式で一致条件を設定することができる。

一致条件のうち、完全一致（「と一致する」）以外は、次のように内部的に正規表現を用いて実現している（str は指定された文字列）。したがって、指定した文字列は、正規表現として解釈される。

- ・前方一致（「で始まる」）：^str
- ・後方一致（「で終わる」）：str\$
- ・を含む（「を含む」）：str
- ・正規表現：str

前文脈、後文脈以外の要素属性に関する絞込み条件は、「フィルタ」タブにまとめてある。図5 に示すとおり、絞込みを行う属性、属性値、一致条件を指定する。絞込みを行う属性、一致条件は、メニュー形式で選択できる。

The screenshot shows a software interface for setting filters. It has three tabs: '検索文字列' (Search String), 'フィルタ' (Filter), and 'オプション' (Option). The 'フィルタ' tab is active. On the left, there is a vertical list of attributes: 年 (Year), 年号 (Year Number), 題名 (Title), 著者 (Author), 欄名 (Section Name), ジャンル (Genre), 文体 (Style), and 話者 (Speaker). The main area contains a table with three columns: '属性' (Attribute), '値' (Value), and '条件' (Condition). The table has three rows, each with a dropdown menu for the attribute and a dropdown menu for the condition. The first row has '年' selected for the attribute and 'で始まる' (Starts with) for the condition. The second row has '年号' selected for the attribute and 'で始まる' (Starts with) for the condition. The third row has '題名' selected for the attribute and 'で始まる' (Starts with) for the condition. Below the table, there is a preview of the filtered results, showing a table with columns: キー (Key), 後文脈 (Context), 年 (Year), 号 (Number), and 題名 (Title). The first row of the preview shows '国語' (Language), 'の構成に由...' (Origin of the structure...), '1895', '01', and '漢字の利害' (Pros and cons of Chinese characters).

属性	値	条件
年		で始まる
年号		で始まる
題名		で始まる

キー	後文脈	年	号	題名
国語	の構成に由...	1895	01	漢字の利害

図5 絞込み条件の入力インターフェイス（各種属性）

5.3 字体変換

5.3.1 字体変換の機能

『太陽コーパス』のように、いわゆる旧字体が用いられている資料に対して文字列検索する場合、利用者が旧字体を指定できずに検索もれが生じる場合がある。『ひまわり』では、この問題を解消するために、字体変換機能を持っている。

『ひまわり』の字体変換機能は、検索キー、および、絞込み条件の文字列指定欄で利用することができる。次の例は、入力され

た文字列「国語」に対して字体変換を実行した例である。

国語 => 國語

字体変換に際しては、単純に字体の置き換えをするのではなく、資料の実態にあわせた変換を行えるようにする必要がある。例えば、新旧の字体でも、「国」と「國」のように、資料の中で使用されていない字体と使用されている字体の関係にあるもの、「國」と「圀」のように、両方が使い分けられている場合もある。また、「仮」、「假」のように両方が等価的に用いられている場合もある。

そこで、『ひまわり』の字体変換機能では、変換前の字体（指定字体）に対して、変換後の字体を三つに分類している。

等価字体：指定字体に対して、等価的に用いられることがある字体

参考字体：異体字だが、指定字体に対しては、通常は等価的に用いられない字体

指定字体：変換前の字体

次の表1は、字体変換辞書の例である。

表1 字体変換辞書

指定字体	等価字体	参考字体
国	國	圀
圀	圀	國
仮	仮, 假	
假	假, 仮	
鬱	鬱, 鬱	
鬱	鬱, 鬱	
弁	弁, 辨, 辦, 瓣, 辯	辯

利用者は、どの字体に変換するかを指定する。字体の指定は、複数指定が可能である。初期設定では、等価字体に変換するように設定されている。変換後の字体は、検索キーや絞り込み条件の文字列指定欄に変換されるので、利用者が適切な字体を選択することができる。

変換後の字体が複数ある場合は、正規表現の文字クラスに類似した表記で表示される。例えば、文字列「憂鬱」に対して、等価字体、参考字体の二つの字体に変換する場合は、次のように変換される。[]の中の文字列が変換後の字体になる。もし、「鬱」を検索対象からはずしたければ、文字列指定欄から「鬱」を削除す

ればよい。

憂鬱 => 憂[鬱鬱]

5.3.2 字体変換辞書

字体変換機能を有効に働かせるためには、検索する資料に応じて適切な字体変換辞書を設定する必要がある。『太陽コーパス』に同梱する『ひまわり』に搭載されている字体変換辞書の内容については、本書の「漢字の実態と処理の方針」（田中牧郎）286頁を参照されたい。

字体変換辞書は、利用者が自由に定義することが可能である。字体変換辞書ファイル jtaidic.xml の文書型は、jtaidic.dtd で規定されている。参考のために、表 1 の字体変換辞書を XML で記述したものを次に示しておく。詳細な内容は、jtaidic.dtd を参照していただきたい。

<字体変換辞書>

<指定字体 等価="國" 参考="圀">国</指定字体>

<指定字体 等価="圀" 参考="國">圀</指定字体>

<指定字体 等価="仮假" 参考="">仮</指定字体>

<指定字体 等価="仮假" 参考="">假</指定字体>

<指定字体 等価="弁辨辦辯辯" 参考="辯">弁</指定字体>

</字体変換辞書>

6

検索結果ブラウザ

検索結果ブラウザは、検索結果を利用者が閲覧しやすいように表示するとともに、利用者が検索結果を分析するのを支援する役割を果たす。3.1に示したとおり、『ひまわり』の検索結果ブラウザには、全文検索結果を閲覧するための KWIC 形式のブラウザと記事全体を閲覧する形式のブラウザがある。

6.1 検索結果ブラウザ (KWIC形式)

6.1.1 概要

KWIC 形式の検索結果ブラウザは、検索エンジンにより作成

された検索結果集合を表形式で表示する（図6）。ここに表示されるのは、検索キーと絞込み条件で指定できるすべての項目（前文脈、後文脈、年、号、題名、著者、欄名、ジャンル、文体、話者、種別など）である。前後文脈の長さは、オプションで指定できる。規定値は、前後文脈とも10文字である。また、検索直後は、年、号、号内での出現位置をキーとして、昇順に並び替えられている。

前文脈	キー	後文脈	年	号	題名	著者	欄名	ジャンル
ず、今もし一國其の	を	表するに新たに文字	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
せて古來海に興りし諸	を	学ぶ、之を漢字を	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
能くすべくば必ず三	に	通ずるを求めんか。	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
を求めんか。一には	、	是れ我國民の思想	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
、斯の國語あり、	を	表出する、もし	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
れ斯の國あり、斯の	あり、	國語を表出す	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
を記す、是れ其の力	を	修むると共に支那語	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
しのみにあらず、又	に	便益する所極めて多	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
の語を採て、之を他	に	同化するに當りては	1895	01	漢字の利害	三宅雪嶺	論説	NDC81
、其語の採て以て我	に入るべき者、その		1895	01	漢字の利害	三宅雪嶺	論説	NDC81
の當時に在て、採て	となされし者の多き、		1895	01	漢字の利害	三宅雪嶺	論説	NDC81
通ありしより、彼の	にして採て我國語と爲		1895	01	漢字の利害	三宅雪嶺	論説	NDC81
彼の國語にして採て我	と爲すに便なるもの決		1895	01	漢字の利害	三宅雪嶺	論説	NDC81
西語を翻譯する者、	の推成に由らずして		1895	01	漢字の利害	三宅雪嶺	論説	NDC81
所とせんか。	研究に就て	上田	1895	01	國語研究に...	上田万年	論説	NDC81
左の一欄は君が客冬	研究會發會式の席上に		1895	01	國語研究に...	上田万年	論説	NDC81
。今日茲に皆様と此	研究會を開きまするに		1895	01	國語研究に...	上田万年	論説	NDC81
は、此大日本帝國の	を尋み愛しみ、殊に		1895	01	國語研究に...	上田万年	論説	NDC81
國は一方に於て、此	の過去に溯り、又其		1895	01	國語研究に...	上田万年	論説	NDC81
に入るべき者、その								

検索総数: 474

図6 検索結果ブラウザ（KWIC形式）

ブラウザの下部の欄には、現在マウスでポイントされている欄の値が表示される（例えば、前後文脈欄など、表中の欄に表示しきれない場合に、この欄で値を参照する）。さらに、最下部には、検索総数が表示される。

KWIC 形式の検索結果ブラウザには、次の機能がある。これらの機能の詳細は、後続の節で説明する。

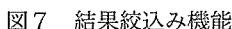
- ・並び替え機能
- ・検索結果絞込み機能
- ・頻度計測機能
- ・著者データベース機能
- ・雑誌一覧、記事一覧機能
- ・他のアプリケーションとのやりとり

並び替え機能は、特定の列をキーにして、検索結果を並び替える機能である。並び替えは、『ひまわり』の列名をクリックすると実行される。通常は、昇順に並び替える。降順に並び替えを行うには、シフトキーを押しながら、列名をクリックする。

複数のキーを指定するには、キーとしたい列を優先順位の逆順にクリックすればよい。例えば、優先順位を「年」、「号」順として、並び替えを行いたい場合には、優先順位の逆順の「号」、「年」の順で並べ替え機能を実行する。

検索結果絞込み機能は、その名の通り、検索結果を絞り込む機能である。検索時の絞込みと異なり、主記憶に保持されている検索結果集合に対して絞込みを行う。この際、絞込み解除のことを考慮し、絞込みを行う前に、検索結果集合のコピーを保存しておく。絞込みが解除されたときは、コピーしておいたデータを検索結果集合に戻す。

絞込みの条件は、列ごとに指定する。絞込みを実行したい列名を右クリックすると、ウィンドウが起動するので、そこで条件を設定する（図7）。条件には、当該列の異なりがリストとして表示される。また、条件を正規表現で記述することもできる。



絞込みを実施した状態で、再度絞込みを行うこともできる。絞込みを解除するには、絞込みを実行する時と同様に、列名を右クリックする。表示されたウィンドウの中に [フィルタ解除] の項目があるので、それを選択する。ただし、複数絞込みを行っていた状態で、絞込みを解除するとすべての絞込み条件が解除される。

6.1.4 頻度計測機能

頻度計測機能は、検索結果から検索対象文字列の出現頻度を計測する機能である。出現頻度は、検索結果集合の内容を集計している。したがって、検索結果の絞込みを実行していれば、その結果は計測結果に反映される。

『ひまわり』では、「検索キー，年」，「検索キー，年，号」を単位として、出現頻度を計測する。それぞれの実行例を図8に示す。

計測結果は、図8のように表形式で表示される。この表は、検索結果ブラウザ (KWIC形式) と同様に並び替え機能、絞込み機能を利用することができる。表の下部にある欄は、表の欄をマウスでクリックした際にその値を表示するためのものである。また、最下段には、頻度の総数が表示される。

キー	年	頻度
国語	1895	150
国語	1901	147
国語	1909	62
国語	1917	74
国語	1925	41
総数: 474		

キー	年	号	頻度
国語	1895	02	13
国語	1895	04	2
国語	1895	05	3
国語	1895	06	4
国語	1895	07	14
国語	1895	08	18
国語	1895	09	10
総数: 474			

図8 頻度計測機能

6.1.5 著者データベース機能

『太陽コーパス』は著者データベースを持っており、『ひまわり』はその内容をコーパスの検索結果と関連づけることができる。『ひまわり』では、『太陽コーパス』の著者データベースを「著者

情報ファイル」に変換して利用する。著者情報ファイルには、次の情報が登録されている。

- ・氏名
- ・所属
- ・分野
- ・生年
- ・没年

これらの情報は、文書型定義ファイル `author.dtd` に則った XML 文書として記述されている。著者データベースの一部を次に示す。「著者データベース」要素が著者データベースを表し、「著者」要素が一人分の情報を表す。『太陽コーパス』の著者データベースの詳細については、本書の「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」（田中牧郎）の「8.2 「著者」属性」（22頁）を参照していただきたい。

<著者データベース>

<著者>

<氏名>浅草坊</氏名><所属></所属><分野></分野>

<生年></生年><没年></没年>

</著者>

<著者>

<氏名>朝倉文夫</氏名><所属></所属><分野></分野>

<生年>1883</生年><没年>1964</没年>

</著者>

『ひまわり』では、上記の形式に則った著者情報ファイル `author.xml` を起動時に主記憶に読み込み、検索結果ブラウザなどから利用できるようになっている。

著者情報ファイルを使った機能が、著者検索機能である(図9)。検索結果ブラウザ(KWIC形式)で閲覧したい著者を左マウスボタンでダブルクリックすると、著者データベースから著者情報が検索される。図9が著者検索結果の例である。

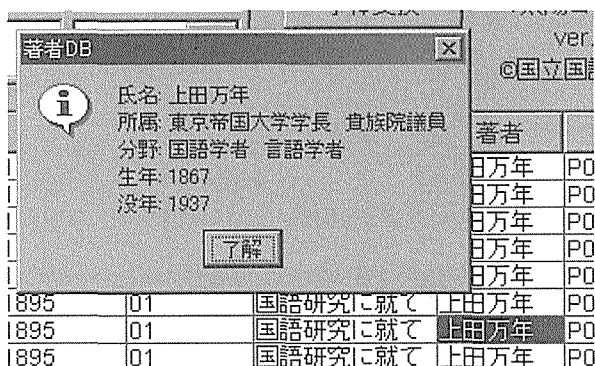


図9 著者検索機能

著者データベースを使ったもう一つの機能は、著者一覧機能である(図10)。この機能は、著者データベースに登録されている全著者を表形式で表示するもので、検索結果ブラウザ(KWIC形式)から独立したウィンドウを起動して表示する。この表も、検索結果ブラウザ(KWIC形式)と同様に、各列の並び替え機能、絞り込み機能を利用することができる。

氏名	氏名よみ	所属	分野	生年	没年
羽峰外史	はほうがいし	東京帝国大学教授		1823	1909
沢雲照	しゃくろんしょう		俳侶 仏教家	1827	1909
トルストイ				1828	1910
西村茂樹	にしむらしげき	日本仏道会創設者	密教思想家教育家	1828	1902
小宮山經介	こみやますすけ		漢学者	1829	1896
飯田武郷	いいだたけさと		国学者	1829	1901
黒川真頼	くろかわまより	東京帝国大学教授	国学者	1829	1906
三島中洲	みしまちゅうしゅう	二松学舎大学創立者	漢学者	1830	1919
川田奎江	かわたおうこう	文学博士	漢学者	1830	1896
黒田清綱	くろだきよつな	枢密顧問官 貴族院議員	政治家 歌人	1830	1917
曳尾晃	ひきおしゅう	元老院議員 貴族院議員	外交官 漢学者	1831	1915
森睦村	もりおうそん		漢学者	1831	1907
福岡美静	ふくばげせい(よし...	元老院議員 貴族院議員	国学者 歌人	1831	1907

検索総数: 1261

図10 著者一覽機能

6.1.6 雑誌一覧, 記事一覧機能

雑誌一覧機能と記事一覧機能は、その名のとおり、コーパスファイルに登録されているすべての雑誌と記事を一覧表示する機能である。それぞれの表示例を図11, 12 に示す。これらの一覧表も、検索結果ブラウザ (KWIC形式) と独立したウィンドウとして表示される。また、検索結果ブラウザ (KWIC形式) と同様に各列の並び替え機能、絞り込み機能を備えている。

雑誌名	年	号	Version
太陽	1895	01	1.0
太陽	1895	02	1.0
太陽	1895	03	1.0
太陽	1895	04	1.0
太陽	1895	05	1.0
太陽	1895	06	1.0
太陽	1895	07	1.0
太陽	1895	08	1.0
太陽	1895	09	1.0

検索総数: 80

図11 雑誌一覧機能

題名	著者	年	号	欄名	文体	ジャンル
〈扉〉	*	1895	01	**	文語	***
太陽の発刊	大橋新太郎	1895	01	**	文語	NDC051
学界の大革新	久米邦武	1895	01	論説	文語	NDC002
戦後の教育	千頭清臣	1895	01	論説	文語	NDC371
戦後の学術	井上哲次郎	1895	01	論説	口語	NDC002
戦争と文学	坪内逍遙	1895	01	論説	文語	NDC901
漢字の利害	三宅雪嶺	1895	01	論説	文語	NDC811
国語研究に就て	上田万年	1895	01	論説	口語	NDC810
事物変遷の研究に對す...	坪井正五郎	1895	01	論説	口語	NDC469

検索総数: 3721

図12 記事一覧機能

6.1.7 他のアプリケーションとのやりとり

大量の検索結果を統計的に処理したり、検索結果を他のアプリケーションにコピーしたいという要求に答えるために、『ひまわり』では次の二つの手段を提供している。

- ・ 検索結果のテキストファイルへの出力
 - － 検索結果を外部ファイルに出力するための機能である。
 - － 出力されるファイルの形式は、タブ区切りのテキストファイルである。検索結果ブラウザ (KWIC形式) のすべての欄のデータがファイルに出力される。
 - － 出力されるファイルの文字コードは、使用している OS 既定の文字コードに設定される。例えば、Windows であれば、Shift JIS である。
- ・ クリップボードへのコピー
 - － 検索結果をクリップボードへコピーできる。これにより、手軽に検索結果を他のアプリケーションに転送することができる。
 - － なお、クリップボードへコピーする際は、上の「検索結果のテキストファイルへの出力」と同様、タブ区切りの形式でコピーしている。したがって、検索結果ブラウザ (KWIC形式) に表示されている全データをコピーすれば、各欄のデータが、Excel などの表計算ソフトウェアの個々のセルにコピーされる。

6.2 検索結果ブラウザ (記事閲覧)

記事閲覧機能は、記事全体を Web ブラウザで表示する機能である。この機能は、KWIC形式のブラウザ、または、記事一覧のウィンドウから起動される。

表示用の Web ブラウザは、次の三つのブラウザのうちのいずれかを選択することができる。初期設定では、Internet Explorer となっている。

- ・ Microsoft Internet Explorer
- ・ Mozilla, Firefox (<http://www.mozilla.org/>)

表示のスタイルは、XSL スタイルシートと CSS ファイルによって設定することができる。『太陽コーパス』に同梱される『ひまわり』には、次の四つの CSS ファイルが付属している。

- ・ 横書き・段落表示

- ・横書き・行番号表示
- ・縦書き・段落表示
- ・縦書き・行番号表示

表示例として、このうち、「横書き・段落表示」と「縦書き・段落表示」の例をそれぞれ図13、14に示す。(この図では色が確認できないが、)文字色が赤の文字列は、検索対象文字である。下線が引いてある部分は「引用」要素部分である。「引用」要素の「話者」属性のように、要素の属性値は、当該要素にマウスを移動させることにより、ポップアップで属性が表示されるようになっている(図13参照)。

このように、XSLスタイルシートとCSSファイルを使うことにより、さまざまな言語的情報を見やすい形で表示できる。また、XSLスタイルシート、CSSファイルを変更することにより、利用者が資料にあわせて、表示の形式を設定することができる。

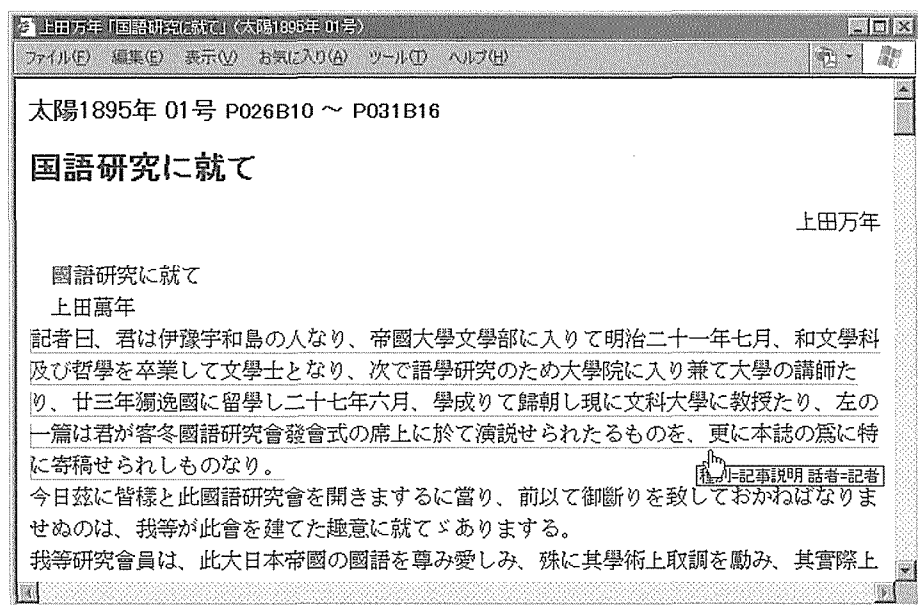


図13 記事閲覧機能 (横書き・段落表示)

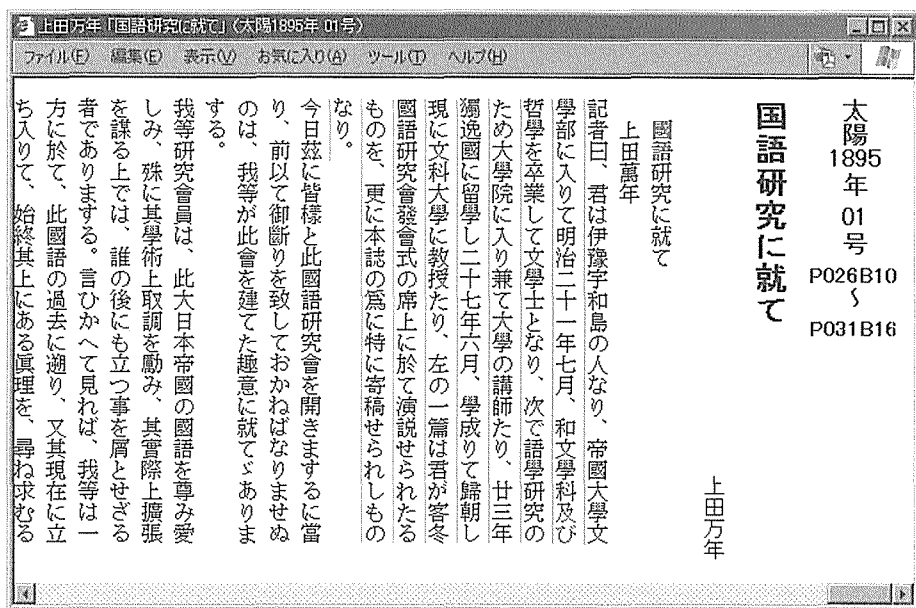


図14 記事閲覧機能（縦書き・段落表示）

7

おわりに

本稿では、全文検索システム『ひまわり』の仕様について解説を行った。ここでは、『太陽コーパス』への適用例という形で記述したが、『太陽コーパス』以外にも『日本語話し言葉コーパス』、『分類語彙表』など、さまざまな構造化テキストに適用できる。詳しくは、山口ら（2004）を参照していただきたい。また、『ひまわり』は現在も開発が継続されている。最新情報については、国語研究所の Web ページ（<http://www.kokken.go.jp/lrc/>）で公開している。

参考文献

- 石畑清（1989）「アルゴリズムとデータ構造」（『岩波講座ソフトウェア科学』3，67-72頁，岩波書店）
 山口昌也・田中牧郎（2004）「多様な構造化テキストに対応した全文検索システム『ひまわり』」（『日本語学会2004年度秋

季大会予稿集』日本語学会)
山下達夫 (2000) 「用語解説 (Suffix Array)」 (『人口知能学会
論文誌』 Vol.15 No.16, 1142頁)

構造化テキストを直接利用するアプリケーション

～『プリズム』と『たんぽぽ』～

———小木曾 智信

1

はじめに

『太陽コーパス』にはXMLで記述した『太陽』の構造化テキストを直接利用するアプリケーションとして、変換・情報抽出用アプリケーション『プリズム』と各種のXSLTスタイルシート、そして検索用アプリケーション『たんぽぽ』が収録されている。いずれも簡易なプログラム言語で記述されたものなので、利用者側で手を加えて利用することもできるようになっている。

XSLTスタイルシートとは、XML文書（XMLで記述された構造化テキスト）の形式を変換するための規則を記述した簡単なプログラムのようなものである（注1）。『プリズム』を使って『太陽』のXML文書に各種のXSLTスタイルシートを組み合わせることによって、構造化テキストを他の形式に変換したり、そこから必要な情報だけを抜き出したりすることが可能になる。

検索用の『たんぽぽ』は、『ひまわり』とは違って構造化テキストを直接検索するプログラムであるため、検索に時間がかかるものの、フリガナを本文と見なしたり、踊り字を展開したりといった小回りのきく検索が可能になっている。

2

利用のための条件と注意事項

アプリケーションはどちらもInternet Explorer 6以降がインストールされたWindows環境で動作する。快適に使うために必要なパソコンの性能の目安は次の通りである。

表1 『プリズム』『たんぽぽ』の動作環境

OS	Windows 98 / Me / 2000 / XP
ブラウザ	Internet Explorer 6以降 (XMLの処理にMSXML3以降を利用する)
CPU	Pentium (互換) プロセッサ・300MHz以上
メモリ	64MB以上
モニタ	256色以上, 800 × 600以上の解像度

『プリズム』と『たんぽぽ』はインストール作業を行う必要はなく、CD-ROM上でそのまま利用することができるため、ハードディスクの空きは特に必要としない。ただし、データとともにハードディスクにコピーして使用したほうが動作は速くなる。

この2つのアプリケーションは手軽に利用できるようにJavaScript (Jscript) とHTMLアプリケーション (HTA) というWeb関連技術を利用して作られている。ところが、一部のウイルス対策ソフトはファイル操作を行うHTA形式を一律に危険なスクリプトとして判定するため、これらのアプリケーションも危険なものとして誤判定される場合がある。しかしウイルスなどの悪意あるコードが混入しているわけではない。

なお、ここで紹介するアプリケーションは、いずれもテキストエディタで開くことでソースを見ることができる。したがって、これを改造して利用目的にあったアプリケーションを作成することも可能であるが、改造したものの再配布については著作権に関する規定に従う必要がある。

3 形式変換と情報抽出

ここでは『太陽コーパス』付属のXSLTスタイルシートと、スタイルシートを用いて形式変換や情報抽出を簡単に行うためのアプリケーション『プリズム』について説明する。

3.1 XSLTスタイルシート

『太陽コーパス』には表2に示すXSLTスタイルシートが収録されている(表の丸付き数字は2.1.1, 2.1.2における説明と対応

している)。

表2 『太陽コーパス』付属スタイルシート一覧

形式変換	Web閲覧用 (HTML)	本文－DHTML	tx2dhtml.xsl	①
		本文－DHTML (行番号)	tx2dhtml2.xsl	②
		本文－シンプルなHTML	tx2html.xsl	③
	テキスト (TXT)	本文－プレーンテキスト	tx2text.xsl	④
		本文－プレーンテキスト (行番号)	tx2text2.xsl	⑤
	印刷用 (LaTeX)	本文－pLaTeX本文	tx2latex.xsl	⑥
		本文－pLaTeX本文 (行番号)	tx2latex2.xsl	⑦
情報抽出	Web閲覧用 (HTML)	外字一覧 (コード順) HTML	htm_gaiji.xsl	⑧
		外字一覧 (出現順) HTML	htm_gaijic.xsl	⑨
		記事情報HTML	htm_note.xsl	⑩
		注情報HTML	htm_kiji.xsl	⑫
		引用情報HTML	htm_quot.xsl	⑭
	表形式 テキスト (CSV)	記事情報CSV	htm_note.xsl	⑪
		注情報CSV	htm_kiji.xsl	⑬
		引用情報CSV	htm_quot.xsl	⑮
HTML形式に変換して閲覧するための共通書式 (他のスタイルシートから呼び出す)			html_mod.xsl	3.1.3 参照

スタイルシートはいずれもXSLT 1.0に準拠しており、その標準的な機能だけで作られているため、対応したXSLTプロセッサを用いればWindows以外の環境でも利用可能である。ただし文字符号化方式がシフトJISになっているため環境に合わせて変換しなければならない場合がある。

XSLTはXML文書の形を変えて出力するためのものである。その役割から、元の構造化テキストの内容を保ちつつ言語研究のために閲覧や分析がしやすい形式に変える「形式変換」と、元の構造化テキストから必要な内容だけを抜き出して閲覧や分析に供するための「情報抽出」に分けられる。以下、それぞれについて説明してゆく。

3.1.1 形式変換

構造化テキストを閲覧・分析しやすい形式に変換するためのスタイルシートとして、テキスト形式・Webブラウザでの閲覧用形式 (HTML) ・印刷用形式 (TeX/pLaTeX2e (注2)) に変

換するものを用意した。それぞれに通常の本文と行番号付きの本文（本文を原文と同じ位置で改行してページ・行番号などの位置情報を付けたもの）の2つの形式がある。さらにWebブラウザでの閲覧用形式では、構造化文書に埋め込まれたほとんどの情報が閲覧できる形式と、情報の多くを省いて単純化した形式の2種類を用意した。

①本文-DHTML

Webブラウザで本文を閲覧するための形式に変換する。表示される目次のボタンを押すと本文が開かれるなど、動きのある表示形式（Dynamic HTML）になっているため、閲覧には新しいWebブラウザ（Internet Explorer 6以降、Netscape7.0以降、opera 7以降、Firefoxなど）が必要である。外字を画像で再現するほか、ルビや注記、引用などの各種情報を再現している。図1は1895年1号の「国語研究に就て」という記事をInternet Explorerで開き、ボタンを押して展開して閲覧しているところである。

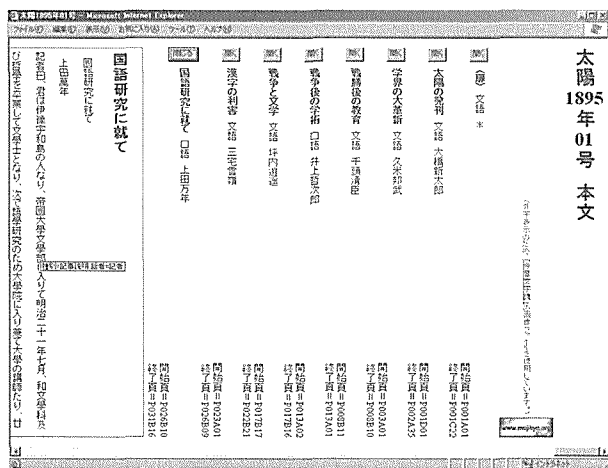


図1 本文-DHTML / ① / Internet Explorer

②本文-DHTML（行番号）

①の本文を原文と同じ位置で改行し、位置情報を付けたもの。同じ記事を閲覧しているところ。

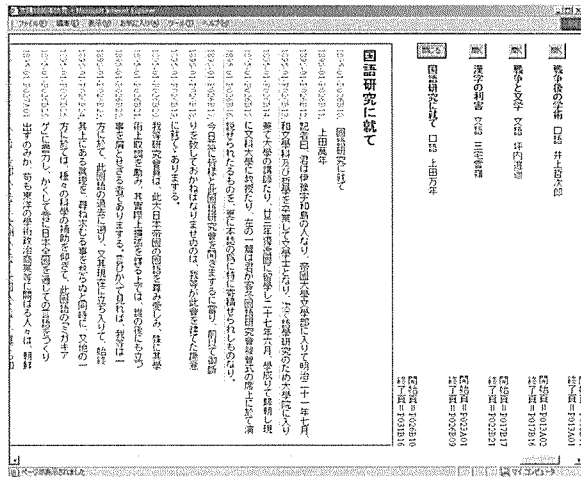


図2 本文-DHTML (行番号) / (②) / Internet Explorer

③本文-シンプルなHTML

どのブラウザでも閲覧できるようにHTMLの基本的なタグだけを使ったHTML本文である。構造化テキストでタグ付けされている情報の多くは省かれている。



図3 本文-シンプルなHTML / (③) / Internet Explorer

④本文－プレーンテキスト

単純な文字列だけのテキストファイルに変換する。構造化テキストでタグ付けされている情報のほとんどが省かれているが、改行タグだけは改行コードとして残されている。図4はファイルを「メモ帳」で開いたところ。

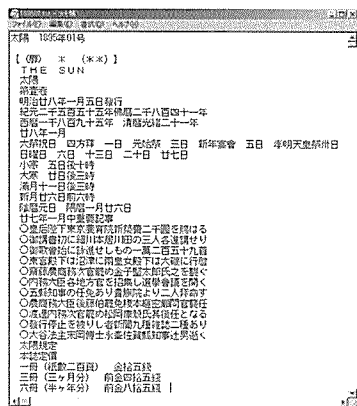


図4 本文－プレーンテキスト／(④)／メモ帳

⑤本文－プレーンテキスト (行番号)

④の本文を原文と同じ位置で改行し、行頭に年・号・ページなどの位置情報を付けたもの。



図5 本文－プレーンテキスト (行番号)／(⑤)／メモ帳

⑥本文－pLaTeX本文

印刷用のpLaTeX形式に変換する。pLaTeXは、構造化テキスト中の情報を紙面に再現するにあたって、現状ではもっとも表現力に富んだ、扱いやすい形式であると思われる。『太陽コーパス』に含まれるフリガナや注、引用などの豊富な情報を、なるべく原文に近く、見やすく再現するためにはたいへん有効である。

この形式のファイルを利用するにあたっては、pLaTeX2eが利用できる環境と、いくつかのpLaTeX用スタイルファイル、特殊なフォント等が必要になる。詳しくは変換によって作られるTeXファイルの冒頭に説明が出力されるようになっている。

図6はTeXファイルを処理することで得られるDVIファイルを、閲覧・印刷用ソフトdviout（注3）で開いた画面。一部分を拡大して示している。

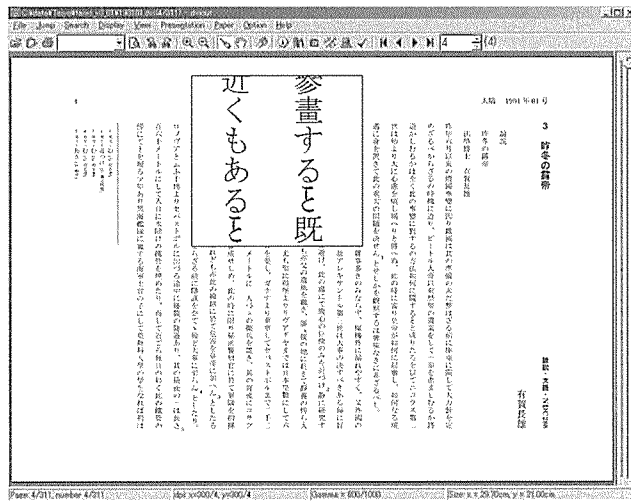


図6 本文－pLaTeX本文／(⑥)／dviout

⑦本文－pLaTeX本文（行番号）

⑥を原文と同じ位置で改行し、行頭に位置情報を付ける形にしたもの。注を脚注にし、改行（論理改行）を矢印で示すなど、他の部分も変更がある。

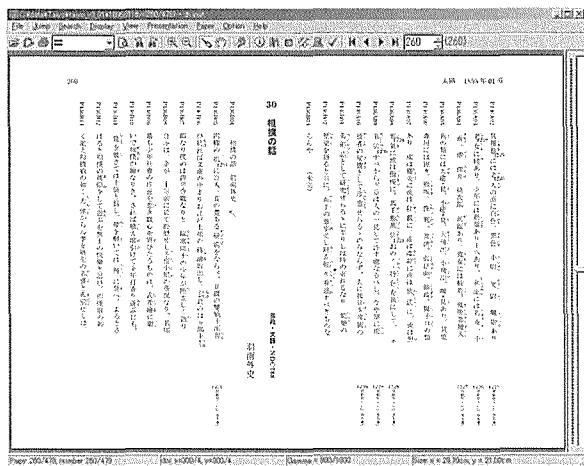


図7 本文—pLaTeX本文（行番号）／(7)／dvipout

3.1.2 情報抽出

構造化テキストの中のタグ付けされた情報を抜き出して、一覧表を作成するためのスタイルシートである。すべて表形式であり、Webブラウザで閲覧するためのHTML形式と、表計算ソフトなどに読み込んで利用するためのCSV形式（コンマ区切りの表形式）が用意されている。

⑧外字一覧（コード順）HTML

録	040921	P175D09 研工研究会の顧問。
間	041262	P001B26 衆議院で眼を開くも閑として静かく。
隄	041740	P109B21 野老呼牛過柳段。
雞	042124	P003B09 祖述といふ人は夜半に雞の鳴をきいて。
雞	042124	P003B11 雞の宵鳴にて世の多事になるを知るとは。

図8 外字一覧（コード順）HTML／(8)／Internet Explorer

⑨外字一覧（出現順）HTML

太陽1895年01号 外字一覧（出現順）

外字 デキスト 文字番号 位置 文脈

熙	熱	001721	F001D02	熙々たる明治二十八年の新春光は至極基なる。皇恩の下に生等をして同胞四千餘萬の愛讀者諸君と共に紙上に掲見を俾せしむ。
磚	=	024597	F002A02	内には浩然たる正気の磅礴するところ禁せんと欲して能はざるあり。
雞	鶏	042124	F003B09	祖述といふ人は夜半に鶏の鳴をきいて。
琨	=	021065	F003B10	同宿の劇裡を呼起し。

図9 外字一覧（出現順）HTML / (⑨) / Internet Explorer

⑩記事情報HTML

太陽1895年01号 記事情報 - Microsoft Internet Explorer

太陽1895年01号 記事情報

年号 No.	タイトル	著者	題名	文庫ジャンル(NDC)	開始位置	終了位置	行番号
1895-01-01	名聞	×	×	文庫 文庫 文庫	F001A01	F001C22	43
1895-01-02	文壇の発展	大橋新次郎	×	文庫 NDC051	F001D01	F002A35	43
1895-01-03	字界の大家新	久米邦武	論説	文庫 NDC002	F003A01	F003B10	294
1895-01-04	戦後の教育	千原清徳	論説	文庫 NDC371	F003B11	F013A01	224
1895-01-05	戦後の文学	井上馨次郎	論説	文庫 NDC002	F013A02	F017B16	250
1895-01-06	戦後の文学	坪内逍遙	論説	文庫 NDC001	F017B17	F022B21	264
1895-01-07	漢字の刊行	三宅雪嶺	論説	文庫 NDC811	F023A01	F026B09	192
1895-01-08	国語研究に就て	上田万年	論説	文庫 NDC810	F026B10	F031B16	267
1895-01-09	事物考證の研究に對する人類學的方法	坪井正五郎	論説	文庫 NDC469	F031B17	F033B12	93
1895-01-10	経済的儲蓄	井上英九郎	論説	文庫 NDC333	F033B13	F037B25	197
1895-01-11	農業教育に就て	横井時敏	論説	文庫 NDC610	F033A01	F041A15	171
1895-01-12	利権の発展	尾崎行雄	論説	文庫 NDC329	F041A16	F047B15	338
1895-01-13	日本帝國の任務	中野重郎	論説	文庫 NDC311	F047B16	F050B20	161
1895-01-14	元寇の奇蹟なる航海	河田忠村	史伝	文庫 NDC209	F051C01	F054A21	166
1895-01-15	オランダローを對する記	戸川幸雄	史伝	文庫 NDC225	F054B01	F066A26	600
1895-01-16	大久保相模守忠雄	福地俊房	史伝	文庫 NDC239	F066B01	F070B25	221
1895-01-17	京都の新築内記	中川四朗	地理	文庫 NDC291	F071A01	F076B26	302
1895-01-18	利根水運新築内記	渡辺千吉郎	地理	文庫 NDC291	F077A01	F082B29	453
1895-01-19	戦艦	尾崎紅葉	小説	文庫 NDC913	F083A01	F089B05	339
1895-01-20	従軍入札	尾崎紅葉	小説	文庫 NDC913	F089B06	F099B04	432
1895-01-21	先時代の権威	幸田露伴	論説	文庫 NDC772	F099C01	F103B23	290
1895-01-22	自然の美	志賀重昂	論説	文庫 NDC701	F109A01	F111A03	111
1895-01-23	露の美	志賀重昂	論説	文庫 NDC701	F111A09	F113B04	39
1895-01-24	日章旗	越日生	論説	文庫 NDC238	F114A01	F115B01	57
1895-01-25	おもしろい	無名氏	論説	文庫 NDC913	F115B02	F117B03	92

図10 記事情報HTML / (⑩) / Internet Explorer

⑪記事情報CSV

年	巻	号	題名	著者	訳者	出版社
1995	1	P001C22	(原)			
1995	1	P002A35	太陽の兵利	大橋新太郎	文芸	文芸
1995	1	P008B10	学問の大軍新	久米邦武	文芸	文芸
1995	1	P013A01	戦時後の教育	千明清臣	文芸	文芸
1995	1	P017B16	戦争後の学問	井上哲次郎	文芸	文芸
1995	1	P022E21	戦争と文学	坪内逍遙	文芸	文芸
1995	1	P026B09	漢学の利権	三宅雪嶺	文芸	文芸
1995	1	P031B16	国語研究に就て	上田万年	文芸	文芸
1995	1	P033B12	事物考証の研究に対する人類学的方法	坪井正五郎	文芸	文芸
1995	1	P037B25	経済的闘争	井上辰九郎	文芸	文芸
1995	1	P041A15	農業教育に就て	後井時敏	文芸	文芸
1995	1	P047B15	対清抗戦	尾崎行雄	文芸	文芸
1995	1	P050B20	日本帝國の任務	中西定郎	文芸	文芸
1995	1	P054A21	紀元前の著なる航海者	森田忠軒	文芸	文芸
1995	1	P066A26	ワートルロー合戦の記	戸川殊花	文芸	文芸
1995	1	P070B25	大久保経緯守忠閣	福地俊房	文芸	文芸
1995	1	P076B26	京都の精華内記	中川四明	文芸	文芸
1995	1	P082B28	利根水原探検紀行	渡辺千吉郎	文芸	文芸
1995	1	P088B05	取説	尾崎行雄	文芸	文芸
1995	1	P099B04	証人失	奥田善村	文芸	文芸
1995	1	P109B23	元時代の補綴	幸田露伴	文芸	文芸
1995	1	P111A08	自然の美	志賀重昂	文芸	文芸
1995	1	P113B04	霧の美	石橋忍月	文芸	文芸
1995	1	P115B01	日暮の美	旭日生	文芸	文芸
1995	1	P117B03	おもしろい	熊谷清	文芸	文芸
1995	1	P119A17	新年	新田武郎	文芸	文芸
1995	1	P126A22	《今昔物語集》	多	文芸	文芸
1995	1	P127B13	《垣歌》	多	文芸	文芸
1995	1	P136B03	能楽(上)	大和田建博	文芸	文芸

図11 記事情報CSV / (⑪) / Excel

⑫注情報HTML

注番号	注対象	注内容
P003E03	其の最後の一は「 <u>医</u> 」	「 <u>医</u> 」は「 <u>医</u> 」の誤り。其の最後の一は「 <u>医</u> 」の誤り。
P005E06	以上は即ち露國皇帝が清國問題に於て適用する所の商約にて又其の <u>伏</u> 「平和の理想と相違れざるものに非ず。	「 <u>伏</u> 」は「 <u>伏</u> 」の誤り。以上は即ち露國皇帝が清國問題に於て適用する所の商約にて又其の「 <u>伏</u> 」は「 <u>伏</u> 」の誤り。
P006A23	砂磧「 <u>板</u> 可ならん。	「 <u>板</u> 」は「 <u>板</u> 」の誤り。砂磧「 <u>板</u> 」は「 <u>板</u> 」の誤り。
P009E01	直く或明に結結する所の <u>雄大なる量</u> 「 <u>雄大</u> 」なからんべからず。	「 <u>雄大</u> 」は「 <u>雄大</u> 」の誤り。直く或明に結結する所の「 <u>雄大</u> 」は「 <u>雄大</u> 」の誤り。
P009E03	根本的 <u>の</u> 「 <u>雄大</u> 」も忘却するが如きこと。	「 <u>雄大</u> 」は「 <u>雄大</u> 」の誤り。根本的「 <u>雄大</u> 」も忘却するが如きこと。
P010A17	龍、 <u>雄大</u> の量「 <u>雄大</u> 」。	「 <u>雄大</u> 」は「 <u>雄大</u> 」の誤り。龍、 <u>雄大</u> の量「 <u>雄大</u> 」。
P010B11	前代未聞の増進を <u>得</u> 「 <u>得</u> 」。	「 <u>得</u> 」は「 <u>得</u> 」の誤り。前代未聞の増進を「 <u>得</u> 」。
P010B21	當節は <u>雄</u> 、其の <u>量</u> を増進「 <u>雄</u> 」。	「 <u>雄</u> 」は「 <u>雄</u> 」の誤り。當節は「 <u>雄</u> 」、其の「 <u>量</u> 」を増進「 <u>雄 </u>
P011A05	以て <u>期</u> の衰微を降去することを得べしとして「 <u>期</u> 」。	「 <u>期</u> 」は「 <u>期</u> 」の誤り。以て「 <u>期</u> 」の衰微を降去することを得べしとして「 <u>期 </u>
P011E01	北米雄洲等の地方に移住を試むるに因て「 <u>北</u> 」。	「 <u>北</u> 」は「 <u>北</u> 」の誤り。北米雄洲等の地方に移住を試むるに因て「 <u>北 </u>
P019A10	そが十九世紀の量詞を自叙し初めたる當代青年の中に <u>量</u> の味方「 <u>量</u> 」得しもの。	「 <u>量</u> 」は「 <u>量</u> 」の誤り。そが十九世紀の量詞を自叙し初めたる當代青年の中に「 <u>量</u> 」の味方「 <u>量 </u>
P021B16	是れ <u>量</u> に「 <u>量</u> 」を得ざる宣言なりき。	「 <u>量</u> 」は「 <u>量</u> 」の誤り。是れ「 <u>量</u> 」に「 <u>量 </u>
P022A02	「 <u>量</u> 」を得ざる也。	「 <u>量</u> 」は「 <u>量</u> 」の誤り。「 <u>量 </u>
P022A15	是の <u>期</u> は <u>量</u> に <u>量</u> を「 <u>量</u> 」。	「 <u>量</u> 」は「 <u>量</u> 」の誤り。是の「 <u>期</u> 」は「 <u>量</u> 」に「 <u>量量 </u>
P022B17	「 <u>量</u> 」を得ず量詞に據りて <u>量</u> を <u>量</u> 。	「 <u>量</u> 」は「 <u>量</u> 」の誤り。「 <u>量量量 </u>
P025A12	「 <u>量</u> 」の <u>量</u> 「 <u>量</u> 」。	「 <u>量</u> 」は「 <u>量</u> 」の誤り。「 <u>量量 </u>

図12 注情報HTML / (⑫) / Internet Explorer

⑬注情報CSV

	A	B	C	D	E	F	G	H	I	J
1	1905	1	P001C00	●本誌ハ前金頃別にアラブレハ一切録述セズ	原文=ワレ	F欄点見				
2	1905	1	P001C00	●本誌ハ前金頃別にアラブレハ一切録述セズ	原文=ス	F欄点見				
3	1905	1	P001D00	世に半を得ず本年を病、たしん	原文=世に	A欄字通用				
4	1905	1	P000A01	世間を疾くもに非ずんば疾くとも已まざる人ぞとす。!	原文=世に	A欄字通用				
5	1905	1	P003A02	而も時事に過ぎざれば非て當世に感ず可らず。	原文=世に	F欄点見				
6	1905	1	P003A02	而も時事に過ぎざれば非て當世に感ず可らず。	原文=世に	F欄点見				
7	1905	1	P003A02	而も時事に過ぎざれば非て當世に感ず可らず。	原文=世に	F欄点見				
8	1905	1	P003A02	而も時事に過ぎざれば非て當世に感ず可らず。	原文=世に	F欄点見				
9	1905	1	P003A03	而も時事に過ぎざれば非て當世に感ず可らず。	原文=世に	F欄点見				
10	1905	1	P003B15	大抵は甲兵を用ふ。	原文=用ひ	G欄点見				
11	1905	1	P003B16	原野に陣すといへり。	原文=陣す	F欄点見				
12	1905	1	P003B16	其處に犯人なければ用ふる所なし。	原文=用ふる	G欄点見				
13	1905	1	P003C01	兵馬動る處をのみ用ふ。	原文=用ひ	G欄点見				
14	1905	1	P003C18	などぞ。	原文=と	F欄点見				
15	1905	1	P004A12	只靜筆の語を以て之を採し。	原文=静筆	A欄字通用				
16	1905	1	P004B03	筆勢とは用ひに於て、世	原文=。	A欄字通用				
17	1905	1	P004B11	其處は静筆を採し、世	原文=。	A欄字通用				
18	1905	1	P005A00	筆勢もの、結語なりし。	原文=結語	A欄字通用				
19	1905	1	P005A06	現物な實情の筆勢を採し。	原文=筆勢	A欄字通用				
20	1905	1	P005B11	入流の世も、世	原文=。	A欄字通用				
21	1905	1	P006A01	今度の世では其筆勢を採し、世	原文=筆勢	A欄字通用				
22	1905	1	P006A05	只今に風俗轉る、世	原文=風俗	A欄字通用				
23	1905	1	P007B11	只今に風俗轉る、世	原文=風俗	A欄字通用				
24	1905	1	P008A01	只今に風俗轉る、世	原文=風俗	A欄字通用				
25	1905	1	P008A06	只今に風俗轉る、世	原文=風俗	A欄字通用				
26	1905	1	P007B20	此の如くならざれば、世	原文=世に	F欄点見				
27	1905	1	P008A16	一世二世三世にて筆勢を採し、世	原文=。	A欄字通用				
28	1905	1	P008A20	然るに今度の世にて正體の筆勢を採し、世	原文=筆勢	A欄字通用				
29	1905	1	P008A24	而も世を著し、世	原文=世に	F欄点見				
30	1905	1	P008B24	而も世を著し、世	原文=世に	F欄点見				

図13 注情報CSV / (⑬) / Excel

⑭引用情報HTML

年号	取記事の題名	取記事の著者	引用の位置	引用の種別	引用の証者	取引用の種別	取引用の証者	引用文
1901-01	明治三十四年	x	P001C07	典拠	孔子			可も道を語り、道を引ばず、
1901-01	昨今の露帝	有賀長雄	P004E06	典拠	『倫敦タイムズ』通信員			是れ露國の滿洲に關、列國聯合以外に於て別國と約束する所あらざるものなり
1901-01	昨今の露帝	有賀長雄	P005A04 ~P005B07	典拠	『タイムズ・オブ・ロンドン』			『タイムズ』及其の雑誌は今日よりアジアに消在せり。……露國政府の作態に於て此の原則に照して變に因り始めて理解すべきもの少からず
1901-01	学政振振と財政	久保田談	P009A06 ~P009A10	典拠	インチヤナ州憲法			金社會二智識・學問・ラ蓄セシムルハ自由ノ政治ヲ保全スルニ必要ナリ故ニ遺言ノ方法ヲ以テ道德智識學術農業ヲ獎勵シ州内一區ナル小學校ノ制度ヲ設ク 英國ニテ中等二教育ヲ受クルコトヲ望シシムルハ則チ然ラズ
1901-01	学政振振と財政	久保田談	P009A11 ~P009A17	典拠	南方カリナ州憲法			州憲法ハ此憲法施行ノ後當テ爾後二テ公立學校維持ノ爲メ金州ノ征伐ニ當テ種族ヲ課シ他ノ州與テ其同一ノ教育ヲ以テ職業シ本州ノ會計局ニ納ムベシ……但シ教育ノ目途ニ使用スルノ外ニ他ノ分額報引課スベカラズ
1901-01	学政振振と財政	久保田談	P009A18 ~P009A20	典拠	アルカソス州憲法			人類ニ依テ種族ヲ課スルハ則チ然ラズナリ
1901-01	学政振振と財政	久保田談	F010A13	典拠	政黨會			今や政治に社會に一大改革を行ひ、
1901-01	学政振振と財政	久保田談	F010A23	典拠	政黨會			加護金公使顧問に久しく駐在公使として遠征に於る將領の閣

図14 引用情報HTML / (⑭) / Internet Explorer

⑮引用情報CSV

A	B	C	D	E	F	G	H	I	J	K	L	M
1	1995	1. 学問の大系	久米邦武	P003A02~P003A03	記事説明	記者				時局を知るは(佐保)の事也。而も時局に造せざる		
2	1995	1. 学問の大系	久米邦武	P003B03~P003B06	記事説明	記者				記者曰。吾字は士族易堂と號す。……其史の		
3	1995	1. 学問の大系	久米邦武	P003B10	会誌	清道				足は登壇にあらず		
4	1995	1. 学問の大系	久米邦武	P003C10~P003C17	心緒	世人				日本は清浄の國なり。 廉潔の民なり。……		
5	1995	1. 学問の大系	久米邦武	P004B26	会誌	徳川家康				光武は高祖時代の孫なり。や。反復の事は		
6	1995	1. 学問の大系	久米邦武	P005A02	会誌	徳川家康				年寄きもの。能記。 傳はたり		
7	1995	1. 学問の大系	久米邦武	P006A09	漢文	『史記』				殺人者死。傷人。 及盜者抵罪		
8	1995	1. 学問の大系	久米邦武	P006B10	漢文	『孟子』				無恒産而有恒心者惟士為能。 吾民則無恒産		
9	1995	1. 学問の大系	久米邦武	P007A11	曲説	足利義満				改輩は士民の所を懸んずべし		
10	1995	1. 学問の大系	久米邦武	P007B06	漢文	『孟子』				天下之本在國。國之本在家		
11	1995	1. 学問の大系	久米邦武	P007B06	漢文	『孟子』				天下國家身心意和		
12	1995	1. 学問の大系	久米邦武	P007B11	漢文	『礼記』				父母尊。不。 有私財		
13	1995	1. 戦後の教育	千鶴清臣	P008B13~P008B15	記事説明	記者				記者曰。吾は昭和三年一月一日を以て土佐高		
14	1995	1. 戦後の教育	千鶴清臣	P009A16	曲説	ビスマルク				我祖の遺訓の今日を致せるはモルトケ將軍の		
15	1995	1. 戦後の教育	千鶴清臣	P009B03~P009B06	会誌	陸軍第一師団長				山崎進教育の根柢。 海軍人の志を勵むるもの		
16	1995	1. 戦後の教育	千鶴清臣	P011B11	会誌	芥川				今世紀の親はるの利に非ずして利の利の親		
17	1995	1. 戦後の教育	井上登次郎	P013A04~P013A11	記事説明	記者				記者曰。吾は親科と號し。……本篇は選紀		
18	1995	1. 戦後の教育	井上登次郎	P014B03~P014B10	漢文	『希書』				地薄者大物不産。水渾者大魚不游。 相先者		
19	1995	1. 戦争と文学	坪内逍遙	P017B10~P017B21	記事説明	記者				記者曰。吾は道徳又孝の命の號あり。……		
20	1995	1. 戦争と文学	坪内逍遙	P018A05	曲説	『孟子』				兵は國の大事。死生の地。……存亡の道		
21	1995	1. 戦争と文学	坪内逍遙	P018A10~P018A17	心緒	世人				他はしばらくは。ひとりの文學藝術は太平の國		
22	1995	1. 戦争と文学	坪内逍遙	P018B06	曲説	俗解1				「戦争とは自然を奉る軍の清濁」		
23	1995	1. 戦争と文学	坪内逍遙	P018B06	曲説	俗解2				「平和を期する軍の清濁」		
24	1995	1. 戦争と文学	坪内逍遙	P018B06	曲説	俗解3				「平和の爲に防戦を争ふ」		
25	1995	1. 戦争と文学	坪内逍遙	P018B07	漢文	*				「聖王は兵を以て争ふ」		
26	1995	1. 戦争と文学	坪内逍遙	P018B09~P018B10	曲説	仏人ゴーチン				「戦争は聖賢の」		
27	1995	1. 戦争と文学	坪内逍遙	P018B12~P018B19	曲説	モーア『無何有島』				「無何有[ノースーピア]の島民は長寿を慕み		
28	1995	1. 戦争と文学	坪内逍遙	P018B20~P018A01	曲説	モーア				「無何有[ノースーピア]の島民は長寿を慕み		
29	1995	1. 戦争と文学	坪内逍遙	P019A12	曲説	バウクル				「世の文明の進むや戦争は漸く消滅せん」		

図15 引用情報CSV / (⑮) / Excel

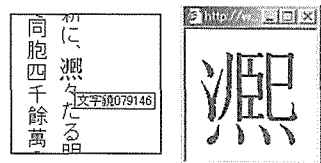
3.1.3 HTML変換時の本文の書式

これまでに説明したXSLTスタイルシートによって構造化テキストをHTML形式に変換してブラウザで閲覧する場合、どのスタイルシートを使ったときにも以下に示すような書式で表示される。これはそれぞれのスタイルシートからhtml_mod.xmlという共通のスタイルシートを呼び出す形で実現している。

検索アプリケーション『たんぽぽ』や全文検索システム『ひまわり』の検索結果表示でも同様の書式が使われている。

● 外字 [図16]

画像を使って表示する。画像をクリックすると別ウィンドウに大きな画像で表示する ([図16] 右)。本文の横に「文字鏡079146」とあるのは画像をマウスでポイントしたときに表示される『今昔文字鏡』(注4)の番号である。



[図16]

● 注 [図17]

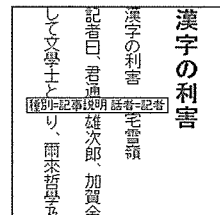
青色の文字で表示される。マウスでポイントするとポップアップで注の内容を表示する ([図17] 左)。クリックすると同内容をメッセージボックスで表示する ([図17] 右)。



[図17]

● 引用 [図18]

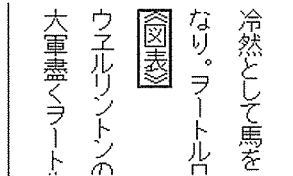
薄青色の背景で表示される。ポイントするとポップアップで引用の出典・話者を表示する ([図18])。クリックすると同内容をメッセージボックスで表示する。



[図18]

● 非入力対象 [図19]

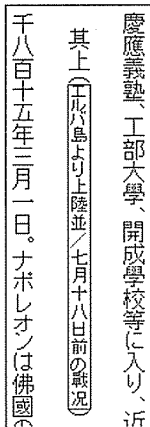
枠内に「《》」に入った文字を囲んで表示する。



[図19]

● 割書 [図20]

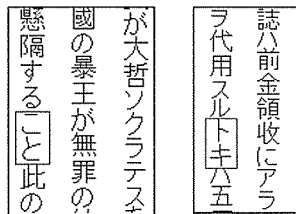
やや小さな文字で、枠内に文字を囲んで表示する。割書の中の改行位置は「/」で示す。



[図20]

● 合字 [図21]

文字を灰色の線で囲んで表示する。



[図21]

この他、「ふりがな (ルビ)」「上付」「下付」「小書」は、HTMLのタグによってそれぞれルビ、上付き文字、下付き文字、小さな文字として表示する。

なお、「注」や「引用」が入れ子になった場合には、ポイントしたときのポップアップ表示ではもっとも内側につけられているものが表示される。クリックすると内側の情報から順にメッセージボックスで表示する。

例：「悉く書を信ぜば書なきに如かず」

(下線部が引用で、傍点部に注がついているとき)

- ・傍点部をマウスでポイントすると「修正注：原文＝信せ 分類＝F濁点脱落」と表示する。
- ・傍点部をクリックすると「修正注：原文＝信せ 分類＝F濁点脱落」とメッセージ。OKを押すとさらに「引用：種別＝典拠話者＝孟軻」とメッセージを表示する。

3.2 変換・情報抽出用アプリケーション『プリズム』

XSLTスタイルシートをXMLによる構造化テキストに適用する方法はいくつかあるが、必要に応じて両者を自在に関連づけることができると便利である。『プリズム』はこれを行うためのアプリケーションである。

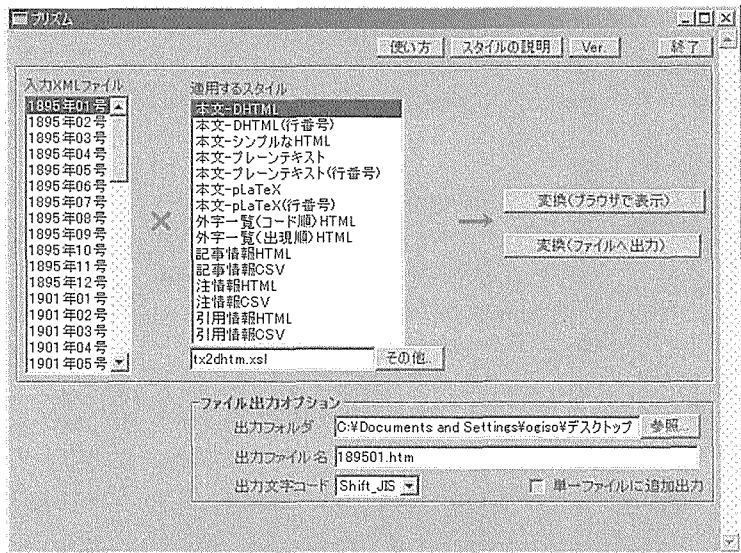


図22 XSLT変換アプリケーション『プリズム』

3.2.1 利用方法

変換結果を画面表示する場合

1. Prism.htaファイルをダブルクリックしてプログラムを起動する。
2. 左側のセレクトボックスから変換元となる「入力XMLファイル」を選択する。
3. 中央の「適用するスタイル」から、目的に応じたHTMLと書かれているものを一つ選択する（注5）。
4. [変換（ブラウザで表示）] ボタンを押す。

まもなくInternet Explorerのウィンドウが開き、変換結果が表示される。

図23はこの手順にしたがって1895年12号の「外字一覧（コード順）」をブラウザで表示する例である。この際、画面下部の「ファイル出力オプション」は無関係なので一切変更する必要はない。

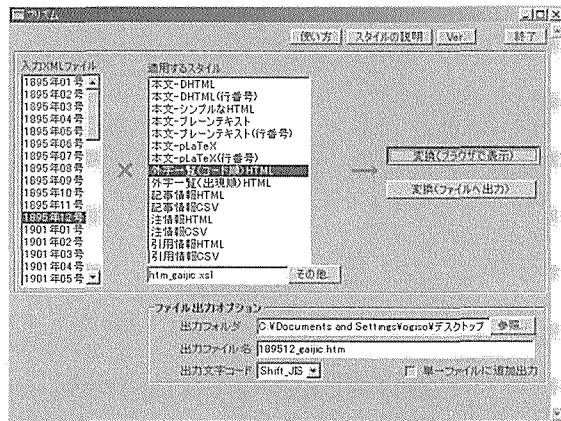


図23 1895年12号の外字一覧（コード順）をブラウザで表示する例

変換結果をファイルに出力する場合

1. Prism.htaファイルをダブルクリックしてプログラムを起動する。
2. 左側のセレクトボックスから変換元となる「入力XMLファイル」を選択する。「ctrl」キー（連続した範囲であれば「Shift」キー）を押しながら左クリックすることで複数選択できる。
3. 中央の「適用するスタイル」を一つ選択（どの形式でも可）。

4. 必要に応じて出力ファイル名などを変更する。（「入力XMLファイル」を複数選択した場合には出力ファイル名は変更できず、既定の名前となる）
5. [変換（ファイルへ出力）] ボタンを押す。

図24は1909年1号の「プレーンテキスト」を「Taiyo190901.txt」という名前のファイルとしてデスクトップに出力する例である。出力ファイル名を変更している。

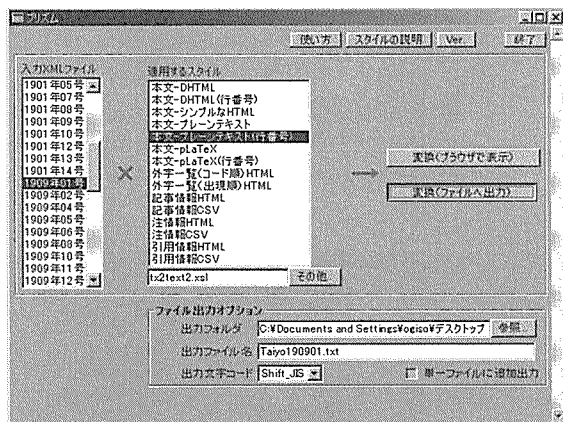


図24 1909年1号のプレーンテキストを「Taiyo190901.txt」という名前で出力する例

「その他..」ボタンによる自作スタイルシートの利用

「適用するスタイル」の下の「その他..」ボタンを押すことで、3.1で紹介した標準スタイルシート以外の、自作のスタイルシートを適用することができる（3.2.3参照）。ボタンを押すとファイル指定のダイアログボックスが現れるので、ここでXSLTスタイルシートを指定する（図25）。その後の操作手順は上記3.以降と同様である。ただし、ファイル出力時にはスタイルシートの仕様にあわせて、出力ファイル名の拡張子を適宜変更する必要がある。

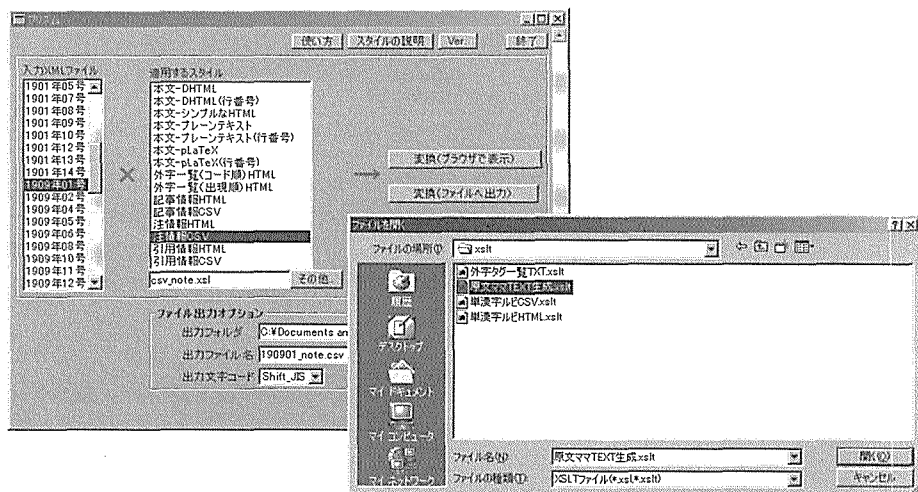


図25 自作スタイルシートの適用

ファイル出力時のオプション

変換した結果をファイルに出力する際には次の出力オプションを指定することができる。

● 「単一ファイルに追加出力」オプション

入力XMLファイルを複数選択したとき、「単一ファイルに追加出力」チェックボックスをオンにすると変換結果が一つのファイルに追加出力される。CSV形式のファイルをまとめて一つの表として出力する場合などに有効である。ただし、単純に追加出力するものであるため、HTML形式の場合にはヘッダ部が何度も出力されることになり、正しいHTML形式にはならないので注意が必要である。

このオプションをオンにした場合には複数ファイル選択時でも出力ファイル名を変更することができる。出力先のフォルダにすでに同名のファイルが存在する場合には、既存のファイルに追加する形で出力される（新しいデータは古いデータの後に次々と追加されてゆく）。


● 「出力フォルダ」の設定

標準では「デスクトップ」が指定されているが、「参照」ボタ

ンを押して選択することで変更できる。ただし、必ず書き込み可能なフォルダを指定する必要がある。この欄を空欄にすると「入力XMLファイル」が存在するフォルダに出力される。

- 「出力文字コード」の設定

ファイルに出力する場合には、出力文字コードとしてシフトJIS・ユニコード (UTF-16) のどちらかを選択できる。TeX形式の場合にはシフトJISを選択するなど出力したファイルを利用するアプリケーションの制限にあわせて適切な文字コードを選択する。

なお、出力文字コードをシフトJISにすると、著者名・引用の出典など一部の属性値に含まれるJIS外字 (ユニコード数値による文字参照 (�形式) で表されていたもの) が「」に置き換えられる。

その他のボタン類

画面の「使い方」「スタイルの説明」をクリックすると、操作方法、各XSLTスタイルシートについての簡単な説明がそれぞれ表示される。また「Ver.」でバージョン情報を表示、「終了」でプログラムを終了する。

3.2.2 自作XSLTスタイルシートの作成と利用

『プリズム』では、「その他..」ボタンを使うことで、自作したXSLTスタイルシートを『太陽』のXMLファイルに適用することができる。XSLTについては多くの参考図書が出版されている。『太陽コーパス』のXMLデータの構造を理解した上で、これらを参考にして自作のスタイルシートを作成することでデータ活用が大きく広がることになる。

全く新しいスタイルシートを作らなくとも、付属のXSLTスタイルシートに簡単な改造を施して利用することもできる。たとえば、次に掲げるのはテキストファイルに変換するスタイルシート tx2text.xml (3.1.1④参照) の中身である (左の行番号と右の説明文はスタイルシートには含まれない)。

1	<?xml version="1.0" encoding="Shift_JIS"?>	XML宣言
2	<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0" xmlns:tx="http://www.kokken.go.jp/taiyo" exclude-result-prefixes="tx">	XSLTスタイルシートの開始、『太陽コーパス』の名前空間の宣言など
3	<xsl:output method="text" omit-xml-declaration="yes"/>	出力形式の指定(テキスト)
4	<xsl:template match="/"><xsl:apply-templates select="tx:雑誌"/></xsl:template>	文書の「雑誌」を処理
5	<xsl:template match="tx:雑誌"><xsl:value-of select="@雑誌名"/> <xsl:value-of select="@年"/>年 <xsl:value-of select="@号"/>号 <xsl:text>
</xsl:text> <xsl:apply-templates/></xsl:template>	「雑誌」にマッチしたら雑誌名・年・号・改行を出力
6	<xsl:template match="tx:記事"> 【<xsl:value-of select="@題名"/> <xsl:value-of select="@著者"/> (<xsl:value-of select="@欄名"/>)] <xsl:text>
</xsl:text> <xsl:apply-templates/><xsl:text>
</xsl:text> </xsl:template>	「記事」にマッチしたら題名・著者・欄名と改行を出力
7	<xsl:template match="tx:br"><xsl:text>
</xsl:text> </xsl:template>	「br」にマッチしたら改行を出力
8	<xsl:template match="*"><xsl:apply-templates/></xsl:template>	その他の要素にマッチしたら中身のテキストを出力
9	</xsl:stylesheet>	スタイルシートの終わり

このスタイルシートを改造して、ルビ入りテキストの形で出力するためには7行目と8行目の間に次のように書き足せばよい。これは「r」要素（ルビ）が見つかった場合、その中身のあとに、rt属性を括弧に入れて出力することを意味する。

	<xsl:template match="tx:r"><xsl:apply-templates/> [<xsl:value-of select="@rt"/>] </xsl:template>	「r」（ルビ）にマッチしたら中身のテキストの後にrt属性を [] に入れて出力
--	--	---

したがって次のような処理が行われ、図26に示すような結果が得られることになる。

例：<r rt="き">黄</r><r rt="いろ">色</r>い<r rt="たんぼ">ぼ</r>蒲公英</r>の花
→黄 [き] 色 [いろ] い蒲公英 [たんぼぼ] の花



図26 改造したtx2text.xslによるフリガナ付き記事の出力例

同様に、次のように書き換えることでルビを本文としたテキストの形で出力することができる。

<pre><xsl:template match="r"><xsl:value-of select="@rt"/> </xsl:template></pre>	<p>「r」（ルビ）にマッチしたらrt属性だけを出力</p>
---	--------------------------------

この場合、次のような処理が行われる。

例：<r rt="き">黄</r><r rt="いろ">色</r>い<r rt="たんぼ">ぼ</r>蒲公英</r>の花
→きいろいたんぼぼの花

このほか、たとえば注一覧のスタイルシート csv_note.xsl (2.1.1 ②参照) を改造し、本書に収録した「仮名遣いについて」(小木曾智信, 351頁から) で利用したXPath式を用いて本文中の仮名遣いに関する注だけを抜き出すためには、次のような変更

を行えばよい。

(元のスタイルシート・10行目)

```
<xsl:for-each select="//tx:注">
```

(変更後)

```
<xsl:for-each select="//tx: 注[./@分類='G仮名遣'] [not  
(contains(./@原文, ' '))] ">
```

4

検索用アプリケーション『たんぽぽ』

『たんぽぽ』は、『太陽』の構造化テキスト本文を直接検索するためのアプリケーションである。『ひまわり』とは違って、インデックスを用意しないgrepタイプの検索なので時間がかかるものの、柔軟な検索方法の指定や、豊富な情報を含む検索結果表示が可能になっている。

『たんぽぽ』は『太陽』の構造化テキストを、その構造にしたがって「記事」ごとにたどってゆく。そして「s」要素（文）を一つずつ読み込んで、そこから「検索対象のテキスト」を生成し、そのテキストの中に検索文字列がないか調査してゆく。「検索対象のテキスト」は、あらかじめ用意しておくのではなく、その場で生成しているので「ルビを開いたテキスト」「踊り字を開いたテキスト」などの形式を指定することができる。

検索文字列に一致する部分がある「s」を見つけると、その「s」を含んでいる「記事」の情報を集め、前後の「s」とあわせて、検索結果として表示する。この検索結果も構造化テキストから直接、動的に生成している。このような仕組み上、検索に時間がかかるため、全号を通しての検索は実用的ではないが、他の方法では検索が難しい例に対処できるなどの利点もある（4.3参照）。



図27 検索用アプリケーション『たんぽぽ』

4.1 利用方法と検索オプション

基本的な利用方法は次の通りである。

- ① Tanpopo.htaファイルをダブルクリックしてプログラムを起動する
- ② 画面中央に並ぶチェックボックスの検索したい号をチェックする
- ③ 「検索文字列」テキストボックスに検索文字列を入力する
- ④ [検索] ボタンを押す

この手順のなかで、以下のような機能を利用することができる。

◆正規表現

③の検索文字列には正規表現を利用することができる。表3に示すような多くの正規表現（注6）が使用可能である。

画面上部の[正規表現について] ボタンを押すことでこれに関する説明を見ることができる。

表3 『たんぽぽ』で利用可能な正規表現

[あいう]	文字クラス（括弧内の文字いずれか）
[あ-ん]	〃 （範囲指定（注7）・ひらがな1文字）
[^あい]	文字クラスの補集合（括弧内の文字以外の文字）
.	任意の1文字
?	直前の文字（正規表現）の0回または1回の繰り返し
*	直前の文字（正規表現）の0回以上の繰り返し
+	直前の文字（正規表現）の1回以上の繰り返し
あ い	論理和（「あ」または「い」）

表4は正規表現を実際に利用する場合の具体例である。

表4 正規表現の使用例

正規表現	対応する文字列
走[つらりるれろ]	「走ら」「走り」など（ラ行五段活用）
国語?研	「国研」または「国語研」
全然[^.] *ない。	（句点付きの記事で）「全然」を含み「ない。」で終わる文
それを 其れ?を	「それを」「其れを」「其を」

◆字体変換

検索手順③の後で「字体変換」ボタンを押すと、入力してある検索文字列中の漢字が、置き換え可能な字体（等価字体）に変換される。『太陽コーパス』で2種類の字体が使われている場合には正規表現の文字クラスを用いた形に変換される。

例： 亜→亞 芦→[芦蘆] 圧→壓

検索文字列として、漢字を含む文字クラスを入力している場合、このボタンを押すと、文字クラスの指定が二重になり、正規表現が正しくない形になることがあるので注意が必要である。

なお、画面上部の「字体変換について」ボタンを押すことで字体変換に関する説明を見ることができる。

◆「検索対象のテキスト」の指定と「踊字をひらく」オプション

検索時のオプションとして検索の対象となるテキストの形式を指定することができる。「検索対象テキストの形式」セレクトボックスでは、「ルビなしテキスト」「ルビ入りテキスト」「ルビを開いたテキスト」のいずれかを指定することができる（表5参照）。

ただし、ルビがついていない部分ではどの形式を選択しても「ルビなしテキスト」を対象に検索することになる。「ルビ入りテキスト」のルビは原則として漢字一字ごとについており、ルビを囲む括弧は全角の大括弧「[]」となっている。

「踊字をひらく」オプションをオンにすると、検索対象テキスト中の踊字（くの字点やゝ、ゞ、ゞ等）を対応する文字に置き換える。これは「ルビを開いたテキスト」の場合にも有効である。

「検索対象テキストの形式」と「踊字をひらく」の組み合わせで、ルビ付き部分の検索対象のテキストは、表5の6種類があることになる。

表5 各種検索対象テキストの例

原文「心のまゝ」(<rt="こゝろ">心</rt>のまゝ)の場合

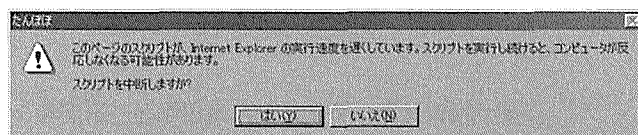
	ルビなしテキスト	ルビ入りテキスト	ルビをひらいたテキスト
踊り字をひらくオフ	心のまゝ	心 [こゝろ] のまゝ	こゝろのまゝ
踊り字をひらくオン	心のまま	心 [こころ] のまま	こころのまま

◆その他のボタン類

画面の「使い方」「正規表現」「字体変換について」をクリックすると、それぞれについての簡単な説明が表示される。また「Ver.」でバージョン情報を表示、「終了」でプログラムを終了する。

◆一括検索時の注意事項

一度に検索する対象が多すぎる場合に、次のようなメッセージが表示される場合がある。



このとき、時間がかかってもさらに検索を続けたい場合には「いいえ」、検索を中断してやり直す場合には「はい」をクリックする。このようなメッセージが頻繁に現れる場合には、検索対象の

記事を減らしてやり直す必要がある。

4.2 検索結果とその利用

検索結果は、見つかった件数とともに図28のように表示される。

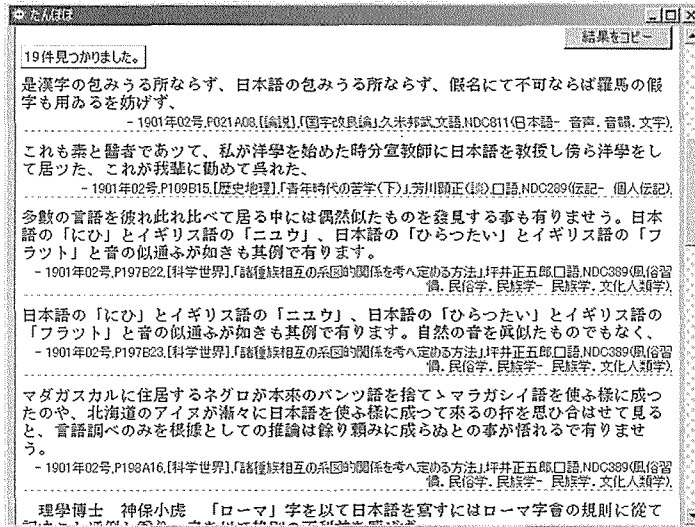


図28 検索結果例 (1901年の「日本語」の用例を検索)

ここに表示される検索結果の本文の表示形式は、構造化テキストに含まれる情報をできるかぎり活かしたものとなっており、別途本文を参照するまでもなく、フリガナや注、外字、引用などの情報を見ることができる。書式は「DHTML本文」XSLTスタイルシートや、『ひまわり』のHTML本文表示と共通である (2.1.3 HTML変換時の本文の書式 参照)。ただし、検索語を切り出してKWICのような形式で表示することはできないなどの制限がある。

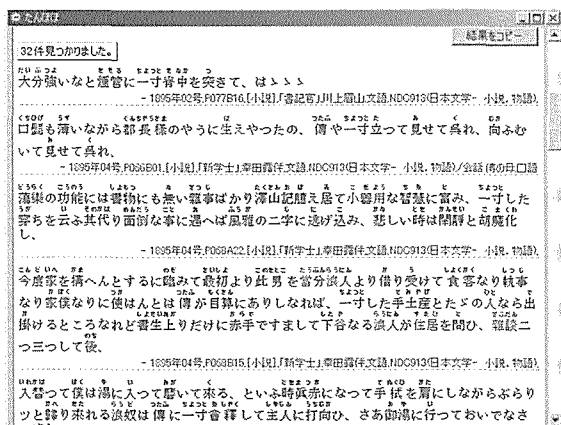


図29 ルビつき部分の検索結果例 (1895年の「ちよつと」の用例を検索)

◆出典情報

本文の下の情報には次の順で並んでいる。最後の「/」以降は検索結果が引用部分にあたる場合にのみ表示される。ジャンルの部分をマウスでポイントまたはクリックするとNDCの二次区分・三次区分が表示される。

年・号	ページ位置	[欄名]	「タイトル」
著者名	記事の文体	ジャンル	/引用種別:話者 引
用文体			

例:

允長老信長老清原極の前にて、光武は高祖幾代の孫なるや、反魂香の事は何書に出たるや、 - 1895年01号,P004B26, [論説], 「学界の大革新」, 久米邦武文語, NDC002, /会話:徳川家康
口髭も薄いながら郡長様のやうに生えやつたの、傳や一寸立つて見せて呉れ、向ふむ
いて - 1895年04号,P066B01, [小説], 「新学士」, 幸田露伴文語, NDC913, /会話:傳の母, 口語

◆検索結果のコピー

検索結果とともに「結果をコピー」ボタンが表示される。このボタンを押すと、検索結果がタブ区切りのテキストデータとしてクリップボードにコピーされるので、そのまま表計算ソフトなどに表として貼り付けて利用できる (図30)。

[illegible]

図30 検索結果を表計算ソフト（Excel）に読み込んだ例

検索結果部分をマウスで選択してコピーし、他のアプリケーションに貼り付けて利用することもできる。この場合は貼りつけられる側がHTML形式に対応していればルビなどを含む形で、そうでなければテキストデータとして貼り付けられる(図31)。

[illegible]

図31 検索結果をコピーして表計算ソフト（Excel）に貼り付けた例

4.3 『たんぽぽ』での検索が有効な例

検索用アプリケーション『たんぽぽ』は、検索に時間がかかるものの、検索対象テキストを柔軟に設定できる。そのため、全文検索システム『ひまわり』では見つけられないような場合でも検索が可能になることがある。以下、そのような例を挙げておく。なお、用例中の下線はわかりやすくするために補ったものである。

◆ルビと本文をつなげた文字列を検索する場合

例：「ひとり」の表記のバリエーション

検索対象テキストの形式：ルビをひらいたテキスト

検索文字列：「ひとり」

他はしばらく措く、ひとり文学藝術は太平謳歌の餘業、静穩なる民心の反映、

- 1895年01号,P018A10, [論説], 「戦争と文学」, 坪内逍遙, 文語, NDC901, /心話: 世人
荒誕だと思ふなら聞き給ふな。僕は單獨で話をする。」「單獨で話をするとは、覺悟を
極めたね。 1895年01号,P086A02, [小説], 「取舵」, 尾崎紅葉, 文語, NDC913, /会話: 学生, 口語
己が何を嫌がらせるものか。お前が獨りで嫌がつて居るのだ。それは最う綱雄は實に
此上もない男さ。

- 1895年02号,P078A25, [小説], 「書記官」, 川上眉山, 文語, NDC913, /会話: 善平, 口語
尚口輕に, 私も一人でのそ〜歩いて直に飽きて仕舞つて詰らんで, 相手欲しやと
思つて居た處へこゝにお出なさつたのは貴娘の因果といふもの,

- 1895年02号,P079A25, [小説], 「書記官」, 川上眉山, 文語, NDC913, /会話: 辰彌, 口語

『ひまわり』では、「單獨」「一人」のような熟字訓は「ひとり」をキーとしてルビを検索することで見つけられるが、「獨り」のように送り仮名込みの例は検索できない。また仮名書きされた「ひとり」はルビではなく本文を検索し直す必要がある。

◆複数の文字にまたがるルビを検索する場合

例：「ほんたう」(本当)の表記のバリエーション

検索対象テキストの形式：ルビをひらいたテキスト

検索文字列：「ほんたう」

これからはわたしも、父様の仰有つた事を眞實にしないからよう御座んす。一躰父様は私を其様に可愛がつて下さらないわ。

- 1895年02号,P078A13, [小説], 「書記官」, 川上眉山, 文語, NDC913, /会話: 光代, 口語
やア訝しいなア, 眞個か, 偽だらう,

- 1895年03号,P058B09, [小説], 「吾妻錦絵」, 須藤南翠, 文語, NDC913, /会話: 紅顔児, 口語
むつと顔をして見せるに, 野澤さんは本當にどうか遊していらつしやる, 何がお氣に障りましたの - 1895年05号, P118B02, [小説], 「ゆく雲」, 樋口一葉, 文語, NDC913, /会話: お縫, 口語
此間新聞に出て居たが, あれはほんたうか。(乙)

- 1901年12号, P105A20, [小説雑俎], 「夢に乙羽君と語る」, 巖谷小波, 口語, NDC914, /会話: 巖谷小波

『ひまわり』では, 「眞實」「眞個」のような熟字訓は「ほんたう」をルビ検索することで見つけられるが, 「ほんたう」(本[ほん]當[たう])のように複数の文字にまたがる通常のルビは一文字ずつ調べないと検索できない。また仮名書きされた「ほんたう」はルビではなく本文を検索し直す必要がある。

◆踊り字をひらいた検索をする場合

例: 「などと」を検索

検索対象テキストの形式: ルビなしテキスト+踊字をひらく

検索文字列: 「などと」

英でも魯でも, 皆な斯の如しだなどと, 我慢をしたら,

- 1895年01号, P148A14, [家庭], 「家庭に於ける第一義」, 三島通良, 口語, NDC498
實は予などは朝鮮や清國位を閉口させたとて, エライダラウなどゝ慢心するやうでは, 到底東洋の覇權を永久に掌握する事は覺束ないと存じます。

- 1895年01号, P148B22, [家庭], 「家庭に於ける第一義」, 三島通良, 口語, NDC498

『ひまわり』では, 検索文字列を「などと」として検索すると「などと」は見つかるが「などゝ」は見つからない。

◆キーになる文字が存在しない正規表現を使った検索

例: カタカナ語を検索

検索対象テキストの形式: ルビなしテキスト (その他の形式でも可)

検索文字列：「[ア-ヴ]+」

山河自然の形勝に據り、三十年の歲月二億テールの巨財を投じ、最新の學術に由りて堅牢の防備を脩め、 - 1895年01号,P009A01, [論説], 「戦勝後の教育」, 千頭清臣, 文語, NDC371 一夫之に當れば百夫も亦其脚を移すに術なき要害となり、實に以太利のスペチュ, 露西亞のクロンスタッドと併稱して世界至險の三大砲臺として數へられしものなり、

- 1895年01号,P009A03, [論説], 「戦勝後の教育」, 千頭清臣, 文語, NDC371 實に以太利のスペチュ, 露西亞のクロンスタッドと併稱して世界至險の三大砲臺として數へられしものなり、千八百七十七年土耳其のオスマンパシヤは十萬の兵を以てブルスへの砲臺に籠り、 - 1895年01号,P009A04, [論説], 「戦勝後の教育」, 千頭清臣, 文語, NDC371

正規表現「[ア-ヴ]+」はカタカナの連続を表すが、この場合、検索のキーになる確定した文字が存在しないので『ひまわり』では検索できない。

5 おわりに

『太陽コーパス』の特長の一つは、さまざまな情報が付与されたXML形式による構造化テキストとして作られていることである。『プリズム』『たんぽぽ』はその構造化テキストから研究に必要な情報を容易に取り出すことができるように開発したものである。その標準的な機能だけを用いても、構造化テキストならではの方法によって、『太陽コーパス』を言語研究に活用することができるものと思う。

さらに『太陽コーパス』の特長を活かすためには、XSLTスタイルシートや正規表現などの技術をうまく使いこなすことが求められる。今後、コンピューターを利用した日本語研究全般においてこうした流れが進んでいくものと思われるが、その先駆けとして『太陽コーパス』においてその実践が広く行われることを期待したい。そのためにも『プリズム』『たんぽぽ』や付属のスタイルシートを活用していただければ幸いである。

注

- (1) XSLTはXML関連の規格として標準化団体W3Cの勧告となっている。<http://www.w3.org/TR/xslt>
- (2) TeXはスタンフォード大学のDonald E. Knuth氏によって開発された文書組版用のシステムで、pLaTeXはそれを扱いやすくするとともに日本語に対応させたもの。スタイルファイルを利用して機能を拡張することにより、目的に応じてさまざまな組版が可能となる。
- (3) dvioutは大島利雄氏（東京大学大学院数理科学研究科）によって開発され・フリーウェアとして公開されているTeX関連の閲覧用ソフト。<http://akagi.ms.u-tokyo.ac.jp/dviout.html>
- (4) 「文字鏡研究会」によって開発された、漢字12万字以上を収録する大規模な文字データベース。<http://www.mojikyo.org/>
- (5) 画面中央の「適用するスタイル」は、2.1.1、2.1.2でのスタイルシートの説明（①～⑮）の順に並んでいる。
- (6) 『たんぽぽ』のプログラムはJavaScript (Jscript) によって書かれており、正規表現もその仕様に準じたものに対応している。ただし、半角の丸括弧「()」は使用できず、先読みや後方参照が使用できなくなっている。
- (7) 文字クラスを範囲指定する場合にはユニコード順で指定する必要がある。『太陽コーパス』で使われているJISコードの範囲内の漢字は[一-龠]であらわすことができる。

漢語「優秀」の定着と語彙形成

—主体を表す語の分析を通して—

——田中牧郎

1 背景と目的

現代語彙の形成過程においては、近代化にともなって新しい漢語が登場し、旧来の和語や漢語との間に新しい体系を形成していく、目立った現象がある。研究史において、この現象を概観する研究や（注1）、個々の新漢語の語誌記述（注2）には蓄積が厚いが、旧来の語彙の中に新しい漢語がどのようにして場所を得、近代的な語彙を形成していったかという、新漢語の定着と語彙形成の視点からの具体的な記述研究は手薄である。この視点からの記述を進めるためには、従来、

（1）バランスのとれた資料にもとづく用例の出現状況の分析

（2）多様な用例から帰納する意味用法の分析

の二つのことが十分に実現できなかったために、研究が進展しなかったのではないと思われる。

（1）については、20世紀初期に定着する漢語に関しては、『太陽コーパス』での出現状況を観察することによって、当時の一般的な状況を推定することができると考えられる。その事例として、本論文では、定着までの過程がとらえられる「優秀」という漢語について、その定着過程と、〈優秀〉を意味する語彙の形成について記述したい。

（2）については、多くの用例をもとに語の意味分析を行うことは、コーパスによらない、これまでの語彙研究でも行われてきたが、その手法は、現代語研究の場合、内省を援用するものが中心であり（注3）、古典語の場合、個々の用例を解釈して現代語への置き換えを考える方法で記述に向かうのが一般的であった（注4）。明治・大正期を対象とする場合は、現代語研究と古典語研究のはざまにあつて、方法論的な自覚が薄かったように見受けられる。また、コーパスを用いて多様な用例から帰納する方法が普及していけば、現代語においても古典語においても、用例から

どのようにして語の意味を分析していくのかということをも、より真剣に問う方向が重視されるようになるだろう。本論文では、内省や解釈に頼らず、統語的に共起する語の分析を積み重ねる方法で帰納する記述を試みたい（注5）。

20世紀初期に定着する漢語「優秀」について、その定着過程を、用例の統語的な分析によって記述するのが、本論文の目的である。

2 漢語「優秀」と〈優秀〉語彙

『日本国語大辞典 第2版』（2000～2002年、小学館）では、「優秀」の初出例に、1906年の石川啄木『雲は天才である』の「石本君の極めて優秀なる風采と態度とは」という例を挙げる。「優秀」は、20世紀初め頃に日本語の文献上に現れるようになった語だと目される。『太陽コーパス』（注6）における「優秀」の年次別の頻度の推移は図1の通りで、1895年から用例はあるが、きわめて少数である。当初はめったに使われない語であったのが、1901年から1917年の間に急速に増加し、1917年から1925年の間では頻度が安定する。20世紀初期に登場し、短い期間で普及に至った語であると推定される。

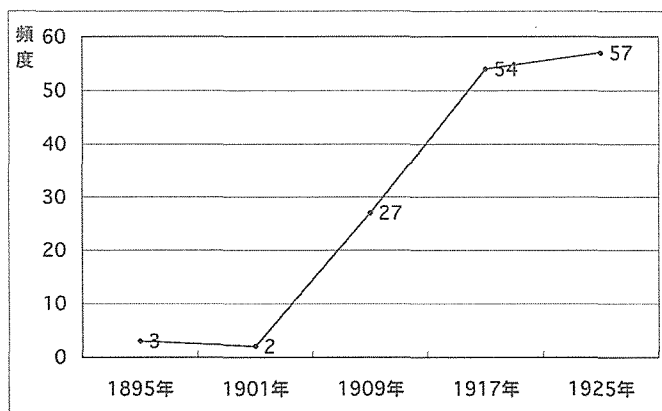


図1 「優秀」の年次別頻度

「優秀」の意味について、現代の国語辞典をみると、和語「す

ぐれる」「ひいでる」を用いて意味を記述するものが多く、これら3語の同義性は高いのではないかと考えられる。

他より一段と目立って(＝秀)優れていること。すぐれひいでていること。(岩波国語辞典第6版・岩波書店・2000年)
特にすぐれているようす(三省堂国語辞典第5版・三省堂・2001年)

他よりいちだんとすぐれひいでているようす(新選国語辞典第8版・小学館・2002年)

〔ほかのどのものと比べてみても〕すぐれている様子だ(新明解国語辞典第6版・三省堂・2005年)

国立国語研究所編『分類語彙表(増補改訂版)』(大日本図書・2004年)においては、「優秀」と「すぐれる」「ひいでる」は、3語いずれも「3.1584限定・優劣」の分類項目に属し、それぞれ03と04の段落に分かれ、次のように配されている。

03 並外れた 人並み外れた ず抜けた ずば抜けた
けた外れ けた違い 段違い 際立った 水際立った
優れて 優れた

04 卓越した 卓絶した 絶倫
拔群 卓抜 秀抜 優秀 秀逸 秀でた 抜きでた
出色 傑出した
圧倒的
有数
錚々〔～たる〕
役者が一枚上

『分類語彙表(増補改訂版)』の段落については、「意味上の語集団」(まえがき)と位置づけられているので、03, 04にあげられる諸語は、それぞれ「すぐれる」、「優秀」「ひいでる」と意味の近い語であると見ることができよう。

これらの語について、表記の違い(「優れる」「勝れる」「すぐれる」など)、品詞の違い(「卓越な」「卓越する」など)、下接する付属語の違い(「優れた」「優れて」「優れず」など)は区別せずに同じ語と認定して、『太陽コーパス』での頻度を調査した。形容動詞または動詞として用いられた例が、『太陽コーパス』全体で頻度10以上の語と頻度をあげると、下の13語の通りである(注7)。

03 際立つ35 すぐれる283

04 卓越53 卓絶26 絶倫20 拔群35 卓抜27

優秀143 ひいでる48 出色30 傑出39 有数94 錚々38

これら13語について『太陽コーパス』から得られた用例を一見すると、ほとんどはよく似た文脈で用いられていることがわかる。

すぐれた人間にたいする尊敬は失なはないのだ。(1917年5号「雑感」武者小路実篤P130B11)

固より伎倆卓越したる人物ならざる可からず。(1901年4号「内閣書記官長」P034B14)

如何に手腕家でも卓越した人物でも、(1901年14号「銀行と預金者」豊川良平(談)P075B03)

此人は古今絶倫の大偉人なり、(1909年2号「春汀君談片」末広一雄P038B08)

各國より技術學術に拔群なるものを聘し、(1909年11号「名士の瑞西観 瑞西の工業」片山正夫(談)P192A14)

俗人の常識は時に卓抜なる美術家の想像よりも賢きことあり。(1901年10号「文芸時評」高山樗牛P037B06)

技能優秀なる者の拔擢奨励の上に(1917年2号「大学程度の工業教育」大河内正敏P076A03)

手腕技術の秀でた醫士が多いのだ。(1909年11号「名士の瑞西観 瑞西の刀圭界」志立富松(談)P195B03)

松村龍雄と佐藤鐵太郎の二人が出色の人物である。(1917年1号「海軍更迭短評」P025A20)

露國には何故傑出した文學者が居るかといへば、(1909年8号「名士の露西亜観 政治家を中心として観たる露國」夏秋亀一(談)P198A02)

將來有数の店員として他から矚目されるやうな者は(1925年9号「研究心に富んでより以上仕事をする特性」浜田西郎P106C20)

小説家の錚々たるものも(1895年2号「文学」P163B13)

ただ、「際立つ」1語のみは、他の12語とは一線を画するのではないかと考えられる。

フランスは何所と云つて際立つて人眼を引くやうな容貌は持つてゐなかつたが、(1917年10号「クラゝの出家」有島武郎P391A16)

上の例は、「人眼を引く」様子について「際立つて」いることをいうもので、「フランス」の「容貌」が〈優秀〉であることを述べるものではない。「際立つ」の他の例を見ても、「見る」「聞く」などの視聴覚の動詞、「色」「音」「青い」「響く」など視覚や聴覚によって認識される名詞や形容詞・動詞の表す様子を述べる

場合が大部分であり、「際立つ」は〈優秀〉の意味とは異なる意味の語と見るべきであろう。

「際立つ」を除く12語は他の用例を見わたしても意味が類似しており、03と04のいずれの段落にあるかによって意味の差異があるようにも見えない。これらはひとつの類語集合と見なすことができそうであるが、意味の微細な差異を見きわめるためにはしるべき方法によって用例を整理していく必要がある。

3

主体の表示形式の整理による〈優秀〉語彙の絞り込み

3.1 全体主体と部分主体

〈優秀〉を意味する語彙12語は、属性を表す形容動詞や動詞であり、文中で述語となって用いられる。この述語に対応する主語の部分には、通常、属性の主体が表されるが、この主体を表す語は、述語にとってもっとも密接な関係をもつものである(注8)。

(1) 繪畫や彫刻の優秀なるを(1909年10号「外人の日本観 日本に於ける洋画の影響」P092A05)

(2) 康熙帝自身が優秀なる文學者にして(1895年6号「清朝 全盛の時代」中西牛郎P027B01)

(1)ではゴチック部の述語に対して、実線部「繪畫や彫刻」が主語となり、その「繪畫や彫刻」は〈優秀〉という属性をもつ主体に相当する。(2)でも実線部の「康熙帝自身」が主体になるが、この例では、点線部の「文學者」も、同じく主体と認められる。この場合、〈優秀〉という属性がどこに直接備わるのかというと、「康熙帝自身」のもつ一つの属性としての「文學者」であると解される。この例と同じように、主体に相当する語句が二箇所を示されているものには、次のような例がある。

(3) そこがまた同君の人の優れた所以で、(1909年8号「故 二葉亭子の性行」坪内逍遙(談)P130A04)

(4) 性質優れたる人に取りては(1909年5号「文芸時評」長谷川天溪P154A02)

(3)は「同君」のもつ「人」(人柄の意)が、(4)は「人」のもつ「性質」が、それぞれ述語「すぐれる」の直接の主体になっているものである。(2)(3)(4)は、いずれも、点線部の表すものは、実線部の表すものの部分を構成していると考えられ、実線部一点線部の間に、全体一部分という関係を認めることがで

きる。このような二種類の主体を、「全体主体」「部分主体」と呼び分けることとしたい。

3.2 全体主体の形式

全体主体に相当する語句が表現された場合、統語的にどのような位置にどのような形式で現れているのかを見ると、次の四つの類型に分類できる。用例における実線部が全体主体、点線部は部分主体である。

A 主格相当の体言

其作如何にすぐれたるも (1901年5号「文芸時評」大町桂月 P054A14)

B 被修飾語の体言

如何に手腕家でも、卓絶した人物でも (1901年14号「銀行と預金者」豊川良平 (談) P075B03)

C 部分主体に組み込み

黒田長政浅野幸長等の軍功は特に拔群なるにも拘はらず、 (1895年4号「加藤清正 (承前)」小倉秀貫 P040B05)

D 間接的な位置

選手となるには、先づ普通の人格に達した上で、その技倆に於て優秀なるを要すべき筈である。 (1917年9号「教育時言」兆水漁史 P012A14)

博士は、どこまでも行き届いた、生丁面な人である。外人との交際が廣く、特に佛獨語に秀でてゐるが、 (1925年12号「書齋から議政壇上に送られる…貴族院の四学者議員」碓氷三郎 P085A13)

上のうち、D「間接的な位置」とは、従属節、後続節もしくは別の文に表現された形式が間接的に述語と結びついているとみなせるものである。

A～Dの各類型が、12語それぞれを述語とする文において、どれだけの頻度と比率で現れているかを調査した結果が、表1である。表1から気づかれることは、D「間接的な位置」の比率が、「すぐれる」から「秀でる」までの8語 (I 群) は、25～50%程度の比率で共通しているのに対して、「出色」から「錚々」までの4語 (II 群) は、25%未満と低い数字で共通していることである。これは、I 群の語よりもII 群の語の方が、全体主体を表す語句を直接要求する度合いが強いことを示すものである。この二つの語群の違いは、B「被修飾語の体言」の数値にも現れており、I 群の8語が15～45%の間におさまるのに対して、II 群の

4語はいずれも45%を超えている。被修飾語は主格に比べて、述語によって属性が規定される性質が強いと考えられるので、Bの数値の出方から、I群の語に比べてII群の語が主体の属性を規定する性格がより強いことを示していると判断される。こうして、DとBの二つの数値の特徴から、全体主体との緊密度が、より弱いI群とより強いII群とに二大別することができる。

表1 全体主体の形式 (上段：頻度，下段：比率)

語 形式		Ⅰ 群								Ⅱ 群			
		すぐ れる	卓越	卓絶	絶倫	抜群	卓拔	優秀	ひい でる	出色	傑出	有数	錚々
A	主格相当 の体言	87	21	6	5	7	11	22	16	4	8	21	2
		30.7	39.6	23.1	25.0	20.0	40.7	15.4	33.3	16.7	22.9	24.7	5.7
B	被修飾語 の体言	79	12	7	4	7	5	62	11	14	16	39	27
		27.9	22.6	26.9	20.0	20.0	18.5	43.4	22.9	58.3	45.7	45.9	77.1
C	部分主体に 組み込み	33	5	5	2	4	2	15	3	3	5	6	0
		11.7	9.4	19.2	10.0	11.4	7.4	10.5	6.3	12.5	14.3	7.1	0.0
D	間接的な 位置	84	15	8	9	17	9	44	18	3	6	19	6
		29.7	28.3	30.8	45.0	48.6	33.3	30.8	37.5	12.5	17.1	22.4	17.1
合 計		283	53	26	20	35	27	143	48	24	35	85	35
		100.0	99.9	100.0	100.0	100.0	99.9	100.1	100.0	100.0	100.0	100.1	99.9

3.3 部分主体の形式

次に、部分主体に相当する語句が表現された場合、どのような位置にどのような形式で現れているかを見ると、次の九つの類型に分類できる。点線部が部分主体に相当する語句、実線部は全体主体に相当する語句である。

a 主格相当の体言

成績が優秀だと (1925年5号「維新志士を輩出せしめたる萩の雰囲気」横山健堂PI69CI3)

b 被修飾語の体言

青年學者の卓拔なる意見に接するは (1895年10号「文学」PO47A01)

c 与格相当の体言

才幹に秀で學術技芸に卓越して居るけれども (1909年11号「独逸文化の世界的影響」湯原元一PI20AI2)

d 主格相当の句

此蛇煙草汁を忌む事拔群で、(1917年14号「蛇に関する民俗と伝説」南方熊楠P191A18)

e 被修飾語の用言

更に一層すぐれて雄大な日向灘が(1917年9号「日向の海岸」田山花袋P124B09)

f 与格相当の句

出産率減少を最も科学的に判断して、それを豫防すべき諸種の組織及設備を立てた點に於て獨逸は著しく勝れてゐる。(1917年8号「独逸戦時の児童保護問題」笠原道夫P124B04)

g 主格相当の体言と被修飾語の用言

京都は氣候のすぐれて良和なるにもあらず(1895年4号「京都は国美の庫なるを論ず」久米邦武P010B12)

h 全体主体に組み込み

優秀なる工業上の技術家と經營家であらねばならぬ。(1917年2号「大学程度の工業教育」大河内正敏P073411)

i 間接的な位置

英吉利に於て國家を論じて最も卓越なる者を、(1909年13号「我が立憲政体と大政党」上杉慎吉P042B06)

j なし

教育を盛にするとは優秀な教育者を多く教育界に引き入れることである。(1917年5号「教育時言」兆水漁史P156B01)

上のように、部分主体として示される語句の位置や形式は、全体主体の場合よりも、種類が豊富である。与格相当の位置や用言や句の形式も見られ、部分主体を示す語は、多様な位置や形式によって述語と関連づいている。また、jのように、部分主体を示す語句が存在しない場合もある。

これらのa～jの各類型が、12語それぞれを述語とする文において、どれだけの頻度と比率で用いられているかを調査した結果が、表2である。

表2によれば、3.2で全体主体の形式の数値から見出したⅠ群とⅡ群との間に、部分主体の形式の数値から見ても明らかな線引きを行うことができそうである。まず、j「なし」の数値が、Ⅰ群は35%以下と低いものに対して、Ⅱ群は45～70%程度と高いことが指摘でき、部分主体を表す語句を要求する度合いは、Ⅰ群により強くⅡ群により弱い。また、a「主格相当の体言」の数値を見ると、Ⅰ群は15%を超えるものに対して、Ⅱ群は、15%以下と低くなっており、部分主体を主格に取り立てる度合いが、Ⅰ群に

表2 部分主体の形式 (上段：頻度, 下段：比率)

語 形式		Ⅰ群								Ⅱ群			
		すぐ れる	卓越	卓絶	絶倫	抜群	卓抜	優秀	ひい でる	出色	傑出	有数	錚々
a	主格相当 の体言	89	10	11	17	11	8	22	17	2	5	1	2
		31.4	18.9	42.3	85.0	31.4	29.6	15.4	35.4	8.3	14.3	1.2	5.7
b	被修飾語 の体言	62	27	4	3	19	17	53	7	5	9	45	7
		21.9	50.9	15.4	15.0	54.3	63.0	37.1	14.6	20.8	25.7	52.9	20.0
c	与格相当 の体言	14	4	4	0	1	0	16	14	0	3	0	0
		4.9	7.5	15.4	0.0	2.9	0.0	11.2	29.2	0.0	8.6	0.0	0.0
d	主格相当 の句	4	0	0	0	1	0	0	0	0	0	0	0
		1.4	0.0	0.0	0.0	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
e	被修飾語 の用言	18	0	2	0	0	0	0	1	0	0	0	0
		6.4	0.0	7.7	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0
f	与格相当 の句	5	0	0	0	0	0	0	0	0	0	0	0
		1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
g	主格相当の 体言と被修 飾語の用言	11	0	0	0	1	0	0	0	0	0	0	0
		3.9	0.0	0.0	0.0	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
h	全体主体に 組み込み	3	2	1	0	0	1	2	0	2	1	1	0
		1.1	3.8	3.8	0.0	0.0	3.7	1.4	0.0	8.3	2.9	1.2	0.0
i	間接的な 位置	1	2	0	0	0	0	0	0	0	0	0	0
		0.4	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
j	なし	76	8	4	0	2	1	50	9	15	17	38	26
		26.9	15.1	15.4	0.0	5.7	3.7	35.0	18.8	62.5	48.6	44.7	74.3
合 計		283	49	26	20	35	27	143	48	24	35	85	35
		100.0	100.0	100.0	100.0	100.0	100.0	100.1	100.1	100.0	100.0	100.0	100.0

より強くⅡ群により弱い。このjとaとの二つの数値の特徴は、部分主体との緊密度がより強いⅠ群とより弱いⅡ群という二大別ができることを示している。

3.4 〈優秀〉語彙の分類

3.2, 3.3における統語的な位置や形式の分析の結果, 全体主体・部分主体との統語的な位置や形式から, 12語は表3のように分類することができた。以下, 「優秀」と類似性の高いⅠ群の8語を〈優秀〉語彙と扱い, これら8語についてさらに考察を進

めていくことにしたい。

表3 全体主体・部分主体との緊密度から見た〈優秀〉語彙の類別

群	語	全体主体との 緊密度	部分主体との 緊密度
I 群	すぐれる 卓越 卓絶 絶倫 拔群 卓抜 優秀 ひいでる	弱い	強い
II 群	出色 傑出 有数 錚々	強い	弱い

4

共起語の整理による意味分析

4.1 全体主体の整理

3で絞り込んだ〈優秀〉語彙8語について、まず全体主体として表現される語句の内容を整理することを通して、語の意味分析を行う。全体主体として表現されている語句について、その形式が3.2で見たD（間接的な位置）であるものを除いたA・B・Cの場合について（注9）、その内容を見わたすと、〔人間〕〔組織〕〔生物〕〔自然〕〔文化〕の五つの意味分野に分けることができる。用例における実線は全体主体、点線は部分主体である。

〔人間〕

すぐれた人間にたいする尊敬は失なはないのだ。（1917年5号「雑感」武者小路実篤P130B21）

〔組織〕

會社の優れた特色の一と数へる事が出来る。（1925年1号「王子製紙会社の事業と社員待遇法」天外散史P254A07）

〔生物〕

シオンの市の優れた牝牛は、ヴァイオレットと闘ふために蒐められた。（1925年13号〈雑学〉ZI36D07）

〔自然〕

私は社殿に詣で、後の二三十分をそのすぐれた海のどよみに見入った。（1917年9号「日向の海岸」田山花袋P126B18）

〔文化〕

たとひ其作如何にすぐれたるも（1901年5号「文芸時評」大町桂月P054A14）

これら五つの分野別の頻度と比率を語別にまとめると表4のようになる。語の配列は、数値の特徴が見えやすいように並べ替えた。

表4 全体主体の内容分類 (左:頻度, 右:比率)

	すぐれる		優秀		ひいでる		卓絶		卓越		拔群		卓抜		絶倫	
人間	112	56.3	48	48.5	16	53.3	14	77.8	32	84.2	16	88.9	18	100.0	11	100.0
組織	8	4.0	22	22.2	5	16.7	2	11.1	5	13.2	1	5.5	0	0.0	0	0.0
生物	7	3.5	1	1.0	0	0.0	1	5.6	0	0.0	1	5.5	0	0.0	0	0.0
自然	8	4.0	0	0.0	8	26.7	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
文化	64	32.2	28	28.3	1	3.3	1	5.6	1	2.6	0	0.0	0	0.0	0	0.0
計	199	100.0	99	100.0	30	100.0	18	100.1	38	100.0	18	99.9	18	100.0	11	100.0

表4によると、まず「卓絶」から「絶倫」までの5語は〔人間〕がほとんどである点で共通しているが、これら5語はあまり用法の広がりを見せておらず、用例数も多くない。これに対して、「すぐれる」「優秀」「ひいでる」はそれぞれ特徴のある数値が出ている。この3語のうち「すぐれる」「優秀」は100～200の用例があるのに対して、「ひいでる」は30の用例数にとどまる。頻度の高低は用例から帰納する分析の質にも影響すると思われる、「ひいでる」は別に扱う方がよいだろう。したがって、以下の考察は、「すぐれる」と「優秀」との比較検討を重点的に行い、「ひいでる」を含めた他の6語については後でまとめて記述する形をとることにする。

さて、「すぐれる」と「優秀」は、〔人間〕が半数程度、〔文化〕が3割程度を占める点でよく似ているが、「優秀」が〔組織〕にも比較的多いのに対して、「すぐれる」は〔組織〕に多いわけではなく、〔生物〕〔自然〕などにも広がっているという相違も指摘できる。「優秀」が〔人間〕〔組織〕〔文化〕といった人間活動に関わる分野に限定されるのに対して、「すぐれる」は〔生物〕〔自然〕といった人間活動以外の分野にも広がっているのである。

「すぐれる」「優秀」がともに多い〔人間〕〔文化〕についてさらに詳しく比較してみよう。表4で〔人間〕を全体主体とする例とした、「すぐれる」112例、「優秀」48例について、全体主体に相当する〔人間〕を表す語句を、種類別に整理すると表5・表6のようになる。語句を一覧する表5では、語句に添えられていた助辞相当の部分は取り去った形で示す。表5・表6からわかる通り、「すぐれる」は、一般名、個人名、代名詞・親族名、職位・

属性名（職業や専門分野，居住地など人物の属性をもとに付けられた呼び名）のいずれにも多く，人間を表す語句を広く主体にとる。これに対して，「優秀」は，職位・属性名に集中する傾向が強く，個人名や代名詞・親族名を主体にとる場合はきわめて少ない。個人名に配した「康熙帝自身」「安保清種少将」も「帝」「少将」の部分に着眼すれば，職位・属性名に分類することも可能である。職位・属性名は「すぐれる」にもある程度見られるが，比率はさほど高くない。また，「優秀」の方には，属性のうち民族を表す語句が多いのも特徴である。

表5 全体主体〔人間〕の語句の種類（一覧）〔数字は頻度，数字のないものは頻度1〕

	すぐれる	優秀
一般名	人12，人々2，人ども，人間2，人物，者3，もの3，男2，君，同君，氏3，此二氏，古人，今代人	人3，人々，者6，もの4，同君，個人，婦人
個人名	重吉，海男，秀吉，諭吉，忠貞，高杉，齋藤隆夫氏，辜鴻銘氏，真龍，金龍，韓愈，コックス，マクヴェー氏，片山廣子，政子，お里，娘お村	康熙帝自身，安保清種少将
代名詞・親族名	余，妾，おの，自分2，各々，自身，我々，御邊，あなた，汝，君，こいつ，彼5，其人，母2，母君，良人，娘2，令嬢，第一子，誰が家の子女	私
職位・属性名	御公達，君子，卿，侯，侯爵，右大臣，薩摩守，三の君，作家，新作家，教授，同教授，探偵，俳優2，畫家，彫師，技術者，閣員，何省書記官正三位，侍，犯人，人傑，花形，上人，上手ども，日の大君，上に在る人，仏蘭西の友人，仏蘭西人，英國々民，日本人，我が民族	帝自身，侯，活動者，科學者・文學者，教育者，學士等，醫師，技術家と經營家，勞働者，我國職工，標準職工，職工，海陸將士，生徒3，卒業生，總代，佛國人民，獨逸人，獨逸民族，日本人，國民，民族3，一民族2

表6 全体主体〔人間〕の語句の種類（頻度と比率）〔左：頻度 右：比率〕

	すぐれる		優秀	
一般名	34	30.4	17	35.4
個人名	17	15.2	2	4.2
代名詞・親族名	28	25.0	1	2.1
職位・属性名	33	29.5	28	58.3
計	112	100.1	48	100.0

同様にして、表4で〔文化〕を全体主体とする「すぐれる」64例、「優秀」28例について整理すると表7・表8の通りである。学術・芸術・文芸・芸能などの分野で創作されるものや素材にあたる語句を〔学芸〕としてひとくくりにし、近代的な技術によって製作されたものや材料は〔製品〕とした。「もの」や指示詞のように指し示す事物がその語句だけでははっきりしないものは、当該例の文脈中からそれに相当する語句を補って（ ）内に示した。本文中にはない語を補った場合は〔 〕内に示した。

表7・表8のように、「すぐれる」は〔学芸〕がきわめて多いのに対して、「優秀」は〔製品〕の方が多いという明瞭な相違が見られる。「すぐれる」は、言葉や芸によって創作されるものの属性を表現する場合に使われやすいのに対して、「優秀」は、近代的な技術によって製作されるものの属性を表現する場合に使われやすいという相違があると言えよう。

このように、「すぐれる」と「優秀」とではほぼ同じ比率を占め

表7 全体主体〔文化〕の語句の種類（一覧）〔数字は頻度、数字のないものは頻度1〕

	すぐれる	優秀
学芸	文明, 文明の思想と様式, ギリシャ美術, 健駄邏彫刻, 和歌, 歌, 日本文, 洋字, 表現法, 考案, 詩, 人物畫, 作4, 此の作, 其作, 作品3, 作物, 通俗小説, もの (壬生踊り・舞楽伎楽), もの (哲學や宗教), もの (本書) 2, 『淡雪』, 『遠州一重切花入』〔ほか〕, 『開墾小屋』など, 治療, 後者 (徳岡神泉氏の同畫題の作), もの (玉置氏の秋圖), もの (繪畫), もの (作品), もの (作), もの (作物) 2, もの (芝居), もの (一枚の繪), もの (袖枕), もの (長塚節氏の『土』), もの (鹽原紀行), もの (書の二幅對), もの (あれ), 君の (『浮雲』), もの {存在}, 單記制, 正月6, 今	國家文明2, 健駄邏美術, 人物畫, 繪畫, 繪畫や彫刻, 作品3, 一つ (朝倉文夫氏の『時の流れ』), 美術品, トルストイ流に多數を標準とした物, 物 (文藝), 是の (佛陀の坐像)
製品	車, 兌換券, 寶, 金銀環, 何れ (銅の種類), 水, 石墨, 牛肉, 生活	蛋白質, 生絲, 機械, 製作品, 航洋潜水艇の一隊, 高周波増幅式受信機, 受話裝置, カタロ及びボラ, 武器, もの (工業製品), もの (ラウドスピーカー〔ほか〕), もの (大砲), もの (製作したる物品), 者 (社會を組織するもの), 倫敦巴里

表8 全体主体〔文化〕の語句の分類（左：頻度 右：比率）

	すぐれる		優秀	
学芸	55	85.9	13	46.4
製品	9	14.1	15	53.6
計	64	100.0	28	100.0

る全体主体の〔人間〕と〔文化〕について、その内訳を見ていくと、かなり明瞭な相違のあることが明らかである。「すぐれる」は、職位や属性に規定されない個人や、言葉や芸によって生み出された伝統的な文化に関して用いられる傾向が強いのに対して、「優秀」は、職位や特定の属性にある個人や、近代的な技術によって生み出された文化に関して用いられる傾向が強いとすることができるのである。

4.2 部分主体の分析

次に、「すぐれる」「優秀」について、部分主体のありようについて分析する。部分主体の表現される形式が3.3で見た、i（間接的な位置）・j（なし）を除いたa～hであるものについて、全体主体の意味分野別に見ていく。

まず、〔人間〕を全体主体とする場合、その部分主体の語句の種類を整理すると、表9・表10の通りである。

「すぐれる」は人間に備わる〔才能〕を表す語句に偏り、一方「優秀」は〔才能〕よりも才能をもとに身に付けられたり生み出されたりする〔技芸〕を表す語句の方が多い。

同様にして、〔文化〕を全体主体とする場合について、部分主体の語句を整理すると、表11・表12の通りである。文化的な営みによって生み出されたものに備わる性質に関わる語句を、〔人間〕の場合の〔才能〕に対応させて〔性能〕とした。文化的な創作の技や芸、あるいはそうやって作り出されたものの機能を言う語句を〔人間〕の場合と同様〔技芸〕としてまとめた。

表11・表12によれば、「すぐれる」は〔性能〕の方により多く、「優秀」は〔技芸〕に偏るという傾向が指摘できる。

全体主体が〔人間〕の場合も〔文化〕の場合も、部分主体として示される語句は、「すぐれる」は主体の内に備わる〔才能〕や〔性能〕が多く、「優秀」は主体から外に現れる〔技芸〕が多いという違いが明らかになった。「すぐれる」と「優秀」はともに人間や文化の〈優秀〉さを意味しながら、内面的な質に焦点を当て

表9 全体主体〔人間〕の場合の部分主体の語句の種類（一覧）

〔数字は頻度、数字のないものは頻度1〕

	すぐれる	優秀
才能	体質及体量, 體格, 健康4, 仕草, 容色, 聲は高い, 気分2, み姿, 美しい, 智慧, 智徳, 智謀, 御ちから, 注意力, 鑑識, 頭脳2, 技倆2, 伎倆, 能力2, 審美的能力, 素質, 武勇, 武勇智略, 腕2, 手腕2, 事務の手腕, 腕前, 右手の利く, 目, 眼, 見ること聞くこと, 學力, 才2, 辯舌の才, 才能, 天才, 才智, 感情に長じ情緒に富む事, 哀れ, 好き, 位・富, 幾字は響渡る, 人格, 人格・品性・伎倆, 人, 性質, 道德, 藝術家, 琵琶の上手, 學者, 操觚者, 劍客	統御力, 技能, 人格2, 國民的精神, 血, 人種, 一職工, 文學者2
技芸	學問2, 音樂, 意匠, 槍・劍, かちゆみ, 擬古文, 見事な表現, 文体が美しい, 創作, 手筆, もの(演説), 言, 筆舌, 物祝ふ, 忠告, 行政顧問, 權威, 勝利, 特長(機略縦横), この點(立憲政治の要素), 常識, 業	成績5, 工業, 技術2, 技工, 勞働能率, 學術, 遊戲或は競技

表10 全体主体〔人間〕の場合の部分主体の語句の種類（頻度と比率）〔左：頻度 右：比率〕

	すぐれる		優秀	
才能	62	72.9	10	45.5
技芸	23	27.1	12	54.5
計	37	100.0	22	100.0

表11 全体主体〔文化〕の場合の部分主体の語句の種類（一覧）〔数字は頻度、数字のないものは頻度1〕

	すぐれる	優秀
性能	品質, 質と内容, 氣品, 價值2, 藝術的價值, 感じ, 楽しい3, 慶ばしい3, 趣向, 眞珠の玉のやうな, 分子, 營養力, 點(到る所豊富に得られる)	雰圍氣, 質
技芸	技, 描出の技巧, 工技用, 繪畫, 藝術, 戀歌, 製作, 用紙, 解剖, 道德的, 役立つ, 進み, 點(政見を發表することができる)	成績, 文學2, 榮養學, 知識上・道德上・審美上・物質上, 補助艦艇, 軍港, 会場の輸送も鐵道の輸送も

るのか、外発的な技に焦点を当てるのか、という意味的な対立があると考えられる。

この意味的な対立は、全体主体に相当する語句がほぼ同じ場合

表12 全体主体〔文化〕の場合の部分主体の語句の種類（頻度と比率）〔左：頻度 右：比率〕

	すぐれる		優秀	
性能	18	58.1	2	20.2
技芸	13	41.9	8	80.0
計	31	100.0	10	100.0

に、部分主体に相当する語句が対立的なありようを示す、次のような用例を比較することで、鮮明に浮かび上がる。

- (1) 性質優れた人に取りては（1909年5号「文芸時評」長谷川天溪P154A20）
- (2) 特に素養のある技術優秀の人を選んだ。（1925年12号「刻下の失業事情と職業補導」安田亀一P098B01）
- (3) そこがまた同君の人の優れた所以で（1909年8号「故二葉亭子の性行」坪内逍遙（談）P130A04）
- (4) 平素同君は非常な優秀な成績を収めて居たから（1925年1号「私の学生時代 二八会を中心とした懐しい思出」土方久徴P141M10）

(1) (2) は「人」を全体主体にとる点で共通しているが、部分主体にとる語は、「すぐれる」の例では「性質」、「優秀」の例では「技術」である。同じく (3) (4) はともに「同君」を全体主体にとるが、部分主体は、「すぐれる」では「人」（人柄の意）、「優秀」では「成績」である。いずれの組においても、内に備わる才能に着眼している場合は「すぐれる」、外に現れる技芸に着眼している場合は「優秀」という相違が明らかである。

- (5) 古い優れた人物画には皆興味を感じ（1917年13号「文部省展覧会日本画家の作意と苦心」平福百穂P117B03）
- (6) 優秀なる繪畫の雰圍氣を藥蒸せしめたるに依るはいふまでもないが（1925年10号「京都画壇の鳥瞰図」黒田天外P138A13）

この2例は、絵画を全体主体とするものだが、(5) の「すぐれる」の例では部分主体に相当する語をとらず、絵画それ自体に着眼しているものと考えられるが、(6) の「優秀」は絵画が外に醸し出す「雰圍氣」を部分主体として取り出してそこに着眼していると見ることができる。(1) ～ (4) の場合と同様の意味的な対立が認められよう。

4.3 周辺のな6語について

「ひいでる」「卓絶」「卓越」「拔群」「卓抜」「絶倫」の6語は、いずれも、〔人間〕を表す全体主体をとる比率がもっとも高いが、「すぐれる」「優秀」で見た分類と同じ枠組みでその数と比率をまとめると、表13のようになる。

表13 全体主体〔人間〕の語句の種類（周辺のな6語）（左：頻度，右：比率）

	ひいでる		卓絶		卓越		拔群		卓抜		絶倫	
一般名	5	33.3	4	28.6	10	30.3	6	37.5	3	16.7	3	27.3
個人名	4	26.7	1	7.1	7	21.2	6	37.5	3	16.7	4	36.4
代名詞・親族名	0	0.0	3	21.4	4	12.1	2	12.5	1	5.6	2	18.2
職位・属性名	6	40.0	6	42.9	12	36.4	2	12.5	11	61.1	2	18.2
計	15	100.0	14	100.0	33	100.0	16	100.0	18	100.1	11	100.1

表13から、職位・属性名の比率がもっとも高い「ひいでる」「卓絶」「卓越」「卓抜」と、一般名・個人名の比率の高い「拔群」「絶倫」との二つのグループに分かれることがわかる。この区分は、漢語に限ってみれば、「卓○」の形式で「○」の部分には「絶する」「越える」「抜く」といった動詞性の成分が来て、「卓^{すぐ}れる」と同義的に並列される語構成をもつ漢語と、「群を抜く」「^{とん}倫を絶する」という動詞性の成分の後に名詞性の成分が来る語構成をもつ漢語との区分に対応する。全体主体の形式上の区分と語構成上の区分とが一致しており、統語形式による区分と語構成による区分との対応が見えやすい。

和語「ひいでる」は、表13では特徴が見えないが、表4（125頁）においては、他の語には見られないか比率の低い〔自然〕を表す語句を全体主体にとる比率の高いことが特徴であった。〔自然〕を表す語句を全体主体にとるのは、他には「すぐれる」に見られるが、「すぐれる」と「ひいでる」の全体主体〔自然〕の語句を比較してみると、次の通りである。

すぐれる

金剛山、マッターホルン・ユングフラウ、折生迫、瀬水岸、
京都、日向灘、海、海中

ひいでる

山、富士峯の形を爲したる一山、抑燧山、阿蘇、広島城の
櫓櫓、帝都の舊跡、皇祖廟地、九江の上流

「すぐれる」にはさまざまな自然を表す語句が見られるのに対

して、「ひいでる」は山に集中し、城や旧跡・廟地・川などの場合も、山地にあるものばかりである。「ひいでる」が山がそびえ立つような形状に焦点をあてて〈優秀〉さを表す意味の語であることを示す現象だと考えられる。

以上のように、周辺の6語も、それぞれ意味的な特徴をもっていたと考えられ、その特徴を生かして〈優秀〉語彙のなかに場所を得ていたものと思われる。

5

語の使用域

5.1 ジャンル

3と4の分析により、主体の形式や内容に反映した意味から見た〈優秀〉語彙の類別や相互の関係を明らかにすることができたが、語彙のありようを把握するには、個々の語の使用域についても調べるのが望まれる。ここでは、『太陽コーパス』に付与されたジャンルと文体の情報を用いて語の使用域を記述する。〈優秀〉語彙8語について『太陽コーパス』に付与されたジャンル情報(NDC)別の頻度と比率を示すと、表14の通りである。

「すぐれる」と「優秀」を比較すると、「すぐれる」は「9文学」に集中し、「優秀」は「3社会科学」に集中するという目立った相違がある。もうひとつの和語「ひいでる」も「9文学」にもっとも多く、漢語「卓越」「卓絶」「拔群」「卓抜」は「3社会科学」にもっとも多いことから、和語は「9文学」に多く漢語は「3社会科学」に多いという一般的な傾向があると考えられる。一方、「9文学」への集中の度合いは「ひいでる」よりも「すぐれる」に著しく、また同じ漢語でも「絶倫」は「2歴史」に多いなど、語による個別的な特徴もうかがえる。「優秀」に目立ち他の語にはない特徴としては、「5技術」に多いことが注目される。近代的な技術によって作り出されたものの属性や、外に現れる技芸の属性を表現する「優秀」の意味的な特徴と関連しそうな使用域の特徴であると言える。

表14 〈優秀〉語彙のジャンル別頻度と比率（左：頻度，右：比率）

	すぐれる		優秀		ひいでる		卓絶		卓越		卓抜		抜群		絶倫	
0総記	22	7.8	1	0.7	0	0.0	1	3.8	6	11.3	2	8.0	2	5.7	0	0.0
1哲学	10	3.5	4	2.8	2	4.2	1	3.8	5	9.4	0	0.0	1	2.9	1	5.0
2歴史	31	11.0	7	4.9	10	20.8	4	15.4	7	13.2	1	4.0	7	20.0	6	30.0
3社会科学	43	15.2	59	41.3	11	22.9	10	38.5	25	47.2	14	56.0	12	34.3	5	25.0
4自然科学	13	4.6	4	2.8	2	4.2	2	7.7	0	0.0	1	4.0	2	5.7	0	0.0
5技術	13	4.6	29	20.3	1	2.1	0	0.0	3	5.7	0	0.0	1	2.9	0	0.0
6産業	2	0.7	10	7.0	1	2.1	3	11.5	2	3.8	1	4.0	1	2.9	0	0.0
7芸術	38	13.4	17	11.9	4	8.3	1	3.8	2	3.8	0	0.0	1	2.9	5	25.0
8言語	1	0.4	0	0.0	2	4.2	2	7.7	0	0.0	2	8.0	0	0.0	1	5.0
9文学	110	38.9	12	8.4	15	31.3	2	7.7	3	5.7	4	16.0	8	22.9	2	10.0
計	283	100.1	143	100.1	48	100.1	26	99.9	53	100.1	25	100.0	35	100.2	20	100.0

* 「卓抜」には他にNDCに分類できない予告記事に用いられた2例がある。

5.2 文体

『太陽コーパス』に付与された文体情報の種類別（文語か口語か）に〈優秀〉語彙8語の頻度と比率を整理すると、表15の通りである。一見して、「すぐれる」と「優秀」は口語に多く、「ひいでる」「卓絶」「卓抜」「抜群」は文語に多いということがわかる。このことから、「すぐれる」と「優秀」は日常語、「ひいでる」「卓絶」「卓抜」「抜群」は文章語、という文体的特徴の違いがあったと見ることができそうである。

「すぐれる」と「優秀」を比較すると、「優秀」の方が口語率が高いが、これは2で見た通り、この語が後の時代になって増加することに関わる現象で、『太陽コーパス』では後の時代ほど口語の文章の比率が高くなっていることによるものである。「優秀」の頻度が安定する1917年・1925年の2年分だけで比率を計算すると、「すぐれる」は文語率0.8%・口語率99.2%、「優秀」は文語率0.9%・口語率99.1%となり、2語に差はない。「優秀」が

表15 〈優秀〉語彙のジャンル別頻度と比率（左：頻度，右：比率）

	すぐれる		優秀		ひいでる		卓絶		卓越		卓抜		抜群		絶倫	
文語	111	39.2	16	11.2	32	66.7	21	80.8	27	50.9	20	74.1	21	60.0	9	45.0
口語	172	60.8	127	88.8	16	33.3	5	19.2	26	49.1	7	25.9	14	40.0	11	55.0
計	283	100.0	143	100.0	48	100.0	26	100.0	53	100.0	27	100.0	35	100.0	20	100.0

普及した後の「すぐれる」と「優秀」の日常語性は同程度であったと考えられる。

6

漢語「優秀」の定着と〈優秀〉語彙の形成

6.1 年次別の頻度の推移

〈優秀〉語彙8語について、『太陽コーパス』における年次別の頻度をグラフにすると図2のようになる。「優秀」の増加は、他の語と比較してみても際立った現象であることがわかる。「すぐれる」は起伏はあるものの、もっともよく用いられる語として不動である。その他の語は、「優秀」「すぐれる」に比較するときわめて低頻度にとどまり、「ひいである」「卓絶」「卓抜」「拔群」のように、減少傾向を見せる語も多い。当初は、「すぐれる」のみが多用されていたが、やがて「優秀」がこれに加わり、この二語が〈優秀〉語彙の基本部分を構成するようになり、他の語は周辺の語としての位置にとどまったと見ることができる。

4で見たような「すぐれる」と「優秀」の意味的な対立関係はこうした頻度の経年変化から見た語彙構成の推移とも関わりをもちながら形成されたのではないかと推測される。「すぐれる」と「優秀」とが意味的な対立関係を形成していく過程について、以下に考察することにした。

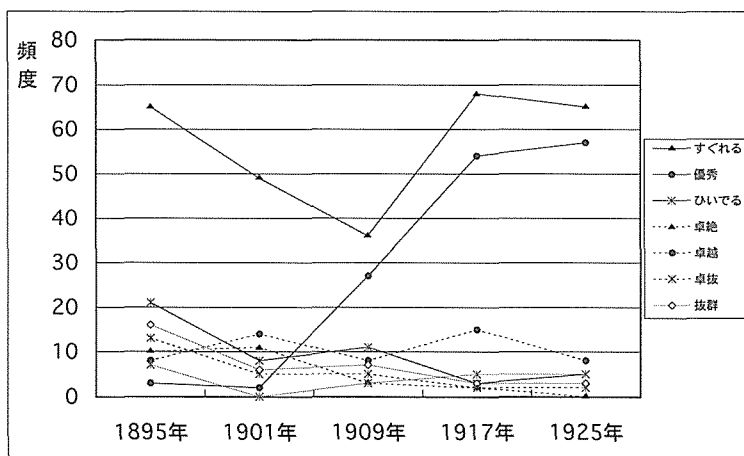


図2 〈優秀〉語彙の年次別の頻度

6.2 年次別の意味用法の変化

「すぐれる」と「優秀」について、全体主体として直接表示される語句を、4で行ったのと同じ方法で整理し、年次別にグラフにまとめると、図3・図4のようになる。

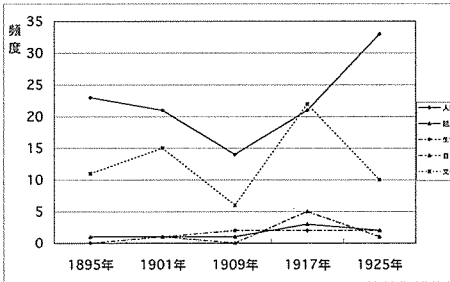


図3 「すぐれる」の全体主体

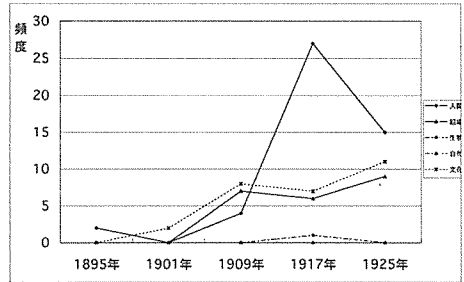


図4 「優秀」の全体主体

「すぐれる」は分野による経年的な変化の方向性ははっきりとは見出だせない。一方、「優秀」は、めったに使われない1895年・1901年では分野が安定せず、増加傾向が見え始める1909年から分野も拡大する。1909年には〔文化〕〔組織〕の分野がまず優勢となるが、1917年には〔人間〕の分野が急伸して優勢になり、1925年には〔人間〕の分野はやや減少し、〔人間〕〔文化〕〔組織〕の3つの分野で同程度に用いられる形になる。語が普及に向かう初期段階では用法にゆれを見せながら、普及が進行し頻度が安定するのにともなって次第に用法も安定していく様子が見て取れる。「優秀」がとる全体主体の語句の推移を、やや詳しく見るために〔人間〕と〔文化〕の分野につきその語句を分類し、年別に一覧にしてみよう（表16）。

1895年の2例は、同一の記事にある用例である。

抑も康熙帝が夷狄の種族を以て、支那國民に愛せられたるは、帝自身が優秀なる文學者にして、又力を極めて文學を奨励したるにあり、抑も支那の如き文學的の國民に文學者が敬愛せらるゝことは、猶ほ俳優が芝居好きの仲間に愛せられ、力士が相撲好きの家に愛せらるゝに異ならず、然らば康熙帝自身が優秀なる文學者にして、又力を極めて文學を奨励したるは或は支那國民を駕馭する一種の政略にあらざるかと疑はるゝものあり（1895年6号「清朝全盛の時代」中西牛郎 P027A24～P027B02）

「文学者」としての中国「康熙帝」を主体とするもので、使用者の中西牛郎は宗教思想家である。1901年に見られる2例は、次である。

吾人は國寶鎌倉大佛畧記と題せる小冊子に接して先づ其國寶の二字に奇異の感を生じぬ、而して其説く所は純然たる大佛製作の由來と其技術上の解説にして、更に佛像として宗教的文字を見ず、〔中略〕彼等（住職等）は本尊として渴仰せらるゝよりも優秀なる美術品なりと稱讃せらるゝに満足し、（1901年3号「宗教時評」龍山学人P051A22）

是の優秀なる健駄邏美術が印度に於ける希臘文明の衰退と共に、遂に三十二相式の壓倒する所となりたるは、實に佛像美術の發達を阻害せる歴史上の一恨事と謂はざるべからず。（1901年10号「文芸時評」高山樗牛P037A15）

別々の記事の用例であるが、鎌倉大仏を指す「美術品」やインドの「美術」を主体としており、仏教美術である点は共通している。また、宗教家龍山学人と思想家高山樗牛という使用者は、宗教や思想の専門家である点で、1895年の使用者であった中西牛郎に通うところがある。このように、1895年と1901年の「優秀」は、中国や仏教の文学や美術に関する人や技芸について、宗教や思想の分野に限って使われており、かなり専門性の高い語であったと推定される。

1909年になると、「優秀」の主体になる語句に1895年や1901年のような限定は見られなくなる。使用者もさまざまな分野の人にわたるように変容しており、1901年と1909年との間で、「優秀」という語の普及は急速に進んだと考えられる。ただし、〔文化〕の方に多く〔人間〕にはあまり多くないという偏りがある点には注意される。1909年において「優秀」という語を頻繁に用いている（3つの号にわたって5回用いている）著者に長谷川天溪（文芸評論家）がいるが、この著者の「すぐれる」と「優秀」の使い方を見ると次のようなことがわかる。「すぐれる」は〔人間〕の場合にも〔文化〕の場合にも使うが、「優秀」は〔文化〕の場合に限って頻用されているのである。

彼れ等外國政治家の大議論を吐き、大文章を草するのは、其の學識の深い故でもなく、其の政治的手腕の優れて居る故でもなく、（1909年12号「文芸時評」長谷川天溪P157B09）

古來優秀なる文學と稱せられた物の相場に就いても亦疑問が生ずる（1909年14号「文芸時評 功利主義の文芸」長谷川天溪P156B17）

表16 「優秀」の〔全体主体〕の語句の推移

		1895	1901	1909	1917	1925
人間	一般名			もの	人, 者3, 人, 個人, 婦人	人2, 人々, 者3, もの3, 同君
	個人名	康熙帝 自身			安保清種少將	
	代名 詞・親 族名					私
	職位・ 属性名	帝自身		侯, 生徒, 學士 等	活動者, 教育者, 醫師, 技術家と經營家, 我國 職工, 職工, 海陸將士, 生徒2, 卒業生, 總代, 佛國人民, 獨逸人, 獨 逸民族, 國民, 民族3, 一民族2	科學者・文學者, 勞働 者, 標準職工, 日本人
文化	学芸		健駄邇 美術, 美術品	繪畫や彫刻, 作 品3, トルスト イ流に多數を標 準とした物, 物 (文藝),	一つ(朝倉文夫氏の 「時の流れ」)	國家文明2, 人物畫, 繪畫
	製品			生絲, 者(社會 を組織するも の)	航洋潜水艇の一隊, カ タロ及びボウ, 武器, もの〔工業製品〕, も の(大砲), もの(製 作したる物品)	蛋白質, 機械, 製作品, 受話裝置, もの(ラウ ドスピーカー〔ほか〕), 高周波増幅式受信機, 倫敦巴里

このような例から1909年の時点では、「優秀」は「すぐれる」とは異なる意味分野の主体について使うという意識があったのではないかと思われる。

ところが1917年になると、2語の使い分けの基準は変容している。例えば、大河内正敏(応用科学者)は、同一の記事に「すぐれる」を1回、「優秀」を3回用いている

其教育制度の如きも、(中略)必ずしも程度の高いのを以て優れたりと爲すのではなく、現在の進歩したる工業に最適となる可き教育制度を確立する事が今日の急務であつて(1917年2号「大学程度の工業教育」大河内正敏P073A17)
工業上の原料を多量に有せざる邦國が、今後唯一の武器とし頼みとす可き者は科學を應用した發明、優秀なる工業上の技

術家と経営家であらねばならぬ。(1917年2号「大学程度の工業教育」大河内正敏P073A11)

上の例の通り、主体自体（「教育制度」）に備わる性質（「程度の高い」）について言う場合は「すぐれる」、主体（「技術家と経営家」）の生み出す技芸（「工業」）についていう場合は「優秀」というような使い分けを行っているものと見ることができる。

さらに1925年になると、2語の使い分けの基準はより微妙で繊細になり、言わば意味の深層において無意識に使い分けられていると見られるものに委容している。

日本人は人種として如何に優秀であつても努力なくしては歐米人と匹敵して、事業を經營することは出来ない。[中略]
我々は僅か五六十年の内に巧みに彼等の工業の外形を模倣し得る丈の優れた素質を有するのであるから (1925年7号「事業経営上より比較したる黄白兩人種の優劣」藤原銀次郎P019B08～I4)

この例は、「日本人」およびそれを言い換えた「我々」を全体主体とする同一の文脈で用いられているものであるが、日本人を「人種として」見る場合は「優秀」が、「素質」に着眼する場合は「すぐれる」が用いられている。「人種」とは他の人種と比較した序列の上で外側から日本人を見るものであろうが、「素質」とは日本人自体の内側に備わる性質を見るものである。

放送無線電話聴取用無線受信機を買ふに就て、注意すべき事は二つある。その一つは感度の極めて鋭敏なるべきこと、次に選擇性の優れたことである。其他廉價を要求することは云ふまでもない。茲に云ふ選擇性と云ふ言葉は初心者に取つて、あまり聞き馴れないものに相違ないが、受信機は單に感度がよいばかりでは未だ完全とは申せないのである。感度が優秀となればなるほど、他より混信妨害を受くる度も多くなる。そこでこの選擇性が最も重大問題となつて来る。(1895年2号「放送無線電話の発達とその聴き方(二)」安藤博P081A18～21)

この例でも、「(放送無線電話聴取用無線)受信機」について述べる一つの文脈のなかで、「すぐれる」と「優秀」とが使われている。「すぐれる」は「選擇性」を部分主体とし、「優秀」は「感度」を部分主体とするという違いがあるが、「選擇性」とは本文に説明があるように混信妨害を防ぐ受信機に備わる性能のことであり、「感度」とは外からの音を受け取る受信機の機能である。これも、内に備わる性質か外に向かう技芸かという違いによって、

「すぐれる」と「優秀」とを使い分けている例にほかならない。

上の二つの記事いずれにおいても、4で見た「すぐれる」と「優秀」との意味的な対立点で使い分けが行われている。こうして、1925年には、「すぐれる」と「優秀」との意味的な対立関係は十分に形成されていると見ることができる。

7

まとめ

本論文では、漢語「優秀」の定着を、〈優秀〉語彙の形成過程をたどることを通して記述した。漢語「優秀」は、主体の外側に現れる技芸などの〈優秀〉さを表す語として、主体の内部に備わる性質などの〈優秀〉さを表す和語「すぐれる」との間に、意味的な対語関係を形成していきながら、日本語の語彙に定着を進めていった。漢語「優秀」は、和語「すぐれる」との間に意味的な使い分けの関係をもったことで、語彙の基本的な部分に深く浸透したものと考えられる。この2語以外の類義の語彙については、文章語などとして周辺的な位置にとどまり、語彙の基本的な部分に深く入り込むものではなかった。定着していく新漢語については、こうした意味的な関係から見た語彙形成や、定着の深度などの観点を加えて記述していくことで、定着過程を具体的にとらえていくことができるのではないだろうか（注10）。

本論文の記述は、『太陽コーパス』から得られる用例を、共起する語の形式や内容を整理する方法を徹底することと、『太陽コーパス』から得られる、使用年、使用者、使用ジャンル、使用文体などの情報を組み合わせて分析することによって、可能になったものである。いずれの分析も、コーパスがなければ実践が困難なものであると考えられる。こうした方法をとることで、コーパスを用いた語彙研究として、従来の研究では達することができなかった新しい研究領域がひらけていくのではないだろうか。

注

- (1) 見通しのよい概観として、宮島（1967）、池上（1984）などがある。
- (2) 進藤（1981）、惣郷・飛田（1986）、森岡（1991）、浅野（1998）、田島（1998）など研究書は多い。
- (3) 柴田他（1976-1982）、森田（1989）、国広（1997）など

- が代表的な研究であり、いずれも用例を多く収集して事例に即した分析を行っているが、用例は分析者の内省を助ける働きをしているものと、見ることができる。
- (4) 大野 (1966), 阪倉 (1978), 宮地 (1979) をはじめ研究は多いが、用例の解釈を積み重ねて語の意味を記述する方法を基本にしているため、研究者によって解釈が分かれることも少なくない。古典語の意味研究においては現代語の場合以上に、用例から帰納する方法が期待されよう。
 - (5) 国立国語研究所 (1972a) (1972b) は、現代語の語彙について用例から帰納して意味用法を記述する方式を徹底した研究であり、コーパスから得られる大量の用例を用いて語の意味用法を研究する際に、参照価値のきわめて高い先行研究である。なお、古典語について解釈によらずに語の意味を導き出す方法を、田中 (2000) では「統語的方法」と呼んで提起したが、本論文もこの問題意識に続くものである。
 - (6) 本論文で用いる『太陽コーパス』は、2004年3月段階のものである。
 - (7) 頻度10というのは、帰納的な分析に堪えうる最低限の用例数として仮に設定したものである。
 - (8) 主体のほかに、程度や情態、比較対象など、〈優秀〉を意味する述語と統語的に関連づく共起語はあるが、主体に比べると表現されることは少ない。意味分析の材料としてもっとも重視すべきものは主体であると考えられる。
 - (9) D「間接的な位置」にある語句の場合、その語句が直接関与する別の述語の影響を強く受けやすいことが想定されるため、〈優秀〉語彙の意味分析の直接的な材料からは除外した。
 - (10) 田中 (2005) では、やはり『太陽コーパス』によって定着過程がとらえられる新漢語「敏感」について、意味的な語彙形成の観点から記述した。

参考文献

- 浅野敏彦 (1998) 『国語史のなかの漢語』 (和泉書院)
池上禎造 (1984) 『漢語研究の構想』 (岩波書店)
大野晋 (1966) 『日本語の年輪』 (新潮文庫)
国広哲哉 (1997) 『理想の国語辞典』 (大修館書店)
国立国語研究所 (1972a) 『動詞の意味用法の記述的研究』 (国立国語研究所報告43, 秀英出版)
国立国語研究所 (1972b) 『形容詞の意味用法の記述的研究』

- (国立国語研究所報告44, 秀英出版)
- 阪倉篤義 (1978) 『日本語の語源』(講談社現代新書)
- 柴田武他編 (1976-1982) 『ことばの意味 1・2・3』(平凡社選書)
- 進藤咲子 (1981) 『明治時代語の研究—語彙と文章—』(明治書院)
- 惣郷正明・飛田良文 (1986) 『明治のことば辞典』(東京堂出版)
- 田島優 (1998) 『近代漢字表記語の研究』(和泉書院)
- 田中牧郎 (2000) 「統語的方法に基づく語の意味研究—万葉集・八代集のカナシの分析を例として—」(『日本語学』19-11, 46-56頁, 明治書院)
- 田中牧郎 (2005印刷中) 「『敏感』の誕生と定着—『太陽コーパス』を用いて—」(『日本近代語研究』4, ひつじ書房)
- 森岡健二 (1991) 『改訂 近代語の成立 語彙編』(明治書院)
- 森田良行 (1989) 『基礎日本語辞典』(角川書店)
- 宮地敦子 (1979) 『身心語彙の史的研究』(明治書院)
- 宮島達夫 (1967) 「現代語いの形成」(『国立国語研究所論集 ことばの研究』3, 1-50頁, 秀英出版)

字順の相反する二字漢語

—「掠奪－奪掠」「現出－出現」について—吉川明日香

1 はじめに

明治期の二字漢語の中には、「掠奪－奪掠」、「現出－出現」などのように字順が逆になって使われている語が目につく。こうした字順の相反する二字漢語の組は、鈴木（1986）・田島（1998 a, b）に多く取り上げられている。この二つの先行研究で指摘された組み合わせのうち、『太陽コーパス』に両方の語が使われているものは、565組という数になる（注1）。それらのなかには、

- ・「国王－王国」「社会－会社」などのように、字順が逆になると、互いの意味が全く異なるもの。
- ・「誕生－生誕」「多数－数多」などのように、字順が逆になると、読みが異なるもの。

も見られるが、これらを除外し、字順が逆になっても読みが変化せず、意味もほぼ同一である二字漢語の中から、用例数が合計で20例以上ある組について、用例数の推移のパターンを観察した。その結果、次の二つのパターンが目立った。

A. 年代がすすむにつれ用例数が片方に偏っていく組。

B. 年代がすすんでも両方の語が使われ続ける組。

表1・表2は、各パターンから代表的な組を挙げ、年別頻度の推移を表にしたものである。

表1 パターンAの年別頻度

	掠奪	奪掠	容姿	姿容	忠誠	誠忠	苦勞	勞苦	往来	来往
1895	17	7	6	3	17	11	36	19	174	16
1901	14	5	6	1	7	1	20	5	78	4
1909	5	0	7	0	7	1	46	13	100	7
1917	9	1	0	0	10	2	53	7	57	5
1925	6	1	5	0	24	1	67	5	48	0
合計	51	14	24	4	65	16	222	49	457	32

表2 パターンBの年別頻度

	現出	出現	見識	識見	応対	対応	情実	実情	威勢	勢威
1895	40	8	22	19	6	0	14	3	10	13
1901	40	8	26	19	10	1	15	7	14	3
1909	41	19	48	22	6	1	33	29	6	5
1917	44	54	20	26	6	11	31	26	8	6
1925	22	46	12	15	7	9	16	18	4	3
合計	187	135	128	101	35	22	109	83	42	30

本稿では、この二つのパターンについて、それぞれの代表例として、「掠奪－奪掠」「現出－出現」の組を取り上げ、この二つのパターンに分かれる事情や背景について、意味・用法の観点を中心に考察することにする。

2

「奪掠」と「掠奪」一年代がすすむにつれて一方に偏るもの—

まず、年代がすすむにつれ、用例数が片方に偏っていく組について考察する。その代表事例として、「掠奪」「奪掠」を対象にする。

2.1 品詞

はじめに、「掠奪－奪掠」が用例の中でどのような品詞で使われているかを分析する。

「掠奪」

動詞

強盗が多く徘徊して、良民を苦しめ豪家を**掠奪**するので、兵隊はナカ〜々尊敬されて、(1901年10号「彼の邊の海老屋扇屋」P127B16)

名詞

肆に亞細亞民族の**掠奪**に委せられぬ。(1895年07号「蒙古大王拔都の西欧侵掠」藤田精一P078B18)

複合名詞

金牛の強力と豹の**掠奪**心とを有し、従順なる羊の如く可憐なる猫に似たり、(1895年7号「蒙古大王拔都の西欧侵掠」藤田精一P074A15)

「奪掠」

動詞

渠は已むを得ず武力を用ゐて以て其の住民より食物を徴求奪掠せり (1895年2号「紀元前の著名なる航海者 (承前)」森田思軒P049B01)

名詞

盧自からは決して奪掠などしなかつた。(1925年1号「夢の蒙古王国 (下)」五重塔人P213C12)

複合名詞

況んや兵を放ちて分割奪掠の事を行ふに於いてをや。(1895年1号「対清政策」尾崎行雄P046A10)

「掠奪－奪掠」の各品詞の割合を整理すると、表3の通りである。「掠奪」と「奪掠」とでは、品詞別の比率に大きな相違はないと判断される。

表3 「掠奪」「奪掠」の品詞

	掠奪		奪掠	
動詞	51%	26例	43%	6例
名詞	29%	15例	43%	6例
複合名詞	20%	10例	14%	2例
合計	100%	51例	100%	14例

2.2 格成分

次に、「掠奪」「奪掠」の動作について、①「誰が」、②「何を」、③「誰から (何から)」奪うのかといった格成分について、用例文に現れる語句の分析を通して整理していく。格助詞などによって格成分であることが明示されるものだけでなく、意味的に格成分に相当すると考えられるものも含めて扱うことにする。

「掠奪」

- (1) 二百四十年冬、①蒙軍は一度留布林 (今魯領に入る) を抄掠して、加里西亞 (今埃領) に歸り更に翌年惟斯拉河 (魯領波蘭の中央を横絶して北海に入る) を氷渡し②散土米爾を掠奪し庫拉加夫に迫りて、再び加里西亞に退き。(1895年9号「十三世紀に於ける蒙古民族の雄図」田岡嶺雲P079A03)
- (2) 詰り殺到せる①獨逸軍の爲めに家屋は破壊され、②財産は掠奪され、生命を斷たれると云ふ慘禍で止まるも進むも到底死を免れない窮地に陥れる爲めに③男子は勿論女子子供等

に至る迄、同じく斃れるならば坐して死を俟つよりも進んで敵を斃すに如かずと云ふ大なる敵愾心となつて今度の如き統一ある立派な働きが出来る（1917年10号「戦後の精神的施設」江木千之P082B03）

(3) ①ベルジスタンの劫掠民が今回の内亂を奇貨として、②波斯の東方を掠奪し初めるそれを抑付ける口實の下に露國が益々干涉の手を伸ばす、(1909年6号「世界之時局 波斯内亂の真相」P091B17)

(4) 留布林、庫拉古夫の固め、もとより強からざるに非ずと雖も、空し蠻民馳驅の巷となりぬ。②散土米爾の勇武いかに意氣壯なりといふとも、肆に①亞細亞民族の掠奪に委せられぬ。(1895年7号「蒙古大王拔都の西欧侵掠」藤田精一P078B18)

(5) 土耳其は洪牙利人と同じくウラルアルタイ民族の分脈なり。性質よりすれば洪牙利人と同じく遊牧武力の民なり、この①土耳其人近く東羅馬帝國を亡して餘力を以て歐洲を席卷せんとす。②歐洲を掠奪せんとするには是非共洪牙利の土地を通過せざるを得ず、(1909年6号「名士の墳洪國觀 洪國の歐洲に介在し得たる原因」P188A04)

(6) ①敵は已に白川を取りて、兩野州を占領する軍略で、先づ十分②白川城へ勢力を集中し、四方を掠奪して勢力が益熾である、③官軍は白川を取られたのは眞に味方の不利益で、殊に白川は奥州野州咽喉の地ぢや、速に恢復しようと云ふので、丁度閏四月廿四日でした、(1901年10号「追懷談」川村純義（談）P128A11)

(7) 開通式に臨んで居つた所が、③内務卿大久保利通から電報が來た、披て見ると『二月二日鹿兒島へ①暴徒が起つて、②彈藥を掠奪された、事に依ると大事になるかも知れぬ』との事だ、(1901年10号「追懷談」川村純義（談）P130B20)

「奪掠」

(8) 野蠻未開時代には①各國が殆ど戦争を以て③他國の侵畧奪掠を企るを自國の利益としたことである故國際道德なるものは殆ど行はれぬことであつたが（1901年8号「國際道德の進歩し難き所以」加藤弘之P014B20）

(9) 聯莊會匪の亂起り①馬賊益猖獗にて九連城、②鳳凰城安東縣附近に奪掠を逞うし、既に國境を越えて韓國に迫らんとす。(1901年9号「各國近事 海外事情」P215B17)

(10) 憐むべき清帝は、遠く西安に播遷して、饑饉の厄に遇ひ、

戦争の難を被りし③地方の良民は、①列國軍隊の爲めに、あらゆる暴行、殺戮、強姦、**奪掠**等の不幸を受け、(1901年1号「政治時評」国府犀東P046B04)

(11) 歐人をして東亞を蹂躪せしむるは、其治安を永遠に保持する所以に非ざるなり。且つや我れ正に反正拯物の洪圖を懷いて、③清國を膺懲するに方つては、①歐米列國は之れに容喙するの權利なし。況んや兵を放ちて分割**奪掠**の事を行ふに於いてをや。(1895年1号「対清政策」尾崎行雄P046A10)

(12) 彼れが關東派の首領として重要な位地を保ちたるは、眞に短日月間のみ。憐む可し、③彼れの馴養したる②關東派は、久しからずして①故星亨氏の爲めに**奪掠**せられたりき。(1901年10号「大井憲太郎氏」P031B21)

(13) 抑蓄妾の俗は上古に腕力を主としたる酋長部落時代に於て、彼俚談小説に大江山の酒顔子を演ずるが如く、①酋長等が暴威を以て領内或は附近の②美女を**奪掠**して、窟室穹廬の中に飲食男女の大慾を恣にしたる殘留物とす、(1895年8号「倫理の改良(三)」久米邦武P150A07)

(14) 後世に下りて、大江山酒顔童子の小説を傳誦するも、京都より山を隔てたる奥には、其比まで①野蠻の住みて②美女を**奪掠**したるによる。(1895年4号「京都は国美の庫なるを

表4 「掠奪」「奪掠」の格成分

	用例番号	①誰が	②何を	③誰(何)から
掠奪	1	蒙軍	散土米爾	—
	2	獨逸軍	財産	男子は勿論女子子供等
	3	ベルジスタンの劫掠民	波斯の東方	—
	4	亞細亞民族	散土米爾	—
	5	土耳其人	歐洲	—
	6	敵	白川城	官軍
	7	暴徒	彈藥	内務卿大久保利通
奪掠	8	各國	—	他國
	9	馬賊	鳳凰城安東縣附近	—
	10	列國軍隊	—	地方の良民
	11	歐米列國	—	清國
	12	故星亨氏	關東派	彼れ
	13	酋長等	美女	—
	14	野蠻	美女	—

論ず」久米邦武 P013A15)

上の用例で、①「誰が」、②「何を」、③「誰(何)から」の下線を付けた語句を、整理すると、表4のようになる。

用例の格成分の観点からわかることは、①は、掠奪集団・為政者(権力者)・国家・軍隊など、②は、具体的な物・人・場所(領地)、③は、民衆・為政者(権力者)・軍隊などというように、「掠奪」と「奪掠」とでほとんど同じであるということである。このことから、「掠奪」と「奪掠」とは等しい用法をもっていることが確認できる。

このように、「掠奪－奪掠」においては、「誰が」「何を」「誰から(何から)」奪うのかについて、使い分けることをしていない。このことから「掠奪」と「奪掠」が、意味・用法を区別して使われていたとは言えない。

2.1の品詞の考察と、2.2の格成分の考察とから、「掠奪」と「奪掠」とは、意味・用法が等しい語の組であることがわかった。等しい意味の語は、いずれか一方があればよい。そのために、年代がすすむにつれて一方の語に限られていくものと考えられる。

3

「現出」と「出現」一年代がすすんでも併存するもの一

次に、年代がすすんでも二つの語が併存し続ける組について考察する。その事例には、「現出－出現」を取り上げる。

3.1 品詞

まず、「掠奪－奪掠」の場合と同様に、「現出－出現」が、どのような品詞で使われているかを分析する。

「現出」

動詞

二項三項の場合を事実に見出す能はざるなり。(1895年1号「対清政策」尾崎行雄P043B09)

名詞

新にして全く新なる世界の現出を見るに至らんことを祈る。
(1901年1号「政治時評」国府犀東P047A17)

複合名詞

象徴は人の活動的現出に堪へざる者である。(1909年5号「劇詩の新傾向とメターリンク」川島金五郎P129A16)

「出現」

動詞

何處かへ姿を晦ましてほとぼりをさまし、漸く選挙前日に**出現**したと云ふ位である。(1917年6号「逐鹿異耳と新代議士」二宮楚川 P151B18)

名詞

已代治内閣の**出現**を豫想する者もあるやうぢやナウ。(1917年8号「政界の表裏 臨時外交調査会の内幕」無名隠士 P053B07)

複合名詞

物質不滅界に於て物質の**出現**消失と云へる二箇相反對せる作用の有り得る理由である。(1909年2号『自然界の三大矛盾』に就て) 遠藤吉三郎P061A03)

「現出」「出現」の各品詞の割合を整理すると、表5の通りである。

表5 「現出」「出現」の品詞

	現出		出現	
動詞	88%	165例	36%	49例
名詞	10%	18例	58%	78例
複合名詞	2%	4例	6%	8例
合計	100%	187例	100%	135例

「現出」は動詞に多いのに対し、「出現」は名詞に多く、動詞がこれに次ぐというように、品詞の分布には明らかに相違が見られる。

なお、「現出－出現」が動詞として用いられる場合は、自動詞か他動詞かの別がある。

自動詞

皇統問題は端なくも眼前に**現出**するに至れり。(1895年10号「ヲット、フォン、ビスマルク公(続)」エスボルンハーク(著) P083B12)

地平線上に帆影の**出現**することもやと佇立凝望したりといふ(1895年6号「亜歴セルカーク(下)」森田思軒P059B05)

他動詞

共に水蝕の著るしき岩石なるを以て、到る處に奇景を**現出**せり、(1895年1号「利根水源探検紀行」渡辺千吉郎P082B11)
到底我邦に歐米を凌駕する經濟時代を**出現**するの望なかるべ

し。(1917年4号「海外に輸出せる我国物品の粗製濫造に対する非難並に其の実例と之が救済策」志田鉦太郎P102A08)
表6は、自動詞・他動詞の比率と用例数である。

表6 「現出」「出現」の自他

	現出		出現	
自動詞	44%	73例	88%	43例
他動詞	56%	92例	12%	6例
合計	100%	165例	100%	49例

「現出」は自動詞にも他動詞にも用いられるが、どちらかというと他動詞の方に多い。「出現」は自動詞への偏りが著しい。動詞の自他の点からみても、「現出」と「出現」は用法を異にしていると見ることができる。

3.2 格成分

「現出－出現」の格成分については、自動詞の場合は、①「何が」②「どこに」現れるか、他動詞の場合は、①「何を」②「どこに」現すか、について分析をおこなうのが有効である。

「現出」

自動詞

- (1) 世界平和維持論は今日始めて生じたものではない。此の②地球上に①國家なるものが**現出**したると同時に經世家、軍人乃至學者、宗教家等によつて畫策せられ尚ほ考慮せられつゝある問題である。(1917年6号「戦後に於ける世界平和耐久策を論ず」泉哲P063A07)
- (2) 百年前ナポレオンとの大戦争後②露國にも①自由思想が**現出**する様になった。(1917年5号「欧州大戦と露国の革命」浮田和民 P007B18)
- (3) 戦争終熄の前後に於て交戦國又は中立國の②經濟社會に**現出**す可き①經濟上の問題は一にして足らず、(1917年4号「戦時並に戦後經濟政策」堀江帰一P061B05)
- (4) 今や①新開諸國無限の沃土は、運搬業の進歩と共に、海洋縮小し、山岳開通せられて、忽然として②歐洲内に**現出**し來りたり、(1901年1号「欧州農業界の大勢を論じ延きて我國農業の前途に及ぶ」横井時敬P014A)
- (5) 機動性の増大火砲威力の要求は②野戰に大口徑砲の使用

となり、所謂①大威力砲が現出した。(1925年1号「近代兵器の進歩並に将来の趨勢」大橋顧四郎P013A11)

他動詞

- (6) 日清兩國の海軍實力(清國北洋水師の全滅と日本艦隊の増加)今や②'東洋の海軍力に①'非常の變態を現出し、延きて其の影響を英露列國の海軍にも及ぼさんとす、(1895年3号「政治」P135A06)
- (7) 益々國債を募集するの止むを得ざるに至らしめ、其の直接の軍費に支出したる金額のみを與ふるも尚ほ殆んど五一、〇〇〇、〇〇〇に達せり然れども戦後の結果は、②'財政上に①'一大變態を現出するに至りたり、(1895年3号「政治」P141B31)
- (8) 此評を會得せば其以後政界事情を批判するに於て、是非自ら分明ならんとす、緒論第十五議會中②'政界に①'二大變象を現出せり、1901年4号「政界の二大變象」P001D11)
- (9) 然れども制限外兌換券發行の策を取れば、其結果は物價騰貴し、一時②'世間に①'好景氣を現出せしむべしと雖も、必ず之に伴ふて輸入超過、正貨流出、從つて兌換制度の動搖を來し、(1901年12号「經濟時評」小松崎筑嶺P055A10)
- (10) 多くは軍需品の製造に使用され、また多くの壯丁は戦場に送られた結果、交戦國の生産は殆んど途絶の有様であつた。この時に際して②'我國の生産界は、頓に活氣を呈し、未曾有の輸出超過の①'盛況を現出した。(1925年7号「禁酒運動の科学的基礎—特に工業能率増進に就て—」伊藤一隆P062A05)

「出現」

自動詞

- (11) 昔しは明の雲棲大師曾て曰へることあり『天堂未だ成らず、地獄先づ成る』と佛を入れずして却て①怪鬼出現す、(1901年4号「宗教時評」龍山学人P055B09)
- (12) 又風の神と稱するものあり、是れ鬼の顔色を有する偶像なるが、此風の神の由來は古代の記録によれば、古代土人の②社會に何れよりか、一人の①仙人出現し、色白く、髯長く、衣服には一面十字章の紋を附け居り、(1901年8号「歴史以前の墨西哥國(承前)」室田義文P024B25)
- (13) 世間往々大人物の拂底を訴ふ、其意蓋し偽政治家の紛々たる跋扈を厭ひ、①眞正政治家の出現を望むに在り。是れ即ち國土的政治家を求むるの聲なり。(1909年1号「政治家の

分類」P013B04)

- (14) 思ふに宗教的道德は人の頭脳に極めて深刻なる印象を與へて、其人を②信仰の野に導かでは已まぬものであるから、①大偉人の出現を俟つて始めて有効なる布教が出来る。(1909年8号「政治家の徳義問題」尾崎行雄(談) P052B07)
- (15) 自分は此の傾向を展覧會の眞の主旨に照して最も當を得たこととして賛辭を呈し、審査員諸氏の勇斷を多とする者である。徒に出品過多呼ばりをする人々に傾聴せず、飽くまで此の方針で未來ある①作家の出現を待たれんことを②斯界の爲めに祈ります。(1917年13号「文展の彫刻」戸張孤雁 P141C09)

他動詞

- (16) アリストファーンが二千年の古に、「②雲裡に①惡鬼を出現せしむ」と云へるが如く、以て衆愚の民を欺くべくして、識者を惑はすに足らざるなり、(1909年13号「我が立憲政体と大政党」上杉慎吉P042A17)
- (17) 犬養は政黨内閣といふ形式を造る爲めに、外交方針の一致した寺内内閣を倒して、外交方針の反對した①加藤内閣を出現せしめようといふのは何ういふ精神だらうか。(1917年2号「政界の表裏 政府の魂胆 政黨の策戦」P025B01)
- (18) 而して多くの貴族及び高級の僧侶は其の同類であつた。此等反逆者の手から佛蘭西を自由にし、①不羈獨立の佛蘭西を出現せんとした熱烈な愛國精神が彼等を縦貫して動かし難い結束力たらしめたのである。(1917年14号「露西亞革命の弱点」浅田江村P020B02)
- (19) 爾來選舉の際に一種の魔力が如何に有効に行はるるに至りしかは世上周知の事實にして心ある者の慨歎する所である。金力さへあれば當選の見込確實であつて、①選舉請負人の如き一種の職業を出現せしむるに至つた。(1925年5号「憲政の危機と対策」添田寿一P006C09)
- (20) 是を以て②瀕死者の尿には殆ど凡ての臓器基體に對する多様の①破壊酵素を出現せしむ、(1925年13号「尿の學問」木内幹 P128A16)

上の用例で、①「何が」、②「どこに」、①「何を」、②「どこに」の下線を付けた語句を、整理すると、表7のようになる。

表7 「現出」「出現」の格成分

	用例番号	①何が	①' 何を	②どこに	②' どこに
現出	1	國家なるもの	—	地球上	—
	2	自由思想	—	露國	—
	3	經濟上の問題	—	經濟社會	—
	4	新開諸國無限の沃土	—	歐洲内	—
	5	大威力砲	—	野戰	—
	6	—	非常の變態	—	東洋の海軍力
	7	—	一大變態	—	財政上
	8	—	二大變象	—	政界
	9	—	好景氣	—	世間
	10	—	盛況	—	我國
出現	11	怪鬼	—	—	—
	12	仙人	—	社會	—
	13	眞正政治家	—	—	—
	14	大偉人	—	信仰の野	—
	15	作家	—	斯界	—
	16	—	惡鬼	—	雲裡
	17	—	加藤内閣	—	—
	18	—	不羈獨立の佛蘭西	—	—
	19	—	選舉請負人の如き 一種の職業	—	—
	20	—	破壊酵素	—	瀕死者の尿

用例文の整理を通してわかることとして、まず、「現出」「出現」の②②'については、②の2「露國」、4「歐洲内」、②'の10「我國」などの国家を意味する語、また、②の3「經濟社會」、12「社會」、15「斯界」、②'の8「政界」、9「世間」などの社会を意味する語というように、対象となるものが具体物である用例文が目立つことがあげられる。よって、②と②'については「現出」と「出現」とで区別がないことが確認できる。

次に、「現出」「出現」の①①'を見ると、「現出」は抽象物、「出現」は具体物という区別があることが指摘できる。

「現出」の①は、例えば、1「國家なるもの」、2「自由思想」、3「經濟上の問題」のように抽象物を指す語が多い。4「無限の沃土」は、具体的場所というよりも新分野・新領域を指す抽象的

な世界であると捉えることができる。また、5「大威力砲」は、物としての大砲を指しているのではなく、新技術としての大砲を指しており、抽象物と解釈できる。「現出」の①'は、6「非常の變態」、7「一大變態」、8「二大變象」、9「好景氣」、10「盛況」など様子や状態をあらわす語が目立ち、対象が抽象物であるのがわかる。

一方、「出現」の①は、対象が、13「眞正政治家」、14「大偉人」、15「作家」など具体物であるのが確認できる。「出現」の①'は、18「不羈獨立の佛蘭西」のように、単に「佛蘭西」とするのではなく、どのような状態の佛蘭西であるかをあらわしていることから具体物と考えられる。また、19「選舉請負人の如き一種の職業」についても、どのような職業であるかを特定して説明することにより、人物を表すものと解釈でき、具体物と考えることができる。

このように、「現出」と「出現」とにおいては、「何が」「何を」に相当する格成分について、抽象物か具体物という区別があることがわかる。このことから「現出」と「出現」は、意味・用法を区別して使われていたと言える。

3.1の品詞の考察と、3.2の格成分の考察とから、「現出」と「出現」とは、意味・用法が異なる語の組であることがわかった。そのために、年代がすすんでも両方の語が使われ続けたものと考えられる。

4

おわりに

『太陽』における字順の相反する二字漢語の年代による推移について、年代がすすむにつれ用例数が片方に偏っていく組と、年代がすすんでも両方の語が使われ続ける組の二つに分け、代表事例を取り上げて、意味・用法を中心に考察をすすめた。「明治初期は漢語の流行・氾濫の時代であった」（注2）が、時間をかけて整理されていったものと推測される。字順の相反する二字漢語の場合も、意味・用法の区別がない組の場合は、一方の語は衰退し、他方の語のみが生き残っていく。これに対して、意味・用法の区別がある組の場合は、両方の語が共存して使われ続けていく。明治・大正期において、漢語語彙が整理され安定していく過程にある事情を、本稿は、具体的に示すことができたと思う。

注

- (1) 本稿の調査に用いた『太陽コーパス』は2004年3月現在のものである。
- (2) 田島 (1987) 2頁。

参考文献

- 鈴木丹士郎 (1986) 「二字漢語の字順についての問題」 (『国語論究 1 語彙の研究』 278-300頁, 明治書院)
- 田島優 (1998 a) 「字順の相反する二字漢語」 (『近代漢字表記語の研究』 316-339頁, 和泉書院)
- 田島優 (1998 b) 「『新説 八十日間世界一周』における字順の相反する二字漢語」 (『近代漢字表記語の研究』 340-374頁, 和泉書院)
- 田島優 (1987) 「江戸中期における漢語表記の安定度—表記意識資料としての『志布可起』一」 (『名古屋大学人文科学研究』 第16号, 1-16頁)

外国地名表記について

—漢字表記からカタカナ表記へ—

——井手順子

1 はじめに

本稿では、『太陽コーパス』に出現している主な外国地名表記の使用状況について調査と考察を行う。

『太陽』以前の外国地名表記については、既に多くの先行研究がある。藤本(1993)は、明治初期について、「地名・人名とも漢字表記を本体として振り仮名をつけており、この方式が当時の大勢を占めていた。」という。勿論、新井白石が西欧語の原音の発音をカタカナによって書き表そうと細かな工夫をしたように(注1)、カタカナ表記によって外国地名を表記する動きもあったが、上野(1981)、藤本(1993)、飛田(1998)などが指摘するように、明治初期、外国地名は漢字によって表記されることが大勢であった(注2)。ところが、ある時期を境に、漢字表記が大勢であったところからカタカナ表記が優勢になる。国立国語研究所(1987)は、雑誌『中央公論』において、1916年から1926年の間で、漢字表記からカタカナ表記への移行が見られることを指摘している(注3)が、『太陽コーパス』を調査すると、同様に、同時期の1917年から1925年の間で、漢字表記からカタカナ表記への変化を鮮明に見ることができるのである。

本稿は、外国地名表記が漢字表記からカタカナ表記へと移行する変化を『太陽コーパス』を利用して調査したものである。なお、調査には2004年2月段階の『太陽コーパス』を使用した。

2 『太陽』における主な外国地名表記の年代別調査

『太陽コーパス』に現れる外国地名のうち、次の基準のすべてにあてはまるものを調査対象とした。

- ・漢字表記とカタカナ表記の両方の表記が使用されている。

・1895年から1925年までの全ての年に(いずれかの表記が)出現している。

・1895年から1925年までの出現記事の総数が100以上。

この基準によって調査対象になったのは、以下に示す21地名の漢字表記とカタカナ表記である。なお、第1類、第2類、第3類の別については後述する。

第1類

- <1> アフリカ・亞弗利加・亞非利加・阿弗利加・亞佛利加
- <2> アメリカ・亞米利加・亞墨利加
- <3> イギリス・英吉利
- <4> シベリア・シベリヤ・西比利亞・西伯利亞・西伯利・西比利
- <5> スペイン・西班牙
- <6> パリ・パリー・パリイ・巴里
- <7> フランス・佛蘭西・法蘭西
- <8> ローマ・羅馬
- <9> ロシア・ロシヤ・露西亞・魯西亞
- <10> ロンドン・倫敦・龍動

第2類

- <11> アジア・アジャ・亞細亞
- <12> イタリア・イタリヤ・イタリー・イタリイ・伊太利・伊太利亞
- <13> オランダ・ヲランダ・和蘭・和蘭陀・阿蘭陀
- <14> ギリシア・ギリシヤ・グリシア・グリシヤ・希臘
- <15> ドイツ・獨逸・獨乙
- <16> ベルギー・白耳義
- <17> ベルリン・伯林
- <18> ヨーロッパ・ヨウロッパ・ヨーロッパ・歐羅巴

第3類

- <19> インド・印度
- <20> エジプト・エジプト・埃及
- <21> ニューヨーク・紐育

これら21地名において、漢字表記とカタカナ表記それぞれの使用の割合を調査した。図1(1~21)は、各地名において、漢字・カタカナの各表記がどれくらい出現するかを記事数と比率で示したものである(グラフ中の数字が記事数、グラフは比率によ

る)。頻度ではなく記事数を指標としたのは、地名の場合、記事の種類によって、同一記事内に同一語が繰り返して出現している場合があり、頻度よりも記事数の方が指標としてまぎっていると考えられるからである。年代別の使用の推移を見ようとするなら、一つの記事内に何語出現するかということは無視し、出現した記事を1として数えた方がより正確にその推移の変化を見ることができると思われる。なお、「アメリカン」などの派生語、「アメリカンインディアン」などのように原語において複合語であるものは除いた。また、グラフ内の表記は、代表的なもので示した。

これら21地名の表記を見わたすと、総じて、次のことが指摘できる。

- ① 1917年までは、漢字表記が大半を占めているが、その中でカタカナ表記の比率は少しずつ増加していく。

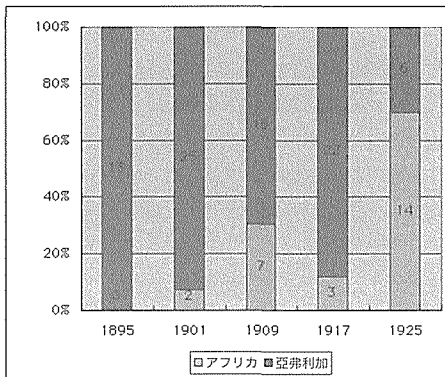


図1-1 アフリカの表記

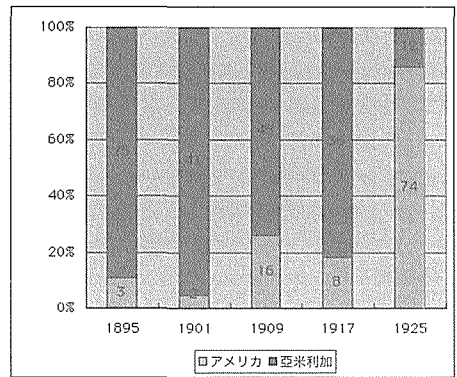


図1-2 アメリカの表記

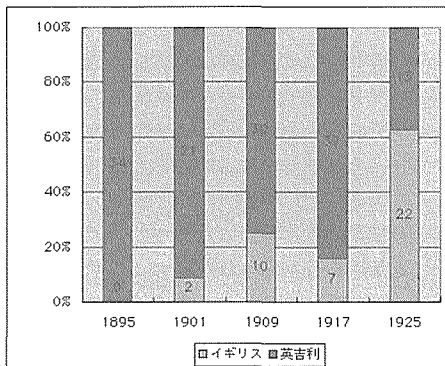


図1-3 イギリスの表記

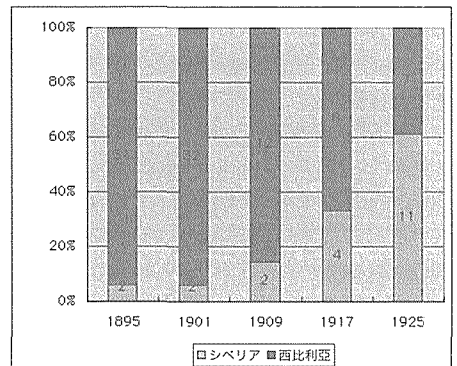


図1-4 シベリアの表記

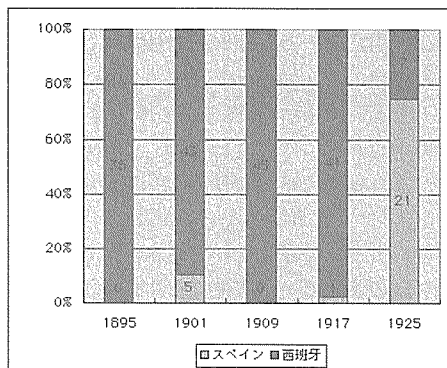


図1-5 スペインの表記

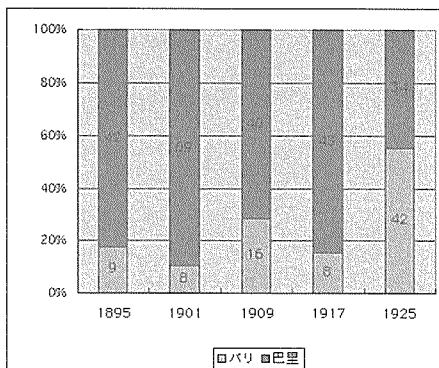


図1-6 パリの表記

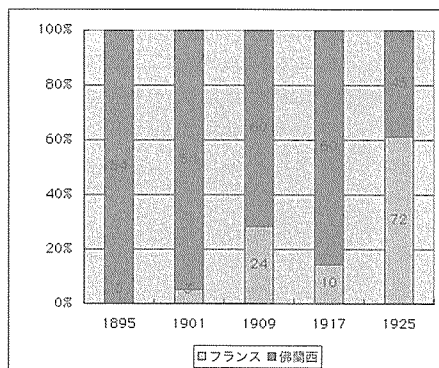


図1-7 フランスの表記

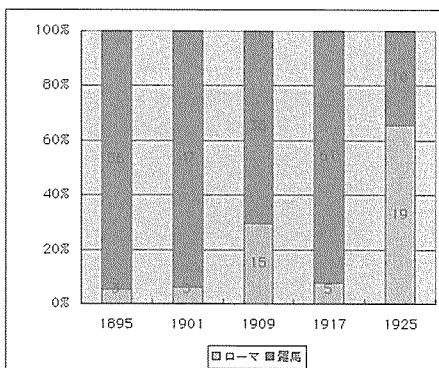


図1-8 ローマの表記

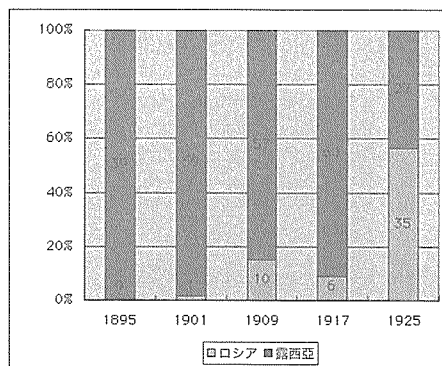


図1-9 ロシアの表記

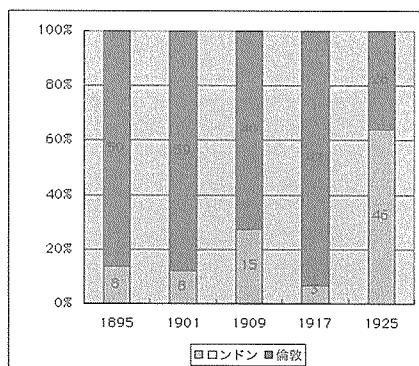


図1-10 ロンドンの表記

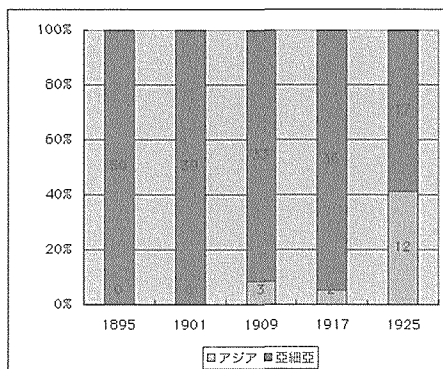


図1-11 アジアの表記

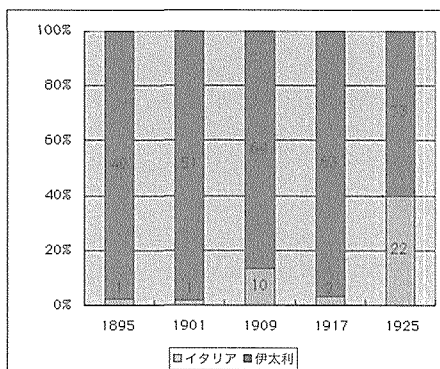


図1-12 イタリアの表記

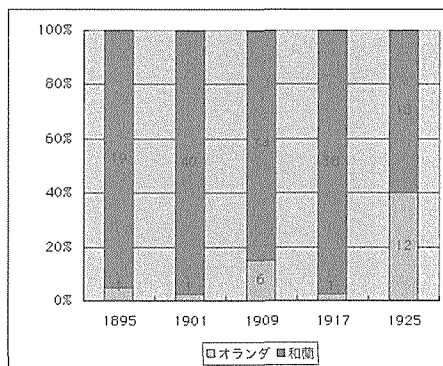


図1-13 オランダの表記

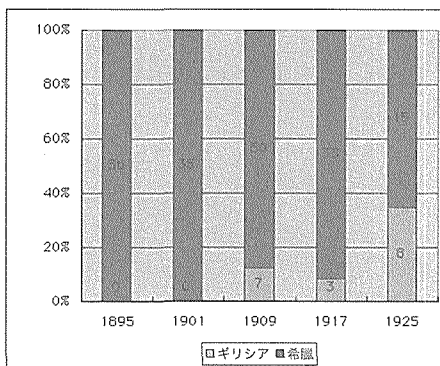


図1-14 ギリシアの表記

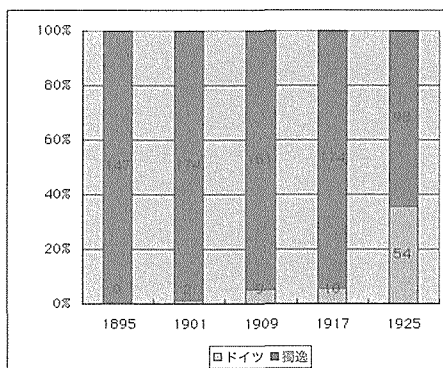


図1-15 ドイツの表記

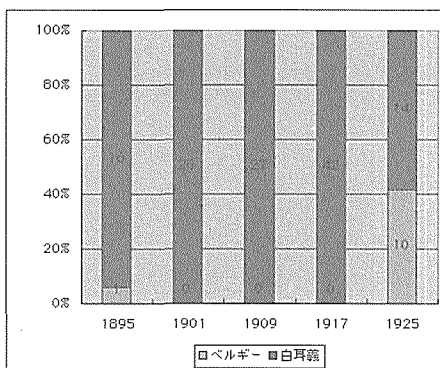
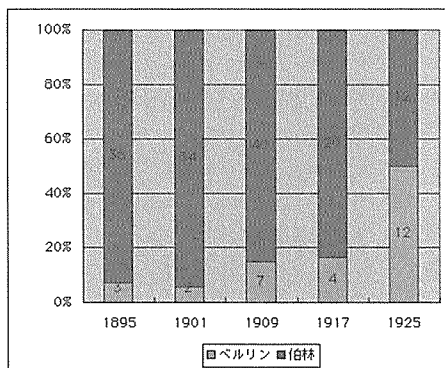
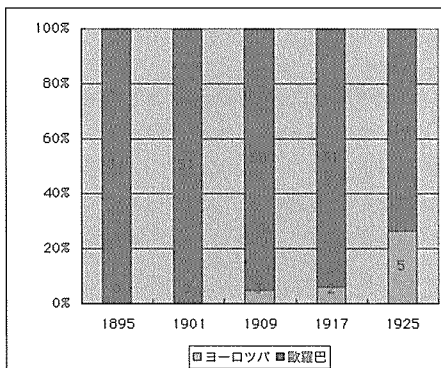


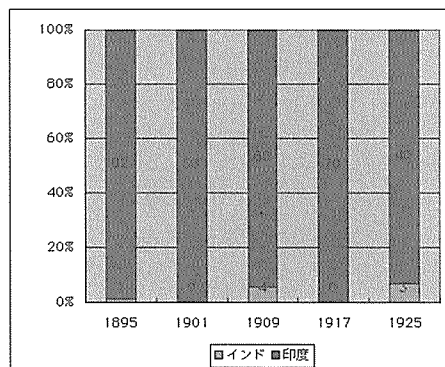
図1-16 ベルギーの表記



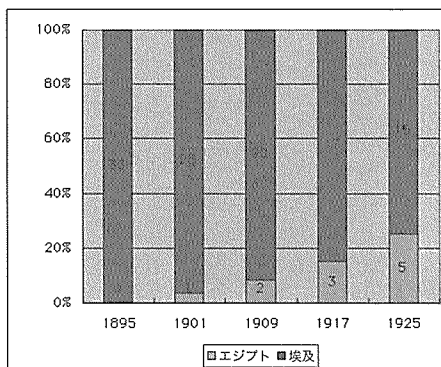
図I-17 ベルリンの表記



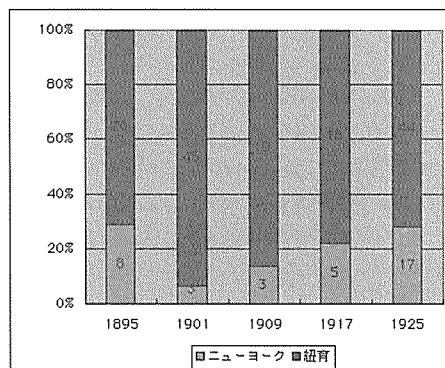
図I-18 ヨーロッパの表記



図I-19 インドの表記



図I-20 エジプトの表記



図I-21 ニューヨークの表記

② 1925年においては、漢字表記の比率は減り、カタカナ表記の比率が急激に増加する。

そして、②について詳しく見ると、21地名は、次の3つに分類できる。

第1類 1925年においてカタカナ表記が急激に増加し、漢字表記よりも優勢になる地名。

第2類 1925年においてカタカナ表記が急激に増加はするが、漢字表記を上回らない地名。

第3類 1925年においてカタカナ表記が増加はするが増加率が少なく、1925年においても漢字表記が大勢である地名(1917年から1925年へのカタカナ表記の増加率が10%以下の地名)。

1～10が第1類、11～18が第2類、19～21が第3類である。第1類、第2類が大半を占め、1925年におけるカタカナ表記の急増は、主な外国地名においては、一般的な趨勢であったことがわかる。

それに対して、第3類は、1925年においても、あまりカタカナ表記が増加せず例外となるが、これらの漢字表記は、「印度」「埃及」「紐育」といずれも漢字二字の表記であることに注目したい。第1類、第2類の地名の漢字表記の多くが漢字三字以上の表記であったことと比較すると特徴的である(注4)。すなわち、漢字二字による地名は、三字以上の地名と異なり、日本語の中で熟語として定着しやすい傾向があったのではないと思われるのである。

第3類の3語を除けば、漢字が大勢であった外国地名の表記が、カタカナ表記へ移行していくという大きな流れとその変化が、1925年において特に際立っていることが認められるのである。以下では、この大きな流れと、1925年における際立った変化について、その背景や要因も含め、詳しく見ていきたい。

これまで大勢であった漢字表記が減少し、カタカナ表記が増加することの背景として、次の三点との関わりを想定した。

- ・ジャンルとの相関
- ・文体との相関
- ・著者の生年との相関

なお、以下の調査では先に述べた2つの特徴(①②)がともにあってはまる第1類と第2類を対象とする。

3 ジャンルとの相関

ジャンルによって表記の使用に偏りはあるのだろうか。『太陽コーパス』に付与されたNDC分類を使用し、ジャンルによって表記の使用に偏りがあるのか調査した。図2（1～10）は、第1類、第2類の地名について、1895年から1925年の5カ年における記事数をNDC分類に基づき数えたものである（グラフの数字が記事数、グラフは比率による）。なお、同一記事内に複数の用例が出現する場合もあるが、記事内の用例数に関わらず、用例のある記事を1として数えた。

総記、産業、言語は年によって極端に記事数が少ないものがあり、傾向が掴みにくいため除外し、それ以外の7つのジャンルに

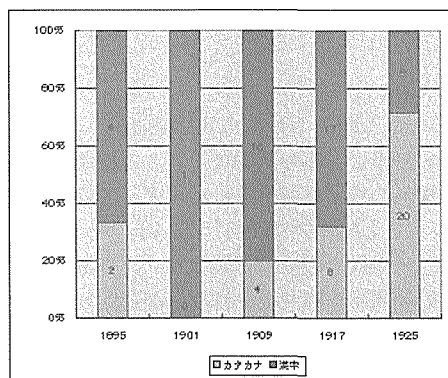


図2-1 総記

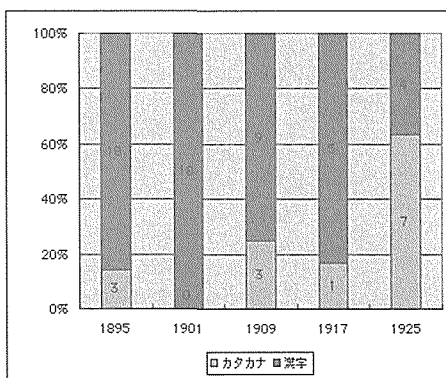


図2-2 哲学

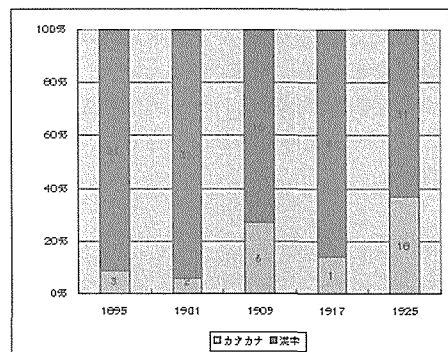


図2-3 歴史

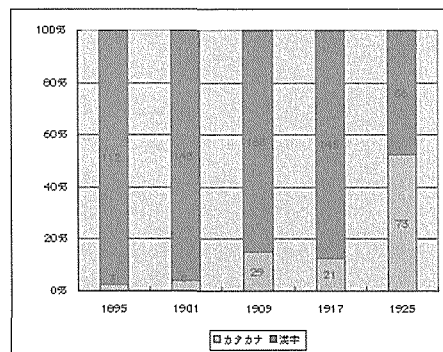


図2-4 社会科学

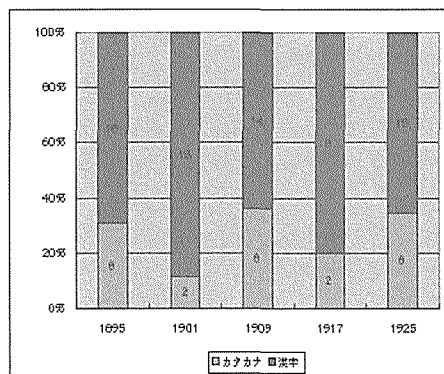


図2-5 自然科学

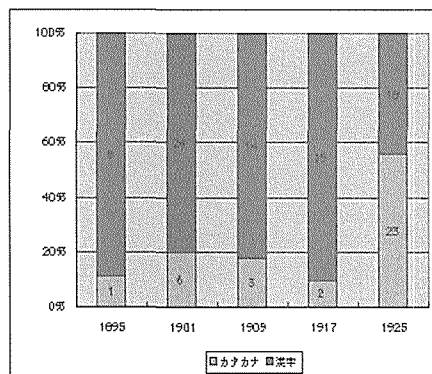


図2-6 技術

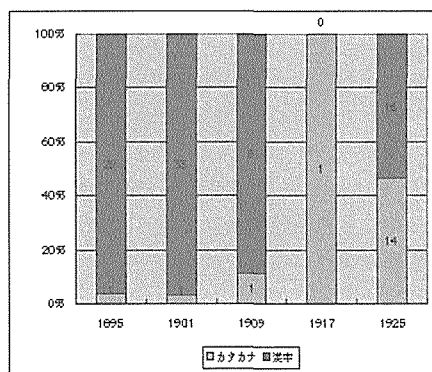


図2-7 産業

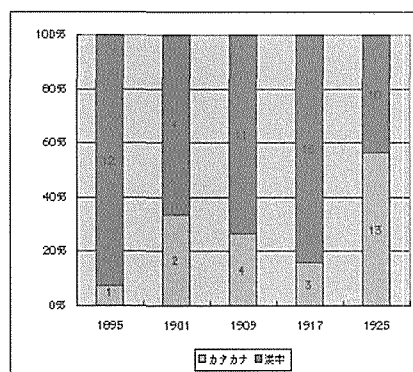


図2-8 芸術

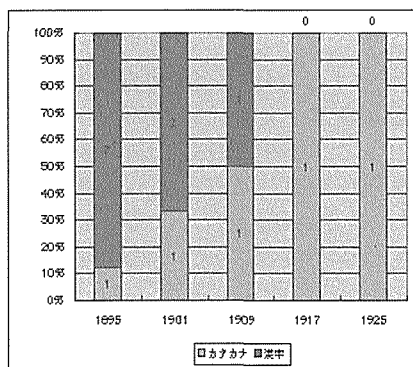


図2-9 言語

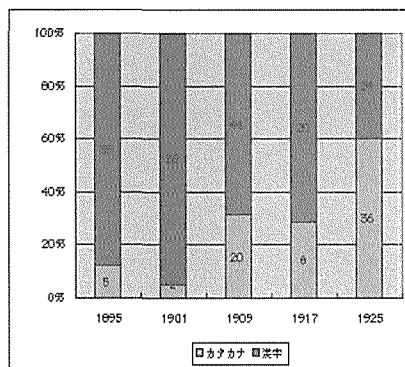


図2-10 文学

ついて見ると、次のようなことが指摘できる。総じて1895年から1917年までは漢字表記が大勢でありながら、カタカナ表記がわずかに増加傾向にあり、1925年になるとカタカナ表記が急激に増加している。この点については、先に見た、外国地名表記の全体的な傾向と一致している。ジャンル別に見ても、傾向は変わらないのである。

しかし、詳細に見ると、ジャンルによって、その増減の傾向には特徴が見られるということも言えそうである。図2によれば、1925年になると、カタカナ表記が半数以上を占めるジャンルと、1925年になっても依然として漢字表記が半数以上を占めているジャンルとに分かれることがわかる。総記、哲学、社会科学、技術、芸術、文学については、1925年でカタカナ表記が半数以上になるが、歴史、自然科学、産業については、1925年におけるカタカナ表記が占める割合は半数未満である。産業については、増加率が大きいので、前者に含めることも可能だと思うが、歴史と自然科学については、カタカナ表記への移行が遅れたジャンルとして見てよいだろう。その理由については不明であるが、ジャンルによっては、1925年におけるカタカナ表記の増加傾向が顕著でなかったものがあつたことを明確に示している。このことは、ジャンルによって表記選択の傾向に差違があるのではないか、ということを示唆している。

4 文体との相関

では、文体についてはどうであろうか。外国地名における漢字およびカタカナの表記選択の問題と文体との相関については、既に深澤(2003)が考察している。深澤(2003)は、『太陽』において、「外国地名が片仮名表記へと移行した理由は、(中略)口語文体からののはたらきかけである」と結論づけた。このことの適否を考えるため、第1類、第2類の地名について、文語体、口語体に分けて、漢字表記とカタカナ表記の比率を調査した。

全体で見ると、文語体よりも口語体の方がカタカナ表記の比率は高く、深澤(2003)の指摘はある程度まで適切と思われる。しかし、図3にあるように、文語においても、口語においても、グラフの形はよく似ている。1で見た2つの特徴(①②)は、同様に、文語においても口語においても指摘できると考えられる。文

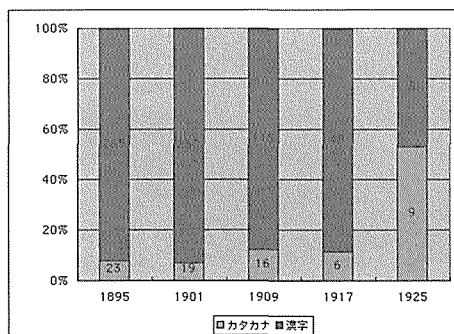


図3-1 文語

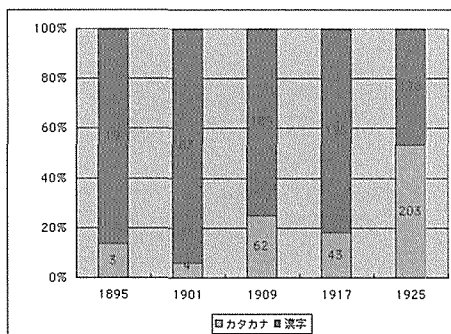


図3-2 口語

体の別が表記の選択の要因になっているとは言い難い。『太陽コーパス』における文語、口語の記事率は、文語が次第に減っていくのに対し、口語は次第に増えていく（注5）。その中でカタカナ表記が増加するので、口語体とカタカナ表記に相関性があるかのように見えるが、カタカナ表記は1925年に急激に増加するのであり、1925年において、漢字表記とカタカナ表記が占める割合は、文語においても口語においても同率である。文体の別と表記の選択とは、実際にはあまり相関性がないと言ってよいだろう。

5

著者の生年との相関

では、著者についてはどうか。図4は、『太陽コーパス』に付与された著者の生年情報を用いて、著者の生年による表記の使用傾向を調査したものである。なお、談話記事と「記者」「編集部」などある『太陽』編集関係者による記事は外した。

生年が遅い著者ほど漢字表記を使用する傾向は減り、カタカナ表記を使用する傾向が強くなっていくことがわかる。このことから、生年が早い著者ほど漢字表記を使用する傾向が強くなり、生年が遅い著者ほどカタカナ表記を使用する傾向が高まっていったことを見ることができるのである。著者の生年と表記の選択との間には強い相関があることは明らかである。すなわち、生年の遅い著者が増えたとき、漢字表記よりもカタカナ表記の使用が増えたのではないかということである。そこで、著者の生年の分布を調べた。

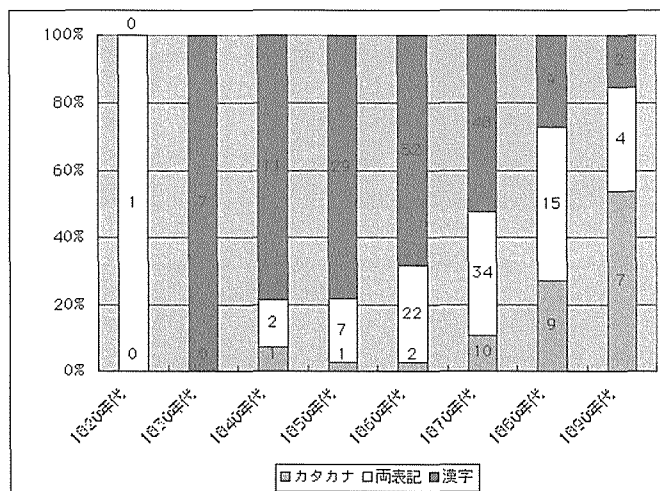


図4 著者の生年別使用率

第1類と第2類の地名を使用している著者の生年を『太陽コーパス』1895～1925年の各年ごとと比較してみると、表1のようになる。

表1 著書の生年の分布の推移 (人数)

生年 \ 出版年	1895年	1901年	1909年	1917年	1925年
1820年代	1	0	0	0	0
1830年代	6	2	1	0	0
1840年代	5	6	1	1	2
1850年代	20	12	3	2	5
1860年代	25	22	15	14	16
1870年代	14	15	29	24	24
1880年代	0	0	1	11	19
1890年代	0	0	0	0	12

生年の遅い著者が、時代が下がるにつれて増加していくことがわかるが、特に、1925年における生年の遅い著者の増加は顕著である。時代が下がるにつれ、生年の遅い著者が増えていくのは当然であるが、1925年での生年の遅い著者の著しい増加は、1925年においてカタカナ表記が急激に増加する傾向と一致する。

つまり、1925年でカタカナ表記が急激に増加したのは、生年の遅い著者が増加したこととの関係が強いのではないと思われるのである。

さらに、1925年でのカタカナ表記増加の要因は、生年の遅い著者の増加だけにとどまらない。図5は、各年代の著者の使用する表記が出版年によってどのように推移していくかを見たものであるが、1925年においては、幅広い年代に、カタカナ表記のみの使用と漢字・カタカナ両表記の使用とが広がっていることがわかる。このことは、カタカナ表記の使用が、生年の遅い著者だけでなく、以前なら漢字表記を主として使用していた生年の早い著者層にまで広がっていることを示している。1925年では、カタカナ表記が、幅広い年代によって使用されるようになっていたと思われる。

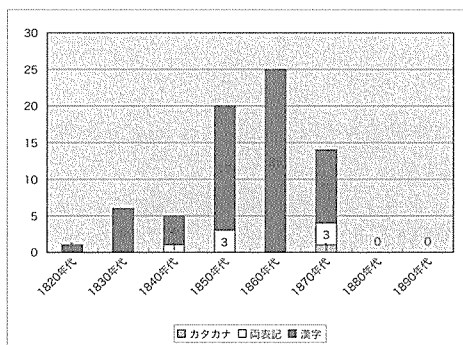


図5-1 生年別使用状況（1895年）

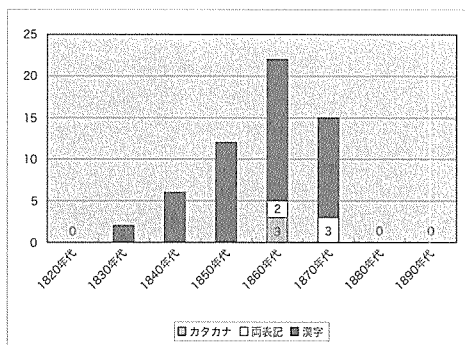


図5-2 生年別使用状況（1901年）

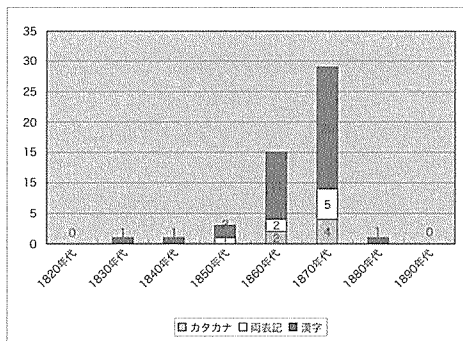


図5-3 生年別使用状況（1909年）

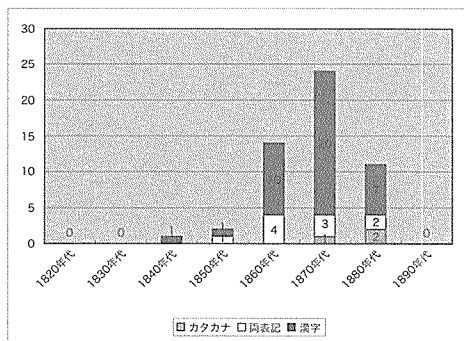


図5-4 生年別使用状況（1917年）

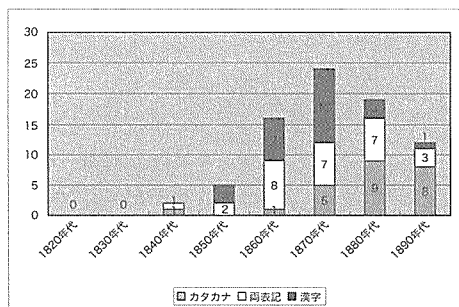


図5-5 生年別使用状況 (1925年)

6 まとめ

『太陽』における主な外国地名表記について考察してきたが、以下のことがわかった。

主な外国地名表記は、

- 1 1917年までは漢字表記が大勢である。カタカナ表記は少しずつ増加する。
- 2 1925年になるとカタカナ表記が急激に増加し、ほとんどの漢字表記は減少する。
- 3 1925年におけるカタカナ表記急増の要因は、生年の遅い著者の増加と、(それまで漢字表記を主に使用していた) 生年の早い著者層へのカタカナ表記使用の広がりによる。

以上、大きく三つのことがわかった。なお、以下の点については、今度の課題としたい。

- 1 ジャンルによる表記選択の傾向の差違は、何によるものか。
- 2 生年の遅い著者の増加がカタカナ表記の増加に結びついた事情としてどのようなことがあったか。また、生年の早い著者層にもカタカナ表記が広がった背景は何か。
- 3 日本語の歴史のなかで2の事情や背景は、どのような意味をもつか。

注

- (1) 松村(1977)による。
- (2) 松村(1977)によれば、古くは、本文が平仮名であれば、外国語・外来語も平仮名、本文が片仮名であれば、外国語・外来語も片仮名であったという。
- (3) 国立国語研究所(1987)は、1906年から1976年の10年ごとの『中央公論』を調査し、「アメリカ」と「阿米利加」のようにカタカナと漢字の両表記が出現する語の用例数を報告している。その結果、1916年と1926年の間で漢字表記とカタカナ表記の使用数が逆転し、1926年以降、漢字表記よりもカタカナ表記の方が多く使用されるようになることを指摘した。
- (4) 第1類、第2類には「巴里」「羅馬」「倫敦(龍敦)」「伯林」「希臘」「獨逸(獨乙)」のように、漢字二字の表記の地名も含まれてはいる。しかし、18地名中12地名が漢字三字以上であり、第1類、第2類では漢字三字以上の地名が多くを占めていると言ってよいだろう。
- (5) 本書「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」(田中牧郎)所収の、「表10『太陽コーパス』の年次別文体別記事数」(28頁)に拠る。

参考文献

- 上野力(1981)「明治初期の外国地名表記」(『常葉学園短期大学紀要』13, 23-30頁)
- 国立国語研究所(1987)『雑誌用語の変遷』(国立国語研究所報告 89, 秀英出版)
- 飛田良文(1998)「外来語の取り入れ方の変化」(『日本語学』17-6, 29-38頁, 明治書院)
- 深澤愛(2001)「雑誌『太陽』創刊号における外国地名片仮名表記」(『国語文字史の研究 6』, 187-218頁, 和泉書院)
- 深澤愛(2003)「漢字平仮名交じり文中における表記の選択」(国立国語研究所編『日本語科学14』, 29-52頁, 国書刊行会)
- 藤本光(1993)「外国地名・人名の表記統一」(『言語生活』266, 65-72頁, 筑摩書房)
- 西浦秀之(1970)「近世に於ける外国地名称呼について」(『皇学館大学紀要』8, 227-324頁)
- 西浦秀之(1971)「幕末・明治初期の新聞にあらわれた外国地名称呼・表記について」(『皇学館大学紀要』9, 151-202頁)

松村明(1977)「新井白石と外国語・外来語の片仮名表記」(『松
村明教授還暦記念 国語学と国語史』,1093-1136頁,明治書院)

逆接の接続詞・接続語句

——馬場 俊臣

1 はじめに

『太陽コーパス』を利用することで、雑誌『太陽』に使用された様々な語彙の状況を把握することができる。本稿は、逆接の接続詞及び接続語句（以下併せて「接続語句」と呼ぶ）（注1）を対象として、出来るだけ網羅的な一覧を作成し、これに基づいて、1895～1925年の時期の、多用される逆接の接続語句の変化（特に文語的な語から口語的な語への変化）及び多用される口語的な語の特徴を、出現頻度の経年変化に基づいて分析することを主な目的としている。

2 一覧作成の意義と方法

2.1 作成の意義

研究を進めるためには、その基礎資料として、どの時期にどのような接続語句が用いられたかが分かる一覧があることが非常に有益である。そのような一覧として、青木（1973）（以下「青木一覧」と呼ぶ）があるが、明治・大正期の調査資料の量は極少ない（注2）。また、この時期の接続語句については、「青木一覧」以外に、個別的な作品（主に文学作品）や作家を対象とした研究において、使用された接続語句の一覧を示した研究はある（注3）。ただし、言語量や文体・ジャンル等の面で、『太陽コーパス』ほど膨大且つ多様ではない。

なお、「逆接」の接続類型に限定したのは、作業量の制約のためである。今後、「順接」など他の接続類型に関する調査も必要である。

2.2 作成の方法

2.2.1 検索語の選定

『太陽コーパス』には、品詞情報が含まれていない。そのため、逆接の接続語句を検索するには、逆接の接続語句と思しき語形(検索語)を検索文字列として指定する必要がある。本調査では、多様な逆接の接続語句が載せられたリストである「青木一覧」を利用して検索語を選定した。

ここで生じる問題は、「青木一覧」に載っていない語句は、調査対象から漏れるということである。この欠点を出来るだけ補うために、正規表現を利用し検索対象の曖昧さを増し、「青木一覧」に載っていない語形も拾えるよう工夫した。また、検索結果の選定の際にも「青木一覧」に載っていない語形に注意を払った。ただし、これはあくまでも部分的な対処に過ぎず、品詞情報が含まれないコーパスの利用上の限界でもある。

さて、「青木一覧」には、接続語句の接続類型(意味分類)は示されていない。そのため、筆者の判断で「逆接」の認定を行った。その際、逆接か否か問題となる語(「それとはんたいに(対比的)」等)、複数の用法を持つ語(「さるを(逆接・転換)」等)、逆接としての用法が見つかる可能性が殆どない語(「それに」等)も含めて、やや緩やかに認定した。

こうして抽出した語は、211語である。さらに、同時期の国語調査委員会編纂『口語法』(1916年)及び同『口語法別記』(1917年)の「接続詞」の項に掲載されていないながらこの211語に含まれていない「そうだけれども、そうですが」の2語を加え、計213語を検索語とした。

2.2.2 検索文字列の作成

この検索語213語に対して、次の要領で正規表現を用いた検索文字列を作った。

まず、表記に関しては、漢字表記の候補を出来るだけ複数想定する、複数の送り仮名の可能性を想定する、平仮名・片仮名の両方を想定するなど、様々な表記の可能性に注意を払った。

次に、短い語形の場合、無駄な検索結果が多数出てくるので、それを予め排除し効率的に作業を進めるために、文頭及び句頭と思われる個所、すなわち「,。?!…——「」『』()《》●◎○▲△▼▽☆★□■◇◆」(◆の後の一字空きは全角スペースを表す)に後続する場合に限定して検索した。ただし、「国定読本」

(国立国語研究所(1997)『国定読本用語総覧 CD-ROM版』)に用いられた接続詞(「が」は除く),『口語法』『口語法別記』の「接続詞」の項に掲載された語(及びそれらに語形の類似する若干の語)は,この時期の代表的な接続詞と考えられ詳細に調査できるように,文頭・句頭に限定せず検索した(注4)。

実際の検索文字列は,例えば次のようになる。

- ・さはいえ [。? ! … — 「」『』()《》●◎○▲
△▽▽☆★□■◇◆][さサ然爾左去][ははわワ]?[いい
ゆユ言云謂]
- ・しかしながら [しシ][かカ][しシ][なナ][がガ][ら
ラ][然併而][かカ]?[しシ]?[なナ][がガ][らラ][しシ][か
カ][しシ]乍[然併而][かカ]?[しシ]?乍[然併而]

2.2.3 文字列検索の実行

文字列検索ツールは『たんぽぽ』を用いた。検索の条件は,「全ての文体 全ての欄 ルビなしテキスト」とした。「ルビなしテキスト」のみを対象としたのは作業量を考慮したためであるが,前項で述べたように様々な表記の可能性を想定し検索文字列を工夫することにより,この方法で殆どの用例を検索できると思われる。

検索に用いたデータは,1895年分は2002年5月段階の『太陽コーパス』,1901・1909・1917・1925年分は2000年11月段階の『太陽コーパス』である。調査時期の違いから異なる版のデータを用いたが,結果には大きな影響は出ないと考えられる。

2.2.4 検索結果の選定

検索の結果得られた用例から逆接の接続語句に該当しない用例を除いた。

該当例か否かの判断に関して補っておく。

まず,句読点のない記事で,次例のように,接続詞か接続助詞かの判定がつかない用例がある。これらは,すべて接続助詞と見做し除外した。

行爲の結果たる文明開化は少しも望ましくないと云ふならば
嗚呼吾復何をか言はんであるが併し幾ら道德學者が左様なこ
とを望んだとて到底其様な事實のあるべきものでない
(1901年8号「国際道德の進歩し難き所以」加藤弘之
PO11B16)

次に、「それが」のような指示詞を含んだ形式では、次例のように、指示詞本来の使われ方をしているのかどうか判断に迷う用例がある。迷う場合は、除外した。また、逆接か否かの解釈に揺れが生じる例もあるが、筆者が個別に判断した。

どんなものに出くはさうとしてゐるのか、すぐお判りになりますよ。』『それが判りさうありませんですね。(1925年4号「生ける死」佐野慶介(訳) P142A13)

最後に、漢字の読みに関して、特に「然れど」「然れども」を「されど・されども」「しかれど・しかれども」のどちらで読むべきか問題となる。この2語に限り、全数調査を行い、ルビが付いている場合はそのルビに従い、ルビが付いていない場合は「然れど」は「されど」、「然れども」は「しかれども」と読むことにした(注5)。

2.2.5 一覧の作成

選定した逆接の接続語句の異なり語数は124語である。

5年分の総出現頻度が100以上の17語については、年次別の出現頻度を表1に示す。また、残りの107語については、5年分の総出現頻度のみを表2に示す。いずれも見出し語(現代仮名遣い)の50音順である。なお、「青木一覧」欄の「◎」「○」印は、「青木一覧」に記載のある語形であることを示す(記号の区別は3.1参照)。

表1 総出現頻度100以上の語の年次別出現頻度

青木 一覧	見出し語	1895年	1901年	1909年	1917年	1925年	合計
◎	が	10	21	138	234	467	870
◎	けれど	6	12	42	68	54	182
◎	けれども	26	29	271	140	323	789
◎	これにほんして	43	75	80	63	28	289
◎	さりとて	20	26	37	22	19	124
◎	さりながら	38	12	47	16	11	124
◎	されど	318	440	339	38	24	1159
◎	されども	86	19	5	2	2	114
◎	しかし	193	382	1178	1312	1663	4728
◎	しかしながら	66	78	291	294	272	1001
◎	しかるに	586	535	706	614	410	2851
◎	しかれども	1311	791	490	138	26	2756
◎	それでも	20	28	71	123	140	382
◎	だが	17	32	46	39	333	467
◎	だって	8	30	34	21	50	143
◎	でも	2	19	69	50	169	309
◎	ところが	22	86	170	155	384	817

表2 総出現頻度99以下の語の総出現頻度

青木 一覧	見出し語	合計	青木 一覧	見出し語	合計	青木 一覧	見出し語	合計	青木 一覧	見出し語	合計
○	かかれど	1	○	しかるあいだ	1	○	それが	95		であるが	2
◎	かくても	9	○	しかるあいだに	2	◎	それがそれ	1		であるにもかかわらず	1
◎	けど	8	○	しかるところ	14		それかというて	1	◎	ですが	28
◎	けども	6		しかるところが	3	◎	それかというて	10	◎	ですけど	3
◎	これでも	6		しかるにもかかわらず	13	○	それだって	4		ですけども	2
	これにはんし	72	◎	しかるも	41	○	それだっても	3	◎	ですけれど	3
	これにもかかわらず	2	◎	しかるを	58	◎	それなのに	23	◎	ですけれども	2
○	さあれ	3	◎	しかれど	2	◎	それでいて	33		ではありますが	1
	さあれど	1	○	したが	8	◎	それととも	11	◎	ではあるが	1
◎	さはあれ	3	○	じゃが	1	○	それとはんたいに	2		ではあるけれども	1
	さはあれど	1	○	そういっても	1	◎	それなのに	9	○	でもございましょうが	1
◎	さはいえ	18	○	そうかつて	1	◎	それにはんし	3	◎	というが	2
◎	さはいえど	4		そうかというて	1	○	それにはんして	9	◎	というて	1
◎	さはさりながら	1	◎	そうかというて	12	◎	それにもかかわらず	50		といったところで	4
◎	さはとて	1	◎	そうかとて	1	◎	それにもかかわらず	1	◎	といいて	44
	さりけれども	3		そうかともうして	2		それはそうだけれど	2	◎	といいても	13
○	さりてでは	20		そうだけれど	1		それはそうでしょうが	1	○	ところを	3
○	さるに	34		そうでなくとも	1	◎	それもそうだけれど	1	◎	とはいひものの	11
○	さるを	27		そうでなくとも	2	◎	それを	50	◎	とはいえ	29
◎	さればというて	15		そうでなくとも	2	◎	だけれど	31		とはいっても	3
◎	さればとて	47	○	そうでもあらうが	1	◎	だけれども	12	○	ともうして	2
◎	さわれ	26		そじゃが	1	◎	だけれど	12		ともうしても	1
○	しかずがに	3	◎	そのくせ	47	◎	だけれども	3		ともうすものの	1
	しかども	1		そりゃあそうですが	1		だけれど	5	○	なれど	2
○	しかはあれど	5		そりゃそうだが	1		だけれども	1	◎	なれども	6
	しかはあれども	1		そりゃそうですが	1		だつても	2		にもかかわらず	5
◎	しかりといえども	34		そりゃそうやれど	1	○	だのに	4			

3

逆接の接続語句一覧の傾向分析

本章では、表1・表2の一覧の語を対象として、『太陽コーパス』における逆接の接続語句の特徴を次の3点から分析する。

- (1) 逆接の接続語句の多様性——「青木一覧」との比較
- (2) 逆接の接続語句全体の経年変化
- (3) 文体からみた高頻度の逆接の接続語句

3.1 逆接の接続語句の多様性——「青木一覧」との比較

一覧の逆接の接続語句が、どの程度多様性に富んでいるかを見るために、「青木一覧」と比較していく。まず、表3に、『太陽コーパス』『青木一覧』及び検索語の異なり語数の内訳を示しておく。

表3 『太陽コーパス』『青木一覧』及び検索語の異なり語数の内訳

『太陽コーパス』にある語 計124語	『太陽コーパス』のみにある語	36語	検索語 計213語
	両方にある語	88語	
	『青木一覧』のみにある語	125語	

3.1.1 『太陽コーパス』にはあるが、「青木一覧」にない語

『太陽コーパス』で用例が見つかった124語のうち、「青木一覧」に記載のない語は36語（124語の約3割）である。表2の「青木一覧」欄が空欄になっている語がこれに該当する。

「これにはんし（72例），しかるにもかかわらず（13例）」以外は，いずれも1～3例程度しか用例がない。連語も多く，また「さはあれ，しかるところ，そうかといって」等の類似の語形が「青木一覧」にある語もある。

ちなみに，この36語のうち，「だけんど，だつても，といったところで」を除く33語は『日本国語大辞典 第2版』（以下『日国2版』と略す）にも立項されていない。また，「だけんど，といったところで」の2語は，『日国2版』での用例よりも早い1901年の用例を『太陽コーパス』で見つけることができる。

以上のように，「青木一覧」及び『日国2版』にない語形がある程度見つけることができること，また，用例の多少の後先はそれほど重要ではないにしても，早い時期の用例を見つけることができることなど，『太陽コーパス』の有用性の一端が窺える。

3.1.2 『太陽コーパス』にも「青木一覧」にもある語

『太陽コーパス』で用例が見つかった語のうち，88語は「青木一覧」にも記載がある。これらの語は，近世以前から使い続けられた語，または明治・大正期に新たに使われた語のいずれかであることが当然予想される。果たして「青木一覧」の時期表示を見ると，近代以後（明治・大正まで）の表示があるのは，63語である。表1・表2の「青木一覧」欄の「◎」を付した語がこれに該当する（注6）。

残りの25語は，「青木一覧」欄の「○」を付した語である。

この25語のうち，「だのに」の1語は，「青木一覧」では欄外に記載され時期表示が書かれていない。

また，「そうかつて，それとはんたいに，それにはんして」の3語は，「青木一覧」の時期表示は昭和だけである。しかし，『太陽コーパス』での初出は，それぞれ，1909年，1909年，1985年である。なお，『日国2版』では，「そうかつて」の用例は「雪

国[1935-47年]」であり、「それとはんたいに」「それにはんして」は立項されていない。

残りの21語は、「青木一覽」の時期表示が近世以前までの語である。「さるに（口語記事1例、文語記事33例）、さるを（文語記事27例）、しかはあれど（文語記事5例）」等のように主に「文語」表示の記事で用いられている語もあるが、「さりとは（口語記事5例、文語記事15例）、しかるところ（口語記事12例、文語記事1例、混在記事1例）、したが（口語記事6例、文語記事2例）、それが（口語記事88例、文語記事5例、混在記事2例）」のように「口語」表示の記事でも多く用いられている。

以上のように、『太陽コーパス』によって、「青木一覽」の時期表示を修正することができ、また、より早い時期の用例を見出すこともできる。

3.1.3 『太陽コーパス』にはないが、「青木一覽」にある語

「青木一覽」に記載がありながらも、『太陽コーパス』に用例のない語は125語であるが、その約8割（102語）（注7）は、「青木一覽」の時期表示が近世以前のみであり、これは十分予想されることである。

一方、近代以後にも時期表示がなされている語は21語ある。この21語について、『日国2版』を参照しながら、特徴を整理してみる。

まず、「そうだけれど（多情多恨[1896年]）、それでいながら（都会の憂鬱[1923年]、暗夜行路[1921-37年]）、といったところが（浮雲[1887-89年]、吾輩は猫である[1905-06年]）、とはいながら（破戒[1906年]）」の4語は、『日国2版』に立項されており、それぞれ括弧内の作品の用例が載せられている。これらは『太陽コーパス』にも現れることが十分期待される語である。また「そうしたところで、そうはいうものの、というものの、としたところで」の4語も、『日国2版』には立項されていないが、同じく現れることが期待されそうな語である。（注8）

また、「そいっても、そういえばといって、そにも、でしたが」の4語は、『日国2版』に立項されていない。「さるは、それだが」の2語は立項されているが、明治以降の用例は載せられていない。「さばれ」の1語は「改正増補和英語林集成[1886年]」の用例が載せられている。これらの7語については、「青木一覽」には記載されているが、明治以降の使用の実態について検討が必要であ

ろう。

さらに、「さらば（逆接・順接）、それに（逆接・順接・添加）」の2語は、複数の用法があり、『日国2版』には明治以降の逆接の用例はない。

残りの「そうじゃないけれど、そうでなかったって、それから、それだとして」の4語は、「青木一覽」の時期表示は昭和のみである。また『日国2版』には立項されていない。

以上のように、「青木一覽」の時期表示に明治以降の記載がありながらも、『太陽コーパス』に用例のない語はいくつかのタイプに分かれ、その中には『太陽コーパス』に現れることが期待される語もある。『太陽コーパス』は膨大且つ多様な語彙を含むが、勿論完全に網羅的ではない。

3.2 逆接の接続語句全体の経年変化

本節では、逆接の接続語句の全体的な経年変化の傾向を分析する。

3.2.1 異なり語数・延べ語数の変化

表4は、各年の異なり語数・延べ語数及びそれぞれの総文字数に対する比率（％）を示したものである。総文字数を用いたのは、語数は算出できないので、その代用とするためである。なお、英数字・仮名・漢字を文字として数えた。

異なり語数の実数の変化には、大きく増加あるいは減少という傾向は見られない。ただし、比率を見るとごくわずかであるが増加傾向にあるようである。

また、延べ語数の実数を見るとやや増加傾向にある。比率でもやや増加傾向にあることが分かる。

このように接続語句の使用は、年次を追うごとに増える傾向にあり、近代語の特徴の一端を示していると思われる。

表4 各年の異なり語数・延べ語数及び総文字数に対する比率

年	異なり語数	比率(%)	延べ語数	比率(%)	総文字数
1895年	63	0.00203%	3009	0.09701%	3101732
1901年	64	0.00208%	2805	0.09112%	3078205
1909年	76	0.00265%	4209	0.14647%	2873581
1917年	56	0.00221%	3516	0.13861%	2536619
1925年	67	0.00232%	4689	0.16247%	2886149

3.2.2 出現頻度の高い上位10語

表5は、出現頻度の高い上位10語を年次別に取り出して示した表である。各年の逆接の接続語句全体の出現頻度に対する比率(%)を示すとともに、上位5位及び10位までの比率の合計も示した。

まず、「上位5位までの比率合計」「上位10位までの比率合計」を見る。「上位5位までの比率」は、年により多少変動はあるが約80%~70%であり、各年とも出現頻度の高い上位の語が頻出していることが分かる。さらに「上位10位までの比率」では各年とも約90%に達している。

前節で『太陽コーパス』には多様な逆接の接続語句が現れることを指摘したが、このように頻出する語は限られている。

表5 出現頻度の高い上位10語（年次別）

順位	1895年	出現頻度	比率(%)	1901年	出現頻度	比率(%)	1909年	出現頻度	比率(%)
1	しかれども	1311	43.6%	しかれども	791	28.2%	しかし	1178	28.0%
2	しかるに	586	19.5%	しかるに	535	19.1%	しかるに	706	16.8%
3	されど	318	10.6%	されど	440	15.7%	しかれども	490	11.6%
4	しかし	193	6.4%	しかし	382	13.6%	されど	339	8.1%
5	されども	86	2.9%	ところが	86	3.1%	しかしながら	291	6.9%
6	しかしながら	66	2.2%	しかしながら	78	2.8%	けれども	271	6.4%
7	これにはんして	43	1.4%	これにはんして	75	2.7%	ところが	170	4.0%
8	さりながら	38	1.3%	だが	32	1.1%	が	138	3.3%
9	しかりといえども	31	1.0%	だって	30	1.1%	これにはんして	80	1.9%
10	けれども	26	0.9%	けれども	29	1.0%	それでも	71	1.7%
		全体の出現頻度	3009	全体の出現頻度	2805		全体の出現頻度	4209	
		上位5位までの比率合計	82.9%	上位5位までの比率合計	79.6%		上位5位までの比率合計	71.4%	
		上位10位までの比率合計	89.7%	上位10位までの比率合計	88.3%		上位10位までの比率合計	88.7%	

順位	1917年	出現頻度	比率(%)	1925年	出現頻度	比率(%)
1	しかし	1312	37.3%	しかし	1663	35.5%
2	しかるに	614	17.5%	が	467	10.0%
3	しかしながら	294	8.4%	しかるに	410	8.7%
4	が	234	6.7%	ところが	384	8.2%
5	ところが	155	4.4%	だが	333	7.1%
6	けれども	140	4.0%	けれども	323	6.9%
7	しかれども	138	3.9%	しかしながら	272	5.8%
8	それでも	123	3.5%	でも	169	3.6%
9	けれど	68	1.9%	それでも	140	3.0%
10	これにはんして	63	1.8%	けれど	54	1.2%
		全体の出現頻度	3516	全体の出現頻度	4689	
		上位5位までの比率合計	74.2%	上位5位までの比率合計	69.5%	
		上位10位までの比率合計	89.3%	上位10位までの比率合計	89.9%	

次に、頻出する語の、年による違いを概観する。各年ともに上位10位に入っているのは、「しかし、しかるに、しかしながら、けれども」の4語である。

「しかし」は、1895年・1901年で4位、1909年～1925年で1位であり、上位で安定して使われている。「しかるに」も1895年～1917年で2位、1925年で3位であり、同じく上位で安定して使われている。「しかしながら、けれども」はそれぞれ3～7位、6～10位の範囲にあり、次いで上位で安定して使われている。

年次を追うごとに順位が下がっていくのは、「しかれども、されど」（それぞれ1917年、1909年まで10位以内）、「これにはんして」（1917年まで10位以内、ただし低位）、「されども、さりながら、しかりといえども」（1895年のみ10位以内）である。

一方、順位が上がっていくのは、「ところが」（1901年以降10位以内）、「だが、が、それでも、けれど」（1901年以降に10位以内に少なくとも2回入っている）である。（注9）

このように、多用される語に変化が見られる。次節で、それぞれの語の文体的特徴を明らかにしながらより詳しく考察していく。

3.3 文体からみた高頻度の逆接の接続語句

本節では、出現頻度の特に高い代表的な語を対象とし、それぞれの語の特徴を考察する。特に文体（口語的・文語的）の違いに注目し、文体別に各語の出現頻度の増減の経年変化のパターンを見ていく。また、口語的な語については使用された欄の違いに注目し特徴を明らかにする。

対象とする語は、5年分の総出現頻度が200以上の語（以下「高頻度語」と呼ぶ）とする。「しかし、しかるに、しかれども、されど、しかしながら、が、ところが、けれども、だが、それでも、でも、これにはんして」の12語がこれに該当する（注10）。

3.3.1 文体（口語的・文語的）の違い

まず12語が、「口語」表示の記事と「文語」表示の記事のどちらに偏って用いられているかを見る。

表6は、各語の5年分の出現頻度を口語記事・文語記事別に集計して示した表である。口語比率・文語比率は、それぞれ口語記事・文語記事での出現頻度の、総出現頻度に対する比率（％）である。「12語の合計」欄の口語比率（65.6％）よりも口語比率が

表6 高頻度語の文体（口語的な語・文語的な語）

見出し語	口語記事	口語比率	文語記事	文語比率	総出現頻度	文体的特徴
が	850	97.7%	12	1.4%	870	口語的
けれども	757	95.9%	9	1.1%	789	口語的
これにはんして	146	50.5%	142	49.1%	289	文語的
されど	134	11.6%	1009	87.1%	1159	文語的
しかし	4440	93.9%	204	4.3%	4728	口語的
しかしながら	854	85.3%	118	11.8%	1001	口語的
しかるに	1529	53.6%	1265	44.4%	2851	文語的
しかれども	170	6.2%	2575	93.4%	2756	文語的
それでも	353	92.4%	17	4.5%	382	口語的
だが	449	96.1%	15	3.2%	467	口語的
でも	279	90.3%	13	4.2%	309	口語的
ところが	803	98.3%	8	1.0%	817	口語的
12語の合計	10764	65.6%	5387	32.8%	16418	

高い語は口語記事に偏って用いられている語、逆に「12語の合計」欄の文語比率（32.8%）よりも文語比率が高い語は文語記事に偏って用いられている語と見ることができる。それぞれ「口語的な語」「文語的な語」と呼ぶことにする。（注11）

口語的な語は「が、けれども、しかし、しかしながら、それでも、だが、でも、ところが」の8語、文語的な語は「これにはんして、しかるに、しかれども、されど」の4語である。

3.3.2 出現頻度の相対的な経年変化の分析

次に、12語の出現頻度の経年変化を見ることにする。

表7は、年次別に、各語の出現頻度の、逆接の接続語句全体の出現頻度に対する比率（%）（以下「出現率」と呼ぶ）を示した表である。語ごとの出現頻度は表1に、各年の逆接の接続語句全体の出現頻度は表4の「延べ語数」に、それぞれ既に表示している。たとえば、1895年の「が」の出現頻度は10であり、同年の逆接の接続語句全体の出現頻度は3009であるので、0.3%となる。出現頻度を直接見るのではなく相対的な出現率を用いることにより、各年の逆接の接続語句全体の出現頻度の違いを均すことができ、より正確に経年変化の様子を分析できる。

この「出現率」に基づいて、出現頻度の相対的な経年変化のパターン（増加、減少、増減なし）を、次のような基準によって分類してみる。

α 1895～1925年で、出現率が増加している語

α 1 極端に（10%程度以上）増加している語

表7 高頻度語の年次別の出現率

文体	見出し語	1895年	1901年	1909年	1917年	1925年
口語的な語	が	0.3%	0.7%	3.3%	6.7%	10.0%
	けれども	0.9%	1.0%	6.4%	4.0%	6.9%
	しかし	6.4%	13.6%	28.0%	37.3%	35.5%
	しかしながら	2.2%	2.8%	6.9%	8.4%	5.8%
	それでも	0.7%	1.0%	1.7%	3.5%	3.0%
	だが	0.6%	1.1%	1.1%	1.1%	7.1%
	でも	0.1%	0.7%	1.6%	1.4%	3.6%
	ところが	0.7%	3.1%	4.0%	4.4%	8.2%
文語的な語	これにはんして	1.4%	2.7%	1.9%	1.8%	0.6%
	されど	10.6%	15.7%	8.1%	1.1%	0.5%
	しかるに	19.5%	19.1%	16.8%	17.5%	8.7%
	しかれども	43.6%	28.2%	11.6%	3.9%	0.6%

表8 高頻度語の文体及び出現頻度の相対的変化パターン

	α 出現率の増加		β 出現率の減少	γ 増減なし
	$\alpha 1$ 極端に増加	$\alpha 2$ やや増加		
口語的な語	が、しかし	けれども、しかしながら、だが、ところが		それでも、でも
文語的な語			されど、しかるに、しかれども	これにはんして

$\alpha 2$ やや（5～9%程度）増加している語

β 1895～1925年で、極端に（10%程度以上）減少している語

γ 1895～1925年を通して、殆ど出現率の変化がない語（ $\alpha \cdot \beta$ 以外）

この基準に基づいて、12語の変化パターンを整理して示すと表8ようになる。

『太陽コーパス』は、1895年では文語記事が圧倒的に多く、次第に口語記事が増えていき、1925年にはほぼ口語記事となっている。したがって、逆接の接続語句についても、口語的な語の出現率が増加し、文語的な語の出現率が減少していくことは当然のことである。「 γ 増減なし」についても、出現率の大幅な増加や減少は見られないが、仔細に見れば口語的な語の「それでも、でも」は増加傾向にあり、文語的な語の「これにはんして」は減少傾向にある。

出現率の増減が特に目立つ語を、個別に見てみる。

「しかし」は、1895年にも出現率6.4%、出現順位4位であ

り、文語的な語に次いである程度は使用されていたが、文語的な語の使用の減少に伴い、1909年には出現率28.0%、出現順位1位となり、その後も出現率が伸びている。「しかし」は現代語のもっとも代表的な逆接の接続詞であるが、その伸長の過程を見ることができる。

「が」は1895年には出現率0.3%に過ぎないが、1925年には出現率10.0%、出現順位2位となっており、急激に増加しており、「しかし」に次ぐ代表的な逆接の接続詞となっている。

文語的な語の「されど、しかるに、しかれども」のうち、「されど、しかれども」の2語は極端に出現率が減少し、1925年の出現率はそれぞれ0.5%（出現頻度24）、0.6%（出現頻度26）となり相対的に殆ど使われなくなっている。これに対して、「しかるに」は1925年の出現率は8.7%であり、出現頻度410のうち口語記事に390回使われている。これは「しかし、が」に次ぐ出現率である。文語的な語といっても、「されど、しかれども」は文語記事の減少とともに使われなくなっていくのに対し、「しかるに」は口語記事の中で使われ続けていく様が見て取れる。なお「されど、しかるに、しかれども」の3語のうち「しかるに」だけが『口語法別記』に載せられている。

3.3.3 口語的な語の特徴——特に「しかしながら」について

前項で「しかしながら」を口語的な語としたが、他の口語的な語と全く同様に扱うことにはやや問題もある（注12）。

均しく「口語的」とであるとは言っても、日常会話で用いられる口頭語的なものから、日常の会話には余り用いられず講演・演説や文章だけに用いられるものまで、様々である。

この点を考慮し、口語的な語とした8語を対象とし、口語記事に限定して出現する欄の偏りを見ることによって、それぞれの語の特徴を、さらに明らかにしたい。なお、1917年及び1925年には欄の種類が特定されていないの記事が多いため、この2年分を除外し、1895年・1901年・1907年の3年分に限って見ていく。

これら8語の欄別の出現頻度を全体的に見ると、「小説」「小説雑俎」欄と「論説」「講演」欄に使われる語に偏りがありそうである。この点をより明らかにするために、小説系の「小説・小説雑俎」、演説系の「論説・講演」及びこれら以外の「その他」の大きく3種類に欄をまとめて、それぞれの出現頻度を集計した。表9がその集計表である（注13）。また、図1～3は、年次別・

語別に、3種類の欄の出現頻度の比率を示したものである。

表9及び図1～3を通してみると、「しかし」の各欄の比率は、3年分ともに「8語の合計」とほぼ同様の比率であり、どの欄にも偏ることなく広範囲に使われる語であることが分かる。

また、「しかしながら」は、「小説・小説雑俎」欄には殆ど用い

表9 高頻度・口語的な語の欄別出現頻度（「口語」記事のみ）

年 欄 名	1895年			1901年			1909年		
	小説・小説雑俎	論説・講演	その他	小説・小説雑俎	論説・講演	その他	小説・小説雑俎	論説・講演	その他
が	1	0	6	12	0	9	1	15	116
けれども	5	8	7	13	0	14	1	62	199
しかし	16	31	39	202	37	97	26	267	798
しかしながら	0	19	21	3	30	30	0	123	81
それでも	4	0	3	20	0	6	1	7	56
だが	4	0	0	29	0	1	0	1	42
でも	0	0	0	16	0	1	13	0	44
ところが	3	9	7	12	0	73	1	47	117
8語の合計	33	67	83	307	67	231	43	522	1453

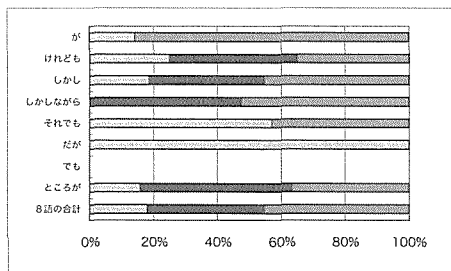


図1 1895年の語別・出現欄の比率

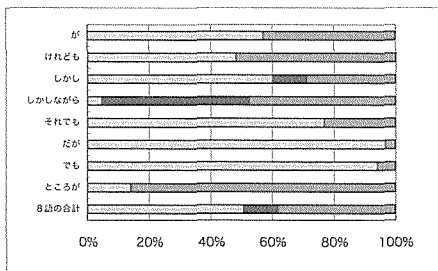


図2 1901年の語別・出現欄の比率

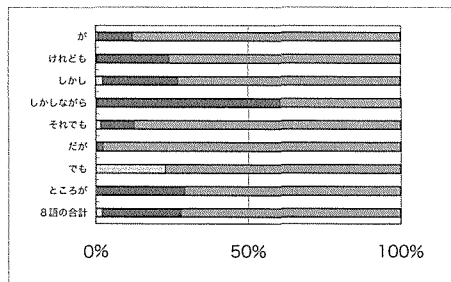


図3 1909年の語別・出現欄の比率

られず、「論説・講演」欄の比率が大きくなっており、演説口調という特徴を持っていることが示されている。一方、特に「だが、でも」は「論説・講演」欄には殆ど用いられておらず、「小説・小説雑俎」欄の比率が他の語に比べて大きくなっており、口頭語的性格を持っていることが窺われる。

4 おわりに

本稿では、まず、『太陽コーパス』の逆接の接続語句の一覧を作成し、今後の研究のための基礎資料として示した。ただし、検索語を予め指定していること、主に文頭・句頭のみを対象とした検索であることなどにより、網羅的な調査ではない。しかし、このような方法であっても、異なり語数124語の多様な逆接の接続語句を示すことができ、『太陽コーパス』に含まれる逆接の接続語句の殆どは抽出できたと思われる。なお、『太陽コーパス』で用例の得られた接続語句には、「青木一覧」及び『日国2版』に載せられていない語も含まれ、また、より早い年の用例を見つけることもできた。このような点から、『太陽コーパス』の有用性の一端を窺うことができる。

次に、本稿では、『太陽コーパス』が対象としている1895～1925年の時期の、多用される逆接の接続語句の変化（特に文語的な語から口語的な語への変化の過程、及び口語的な語の特徴）を、出現頻度の経年変化に基づいて明らかにした。多彩な記事を含む、一定間隔の言語資料である『太陽コーパス』の特長から見て、一般性の高い結論が得られたと思われる。主な点を列挙しておく。

- ・逆接の接続語句が出現する比率は、年次を追うに従い増えていっており、近代語の特徴の一端を示している。
- ・5年分全体の異なり語数は124語と多いが、各年ともに出現頻度の高い上位数語が延べ語数の大半を占めており、頻用されるのはごく一部の語に限られる。
- ・口語的な語である「が、しかし」「けれども、しかしながら、だが、ところが」は、口語記事の増加に伴って、年次を追うごとに出現頻度が高くなっていく。特に「しかし」、次いで「が」の増加が著しく、現代語の代表的な逆接の接続詞とし

て伸長する過程が見られる。

- ・文語的な語である「されど、しかれども」は、文語記事の減少に伴って年次を追うごとに出現頻度が低くなっていき使用されなくなっていく。これに対し、同じく文語的な語とした「しかるに」は、口語記事の中でも出現頻度が高く、口語記事の中で使用され続ける。
- ・口語的な語のうち、「しかし」はどの欄にも偏りなく広範囲に用いられている。「しかしながら」は「論説・講演」欄での出現率が高く、演説口調という特徴を持っていることを示している。「だが、でも」は「小説・小説雑俎」欄での出現率が高く、口頭語的性格を持っていることが窺われる。

注

- (1) 本稿では、接続詞及び接続語句（接続詞と同様の文連接機能を果たす連語）をまとめて「接続語句」と呼ぶ。接続詞に関しては、品詞としての位置付け、接続詞と認定される語の範囲、接続詞的な機能を果たす連語との異同等、種々の根本的な問題がある。こうした問題は重要であるが、本稿では立ち入らない。本稿では、調査漏れを出来るだけ避けるために、連語も含め調査語形の範囲を広めにとることにした。したがって、個々の接続語句の認定に関しては、若干の疑義が残る場合もありうる。
- (2) 青木（1973）の概要を記しておく。この一覧は、「上代・中古・中世・近世・近代・現代に分ち、それぞれの時代にいかなる語が存したかを知り、更にある時代に新しく現われたものと、前代からひき続き存するものとが直ちに識別し得るように工夫」（青木1973、以下同じ）された一覧である。各時代の代表的文献及び研究書から「接続詞（的語彙）」を抽出して作られている。明治時代は「明治初期の新聞の用語」（国立国語研究所報告）「牡丹灯籠」「福翁自伝」のほか小説8作品、大正時代は小説3作品、昭和時代は小説3作品が対象となっている。なお、接続詞の範囲に関しては、「意見の異なる多くの人々の利用に供する」目的から「能う限り広く採取」することを考え、表題にも「接続詞および接続詞的語彙」（原文傍点）を用いたと述べられている。
- (3) たとえば、次のような研究がある。
京極興一・松井栄一（1973）「接続詞の変遷」（『品詞別日本

文法講座6 接続詞・感動詞』89-136頁, 明治書院)

木坂基 (1987) 「近代文学の接続語」(『日本語学』6巻9号, 65-83頁, 明治書院)

(4) 文頭・句頭に限定せず検索した検索語44語の内訳は次の通りである。「けれども, しかし, しかるに, それでも, でも」(「国定読本」, 『口語法』・『口語法別記』両方に用いられた5語), 「けれど, さはあれ, されど, されども, しかれども, それなのに, それでいて, それなのに, だが, だけど, だって, といつて, ところが」(「国定読本」のみに用いられた13語), 「しかしながら, そうだけれども, そうですが, それに, だけれども, ですが, とはいうものの」(『口語法』・『口語法別記』のみに用いられた7語), 「けども, さりとて, さりとては, さりとて, さりながら, さわれ, しかりといえども, しかりといえども, しかるを, しかれど, それにもかかわらず, だけども, だけれど, ですが, ですが, ですけど, といつて, といつても, にもかかわらず」(語形の類似する19語)。

(5) 「然れど・然ど」の総数は30例であり, そのうち「さ」「され」のルビが付いている用例が11例, 「しか」のルビが付いている用例が2例, ルビが付いていない用例は残りの17例である。一方, 「然れども・然ども・然れ共・然共」の総数は3029例であり, そのうち「さ」「され」のルビが付いている用例が15例, 「しか」「しかれ」のルビが付いている用例が412例, 「け」のルビが付いている用例が1例で, ルビが付いていない用例は残りの2601例である。こうした状況から, ルビのない「然れど・然ど」は「されど」, 「然れども・然ども・然れ共・然共」は「しかれども」と読む蓋然性が高いと判断した。なお, 上記以外の「然」の読みに関しては, 「然は然りながら, 然りとて, 然りながら, 然ればとて, 然ればと言つて」に「さ」, 「然とて」に「さり」, 「然うかつて, 然うかとて」に「そ」の読みを与えた以外は, 「し, しか, しかり」(送り仮名等によって異なる) 及び「しかし」(「ながら」に続く場合) の読みを与えた。また, 「然雖」(1例) に関しては原本で確認したところ, 「しかども」のルビが振られていた。「しかれども」の誤植の可能性もあるが, 一覧ではそのまま見出し語とした。

(6) 「青木一覧」の「さはれ」「さわれ」の二語の時期表示に関

しては、それぞれ「さわれ」「さはれ」に置き換えて分析を行った。

- (7) 102語は次の通りである。なお「そうだけれども、そうですが」の2語は『口語法』『口語法別記』から採った検索語であり、以下の分析から除いている。

あるが、あるを、あれども、いうても、かかりしかども、かかるを、かかれども、かくてあれど、かくはあれど、かくはあれども、かようにあれど、かようにあれども、かようにあれど、こうあるに、こうあれども、こうはいうものの、さいういう、さいういうも、さいえど、さこそいえ、さながら、さはありとも、さはありながら、さはあれども、さはそうろうとも、さはなくとも、さはれ、さようだが、さりけるに、さりけれど、さりというとも、さりといえども、さりとても、さりとも、さりながらも、さるがなかにも、さるから、さるにつけても、さるにても、され、さればてって、されや、しかありとも、しかあるに、しかあるを、しかあれど、しかしながらも、しかすれど、しかりとはいえども、しかりとはもうせども、しかりとも、しかるが、しかるところに、しかるところへ、しかるは、しかるものを、したけれど、したところが、したに、したれども、したを、じゃというて、じゃともうしまして、するが、するに、すれども、そうあるところで、そういったって、そうかとして、そうしても、そうじゃてて、そうだが、そうだけれども、そうではあるけれど、そじゃてて、そじゃもの、そながら、そのくせに、そりゃあそうだけれど、それじゃて、それじゃというて、それじゃともうして、それじゃともうしましても、それだあとも、それだけれど、それだけれども；それだというて、それだといって、それだとして、それだものを、それでもねえが、それにしたところが、それはさることでござれども、それはそうじゃが、それはそうだろうが、ではあろうが、ではござりましょうが、ではござりますが、ところに、もうしながら、とはもうすものの、ほれだって

- (8) 「といったところが、とはいいいながら」「というものの、としたところで」は、接続助詞相当の複合辞としての用例は『太陽コーパス』にあり、接続語句として用いられる可能性は窺える。

- (9) 各語の年次別の出現頻度は表1にも示している。表1にな

い「しかりといえども」の1901年・1909年の出現頻度は、それぞれ1及び2である。

- (10) 「これにはんして」の接続助詞「て」の付かない「これにはんし」という語形も使われている。両語を同一語としてまとめて扱うこともできようが、本稿では「これにはんして」のみを対象とした。両語をまとめて集計しても、数値や順位の些細な変動はあるものの、本稿で示した分析や考察内容は同じである。
- (11) 表6には「韻文」及び「混在」表示の記事の出現頻度を示していないため、口語記事と文語記事の出現頻度の合計は、総出現頻度とは合わない。なお、124語全体の口語比率は65.5%，文語比率は32.9%であり、「12語の合計」とほぼ同じである。
- (12) 田中（1984）は、現代語を対象として、「現代の接続表現」（接続詞・接続助詞）を、文体的特徴により、会話的・話ことば的・一般・書きことば的・文語的の5種類に分類している。「だが、でも」は「会話的」，「ところが」は「話ことば的」，「けれども」は「一般」，「しかし、が」は「書きことば的」にそれぞれ分類しているが、「しかしながら」は「文語的」に分類している。「文語的」には他に「さりとして、されど、しかるに」が例示されている。なお、「それでも」は載せられていない。また、『日国2版』の「しかしながら」の「語誌」には、「(4)近世中期頃までには完全に接続詞化したが、近世初期から、「しかし」という形も用いられるようになり、「しかしながら」が文章に、「しかし」がくだけた会話文に用いられた。(5)現代語の場合、日常の会話では「でも」が優勢で、「しかし」は構えた話し方、「しかしながら」は演説や改まった固苦しい話し方に用いられる。」とある。このように「しかしながら」は他の口語的な語とやや性格が異なるようである。
- (13) 「文芸」欄は、文芸論、小説などが混じっているため、「その他」として扱う。

参考文献

- 青木伶子（1973）「接続詞および接続詞の語彙一覧」（『品詞別日本文法講座6 接続詞・感動詞』210-253頁，明治書院）
- 田中章夫（1984）「接続詞の諸問題—その成立と機能—」（『研究資料日本文法第4巻 修飾句・独立句編』81-123頁，明治書院）

「そして」の用法について

—用例に基づく類型の分類と分析—

——島田 泰子

1

はじめに

現代日本語の確立期である明治・大正期のまとまったデータが収められた『太陽コーパス』には、研究利用上さまざまな可能性が期待できる。一定の量的なまとまりを持つ用例群が得られることで、そこから帰納的に導き出される語義・用法の分類がより適切となる、という利点は、その一つであろう。本稿では、この可能性を試すものとして、以下に述べるようなある固有の問題を含む接続詞「そして」について、その用法をめぐって記述的な整理と分析を試みる。

2

研究対象ならびに問題の所在

指示語「そう」がサ変動詞と接続助詞を伴った「そして」の短呼形である「そして」は、今日接続詞として多用される。「本、鉛筆、そしてノートを持参下さい。」のように名詞と名詞とを並べてつなぐのはその用法のひとつであるが、西谷（1973）はこれについて、「従来の日本語の表現にはなかったもの」で「いわゆる翻訳調」、しかも「誤訳」であって日本語としては「誤り」とであると示した。指示詞とサ変動詞とで先行する動詞文を受け、さらに別の動作を並べる「図書館へ行き、そして読書をした。」のようなものが、接続詞「そして」の本来の用法である、と言うのである。一方、梅林（2000）は、大正年間の小説に見られる例を示しつつ西谷（1973）の誤訳・誤用説に疑義を呈し、近代初頭における「そして」の用法の実態調査の必要性を指摘している。

本稿は、この問題提起を受け、明治・大正期における接続詞「そして」の用法の実態について、『太陽コーパス』のデータを用いた分析と考察とを行うものである。梅林（2002）では、明治

期の翻訳児童文学21編を対象として「そして」の用法を調査しているものの、この期における「そして」の用法を類型的に分類・整理しその消長について分析するに十分なデータは、得られていない。本コーパスのデータによってこれを補い、明治・大正期における「そして」の用法についての記述として、展開や消長についての研究に役立てることへとつなげたい。

3

用法の分類基準についての検討と批判

梅林（2000）（2002）では、「そして」の用法分類に関する長田（1970）・松井（1970）の記述を参考にしつつ、「そして」によってつながれる前件Aと後件Bとの意味関係に着目して、その意味用法を2種5類に分類している。梅林（2002）に示されたその類型ならびに代表例を引用すると、以下の通りである。（用例の出典として、以下の記号をいま私に付す。※＝三省堂『例解国語辞典』第五版、＊＝長田（1970）掲載例、＃＝梅林（2002）での作例）

I. 「そして」の前件と後件との間に順序性がある場合。

(a) 前件Aの結果として順当に後件Bがでてくることを示す。

「Aその結果B」と解釈できるもの。

例；雨がやんだ。そして青空が広がった。※

(b) 前件Aと後件Bの時間的なつながりを示す。「Aその次にB」と解釈できるもの。

例；彼は、玄関の前に立った。そしてベルを押した。※

II. 「そして」の前件と後件との間に順序性がない場合。

(a) 一つのことがらの上にもう一つのことらが重なることを示す。「Aそれに加えてB」と累加的に解釈できるもの。

例；くじらは、魚にそっくりな形をしています。けれどものなかまである証拠をたくさん持っています。そして、初めからあのような形ではなかったのです。＊

(b) 二つのできごとを対等の関係でならべる。「主語1…述語1＋そして＋主語2…述語2」のようにできごとを対等の関係でならべ、「A一方（また）B」と解釈できるもの。

例；フランス人は考えたあとで走りだす。そして、スペイン人は、走ってしまったあとで考える。＊

(c) 複数のものごとを列挙する。

例；京都、大阪、そして神戸を旅してきた。#

こういった分類基準について、ここではまず、検討を試みた上で問題点を指摘し、批判を加えておきたい。

上に引用した分類におけるⅡの例文は、「くじらは、初めからあのような形ではなかったのです。そして、魚にそっくりな形をしています。けものなかまである証拠をたくさん持っています。」のように「そして」の前後を入れ替えても同一の事実を意味する文として成り立ち得る（注1）が、Ⅰの例文ではそれが不可能である。こういった操作で容易に確認されるように、前後件の間の順序性の有無は、述べられてある事態の側の問題として動かし難い事実であり、これに基づいて個々の用例は、必ずⅠかⅡのいずれかに截然と分かれたる。ところが、その下位分類にあたる(a)や(b)などの区分については、その限りではない。Ⅱの下位にある分類の基準として、例えば(b)と(c)とにおいては、「二つ」かそれより多い「複数」か、そして「できごと」か「ものごと」か、という二つの観点が示されているが、両者は、文に述べられてある事態にそれぞれ独立した別個の問題として備わる要素であるから、これでは異なる基準が交錯していることになる。この二つは必ず連動して決まった取り合わせとなるわけではないし、さらに、その取り合わせは、「対等」「列挙」という前後件の意味上の関係性と、常にそれぞれ対応するとも限らない。

分類基準が複数混在すれば、実際の用例を分類項目にあてはめるに際して混乱と困難とが伴われることになる。例えば、(b)の例文について言えば、長田（1970）が出典とした警句「イギリス人は歩きながら考える。フランス人は考えた後で走りだす。そしてスペイン人は、走ってしまった後で考える。」（注2）はもともと三件が並ぶものであるから、この「そして」は上の分類ではむしろ(c)に近い。その意味では「列挙」の用法ともされ得るものであるが、「そして」以下を前二件に対する「累加」として説明することもあながち誤りとは言えまいし、あるいは三件が対比的に「対等」な関係で並んでいると見ることもまた可能である。

用法の分類におけるこういった解釈のゆれや迷い、すなわち分類基準の有効性についての疑わしきは、Ⅰの(a)と(b)の間にも、同様に指摘できる。前件Aと後件Bとの間の因果関係の有無は、判断に迷う微妙な場合もあろう（注3）。つまり、上の分類における(a)や(b)などは、ⅠなりⅡなりの下位分類としてお互い等価に並ぶくくりであるとは言えない側面を持つのであって（注4）、こ

ういった文脈上の意味は、分類に際して必ずしも有効な指標とはなり得ない。

そもそも、用法の分類に際して行われる旧来の項目立て、つまり小林典子(1984)に言う「ラベル」の措置(注5)は、「そして」によって成立する接続関係に付随して看取されるさまざまな側面としての含意を切り出そうとすることであって、文脈的に伴われる意味の諸相を多角的に観察するには有効でも、一つの用例をそのいずれかに認定するための分類項目としては甚だ不適切なものである。したがって、接続における意味関係を手がかりとする分類では、いかにラベル名をあれこれ工夫したところで、その方法じたいを変えない限り、何らかの不都合と困難とを伴うことは免れない。よって本稿では、多くの先行研究に見られるこういった意味による分類のあり方から離れ、次節に示すような独自の観点と基準とを用いて、用例の分類を試みることにする。

4

『太陽コーパス』に基づく用法の類型認定

4.1 データの扱いについて

『太陽コーパス』中の仮名書きおよびルビ付きの「そして」の例は、以下の通りである。表1は、ルビのない「而して」「然して」「然而して」の例(計7909例)を含まない。「そして」か「そうして」か判別できず、あるいは「しかして」「しこうして」などである可能性もあるため、これらは全て本稿の扱いからは除外することとした。

表1 太陽コーパス中の「そして」の確例

そして	ソシテ	ソして	而[そ]して	合計
2633	3	1	300	2937

分析の対象を、よみの問題を含まない確例に限定する以上、ここに示した用例数はこの語の実際の使用数とは一致せず、年代別の用例数の増減を手がかりとした用法の推移などは扱えないことになる。よって本稿では、通時的な用法の消長や展開を論じることとは控え、以下の記述と考察は、本コーパスのデータから帰納される用法のヴァリエーションに対する類型的整理と、類型間の本質的な異なり並びに関連性(連続性)についての論理的な分析に

留めることとしたい。

なお、表1に分けて示したような表記の異なりには、語の意味用法に対応して書き分けられるような傾向が特に認められなかったので、以下では表記の異なりを捨象して全て同一に扱う。

4.2 用例の分類と分析

先に3.の終わりで触れた通り、本稿では、「そして」の前後件における文脈上の意味関係による分類を行わない。そもそも、西谷(1973)における誤用との指摘は、「そして」によってつけられる前後件が動詞であるか名詞であるかという文法的な異なりに着目してのものであった。また、こういった品詞性は、連接上の意味関係に比して、より客観的な指標である。さらには、先にその有効性を確認した、前後件の間における時間的な順序性の有無という基準も、以下に述べる通り、実はこの品詞性の違いに強く関連するものである。よってここからは、こういった品詞性の異なりを主たる手がかりとして用例の分類を試み、「そして」の用法の類型を提示することにする。

以下、その類型と代表的な例を示し、それぞれについての分析と考察とを行う。用例の引用には、2003年3月段階の『太陽コーパス』を用いる。

【類型ア】動作性の用言を並べるもの（時間的継起性を有する）

「そして」の前後が動詞文である場合のうち、次の(1)～(3)のようなものを、まずアの類型とする。この類型には、文形式上のヴァリエーションとして、(2)(3)のように主に動詞による用言性の文がひとたび終止した後に用いられる場合と、(1)のように連用形中止法などで並列的に述べられる場合とが見られるが、いずれもここに含めて扱う。

(1) 「《前略》何も妻が出来たからって英氣の挫けるものでもなし、それが爲めに事業の出来ないなんて、そんな人なら無くつても矢張出来ないのだ」と幕無しに饒舌^{しゃべ}つて、そしてハ、、、と高らかに笑った。(1895年9号「朝顔」江見水蔭 P111A05)

(2) 《略》と寛三は力なさうに答へた。そして彼はジツと何か考へ込んだ。(1909年11号「漂泊」相馬御風 P121B19)

「そして」に含まれる「そ(う)」の持つ指示性は、実際の文脈では、専ら前出の記述や発話を直接指す(文脈指示)が、「し」

が動詞に由来することにふさわしく、指示される対象は、動作性の備わったもの（「幕無しに饒舌る」「答へる」）である。先に触れたように、接続詞としての「そして」の最も基本的な用法は、こういった、動詞文による記述なり発話なりを受け、その動作に引き続いて行われる次の動作（「笑う」「考へ込む」）を「そして」以下に述べるものであると考えられる。これは、次の例のように、人間の行為以外の事態を表す用言の文でも同様である。

（３）地球は破裂するだらうか？ 《略》 いつかは一瞬にして地球も無数の破片となつて散亂するかも知れない。そして太陽の周囲を無限に旅しつゞけることだらう。（1925年13号「爆発した惑星」 V136B11）

この場合、述べられる事態の側における事実関係として、「そして」の前後件には時間的な先後関係が存在している。（３）のように現実には未実現の、想像・想定される事態であっても、それは変わらない。このことは、既に述べた通り、前後件の順序を入れ替えて述べることが出来ないという点によって確認される。

アの類型に認められるこういった先後関係を、いま「時間的継起性」と呼ぶことにしよう。このアの類型に属する用例には必ず「時間的継起性」が備わる。

【類型イ】状態性の相言的用言を並べるもの（時間的継起性は希薄）

動詞により表現された事態でも、先のアの場合と異なり、恒常性や継続性を備えていて、むしろ状態を言うに近い場合のものもある。動詞文の終止または中止の後に来る「そして」のうち、こういったタイプのものを、イの類型とする。

（４）降ること八千尺の所で、タルナ社は見えた。そして其蕃社の附近には盛に畑を作つてある。こゝにて又しばし休息す。（1901年13号「台湾中央山脈横断記（承前）」鳥居龍蔵 P139A04）

（５）彼は殆んど祈りに似た單純な願ひを感じた。それほど無意識に亢奮し、そして一方には錯亂してゐた。（1925年14号「乱華（第五回）」山中峯太郎 P092B25）

この類型の例には、（４）「作つてある」、（５）「亢奮し（てゐる）」「錯亂してゐ（る）」のようなテイル形に代表されるアスペクト表現がしばしば共起するが、これは、用例をイの類型として分類するにあたっての重要なマーカーである。

テイル形などで表された状態に近い事態では、「そして」の前

後件が同時に存在することもあり、上のアの類型に見られたような「時間的継起性」は、必ずしも明らかでない。(5)における「一方には」の語は、前後件の同時性を明示するものであろう。(4)における「見えた」は、目標物を視認したという行為を指しながらも、目標物そのものの存在を言うものでもあるから(「タルナ社は見えた」は「タルナ社はあった」にほぼ等しい)、時間的な継続性や同時性という点では、やはり共通していると言える。

動詞文による描写であっても、次の(6)のように属性に関する表現であるものは、やはりこのイの類型に分類されよう。

(6) 是等砲兵士官の謙遜なる談話は、彼をして彼等に尊敬と愛情とを起さしめた、而して又彼自身も、其美しい無邪氣の容貌、と其控 自勝^{ひみえ}の舉動とを以て、充分彼等の同感^{どうかん}を惹きた、(1901年14号「セバストウポルの落城」嵯峨の屋おむろ P100B13)

(6)での前後件は、「彼等」(砲兵士官)と「彼」とが相互^{きあひ}に与え合う影響であるから、どちらかがどちらかに先駆けて実現するというような時間的な先後関係は、両者の間に定め難い。先の(5)における「一方」同様、「又」の語が用いられていることは、それを表すものと捉えられよう。

【類型ウ】相言を並べるもの(叙述的継起性を有する)

上のア・イの類型が動詞文を前後件に持つのに対し、「そして」の前後に、形容詞・形容動詞など相言性の語による文が来るものがある。それら相言性の文を終止形で終止した後、または連用形などで中止した後「そして」が用いられ、さらにその後にも相言性の記述が続くものを、ウの類型とする。

文形式上のヴァリエーションとしてはさまざまなタイプの用例がある。以下の例では、(7)が連用形での中止の後に、(8)が終止形で終止の後に、それぞれ「そして」が続く文型であるが、(9)(10)のように、連体形で修飾語として並べられるものが「そして」によってつなげられた上で、全体として一つの被修飾語(「クラシックス」「人道主義者」)へかかる構造を取るものもある。(8)(10)のように並べられるものが二項であれば、一般には「累加」や「添加」、あるいは「対等」な関係での「並列」などと説明されるところであろうし、(7)のように三項の場合やそれ以上((9)では5項もの連体修飾語が並べ立てられている)であれば、「列挙」に近い意味に取ることが出来る。「並列」

にせよ「列挙」にせよ、(7)(8)が述定の並べ立てであるのに対し(9)(10)は装定の並べ立てであり、ウの類型における「そして」の用いられ方の自在性は特に注意されるものである。

(7) ホンタウに此處^{こゝ}がいゝワ、涼しくつて、蚊が居なくつて、
そしてお祖母^{ばあ}さんも居なくつて、妾^{めかけ}何時^{あたしい}までも居ようかしら、
(1895年7号「子煩悩」大橋乙羽 P105B21)

(8) 陸地のある所のそれらの地震帯は極めて規模が狭い。そして小さい。(1925年2号「余震論」今村明恒 P043C13)

(9) 文學とは遠いかも知れないが、私は漢籍を読むことを勧める。《略》それは文學の方から見ても、極めて簡潔な、豪宕な、勇健な、そしてナイーブな、素樸な、クラシックスを味ふことが出来る。(1909年10号「歴史物、手習ひ及び其他」藤井健次郎 P129A16)

(10) 森戸辰男氏は、情熱の豊かな、そして詩藻の溢れた、人道主義者である。(1925年2号「学界及思想界の花形」鳥角巾 P134A07)

ここでの相言性の語とは、「涼しい」「狭い」「簡潔だ」「ナイーブだ」のような品詞名としての形容詞・形容動詞に限られるものではない。(7)の「蚊が居ない」が「此處」の快適さについて言うものである以上、これにも相言性が認められるし、「お祖母さんが居ない」ことの快適さは前後の文脈に示される祖母による仕打ち(直後に「早く歸らないと、また叱られて、今夜の御飯^{おまんま}を食べさせられない」とある)によって逆説的に表されているから、これもまた相言性の文であると見てよかろう。(10)の後件「詩藻の溢れた」が、意味はもちろん文法的にも連体修飾に働く点でやはり相言性の表現であることは、言うまでもない。

先にイの類型とした例において、「そして」の前後にある状態((5))や属性((6))が同時性を有することを指摘したが、それと同様に、このウの類型に分類される、物や人の様子・性質を言う相言性の文においても、「そして」の前後に時間的な順序性は希薄である。「そして」で結ばれた前後件は、本質的には同時性を備えており、前後を入れ替えて表現しても、事実レベルでは同じである。

同時性を備えた二項またはそれ以上のことがらは、しかしながら、同時に述べられることは出来ない。一度に二つ以上のことは言い表せないという、いわゆる言語の「線状性」によって、述べられるにあたっては、必ず何らかの順序を与えられて一つずつ並べられなければならない。その場合、このウの類型において与え

られる順序とは、先のアのように事態の側に存在する順序に基づくものではなく、話し手・書き手にとっての、その順序で述べなければならないという認識上または修辞上の序列ということになる（注6）。小林千草（1983）では、低学年児童の作文において「書き手の気持ちの上で」例えば「脳裏に浮かんだ順序として」前件の後に後件が位置付けられるような場合に「そして」が用いられる、という傾向を指摘している。（7）の例（発話者について、場面の始めに「年の比は、十歳あまりにもなるらん小娘」との描写がある）についても同様のことが言え、それはごく幼児児童の拙い作文に限られることでもないと考えられる。

述べられる事態の側に存在する先後関係を指して用いた先の「時間的継起性」に対して、事態を述べる側にとっての序列付けによるこの先後関係を、いま「叙述的継起性」と呼ぶこととする。この「叙述的継起性」は先のイの類型における（5）や（6）のような用例においても見られるものであり、このウの類型は、イとあわせて、最初に示したアとは大きく異なるものと捉えられる。

なお、次のような例は、テイルを伴う動詞の文ではあるが、動作や状態というよりは画風の様子を表すものであり、その点で、先のイではなくこのウの類型に分類される用法であろう。

- (11) 片田君の「妓女舞踊の圖」は漫畫的、智解が伴なつてゐる。そして種々の興味がよれ合つてゐる。（1917年13号「文展の洋画（青年作家の作品に対する寸評）」正宗得三郎 PI38C10）

このようなものを介して、イとウは本質的に連続しているとも言える。

【類型エ】体言を並べるもの（叙述的継起性を有する）

分類の最後として、「そして」が、名詞やそれに準ずる名詞的なもの、つまり体言を並べて結びつける場合のものを、エの類型とする。日本語として誤用であるか否か先行文献で問題とされている用法は、このエの類型に分類されるが、『太陽コーパス』からは以下のようなものが得られる（注7）。

- (12) つく〜 その姿を見ると、身には黒のフロックコートに、同じ色のズボン、同じ色の高帽子、そして例の金椀の眼鏡、何れも見覺のある物計りだが……、（1901年12号「夢に乙羽君と語る」巖谷小波 PI03B22）

- (13) 寛三は黙つて、垂頭れた。此の瞬間彼の胸中には、云ふべからざる憤激と、悲哀と、そして不安とが漲つた。

(1909年11号「漂泊」相馬御風 P120B16)

(14) 北海道の秋。玉蜀黍と豆と馬鈴薯と、そして林檎の熟する北國の秋。《略》僕は吃りやすい舌を熟してその秋の景色を語つた。(1909年14号「一室内」真山青果 P103A03)

(15) 二人は床の上でも話し續けた、若い文學好きの人達が話すやうな色々な事を、旅だの、戀だの、そして日本の議院政策、(川島は昨日わざ〜衆議院の傍聴に出掛けたのであつた) 兒戯に類する事だの、支那の騒がしさだのロシアの革命の事までを。(1917年09号「小説みじか夜」中村星湖 P238A06)

「叙述的継起性」が最も顕在化するのは、このエの類型であろう。体言同士をつなげるこの用法では、実際の出来事として話題に上がった順に対話の内容を述べる(15)のような場合を除いては、やはり述べられる順序に「時間的継起性」は認められない。先に3.において見た梅林(2002)におけるII(c)の例「京都、大阪、そして神戸を旅してきた。」にも、都市の名前の述べられる順を入れ替えにくい場合がある((注1)を参照)が、それは、この(15)とともに、「回想」の場面という特定の条件のもとで事実レベルの時間的な先後関係に即して叙述上の序列が決定されるものであって、エの類型の中ではいささか特殊なものである。このような例は量的にもむしろまれであるようで、体言をつなげる用法での「そして」は、同時に存在する別個の物を並べる場合に用いられるのが一般的であり、そういった「そして」の前後件における本質的な同時性、順不同な対等性の面で、エの類型は、先に分析したイ・ウの延長線上に捉えられる。

ウと同様、このエの類型でも、上の例に確認される通り、二つだけでなく三つ以上の物を枚挙的に羅列する文脈での使用も見られるが、その順序は、やはりウの場合と同じく叙述上の序列を反映したものであると言える。例えば(13)は、「此の瞬間」とあることから、現実には同時多発的な三種の感情が、あくまで叙述上の序列においてこのように並べられているのであって、胸中に沸き上がった順に記されているものではないことは明らかであろう。

叙述上の序列としては、(12)は目に入った順、(14)は想起される順、といった具合に、述べられる言表事態の側ではなく、あくまで述べる側の言表主体である話し手なり書き手なりに属する、認識上の問題である。よって、場合によっては、前置きのものの後に主となる重要なものを配したり、逆に、主となるもの

を述べた後についてに補足的な付け足しを行うなど、述べられる項目に主従や軽重などの価値付けが伴われ、その意味では、序列に修辭的な要素も含まれることになる。事実レベルで順不同とは言っても、やはり叙述レベルでは、その順序で述べられることに、それなりの必然性があると言える。

4.3 類型の整理と比較

以上、『太陽コーパス』から得られた「そして」の用例を、この語によってつなげられるものが用言か相言か体言かという品詞性の観点から整理して、ア～エの四種の類型を認定した。二つまたはそれ以上のものごとを、順序立てて一つずつ述べる際に、その順序が何にしたがって決定されるかが、表2のように、類型ごとに異なってくることになる。

表2 それぞれの類型と二つの「継起性」

	「時間的継起性」 (事態に備わる先後関係)	「叙述的継起性」 (話し手・書き手にとつての序列)
ア 動作性の用言	○ (あり)	× (潜在)
イ 状態性の用言	△ (希薄)	△ (一部顕在)
ウ 相言	× (なし)	○ (顕在)
エ 体言	× (なし)	○ (顕在)

事実¹⁾に即した「時間的継起性」は、認識上・修辭上の「叙述的継起性」に優先される(現実においてある順序に従つて生起した事態は、叙述においても、その通りの順序で述べればよいだけの話である)。「時間的継起性」は、アにおいて必ず見られるが、ウ・エには認められない。一方、「叙述的継起性」はア～エの全てに備わるが、アでは、事態の側の「時間的継起性」に一致するために潜在的であつて、ウ・エにおいて顕在化してくると言える。

これら四つの類型における「そして」の用法は、連続性を備えつつ通底している。つまるところ、複数のものごとを順序立てて述べるにあたり、先行するものへ後続のものをただ言い継ぐ(書き継ぐ)のが、「そして」の働きである、と言つてよいのではない。言い換えればこの語は、例えば前件に対する後件^{しつたひ}の出来の順当さ・項目間の対等性などいった、述べられるものごとの関係性を示すために用いられるものではない。この点で、「そして」の語は、逆接や因果関係などといった論理的な関係性を明示する働きを担う接続詞とは性格を異にしている。用例の分類において、

他の接続詞のように意味的な接続関係を基準とすることが出来ないのも、このためであると考えられる。

5

名詞をつなげる「そして」の用法をめぐって

では、このように分類される接続詞「そして」の用法の中で、体言をつなげるものは、他に比してことさらに特異なものであるのか。以下では、エの類型での用法について、さらに詳しくその実態を分析してみたい。

先にエの類型の用例として挙げたものは、(12)のような体言止め相当の場合を除けば、(13)～(15)のように、並べられる体言が同一の格助詞を共有し、共有される格助詞が一度だけ用いられる形式のものであった。体言をつなげる「そして」の用法として真つ先に思い浮かべられるのはこの形式による例文であろうし、実際に、西谷(1973)、梅林(2000)(2002)などが挙げる作例は、もっぱらこの形式を取るものである。その意味でこれらは代表的な例であり、上にもまずはこれらを示したが、現実には用いられる「そして」の用法は、この限りではない。

『太陽コーパス』からの用例を概観すれば、広く体言をつなげる働きにあると認められる「そして」の例文は、おおよそ次の三種に整理出来る。

- ・ 格助詞を共有する形でつなげるもの
- ・ 体言を核とする連用成分をつなげるもの
- ・ 体言を核とする述語成分をつなげるもの

以下それぞれについて考察し、体言をつなげる用法を、接続詞「そして」の働きの中へ最終的に位置付けてみたい。

5.1 格助詞を共有する形で体言をつなげるもの

先の(12)のように、体言が文中で述語に対する格成分としての役割を担わず、体言としての資格のまま並べられるものは、むしろ珍しい。エの類型では、文中で何らかの格に立つべきものを複数並べてくくり、述語動詞に対しまとめて格成分に立てるパターンのもが目立つ。先の(13)や次の(16)はガ格((14)は助詞「の」によるものであるが文法的にはガ格相当)、(17)はヲ格に立つものが、それぞれまとめられて「漲つた」「ござい

ました」「熟する」「異にする」などに対し格成分に立っている。

(16) 大状師『^{テーブル}卓上にあつたのはそれだけか?』

證人『その他^{ほか}ペンとインキと紙と、そして本が一冊ございました。』(1925年9号「ハートの九(第三回)」延原謙(訳) P128B10)

(17) 凡ての蛋白質は、約二十種のアミノ酸が、その量とその種類と、そしてその排列とを異にして組合つて出来てゐるからだといふことがわかつた。(1925年2号「最近栄養学上の進歩」三浦政太郎 P067A08)

(18) 印刷業と爆彈製造、そして胡弓^{こきゅう}、之れ^こが今の郁生晶^{ユーシヤンチャン}の生活なのか?(1925年10号「乱華」山中峯太郎 P040B22)

(18) も、一見体言の言いさしのようで(12)に近いが、「之れ」が、直前に並べ立てられた三項をまとめて一旦受けた上で、がを伴って述語「生活なのだ」に対する主格に立っていると取れば、やはり格助詞を共有する、エの類型での一般的な用法の例として扱うことが出来る。また先の(15)は、「だの」を伴って並べ立てられる点でやや変則的ではあるものの、最終項「ロシヤの革命の事(まで)」の後にあるヲ格でくくられて、述語「話し續けた」の目的格に立つ、と見ることも出来、やはりこれらと同様のパターンとして分析される。

5.2 体言を核とする連用成分をつなげるもの

いま見たようなガ格・ヲ格のものと違って、格助詞がカラ・へ・ニなどの場合には、個々の体言のそれぞれに助詞が伴われがちである。

(19) 《略》浅草は相もかはらず、《略》ごつた返しに煮えたぎつて居る。《略》何しろ雷門に集まるところの電車が一日に千二三百臺、自動車^{自動車}が八百臺、これが南千住の方^方からと、吾妻橋の方^方からと、浅草橋の方^方からと、そして上野の方^方からといふ様に四方八方から、一日に二三十萬の人を運んで来るのだからたまらない。(1925年11号「浅草放浪記」記者 P224A13)

(20) 其處は、一方は居留地^{居留地}へ、も一方は遊廓町^{遊廓町}へ、そしても一方は大商店の軒を並べた大通り^{大通り}へと、云つたやうな地點なので、それに宵の口で人の出盛りなので、瞬く間に彼の周囲には十重^と二十重^はの人垣^{人垣}が出来た。(1917年8号「蠢動」村上勇三 P212A16)

(21) 我日本帝國をよりよく大きく育くむためには、我日本

國民は大なる愛國心に、そして、國際心に目ざめなくてはならぬ。(1925年2号「我國民に訴ふ」新渡戸稲造 P019C17)

(19) は、「南千住の方」「吾妻橋の方」「浅草橋の方」「上野の方」がそれぞれ助詞^カを伴って並べられ、いずれも人の移動の起点としての方向を示す働きで述語「運んで来る」にかかっている。(20) は、助詞^へによって (19) とは逆に行き先としての方向を表示しつつ、係り先の述語動詞(「通じる」など)を省略するものであり、(21) では、「愛國心」「國際心」が二格で「目ざめる」を連用修飾している。

これらは、「そして」によってつながられる体言が、格助詞を伴って文の述語に対する連用成分となるものである。先に 5. 1. で見たものも含めて、体言をつなげる「そして」の用法とは、厳密には体言そのものをつなげる働きにあるのではなく、体言を核とする文の成分をつなげる働きにあると言えよう。その意味では、エの類型の用例は、次のような用法での「そして」との間に、決定的な断絶を持たない。

(22) 《略》 苟くも、有識者流に於ては、そして恒産あり、恒心ある人々に於ては、選舉界の低級云々に愛憎をつかすことなく、自分は自分として、なすべき義務と責任とを果すに努めねばならぬ。(1925年3号「普選実施の國民思想に及ぼす影響」林毅陸 P007B06)

「有識者流、そして、恒産あり、恒心ある人々に於ては」とある場合との連続性を考えれば、(18) までのような典型的かつ代表的な体言をつなげる「そして」の用法が、この (22) のようなものに対しことさら特異なものであるとは言えなくなるのではないか。さらにこの (22) は、条件句の並列に用いられる次の (23) のような「そして」の例へとつながっていく。

(23) 若し果して星辰が無数であるとするならば、そして宇宙の萬物間に果して引力が働くものとするならば、其の無数の星辰は何れも多少とも我が地球に對し相引く力が働く理で、《略》地球の物質には重量といふものがなく、否宇宙間の總ての天體も同様一切重量なく、雲霧の形を以て互に無限の距離に排斥してふ筈である、(1909年1号「学芸小話」P176A11)

先に見たように、いずれの類型においても、文の終止や中止の後に続けて別のことがらを言い継ぐ(書き継ぐ)のが、接続詞「そして」の働きであるとされるが、特に、完結した文の終止に続くのではなく、「そして」が文の成分同士をつなげる場合、品

詞性の異なりを超えて、それらの用法には共通性と連続性とが認められると言える。

5.3 体言を核とする述語成分をつなげる場合

用例には、述定の成分であるものがつなげられる場合も見られる。以下のように、名詞述語文の述部相当のものを「そして」でつなげるものである。

(24) 《略》直ぐ目についたものは、《略》古い姫路革の手文庫^{てぶんこ}だった。それは、前年^{ぜんねん}の春亡^{はるな}つた姉の富子^{とみこ}が、嫁く時に随^{したが}かに持つて行つた筈^{はず}の品で、そしてまた彼等^{かれら}の生母^{なむ}の嫁入道具の一つとして、姉が片身に貰^{もら}つてゐた品^{もの}だった。
(1917年4号「恐ろしき結婚」里見弴P216A10)

(25) けれども彼女^{かのう}が彼^{かれ}の初めて知つた女^{おんな}であり、而して最後まで知つた唯一^{たひひとり}人であつた事は、一種微妙な因縁感^{いんえんかん}を彼女の心^{こころ}に齎^{もたら}した。
(1917年10号「感謝」久米正雄 P274A13)

(24) は「手文庫」にまつわる経緯を説明し、(25) は「彼女」について「彼」との関係性を述べるもので、いずれも、いわば物の素性を述べるにあたって、体言が断定辞「だ」「である」を伴つた述語成分が複数並べられ、「そして」によってつなげられたものである。

この場合、前件は連用形による中止法となるが、素性を述べる文が「そして」によってつなげられているという観点から、これらを、体言をつなげる「そして」の用法とはしないのが通常の判断であろう。「品＋そして＋品」または「女＋そして＋唯一人(の女)」と取り出してみても、「そして」の前後件が、同じ品なり同じ女なりといった具合に同一の存在であるという点で、存在として別個のものを並べる、先に見た代表的かつ典型的なエの類型のものと異なるからである。

通常では体言をつなげる用法と見なさないこれら (24) (25) の例であるが、エの類型の「そして」の用法が、実際の文中においては「体言を核とする文の成分をつなげる働きにある」と考えられたことの延長線上に、これらも位置づけられる。述語成分もまた文における何らかの成分であり、その成分が「そして」によってつなげられたものとして、これらの例を捉えることが出来るからである。

実際、(24) (25) のような例において、体言を修飾する連体部分を一括し、例えば (24) において「そして」の前後件に共

通する「品」をくくりだして「…持って行った筈の、そしてまた…もらってゐた品だった」とすることも出来る。そうすると、これは、相言をつなげるウの種類のうち、(9)(10)のような、連体修飾語を並べる「そして」の用法となる(注8)。文の中止の後に、「そして」によって残りの述語成分が続けられるこれらの用法には、体言をつなげるエの種類との連続性も、相言をつなげるウの種類との連続性も認められることになる。

5.4 体言をつなげる「そして」の位置付け

以上見たように、体言を核とする連用成分をつなげる「そして」は、(22)のようなものを介して条件節を並べる(23)のような例へと連続し、連体修飾に働く成分を並べるウの種類と対をなしつつ「装定の並べ立て」を構成するものと言える。一方、体言を核とする述語成分をつなげるものは、「述定の並べ立て」を構成するものであって、動詞文によって述べられた前件を受けて後件をつなげるアやイの種類での用法と、本質的に通底するものである。こういった他の種類との連続性において捉えるならば、体言をつなげる「そして」の用法は、これだけがとりわけ異質なものであるとは言えないことになろう。

ウの種類における「装定の並べ立て」が、あたかも数学でいう因数分解のように、被修飾語である体言を共通項としてくくりだして修飾語である相言をならべるのと同じように、エの種類での典型的な「そして」の用例は、助詞「が」「を」や述語動詞などといった共通因数がくくりだされることによって、形式上は体言と体言が並べられ「そして」によってつなげられているように見える。しかしその体言は、文中において、連用成分(副詞的)であつたり述語成分(形容詞・形容動詞的)であつたりして、厳密な意味での体言(名詞)の資格に留まるものでは必ずしもない。こういったさまざまな文の成分をつなげるのに接続詞「そして」は広く用いられるのであって、「そして」に構成要素として含まれる「し」は、動詞としての性質をすでに失い、ほぼ形骸化していると見てよい。

これに加えて、例えば「青い目をしたお人形」「真つ赤な顔をして訴える」「彼はいつもぼおつとしている」などのように、装定(連体・連用)・述定を問わず、語としての品詞(名詞・副詞)を問わず、動作性から離れた表現に「する」が広く用いられることを思い合わせれば、体言をつなげる「そして」の用法は、誤用でも誤訳でもなく、日本語として自然なものであると言うことが

出来る。

6 おわりに

本稿では、『太陽コーパス』を利用して明治・大正期における接続詞「そして」の用法のヴァリエーションを広く見渡し、記述的な整理と用法の類型の提示を通じて、その異なりや関係性などについて論理的な観点からも分析と考察を行った。

新たな分類基準への着眼をもたらすという点で、本コーパスの量的な有利さは、それを質的な面に転換することにおいても機能したと言える。本稿が提示した類型によって分類した場合に、用例の年代的な面で実際にどのような消長が見られるかについては、通時的な観点からの扱いによって、また別の研究として成り立つものであろう（注9）。

注

- (1) (c)の例文においては、都市の名前の述べられる順序が、同一の旅行における旅程や、別個の旅行の実現順に一致する場合には、並び替えに多少の困難さが伴われる。ここでは、同じII(c)の類型に分類されるであろう「一度は訪ねてみたい関西の大都市として、京都、大阪、そして神戸が挙げられる。」のような例文で、順序を変えて「神戸、大阪、そして京都が…」とあっても事実レベルでの意味が変わらないことを参照されたい。
- (2) 笠信太郎『ものの見方について—西欧に何を学ぶか—』河出書房 市民文庫（1951.3）による。
- (3) 例えば、「彼は動揺した。そして真っ赤になって彼女をにらみつけた。」のような例文の場合。後件Bは前件Aの次に取った行動であるが、Bの行為を、動揺したことの表れ（Aの結果）と取ることも可能であって、この例はIの(a)とも(b)とも言えるものである。
- (4) そもそも、「累加」「対等」などの用語や概念じたい、厳密な意味で相互に排他的に定義されたり、相互の包含関係などを明確に整理されたりした上で用いられるものではない。それゆえ、論者によってその扱いが不統一な部分も多く見られ、例えば、梅林（2002）では「累加的」なもの（II(a)）と並ぶものとして「対等の関係」（II(b)）が項目立てされているのに対し、

田中(1984)では、「対等の接続」の下位分類として「語句の累加」が整理されてある(第六節「対等の接続(並立の接続)」のうちのソシテほかの項)。

- (5) 小林(1989)では、「添加」「累加」とされる「そして」の働きについて、他の接続詞などへの置き換えを手がかりとしてさらに細かく10種の接続のありようを認定している。それらは単一の用例において一義的に認定されるのではなく、〈推移〉〈並列〉〈序列〉〈対比〉〈強調〉〈転換〉〈補強〉〈要約〉〈結果〉〈典型的並列〉と名付けられた「ラベル」は、一つの用例につき複数を同時に付ける扱いとなっている。
- (6) 杖下(1976a)では、「そして」の前後を入れ替えても文全体の伝える情報に変わりのない場合の最たるものとして、「 $a=2$ であり、そして $b=3$ である」のような命題をつなげる論理の接続詞「そして」を示し、それとの対比で『日常の『そして』』における「修辭的・心理的ファクター」を示唆している。『算術や代数の命題はもともと、それを主張する『人間』、主張される『場所』、そしてとくに『時間』といったファクターへの配慮をすべて切り捨て、このかぎり、はじめから超時間的・普遍的にできている』。その意味で「この種の命題を結ぶ『そして』」は非時間的であり、それで差し支えなく、またそうあらねばならないと言う。これを逆に言えば、ウの類型において、どの順番で述べられても構わないはずの二項(以上)のことがらが、「そして」の前後件として順序を与えられる場合、その順序は、述べられる事態の側の問題ではなく、それを述べようとする書き手なり話し手なりの側の問題として捉えられるものであることになる。
- (7) 本コーパスにおけるこの類型での初出例は、1901年のデータから一例だけ得られた(12)であるが、次に古い1909年とその次の1917年のデータ(つまり、梅林(2000)に挙げられた大正期の用例よりも年代を遡るもの)からそれぞれ複数の例が得られ、後に挙げるように1925年のデータにおいても同様であった。また、出典は後出の(16)のような翻訳ものに限られる訳ではなく、あるいはジャンルや文体的な偏りも特に認められず、特徴的な位相性も有しない。これらのことから、少なくともすでに明治期において、このエの類型での「そして」の使用は特殊なものではなかったと考えられる。
- (8) これらの「連体修飾語+体言+断定辞」が相言文的であることは、また、並べられる順序があくまで叙述上のものである

ことから示すことが出来よう。例えば (24) では、事実は述べられてある順序と逆であり、「姉の富子」は「手文庫」を「生母の嫁入道具の一つとして」「片身に貰つた後に、自らが「嫁く時に慥かに持つて行つた」のである。これを遡るかたちで逆の順序に述べられることが可能なのは、ここで「品」にかかる連体修飾部が（動詞文相当でありながら）相言的な性格で機能していて、そのような場合に、叙述上の序列と事実上の先後関係とが別ものであるからに他ならない。

(9) ただしその場合には、類似の用法を揃える非短呼形「そして」を含め、さらに遡って近世期の用法を調査する必要がある。

参考文献

- 梅林博人 (2000) 「近代小説にみる接続詞『そして』—翻訳調といわれる『AそしてB』をめぐって—」(『国文学 解釈と鑑賞』65-7, 36-42頁, 至文堂)
- 梅林博人 (2002) 「明治期の接続詞『そして』について—翻訳児童文学作品を資料として—」(『日本語研究』22, 89-101頁, 東京都立大学国語学研究室)
- 小林千草 (1983) 「児童の文章における接続詞—小学一年生の場合—」(『言語と文芸』94, 119-152頁, 東京教育大学国語国文学会)
- 小林典子 (1989) 「『そして』による接続詞の連接類型」(『筑波大学留学生教育センター日本語教育論集』4, 19-31頁)
- 田中章夫 (1984) 「接続詞の諸問題—その成立と機能」(『研究資料日本文法』4, 81-123頁, 明治書院)
- 杖下隆英 (1976a) 「日本語と論理 1 『そして』」(『UP』42, 1-6頁, 東京大学出版会)
- 杖下隆英 (1976b) 「日本語と論理 2 『そして』の詩と真実」(『UP』43, 16-21頁, 東京大学出版会)
- 長田久男 (1970) 「接続詞小辞典口語編 そして②」(『月刊文法』2-12, 74頁)
- 西谷元夫 (1973) 「表現上の問題点二つ その一『…ならないさきに』 その二 接続詞『そして』の用法」(『解釈』19-5, 70-72頁, 教育出版センター)
- 松井利男 (1970) 「接続詞小辞典口語編 そして①・そして」(『月刊文法』2-12, 73-74頁)

副詞「とても」について

— 陳述副詞から程度副詞への変遷 — 中尾比早子

1

はじめに

現代語の「とても」には次の二つの用法がある。

- (1) これ以上はとても食べられない
- (2) この映画はとても面白い

前者は否定辞「ない」と呼応する陳述副詞の用法で、「どんなにしても」「どのようにしても」などに相当する意味で用いられる。後者は程度副詞の用法で「とても」がかかる語「面白い」を強調する表現で、程度が大きいさまを表し、「非常に」とほぼ同じ意味に用いられる。

先行研究によると、「とても」の程度副詞の用法が登場するのは、大正時代頃であるとされている（注1）。また、変化のきっかけを推察した、松井（2004）、「とても」がかかる語句に着目して変化の過程を記述しようとした、涌井（1988）、三段階の用法変遷がみられる副詞として「とても」を取り上げた播磨（1993）などの研究もある（注2）。しかし、そうした先行研究では、調査されている用例数やその範囲は限定されており、変化の過程が十分に記述されているとは言い難い。『太陽コーパス』によって、広範囲の多くの用例を経年的に調査することを通して、この現象に関して、変化の過程を詳細に記述できるのではないかと考えられる。

2

陳述副詞と程度副詞

『太陽コーパス』（注3）で「とても」を検索すると、合計822例（注4）の検索結果が得られた。このうち副詞の用法ではない例は対象から外す（注5）。その結果、「とても」の副詞としての例は447例となった。

否定辞と呼応する陳述副詞の「とても」は、例えば(1)の例では、「ない」と呼応するほか、「食べられない」にかかる。

(1) これ以上は とても 食べられない

(2) この映画は とても 面白い

(1)(2)は副詞の種類が異なると捉えられているが、この区別を呼応の有無で捉えてみると、呼応する否定辞をもつものが陳述副詞、呼応をもたないものが程度副詞であると考えられる。

さて、この観点から見た「とても」の陳述副詞と程度副詞の、年代別の用例数を、表1にあげた。

表1 陳述副詞・程度副詞の年代別用例数

年代	陳述副詞 〔否定辞と呼応する〕	程度副詞 〔否定辞と呼応しない〕	合計
1895	55(91.7%)	5 (8.3%)	60
1901	75(80.6%)	18(19.4%)	93
1909	66(85.7%)	11(14.3%)	77
1917	69(85.2%)	12(14.8%)	81
1925	102(75.0%)	34(25.0%)	136
計	367	80	447

「とても」の用例数は、1901年で増加し、その後しばらく停滞するが、1925年で再び増加している。また、陳述副詞と程度副詞の比率に着眼すると、全年代を通して陳述副詞が優勢であるが、次のような推移が指摘できる。程度副詞の比率は、1895年当初は1割に満たなかったが、1901年で増加し、その後しばらく停滞するが、1925年で再び増加する。「とても」が増加する様子と、程度副詞の比率が増大していく様子とが、連動しているように認められる。

このように、「とても」が増加していく過程と、そのなかでも特に程度副詞が増大していく過程を、具体的に数字で示すことができる。以下では、程度副詞が増大していく過程を、「とても」の機能の面から記述していきたい。

3

陳述副詞「とても」

3.1 呼応する否定辞

次の表2は呼応する否定辞と年代別用例数である(注6)。推

量意志系の否定辞、通常の否定辞の順で、それぞれ合計用例数の多いものから示してある。全般に、「まい」「まじ」「じ」などの推量意志系の否定辞は少なく（注7）、「ない」「なし」「ぬ」「ず」「ん」などの通常の否定辞が大半である。年代別の推移を見ると、口語形式の「ない」は徐々に用例数が増加し、文語形式の「ず」は徐々に減少するが、これは、『太陽コーパス』に含まれる文章が、文語文主体から口語文主体に推移していくことによるものと解される。それ以外の点では目立った推移は認められない。

表2 呼応する否定辞 年代別用例数

年代	まい	まじ	ない	なし	ぬ	ず	ん	省略	合計
1895	2	1	9	1	17	19	5	1	55
1901	9	1	33	3	13	9	6	1	75
1909	8	2	22	1	23	6	4	0	66
1917	5	0	41	0	19	1	3	0	69
1925	7	1	67	1	10	1	15	0	102
合計	31	5	172	6	82	36	33	2	367

3.2 陳述副詞「とても」のかかり先

「とても」の機能の分析にあたっては、そのかかり先の語句に着眼したい。まず陳述副詞の場合について、かかり先の語句のうち否定辞に上接する語の品詞によって、次の2種類に分けられる。

(1) 動詞句

①「育たない」など、動詞＋否定辞の例

②「行かない」など、動詞＋助動詞＋否定辞の例

(2) 名詞句

③「できるものではない」など、形式名詞＋（助詞・助動詞）＋否定辞

④「余裕はない」など、名詞＋（助詞・助動詞）＋否定辞

否定辞に接続するパターンのうち、(1)動詞句は動詞の種類により、可能表現、自動詞、他動詞に分類することができる。(2)③のタイプは「～ことはない」「～わけではない」など形式名詞で受けるものであり、名詞句に上接する語により再度分類した(例：「できるわけではない」の場合は可能表現)。このとき動詞であれば(1)の動詞の種類に計上した。④は名詞に計上している。以上の類別の結果が表3である（注8）。

表3 陳述副詞「とても」のかかる語の表現の年代別用例数

年代	可能表現	自動詞	他動詞	名詞	形容動詞	合計
1895	20(37.0%)	26(48.1%)	7(13.0%)	1(1.9%)	0	54
1901	34(45.9%)	25(33.8%)	8(10.8%)	7(9.5%)	0	74
1909	33(50.0%)	22(33.3%)	7(10.6%)	4(6.1%)	0	66
1917	35(50.7%)	26(37.7%)	2 (2.9%)	6(8.7%)	0	69
1925	41(40.2%)	34(33.3%)	14(13.7%)	12(11.8%)	1(1.0%)	102
合計	163(47.7%)	133(36.4%)	38(10.4%)	30(8.2%)	1(0.3%)	365

*このほか「省略」が1895年、1901年に各1例あり

表3から、陳述副詞「とても」がかかる語には、可能表現が最も多いこと、次いで自動詞が多いことが明らかである。そして、この二つで、全体の8割をも占める。この数字は、かかり先が可能表現や自動詞に大きく偏ることを示すものであり、陳述副詞「とても」は、否定辞と呼応するだけでなく、かかり先にも際だった特徴を示していると解するべきであろう。陳述副詞「とても」は、可能表現＋否定辞、自動詞＋否定辞にかかるという、目立った特徴を有していると認められる。

また、年代による推移を見ると、大きな変動はないが、1925年で、他動詞や名詞の比率がやや大きくなっていることは注意される。先に述べた特徴が、1925年に至って、やや薄くなっているのではないかと認められる。

なお、可能表現の内訳は表4のとおりである。「できる」が最も多く、次いで助動詞「れる・られる」を用いる割合が高い。

表4 可能表現の内訳

年代	できる	可能動詞	一得る	一能ふ	れる・られる	合計
1895	4	5	2	3	6	20
1901	20	2	4	2	4	32
1909	19	2	4	1	4	30
1917	17	2	0	0	12	31
1925	18	6	4	0	8	36
計	78	17	14	6	34	149

4

程度副詞「とても」

4.1 程度副詞「とても」のかかり先

程度副詞「とても」について、かかる語について調査した結果が表5である。品詞ごとに年代別用例数を示してある。品詞ごとの合計用例数でいえば形容詞が最も多く、次いで形容動詞が多い。形容動詞は年を追うごとに用例数が増加している。1925年になって、動詞が用例を増やし、名詞、副詞にもかかるようになる。

表5 程度副詞「とても」のかかる語の年代別用例数

年代	動詞	形容詞	形容動詞	名詞	副詞	合計
1895	0	4(80.0%)	1(20.0%)	0	0	5
1901	1(5.6%)	13(72.2%)	3(16.7%)	1(5.6%)	0	18
1909	0	5(45.5%)	6(54.5%)	0	0	11
1917	1(8.3%)	5(41.7%)	6(50.0%)	0	0	12
1925	4(11.8%)	15(44.1%)	13(38.2%)	1(2.9%)	1(2.9%)	34
計	6(7.5%)	42(52.5%)	29(36.3%)	2(2.5%)	1(1.3%)	80

さて、程度副詞としての「とても」は、例えば次のように現れる（下線部）。

従来木造で現はし得た通りの表現は石材ではとても駄目であるし、(1909年11号「日本建築の将来と仏寺の再建」黒田鵬心P134A25)

家賃が馬鹿に高くつて、僕なんぞは家を持ちたくても逆も遣り切れないから據なしの下宿住居だ。(1901年3号「丸之内」内田魯庵P092B24)

最初の用例の「駄目」は「できない」という解釈ができ、実現の難しさを表現する語であるといえる。「駄目」は『太陽コーパス』において17例みられ、程度副詞のかかり先として最も用例が多い語である。

二つ目の用例の「やりきれない」は「やりとげることができない」の意味に解釈でき、こちらも実現の難しさを表す。「一きれない」の用例は11例みられ、「駄目」の次に多い語である。

ここにあげた「駄目」「一きれない」をはじめ、程度副詞のかかり先となる「むずかしい」「困難なり」「不可能」「無理」、「一がたい」「一にくい」はいずれも実現の難しさを表現するもので

ある。これらの語は合計すると49例であり、程度副詞の用例のうち61%を占める。

3.2で見たように、陳述副詞「とても」のかかり先は、可能表現＋否定辞、自動詞＋否定辞が大部分を占めていた。可能表現や自動詞に否定辞が付いて表される意味は、実現の難しさの意味とつながるところがあると考えられる。陳述副詞のかかり先の特徴と、程度副詞のかかり先の特徴とは、通底するところがあるとみてよいのではないだろうか。

4.2 かかる語の意味分類

次に、程度副詞「とても」のかかる語について、否定的な意味をもつグループと肯定的な意味をもつグループとに分類してみた。以下がその表である。

表6 程度副詞「とても」のかかる語の品詞別意味分類

品詞	否定的意味	肯定的意味
動詞	余る, 窮する	似る, 澄ます, 費用がかかる, たまらなくなる
形容詞	むずかしい, 酷い, おぼつかない, しかたない, うだつがあがらない, 一きれない, 一がたい, 一にくい	強い, (知識が) うすい, 素晴らしい, 忙しい, 怖い
形容動詞	駄目, 困難なり, 劣等, 不自由, 不可能, 無理	素敵, 丈夫, 膨大
名詞	沙汰の限り	閉口
副詞	——	うまく

否定的意味のグループには、単語自体に否定要素が含まれるものを選択している。一方、肯定的意味のグループにはそれ以外のものを選択したが、意味は多様化している。例えば形容詞を例にとってみると、属性形容詞、感情形容詞のいずれも出現しており、意味として共通性は見出せない。

動詞の例はそれぞれ「とても驚くほどマリヨールに似てゐますね」(1925年12号「院展の彫刻の印象」畑正吉, デルスニスP188C03)「とても皮肉に澄ましてゐるやうだが」(1925年13号「乱華」山中峯太郎P132B03)「とてもあの聲を聞いたゞけでもう日本の女を思ひ出してたまらなくなるものだからね」(1925年7号「文壇風聞記」水上渉P217A23)「とても費用がかゝる」(1925年3号「時事漫言白眼倒視」水島爾保布P148A21)であ

り、4例中2例がアスペクト表現を伴っており、ほかの例もまた動作性というより状態性をいうものである。

なお、副詞にかかる例が1例ある。「とてもうまく描いてゐる」(1925年12号「二科、漫評・粗見・冷拝」矢部友衛P193A10)はかかり先の「うまく」がさらに「描く」という動詞にかかっており、アスペクト表現「てゐる」が接続しているため、動詞でありながら状態性を表す表現となっている。

表7は意味別の年代推移を調査した結果である。

表7 程度副詞「とても」のかかる語の意味別年代別用例数

年代	否定的意味	肯定的意味	合計
1895	5	0	5
1901	18	0	18
1909	11	0	11
1917	11	1	12
1925	20	14	34
計	65	15	80

否定的意味の用例は当初からあり、次第に増加する傾向にある。一方、肯定的意味の用例がはじめて出現するのは1917年であり、1925年で急激に増加する。

5

陳述副詞から程度副詞へ

5.1 かかり先の語からみた陳述副詞と程度副詞の連続と区分

「とても」がかかる語の品詞を整理してみると、主要なものとして大きく三種類に分類できる。

- (1) 陳述副詞【動詞＋否定辞】
- (2) 程度副詞（否定的意味の語にかかる）【形容詞，形容動詞】
- (3) 程度副詞（肯定的意味の語にかかる）【動詞，形容詞，形容動詞，副詞】

本稿では否定辞との呼応のあるタイプ(1)を陳述副詞とし、呼応のないタイプ(2)(3)を程度副詞として論を進めてきた。

陳述副詞と呼応する否定辞で最も多いものは「ない」であった。「ない」が接続すると呼応部分「動詞＋否定辞」が形容詞化する。これに影響され、「とても」が形容詞にかかることにそれほど抵抗なく、品詞使用の拡大に至ったのであろう。

(2)は否定辞と呼応しないが、かかり先の語の意味合いに否定的意味を含むものである。否定的な意味合いの強いものでありながら、品詞による変化を遂げており、陳述副詞と程度副詞の両方の性質を兼ね備えたタイプであると考えられる。

(3)は(2)の品詞の多様化を受け、かかり先の語に使用される単語の意味合いに制限をかけないものである。現代語においては、程度副詞の基本的な性格は、静的な性質や状態を表す形容詞を限定するものである。仁田（2002）では、純粹程度の副詞の中心用法について「形容詞（相当）が表す属性（質）・状態の有している程度性を指定し限定している」と規定している。『太陽コース』の検索結果において、属性形容詞あるいは性質・状態を表す形容動詞、動詞の場合アスペクト表現を伴うなど状態性を表すものにかかっていることから、(3)の肯定的意味の語にかかるタイプにその機能が認められる。ここで現代と同じ用法の程度副詞が確立したといえよう。

(2)の否定的意味のタイプは形態の面からは変化しているが、意味の面からみると陳述副詞の用法を引き継ぐものである。こうしてみると程度副詞とした(2)の用法は、陳述副詞と程度副詞の中間の用法にあたり、過渡的用法として位置付けられるものであるといえる。(1)(2)は意味の連続、(2)(3)は形態の連続があるものと思われる。

次の図は、この三つの区分を各年ごとに割合で示したものである。グラフ中の数値は用例数である。

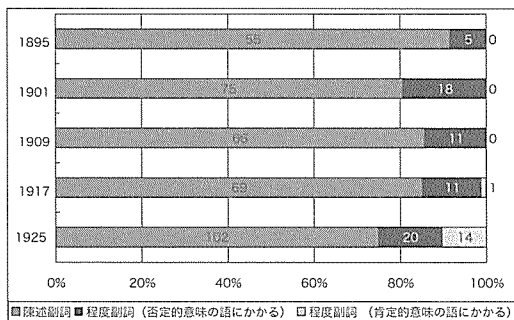


図1 「とても」の用法の推移

このグラフから概観できるように、1895年には程度副詞化がすでに始まっている。グラフの推移からすると、「とても」は次第に程度副詞の用法を拡大していったものと推察される。

程度副詞としての用法が多く用いられるようになった1925年(大正14年)は、ちょうど坪内や芥川が程度強調の用法を指摘したところと時期が一致する。

全体の用例数を増加させながら、程度副詞の用法も増加させつつあることが確認できる。そして1917年に至って程度副詞(肯定的意味の語にかかる)の例が出現した。『太陽コーパス』で最初の用例がみられた1917年(大正6年)は、坪内や芥川の指摘より少し早い時期である。1917年の用例は会話文中での「とても素敵だわ」(1917年5号「第一印象」田村俊子P214A02)であり、完全な程度副詞である。

5.2 程度性の表現

これまで述べてきたとおり、陳述副詞としての「とても」はかかる語に変化をきたし、結果として程度副詞の新たな用法を生み出した。この程度副詞化を別の観点からも捉えてみたいと思う。着目したのは副助詞である。

陳述副詞「とても」には「とてもできない」のように「とても」の直後で否定辞と呼応するものや「とてもあなたの真似はできない」のように、「とても」と呼応する否定辞との間に主語・補語・目的語等の要素が組み込まれる文型もある。「とても」に程度副詞の用法が出てくるとすれば、変遷過程において「とても」と否定辞との間に句として程度性を表現するものが出てきてよいはずである。用例を検討してみると、「ほど」「くらい」等の副助詞を含む表現に特徴のあることが発見できた。

程度性の表現のある文について、どの位置に否定辞がくるのかという観点から二類に分けた。第一類(副詞節中に否定表現が含まれていないタイプ)と第二類(副詞節中に否定表現が含まれているタイプ)である。「とても」には下線を付し、程度性の表現については四角で囲み、強調する対象となるであろう部分(がある場合)には波線を引いている。

第一類 (副詞節中に否定表現が含まれていないタイプ)

①是が弊害を云へばとても一場の話位では盡きない(1901年12号「風俗改良問題(其二)」井上哲次郎P076A09)

②外文士は(佛國側より見れば)謂はゞ烏合の衆に外ならざれば、迥も佛國文壇を壓する程の勢力を有するものでは

ない、(1909年6号「仏国近代の文芸界(ランソン著)」
前田越嶺P137B12)

③今日ではとてもそれ位では相談に乗らぬといふ決心で居る。(1917年9号「独逸政変の側面観」某將軍P089B08)

④もう逆もあれ程精しいものは出来さうもない。(1925年9号「印籠の話」武井守正P206B09)

第二類 (副詞節中に否定表現が含まれているタイプ)

⑤毛彫の緻密精巧なこと、到底も老眼の手藝とは思はれないほどであつた(1895年12号「睡眠の節減」石橋思案P178B09)

⑥その時分の二人の愛情は、とても普通のきやうだいに見られないほど、篤い、美しいもののだつた。(1917年4号「恐ろしき結婚」里見葦P222A05)

⑦工藝上の事などは、これを従來の儘にして置けば、とても十年や十五年で發達することの出来ぬ程度に發達もし…(1917年4号「歐州戦争を中心として」某將軍P116A21)

⑧聞いてみてもとてもフオアラとは思はれないほどだつた。(1925年14号「ハートの九」延原謙P163A15)

⑨すると『フィラメント』からとても數へ切れない程澤山に、『電子』と云つて陰性の電氣を持つた…(1925年4号「無線電話の真空管に就て」佐野昌一P212B16)

⑩我々人間が作り上げ放散する電力とは、とても比較にならない位強力大量なものなのであつて…(1925年4号「ラヂオで天氣を支配し得るか」V295C02)

第一類に該当するものは、「とても」が文末の否定辭に呼応する例である。「ほど」「くらい」に上接する語を並べてみると半数が「それ位」「あれ程」などの指示詞、「前の時ほど」と時を表す語である。ほかには「一場の話位」(①)「一席のお話位」があり、最小であることを示す。ここに含まれる用例のうち、②の波線部分「勢力を有する」のように、強調する対象のある例は、全部で4例ある。「前の時ほど油の乗つたものは出来ず」(1917年1号「失はれた原稿」里見葦P340A06)「あれ程精しいものは出来さうもない」(④)など。しかし、文末に否定辭があるため、「とても」が程度強調の意味合いにはならない。

第二類に該当するものは、程度性の表現の中に否定辭が組み込まれているために、陳述副詞であるのか程度副詞であるのか判定が難しい例である。第一類と異なるのは強調するものがあり、それが否定されない点である。⑦は「程度に」が「發達」の程度を

表現し、⑨は「ほど」が「沢山」の程度を表現し、⑩は「くらい」が「強力大量」の程度を表現しているといえる。このような強調するものが現れている例が10例中7例ある。これらの用例は程度性を表す囲み部分を取り去っても、「とても」が程度副詞として十分に意味が通じるものである。

こうした程度性を表すものは1895年のデータから用例が見つかり、この段階で程度副詞化が起っていることが考えられる。⑤の例では「緻密精巧」が「老眼の手藝とは思はれないほど」のものであったということが書かれており、程度性を表現するものであった。「ほど」節中に否定表現が含まれ、被修飾部分は「とても」より前に表現されている。この例から1895年にすでに程度副詞化への変化の第一歩が見出せるのではないか。

また⑧のように強調するものがない例もほかに2例みられた。

以上の二分類で年代別に用例数をあげると表8のとおりである。程度性を表す語句は多様であるが、明らかなものとして「ほど」「くらい」に絞った。あまり用例数は多くないが、合計数でいうと徐々に用例が増えていることが確認できる。なお、括弧内は強調する語をもつ例である。

表8 「ほど」「くらい」節を含む用例数

年 代	第1類	第2類	合計
1895	1	2(1)	3
1901	2	0	2
1909	2	0	2
1917	2	2(2)	4
1925	2	6(4)	8
計	9	10(7)	19

第一類は用例数に変化はみられない。それに対し第二類は1895年にも用例がみられるが、程度副詞の用法が出現した1917年、1925年にこれらの表現も増加している。1925年では強調する語をもつ例が6例中4例で、大正期に入って増加している。程度性が顕著に現れる例であるといえる。この副助詞からみても1917年、1925年は陳述副詞の用法から程度副詞の用法へと移行する過渡期にあるといえるだろう。

上記のほか、「とても」が程度性を表す決定的な例として次のものがある。「程度はとても～に非ず」という構文で、「とても」

が程度を表す語になっていることが見出せる例である。

「慢り難き勢力なりとする其信仰の程度」はとても強國の臣民を以て自ら誇り、意氣軒昂、何ものも恐ろしきものなしとする青年日本魂と比較し得べきものに非ず。(1909年10号「伊藤公と韓国経営」山路愛山P036A13)

6

おわりに

以上、本稿では副詞「とても」を取りあげ、かかり先の語の分析から用法が新たに成立する過程を述べてきた。その結果、形態、意味の面から段階を追って変化していることが判明した。

「とても」を取り上げるにあたり、『太陽コーパス』はちょうど用法の変化がみられた時期と一致するため、用法の時系列的変化を調べるのに有用な資料である。

「とても」は、30年という短期間のうちに用法の変化する実態が明らかとなった。「とても」に限らず、程度副詞は入れ替わりが激しいため、ほかの副詞も併せ、その実態を検討すべきであると思われる。

注

- (1) 肯定的な表現に用いる「とても」の存在を指摘したのは坪内逍遙(1923)=大正12年、芥川龍之介(1924)=大正13年である。
- (2) 涌井、松井ともに「とても駄目」「とてもむずかしい」等の否定的意味の語を媒介にして程度副詞化がすすんだものと推察しているが、可能性があるという指摘にとどまる記述となっている。播磨は大まかに形式の変化を分類したうえで、意味から程度副詞化の説明を加えているという点で涌井、松井より踏み込んだ考察を行っているが、変化の詳細までは記されていない。
- (3) 本論文で用いた『太陽コーパス』は、2002年10月段階のものである。
- (4) 検索対象とした表記は、「とても」「トテモ」「是連も」「兎ても」「到底も」「連も」で、用例数はそれぞれ、「とても」569例、「トテモ」14例、「是連も」1例、「兎ても」12例、「到底も」16例、「連も」210例の計822例である。
- (5) 陳述副詞でも程度副詞でもない次のような用法は、今回は対象としない。

又如何に寒き時節とても、雨中なりとて、 外套如き物を用うるを許さず、(1901年2号「御店の小僧手代の現状」麴町坊P207A10)

「時間 (今後、従来、今など)」+「とても」、「国名、地名」+「とても」、「人名」+「とても」のような例がここに含まれる。

(6) 呼応する否定辞かどうかを区別するために、以下の点に留意して行つた。工藤 (1999) を参考に行っている。

(1) 否定接頭辞のついた語彙的な派生のものは陳述副詞に呼応する否定辞とは見なさない

例：不可能、不自由、無理

(2) 形式上否定表現であるが、意味上は肯定の派生形容詞は、陳述副詞に呼応する否定辞とは見なさない

例：やりきれない、望みがたい、近寄りがたい、出来がたい、覚束ない

(3) 語彙的に否定的意味をもった形式は、陳述副詞に呼応する否定辞とは見なさない

例：難しい、駄目、困難、酷い

本稿でいう陳述副詞とは、形式上否定であるものを扱い、意味的に否定の場合を含まないこととする。

(7) 吉井健 (1993) の語史には、「(江戸時代は)「とても」がかかる句が、否定を含むものである割合が増加していること(増加の理由は不明)、その否定は、「まい」「じ」で表されることが多く、不可能の意味をもっていることには、現代語との関わりにおいても、注目される」とある。『太陽コーパス』の検索結果では「まい」は数が少なく、「じ」の用例はみられない。

(8) 形容動詞の例は名詞句に上接するものである。用例は「とても容易なことではない」。

参考文献

- 芥川龍之介 (1924) 「澄江堂雜記」(『芥川龍之介全集』4巻、144-145頁、筑摩書房)
- 工藤真由美 (1999) 「否定と呼応する副詞をめぐって」(『大阪大学文学部紀要』39号、69-107頁)
- 坪内逍遙 (1923) 「所謂漢字節減案の分析的批判」(『逍遙選集』11巻、589-655頁、第一書房)
- 仁田義雄 (2002) 『副詞的表現の諸相』(169-180頁、くろしお出版)

-
- 播磨桂子 (1993) 「『とても』『全然』などにみられる副詞の用法
変遷の一類型」(『九大 語文研究』75号, 11-22頁)
- 松井栄一 (2004) 「『とても』と『全然』について (講演 近代
文学と国語辞典)」(『「のっぺら坊」と「てるてる坊主」－現代
日本語の意外な事実－』237-246頁, 小学館)
- 吉井健 (1993) 「国語副詞の史的研究『とても』の語史」(『文林』
27号, 1-30頁, 松蔭女子学院大学国文学研究室)
- 涌井澄子 (1988) 「程度副詞「とても」の研究－陳述副詞から程
度副詞への用法の変化を中心に」(『上越教育大学国語研究』第
2号, 30-34頁)

尊敬待遇表現

——動作性の名詞や動詞連用形に付く形式について——

——近藤 明日子

1

はじめに

雑誌『太陽』刊行時期において、待遇表現変遷史上、顕著な現象の一つとして、動作性の名詞（サ変動詞語幹）や動詞連用形に付いてその動作主体を高める尊敬待遇表現形式が、「御～なさる」（注1）から「御～になる」へ移行したことがあげられる。両形式を比較して論じる先行研究も辻村（1951）・原口（1974）・山田（1959）と複数あり、本稿筆者もまた、小椋・小木曾・近藤（2002）で『太陽コーパス』を資料として両形式の比較を行い、「御～なさる」から「御～になる」への移行についての考察を一部行った。しかし、「御～なさる」「御～になる」だけでなく、この時期の多種の尊敬待遇表現形式をより広く概観し、それぞれの性質を考察する研究は、まだほとんどないといってよいであろう（注2）。

そこで、本稿では、動作性の名詞（サ変動詞語幹）や動詞連用形に付く尊敬待遇表現形式として、「御～なさる」「御～になる」の他に「～なさる」「御～くださる」「～てくださる」「御～あそばさる」「～あそばさる」「御～あそばす」「～あそばす」を加えた計9種の形式を『太陽コーパス』から抽出し、当時の尊敬待遇表現形式を概観することを試みる。考察には、「～」に入る語の種類、文章の種類、話し手の性別の3種の観点を導入し、各形式の性質を考察することとする。

2

調査対象および用例検索方法

調査対象とするのは、『太陽コーパス』1895年・1901年・1909年・1917年・1925年の5年分とし、データは2002年2月時点のものとする。用例検索は、全文検索システム『ひまわり』

を用いた。検索に用いた正規表現を待遇表現形式ごとに記すと表1のようになる。検索範囲はすべて6とした。

これらの正規表現にマッチした文字列から、本稿の調査対象とする尊敬待遇表現形式を本稿筆者の判断により抽出し、以下の考

表1 検索に用いた正規表現

検索対象待遇表現形式	正規表現
御～になる	[おご御][^。！？「」『』]+[あ-ん]*な[ら-れつつツツ] [おご御][^。！？「」『』]+[あ-ん]*成
御～なさる	[おご御][^。！？「」『』]+[な成]さ[ら-れいつつつツ] [おご御][^。！？「」『』]+[な成]す[つつツツ] [おご御][^。！？「」『』]+[な成]せ[いいイイエエエエ] [おご御][^。！？「」『』]+成[ら-れつつツツいいイイエエエエ] [おご御][^。！？「」『』]+被[成為為] [おご御][^。！？「」『』]+[成為為]被 [成為為]御
～なさる	[^。！？「」『』]+[な成]さ[ら-れいつつつツ] [^。！？「」『』]+[な成]す[つつツツ] [^。！？「」『』]+[な成]せ[いいイイエエエエ] [^。！？「」『』]+成[ら-れつつツツいいイイエエエエ] [^。！？「」『』]被[成為為] [^。！？「」『』]+[成為為]被
御～あそばす 御～あそばさる	[おご御][^。！？「」『』]+あそば [おご御][^。！？「」『』]+遊 遊御
～あそばさる ～あそばす	あそば 遊そ*ば[さしすせ] [亜-熙]遊[^あ-んア-ヴ] [^あ-んア-ヴ]遊[亜-熙]
御～くださる	[おご御][^。！？「」『』]+下だ*さ[ら-れつつツツ] [おご御][^。！？「」『』]+下だ*す[つつツツ] [おご御][^。！？「」『』]+下だ*せ[いいイイエエエエ] [おご御][^。！？「」『』]+くだ[さすせつつツツ]
～くださる	[てで]下だ*さ[ら-れつつツツ] [てで]下だ*す[つつツツ] [てで]下だ*せ[いいイイエエエエ] [てで]くだ[さすせつつツツ]

察に利用する（注3）。なお、『太陽コーパス』内に引用されている前時代の資料内に見いだされる用例は、考察の対象から外した。

3

通時的変化の概観

まず、各形式の用例数の通時的変化を概観する。表2は、形式ごとに年別の用例数をあげ、（ ）内に年ごとの構成比率を示したものである（注4）。

表2 年別の用例数と構成比率

	1895	1901	1909	1917	1925	通年
御～になる	75 (14.3%)	41 (11.0%)	70 (10.2%)	86 (20.5%)	237 (30.6%)	509 (18.3%)
御～なさる	191 (36.3%)	157 (42.0%)	272 (39.5%)	108 (25.7%)	173 (22.4%)	901 (32.4%)
～なさる	67 (12.7%)	53 (14.2%)	69 (10.0%)	69 (16.4%)	78 (10.1%)	336 (12.1%)
御～くださる	50 (9.5%)	34 (9.1%)	68 (9.9%)	32 (7.6%)	84 (10.9%)	268 (9.6%)
～てくださる	95 (18.1%)	75 (20.1%)	166 (24.1%)	109 (26.0%)	157 (20.3%)	602 (21.6%)
御～あそばさる	12 (2.3%)	9 (2.4%)	13 (1.9%)	0 (0.0%)	13 (1.7%)	47 (1.7%)
～あそばさる	4 (0.8%)	1 (0.3%)	2 (0.3%)	1 (0.2%)	4 (0.5%)	12 (0.4%)
御～あそばす	31 (5.9%)	4 (1.1%)	27 (3.9%)	14 (3.3%)	24 (3.1%)	100 (3.6%)
～あそばす	1 (0.2%)	0 (0.0%)	2 (0.3%)	1 (0.2%)	4 (0.5%)	8 (0.3%)
合 計	526 (100.0%)	374 (100.0%)	689 (100.0%)	420 (100.0%)	774 (100.0%)	2783 (100.0%)

もっとも顕著な通時的変化は、やはり「御～になる」の増加と「御～なさる」の減少である。構成比率は、「御～になる」は1895年・1901年・1909年はそれぞれ14.3%・11.0%・10.2%と大きな変動はないが、それ以降の1917年・1925年はそれぞれ20.5%・30.6%と急激に増加する。それに対し「御～なさる」は、1895年・1901年・1909年はそれぞれ36.3%・42.0%・39.5%とこれも大きな変動のなかったのが、それ以降の1917年・1925年には25.7%・22.4%と減少する。一方、「御～になる」「御～なさる」以外の形式では、これほど明確な通時的変化は見出されない。つまり、1917年・1925年の「御～になる」の増加と「御～なさる」の減少とは相関関係にあり、それ以外の形式の増減の影響は考えにくい。先行研究で行われてきた「御～に

なる」と「御～なさる」の両形式を比較する考察の妥当性も確保されることになる。

結果、1895年・1901年・1909年では、最も大きな構成比率を占めるのが「御～なさる」で40%前後を占め、以下、「～てくださる」の20%前後、「御～になる」「～なさる」「御～くださる」の10%前後が続く。それが、1917年・1925年の「御～になる」の増加および「御～なさる」の減少により、構成比率も変化し、1925年では、「御～になる」が約30%と「御～なさる」「～てくださる」を抜いて最も大きな構成比率を占め、以下、「御～なさる」「～てくださる」の約20%、「～なさる」「御～くださる」の約10%が続く。1895年・1901年・1909年は構成比率第1位の「御～なさる」と第2位の「～てくださる」とを併せて60%前後、1925年は第1位から第3位までの「御～になる」「御～なさる」「～てくださる」を併せて70%強となり、これらが当時一般的に用いられる形式であったことになる。一方、5年間通じて構成比率が第6位から第9位までの「御～あそばさる」「～あそばさる」「御～あそばす」「～あそばす」の4形式は、構成比率の合計が最も多い1895年で全体の9.1%にしかならない。これら「あそばす」系とでも呼べる4形式は、少数派としてその特異性に注意して考察を進める必要がある。

そこで、以下の観点別の考察では、年別の構成比率の通時的変化が相関関係にある「御～になる」「御～なさる」と、それに続いて構成比率の高い「～なさる」「御～くださる」「～てくださる」と、最後に少数派の「あそばす」系の4形式との3グループに分けて述べてゆくこととする。

4 「～」に入る語の種類から見る各形式の性質

では、それぞれの尊敬待遇表現形式の性質を考察するため、はじめに「～」に入る語の種類を観点として考察する。語種等により「～」に入る語を、①和語動詞連用形、②漢語名詞（漢語サ変動詞語幹）、③漢語サ変動詞連用形、④その他（①以外の和語や混種語）の4種類に分けることとする。以下、参考に語の種類ごとに用例をあげる。

①和語動詞連用形

皆様、これちや耐らん。ちと甲板へお出なさい。涼くツて奈何なに心地が快か知れん。(1895年1号「取舵」尾崎紅葉 P084B20)

いゝ加減にあしらつて居れば可いではありませんか。え、何うかして左様おしなさいな。(1895年2号「書記官」川上眉山 P082A15)

②漢語名詞（漢語サ変動詞語幹）

しかも其大和魂の、外にあらはれたとも申すべき、大和言葉は未だ其價直がきまりませぬ、東洋に於てどころか、我國中でさへきまりませぬ。試に教育界を御覧になればわかります。(1895年1号「国語研究に就て」上田万年 P028B10)
お母さんは御承知なざるまいが世間では誰知らぬ者は無い。(1901年5号「投機」内田魯庵 P105B18)

③漢語サ変動詞連用形

必ずお案じなさらぬやう旦那様からも御詞を、(1895年1号「従軍人夫」饗庭篁村 P092B11)

眞個に國を憂ふる人は、希くは爰に猛省して下さい(1895年2号「家庭に於ける第一義（承前）」三島通良 P141A14)

④その他

『しつかりなさいよ。』肥つた女はドカリと彼女の脊をぶつた。(1925年9号「蛇人（第七回）」三上於菟吉 P149A12)

金は入らぬ………子が欲しい………と云ふ方なれば大切がツて下さるに相違ない………(1895年1年「従軍人夫」饗庭篁村 P091B15)

表3は、「～」に入る語の種類ごとに年別の用例数をあげ、()内に年ごとの構成比率を示したものである。

表3 語の種類別の用例数と構成比率



	1895	1901	1909	1917	1925	通年
①和語動詞連用形	423 (80.4%)	290 (77.5%)	536 (77.8%)	351 (83.6%)	599 (77.4%)	2199 (79.0%)
②漢語名詞	91 (17.3%)	72 (19.3%)	126 (18.3%)	61 (14.5%)	147 (19.0%)	497 (17.9%)
③漢語サ変動詞連用形	9 (1.7%)	11 (2.9%)	20 (2.9%)	6 (1.4%)	24 (3.1%)	70 (2.5%)
④その他	3 (0.6%)	1 (0.3%)	7 (1.0%)	2 (0.5%)	4 (0.5%)	17 (0.6%)
合計	526 (100.0%)	374 (100.0%)	689 (100.0%)	420 (100.0%)	774 (100.0%)	2783 (100.0%)

ここからわかるように、①和語動詞連用形の構成比率が最も高く（通年79.0%）、次に②漢語名詞（通年17.9%）が続き、この2形式で95%以上（通年96.9%）の高い比率を占めており、残る③漢語サ変動詞連用形（通年2.5%）と④その他（通年0.6%）の構成比率は併せてもごくわずかである。そして、この構成比率には明確な通時的变化は見られず、通年で一貫してこの比率が保たれていると見なされる。よって、用例の大半を占める①和語動詞連用形と②漢語名詞を中心に考察することで、各待遇表現形式の性質が見えてくることになるであろう。つまり、①和語動詞連用形の入りやすい形式が和語表現と結びつきやすい性質を持ち、②漢語名詞の入りやすい形式が漢語表現と結びつきやすい性質を持つとすることができるのである。

本稿では、和語表現あるいは漢語表現との結びつきやすさを、特化係数という値によって評価する。特化係数とは、部分集合に対する構成比率を全体集合に対する構成比率で割って求めた値で、例えば、1895年の「御～になる」の特化係数は次の計算式で求められる。

$$\begin{array}{c} \text{1895年「御～になる」の} \\ \text{語の種類別の特化係数} \end{array} = \frac{\begin{array}{c} \text{該当の種類の語の入る1895年「御～になる」の用例数} \\ \text{1895年「御～になる」の用例数} \end{array}}{\begin{array}{c} \text{該当の種類の語の入る1895年の全形式の用例数} \\ \text{1895年の全形式の用例数} \end{array}}$$

つまり、特化係数が1より大きく、かつ1から値が離れるほど、全体の傾向と比較して該当の待遇表現形式が該当の種類の語と結びつきやすいことを意味し、逆に特化係数が1未満でかつ1から値が大きく離れるほど、全体の傾向と比較して該当の待遇表現形式が該当の種類の語と結びつきにくいことを意味する。

この特化係数を①和語動詞連用形と②漢語名詞について年ごとに示し、用例数を（ ）内に示したものが表4である。表4では特化係数が1より大きく、かつ1から大きく値の離れる場合として1.20以上を  のように太線の枠で囲み、1未満で、かつ1から大きく値の離れる場合として0.80以下を  のように欄内に色を付けて示した。

この特化係数により、各形式が和語表現と漢語表現のいずれと結びつきやすいのか評価を行い、そこから導かれる考察を形式ごとに以下に述べてゆく。

表4 語の種類別の特化係数と用例数

	語の種類	1895	1901	1909	1917	1925	通年
御～になる	①和語動詞連用形	0.90 (54)	1.04 (33)	0.97 (53)	1.07 (77)	1.16 (212)	1.07 (429)
	②漢語名詞	1.62 (21)	0.89 (7)	1.17 (15)	0.72 (9)	0.51 (23)	0.83 (75)
御～なさる	①和語動詞連用形	1.05 (161)	1.00 (122)	0.96 (204)	0.90 (81)	0.78 (104)	0.94 (672)
	②漢語名詞	0.82 (27)	1.09 (33)	1.27 (63)	1.66 (26)	2.10 (69)	1.35 (218)
～なさる	①和語動詞連用形	1.06 (57)	1.12 (46)	0.93 (50)	1.01 (58)	0.88 (53)	0.99 (264)
	②漢語名詞	0.69 (8)	0.59 (6)	1.19 (15)	0.90 (9)	1.15 (17)	0.92 (55)
御～くださる	①和語動詞連用形	0.80 (32)	0.61 (16)	0.98 (52)	0.82 (22)	0.97 (63)	0.87 (185)
	②漢語名詞	2.08 (18)	2.60 (17)	1.21 (15)	2.15 (10)	1.19 (19)	1.65 (79)
～てくださる	①和語動詞連用形	1.16 (89)	1.17 (68)	1.18 (152)	1.14 (104)	1.16 (141)	1.16 (554)
	②漢語名詞	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
御～あそばさる	①和語動詞連用形	0.21 (2)	0.14 (1)	0.40 (4)	— (0)	0.60 (6)	0.35 (13)
	②漢語名詞	4.82 (10)	4.62 (8)	3.79 (9)	— (0)	2.84 (7)	4.05 (34)
～あそばさる	①和語動詞連用形	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
	②漢語名詞	5.78 (4)	5.19 (1)	5.47 (2)	6.89 (1)	5.27 (4)	5.60 (12)
御～あそばす	①和語動詞連用形	1.12 (28)	1.29 (4)	1.00 (21)	0.77 (9)	0.97 (18)	1.01 (80)
	②漢語名詞	0.37 (2)	0.00 (0)	1.01 (5)	2.46 (5)	1.32 (6)	1.01 (18)
～あそばす	①和語動詞連用形	0.00 (0)	— (0)	0.00 (0)	0.00 (0)	0.65 (2)	0.32 (2)
	②漢語名詞	5.78 (1)	— (0)	5.47 (2)	6.89 (1)	2.63 (2)	4.20 (6)

4.1 「御～になる」「御～なさる」

本稿筆者は小椋・小木曾・近藤（2002）で、「御～なさる」から「御～になる」への移行は「～」に和語動詞連用形の入る用法で進み、漢語名詞の入る用法での移行はほとんど進んでいないことを指摘した。本稿の特化係数を用いた考察では、「御～になる」の①和語動詞連用形の特化係数は年を経るにつれて値が大きくなり（1895年0.90→1925年1.16）、②漢語名詞の特化係数は年を経るにつれて値が小さくなる（1895年1.62→1925年0.51）のに対し、逆に「御～なさる」の①和語動詞連用形の特化係数は年を経るにつれて小さくなり（1895年1.05→1925年0.78）、②漢語名詞の特化係数は年を経るにつれて値が大きくなる（1895年0.82→1925年2.10）ことに、その傾向が現れている。このような特化係数の明確な通時的変化は他の形式では認められず、

「～」に入る語の種類を観点とする場合、「御～になる」と「御～なさる」の通時的変化には相関関係があると考えられる。1925年に至って「御～になる」は和語表現と結びつきやすくなり、「御～なさる」は際立って漢語表現と結びつきやすくなっていることになるが、この変化の延長線上に、「～」に入る語が②漢語名詞の場合にはほぼ限定される現代語での「御～なさる」の使用状況（注5）があり、『太陽コーパス』ではその現代語に至る過渡期の姿が示されていると言える。

4.2 「～なさる」「御～くださる」「～てくださる」

「～なさる」は、①和語動詞連用形や②漢語名詞に関しては年により特化係数が1より大きかったり1未満であったりと値に揺れがあり、通時的変化や通年での一貫した性質は見出しがたい。なお、表中には示していないが、「～なさる」は④その他の特化係数が1.20以上の特に高い値（通年3.90）で、「～なさる」は他の形式と比較して「～」に入る語に制限の少ない、幅広い表現を取り込むことのできる形式であったと言える。

「御～くださる」は、各年で①和語動詞連用形の特化係数が1未満（通年0.87）で、②漢語名詞の特化係数が1より大きい（1925年以外は1.20以上。通年1.65）。漢語表現と結びつきやすい形式であったと言える。

「～てくださる」は、「～」に入る語が動詞連用形に限られるため、②漢語名詞の特化係数は0となる。漢語表現を「～」に入れるには③漢語サ変動詞連用形を用いるしかない。ゆえに、表中には示していないが③漢語サ変動詞連用形の特化係数が1.20以上の特に高い値（通年2.73）を示す。ただし、これも表中には示していないが、漢語表現である②と③の和の特化係数は通年で0.80以下の特に低い値（通年0.34）で、③漢語サ変動詞連用形で補うことをもってしても漢語表現とは結びつきにくい形式であったと言える。逆に①和語動詞連用形の特化係数は各年で1より大きく（通年1.16）、和語表現と結びつきやすい形式であったと言える。

4.3 「あそばす」系

「あそばす」系の4形式のうち、「御～あそばさる」「～あそばさる」は、通年で②漢語名詞の特化係数が1.20以上の特に高い値（それぞれ通年4.05、5.60）であり、際立って漢語表現と結

びつきやすい形式であったと言える。特に「～あそばさる」は②漢語名詞の入る用法しかない。「～あそばさる」に入る語は、全12例中9例までが「行啓・環啓・還幸・入御・上意・崩御」といった尊敬待遇の働きを持つ「御～」の形をとり得ない（とる必要のない）語である。表2にあるように「～あそばさる」は「御～あそばさる」と比べ『太陽コーパス』全体での用例数は少ない。「～あそばさる」は語法上「御～あそばさる」に入り得ない語を用いた表現をとる際に、「御～あそばさる」の代わりに補助的に用いられた形式であったと考えられる。

一方「御～あそばす」は、年により特化係数が1より大きかったり1未満であったりと値に揺れがあり、通時の変化や通年での一貫した傾向は見出しがたい。「～あそばす」は、各年で②漢語名詞の特化計数が1.20以上の特に高い値（通年4.20）であり、漢語表現と結びつきやすい形式であると言える。ただし、「御～あそばさる」に対する「～あそばさる」のように、「御～」の形をとり得ない語が入るという傾向はなく、「結婚」「興奮」などの通常の漢語名詞が入る。なお、「～あそばす」に用いられる①和語動詞連用形の2例はいずれも同一の記事（1925年2号「歴史情話のめぐみ（第一回）」）中の用例で、「仰せ」という「御～」の形をとり得ない語が入るという極めて限定的なものである。「～あそばす」は表2にあるように、用例数は「御～あそばす」に比べてごく少数である。「～あそばす」は、②漢語名詞の入る場合に「御～あそばす」の代わりにごくまれに用いられる形式であったと考えられる。

5

文章の種類から見る各形式の性質

次に、各待遇表現形式が用いられる文章の種類を観点として各形式の性質を考察する。『太陽コーパス』の資料としての性質として、小説以外に論説・演説速記・談話速記などの多ジャンルの文章が混在していることがあげられる。従来の近代語での待遇表現研究では、待遇表現が主に会話で用いられるため、会話の分量の多い小説を資料とすることが多く、それ以外のジャンルの文章を資料とする研究は未だ少ないのが現状である。『太陽コーパス』を用いることで、当時の会話以外の文章での待遇表現の使用状況

も概観できることが期待される。

本稿では、文章の種類を、①会話（心話を含む）、②消息、③談話速記、④演説速記、⑤地の文（①～④⑥以外）、⑥韻文に分けることとする（注6）。表5はその文章の種類ごとに年別の用例数をあげ、（ ）内に年ごとの構成比率を示したものである。

表5 文章の種類別の用例数と構成比率

	1895	1901	1909	1917	1925	通年
①会話	422 (80.2%)	314 (84.0%)	629 (91.3%)	358 (85.2%)	664 (85.8%)	2387 (85.8%)
②消息	23 (4.4%)	21 (5.6%)	13 (1.9%)	9 (2.1%)	27 (3.5%)	93 (3.3%)
③談話速記	0 (0.0%)	8 (2.1%)	22 (3.2%)	19 (4.5%)	2 (0.3%)	51 (1.8%)
④演説速記	57 (10.8%)	3 (0.8%)	0 (0.0%)	12 (2.9%)	5 (0.6%)	77 (2.8%)
⑤地の文	24 (4.6%)	28 (7.5%)	25 (3.6%)	22 (5.2%)	70 (9.0%)	169 (6.1%)
⑥韻文	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (0.8%)	6 (0.2%)
合 計	526 (100.0%)	374 (100.0%)	689 (100.0%)	420 (100.0%)	774 (100.0%)	2783 (100.0%)

ここからわかるように、もっとも構成比率の高いのは①会話である（通年85.8%）。以下、⑤地の文（通年6.1%）、②消息（通年3.3%）、④演説速記（通年2.8%）、③談話速記（通年1.8%）、⑥韻文（通年0.2%）の順で続く。①会話の構成比率が特に高いのは、待遇表現というものが聞き手や読み手を明確に意識しやすい文章において用いられやすく、特定の少人数の聞き手を目の前に話される会話はその要件を満たす文章であるからである。聞き手や読み手を明確に意識しやすい文章としては、他に、特定の少人数の読み手に対して書かれる②消息や、特定の記者を目の前に話される④談話速記、目の前の聴衆を聞き手とする④演説速記もあげられようが、『太陽コーパス』ではそれらの種類の文章の占める比率が元来少ないので、本稿調査での構成比率もそれに応じて低いものになっていると考えられる。以上の文章の種類別の構成比率に明らかな通時的変化は見出されない。④演説速記が1895年のみ10.8%と他の年に比べ構成比率が高いが、1895年は演説速記を取り上げる記事自体の分量が他の年より多いことを反映するものと考えられる。

さて、①から⑥に分けた文章の種類の内、⑥韻文を除いた①から⑤は、表6に示すように、（Ⅰ）媒体——音声か文字か——，

および（Ⅱ）聞き手や読み手の多寡——特定の少人数か不特定多数か——によって分類することができる。

表6 文章の種類の分類

		(Ⅰ) 媒体	
		音声	文字
(Ⅱ) 聞き手や読み手の多寡	特定の少人数	①会話 ③談話速記	②消息
	不特定多数	④演説速記	⑤地の文

（Ⅰ）の観点からの分類について、『太陽コーパス』は雑誌『太陽』に基づいた文字資料であるから、①から⑤のすべてが真の意味では文字を媒体とする文章でしかあり得ない。しかし、例えば①会話として『太陽コーパス』中に引用される文章は、たとえそれが著者の創作であっても、実際の音声による文章に近い性質を持つと考えられる。また、③談話速記や④演説速記は、文字化の過程で速記者による変改を余儀なくされるとはいえ、やはり実際の談話者や演説者の音声による文章に近似した性質を持つものと考えられる。したがって、①③④は音声を媒体とする文章として扱うこととする。（Ⅱ）の観点からの分類については、③談話速記は特定の記者を聞き手とする文章として、①会話と同様、特定の少人数を聞き手とする文章として扱うこととする。

さて、これら（Ⅰ）（Ⅱ）の観点からの分類によって、①から⑤それぞれの文章で用いられる待遇表現形式のどのような性質が明らかになるのであろうか。（Ⅰ）から見えてくるのは、音声による文章である①③④で用いられやすい形式はより話し言葉的であり、文字による文章である②⑤で用いられやすい形式はより書き言葉的であるということである。さらに（Ⅱ）からは、特定の少人数を聞き手や読み手とする①②③で用いられやすい形式は私的な述べ方であり、不特定多数を聞き手や読み手とする④⑤で用いられやすい形式は公的な述べ方という性質が見えてくると考える（注7）。以下、この（Ⅰ）および（Ⅱ）によって分類された文章の種類での用いられやすさを特化係数で評価し、各待遇表現形式の性質を考察する。なお、⑥韻文での用例については、表5にあるように用例数がほとんどないことと、また表6に示さなかったように、（Ⅰ）（Ⅱ）の観点による分類にはそぐわないことか

ら、以下記述を省略する。

5.1 媒体から見る各形式の性質

まず（Ⅰ）の観点に従って、各形式の性質を話し言葉的か書き言葉的か判断する。音声による文章である①③④と文字による文章である②⑤とに用例を分け、用例数とそこから求められる特化係数を年ごとに示したものが表7である（表の書式は表4に準ずる）。

表7 媒体の手段別の特化係数と用例数

	媒体	1895	1901	1909	1917	1925	通年
御～になる	音 声	1.05 (72)	0.84 (30)	1.01 (67)	0.92 (73)	1.04 (213)	0.99 (455)
	文 字	0.45 (3)	2.05 (11)	0.78 (3)	2.05 (13)	0.81 (24)	1.13 (54)
御～なさる	音 声	1.04 (181)	1.08 (147)	1.05 (271)	1.06 (106)	1.04 (156)	1.06 (861)
	文 字	0.59 (10)	0.49 (10)	0.07 (1)	0.25 (2)	0.78 (17)	0.47 (40)
～なさる	音 声	1.05 (64)	1.13 (52)	1.03 (67)	1.03 (66)	1.08 (73)	1.06 (322)
	文 字	0.50 (3)	0.14 (1)	0.53 (2)	0.59 (3)	0.10 (1)	0.32 (10)
御～くださる	音 声	0.72 (33)	0.58 (17)	0.78 (50)	0.91 (27)	0.87 (63)	0.78 (190)
	文 字	3.81 (17)	3.82 (17)	4.80 (18)	2.12 (5)	1.80 (19)	3.01 (76)
～てくださる	音 声	1.09 (94)	1.14 (74)	1.03 (161)	1.01 (102)	0.98 (133)	1.04 (564)
	文 字	0.12 (1)	0.10 (1)	0.55 (5)	0.87 (7)	1.22 (24)	0.67 (38)
御～あそばさる	音 声	0.46 (5)	0.13 (1)	0.49 (6)	— (0)	0.35 (4)	0.38 (16)
	文 字	6.53 (7)	6.78 (8)	9.76 (7)	— (0)	5.52 (9)	7.01 (31)
～あそばさる	音 声	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.29 (1)	0.09 (1)
	文 字	11.19 (4)	7.63 (1)	18.13 (2)	3.55 (1)	5.98 (3)	9.74 (11)
御～あそばす	音 声	1.06 (30)	1.15 (4)	1.06 (27)	1.08 (14)	1.15 (24)	1.10 (99)
	文 字	0.36 (1)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.11 (1)
～あそばす	音 声	0.00 (0)	— (0)	1.06 (2)	1.08 (1)	1.15 (4)	0.97 (7)
	文 字	11.19 (1)	— (0)	0.00 (0)	0.00 (0)	0.00 (0)	1.33 (1)

4.での考察同様、この特化係数を、各形式が音声による文章と文字による文章とのいずれで用いられやすいのかを評価する指標とし、そこから導かれる話し言葉的あるいは書き言葉的性質を形式ごとに述べてゆく。

5.1.1 「御～になる」「御～なさる」

「御～になる」「御～なさる」2形式の間に4.1で見られたような、通時的変化における相関関係は見出すことはできない。

「御～になる」単独では、特化係数が年により1より大きかったり1未満であったりと値に揺れがあり、通年での一貫した傾向も見出しがたい。

一方「御～なさる」は、各年で音声の特化係数が1より大きく（通年1.06）、文字の特化係数が0.80以下の特に低い値（通年0.47）であることから、話し言葉的な性質を持つ形式と考えられる。

5.1.2 「～なさる」「御～くださる」「～てくださる」

「～なさる」は、各年で音声の特化係数が1より大きく（通年1.06）、文字の特化係数が0.80以下の特に低い値（通年0.32）であることから、話し言葉的な性質を持つ形式と考えられる。

「御～くださる」は、各年で音声の特化係数が1未満（前半3カ年は0.80以下の特に低い値。通年0.78）で、文字の特化係数が1.20以上の特に高い値（通年3.01）であり、際立って書き言葉的な性質を持つ形式と考えられる。加えて「御～くださる」は、文字による文章での全76例中、実に56例が候文で用いられるという性質を持つ。また、本稿調査対象の待遇表現形式の全用例中、候文に出現するものは全75例あるが、その内の7割以上が「御～くださる」で占められることになり、『太陽コーパス』の候文で主要な地位を占める尊敬待遇表現形式が「御～くださる」であったと捉えることもできる。「御～くださる」は、書き言葉的であるに加えて、候文という定型的な文語の文章でよく用いられるという極めて特徴的な性質を持つ形式である。ただし、前半の3年と後半の2年とを比較すると、音声の特化係数が増加し、文字の特化係数が減少しているように見える。これは、候文そのものが『太陽コーパス』でしだいに用いられなくなっていったことと関連している可能性がある。

「～てくださる」は、年を経るごとに音声の特化係数が1より大きい値から1未満に減少し（1895年1.09→1925年0.98）、文字の特化係数は0.80以下の特に低い値から1.20以上の特に高い値に増加する（1895年0.12→1925年1.22）という通時的変化が見出される。1895年当時は話し言葉的な性質のあったものが、1925年に至って書き言葉的な性質を持つようになったと考えられ

る。

5.1.3 「あそばす」系

「あそばす」系の内、「あそばさる」を用いた「御～あそばさる」「～あそばさる」は、各年で音声の特化係数が0.80以下の特に低い値（それぞれ通年0.38, 0.09）で、文字の特化係数が1.20以上の特に高い値（それぞれ通年7.01, 9.74）であり、両形式とも共通して際立って書き言葉的な形式と言える。

一方、「御～あそばす」は各年で音声の特化係数が1より大きく（通年1.10）、文字の特化係数が0.80以下の特に低い値を示す（通年0.11）。「～あそばす」は全体の用例数が8例と少なく、慎重な検討が必要であるが、その内の7例が音声による文章に出現していることから、およそ「御～あそばす」に似た傾向を示していると考えてよからう。「あそばす」を用いた「御～あそばす」「～あそばす」は、「あそばさる」を用いた「御～あそばさる」「～あそばさる」とは逆に、話し言葉的性質を持つ形式であると言える。

5.2 聞き手や読み手の多寡から見る各形式の性質

次に、(Ⅱ)の観点に従って、各形式の性質が私的と公的のいずれかを判断する。特定の少人数を聞き手や読み手とする文章である①②③と不特定多数を聞き手や読み手とする文章である④⑤とに用例を分け、用例数とそこから求められる特化係数を年ごとに示したのが表8である（表の書式は表4に準ずる）。

ここまでの考察同様、この特化係数を、各形式が特定の少人数と不特定多数とのどちらを聞き手や読み手とする文章でより用いられやすいのか評価する指標として用い、そこから導かれる形式の性質をそれぞれ以下に述べていく。

表8 聞き手や読み手の多寡別の特化係数と用例数

	聞き手や 読み手の多寡	1895	1901	1909	1917	1925	通年
御へになる	特定の少人数	0.54 (34)	0.72 (27)	0.99 (67)	0.81 (64)	1.00 (213)	0.87 (405)
	不特定多数	3.55 (41)	4.12 (14)	1.18 (3)	3.16 (22)	1.05 (24)	2.31 (104)
御へなさる	特定の少人数	1.11 (180)	1.06 (152)	1.03 (271)	1.06 (105)	1.01 (156)	1.05 (864)
	不特定多数	0.37 (11)	0.38 (5)	0.10 (1)	0.34 (3)	1.01 (17)	0.46 (37)
～なさる	特定の少人数	1.15 (65)	1.07 (52)	1.01 (67)	1.06 (67)	1.05 (73)	1.06 (324)
	不特定多数	0.19 (2)	0.23 (1)	0.80 (2)	0.36 (2)	0.13 (1)	0.27 (8)
御へくださる	特定の少人数	0.90 (38)	1.03 (32)	0.95 (62)	0.99 (29)	1.01 (76)	0.97 (237)
	不特定多数	1.56 (12)	0.71 (2)	2.43 (6)	1.16 (3)	0.74 (6)	1.22 (29)
～てくださる	特定の少人数	1.17 (94)	1.09 (75)	1.01 (161)	1.06 (106)	1.02 (143)	1.06 (579)
	不特定多数	0.07 (1)	0.00 (0)	0.83 (5)	0.34 (3)	0.92 (14)	0.43 (23)
御へあそばさる	特定の少人数	0.30 (3)	0.12 (1)	0.56 (7)	— (0)	0.26 (3)	0.33 (44)
	不特定多数	4.87 (9)	10.72 (8)	12.72 (6)	— (0)	7.94 (10)	7.94 (33)
～あそばさる	特定の少人数	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.28 (1)	0.09 (1)
	不特定多数	6.49 (4)	12.06 (1)	27.56 (2)	12.35 (1)	7.74 (3)	10.37 (11)
御へあそばす	特定の少人数	1.14 (30)	1.09 (4)	1.04 (27)	1.09 (14)	1.12 (24)	1.09 (99)
	不特定多数	0.21 (1)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.11 (1)
～あそばす	特定の少人数	1.18 (1)	— (0)	1.04 (2)	1.09 (1)	1.12 (4)	1.10 (8)
	不特定多数	0.00 (0)	— (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)

5.2.1 「御へになる」「御へなさる」

2形式の間に4.1で見られたような通時的変化における相関関係は見出されない。

「御へになる」単独では、年によって特化係数の値に揺れがあるものの、特に1895年・1901年に特定の少人数の特化係数が0.80以下の特に低い値（それぞれ0.54, 0.72）で、不特定多数の特化係数は1.20以上の特に高い値（それぞれ3.55, 4.12）であることが注意される。本稿筆者は小椋・小木曾・近藤（2002）で、「御へになる」は演説などの不特定多数を聞き手とする発話で使用が始まり、しだいに日常的な会話での使用が広まったことを指摘したが、特化係数を指標とすることでそれがより鮮明に示された。「御へになる」は当初際立つて公的な性質を持っていたが、しだいにその特徴を失って私的な文章でも用いられるように

なったものと考えられる。

一方「御～なさる」は、各年で特定の少人数の特化係数が1より大きく（通年1.05）、不特定多数の特化係数は1925年を除き0.80以下の特に低い値（通年0.46）であり、1917年までは私的な性質を持っていたと言えるであろう。1925年はそれ以前の4年と比較して、不特定多数の特化係数が増加し1より大きくなる（1.01）が、これは1925年以降にも続く通時的変化を示すものなのか、1925年の個別事情による揺れなのか、不明である。

5.2.2 「～なさる」「御～くださる」「～てくださる」

「～なさる」は、各年で特定の少人数の特化係数が1より大きく（通年1.06）、不特定多数の特化係数は0.80以下の特に低い値（通年0.27）であり、私的な性質を持つ形式であったと考えられる。

「御～くださる」は、年により特化係数が1より大きかったり1未満であったりと値に揺れがあり、通時的変化や通年での一貫した性質は見出し難い。

「～てくださる」は、各年で特定の少人数の特化係数が1より大きく（通年1.06）、不特定多数の特化係数が1未満（内3カ年は0.80以下の特に低い値。通年0.43）であり、私的な性質を持っていたと考えられる。

5.2.3 「あそばす」系

「あそばさる」を用いた「御～あそばさる」「～あそばさる」は、各年で特定の少人数の特化係数が0.80以下の特に低い値（それぞれ通年0.33, 0.09）で、不特定多数の特化係数は1.20以上の特に高い値（それぞれ通年7.94, 10.37）であり、際立って公的な性質を持つ形式であったと言える。

一方、「あそばす」を用いた「御～あそばす」「～あそばす」は、各年で特定の少人数の特化係数が1より大きく（それぞれ通年1.09, 1.10）、不特定多数の特化係数は0.80以下の特に低い値（それぞれ通年0.11, 0.00）であり、私的な性質を持つ形式であったと言える。

5.1.3で見たのと同様に、「あそばす」系であっても、「あそばさる」を用いた「御～あそばさる」「～あそばさる」と、「あそばす」を用いた「御～あそばす」「～あそばす」とでは、正反対の性質が見いだされる点に注意する必要がある。

6

話し手の性別から見る各形式の性質

次に、各尊敬待遇表現形式の用いられる文章の話し手の性別を観点として考察する。男性の話し手に用いられやすい形式は男性語的性質を持ち、女性の話し手に用いられやすい形式は女性語的性質を持つと考えられる。本稿で「あそばす」系と呼ぶ4形式を含めた「あそばす」を用いた尊敬待遇表現が「あそばせ言葉」と通称され、当時の女性語の典型とされているが、『太陽コーパス』での実態も把握しておく必要がある。また、他の形式で男性語的または女性語的性質が見いだされるのかも考察したい。本稿では、話し手の性別による位相がより強く現れるであろう5.で分類した①会話での用例を考察対象とする。話し手の性別別に年ごとの用例数をあげ、() 内に年ごとの構成比率を示したものが表9である。話し手の性別が不明の若干数の用例は表に含めていない。

表9 話し手の性別別の用例数と構成比率

	1895	1901	1909	1917	1925	通年
男	181 (43.7%)	176 (56.6%)	340 (54.5%)	130 (36.8%)	394 (60.2%)	1221 (51.8%)
女	233 (56.3%)	135 (43.4%)	284 (45.5%)	223 (63.2%)	260 (39.8%)	1135 (48.2%)
合 計	414 (100.0%)	311 (100.0%)	624 (100.0%)	353 (100.0%)	654 (100.0%)	2356 (100.0%)

これによれば、性別別の構成比率は男性の多い年もあれば女性の多い年もあり、通時的変化や通年の一貫した傾向は特に見出されないことがわかる。

さて、各形式の会話での話し手の性別別特化係数と用例数を年ごとに示したものが表10である（表の書式は表4に準ずる）。

これまでの考察同様、この特化係数を使って、各形式が男性と女性のいずれの話し手に用いられやすいのか評価し、そこから導かれる男性語的あるいは女性語的性質を形式ごとに判断していく。

表10 話し手の性別別の特化係数と用例数

	話し手の性別	1895	1901	1909	1917	1925	通年
御～になる	男	1.68 (25)	1.34 (19)	1.07 (31)	0.91 (21)	0.78 (95)	0.97 (191)
	女	0.47 (9)	0.55 (6)	0.91 (22)	1.06 (42)	1.34 (108)	1.03 (187)
御～なさる	男	0.73 (53)	0.80 (63)	0.85 (121)	0.81 (30)	1.07 (99)	0.86 (366)
	女	1.21 (113)	1.26 (76)	1.18 (140)	1.11 (71)	0.90 (55)	1.15 (455)
～なさる	男	1.81 (49)	1.26 (37)	1.26 (46)	1.64 (29)	1.27 (55)	1.38 (216)
	女	0.37 (13)	0.66 (15)	0.69 (21)	0.63 (19)	0.59 (17)	0.59 (85)
御～くださる	男	1.52 (18)	1.66 (16)	1.17 (32)	1.30 (12)	1.07 (40)	1.26 (118)
	女	0.59 (9)	0.14 (1)	0.79 (18)	0.82 (13)	0.89 (22)	0.72 (63)
～てくださる	男	0.61 (25)	0.98 (41)	1.24 (109)	0.96 (36)	1.28 (101)	1.07 (312)
	女	1.30 (69)	1.03 (33)	0.71 (52)	1.02 (66)	0.58 (30)	0.92 (250)
御～あそばさる	男	0.76 (1)	— (0)	0.00 (0)	— (0)	0.55 (1)	0.43 (2)
	女	1.18 (2)	— (0)	2.20 (3)	— (0)	1.68 (2)	1.61 (7)
～あそばさる	男	— (0)	— (0)	— (0)	— (0)	1.66 (1)	1.93 (1)
	女	— (0)	— (0)	— (0)	— (0)	0.00 (0)	0.00 (0)
御～あそばす	男	0.82 (10)	0.00 (0)	0.07 (1)	0.42 (2)	0.14 (2)	0.30 (15)
	女	1.14 (18)	2.30 (4)	2.12 (26)	1.34 (11)	2.31 (22)	1.75 (81)
～あそばす	男	— (0)	— (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
	女	— (0)	— (0)	2.20 (2)	1.58 (1)	2.52 (4)	2.08 (7)

6.1 「御～になる」「御～なさる」

「御～になる」は、年ごとに男性の特化係数が小さくなり(1895年1.68→1925年0.78)、女性の特化係数は大きくなる(1895年0.47→1925年1.34)。5.2.1に述べたように、「御～になる」は不特定多数を聞き手や読み手とする文章で用いられ始めた形式と考えられる。その類の文章の担い手は当時主に男性であり、そこで「御～になる」に早くから親しんだ男性が会話でも先駆けて「御～になる」を用いたため、当初は際立って男性語的な性質を持っていたが、「御～になる」の使用が拡大するにつれ男性語的な性質が薄れ、逆に際立って女性語的な性質を獲得していったと考えられる。

また、「御～なさる」にも「御～になる」のように通時的変化が見られるようである。男性の特化係数がしだいに大きく(1895年0.73→1925年1.07)、女性の特化係数はしだいに小さ

くなっている (1895年1.21→1925年0.90)。結果、1895年では際立って女性語的な性質であったものが、1925年には男性語的な性質に変わっていったものと考えられる。

6.2 「～なさる」「御～くださる」「～てくださる」

「～なさる」は各年で男性の特化係数が1.20以上の特に高い値 (通年1.38) で、女性の特化係数が0.80以下の特に低い値 (通年0.59) であり、際立って男性語的な形式であったと考えられる。

「御～くださる」は、各年で男性の特化係数が1より大きく (内3カ年は1.20以上の特に高い値。通年1.26)、女性の特化係数が1未満 (前半3カ年は0.80以下の特に低い値。通年0.72) であり、際立って男性語的な形式であったと考えられる。

「～てくださる」は、年によって特化係数が1より大きかったり1未満であったりと値に揺れが見られ、通時の変化や通年での一貫した傾向を見出しがたい。

6.3 「あそばす」系

「御～あそばさる」「御～あそばす」「～あそばす」の3形式は、ほとんどの年で男性の特化係数が0.80以下の特に低い値 (それぞれ通年0.43, 0.30, 0.00) で、女性の特化係数が1.20以上の特に高い値 (それぞれ通年1.61, 1.75, 2.08) である。会話中の用例が1例しかなく慎重な検討を要する「～あそばさる」を除外すれば、「あそばす」系は総じて際立って女性語的であると言える。加えて、1895年における「御～あそばさる」の男性の用例全1例および「御～あそばす」の男性の用例全10例中8例が同一の作品 (12号「滑稽道中すごろく」) に現れている。このため、1895年の「御～あそばさる」「御～あそばす」の男性の特化係数は他の年より高く、女性の特化係数は低くなっているのであるが、これを特定の記事に起因する特殊例として勘案すれば、「あそばす」系の女性語的な性質はさらに確固たるものとなる。

ただし、会話での「あそばす」系に女性語的な性質が認められるとはいえ、「あそばさる」を用いた「御～あそばさる」「～あそばさる」は表10にあるように会話での用例はごくわずかであることは注意しなければならない。本来「御～あそばさる」「～あそばさる」は5.1.3と5.2.3で見たように、不特定多数を読み手とする文字による発話で多用される形式である。不特定多数を読み手

とする文字による文章とは、つまり文章の種類で言えば⑤地の文の類になるが、『太陽コーパス』において、地の文の書き手はほとんどが男性であり、そこに用いられる「御～あそばさる」「～あそばさる」も当然のことながら男性の使用によるものである。「御～あそばさる」「～あそばさる」は女性語「あそばせ言葉」としてはほとんど用いられなかったことになる。一方「御～あそばす」「～あそばす」は5.1.3と5.2.3で見たように、特定の少人数を聞き手とする音声による文章、つまり文章の種類で言えば①会話で多用される形式である。その会話で女性に多く用いられる「御～あそばす」「～あそばす」こそが、女性語「あそばせ言葉」の役割を主に担っていたと考えられる（注8）。

7

おわりに

以上、9種の尊敬待遇表現形式をとりあげ、『太陽コーパス』での全体の通時的変化を概観した上で、「～」に入る語の種類、文章の種類、話し手の性別の3つの観点に基づき、各形式の性質を考察してきた。そこから見えてきた各形式の性質を改めてまとめると表11ようになる。性質に明らかな通時的変化が見られる「御～になる」「御～なさる」「御～くださる」は、変化前と変化後を分けて示した。また、ある性質を評価する特化係数が1.20以上の特に高い値で、かつその性質と相対する性質を評価する特化係数が0.80以下の特に低い値の場合、際立った特徴を示しているものとし、該当するものを太字で示した。

『太陽コーパス』という広範な種類の文章を含む資料において、9種もの尊敬待遇表現形式について考察することで、従来論じられることのなかった各形式の特徴が種々浮かび上がってきた。その内、注意すべきものについて3点以下にまとめる。

- (1) 9種の待遇表現形式の中で、互いに影響しあいながら量的にも質的にも大きな通時的変化を見せるのが、「御～になる」と「御～なさる」である。それは「御～なさる」から「御～になる」へ移行する様相を具体的に示すものである。質的变化で両形式が相関関係にあるのは、「～」に入る語の種類においてであり、その変化の延長線上に、「～」に入る語が漢語名詞の場合にほぼ限定される現代語での「御～なさ

表11 各形式の特徴

		和語／ 漢語	話し言葉的／ 書き言葉的	公的／ 私的	男性語的／ 女性語的
御～になる	変化前	漢語	(不明)	公的	男性
	変化後	和語		(不明)	女性
御～なさる	変化前	和語	話し言葉	私的	女性
	変化後	漢語			男性
～なさる		(不明)	話し言葉	私的	男性
御～くださる		漢語	書き言葉	(不明)	男性
～てくださる	変化前	和語	話し言葉	私的	(不明)
	変化後		書き言葉		
御～あそばさる		漢語	書き言葉	公的	女性
～あそばさる		漢語	書き言葉	公的	(不明)
御～あそばす		(不明)	話し言葉	私的	女性
～あそばす		漢語	話し言葉	私的	女性

る」の姿が見出される。

- (2) これまで論じられることの少なかった会話部分以外での各形式の用例を『太陽コーパス』を資料として考察することで、書き言葉の性質を持つ形式が見出された。一つは「御～くださる」で、漢語表現とともに用いられやすく、際立って男性語的で、特に候文に多用される特徴を持つ。他には「御～あそばさる」「～あそばさる」があり、際立って漢語表現とともに用いられやすく、かつ極めて公的な性質を持つ形式である。

- (3) 「あそばせ言葉」として同一視されがちな同じ「あそばす」を用いた形式であっても、「あそばさる」を用いた形式と「あそばす」を用いた形式とは異なった性質を持つ。「あそばさる」を用いた「御～あそばさる」「～あそばさる」は、上述のように公的で書き言葉的特徴を持ち、会話で女性語として用いられることはほとんどない。一方、「あそばす」を用いた「御～あそばす」「～あそばす」は逆に私的で話し言葉的であり、会話で女性語的な性質を強く持つ。よってこの「御～あそばす」「～あそばす」こそがまさに「あそばせ言葉」と呼ぶにふさわしい形式であると言える。

最後に、待遇表現を論じる上で忘れてならない観点として、該当の待遇表現形式の用いられる文の表現意図や形式の間に見られる待遇差など、他にもいくつか考えられるが、本稿では取り上げきれなかった。今後の課題としたい。

注

- (1) 以下、各尊敬待遇表現形式の「御」は、「～」に入る語によって「お」や「ご」となる語形を併せて示す。
- (2) 『太陽コーパス』の扱う明治期から昭和初期に重なる時代の資料を広範囲に調査し、多種の待遇表現形式を相互に比較し概観した先行研究として小島（1966）・辻村（1974）があげられる。ただし、いずれも資料は小説に限定され、小島（1966）は待遇差、辻村（1974）は用例数の通時的変化について述べるにとどまる。
- (3) 表1にあげた正規表現からわかるように、「御～なさる」「～なさる」の「なさる」は「て」「た」に続くときに現れる「なすつ」や命令形「なせい」「なせえ」を用例に含み、かつ下一段活用・五段活用両活用の用例を併せて示すものである（「御～くださる」「～てくださる」に用いられる「くださる」も同様）。また、「～てくださる」は、「～」に入る語によって「～でくださる」となる場合を併せて示すものである。
- (4) 本稿で扱う形式同士が複合したものとして、「御～なさってくださる」が19例、「御～あそばしてくくださる」が5例、「御～になつてくださる」「御～なされてくださる」「～なされてくださる」が各1例ずつあるが、表中には含めない。また、本稿では扱わない尊敬待遇表現形式と複合したものとして「御～になられる」が2例、「御～あそばされたまふ」「～あそばしたまふ」「御～れてくださる」が各1例ずつあるが、これも表中には含めない。
- (5) 「～」に漢語名詞を入れる場合、坂本（1992）は筆者の内省から「御～になる」より「御～なさる」がよく使われるとし、菊地（1994）も「御～になる」よりも「御～なさる」のほうが好まれるのではないかとする（p.149）とする。本稿筆者もまた「御～なさる」をより自然な表現と感ずる。その理由について菊地（1994）に、漢語サ変動詞「～する」の尊敬待遇表現として、「する」の尊敬待遇表現「なさる」を用いた形式が好まれるのは「自然なこと」（p.149）とある。

- (6) ③談話速記, ④演説速記は『太陽』の記事中にそれぞれ談話・演説の速記と判断できる記述のあるものに限って分別した。それ以外の①会話, ②消息, ⑤地の文, ⑥韻文は, 本稿筆者の判断により分別した。
- (7) 本稿の考察では, ②会話のなかに心話を含めたが, 心話は聞き手の全くない文章であり, (1)の視点から見れば②会話や③談話筆記よりさらに私的な文章ということになる。
- (8) 高澤(2004)は, 明治期から昭和期の婦人雑誌を資料として「あそばす」を用いた尊敬待遇表現形式の用法を考察し, 「あそばせ言葉」という女性語としての用法(副流の系譜)が見られる一方で, 高位の人物を高く待遇する「あそばす」本来の用法(本流の系譜)も連綿と続いていたとする。高澤(2004)は「あそばさる」と「あそばす」を区別しない点で本稿筆者とは立場を異にするが, 『太陽コーパス』の主に地の文に見られる, 「あそばせ言葉」とは性質を異にする「御〜あそばさる」「〜あそばさる」の用例は, 高澤(2004)の言う「本流の系譜」に通じるところがあると考ええる。ちなみに, 地の文での「御〜あそばさる」全37例中35例, 「〜あそばさる」全11例中10例までが日本の皇室および外国の王室を待遇対象としており, 非常に高い待遇の程度を持つ表現として用いられている。

参考文献

- 小椋秀樹・小木曾智信・近藤明日子(2002)「『太陽コーパス』を使った近代語表現の通時的研究——口語文体・可能表現・待遇表現について——」(『国語学会2002年度春季大会要旨集』177-184頁, 国語学会)
- 菊地康人(1994)『敬語』(角川書店)
- 小島俊夫(1966)「「アナタ」・「オ前サン」・「オ前」——一八〇九〜一九〇〇年における滑稽本・人情本・言文一致小説にあらわれた待遇表現体系の変化——」(『言語と文芸』49, 東京教育大学国語国文学会) ※再録(小島俊夫『後期江戸ことばの敬語体系(新装版)』65-89頁, 笠間書院, 1998)
- 坂本恵(1992)「「お〜になる」と「れる・られる」」(『国語学研究と資料』16, 11-21頁)
- 高澤信子(2004)「「あそばせことば」の待遇性について——明治期から昭和期へ——」(『日本語学会2004年度春季大会予稿

-
- 集』45-52頁，日本語学会)
- 辻村敏樹 (1951) 「「お……になる」考」(『国文学研究』復刊4
早稲田大学国文学会) ※再録 (辻村敏樹『敬語の史的研究』
251-275頁，東京堂出版，1968)
- 辻村敏樹 (1974) 「明治大正時代の敬語概観」(『敬語講座5 明
治大正時代の敬語』7-33頁，明治書院)
- 原口裕 (1974) 「『お——になる』考」続貂」(『国語学』96
23-32頁，武蔵野書院)
- 山田巖 (1959) 「明治初期の文献にあらわれた尊敬表現「お(ご)
——になる」について」(『国立国語研究所論集1 ことばの研
究』201-214頁，秀英出版)

漢語サ変動詞の可能の形

——「～できる」の展開——

———小木曾 智信

1 はじめに

本稿では、漢語サ変動詞の可能を表す形、なかでも「できる」を用いる形が、『太陽コーパス』においてどのように用いられていたのかを見てゆきたい。漢語サ変動詞に「できる」がつく形といってもいくつかの形式があるが、歴史的にみると、それらの形式のうちよく用いられるものが徐々に移り変わってきたことが先行研究から明らかになっている。この変化をコーパスの豊富な用例をもって裏付けるとともに、文体や上接語などの様々な観点から分析し、記述を行うことが目的である。『太陽コーパス』に収録された資料は、現代日本語の確立期のものであり、ここにおける通時的な変化を追うことは、そのまま現代語の成り立ちを探ることになるといえる。したがって、これは今日使われている漢語サ変動詞の可能の形の広がりがいかにして成り立ったのかということを探る試みでもある。

2 漢語サ変動詞の可能の形のバリエーション

漢語サ変動詞の語幹をXとすると、その可能の形の主なものとして『太陽コーパス』では、表1の諸形式があげられる。

このほかにも、文語文であれば助動詞「べし」や「能く～す」「～かねる」などの形があるほか、特定の語については「～なる／ならない」の形で可能を表すものもある。だが、これらは用法の広がりや用例数の上からみて特殊なものと考え、ひとまず考察の対象からは除外することとする。

用例数が多いものを例にとると、その可能の形として表1のすべての例が『太陽コーパス』中に確認できる。続けて示すのは「想像する」の可能形のバリエーションの実例である。

表1 『太陽コーパス』における漢語サ変動詞可能形のバリエーション

助動詞「れる」「らる」によるもの	Xされる (Xせらる)
「能ふ」によるもの	Xすることあたふ Xするあたふ Xしあたふ
「得る」によるもの	Xすることを Xするを得 Xし得る
「出来る」によるもの	Xすることができる Xができる Xできる

[Xせらる]

人々の平生の心づかひ又は品位の高卑までも、想像せらるゝ
やう書かるゝやうになりました、(1895年11号「国楽改良意
見」松本操貞P138B04)

[Xされる]

それが爲に人が墮落するとかいふやうなことは想像されな
い。(1909年1号「文芸取締問題と芸術院」上田万年(談)
P132A16)

[Xすること能ふ]

是等の事實は現代國家に於ては殆んど想像すること能はざる
所なるも、(1909年6号「財政、経済」本多精一P012A19)

[Xする能ふ]

状況如何なりしやを想像する能はずといへども (1895年1号
「元時代の雑劇」幸田露伴P107A23)

[Xし能ふ]

此の如き良好の結果を得べしとは蓋し何人も想像し能はざり
し所ならん (1895年11号「琵琶湖連合競漕会」志田鈿太郎
P182A12)

[Xすることを]

之に依て戸倉に至るを得べき日數も豫め想像することを得、
(1895年1号「利根水源探検紀行」渡辺千吉郎P079B35)

[Xするを得]

必ず數百年の古代に鑑み或は數百年の未來を想像するを得る
ものにて、(1895年12号「人生觀に就て」元良勇次郎
P033B18)

[Xし得る]

予が此の梗概を記せるを讀みても蓋し人の想像し得るところ
ならん、1895年06号「元時代の雜劇 関漢卿（下）」幸田
露伴P105B21)

[Xすることができる]

雑誌と福澤諭吉先生の遺編とを覗いた丈でも十分に想像する
ことが出来ます。(1917年2号「心頭雜草」与謝野晶子
P039B15)

[Xができる]

多数同時に同一所に生じたとは想像が出来ぬ。(1909年2号
『自然界の三大矛盾』に就て「遠藤吉三郎P057B07)

[Xできる]

如何に根強く露西亞に侵入して居たかゞ想像出来るであら
う。(1917年1号「戦時欧米産業界の活動」記者(文責) ;
鶴見左右雄P067A15)

これら諸形式のなかでも、本稿で注目するのは、今日広く用い
られている「出来る」による形である。表1の各形式のうち「能
ふ」によるものと「得る」によるものはサ変動詞に限らず、他の
動詞であっても用いられる可能の表現形式である。それに対して
「Xができる」「Xできる」はサ変動詞専用の形であって、この形
が存在するためにサ変動詞は他の動詞よりも「出来る」による可
能の形のバリエーションが多いことになる。サ変動詞の「出来る」
に焦点を当てる理由の一つはここにある。

今回の調査では対象を漢語サ変動詞に絞り、和語のサ変動詞に
ついては除外した。これは問題の範囲を限定して扱いやすくする
ためであるが、和語サ変動詞の用例数は少なく(注1)、これを
除外しても全体の用例数には大きな影響はない。また、語幹が一
字の漢語サ変動詞も対象から外した。「愛する」「適する」「信ず
る」など一字語幹の漢語サ変動詞は数多いが、これらは「Xでき
る」「Xができる」の形になることがなく、一部の動詞について
は活用の面でも違いを見せるなど、可能形を作るにあたっての振
る舞いが大きく異なるからである。

なお、「Xができる」という形については、漢語サ変動詞の可
能形というよりも、単に漢語名詞が「できる」という動詞ととも
に使われたものとみるべきだという考え方もあるだろう。少なく
とも構文的には、漢語サ変動詞の可能形と見なければならない理
由はない。しかしながら、この「Xができる」という形は、「X

できる」という形が用いられるようになる以前に「Xすることができる」の短い形として用いられていたという経緯がある。つまり「Xができる」には「Xできる」の前身と見てよい一面があるのであって、今回はこの立場から、漢語サ変動詞の可能形の一種として扱うことにした。

3

太陽コーパスにおける可能表現形式

可能表現形式と文体

「出来る」(注2)以外の可能表現形式についての詳しい検討は他の機会に譲るが(注3)、これら諸形式の中における「出来る」の位置を見定めるために、各形式が漢語サ変動詞について用いられた用例数についてまとめ、あわせてこれらの形式と文体との関わりについて見ておきたい。

表1に挙げた形式のうち、「能ふ」の各形式(“連体形+ことあたふ”“連体形+あたふ”“連用形+あたふ”),「～を得」の形式(“連体形+ことを得”“連体形+を得”)は、『太陽コーパス』では文語文における可能表現形式としてもっとも広く使われた形式である。出現するのはほぼ文語文に限られ、慣用句などをのぞけば口語で用いられることはほぼない。これに対して、今回取り上げた「出来る」形式は文語文中でも用いられないわけではないが、多くは口語文中で用いられている。

一方、『太陽コーパス』の記事の文体は、ちょうど1895年から1925年にかけて、文語文から口語文へと大きく移り変わっていく。そのため、可能の形も文体にあわせて大きく変化していくことになる。図1は漢語サ変動詞に限って、可能表現形式の用例数の移り変わりの記事の文体の割合を一つのグラフにまとめたものである(注4)。『太陽コーパス』において「出来る」の用例数は一貫して増加しているのであるが、この図からもわかるとおり「出来る」の増加の最大の要因は文体の変化にある。口語文の中で「出来る」の使用頻度が増しているわけではないのである。

「出来る」の勢力の伸張は、口語文が増加してゆく中で、文語文の「～能ふ」「～を得」に代わって「出来る」が用いられた結果であると考えられる。これは単に両者の増減が相関関係にあるというにとどまらず、「出来る」が文語文における「能ふ」などの代わりの形、いわば口語形として用いられるようになったものと

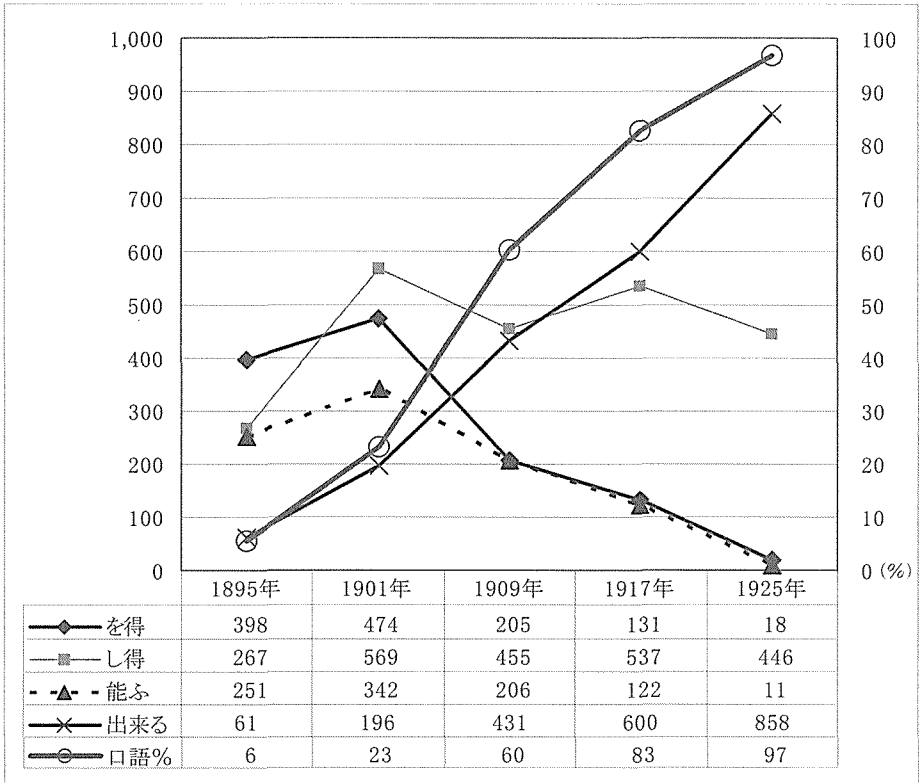


図 1 漢語サ変動詞可能形の用例数と口語記事の割合の推移

考えられる（小椋・小木曾・近藤2002参照）。

なお、助動詞を用いる「される（せらる）」については、その語形がコーパスから取り出しにくいこと、用法の分類が難しいことから、全数調査は出来ていない。概観したところでは、「せらる」「される」はともに可能の意味でも用いられている。「せらる」が文語文、「される」が口語文に用いられるのはいうまでもないが、両者をあわせてみるならば文語文・口語文のどちらでも用いられており、「能ふ」のような極端な文体差はないようである。

「出来る」諸形式の成立と普及

つづいて、先行研究を元にして「出来る」諸形式の成立時期やその後の推移について概観しておきたい。「出来る」全般について概括的にいえば、江戸時代後期より可能の意味を担うようになり、その後、使用が広がるなかで新しい形が広がっていったものとみられる。

渋谷 (1993) は、これら「出来る」諸形式について、形態的特徴からみて発生順序は「Xができる」→「Xすることができる」→「Xできる」であったとする。そして、「出来る」以前に使われた可能表現形式「なる」において既に「Xがなる」「Xすることがなる」「Xなる」の形が認められることから、これを前提として、どの形も早い時期から存在する可能性があったのだと述べている。

『太陽コーパス』の時代は既にどの形式も成立した後ではあるが、各形式が普及してゆく時期であり、ちょうどその使用の移り変わりが観察できる時期にあたる。

「Xすることができる」

「することができる」について 鶴岡 (1967) は、江戸時代末期、文化文政期頃に生まれ、明治10年頃から明治末にかけて急速に普及した語であると述べている。また、神田 (1964) はこの形式が明治中頃の資料から急に多くなり、その後次第に増加したと指摘している。

原口 (1985) は「江戸後期語において、スルコトガデキルの可能の表現の発達は文章語で先行した節がある。宝暦から化政期にかけて、ナルとデキルの交替が平俗な文章語で実現した」と、その成立が文章語に関わるものであったと述べるとともに、明治以降の普及にも外国語学習の影響とともに文章語が関与したことを指摘している。この文章語の関与という点に注意を払いたい。

『太陽コーパス』の時代はこの形がかなり普及してきた時期ということになるが、前に見たとおり、文語文においては「能ふ」「得る」などが勢力を持っていたこともあり、この形は当初はそれほど用いられていない。

「Xができる」

形から想定される発生順序のとおり、神田 (1961) の調査結

果によると、「Xができる」という形は「することができる」と同じく古くから用いられている。この調査はすべての動詞を対象としたものであるが、漢語サ変動詞に限っても、「Xができる」は「Xすることができる」と同じ頃から用いられていたとみてよい。しかし、この形は後に使われなくなってゆく。神田（1964）では次のように述べている。

「名詞+ができる」の形は、明治四十年代の資料までは割合に使われているが、昭和初年以後はほとんど使われていない。この「説明ができる」式の言い方は短い間のもので、次第に「説明できる」という直接「できる」を接続させる形に代わってしまうのである。現在、この形は、サ変動詞になる漢語の場合は次第に少なくなり、〈話ができる〉〈身動きができる〉のように、和語の可能形の場合に残っているように思われる。

このように、「Xができる」は明治末頃から徐々に「Xできる」にその座を明け渡していったものと考えられる。

「Xできる」

原口（1985）によれば、すでに幕末の文章語（海保青陵らの文章）にこの形が用いられているというが、この形が一般に普及するのは、先の神田の指摘通り、かなり時期が下るようである。

神田（1961）ではこの語について次のように述べている。

名詞に直接「できる」をつける形は、明治末頃から一般に使われ出し、大正から昭和の初めまでに非常に多くなったようである。現在では二字の漢字熟語は、サ変動詞として「連体形+ことができる」の形を使うことは少なく、ほとんどこの形を使う。

つまり、「Xできる」は、先に見たとおり明治40年代までの「Xができる」のあとを継ぐようにして広まってきたものであるが、漢語サ変動詞の場合には、さらに「Xすることができる」という形をも「Xできる」が置き換えてゆくというのである。

早い時期から存在したはずの「Xできる」がずいぶん遅れて広まっていった理由は定かでないが、口頭語としては使われていたものの、なかなか文章語に取り入れられなかったという可能性が考えられよう。

以上を元に、『太陽コーパス』での漢語サ変動詞+「出来る」諸形式の現れ方を予想するならば、次のような変化が観察できるものと期待されよう。

- ・「Xができる」の減少
- ・「Xすることができる」の減少
- ・「Xできる」の増加

この3つの形は現代語でもそのままの形で使われるものであり、文脈や微妙な文体的価値の違いなどによって使い分けられているものと考えられる。しかし通時的に見るならば、現代語のようなあり方に行き着くまでに、それぞれの使われ方はかなり変化を遂げてきたものと考えられるのである。

5

用例の検索・絞り込みの方法

ここで、用例の検索に用いた方法と、検索にあたっての問題点や注意を要する用例の扱いなどについてまとめておく。

検索の方法

検索には全文検索システム『ひまわり』を使用し、データは2004年2月時点の『太陽コーパス』を対象とした。検索にあたっては、表記の揺れに対応して検索漏れを防ぐとともに、調査対象外の文字列が含まれないように正規表現を用いて検索した。

「Xすることができる」 [一-脛]する(こと|事)[あ-ん](でき|出来)

「Xできる」 [一-脛](でき|出来)

「Xができる」 [一-脛] [あ-ん] (でき|出来)

検索結果は表計算ソフトに読み込み、調査対象外の例を手作業で排除したのち、集計を行った。

「Xすることができる」

「Xすることができる」の用例を検索するにあたっては、「Xすることができる」だけではなく、「Xすることはできる」「Xすることもできる」のように、係助詞「は」「も」に置き換わった形も用例に加えた。ただし、「こそ」や「だけ」など他の係助詞・副助詞に置き換わったものは除外している。

さらに、連体節中にある「Xすることのできる」も含めた。「Xできる」などの他の形式では、連体節中でもそのままの形で使われるため、用例数に含まれている。これに対し、連体節中の「Xすることができる」相当の形を排除してしまつては偏りが生

じることになるためである。

「Xができる」

「Xができる」は、サ変動詞の可能の形として認めるべきかどうかの判断が難しい場合がある。特にこの時代の「できる」は今日用いられないような用法も備えているため判断は容易ではない。

「Xができる」形式としての認定にあたっては、他の「出来る」と相互に置き換えがきく例であることを原則とした。そのため、次のように「Xすることができる」「Xできる」とは置き換えがきかない例は一律に除外した。

- ・ Xが「～の」「～な」などの連体修飾を受けている場合
- ・ Xが現代語においてサ変動詞として用いられることがなく、太陽コーパスにおいてもサ変動詞としての用例が認められない場合

このように構文や語形から判断することのできないものもある。たとえば「試験ができなかった」という例では、文脈から見て「試験が思惑通りにいかなかった／いい点が取れなかった」という場合は「試験」を単なる名詞とみて除外し、「試験することができる」と解される場合にはサ変動詞の可能形として判断することになる。こうした例は個々の文脈を見て判断した。

「Xはできる」「Xもできる」

扱いが難しいものとして、「Xはできる」「Xもできる」という形がある。これらは「Xができる」の「が」が「は」「も」に置き換わったものとも解釈できるが、「Xできる」に「は」「も」が介入したものと見ることも可能である。係助詞「は」の性質上、「Xできる」に「は」が介入したものであれば、「Xができる」の「が」が「は」に置き換わったものよりも対比の意味合いが強くなるといったことは考えられるが、用例を見ても事実上区別ができない場合が多い。そこで、これらは「Xができる」「Xできる」の両方の可能性を持つものとしてひとまず別扱いすることとした。

このほかの係助詞・副助詞が介入しているもの（「Xこそできる」「Xだけできる」）や、それらが複合しているものの存在も考えられるが、今回は調査の対象外とした。予備的な調査では、こうしたものの実数はさほど多くなく、全体に与える影響は小さいものと考えられる。

「Xをすることができる」

以上のほかに、漢語サ変動詞+「できる」に関連する形式として、用例数は非常に少ないものの、次のような「Xをすることができる」という形の用例がある。このうち他の形式と置き換え可能な例が1925年に2例認められた。

蟻はたしかに決心をすることが出来る。(1925年7号「蟻の戦争」橋爪生P181A01)

これは漢語サ変動詞語幹の名詞性に関わる問題として、可能の形以外の漢語サ変動詞とあわせて考えてみなければならないものであるが、今回の集計には含まれていない。

6

「出来る」諸形式の用例数の推移

以上の条件の下に、漢語サ変動詞+「出来る」諸形式の用例数の推移をまとめると図2のようになる。

この図からわかるとおり、すべての「出来る」形式で用例数が増加している。「出来る」全体の合計では、1895年には61例にすぎなかったものが1925年には858例と14倍もの数に上っている。たいへん急激な増加であるが、これは前述したとおり、文体の変化によるものであり、文語における「～能ふ」「～を得」などの形式を置き換えるように「出来る」が用いられたためであると考えられる。

用例数だけを見ると、すべての形式で常に増加していることから、どの「出来る」も使用が盛んになっていったように思える。しかし、図3のように割合の変化として見てゆくと、「Xができる」「Xは／もできる」は割合の上では減少しており、その反対に「Xできる」が割合の上で増加している様子が浮かび上がってくる。

「出来る」全体の増加が文体の変化によるものであることを考えると、割合の上で減少してゆく「Xができる」は、実のところ、使用が増えているというよりは勢いを失いつつあることを示しているといえることができる。

これに代わって勢いをもち始めるのが「Xできる」である。1909年以降、割合が急激に増加していく。最終的に「Xができる」を逆転するまでには至らないものの、1925年には同数にな

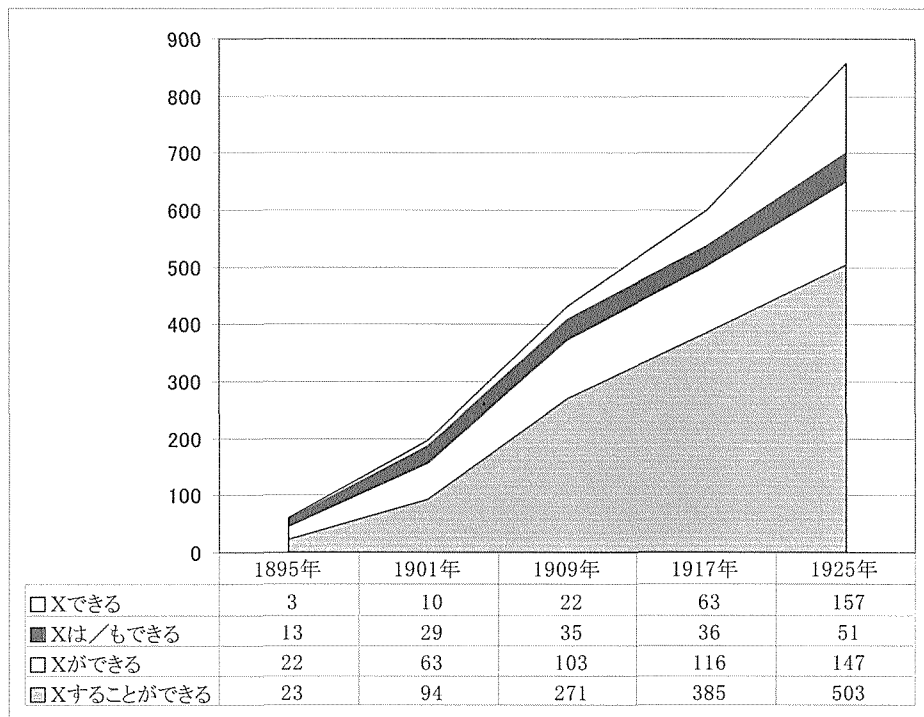


図 2 漢語サ変+「出来る」諸形式の用例数の推移

っている。

「Xできる」「Xができる」の両方の可能性を含んでいる「Xは／もできる」は、図3を見ると「Xができる」とほぼ同じように推移しており、「Xができる」の文脈に応じたバリエーションとして考えておいてよさそうである。

「Xができる」の割合が1909年から減少しはじめるのに続いて、1917年から「Xすることができる」も減少に転じる。「Xできる」は、まず「Xができる」から置き換えはじめ、やや時期が遅れて「Xすることができる」を置き換えていったようである。

このように「出来る」形式が多用されるようになる中で生み出されたいくつもの形が、徐々にその全体に占める割合を変化させていく様子が用例数から見て取れる。「Xすることができる」「Xができる」に代わって「Xできる」が伸張してゆくこの変化は、先行研究をもとにした予想と一致している。これまでの研究で論じられてきたことが、コーパスの大規模なデータによって裏付け

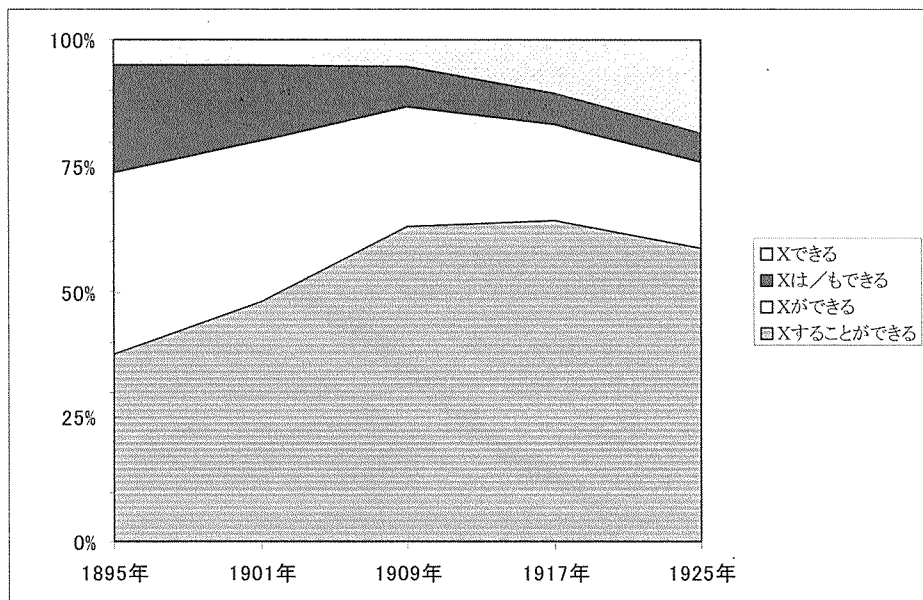


図3 漢語サ変+「出来る」諸形式の割合の推移

られたといつてよいであろう。

「Xすることができる」から「Xできる」へ

このような変化の中で、特に注目しておきたいのは初期の「Xできる」についてである。「Xできる」は、1895年に3例、1901年に10例みられるが、1895年の用例は講演が2例、発話の引用が1例であり、1901年の用例は6例が小説の会話部分、3例は新しい言葉遣いが目立つ福地信世・鳥居龍蔵の紀行文である。これに対して、「Xすることができる」は1895年に23例、1901年に94例みられるが、当初より論説文でも用いられており、ジャンルの偏りはみられない。もともと1895年の口語文の多くが講演速記や小説の会話部分など口頭語的なものであるとはいえ、「Xできる」はやはり口頭語的な性格が強かったものと考えられる。1909年以降の用例ではこのような大きな偏りはなく、ごく一般的な論説文でも広く用いられるようになっている。

このことから、「Xできる」は、「Xすることができる」と同じように口語文で用いられはじめた形式ではあるが、「Xすることができる」が「能ふ」などを置き換えるような文章語的な性格の

ものであったのに対して、「Xできる」はもともとかなり口頭語的な性格が強いものだったと考えられる。だとすれば、1909年以降、「Xできる」が他の形式を押しつけるようにして広まっていく過程は、口頭語的なものが文章語へと取り込まれていく過程だったということができよう。

同じような口頭語的な傾向は「Xができる」についても認められ、早い例の半数近くは小説の会話文中にでている。ところが、こちらは文章語に取り込まれて一般化することではなく、後述するように和語や一部の口頭語的な語について用いられる形式として終わったものとみられる。

漢語サ変動詞の可能形としての「Xできる」

「出来る」各形式の変化は、大きくいえば「Xできる」という簡潔な形が定着してゆく流れだといえる。しかし、「Xができる」と比較したとき、「Xできる」は単に助詞「が」が落ちたものではなく、構文的な変化を伴っていることに注意する必要がある。

「Xができる」は、たとえば「製錬夫の手に委せれば完全に銀の分離が出来る」(1895年8号)のように、Xに対する連体修飾の形でXの格的要素をとることができる。したがって「Xすることができる」と同じように、いったんまとめ上げられた格的要素を含む「こと」(事態)が「出来る」(出来る)という形をとっているといえる。

しかし「Xできる」の場合には、Xが連体修飾を受けることはなく、格的要素を承けるのはあくまでも「Xできる」全体であって、「こと」と「出来る」とを分離することができない。逆にいえば、このときの「出来る」は「こと」がどうしたかというよりも、述語動詞に直接可能の意味を付加する形で働いているとみることができる。

同じような変化は「能ふ」でも見られる。「すること能ふ」に加えて「するを能ふ」「する能ふ」などの「こと」を含まない連体形を受ける形が用いられるようになり、ついには連用形に直接する「し能ふ」という形を生むに至るのだが(小木曾2003)、連用形に直接するというのは単に接続の問題ですまされることではなく、「能ふ」が述語動詞に対して可能の意味を付加する形となっていることの反映だと考えられる。

もともと可能表現は出来事の成否に関わるものである。「ことが出来る」という、文字通り事態の出来をあらわす言葉が可能を表す形式となるのもそのためであって、「～こと」という事態に

まとめ上げた上でそれが実現するか否かという形で可能の意味は表されてきた（注5）。「すること能ふ」にせよ「することができる」にせよ、古い長い形はこれを忠実になぞった語形をしている。それは単に述語動詞に可能の意味を加えるというのではなく、事態をいったん「こと」にまとめてその成否を述べているのである。

これに対して、「Xできる」や「し能ふ」のような語形は、事態の実現の成否という形を経由せず、述語動詞に直接可能の意味を付加する形式となっている。そういう意味で、「Xできる」はまさに「Xする」の「可能形」と呼ぶべきものになっているといえるであろう。このような、語形の上で「こと」としてまとめ上げない、動詞に可能の意味を付加するような可能表現形式の広がりには、単に短い形が新しく現れたというだけでなく、近代語における可能表現の新しいあり方を反映したものであるように思われる。

7

上接語（漢語サ変動詞語幹）と「出来る」諸形式

「出来る」の使われ方が、時代とともに移り変わっていることを見たが、ここでは、「出来る」諸形式の用例における「X」、すなわち上接語（漢語サ変動詞語幹）の違いという観点から分析を加えてみたい。表2は、形式ごとに最も多く用いられている「X」の上位10語を挙げたものである。この表だけでははっきりした違いは見えてこないが、以下でより詳しく見てゆくことにする。

用例数の推移は、語幹漢語を特定の語に絞って見ても全体とほぼ同じような傾向を示すものが多い。図4・図5は、各形式の用

表2 漢語サ変＋「出来る」で多く用いられる上接語（漢語サ変動詞語幹）

形式	異なり語数	用例の多い語（括弧内は用例数）
「Xすることが できる」	583	発見(33),想像(22),維持(21),断言(21),使用(19),満足(17),否定(16),説明(13),利用(13),増進(12),實現(12)
「Xができる」	238	想像(18),我慢(13),安心(13),了解(11),賛成(10),満足(9),説明(8),試験(7),信用(7),生活(7)
「Xできる」	129	想像(13),實行(13),満足(9),安心(8),断言(8),了解(7),理解(5),使用(5),生活(5),適用(5)

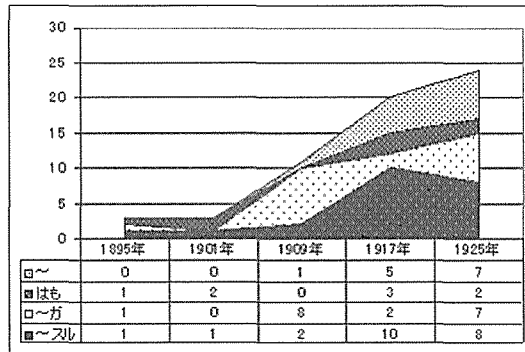


図4 「想像+出来る」(計61例)

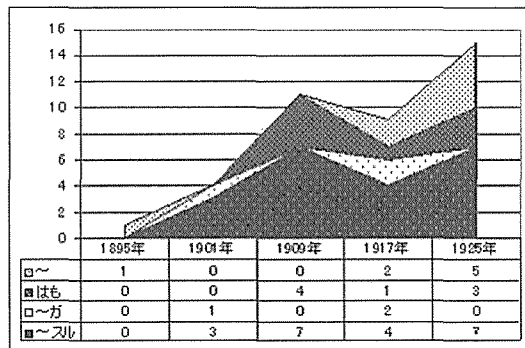


図5 「断言+出来る」(計40例)

例数の合計がもっとも多いもの2語「想像」「断言」について用例数の推移をまとめたものである。

このように、全体の推移と同じような用例数の推移を示す語が目立つが、中には用例が一つの形式に偏るなど、特異な分布を見せる語も存在する。

その一つは、「Xすることができる」では用いられていない語である。「Xすることができる」形式は、他の形式と比べて用例数が多いため、この形で用いられる語は数多い。にもかかわらず、この形をとることのない語が認められるのである。「Xすることができる」の形をとらない語のうち、用例数の多いもの上位3語についてみると、表3のようになっている(注6)。

表3 「Xすることができる」では用いられない漢語サ変動詞

	Xができる	Xは／もできる	Xできる
我慢	13	2	4
感服	4	1	4
承知	6	0	2

これらの語は、他と比較すると、会話などの口頭語において使用頻度が高い日常的な語であるといえそうである。また、これらの語では「Xができる」の用例が多いことも特徴としてあげられる。

こうした語は、文章語で用いられる以前に口頭語において「Xができる」の形が定着していたものと考えられる。『太陽コーパス』中の用例も、小説の会話文中に用いられているものが多く、たとえば「我慢が出来る」は13例中7例までが小説の会話文であり、残りも口語体の随筆類である。「Xすることができる」という形はあまりに長く分析的な表現であって、やはり会話にはなじまない形であると思われる。これに代わって口頭語では、短い形である「Xができる」が早くから用いられていたであろう。

上接語のこのような偏りは、「～ができる」形式が和語について用いられるという先の指摘（神田1964）とも重なるところがある。『太陽コーパス』において和語のサ変動詞である「仕事」＋「出来る」の用例を調査すると「仕事することができる」の用例は皆無であり、「仕事ができる」が20例見つかる。このように、和語や日常的な漢語については「Xができる」が早くから用いられ、そのために「Xすることができる」は使われることがなかったのだと考えられる。

なお、ここで取り上げた「我慢」「感服」「承知」の用例には、「我慢ならない」「我慢しきれない」「感服しがたい」「承知しかねる」など、不可能を表すさまざまな形が認められ、ここでも他の語と際違った違いが見られた。これには語の意味も関与していると思われるが、語の文体的価値や使用頻度とも関わっている現象であろう。

以上のような「Xすることができる」で用いられない語とは逆に、用例数が多いにもかかわらず「Xすることができる」でしか用いられない語も存在する。用例数が比較的多いものをあげると次の通りである。

増進(12),期待(11),繼續(8),發揮(7),輕減(7),除去(6),製造

(6),設立(6),増加(6),従事(6)

(括弧内は用例数)

こちらは口頭語では用いられにくいものが多く、実際、小説の会話文中で用いられた用例は存在しない。「Xすることができる」が「Xできる」よりも文章語的であることと平行して、上接語も文章語的なものが現れているといえる。

なお、上接語以外に、「著者」や「ジャンル」などの情報を元にした分析を行ったが、「出来る」の使われ方についてはっきりとした傾向は認められなかった。著者ごとに「出来る」の用い方に違いがあるということは言えそうであるが、個々の用例数が少なくなってしまうこともあって、そこから何らかの理由を見いだすことはできないようである。

8

まとめ

以上、『太陽コーパス』の1895年から1925年までのデータの調査から、当時の「出来る」諸形式の使用実態について、次のようにまとめられよう。

- ・「出来る」諸形式はすべての形式で用例数が急増した。これは文語文から口語文へという文体的変化によるものであり、文語文における「能ふ」「得る」などの形式に代わって、口語文において「出来る」が用いられるようになったものである。
- ・「出来る」諸形式はすべて用例数が増えているが、「Xすることができる」「Xができる」の用例数の伸びは1909年頃から鈍化しており、割合で見ると「Xできる」形式に押されている。もともと口頭語的であった「Xできる」形式は1909年頃から文章語においても勢いを増し、「Xができる」「Xすることができる」の一部を置き換えるようにして勢力を広げていった。
- ・「Xができる」形式は、日常的で使用頻度の高い口頭語的な性格の強い語で多く用いられ、こうしたものの中には「Xすることができる」形式で用いられない語もあった。逆に「Xすることができる」形式は文章語的な性格が強い語で多く用いられた。

注

- (1) 和語のサ変動詞の形として謙譲語形「御へできる」があるが、この用例はすべての「出来る」形式を合計しても10例以下にすぎない。
- (2) 以下、「Xすることができる」などの個別の形式を指す場合には、かな書きの「できる」を、「できる」を用いる諸形式全体を指す場合には漢字を使った「出来る」を用いる。
- (3) 『太陽コーパス』における「能ふ」のバリエーションについては、小木曾（2003）で漢語サ変動詞に限定せず論じたことがある。
- (4) 小椋・小木曾・近藤（2002）では漢語サ変動詞に限らず、『太陽コーパス』の文体と可能表現形式との関係をまとめている。
- (5) 「れる・られる」による可能表現もまた事態の出来ということを基盤とした表現であると考えられる（尾上1998-1999）。可能性を表す形式として発達した「し得る」についてはまた別に考える必要があろう。
- (6) 用例数4位以下は計5例以下となり、特定の記事に偏って現れる語（手術5例・宿泊4例・歩行4例など）であるため取り上げなかった。

参考文献

- 小木曾智信（2003）「近代日本語における『能ふ』の用法 —『太陽コーパス』の用例から—」（『明海大学外国語学部論集』第15集，1-11頁）
- 小椋秀樹・小木曾智信・近藤明日子（2002）「『太陽コーパス』を使った近代語表現の通時的研究—口語文体・可能表現・待遇表現について—」（『国語学会2002年度春季大会発表要旨集』177-184頁，国語学会）
- 尾上圭介（1998-99）「文法を考える5・6・7 出来文（1）（2）（3）」（『日本語学』17巻7号，76-83頁，17巻10号，90-97頁，18巻1号，86-93頁，明治書院）（再録：尾上2001『文法と意味Ⅰ』くろしお出版）
- 神田寿美子（1961）「現代東京語の可能表現について」（『東京女子大学 日本文学』第16号，70-84頁）
- 神田寿美子（1964）「見れる・出れる—可能表現の動き—」（時枝誠記他編『口語文法講座3ゆれている文法』81-91頁，明治

書院)

小松寿雄 (1982) 「近代の文法Ⅱ (江戸篇)」 (『講座国語史4文法史』 537-614頁, 大修館書店)

渋谷勝己 (1993) 「日本語可能表現の諸相と発展」 (『大阪大学文学部紀要』 33巻第1冊)

鶴岡昭夫 (1967) 「江戸語・東京語における可能表現の変遷について」 (『言語と文芸』 54, 54-63頁)

原口 裕 (1985) 「可能表現『スルコトガデキル』の定着」 (『国語と国文学』 62-5, 56-66頁)

漢字の実態と処理の方法

——田中 牧郎

1

はじめに

本論文は、『太陽コーパス』の漢字処理の方法について考察するものである。その概要は、本書の「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」（田中牧郎）の、「3.5漢字字体」（8頁から）、「13.外字タグ」（34頁から）に述べたが、本論文では、より広い視野から『太陽』における漢字の実態とそれをふまえた『太陽コーパス』の設計における漢字処理の方法について、詳しく考察する（注1）。上記論文の該当部分と読み合わせていただければ、幸いである。なお、本論文で利用する『太陽コーパス』のデータは、2004年3月現在のものである。

2

基本方針

2.1 文字の電子化

文献資料を電子的なテキストとするための最も基本的な作業のひとつとして、文献上の文字の一つ一つを、コンピューター上に用意された文字集合の区点位置の一つ一つに対応づけていく作業があげられる。特に、日本語の場合、漢字を中心に文字の種類が多く字体のゆれが大きいいため、電子テキスト作成におけるこの作業の占める位置は大きい。原資料となる文献（原文）の文字がコンピューターの文字集合に用意されていない場合や、原文の文字とそれに対応しそうなコンピューターの文字との間で、文字の形が一致していない場合などに、問題が表面化しやすい。

明治・大正期の雑誌『太陽』をもとに、日本語研究を目的として電子化テキストが作成される『太陽コーパス』の

設計においては、原文をいちいち参照しなくても、幅広く高度な目的に使用できる共有の資料としての電子テキストを作成することが求められる。こうした目的を達成するためには、文字の電子化作業において、できるだけ多くのコンピュータで扱えること（汎用性）、原文の文字をできるだけ確実に再現すること（再現性）、目的とする文字や言葉をできるだけ確実に検索できるようにすること（検索性）の三つの方向を追求することが望まれよう。まずこの三つを実現させるための基本方針から検討したい。

2.2 汎用性

現在、コンピュータで日本語を処理する規格としての文字集合には、一般的なものとしてJISとユニコードとがある（注2）。その規格名と文字数、普及状況は表1の通りである（注3）。

表1 日本語文字集合の現状

文字集合の種類	文字数	日本での普及状況	世界での普及状況
JIS X0208 1997	6879	広く普及	—
JIS X0213 2000・2004	4344	普及せず	—
ユニコード 4.0	70195	普及途上	普及途上

単純に文字集合の汎用性という点から見れば、世界標準として開発が進められもつとも普及範囲の広いユニコードがあげられる。しかし、日本における普及状況を見ると、ユニコードはまだ限定的である。日本語の文字入力に用いる仮名漢字変換の一般的なソフトウェア（MS-IME、ATOKなど）はユニコードに対応しているが、初期設定など普通の利用環境では、簡単にユニコードの文字が使える環境にはなっていないことが多い。『太陽コーパス』は日本語研究のために作られるものであり、その研究は日本語用のコンピュータ上で行われることが最も一般的である。こうしたことから、『太陽コーパス』で採用する文字集合に、ユニコードを用いることは、現段階の利用状況に照らし合わせると、汎用性を満たさない。

JISに目を転じると、現行の主要な規格として、JIS X0208 1997（第一・第二水準）とJIS X0213 2000・2004（第三・第四水準）（注4）とがあり、後者は前者を

補う形で出されたものである。JIS X0213を採用すれば第一水準から第四水準までの11223字が利用できるが、JIS X0213を実装するコンピュータは、現段階ではほとんどなく、今後も普及する見通しはない。この状況は、JISX0213に含まれる文字種はほとんどすべてユニコードに包含されており、近い将来ユニコードが日本においても普及していくことを想定して、コンピュータメーカーがJISX0213を実装しないためだと考えられる。現在の日本語での汎用性という点では、JISX0208が最も安定している。JISは日本の規格であり、そのまま世界に普及するものではないが、ユニコードに包含されることで、JISの各区点位置はユニコードの各区点位置に対応づけられているので、ユニコードを通してJISの文字が世界に普及していくと見ることもできる。ユニコードが将来国内でも普及していけば、JISで作られた電子テキストをユニコードに変換することはさほど困難なことではない。したがって、電子テキスト作成作業においては、現在もっとも安定し普及しているJISX0208を採用し、普及状況の変化に対応できる形に設計しておくことが、汎用性を満たすことになると考えられる。

2.3 再現性

雑誌『太陽』は、使用されている文字の種類が非常に多く、また字体のゆれの範囲も相当に大きく、こうした実態に対応させて、原文の文字をできるだけ確実にコンピュータ上に再現するためには、コンピュータの文字集合は次の二つの要件を満たしていることが望まれる。

- ・収録文字が豊富であること
- ・字体のゆれ幅の認定規準が示されていること

このうち収録文字の豊富さについては、2.2の表1で見たように、JISよりもユニコードが勝っている。一方、字体のゆれ幅の認定規準の観点から見ると、ユニコードではそれが明記されていないのに対して、JISでは「包摂規準」として細部にわたって明確な規定があり、JISの方が勝っている。どちらの要件を優先するかで、望ましい文字集合は変わってくることになる。

ユニコードによれば多くの文字をそのまま入力することができるという利点があるとはいえ、『太陽』の場合、ユニ

コードによってもそのままでは入力できない文字が残る。JISの示す包摂規準にしたがえば、ゆれ幅を大きくもつ多くの文字をJISの区点位置に包摂して入力していくことが可能になるが、その規準では対応できないゆれをもつ文字が残る。このように、文字の種類が多く字体のゆれ幅の大きい『太陽』では、JISによってもユニコードによっても文字集合をそのまま用いるだけでは再現性を十分に満足させることは困難なのである。JISによるにしてもユニコードによるにしても、再現性を確保するためには原文の文字とコンピューターの文字との間の対応づけの方法について独自に策定することが不可欠となる。その独自の方策を工夫する際、現実の文字とコンピューターの文字との対応づけの規準が明示されているJISは、参照価値が高い。

2.4 検索性

原文の文字に対する再現性を保てる形に電子テキストを作成できたとしても、利用者が目的とする文字や言葉を、電子テキストから無理なく的確に検索できるようになっていなければ、コーパスとしての価値を十分に発揮することはできない。文字列検索によって利用することを想定している『太陽コーパス』の場合、利用者が調べたい言葉の表記として想定した検索文字列が、電子テキスト内のその語の表記である文字列と一致することが期待される。『太陽』には現代人として通常想起される表記と異なる表記が行われている場合も多いので、利用者の想起する文字と電子テキストの文字との対応づけを支援するしくみが必要になる。

文字数の多いユニコードで入力してあると、利用者が想起する文字よりもかなり広範囲の文字が使われていることになり、対応づけのためのしくみは、かなり複雑なものになる。文字数が比較的限られ、文字の同定の規準が明示されているJISでは、対応づけのしくみはユニコードの場合よりも簡単なもので済む。この点から見るとユニコードよりもJISの方が扱いやすい。JISが扱いやすいといっても、2000字弱の常用漢字を基本としてそれにいくらかの文字を加えた漢字の範囲で読み書きを行っている一般的な現代人にとって、JISX0208の7000字近くを識別して使いこなすことは容易でない。利用者が想起する文字と電子テキスト上の文字とを対応づけるしくみは、JISで入力した場合

も、不可欠であろう。何らかの支援装置をコーパスの側で用意しておき、検索性を高めることが望まれる。

2.5 独自規準の必要性

以上の検討により、雑誌『太陽』を電子化してコーパスを作成する目的においては、汎用性・再現性・検索性、いずれの観点から見ても、文字集合には、ユニコードよりもJISを用いる方が利点が大きいことが導き出せた。JISのうちでもJISX0213よりもJISX0208によるのが、現状ではもっとも現実的である。一方で、JISX0208によっても、文字の電子化において数々の問題が残ることも明らかになり、独自に処理規準を立てるなどして工夫を行うことも不可欠であることもわかった。したがって、『太陽コーパス』で用いる文字集合はJISX0208とし、不都合の生じる点については独自の処理方法を工夫することとした。

以下では、JISの文字集合や処理規準と、『太陽』の漢字の実態とを、具体的に突き合わせ、電子テキストの作成において補っておくべき規準や施すべき工夫について、具体的に考察する。

3 包摂

3.1 JIS包摂規準による包摂

JISX 0208、およびこれを発展させたJISX 0213で明示された「包摂規準」は、字体のゆれ幅を一つの区点位置に包摂する規準として、現段階で最も整備されたものである。例えば、『太陽』で用いられている「青」の字体は、JIS包摂規準の146番によって「青」と同一視される。したがって、コンピューターの文字集合にJISを採用した場合、原文の「青」はコンピューター上では「青」で表現されることになる。また、『太陽』では、「温」と「溫」とは、どちらの字体も使われているが、この二つは、JIS包摂規準の147番によって「溫」に同一視される。したがって、『太陽』でどちらの字体が使われていても、コンピューター上では、「溫」で表現されることになる。このようにしてJIS包摂規準にしたがうことで、『太陽』の多くの字体をコンピューター

一上のJIS漢字に明示的に対応づけることが可能になる（注5）。

3.2 追加包摂規準による包摂

『太陽』の場合、JIS包摂規準によって字体の包摂を行っても、JIS漢字に包摂できず入力できない文字が残る。明治・大正期の漢字はJIS包摂規準が想定する字体のゆれ幅を超えるゆれ幅をもつ場合も多く、この規準をそのまま適用するだけでは包摂できない文字が大量に残るのである。こうした漢字とJIS漢字とを照らし合わせ、実態を観察してみると、JIS包摂規準の適用外にはあるが、包摂を行うことができるのではないかと考えられる漢字も少なくないことがわかってくる。

一例として、「怨」（「怨」の右上が「巳」）という漢字について検討してみよう。JIS包摂規準67番には、部分字体（漢字の構成要素になる字体）「己」と「巳」とが包摂されることが規定され、同じく70番には、「己」と「巳」とが包摂される規定をあげている。『太陽』に見られる「怨」は、下の例のようにJIS漢字の「怨」と同じ読みと意味で用いられ、包摂可能な字体の変異ではないかと思われるが、「怨」と「怨」とを同一視する「巳」と「巳」とを包摂する規定は、JIS包摂規準には存在しない。三つの部分字体の関係は図1の通りである。

多年の怨みを晴らさせた。（1895年1号「流行」流行記者PI58A26）

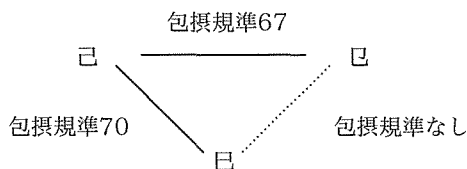


図1 己・巳・巳の包摂関係

「怨」と「怨」とを包摂せずに別の漢字として区別すべき積極的な根拠は、この二字を見る限りでは想定しにくく、この二字についても包摂の扱いにすることが現実的であると思われる。これと同じように、JIS包摂規準の個々の規定の組み合わせや類推によって包摂できるものに、「記」（「己」

が「巳」などがある（「記」に包摂する）。

別の例を見てみよう。『太陽』には、JIS漢字の中にはない「款」（左上が「上」）、「欸」（左上が「止」）という字体が見られるが、読み・意味の観点からは、JIS漢字の「款」と等しいものである。

最惠國條款を有する歐米人も（1895年5号「商業」PI63B25）

第六款（1901年7号「各国近事」P217A20）

JIS包摂規準には「土」「上」「止」などの部分字体を包摂する規定はない。ところが、JIS包摂規準の中には、「魚」と「𩺰」における「𩺰」と「大」のように、特定の文字の部位において、その形が大きく異なり意味も異なる部分字体が、包摂の關係に規定されている場合がある（包摂規準149番）。

魚	款	欸
┆ 包摂規準149	┆ 包摂規準なし	┆ 包摂規準なし
𩺰	欸	欸

図2 魚・𩺰 款・欸・欸の包摂關係

こうした事例に鑑みれば、「款」「欸」「欸」における、「土」「上」「止」などの部分字体も包摂し、JIS漢字である「欸」で代表させることも、問題ないのではないかと考えられる。この例のように、JIS包摂規準と大体同程度の字体の変異と認められそうなものは、「陰」（「陰」に包摂）、「糾」（「糾」に包摂）など、枚挙に暇がない。

さらに次のような例もある。『太陽』には「棲」の字体が見えるが、使用されている読みや意味は「棲」と等しい。

鹿多く此處に棲息し（1895年2号「広島の形勢」野口勝一P070A24）

魚介棲息するに至る。（1895年11号「地質学及び地質学者」佐藤伝蔵P142A21）

これらを包摂することはできないだろうか。この二つの漢字は、手偏と木偏とで相違しており、形の上からは微細な差異とは認めにくい。「揚」と「楊」，「推」と「椎」などのように、手偏と木偏の相違によって別の漢字であることが区別される事例も多い。ところが、「棲」については、『大漢和辞典』が「栖・棲に同じ。〔正字通〕棲，棲・栖通，

从棲好為正」とするように、意味や読みの面では、「棲」と「棲」とは通じるところがある。手偏と木偏については、手書きの場合、くずし字の運筆でほぼ同じ形になることもあり、こうした場合は、偏の異なる漢字間で、包摂の扱いをすることも可能であると考えられる。同様のものに「扑」と「朴」などあり、二水と三水についても「減」と「減」などの例が指摘できる。

包摂		別字	
棲	—— 棲	揚	…… 揚
扑	—— 朴	推	…… 椎

図3 手偏と木偏の関係

以上三つの類型にわたり、JIS包摂規準を補うようにして独自に追加する規準（追加包摂規準）によって包摂できると考えたものを述べた。『太陽コーパス』の追加包摂規準によって包摂した『太陽』の漢字は、約300種類ある。

4 代用

3で述べた手続きによっても包摂できない漢字は「外字」となるが、外字は、コンピューター上で処理できない漢字として、再現も検索も困難になり、コーパスを活用する上での障害になる。コーパスの設計において、この障害はできるだけ軽減しておくことが望まれる。そのための方策として、包摂できない漢字のうちで、JIS漢字で代用できるものは、問題のない範囲で代用するという考え方がある。字形上の差異が軽微であるとはいえず、運筆によってそれが軽微になるわけでもない場合についても、読みと意味とが同じであるなど機能的には等価な関係にある場合は、外字をJIS漢字で代用する方式をとるのである。

例えば、先に3.2で例に取り上げたJIS漢字「款」には、包摂した漢字のほか、対応づけられそうな『太陽』の漢字として、「欸」がある。この「欸」について、『大漢和辞典』は、「款の俗字。〔字彙〕欸，俗款字」と、「款」の俗字であることを示している。『太陽』の使用例を見ても、次のよう

に「款」と同じ語の表記に同じ意味で用いられていると認められる。

俵屋宗達の落款を入れた贋物人物畫ナ（1901年10号「古物家」内田魯庵P102B06）

『つねのふ』と假名の落款のある浮世繪（1925年9号「我楽多雜記」高橋義雄P077B25）

「款」と「款」とは、「欠」を共通させながらも、偏の部分は全く異なる字形であり、包摂の扱いはできない。しかし、『太陽』での使用実態を見る限り、「款」と「款」との間に機能上の区別はなく、等価の関係にあることが認められる。このような場合、外字「款」もJIS漢字「款」で代用して入力することにすれば、検索は可能になる。また、再現性という観点からすれば当該例の漢字が本来は包摂されない別字「款」であったことを示す情報を付与することが望まれるが、次項で詳しく述べる「外字」タグによって、原文での字体を再現できるしくみを用意する。こうした処理の方法を「代用」と呼ぶことにする。

代用を適用する場合の条件として、次の三つを立て、これらの条件のすべてを満たす場合に、代用の処理を適用した。

- (1) 同じ語の表記に用いられること
- (2) 意味による書き分けがないこと
- (3) 字体の変異が次の五つのいずれかの類型にあること

A 配置変換型（偏・旁・冠・脚など、漢字の部分字体の配置が入れ替わる型）

B 部品変異型（偏・旁・冠・脚など、漢字の部分字体が異なる型）

C 部品加除型（偏・旁・冠・脚など、漢字の部分字体に加除のある型）

D 混合型（上記の型を複数組み合わせた型）

E その他（字体上の類似はないが、歴史的に通用が明らかなもの）

(3) の五つの類型別に、具体例を二組ずつ示す。矢印の左側が『太陽』原文の漢字、右側が代用するJIS漢字である。

A 配置変換型

槩 → 概

一槩にこれを視ず（1895年3号「彰義隊 上」曳尾叟P037A16）

一概に反亂人を悪とせず (1895年9号「詩人パイロンの海賊、及び「サタン」主義」木村鷹太郎 P033A17)

毳 → 氈

絨毯, 毛氈, 皮貨の類 (1895年3号「天津港」曾根俊虎 P046A11)

沿道の富家は毛氈を敷き (1895年1号「第一祝捷大会を観る」T O生 P189A18)

B部品変異型

疏 → 疏

一種の誤解又は疑惑を疏通融解し, (1901年14号「外資輸入国としての日本」近藤廉平 (口述) P021A24)

能く政府の意と西郷の意と疏通融和する様に (1901年10号「追懷談」川村純義 (談) P131A13)

輓 → 軟

太子が幼時の輓弱なる体質と (1901年10号「露国の宮廷」日下逸人 (訳) P112A05)

軟弱なる身体を有するに至りしは (1901年12号「日本女子と体育」来住素子 P223B23)

C部品加除型

奄 → 奄

氣息奄々殆んど死に垂んたる有様なり (1895年6号「樺太探検記 (承前)」関口信篤 P072B11)

氣息奄々として將に絶えなんばかりにながら, (1901年4号「鎮西遊記 (接前号)」久保天隨 P114B17)

厲 → 勵

進取の念を厲まし (1895年1号「戦勝後の教育」千頭清臣 P010A07)

勇を鼓し氣を勵まし (1895年1号「日本帝国の任務」中西牛郎 P048B16)

D混合型

欸 → 咳

ボ氏の聲欸に接したることなきも (1901年13号「『政党及議院政治の弊』に対する所感」清野長太郎 P022B09)

直接先生の聲咳に接するが如き (1917年12号

「〈新刊紹介〉」T11C19)

窓 → 窗

悲しげな窓には雨の滴 (1909年1号「幻覚」蒲原
有明 (訳) P117B01)

窗に懸ける琴を下して (1895年12号「韻文に就て」
島崎藤村P183A30)

E その他

开 → 其

开は兎も角も (1895年7号「時事」P023A13)

其は兎も角も (1901年9号「宗教時評」龍山学人
P051B08)

煞 → 殺

煞煞笑倒 (1901年14号「笑殺笑倒」璵珞上人
P153A01)

休々に枉げて笑殺せん (1895年9号「元時代の雑
劇 (七)」幸田露伴P042A16)

代用の処理を行ったものは異なりで200字余りになる。
頻出する上位10字を、頻度と用例とともにあげると、表2
の通りである。

表2 代用とした主な漢字 (頻度の上位10字)

順位	頻度	原文の文字	文字番号	代用字	JIS区点番号	用例	所在
1	272	欸	16085	欸	8ABC	定欸改正の決議を爲し	1895-02-P185A10
2	266	漚	18019	汽	8B44	汽船の出發は	1895-01-P088A04
3	170	雞	42124	鷄	EA50	夜半に鷄の鳴をきいて	1895-01-P003B09
4	85	熙	1721	熙	E087	熙々たる明治二十八年の新旭光は	1895-01-P001D02
5	71	开	51106	其	91B4	其は信ずべからざることなり	1895-08-P018A15
6	70	巷	T003	巷	8D4A	社會は正邪ならび存するの巷	1895-01-P019B01
7	69	厲	56009	勵	99AD	進取の念を勵まし	1895-01-P010A07
8	55	鹽	47550	鹽	EA64	料理獻立の鹽梅は	1895-02-P155A14
9	38	咤	50005	咤	9A42	三軍を叱咤し	1895-01-P054B10
10	32	牒	19790	牒	92AB	韓國宮内大臣に牒を通じしかば	1901-01-P139A16

表の中の「文字番号」については「5.外字」のところで
説明する。代用とした漢字全体は、『太陽コーパス』におい
て外字一覧を生成することによって見るができるが、
この外字一覧についても、「5. 外字」で言及する。

包摂や代用の処理によっても、JIS漢字で表現することが不可能な漢字は、電子テキストとして入力することができない文字である。『太陽コーパス』では、電子テキストにおけるこのような文字の部分には「𪛗」を入力した。しかし、「𪛗」だけでは、再現性も検索性も全く確保できない。そこで、「𪛗」に対して次のような外字タグを付与することで、限定的ながら再現性と検索性に配慮することとした。

龐 <外字 文字番号="048824">𪛗</外字>

3に述べた「代用」の場合もこの外字タグによって再現性と検索性に対応している。

歎 <外字 文字番号="016085">歎</外字>

外字タグの「文字番号」属性には、『今昔文字鏡』の漢字番号を記入する。『今昔文字鏡』は、『大漢和辞典』の全文字5万字強を含んだ10万字にのぼる文字をおさめた文字集合であり、漢字を多く扱う資料の電子的流通や印刷等に、ある程度普及している（注6）。『太陽』には、『今昔文字鏡』にも収録されていない漢字も用いられているが、その場合は当該字を原文から複写し、画像として保管し、その画像番号を「文字番号」属性に記入する方法をとった。

燄 <外字 文字番号="T002">燄</外字>

鯽 <外字 文字番号="T125">𪛗</外字>

『今昔文字鏡』を用いる場合も原文文字の画像を用いる場合も、記事本文を閲覧する場合などは画像ファイルを用いることで『太陽コーパス』利用者のコンピューターに字形が再現されるようにした。

『太陽コーパス』において、外字とされる文字（「代用」を除く）は異なりで1000字、延べで3000字を超える。頻出する外字上位10位までをあげると、表3の通りである。

『太陽コーパス』における外字の一覧（「代用」の一覧も含む）は、コーパスに添付したソフトウェア『プリズム』によって生成することができる。その具体的な方法については、本書の「構造化テキストを直接利用するアプリケーションー『プリズム』と『たんぽぽ』ー」（小木曾智信，83頁から）を参照してほしい。

表3 『太陽コーパス』における主な外字（頻度の上位10字）

順位	頻度	原文の文字	文字番号	用例	所在
1	242	龐	48824	居士世推 _三 鹿門	1895-04-P126A31
2	55	詹	35458	學士，九卿， _三 事，科道等の官を召して	1895-06-P026A03
3	38	睜	23439	涙堪へて眼は _三 れども	1895-01-P092A01
4	28	筭	26062	殆ど十四呎の _三 隙あり	1901-08-P169B11
5	27	弇	9610	輻射輪の形状に於ける環状 _三 が	1901-02-P183B15
6	27	蔡	67692	_三 婆々といふもの出でゝ	1895-06-P101B06
7	26	鉸	40346	蝶 _三 附の扉は無功なるに依り	1901-03-P177B15
8	25	噤	53267	北は準 _三 爾降伏し	1895-06-P026B05
9	24	颺	58594	紙鳶を _三 げたり	1895-03-P061A12
10	22	閩	41315	_三 浙の海寇を綏靖し	1895-03-P094A12
10	22	葶	9808	里見 _三 著	1917-08-T104B13

6

参照

6.1 JISの参照字

ここまで、JIS漢字に対応づけていく『太陽』の漢字を、どのようにして対応づけて電子テキストに入力していけばよいかという方向で、問題点とその解決策について述べてきた。ところで、JIS漢字への対応づけと検索性の問題については、もうひとつ別の方向から工夫が必要な問題がある。それは、等価と考えられる漢字が、JIS漢字のなかで異なる区点位置を与えられているために、コーパス利用者がそれに気づかず、目的とする言葉や文字について検索もれを起こしてしまう問題である。

例えば、JISでは「亜」と「𠂔」とは、それぞれ889Fと98B1の二つの区点位置に別々の漢字として登録されている。この二つの漢字は、いわゆる新字と旧字の関係にあるもので、『太陽』では旧字の「𠂔」の方しか用いられておらず、『太陽コーパス』でも「𠂔」で入力されている。ところが、『太陽コーパス』の利用者は、現代通行の新字「亜」の方を指定して検索を試みる可能性も高い。この場合、コーパスの側で何も手立てを設けておかなければ「𠂔」は検索対象にならず、目的の言葉が検索できないおそれがある。

利用者が指定した「亜」とコーパス内の「亜」とを、あらかじめ関連づけておき、利用者が「亜」と指定しても「亜」を検索対象にすることができるようにしておくことが望まれるのである。

JISでは、「亜」と「亜」のような関連の深い漢字を相互に参照させるように、「参照字」を規定している。こうした参照字の関係にある組み合わせとして、600組余りが規定されているが、これらの組の一つ一つについて、どのような規準で参照字と認定したかについては、規格票等には記されていない。上記の「亜」「亜」のような、新字・旧字の関係にあるもののほか、『大漢和辞典』の記述によれば、俗字（「芦」は「蘆」の俗字）、略字（「壺」は「壺」の略字）、譌字（「𠂔」は「肉」の譌字）、古字（「弌」は「一」の古字）、あるいは単に「同じ」と言われるもの（「菴」は「庵」に同じ）など、さまざまな関係の場合がある。

これら参照字とされるものは、広い意味で異体字と見なせるが、各組の漢字同士の字形上での関係を見わたすと、ちょうど、4で代用とした漢字同士の関係と重なり合うものが多い。

A 配置変換型（偏・旁・冠・脚など、漢字の部分字体の配置が入れ替わる型）

（例） 峰・峯 蓐・蓐

B 部品変異型（偏・旁・冠・脚など、漢字の部分字体が異なる型）

（例） 嘩・譁 煙・烟 鬱・鬱

C 部品加除型（偏・旁・冠・脚など、漢字の部分字体に加除のある型）

（例） 余・餘 叡・睿 一・弌

D 混合型（上記の型を複数組み合わせた型）

（例） 睹・覩 杯・盃

E その他（字体上の類似はないが、歴史的に通用が明らかなもの）

（例） 欠・缺 万・萬 村・邨

このような類型のほか、参照字とされるものの中には、包摂関係にある漢字の組み合わせにおける字形上の関係と重なるものもある。

（例） 獎・獎 飲・飲

これらは、JISでは当該包摂規準に「適用除外」と規定され、

双方に区点位置を与えているものである（上の例では、包摂規準連番124と155）。

参照字とされる漢字の組のうち、『太陽』において、「椿」「檣」の組の場合など、いずれの漢字も使われていない場合は問題ない。また、先に見た「亜」「亞」のように、参照字のうち一方の字が『太陽』では使用されていない場合は、コーパス利用者が使用されていない漢字を指定した場合、使用されている方の漢字に変換させることで問題は解決する。ところが、参照字の組の両方の漢字が『太陽』で使われている場合は、問題が複雑になる。この場合、二つのタイプがある。

第一のタイプとして、二種類の漢字が同じ語の表記に同じ意味で使われている場合がある。例えば、「鬱」と「鬱」の組について見ると、『太陽』では次のように用いられている。

天気は陰鬱なること多くして（1895年7号「梅雨に就き頓野新説（上）」頓野広太郎P138B08）

風土の陰鬱なと（1917年1号「蕈」徳田秋声P388A14）

このほかにも、「沈鬱」「沈鬱」、「憂鬱」「憂鬱」、「鬱蒼」「鬱蒼」、「鬱陶しい」「鬱陶しい」などのように用いられおり、「鬱」と「鬱」との間に意味上の書き分けは認められない。この場合、検索者が「鬱」を指定しても、「鬱」の字体も用いられていることを示し、逆に「鬱」を指定すれば「鬱」の字体も用いられていることを示し、検索目的とする語について両方の漢字で検索を実行できるようなくみを用意しておくのが望ましい。

第二のタイプとして、二種類の漢字が相互に通用されず書き分けられている場合がある。例えば、「育」と「毓」の組では、「育」が、和語「そだつ」「そだてる」「はぐくむ」の表記のほか、漢語「教育」「生育」「発育」「育英」や、地名ニューヨークの表記「紐育」などと用いられているのに対して、「毓」は、人名の表記に限られており、『太陽』における実態を見る限り、明らかに異なる価値をもった漢字である。

師匠の教育の如何では（1917年2号「悪人」小川未明P213A09）

紐育へ出かけた。（1917年2号「無言劇の復興」厨川白

村P126B09)

孫が、清代の孔毓圻である。(1917年3号「支那学研究者の任務」桑原隲蔵P103B09)

このような場合は、参照字とされるものであっても、相互に参照させるしきみを用意する必要はないものと考えられる。

また、JISにおいて参照字の規定がない漢字についても、お互いに通用され、意味による書き分けのない漢字の組がある。例えば、「踏」と「蹈」は、参照字とはされていない関係にあるが、『太陽』では、和語「ふむ」「足ぶみ」などの表記にはどちらも用いられるほか、「踏査」「蹈査」、「踏破」「蹈破」、「雑踏」「雑蹈」、「舞踏」「舞蹈」など、同一語の表記に同じように用いられている。

實地踏査した所によりて (1901年3号「日本の製糖業」鈴木藤三郎(談) P077B02))

實地蹈査して (1909年1号「名士の独逸観」窪田空々閑人P196A07)

「踏」と「蹈」についても、「鬱」と「鬱」などの場合と同様に、相互参照の扱いをすることが望ましいと考えられる。

以上のように、JIS漢字の中にある漢字同士で、『太陽』において通用される組み合わせの場合、検索時に相互参照できる形にしておき、検索性を確保する工夫が望まれる。

6.2 字体変換辞書

相互に通用する漢字の参照関係をコーパスの設計において規定しておく具体的な方策として、『太陽コーパス』では、参照関係を定義した「字体変換辞書」を作成し、検索時に、この辞書を用いるようにした。

字体変換辞書は、JISの参照字の関係にある漢字と、6.1で例にあげた「踏」と「蹈」のような参照字と同様の関係にあると考えられる漢字をおさめ、表4のような構成とする。

「指定字体」とは、利権者が検索時に指定することを想定した字体であり、見出し字の役割をもつものとして、この辞書におさめるすべての文字を一字ずつ掲げたものである。「等価字体」とは、指定字体に対して『太陽コーパス』において等価の関係で用いられている字体を指す。指定字体そのものが『太陽コーパス』でも用いられている場合は、

表4 字体変換辞書の例

指定字体	等価字体	参考字体
亜	亞	
亞	亞	
鬱	鬱 鬱	
鬱	鬱 鬱	
育	育	毓
毓	毓	育

これも等価字体としておさめた。「参考字体」は、JISの参照字として参照関係とされているものの、『太陽コーパス』においては等価でないものをおさめた。

例えば、指定字体「亜」については、これと等価の関係で『太陽コーパス』で用いられている「亞」を引き出すことができるように等価字体に「亞」をおさめた。『太陽』で用いられていない「亜」は等価字体にはおさめない。また、「鬱」については、等価の関係で『太陽』に用いられている「鬱」と「鬱」を等価字体におさめた。「鬱」を指定字体とする場合にも、この二字を等価字体におさめた。「育」「毓」については、それぞれ「育」「毓」を1字ずつ等価字体とするが、それぞれ参考字体として「毓」「育」をおさめた。「参考字体」は、等価でないものについても参照関係にある字体の確認が必要とされる場合を想定して設けたものである。

『太陽コーパス』において検索を実行する際に、指定字体は等価字体に変換できる。利用者の目的によっては参考字体も検索したい場合もあることを考慮して、参考字体にも変換可能なオプションを用意した。検索時の字体変換辞書の運用方法については、本書の「構造化テキストに対応した全文検索システム『ひまわり』」（山口昌也、70頁）を参照してほしい。

字体変換辞書は1300字余りの見出しからなっており、『太陽コーパス』CD-ROMのなかの、jitai.xmlファイルに、xml形式で格納されており、一覧形式で閲覧することも可能である。

漢字の代用や参照の処理は、コーパスの検索性をあげるための工夫であるが、この方向にある工夫として、もうひとつ、以下に述べるような方策を講じた。

例えば、『太陽』原文には次のような例が認められる。

遊客の氣嫌を取るを職業とする (1901年8号 高山樗牛「文芸時評」P038A19)

日本の政府、日本の外交家に信用がないものとすれば、實に言語同斷の沙汰である。(1909年1号「大流小流」P014A26)

其潯熱に避易して (1895年1号 尾崎紅葉「取舵」P084B12)

これらは、現代語の規範から見れば、「機嫌」「言語道斷」「辟易」とあるべき表記であり、その規範形にあたる表記も、『太陽』において次のように認められる。

勇吉の機嫌をとるためでもあつた (1917年10号 久米正雄「感謝」P267A12)

濱松鐵工所を廢止せんとする如き言語道斷の沙汰なれど (1909年4号「彙報」P219A20)

外人の競争に辟易して逡巡するは (1895年10号「工業」P063A24)

『太陽コーパス』全体での、その用例数と規範から外れる表記の占める比率を示すと表5の通りである。

表5 通用表記と規範表記 (左：頻度, 右：比率)

	通用表記		規範表記		合計	
氣嫌／機嫌	5	2.7%	182	97.3%	187	100.0%
言語同斷／言語道斷	10	27.8%	26	72.2%	36	100.0%
避易／辟易	19	48.7%	20	51.3%	39	100.0%

いずれも、通用表記を単純な誤字とするには比率が高く、30%から50%にものぼるものもあり、こうしたものは、当時の通用として普通に用いられた表記であったと判断される。

こうした通用表記の例について再現性を重視する立場からすれば、原文表記のままで電子化を行うことが考えられ

る。しかし、検索性を重視する立場から見れば、こうした例を原文のまま電子化すると、現代語の規範を背景にもって検索語の表記を指定する利用者は、これらの例を発見することができない（検索もれを起こす）おそれがあり、こうした検索もれを防ぐ手立てを用意しておくことが望まれる。そこで『太陽コーパス』では、このような例は規範的な表記に修正して電子テキストに入力し、当該部分に「注」タグを付与し、「A誤字通用」の分類を与え、注タグの「原文」属性に原文の漢字を記入する方法をとることにした（注7）。次は、原文に「氣嫌」とある場合について、「機嫌」と修正し、注タグを付与して必要な情報を記入した例である。

<注 原文="氣嫌" 分類="A誤字通用">機嫌</注>

この「原文」属性の内容は全文検索システム『ひまわり』による検索結果のKWIC表示画面にも表示されるようにし、利用者が原文の表記を容易に知ることができるようにした。これによって、検索性と再現性とを両立させることができる。

ただし、濁音表記や仮名遣いなどのように規範が明らかなものとは異なり、漢字表記の選択における規範の範囲を判断することは、現実には難しい。便宜的な方法として、現代の書き言葉の規範が比較的広範囲にわたって採られていると見られる『広辞苑』（岩波書店）などの中型国語辞典を目安として規範の範囲を認定した。どのようなものを通用表記と認定したかについては、構造化テキストから情報を抽出するアプリケーション『プリズム』で注一覧を作成し、「A誤字通字」と類別が与えられたものの中に見ることができる。その方法については、本書の「構造化テキストを直接利用するアプリケーション—『プリズム』と『たんぽぽ』—」（小木曾智信）（83頁から）を参照してほしい。

8

コーパスにおける漢字処理

『太陽コーパス』の設計において、汎用性、再現性、検索性を確保することを基本方針として、漢字の電子化とその周辺において施した種々の工夫について述べてきた。結

果的に、検索性を確保するための方策を述べる部分が多くなった。コーパスがコンピューターによって管理された大量の電子テキストをもとに、言語研究の広範な目的のために的確な言葉の用例が自在に検索できるようにすることを目指して作られるものであることを考えれば、特に高い検索性を確保することの重要性は明らかであろう。

検索性を追求して、追加包摂規準の策定、代用、通用表記の修正などの処理を行うことにしたが、この方向は、漢字仮名選択の表記のゆれ、同訓異字選択のゆれ、送り仮名のゆれなど、さまざまにゆれる日本語の表記のゆれ全般について、検索性を向上させる手立てを講じる方向にもつながるものかもしれない（注8）。しかし、これらの表記のゆれは、現代語においても一般的なものであるので、コーパス利用者がゆれの範囲を想定することは比較的容易である。これに対して、本論文で扱ったような異体字や類似の漢字間での通用の現象は、現代語ではあまり目立たないものであり、利用者が見逃す可能性の高いものである。現代語の規範と『太陽』における規範とがずれる部分を中心に、現代人であるコーパス利用者が、『太陽コーパス』の検索に支障を生じないような、検索支援の方策を具体的に用意したわけである。

再現性についても、検索性を確保した上で外字タグや注タグを付与し情報を書き込むことによって、できるだけ原文の形が再現できるように努めた。汎用性については、現状におけるJISの優位性を確認し、JISを基本として独自の規準や処理方法を工夫したことを述べた。文字集合をめぐる状況は近い将来変わることも予想されるが、JISを基本としつつ独自の処理規準を明示的に設けておくことで、将来的な変化にも対応できるものと考えられる。

本論文では、『太陽コーパス』の設計における漢字処理の方法を具体的に記したが、ここで検討した問題は、日本語で書かれた文献を資料としてコーパスを設計する際の一般的な問題として普遍性をもつものと思われる。資料における漢字使用の実態と、コンピューターにおける文字集合の実態とを突き合わせるところに、豊かな漢字研究の領域も広がっていきそうである。

注

- (1) 本論文で述べる研究成果は、木村・田中・飯島・笹原 (1999) における研究に端を発している。この文献での調査は、『太陽』の1901年分12冊のみを対象にしたものであるが、本論文よりも詳細なデータを示している。また、『太陽コーパス』の漢字処理についての基本的な考え方は田中・小木曾 (2000) でも述べたことがある。
- (2) コンピューターによる日本語処理のための文字集合の状況を論じた研究として、最近の状況をまとめたものとしては、加藤 (2004)、笹原・横山・エリク (2003) (2004)。この問題は、1990年代に国語審議会 (第18～22期) でも取り上げられ、今後も重要な日本語問題のひとつとして、議論が続くものと考えられる。田中 (2001) では、国語審議会での議論を中心に、この問題を展望した。
- (3) 各規格の詳細は、JISについては日本規格協会 (1997) (2000) (2004)、ユニコードについては、トニー (2001) などを参照。
- (4) JISX0213は2000年に制定され、その追補の形で2004年に改正され10字追加された (日本規格協会 (2000) (2004))。
- (5) ユニコードによった場合、「青」と「青」, 「温」と「温」は、それぞれ別々の区点位置で区別されるため、『太陽』の原文に即して「青」「温」「温」と入力することができる。この点はユニコードの利点であるが、「羽」「羽」など、ユニコードでも包摂してしまう字体は多く、ユニコードによれば包摂がなくなるというわけではない。
- (6) 『今昔文字鏡』の有用性については、文字鏡研究会 (2002) で知ることができる。
- (7) 注タグの形式や位置づけについては、本書の「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」(田中牧郎, 38頁から) を参照。
- (8) この方向で検索性を高める具体的な方策として、山口・田中 (2002) では「表記辞書」の策定と利用を提案したことがある。表記辞書は実用段階にはいたっていないが、さらに研究を進めていきたい。

参考文献

- 加藤弘一 (2004) 「インターネットと漢字」 (『朝倉漢字講座5漢字の未来』, 34-77頁, 朝倉書店)
- 木村睦子・田中牧郎・飯島満・笹原宏之 (1999) 『『太陽』コーパス』の漢字処理—『太陽』1901の漢字調査— (科学研究費研究報告書・新プロ「日本語」, 国立国語研究所)
- 笹原宏之・横山詔一・エリク=ロング (2003) 「文字コードの現状と将来」 (『日本語学』22-5, 99-110頁, 明治書院)
- 笹原宏之・横山詔一・エリク=ロング (2004) 『現代日本の異体字—漢字環境学序説—』 (国立国語研究所プロジェクト選書2, 三省堂)
- 芝野耕司 (2002) 『増補改訂 JIS漢字字典』 (日本規格協会)
- 田中牧郎 (2001) 「情報化社会の言語政策」 (『現代日本語講座1言語情報』51-70頁, 明治書院)
- 田中牧郎・小木曾智信 「総合雑誌『太陽』の本文の様態と電子化テキスト」 (『日本語科学』8, 141-152頁, 国書刊行会)
- トニー=グラハム (2001) 『Unicode標準入門』 (翔泳社)
- 日本規格協会 (1997) 『7ビット及び8ビットの2バイト情報交換用符号化文字集合 JISX0208 1997』 (日本規格協会)
- 日本規格協会 (2000) 『7ビット及び8ビットの2バイト情報交換用符号化拡張漢字集合 JISX0213 2000』 (日本規格協会)
- 日本規格協会 (2004) 『7ビット及び8ビットの2バイト情報交換用符号化拡張漢字集合 (追補1) JISX0213 2004』 (日本規格協会)
- 文字鏡研究会 (2002) 『パソコン悠悠漢字術2002 今昔文字鏡徹底活用』 (紀伊國屋書店)
- 山口昌也・田中牧郎 (2002) 「言語研究のための構造化テキストと検索支援システム」 (『国語学会2002年度春季大会要旨集』, 165-172頁)

漢字文字列における字体の同化と衝突

—— 笹原 宏之

1

はじめに

明治期の活字で組まれた本文を電子化することは、すなわち電子的な翻刻を行うことであり、そこには解釈に基づいた本文校訂が加わらなくてはならない。校訂を行うためには、校訂の基本的な基準を設定し、個別にその時代の慣用と認められるものは何であり、誤りとみなすべきものは何であるかを把握しておく必要がある。

その基準となる概念の中に、字体に生じる「同化」と「衝突」というものがある（笹原1990・同1997）。『太陽』1901年のデータから「同化」とみられる文字列を抜き出して、その実態の記述を行う。さらにそれらを、他の字と「衝突」などの現象を起こしたか否かなどの点から分類を試み、同化の発生傾向を探索し、電子化の方法を考える。

2

同化について

雑誌『太陽』の原誌本文は、原則として原稿の手書き文字が、活字に置き換えられることによって形成されているものである。その手書き字体と活字字体とが異なる場合には、一般に既存の活字が植字されることで、活字字体で「正しく」印刷されるのが通例であるが、中には原稿の字体を忠実に、あるいは無批判にそのまま活字化しようとしたと推測されるものが含まれている。

その結果、筆者が誌面に出すことを意図しない単なる臨時的な書き誤りが、そのまま活字化されてしまったものもあると考えられる。つまり、電子化本文を作成する上で、下記の両者が混在している可能性があるのである。

・一般的なないし位相的な慣用が認められるものとして原文の

ママとしうるもの ・校訂の対象としうるもの

一般に、手書きに際して、字体Aが、近くにある別の字体Bに干渉して、字体Bを字体Aや字体 β へと変化させるという現象があることはすでに指摘されてきた（笹原1997・木村ほか1999）。ふだんは正確に書けている「干渉」を書こうとして「汗渉」と書いてしまったり、「衝動」と書こうとして「働動」、「講義」と知っていたとしても「講議」と書いてしまうことは、多くの人が経験するところであろう。これらには、他の語や字との混淆（「講義」と「会議」との混淆）や俗解に基づく再構成（「義」の意味の不確かさに比して「議」の意味が一見適合するよう感じられる）も介在しているようだが、混淆も想定しがたい例、姓の「鮎沢」を「鯨沢」と書くたぐいも散見される。

これは、音声・音韻レベルではつとに指摘され、熟語文字列においては文字・字体レベルでも断片的、個別的には古くから言及されてきたものである。ともに「同化」と呼ばれることがあり、一般に定着するケースもある。字体に関しては、「牽引」、「扁ぞろえ」などの用語も同義ないし類義で行われており、字体変化と異体字の定着の要因として重視すべき現象といえる。

文字・字体の同化に関して、用例を採集し、分析していくと、一口に同化といっても、実際には形態、頻度、別の字との関係、使用心理など各面において、さまざまなケースが存在していることがうかがえる。ここでは、文字全体が入れ替わる同化は除外し、字体レベルでの同化について考察を行う。

3 衝突について

手書きの際に発生した字体が、それまでに存在した字体とたまたま一致するケースを「衝突」と呼び、さらに字体だけでなく音義までも一致するケースを「暗合」と呼ぶ（笹原1990・同1997・「JIS X 0208-1997」）。衝突という概念は、元来、言語地理学において同音異義語の語形・音韻について用いられたもの（同音衝突）であり、それを文字の字体に適用したものである。日本における漢字使用の歴史は、字書上だけの静的な存在ではなく、きわめて流動的ないわば動態の中に把握すべきものであり、衝突や暗合という概念を用いることで初めて理解される事象が少

なくない。

4

同化・衝突の分類

4.1 『太陽』に存在する同化

表は、『太陽』1901年における同化の具体例をまとめたものである。形態・用法・使用原因という3つの側面から分類を行い、同化現象の発生傾向を探る。なお、複数出現したものについては、頻度を掲げる。

表 『太陽』における同化

<形態> <用法> <使用原因>

文字列	頻度	順逆	付加	遠隔	衝突	前例有無
輻輳 (←湊)	10回	順行	置換	連続	衝突	定着
饑渴 (←渴)	2回	順行	置換	連続	衝突	定着
味咻 (←酖)		順行	置換	連続	衝突	臨時
紡錘 (←錘)		順行	置換	連続	衝突	臨時
錯綜 (←綜)		順行	置換	連続	衝突	臨時
瑜珈 (←伽)		順行	置換	連続	衝突	臨時
研鑽 (←鑽)		順行	置換	連続	衝突	臨時
肯綮 (←綮)		順行	置換	連続	衝突	臨時
爛熳 (←漫)	9回	順行	置換	連続	新字	定着
人口に膾 (←膾)炙		順行	置換	遠隔	衝突	臨時
眼光の炯 (←炯)々たる		順行	置換	遠隔	新字	臨時
澆淳 (←季)		順行	付加	連続	衝突	臨時
萎靡 (←靡)		順行	付加	連続	衝突	臨時
潑刺 (←刺)		順行	付加	連続	新字	臨時
坭 (←泥) 坊		逆行	置換	連続	暗合	臨時
斟 (←斟) 酌	7回	逆行	置換	連続	衝突	定着
縉 (←搢) 紳	5回	逆行	置換	連続	衝突	定着
蠟 (←蠟) 燭	3回	逆行	置換	連続	衝突	定着
麴 (←麴) 包		逆行	置換	連続	衝突	定着
闕掖 (←掖) (狩衣の…)		逆行	置換	遠隔	衝突	臨時
模 (←模) 糊	8回	逆行	置換	連続	新字	定着

瞞 (← 爛) 眼		逆行	置換	連続	新字	臨時
娼 (← 保) 姆		逆行	置換	連続	新字	臨時
湮 (← 垂) 涎		逆行	付加	連続	暗合	臨時
蘆 (← 塵) 芥	5 回	逆行	付加	連続	衝突	定着
繻 (← 彌) 縫	10 回	逆行	付加	連続	新字	定着

まず、字や文字列の＜形態＞の面から見ていく。

「順逆」では、「順行同化」つまり前の文字の字体が後ろの文字の字体に影響を及ぼすものか、その逆の「逆行同化」であるのかを示す。つまり、前者は先に書いた字についてられて生じるものであり、後者は次に書く字に先走り、いわば先取りすることで生じるものである。

「付加」とは、同化によって字体が付加されることを指し、「置換」とは、同化の影響で本来の字体（字全体か構成要素などの部分字体）が別の字体に入れ替わることを指す。

「連続」、「遠隔」は、筆記の際に、同化の影響を与える文字とその受ける文字とが連続しているか、連続していないかを表すものである。厳密には、字の構成要素レベルで、同化部分が連続しているか否かも区別することが可能である。

次に、字の＜用法＞の面から見る。

同化によって字体が変化した字が、それまでに存在していた字の字体と一致する場合、それが「衝突」なのか「暗合」なのか、あるいは過去に存在したことが確認できず、同化などによって発生したと考えられる字体（ここではそれを「新字」とよぶ）なのかという点については、『大漢和辞典』などの辞書や過去からの文献資料に、そのような字体や音義・用法が掲載されていたり、使用されているかどうかによって判断する。衝突・暗合にも、一致した字が、一般的に使用・理解されていた場合と、辞書上のみ見出せる程度の場合（つまり非使用字・非理解字・死字）とがあることになる。

最後に、字・熟字の＜使用原因＞の面から見ておく。

同化を起こした字体が、当時、「定着」していたものか、「臨時」に発生しただけのものなのかについては、ここでは、同時期までの用例がほかに確認できるか否かによって判断した。

先の表に、「前例有無」としたもののの中には、筆記行動という立場から見ると、実際には、筆者が臨時に同化を起こしてしまったものが、過去や同時期の他者の書証と暗合したケースもある

と考えられる。例えば、「蠟（←蠟）燭」は、「蠟燭」と正確に書くようにして無意識に1字目に2字目の火偏を先取りして書いてしまった人が、複数存在する、という可能性がある。つまり、先人による「蠟燭」という書き方を理解字彙にもたない人でも、臨時的にそれを生み出すことがありうるのであるが、他の資料での当人による使用実態との比較照合も、それを探る手段として考えられる。これらの蓄積自体が、何が一般的であるのかを把握することにつながり、今後の研究の基礎となるはずである。

誤りが個別的に反復して発生する場合は、共通誤字の段階である。それに対して、誤りの用例が後に影響を与え、さらにそれが別の者に影響するというように、影響関係により継続的に使用されるようになっている状態をここでは「定着」と呼ぶ。

また、意識して元の字体に改造を加えたものか、無意識のうちに字体が変わってしまったものなのかについては、筆者本人によるメタレベルの言及がその前後にない限りは断定しがたい。

中には、単に定着していた字体を使用した場合のほか、熟語であることを明示したり、熟語の語義を鮮明にするためや、極端な場合には字面の美観を求めて、意識的に字体を揃えることもあった可能性がある（例：鳳凰、葡萄、蝌蚪）。

使用原因としては、書き手の意識だけでなく、編集者・校正者の意識も介在しているが、それらを探ることは容易ではない。さらに、文選工が活字を選び出す段階で発生する、誤解や目移りによる同化さえも考えられる。

上表を通して、一口に「同化」といっても、さまざまなパターンが内包されていることが明らかとなったが、おおまかな傾向としては、形態では「順行、置換、連続」型が多く、用法では「衝突」型、原因では、「臨時」型が優勢だといえ、無意識に発生しているものが多いとみられる。当然のことだが、遠隔と定着は共起しない。「鰯」という新字（「魚の魚卵は」9号P146B13、「鰻の鰻は」9号P147A01）もこの一種ととらえられる可能性がある。

4.2 『太陽』にみられる定着していた同化

以下、個々の例について、簡単に説明を加える。

まず、個々の同化形のうち、当時までに、「定着」していたと判断されるものについて、考察する。

【輻輳】

若し然らざる事ならば正に煩雑なる改良事件の輻輳したる際には放棄して可なり。儒教は神教に非ず、(1901年2号「国字改良論」久米邦武P019B10)

ほかに、小松崎吉雄、佐藤伝蔵、金子篤寿、肝付兼行(談)、近藤廉平(口述)などにも見られ、広まっていたことがうかがえる。

門内に馬車、腕車の輻湊し得るほどの玄関を備へ、内に毎月二三十圓の手當を受けて、(1901年3号「都会の誘惑物」麴町坊P204A11)

吉沼又右エ門及び松井菊野の被告事件等輻湊せるを以て伊庭の控訴公判は本年中には開廷の運びに至ざるべしと云ふ。(1901年14号「海内彙報」P225C06)

これは、上の例にある「輻湊」が元のものであるが、中国古典においても、この同化した熟字が見られる(『大漢和辞典』)。

【饑餧】

彼又思へらく宗教は恰も人が饑餧を感じて之を満さんとするが如く人の習慣より出で只フョールスピールvorspielたるに過ぎざるのみ、饑餧なくんば即何の要かあらんと、(1901年5号「宗教時評」龍山学人P049A04)

これは、近世期から見られる文字列であるが、「饑(飢) 饉」「饑餓」という熟語が混淆を促したとも考えられる。また、「餧」を単字で「うえる」に当てる用例も存しており、相互の影響を想定させる。同じ筆者が後の号では本来の文字列を用いている。

基督教信仰の何者たるやを認識し得られざりし彼等が、漸く其饑餧の苦境を過ぎて、稍平然たる時必然起るべき智識的信仰の疑問に想到せば、(1901年9号「宗教時評」龍山学人P049B17)

【爛熳】

天真爛熳なる人也、理想の高き人也、(1901年1号「文芸時評」大町桂月P034A12)

爛熳松竹古。山腰有危亭。(1901年12号「天平山遊記」結城蓄堂P139A14)

これは、本来の字体による「爛熳」は1回のみであり、かえて同化によって新たに生じた字体(新字)の方が頻度が高くなっている。当時の漢字字典にはほとんど載っていない字体であるが、

使用の定着度が高かったことがうかがえる。大町桂月は繰り返し使っているほか、佐藤伝蔵、横山健堂、嵯峨の屋おむろ、日下逸人も用いる。

「爛漫」は、偏がともに「シ」である「瀾漫」となることは多くなく、『太陽』1901年にも出現しなかった。一方、この字書にない「爛漫」は、漢籍では古くは北周の典拠があり（『漢語大字典』）、敦煌文書（ペリオ2292 張涌泉（1996）『敦煌俗字研究導論』60頁）や『全唐詩』、南宋金鈔本（清代 錢大昕『十駕齋養新錄』（1957版）巻14）などにしばしば見られる。

この定着には、一語であることの明示という意図のほかにも、「燦爛」などの既存の熟字の影響が考えられる。『太陽』で、漢詩文にも現れたのはその流れにとらえることができる。

朝鮮でも、李氏朝鮮の丁若鏞『雅言覚非』（金鍾権訳注）巻1, 3にも「爛漫」が「爛漫」となる例が挙げられ、「爛漫偽之為（誤作）爛漫（字典無爛字 古文或有瀾漫、無爛漫也）」、「若此類、又不可勝数」という。しかし、この字源や発生理由については述べられていない。『五洲衍文長箋散稿』巻11、『皇城新聞』1898年9月6日にも用いられ、『韓国固有漢字研究』にも言及がなされている。

日本でも、古くからこの文字列が見られる（『日本書紀』巻16 武烈天皇、津阪孝緯（1836刊）『夜航詩話』「此類甚多皆一時趁筆之誤後多沿其失而不攷耳」など）。北慎言『梅園日記』（1845年 百家説林・日本隨筆大成）巻5は、「俳諧」について、「上の俳字にうつりて」「俳諧」となったといい、「正字にはあらねども」「久しく書来りつれば」、「俳言俳文など」を除いて「故実にはかなふべき」という。この類として『続筆乗』と他の漢籍などから「爛漫」→「爛漫」など、多数の実例を加え傍証とする。

大槻文彦（1902）「文字の誤用」（『復軒雜纂』）277頁には、「摺紳輻湊なるべきを、縉紳、輻輳とするは、他の偏に釣り込まれての事であるが、正しく書く方がよい、爛漫を、爛漫など書くも同じ理窟であるが、最も悪るい」と評されている。大町桂月・佐伯常麿（1911）『誤用便覧』420頁も、「爛漫」→「爛漫」は「偏につり込まれてうつかり」と臨時的、無意識に、つまり個別的に生まれるが、定着して活字にもなっていることを指摘する。大町桂月（1918）『正しき用字法』260頁には、「爛漫」について、字書に見えない第二字は、「爛」字の「火」につりこまれて誤りしのみ」という。後藤朝太郎（1912）『明治の漢字』192～193頁では、「字畫の誤りは多く類推によりて起る。殊に二字結

合せられる時に起る」,「接近聯想によりて起る誤り」に「搢(搢紳)→「縉紳」,「輻湊」→「輻輳」などがあるという。後者は、211頁に「無意識にやつた誤謬」ともいう。205頁では、「規矩」→「規矩」のように義符にも起こることを指摘する。また、213～214頁には、「字が全く誤畫となることがある」とし、「前者が後者によりて類推せられること」と「後者が前者の爲めに類推せられること」、つまり順行同化と逆行同化とがあり、「俳諧」→「俳諧」,「蠟燭」→「燭燭」,「輻湊」→「輻輳」,「爛漫」→「爛漫」などのほか、二字熟字でない、本稿でいう「遠隔型」の、
團樂ノ樂→團樂ノ樂
清涼劑→清涼濟
という例も挙げる。

【斟酌】

米國杯の例に據れば經驗上大概の標準を得べきに依り、克く彼我の事情を核酌し確實に事を爲すに於ては、前述の困難は庶幾くは之を避くるを得ん。(1901年1号「商業世界」祖山鍾三・佐野善作P176B01)

「斟酌」は、「斟酌」「斟酌」「斟酌」「斟酌」といったさまざまな種類の同化が起こる熟字だが、ここでは「斟酌」のほかのパターンはない。岡田三面子は2号において、「斟酌」とともに使用し、後の号ではすべて「斟酌」のみとなっている。佐野善作も1号において使用しているものの後には「斟酌」となっている。ほかに、関根正直、水島鉄也、佐野善作、大倉喜八郎(談)の各人が使用する。大町桂月のように「斟酌」しか使わない者もある。

これは、時代を超えて、個別に散発的に生ずる誤用とも考えられる。頻出するが、後述する「蠟燭」の場合と同様に、語構成あるいは構成要素の変動が目立つためか誤用意識が強くもたれるため、定着とまではいえないようである。

「斟酌」の同化形には、以下の4種の組み合わせが実際に起きている。

- 斟酌 書陵部藏1528年の写本の写し『閑吟集』・『夜航詩話』4-3才
- 斟酌 『近世の古文書』20頁
- 斟酌 古文書古記録(鈴木眞喜男・大熊久子(1993)『有坂本和名集』59頁「開題」「扁ぞろえ」と呼ぶ)・『漢字の形音義』31頁・『明治前期家族法資料』2-2(1)58頁 1878年「御斟酌ノ上」

斟酌 『和字正濫鈔』『語学叢書 仮名文字遣』 解題5頁・
『色道大鏡』1811頁 (斟酌)

『武辺咄聞書』には「醇酌」も見える。それには、「斟酌」とも書かれており、臨時的かつ部分的な同化により生じた別字による表記かと思われる。

直接の同化ではない「斟量」(酌量)は、意識上の文字列による同化といえようか。これも、岡田三面子に見られた。

裁判官の見込を以ては如何なる場合にも加重するを許さざるが故に兇惡慘忍の徒にして確に刑の輕きに失するの止むを得ざる不都合を生じ、反對の方面に於ては出來得るだけの斟量減刑を與ふるも猶且つ酷に失し、求刑したる檢察事も宣告したる判事も心竊に被告を憫み、(1901年1号「法律時評」岡田三面子P052A08)

こうした表記は、原稿段階では、「詐欺師」(夏目漱石『坊っちゃん』)のようなものが生じていても、活字化される際に直されるのが通例である。しかし、先に、固定化した例を掲げたが、その一因として、活字などによる印刷物・刊行物に使用されることが挙げられよう。坪内逍遙(1923)『所謂漢字節減案の分析的批判』(『早稲田文学』1923年8号12頁)、『誤用便覧』420頁にも、「斟酌」について言及がある。

【縉紳】

古蹟取調會 九條公其他朝野縉紳の設立也調査目的は皇祖神蹟、皇宮舊蹟、(1901年01号「海内彙報」P228C27)

『百敷の大宮人はいとまあれや』の歌は當時縉紳の生活を唱へるもの、華山一條以降、(1901年2号「王朝の絵画」高山樗牛PI07A25)

中国古典においても、同化した熟字が見られる(『大漢和辞典』。前述参照)。本来の形は現れなかった。

【蠟燭】

甚だ簡單にして、十六蠟燭之火力と全等なる光を、綠色硝子を通じて、(1901年9号「世界紀聞」P213B14)

ほかに、嵯峨の屋おむろが用いている。

「蠟燭」(「蠟」は原文では、字体を「蠟」に作る)には、他の文献においては、順行、逆行による2種の同化形のほか、転倒形さえも現れる。そのうち、「蠟燭」となる例は、広大本『和名集』『北野社家日記』『初心要抄』黒本本『節用集』『有坂本和名集』

「開題」「札記」)のほか、西鶴作品、『明治の漢字』に見られる。「蠟燭→蠟燭」も『有坂本和名集』のほか、内閣文庫蔵江戸初期写『庭訓往来』、弘治二年本・天正十八年本『節用集』(同)などに見られる。『舜旧記』には、「蠟燭→蠟燭」もある(『有坂本和名集』178頁参照)。

【麵包】

是れ皆歐發和傳の西唱東隨法であつて主食品の米たと麥たと薯たとと麵包たるとに關はらず只肉食と蔬食との善惡を解説するは恰も下女下男が御膳の菜で兎や角云ふと同じく言はゞ必竟御膳の御菜論に過ぎないのでチト賤しく聞える所がある孟子が(1901年10号「食物養生法に就て」石塚左玄(談) P152A13)

この文字列は、『日本国語大辞典』『大漢和辞典』『漢語大詞典』にともに見られない。しかし、1895年の『太陽』にもすでに2回現れている。「パン」を漢語に訳した「麴包」ないし「麵包」は、各辞典に古い出典を欠くが、この同化形とみられる。あるいは、「パン」の当て字に「麴麴」「麴麴」という、2字めに「麴」を当てる、これも同化形とみられる表記が存在するため、それとの混淆という可能性もある。また、「麴」という字体が「麴」とやや類似していること、「パン」と「麴」という意味や「メン」という音との関連についての意識の薄さに比して「包」という字の音義との類似性も、この文字列を生み出す素地となっている可能性がある。

【模糊】

朝來大霧、遠山近水すべて模糊の中にあり、六里の新道、(1901年4号「鎮西遊記(接前号)」久保天随P130B03)

ほかに、大町桂月、室田義文、下岡蓮杖(談)、上野英三郎の文章にも見られる。本来の「模糊」が1例も現れず、『太陽』1895年に、熟語だが「糊」を含まない「模視」まであることから、この熟語に基づく同化字体の定着の実態がうかがえる。これは、部首レベルでは「木」→「米」という置換と考えられる。ただし、手書きではより細かな部分が同化するケースもあり、「木」に「ソ」が足された付加とも解釈しうる。

「模糊」の第1字は、古い辞書にないが、やはり中国で現れたものである。『青箱雜記』(叢書集成)に見える。「爛漫」も「模糊」もいわゆる疊韻の語であり、2字で一まとまりの語として音

と意味をもつものと意識され、語構成や単字の字義に関する意識が薄かったのであろう。また、疊韻の語には「混沌」のように偏が揃ったものがあり、揃えようとする意識も働いた可能性がある。

先に引いた朝鮮の『雅言覚非』には「模糊」が「模糊」となる例も「模糊譌（誤）作模糊（字典無模字）」と挙げられており、また先に引いた『梅園日記』や『正しき用字法』にも「模糊」が挙げられていた。そのほか関子徳・関世道『行書類纂』（1829刊 早稲田大学図書館蔵版本・異体字研究資料集成影印）已集挙要にも、「模糊」に関連して、「模」は「字書不載」、「蘭亭叙」から「俚情」が「倦情」となった例などを挙げて、「上取下也模糊之模本模字盖亦此例」という。尾崎紅葉『多情多恨』にも「模糊」が用いられている。

さらに、後藤朝太郎（1909）『漢字音之系統』・『線音双引 漢和大辞典』2266頁・『同文新字典』のように、「模」一字で「ミダリ」と読む字だとする辞書が現れた。

【塵芥】

○市街塵芥の肥價 / 瑞典國ストックホルム市に於ける塵芥の性質に関する研究、

ローザン氏によりて行はれたり、氏は各國に於ける市街塵芥の性質及肥價を調査せる頗る密、而して何れも大差なきことを認めたり、而して何れも大差なきことを認めたり、ストックホルム市塵芥の九十一％は馬肥、木灰、鐵片（○、四％）等より成るを檢せり、而して市街塵芥の肥價を決定せんが爲めに、各種厩肥と肥料上の効力を比較せり（1901年2号「農業世界」上野英三郎P185A19～P185B06）

これは、「塵」が崩し字を介してこのように変わる現象が各種の文字資料でそれぞれ別個に見られるため、これは崩し字が関与したものである可能性が高い。例えば、『宝暦新撰早引節用集』に「塵芥」（じんがい・ちりあくた）に草冠があるように見える崩し字がある。これは、「鹿」の崩し字が元々そうなっているためである。

この文章にしか現れず、しかも途中から「塵芥」となっている。これは、原稿では草冠があるように見えたものでも、活字が不足して代替されたものとも考えられる。

但し厩肥一立方米は六百八十七基の重を有し、塵芥同量は四百七十一基の重を有せり、此成分上より一立方米の肥價を瑞典に於て算するに、八三弗、塵芥一、一七弗なりと云ふ、一

七弗なりと云ふ、更に栽培試験によりて同積の厩肥及塵芥を給して其効果を檢したるに、(1901年2号「農業世界」上野英三郎,P185B10~P185B14)

【繻縫】

一時を繻縫することの度重なるに及びては、如何に幼稚なる社會と雖も永く是に勝ふべくもあらず。(1901年1号「文明批評家としての文学者(本邦文壇の側面評)」高山樗牛P022B10)

ほかに、添田寿一、加藤政之助、国府犀東、坪谷水哉、小松崎吉雄、斎庭、記者(抄録)、金子堅太郎(談)の文章に用いられている。これも伝統的な漢字字典にない文字、つまり新字を含むものである。

1878年『万国史字解』に「繻縫シテ」(山田忠雄(1981)『近代国語辞書の歩み』上148頁「上字は正しくは彌、下字に牽引されて糸篇を増画した臨時の造字である」という)、1895(明治28)年脱稿陸奥宗光『蹇蹇録』(岩波文庫)にも「繻縫の策」が現れる。なお、単字で音仮名として使用された例が『日本書紀』に見えるという(大塚毅(1978)『万葉仮名音韻字典』など)。

4.3 『太陽』にみられる臨時的な同化

以下は、当時定着したいと判断しうる根拠に乏しいものの例である。

【味淋】

下迄柔かくさし得る様になれば、砂糖のあくひき(むき栗一升に砂糖三百匁味淋三合入れ)一寸煮たて、金とんのころもには、(1901年12号「精進料理」赤堀峯吉P153 A17)

なお、現在一般化している「味醂」は、用いられておらず、ほかには当時しばしば見られた「味淋」(大町桂月ほか)のみである。

【紡緋】

頻出する「紡績」との混淆も起こった結果である可能性がある。本来の「紡緋」は、出現しなかった。

錢石等の石灰岩を出すが、鮫石とは有孔蟲の一種紡緋虫Artesianの化石にして、錢石は海百合の化石なり、(1901年2号「地殻中の産物(承前)」佐藤伝蔵P200B22)

【錯綜】

其貨物を自在に此地方に入れしめば、列國の利益は互に錯綜して、自ら殺伐の氣象を銷滅し、(1901年5号「満洲問題」島田三郎(談) P076B07)

「錯綜」(龍山学人ほか)が元になっている。

【瑜珈】

昔者印度に瑜珈と稱する苦行の學徒ありき。彼等の爲すところは實に今の人を戰慄せしむるに足るものなりき。(1901年9号「文芸時評」高山樗牛 P038A03)

これは、2字めの活字が存していたならば、あるいは印刷段階での誤植であろうか。この表記は、『漢語大詞典』『大漢和辞典』『日本国語大辞典』に見られない。

【研鑽】

この粗朴なる屋内にてアベ教授は多年光學數學社會經濟學及び新工夫につき辛苦研鑽するを以てその常務とし孜孜として老の既に至れるをも知らず傍彼が創設せる事業の日を逐ふて盛大なるを欣びつゝ静和なる生活を過せり(1901年10号「世界唯一の模範工場」南船北馬樓主人 P172A04)

「研鑽」(研鑽)の同化した形である。同じ文章で、直前では「學理の蘊奥を研鑽し」となっている点から、臨時的な誤記かと考えられる。

【肯綮】

彼の國現下の國力を論ずる、吾人は頗る肯綮を得たるものなるを信ず。讀者は之を見て如何に感ずるか。(1901年10号「經濟時評」坪谷水哉 P067B07)

「肯綮」が変化したものである。なお、「肯綮」とする箇所(1901年07号「大江広元論(接前号)」笹川臨風 P109A07)は、誤植であろう。

其の場合には言へば必ず肯綮に中り、一擧して一世を警醒する様にせねばならぬ、(1901年1号「社会の腐敗救済意見(其一)」岡部長職 P074A03)

【人口に膾炙】

千八百四十三年は實に天が博士と彼れロースの名と共に人口に膾炙せるギルバート博士とを結合せし年にして、(1901年

05号「農業世界」上野英三郎P171A05)

これは、「膾」が2字前にある「口」に干渉を受けたものであろう。「灸」は誤植である。

【咽々】

後述する。

【澆淳】

法の欠缺を恃んで監督權に抵抗するが如きもの滔々たる澆淳の世界其實例に乏からず風教徳化と相待たざる自治制度の運用豈其骨髓を得たるものと云ふを得べけんや。(1901年13号「『政党及議院政治の弊』に対する所感」清野長太郎P025A08)

元は「澆季」である。以下のほか、3例ある。

見よ世は次第に澆季に流れ、人は漸く猾智に走り、(1901年5号「北海道に於ける無言の行者(接前号)」金山尚志P204A17)

【萎靡】

人心萎靡の社會に、剛健の著作出でざるが如し、(1901年5号「太陽」P001C06)

これは、次にあるような「萎靡」に、同化が発生したものである。

苟も前途に快心の望あらば、人間の意氣決して萎靡せざるなり、(1901年3号「憲政の一大危機」P001D16)

この形は、『大漢和辞典』(「委靡」はあり)『漢語大詞典』『日本国語大辞典』に見られない。「委靡」のように、草冠がないことで揃ったものもあることが、揃えるという意識を生み、草冠で揃えられたことの原因となったとも考えられる。

次も、辞典に見えないが、「萎微」にこれと類似する同化が起こったものであろう。

通商貿易の隆盛得て望むべからず、國家は悉く萎微不振の苦境を脱すること能ざる可し。商業道德の重んぜざるべからざるや夫れ斯くの如し、(1901年7号「商業世界」祖山鍾三・佐野善作P194A08)

【潑刺】

遠く彼より吾が國民の氣慨を造り出し、活動潑刺として今や

東洋大陸の大勢力を支配しまして、漸く其光氣を發揚せんとする時期に到達したのであります、(1901年13号「人種強健策」井上豊太郎P202A25)

これは、後にJIS第2水準に採用されるほど出現する「濶」の古例である。発生の新しい「濶濶」の「濶」は、ここでは旁を「刺」ではなく、誤って「刺」に作る。

頭白くして雪の如し。左手に濶刺たる鯉を握り、右手に快刀をとりて、(1901年8号「利根川の一日」大町桂月P110A06)

この「濶刺」は、疊韻の語であり、かつ「活濶」との混淆によるものであろう。1905年蒲原有明『春鳥集』9頁自序などに見え、1894年の築地5号活字に見られるが、これより古い使用例が見当たらないため、仮に臨時としておく。

【堀坊】

後述する。

【闕被（狩衣の…）】

其の證據には、手向山の競馬乗の闕被（狩衣の如き衣）の上
に綾頭を着て居る、これが矢張り競馬乗に鎧を着た態である、
(1901年13号「甲冑の話」森大狂（記）・久保田米僊（談）
P141A07)

これは、あるいは「腋」に基づく意図的な会意化であろうか。

【睨眼】【眼光の睨々たる】

「炯」の異体字「炯」には、逆行同化と順行同化の両方の結果として現れており、新字を形成している。執筆者は金山尚志と高山樗牛で、別である。「睨々」は、意識的に会意化した、意味が関与する造字か。「眼光炯々として」は佐々友房が用いている(13号)。

あるいは、当時、活字ケースに準備されていたものが反映したものか。見出し用の活字を並べた『明朝体活字字形一覧』にはない。

茲に清新の氣運を喚起せんとする時代也。若し睨眼なる史家の文章を能くするものあらば、是の時代の如きは眞に好箇の題目たらずんばあらず。(1901年8号「文芸時評」高山樗牛P037A05)

一種犯すべからざる氣風を具へ、眼光の睨々たるは、彼女が意志の如何に頑強なるかを現はせり、(1901年7号「北海道

【娼婦】

後述する。

【垂涎】【坭坊】

これらは、同化後の字体が、それぞれ同化前の字と同じく「つば」や「どろ」という意味も有しており、暗合であるのか、意識的に字を改めたのが不明である。「垂涎」は4号などに見える。「坭坊」(14号に見られる表記)には、また岡田三面子がかかわっている。

あゝビザンツに住する露國の民が讚美し歌を聞きたるはそも幾年月ぞ、あゝ限りなき神の恵を讚美してコンスタンチノール街頭に巍然たるソフヒイン寺院の占領に垂涎したること夫れ幾年ぞや、一世紀は殆んど已に過ぎ去りて彼等の夢は未だ覺めざるなりし、(1901年14号「露國の宮廷（承前）」日下逸人（訳）P134A04)

一見同一の實害を醸しながら甲は軽く乙は重く不正不條理なる皮想ありと雖も憫むべき貧の盜もあるべし惡むべき坭坊營業者もあるべし客觀的の實害のみを顧慮せず主觀的本人の危険をも併せ察せざる可らず（1901年2号「刑法改正非改正」岡田三面子P036A12)

ほかにも、こうした同化のたぐいである可能性のあるものを挙げると【蹠蹠】は、「涉獵（躋）」との混淆か。【涉蹠】は、『大漢和辞典』『漢語大詞典』にない。

苦役、蹠蹠を憚らず。反之新世界の自由、(1901年01号「政治時評」国府犀東P045A02)

【粗糙】

『大漢和辞典』に出典を欠くが、ソサウと読み、きめの粗いこととある。「糙」という字が、きめが粗いという意味をもつ。【粗造】も『大漢和辞典』では出典がなく、ソザウと読み、雑なこととする。本文では、後者の意味のようで、同化の可能性もあるが、今しばらくここに記すにとどめる。『漢語大詞典』には、前者に後者（立項せず）の意味（粗笨拙劣、粗暴魯莽）も認めている。

而て鐵板は前の一工程に於て粗糙に機械の作用を受くることは已に記せる如くなるが、之を爲すに諸器具の作用は獨り炭化面の顯はるゝに至る點までにて、(1901年2号「工業世界」

金子篤寿P181A24)

4.4 『太陽』にみられる同化の暗合

同化には、他に同じ用例があっても、上にも触れた通り、たまたま同じ現象が臨時的に発生したもの、すなわち暗合という可能性がある。例えば、「保母」に関して、次の例がある。この「嫗」は、かつて辞書に掲載されなかった漢字である。『漢語大詞典』が引く柳宗元の「嫗傳」（保傳）のような唐代中国の話ではなくドイツの話であり、少なくともそれを直接踏まえた表記ではない。

殊に朝御顔を洗ひ給ふを厭ひて、日々御部屋にて嫗母なる宮・女と争ひ、(1901年7号「世界紀聞」,P214B03)

これから半世紀ほど後の1956年の雑誌九十種調査（国立国語研究所）においては、置換でなく付加型の「嫗母」という形態が出現している。

思えば……幼い由夫のことは、あの若い、美しい嫗母、川村千代子に預かって貰って、どうやら心配なくなったものの、しかし、もっと大きな心配は、あるのだった。（『明星』1956年4月号P115 北條誠「この世の花」）母、「保母」という語の表記に、これらの暗合に類する現象が発生する原因は、両字の熟合度によるだけでなく、「保母」という表記も影響していると思われる。

家族と共に舍内に居住し、且つ保母を聘用し、父兄の依頼に應じ、(1901年07号「教育時評」大町桂月P058B21)

なお、同じく雑誌九十種には、「噴煙」の同化形である「煩煙」という熟字も見られる。こうした同化は、手書きにおいて発生しやすい現象であり、手書きの原稿の通りに活字を拾ったり、作つたためと思われる（注1）。

同化という現象は、仔細に観察すれば、一つの字（単字）の内・部でも発生することがあるほか、仮名文字列などにも発見される。さらに、まれに、熟語においてそれぞれの文字や文字の構成要素が入れ替わるという「転倒」も観察される（注2）。

5

おわりに

ここまで述べてきたように、文字列における同化によるとら

えられた字の中には、「臨時」的な字体であり、かつ意識的とも解しがたいもの、つまり筆者の不注意による誤用、誤字ないし誤字体で、そのままの活字化を望んでいなかった可能性が高いものがあつた。ここでいう誤用とは、その字体には本来その意味がないにもかかわらず、字体変化によって新たな意味が生じた衝突のことである。

これは、明治期という時代や、月刊雑誌という執筆者や口述者自身が校正をじっくりと行えたとは考えにくいメディアの制約によるところが大きい。それらについては、原文を尊重する立場よりも、文の読解や語の使用実態の記録などを主目的とするコーパスの電子化本文を作成するのであれば、原文のママとしておいたり、あるものは「𠄎」（ゲタ）として処理しておいたりするのではなく、本文校訂の対象として、本来の文字に置き換えて本文を作成し、表記レベルのタグに原文の字体や同化である旨を注記をするという対応でも差し支えないと考えられる。

注

- (1) 近年、DTP印刷においては、オペレーターが手書き原稿の字を書かれたとおりに安易に作字するという指摘が少なくないが、国立国語研究所が行った調査の対象となった1994年の雑誌七十種では、ワープロ入稿が増えたことが強く影響しているであろう、こうした臨時的な同化はほとんど現れなくなっている。
- (2) 「土碓」という例は、「同化」と対比しうる「異化」の例であろうか。『漢語大詞典』『大漢和辞典』に、この熟字はない。あるいは、意味によって字体が変わったものか。問題が多岐にわたるので、このたぐいについては別に扱うこととする。

これは山脚崩れて、土碓は谷間に落ち、纔に巨岩を積み残したばかりの處で、(1901年14号「日光山の奥」長谷川天溪 P141B14)

「土塊」は、1901年01号「燕歌鴻吟（第一集）」巖谷小波 P114B06、1901年03号「世界紀聞」P211B01に用いられている。

参考文献

- 木村睦子・田中牧郎・飯島満・笹原宏之 (1999) 『『太陽』コーパスの漢字処理—『太陽』1901の漢字調査—』(科学研究費報告書・新プロ「日本語」、国立国語研究所)
- 笹原宏之 (1990) 「国字と位相—江戸時代以降の例に見る「個人

文字」の、「位相文字」、「狭義の国字」への展開一」（『国語学』163号,14-26頁,武蔵野書院）

笹原宏之（1997）「字体に生じる偶然の一致—「JIS X 0208」と他文献における字体の「暗合」と「衝突」—」（国立国語研究所『日本語科学』創刊号，7-23頁，国書刊行会）

異体仮名について

——中川 美和

1

はじめに

雑誌『太陽』の仮名（平仮名・片仮名）は、現代の雑誌とそうかわらないようにみえる。

確かに、『太陽』の漢字が、膨大な数の異字体を含むのに比べて、平仮名や片仮名の異字体は少数である。『太陽』の仮名は問題ないのだろうか。

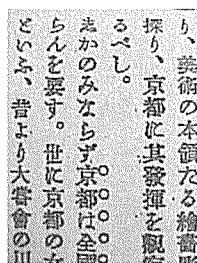


図1 「しかのみならず」（1895年4号「京都は國美の庫なるを論ず」久米邦武P012A15 図中4行目）

『太陽』創刊時とほぼ同じ時期の明治33（1900）年、「文部省令第14号小学校令施行規則中教授用仮名字体に関する規定」（以下、「小学校令」と略す）が施行されている。ここをもって、異体仮名は廃止され、現在私たちが目にするのできる五十音図のような、一文字が一音節に対応する仮名字表に統一された。このような時代の流れと、『太陽』が無関係であったとは思われない。

しかしながら、当時の仮名字体のありようが、「小学校令」発行日を境に一斉に切り替わったとも考えられない。『太陽』の仮名を仔細に観察してみると、異体仮名のなごりともいえるべき、字体の異なる仮名がいくつかみられる。実際の活字媒体においては、

仮名字体の統一は徐々に行われたと推定するほうが自然であろう。

「小学校令」の意味するところは何だろうか。それは、複数の仮名字体が同じひとつの音節に対応する、という非合理的な書記のしくみを廃止し、合理的な一文字一音節の書記システムに変更する、というものであったのではないか。そうだとすれば、たとえば、〔し〕の音節に相当する仮名が〈し〉と〈志〉の二種類あり、その使い分けが何らかの役割を果たしていたそれまでの時代と比べて、『太陽』の時代には、異なる字体で同じ音節を示す仮名の使い分けが、意味をもたなくなったのだと推定できる(注1)。

先にも述べたように、『太陽』の仮名の異字体は数が少ない。そのことは、仮名の字体が統一される末期の状況を『太陽』が反映していることを予測させる。異体仮名の統合は、実際の活字媒体においては、どのように行われ、またどのような使い分けが最後まで残ったのか。『太陽』の仮名字体の観察を行い、その位置づけを行うことによって、異体仮名が消滅する過程の側面をも捉えることができるのではないか。本稿はこのようなもくろみによって書き起こされた。

調査対象資料について述べておく。『太陽』では、異体仮名の数は、年を経るに従って減っていく。そこで、異体仮名が比較的多くみられる、創刊年1895(明治28)年1号から12号を中心に調査を行った。また、具体的な調査方法は次の通り。異体仮名の抽出は手作業で行った。『太陽コーパス』(2003年3月段階)を利用した。集計にはmozil.awk(中野洋作成)等を用いた。

なお、仮名の問題は、字体ばかりではない。仮名遣いの問題(注2)や、仮名の活字の字形(注3)の問題もある。このような問題も含めて、全体で『太陽』の仮名の問題といえるわけだが、ここでは、『太陽』の仮名字体の問題に絞って報告を行いたい。

2

創刊年『太陽』の異体仮名の概観

創刊年『太陽』の平仮名および片仮名の異体仮名を、表1として示す。

実際に『太陽』創刊年に使用されている仮名字体をあげた。空欄になっている仮名字体には、たまたま創刊年に出ていない可能性も含まれている。

仮名は五十音順に並べた。また表中〈〉で括ったものは、現在目にする五十音図と字体の異なるものを指す。〈〉内に複数の異字体がある場合はそれらの異字体が共存していることを表す。

全体的に見て、異字体の種類が多くないことがわかる。平仮名では〈し〉〈志〉の使い分けが見られるほか、や行の〈江〉が一部用いられている。また、片仮名では〈ネ〉のかわりに〈子〉を使用していること、ワ行の〈ヰ〉のかわりに〈井〉が用いられることがわかる。

なお、※は五十音図には通常は含まれないが、実際に『太陽』創刊年に使用されている字体である。

表1 『太陽』創刊年平仮名片仮名の字種
(明治28年(1895年)1号～12号)

平仮名	片仮名
あいうえお	アイウエオ
かきくけこ	カキクケコ
さ〈志し〉すせそ	サシスセソ
たちつてと	タチツテト
なにぬねの	ナニヌ〈子〉ノ
はひふへほ	ハヒフヘホ
まみむめも	マミムメモ
やゆ〈(江)〉よ	ヤ ユ ヨ
らりるれろ	ラリルレロ
わゐ ゑを	ワ〈井〉 エヲ
ん	ン
がぎぐげご	ガギグゲゴ
ざじずぜぞ	ザジズゼゾ
だぢづでど	ダヂヅデド
ばびぶべぼ	バビブベボ
ぱ ぽ	パピプペポ

※アイエ

ツ

ヴ

ケ

次に、『太陽』創刊年の異体仮名の状況と、同時代に施行され

た「小学校令」の仮名表を比較してみよう。

「小学校令」は片仮名のワ行の〈ヰ〉のかわりに〈井〉を採用している点をのぞいては、現在の五十音図とほぼ同じであることがわかる。仮名字体は一音節一文字に統一されている。

表2 文部省令第一四号小学校令施行規則中教授用仮名字体に関する規定（官報5141号 明治33年8月21日）
（国語教育研究会編1969『国語国字教育史料総覧』風間書房p787）

第一号表（〈〉は稿者による）

平仮名	片仮名
あいうえお	アイウエオ
かきくけこ	カキクケコ
さしすせそ	サシスセソ
たちつと	タチツテト
なにぬねの	ナニヌネノ
はひふへほ	ハヒフヘホ
まみむめも	マミムメモ
やいゆえよ	ヤイユエヨ
らりるれろ	ラリルレロ
わゐうゑを	ワ〈井〉ウエヲ
ん	ン
がぎぐげご	ガギグゲゴ
ざじずぜぞ	ザジズゼゾ
だぢづでど	ダヂヅデド
ばびぶべぼ	バビブベボ
ぱぴぷぺぽ	パピプペポ

3

使い分けのある異体仮名

創刊年『太陽』に見られる異体仮名は少数である、と述べたが、少数ながらも異体仮名を残すそれなりの理由や背景があったことがうかがわれる。

本文中で、複数の異体仮名を使い分けているといえるのは、〔し〕の音を表す〈し〉と〈志〉である。創刊年『太陽』には、

全部あわせて282例の〈志〉の仮名字体が見られる。ただし、満遍なく見られるわけではなく、ある記事に偏っていることがわかる。

たとえば、『太陽』1895年2号の〈志〉の例は、坪内逍遙・落合直文・川崎千虎・大橋乙羽らの記事に集中している。以下に全例を挙げる（以下用例の踊字は「々々」に改めた）。

〈志〉か名け得べし。（「戦争と文学」坪内逍遙P011A05）

よし〈志〉からざるも、（「戦争と文学」坪内逍遙P012B05）

〈志〉ばらくは此れが爲に影を藏さん、（「戦争と文学」坪内逍遙P012B10）

グリフィウス等の愛國的詩歌は〈志〉ばらく措き、（「戦争と文学」坪内逍遙P013A05）

太古激闘時代の吟客は〈志〉ばらく措き、（「戦争と文学」坪内逍遙P014B07）

〈志〉ば々々愛國の熱誠を吐露せし（「戦争と文学」坪内逍遙P015A19）

豫言せしにも〈志〉るけし。（「戦争と文学」坪内逍遙P015A21）

〈志〉かりと雖も、（「戦争と文学」坪内逍遙P015B09）

〈志〉ばらく大戦敗後の文學は之れを別問題として取除き、（「戦争と文学」坪内逍遙P015B13）

〈志〉ば々々之れを證す（「戦争と文学」坪内逍遙P016B01）

〈志〉ら雪物語（「しら雪物語」落合直文P107B23）

〈志〉ばしにらみあふほどに、（「しら雪物語」落合直文P111A11）

〈志〉たりがほなるものゝ段に（「小弓の勝負」川崎千虎P138B18）

〈志〉はぶきをしまざらはして（「小弓の勝負」川崎千虎P138B19）

〈志〉やうじをかなたこなた（「小弓の勝負」川崎千虎P140A16）

盆に〈志〉めを結ぶ中に（「新年の禮式（承前）」有住齊P142B22）

貌姑射山人の「〈志〉のぶの露」（懸賞歴史小説）あり、（「明治二十七年の文学界」著者不明P164B11）

不知彼の〈志〉がらみ草紙再誕の曉、（「明治二十七年の文学界」著者不明P165B35）

〈志〉かし今の世のありさまにくらぶれば、（「元禄風俗と英

一蝶」大橋乙羽P176A04)

〈志〉のゝめの唄に（「元禄風俗と英一蝶」大橋乙羽P177A16)

「〈志〉のゝめ々々いつ夜が明けた、（「元禄風俗と英一蝶」大橋乙羽P177A16)

外面に〈志〉げき市人と（「元禄風俗と英一蝶」大橋乙羽P177A18)

このように、ある特定の筆者の記事に集中しているということは、元原稿にあった〈志〉の字体をそのまま反映した可能性が高い。

また、これらの記事においては、〈志〉は語頭、〈し〉は非語頭、という使い分けがなされている。このような語頭〈志〉・非語頭〈し〉の対立は、他の記事にもみられる。

尾崎紅葉「取舵」「〈志〉なしたり！」（1895年1号P089A05)

饗庭篁村「従軍人夫」「〈志〉ばし」（1895年1号P094B14)

無名齋「おもしろい」「〈志〉くじった」（1895年1号P116B15)

音羽庵主人「正月」「〈志〉たしむ」（1895年1号P154A15)

あるいは、〈志〉は形態素の切れ目に用いられている、といったほうがよいかもしれない。

饗庭篁村「従軍人夫」「立たッ〈志〉やれ」（1895年1号P092A25)

むろん、語頭〈志〉・非語頭〈し〉の対立がみられるのは、〈志〉の仮名字体がみられる記事においてだけである。〈志〉の仮名字体をそもそも使用しない記事においては、語頭でも〈し〉の仮名字体を使用している。

『太陽』において、複数の異体仮名を本文で使用しているといえるのは、〈し〉〈志〉においてだけであるといつてよい。なぜ〈し〉〈志〉の使い分けだけが残っているのであろうか。また、語頭が〈志〉、非語頭が〈し〉という形で使い分けられているのは、根拠があるのだろうか。

同時代の他の資料と比較してみると、たとえば、矢田（1996:440頁）では、「小学校令」以降も、『増補俚言集覧』（明治33年刊）、『類聚名物考』（明治36～8年刊）のような辞書類で〈志〉〈し〉の仮名字体が使い分けられ続けたことが指摘されている。

こうした当時の状況を反映したものであろうか、明治33年1月には、次のような決議がとられている。（安田章（1980）がすで

に指摘するとおり)。

明治三十三年一月帝国教育会仮名調査委員 国語国字に関する決議

(小数説)「志」を存して「し」の方を一語の中又は末に用ゐる仮名とすること(明治三十三年十月十五日「教育公報」第240号所載)

このように〈志〉の仮名を残すとすると、仮名字体の統一に反することになるはずである。にもかかわらず、「小学校令」の施行の前に、「小数説」とはいえ、〈志〉と〈し〉の複数の仮名を併存させる説が出たということは、当時の異体仮名の使用状況に、語頭の〔し〕には〈志〉の仮名字体を用いる、という使い分けが根強く残っていたことを反映したものと考えられる。

雑誌以外では、三好(1977)が指摘するように、『若菜集』(島崎藤村・明治30年)にも同様の語頭の〈志〉がみえる。用例は次の通り。

志る・志らべ(34頁), 志れず・志らぬ(51頁), 志るせし(67頁), 志ろき(70頁), 志づみし(75頁), 志は々々(79頁), 志ろし(80頁), 志づかなる(80頁), 志づかに(86頁), 志かじ(86頁), 志れかし(86頁), 志たへる(親へる)(92頁), 志ぐれ(95頁), 志らず(106頁), 志るす(119頁), 志げくして(124頁)

このような語頭の〈志〉は、さかのぼっては、『当世書生気質』(注4)にみえ、また『言海』『日本大辞書』も語頭の〈志〉を使用している(注5)。

『太陽』において〈志〉の仮名字体が使用されているのは、このような伝統的な、慣習的な異体仮名の使い分けを残している筆者の元原稿を反映したものであると見てよいだろう。

さて、それでは、なぜ〔し〕のみが複数の異体仮名を用いているのだろうか。この問題は『太陽』のみにとどまらない。他の資料においても〈志〉は根強く残り続ける。

たとえば他の仮名,〔か〕は〈加〉(〈可〉ではなく),〔た〕は〈太〉(〈多〉ではなく)など、語頭多用字を採用している。それに対して〔し〕の語頭多用字である〈志〉を採用せず、〈し〉を採用するのでは、他の仮名とのバランスが悪かったのではないか、という推測も成り立つ。なお検討したい。

いずれにしても、創刊年『太陽』において使い分けの見られる異体仮名〈志〉〈し〉は、その語頭・非語頭という使い分けの基準も含めて、同時代の資料の活字の状況を反映したものといえる。

4

字体の異なる異体仮名

〈志〉〈し〉のように、複数の異体仮名を使い分けている例とは異なって、〔ネ〕を表す片仮名の〈子〉や、ワ行の〔ヰ〕を表す〈井〉は、一音節が一文字に対応するよう統一されていることには変わらない。ただし、その統一された字体が現代の仮名文字表とは異なるのである。以下その例をあげる。

まず、片仮名の〔ネ〕を表す〈子〉だが、創刊年『太陽』には169例がみられる。

逅して之と和親を訂したる後ち更らに復た南してセル〈子〉
島に(1895年1号「紀元前の著名なる航海」森田思軒
P051E13)

このように、『太陽』における片仮名〈ネ〉はすべて〈子〉表記に統一されており、例外はない。仮名字体の統一の際に、どの字体を選択するかも、興味深い問題であるが、「小学校令」と比較してみると、「小学校令」は〈ネ〉を採用している。しかしながら、「小学校令」に先立つ、明治三十三年一月帝国教育会仮名調査委員「国語国字に関する決議」（明治三十三年十月十五日「教育公報」第240号所載）の仮名調査委員議決が、〈ネ〉ではなく〈子〉を採用しているところを見ると、どちらの仮名字体を採用するかについては、議論があったことがうかがわれる。

他資料で〈子〉の仮名字体がどのように用いられていたかをみると、『太陽』創刊年よりややさかのぼるが、『言海』（明治22年～24年）や『日本大辞書』（明治25年～26年）なども片仮名の〈子〉を使用している。〈志〉の仮名字体の場合と同じように、『太陽』の〈子〉の仮名字体も、当時の通行の異体仮名を選択したとみてよいようである。

次に、ワ行の〔ヰ〕を表す仮名〈井〉だが、189例みられる。

ヴ〈井〉ヴ〈井〉エンの説に従へば(1895年1号「紀元前の著名なる航海者」森田思軒P054A12)

このように、本文にみられるほか、ポイントの小さい活字においても同様の仮名字体を使用している。

ポイント小例) デーリグラフ〈井〉ツクに見えたるが(1895年1号「檳榔子」市村塘P175B13)

ポイント小例) 軍艦ロヤリスト號艦長ダヴ〈井〉ス氏(1895年1号「輿論一斑」著者不明P198C07)

また、ルビにおいても〈井〉を使用している。

戯場（シバ〈井〉）（1895年1号「京都の新案内記」中川
四明P073A24）

〈井〉の仮名は、〈子〉と同じように、現代の仮名字体表とは異なる字体に統一されている。

しかし〈子〉と違って、「小学校令」とは一致しているのである。とすれば、〈井〉の仮名字体はむしろ〈𛄁〉よりも一般的な仮名であったのかもしれない。

ただし、これも議論の余地がないわけではない。宇野(1993)によれば、「小学校令」の出た翌日22日に「…井ハ𛄁（中略）ノ孰モ誤」という「正誤」が示されているという（宇野論文より孫引き）。〈𛄁〉か〈井〉かの選択について、「小学校令」も揺れていたということであろう。『言海』や『日本大辞書』でも〈井〉の仮名を使用している（『言海』は〈𛄁〉〈井〉併用）。『太陽』でも、後年のものは〈𛄁〉の仮名を採用するようになる。創刊年では、より伝統的な古い〈井〉の仮名を使用しているということだろうか。

5

小さい活字のみに現れる異体仮名

これまで報告してきた仮名字体は、本文において用いられる仮名字体であった。〈志〉〈し〉にしても、〈子〉〈井〉にしても、複数の異体仮名を使い分けているか、一音節一文字に統一されているかの違いはあるが、本文に現れる仮名字体である。しかし、以下（5・6）に述べる異体仮名は、本文には現れない。ポイントが小さい本文にしか現れなかったり、ルビにしか現れなかったりする、いわば特殊な仮名字体である。そこで、表1にはのせていなかったり、あるいはかつこつきでのせている仮名字体である。

たとえば、〔ず〕を表す仮名字体〈春^ㇿ〕は、本文中には現れず、ルビにだけ現れる。本文中では〔ず〕は〈ず〉と表記されているのである。このような仮名字体は、元原稿の異体仮名の使い分けを反映したというよりは、おそらく印刷側の活字の種類事情などによるものと考えられる。なぜなら、ルビ内においては〈す〉〈春〉は厳密な使い分けがなされているものの、本文ではそうではなく、〈志〉〈し〉の使い分けのように、筆者による偏りもみられず、他資料にも似たような使い分けを見ることができないからである。〈す〉〈春〉の使い分けについては後で詳しく述べる。

こうした事情から、以下5, 6に述べる仮名字体を、本稿では、創刊年『太陽』に独自の問題と捉えている。ただし、単に例外的な事象として切り捨てることができないのは、用例数が少なくないからである。

たとえば、〈江〉の仮名字体だが、他の本文と比べてポイント数が小さい活字にしか現れないものの、全部で103例みられる。

沼猶見〈江〉ず(1895年1号「利根水源探検紀行」渡邊千吉郎P081A30)

音楽らしく聞〈江〉ず(1895年1号「海外思想」著者不明P195B36)

このような〈江〉の仮名字体が用いられるのは、ヤ行活用動詞(覚ゆ・見ゆ・聞こゆ)に限られる(「見え」の〈え〉は一例を除いてすべて〈江〉)。通常の大きさのポイントの本文においては、あ行の〈え〉を用いたり(〈見え〉)、は行の「へ」を用いたりしている(〈見へ〉)。

や行の〈江〉の字体の使用には、活字の大きさが関わっていることが推測される。

なお、ポイントが小さい本文中で、ヤ行活用動詞以外の〔え〕の音節をどの仮名字体で表記するのかは、実例がなく確認できなかった。〈江〉がヤ行活用動詞だけに用いられているのか、あるいは、〔え〕を表す仮名字体としてもっぱら用いられているのかは、わからない。

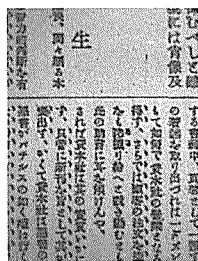


図2 「殖〈え〉行く」図中最終行、上段右上が通常のポイントの大きさ
(1895年4号「文學と英國の書肆」挹翠生P734B15)

6

ルビにのみ現れる異体仮名

次に、ルビにしか現れない異体仮名について報告する。

まず、〔ず〕を表す〈春〉である。創刊年『太陽』には、総数261例がみられる。〔ず〕を表す仮名字体としては〈春〉に統一されており、例外はない。ただし、ルビにおいてだけであつて、本文においてはこの限りでない。

蚊相撲（か〈春〉まふ）（1895年1号「相撲の話」羽南外史 P139B10）

さらに、〈春〉が用いられるのは、濁音の場合だけであつて、同じルビであつても、清音の場合は〈す〉の仮名を用いている。いわば、清音仮名は〈す〉、濁音仮名は〈春〉という区別が存在しているのである。

このような清音仮名と濁音仮名の対立が、〔す〕〔ず〕においてだけ現れるのか、はよくわからない。

〈す〉〈春〉の異体仮名の使い分けについて、他資料に目を向けると、たとえば、『文学界』明治28年1月～29年1月掲載の「たけくらべ」にもルビの〈春〉の異体仮名があるという（前田（1994）による）。また辞書『言海』では、〈春〉を用いるのは、漢字音の項目（「すくせ（宿世）」「すさい（秀才）」など）で、拗音の直音表記の場合であり、〈す〉は、それ以外（「すこしく（少）」「すくふ（救）」）で用いられている。〈春〉の字体は、〔す〕の音を表すものと区別され、特別に用いられた、いわば有標の字体であることがわかる。

創刊年『太陽』における〈す〉〈春〉の使い分けは、清音対濁音という点では、表音機能に基づいた使い分けをしているといえるが、ルビにしか現れないという点で特異である。先に、ポイントが小さい本文にしか用いられない異体仮名について述べたが、同じように活字の大きさが関わっていることがうかがわれる。

このように、5および6で触れた小さいポイントにしか現れない仮名字体や、ルビにしか現れない仮名字体は、当時の仮名の使用状況を反映したというよりは、『太陽』の印字を組んだ博文館の印刷技術や道具の制限によるもので、それ故『太陽』独自の問題であるといえる。

7

その他の異体仮名

そのほか、数ごく少ない異体仮名の例をあげておく。

まず、〔は〕を表す〈者〉の用例が、創刊年には16例みられた。

小金井喜美子「浮世のさが」(1895年7号・8号)に集中して表れるほか、以下の記事に見られる。

幸堂得知「古今演劇談(承前)」(文芸)(2号)

遅塚麗水「芦花」(小説)(4号)

著者不明「夕すゞみ」(雑文)(8号) 同字忌避

依田学海「阿新」(小説)(10号)

三宅花圃「筆のまに々々(其二)其三)」(11・12号) 文語小説や文芸作品に多く、元の原稿との関係が推定される。またほとんどが、「は」との同字回避による使用である。

人造ひは〈者〉だしなり、(1895年2号「古今演劇談(承前)」

幸堂得知P134A05引用「芝居訓蒙圖彙」)

男〈者〉はだか女はてゝ茶代も入らぬ(1895年8号「夕すゞみ(雑文)」著者不明P181B08 ※助詞のハを〈者〉表記する。)

また、〈そ〉の仮名(注6)が一例見られた。

御起居供御の御事までもかり〈そ〉めなることを爲させ玉はず、(1895年10号 海内彙報 P188B05)

1895年分には1例だけだが、1901年発行分のルビに13例「ぞ」の異体仮名がみえる。

「俗人(ぞくにん)」(1901年4号「家庭談叢 俳諧新旧派の異同」正岡子規談P141A23)

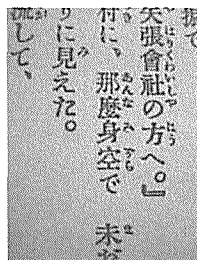


図3 ルビ「身空」(1901年4号「黄昏」川上眉山P084A15)

8

創刊年『太陽』の異体仮名の位置づけ

このように、『太陽』1895（明治28）年1号から12号の異体仮名を調査した結果、以下のようなことがわかった。

この時期の『太陽』の仮名は平仮名片仮名ともに、原則として一音節に一字体が対応するような形式に統一されている。1901（明治33）年に施行された小学校令施行規則による仮名字体の統一（異体仮名廃止）に先駆けて、『太陽』本文では異体仮名の収斂がほぼ完結していたとみてよい。

異体仮名を廃止する「小学校令」が施行されたのは、『太陽』創刊から五年後である。また同時代の雑誌（『新小説』明治29年分・『帝国文学』明治29年分）もほぼ一音節一仮名字体に統一されている。『太陽』における異体仮名の種類の少なさは、『太陽』に限ったものではなく、明治28年から30年初頭にかけての雑誌一般の傾向を反映したものとみてよいだろう。

ただし、『太陽』の異体仮名は、種類は少ないが、用例数は決して少なくない。ここに、異体仮名の収斂の末期状況をみることができる。

異体仮名のなごりともいうべきこれらの仮名字体についてまとめて示す。

（1）語頭・非語頭で使い分けのあるもの

〔し〕を表す異体仮名〈志〉〈し〉は、語頭が〈志〉、非語頭が〈し〉というように使い分けられている。

（2）ある特定の著者などに限定的に用いられるもの

〔は〕を表す〈者〉は、小説や文芸作品に主に用いられている。また、〔し〕を表す〈志〉も、ある特定の著者や記事に限定的に用いられている。これらは、元の原稿の字体を反映したものであろうことがうかがわれる。

（3）現代の仮名字表とは違った字体を採用しているもの。

片仮名の〔ネ〕を表す〈子〉や、ワ行の〔ヰ〕を表す〈井〉などは、一音節一文字に統一されているものの、現代の字体とは違う字体が採用されている。

（4）活字の状況による違い

〔ず〕を表す〈春^ゝ〉が、ルビにのみ現れるとか、〔え〕を表す〈江〉が、本文の文字ポイントが小さい場合にしか現れない、などのような例は、活字による制限が大きく影響していることが推測できる。

(1)(2)(3)は、仮名字体統一の流れの中に残った、伝統的・慣習的な異体仮名の使用のなごりといえるだろう。先に述べたように、異体仮名廃止の動きは、一日で完成したものではなく、実際の活字媒体においては、徐々に収斂されていったものと考えるのが自然である。そのように仮定した場合、果たしてどのような異体仮名の使い分けが、どのような状態で残り続けるのか。統一は問題なく進むのか。その一側面を、『太陽』の異体仮名の状況(1)～(3)は示している。

すなわち、「小学校令」の異体仮名廃止が意味するところは、複数の字体が、ひとつの音節に対応するという、非合理的な書記のありかたから、一音節には一字体が対応するという、合理的な体系への切り替えだ、とすると、使い分けに意味のない異体仮名から廃れていき、いまだに何らかの役割を担っている異体仮名のみが残ることになるだろう。(1)のように〔し〕を表す〈し〉〈志〉の異体仮名の残存は、語頭・非語頭という位置による使い分けといわばセットになっているのである。

また、異体仮名の統一が行われたのは、活字が先であり、手書きはずっと遅れたことが推測される。(2)のように、ある特定の著者に限定的に用いられる字体や、異体仮名の使い分けは、元の手書き原稿を反映したものであると考えられる。前田(1994)は、『たけくらべ』において、手書き原稿と雑誌掲載時の活字における仮名字体を比較している。そこでは手書き原稿がより多く異体仮名を残していることが知られる。伝統的・慣習的な異体仮名の使い分けや字体の選択は手書きのほうにより残りやすく、統一は活字のほうに起こりやすいという状況を表している。

(3)のように、現代とは異なる仮名字体を使用している『太陽』だが、〔ネ〕を表す〈子〉や、〔𠂔〕を表す〈井〉は、むしろ当時の通用の字体であったことが、他の資料をみるとわかる。

(1)～(3)は『太陽』に独自の基準によって使用されたのではない。むしろ、当時の活字の状況を反映しているものといえる。『太陽』におけるいくつかの「例外」は、文部省による規則制定までの試行錯誤段階との興味深い一致を示している。両者の連続性は、互いに当時の活字の状況と無関係でないところから生じている。

一方、(4)は、『太陽』に独自の問題である。これらは、当代の一般的な状況とはくい違っており、むしろ『太陽』を出版している博文館の印刷技術や活字による制限に負うところが大きかつ

たことが推察される。元々は伝統的・慣習的であった異体仮名の使い分けや仮名字体の選択—や行の〈江〉や〔す〕を表す仮名字体〈春〉など—が、おそらくたまたまルビやポイントが小さい活字には残ったのではないだろうか。こうした偶然の事情から、異体仮名のなごりが当時の雑誌の活字に現れることもあったのであろう。

9

まとめと課題

『太陽』の異体仮名の選択・決定要因となっているものは、

A 元の手書きの原稿の字体

B 『太陽』印刷所の活字の状況（注7）

であると思われる。

たとえば、手書き文字の異体仮名の例として、漱石書簡をみると、〔か〕をあらわす仮名として〈可〉が、〔た〕を表す仮名として〈多〉が、それぞれ多用されていることがわかる。1900年の字種制限以降も、手書き文字においては、異体仮名の存在が許されていた。手書き文字と活版印刷とにおける異体仮名のありようには違いがあったことが知られる。

重要なのは、このような手書き原稿における異体仮名の使い分けを、活字にした段階でどの程度反映させたか、であろう。

そもそも、同一の音を表す仮名が複数の字母をもつという書記システム（＝異体仮名）は、現在の目から見ると不便で読解効率も低いように思われる。しかし、古くは、音節文字列である平仮名文において、異体仮名が、境界表示という重要な役割を果たしていたことは、定家筆文献の調査等を端的な例としてかなり詳細に記述されている。日本語の仮名文における異体仮名は、アルファベット文における分かち書きのような機能を果たしていたといえる。

のちに、この境界表示の役割は、漢字と仮名という異なる文字体系を混ぜて使用するというシステムにとってかわられる。異体仮名の役割は終わるのである。こうして平仮名の異体仮名が、時代が下るにつれて収斂される傾向にあることは、すでに先行研究により指摘されている通りである（浜田1979・前田1971など）。

夙に、三好（1977）は、印刷所の活字字母の変更によって、明治二十二年を境に、雑誌の紙面に、「変体仮名から近体仮名へ」

という「活字字母の近代化」が急速に進んだことを指摘している。三好（1977）によれば、「字母の近代化」は印刷所によってかなりの開きがあり、『文学界』の印刷所であった秀英社工場は、平仮名字母を比較的早く近代化した印刷所のひとつだった」という。

いわゆる活版印刷が、異体仮名の収斂を後押ししたことは想像に難くない。

しかしそれは、単に、異体仮名がもはや機能を失ったゆえに収斂されるという変化の先と同一線上にあるだけではない。むしろ、『太陽』においても、異体仮名の状況は、ある程度は手書き原稿を反映したものである。たとえば〈志〉の残存状況が示すように、異体仮名の役割の有無と全く無関係ではない。しかし、一方で、活字の制限により、たまたまルビやポイントが小さい部分にだけ残った仮名字体も存在する。このような活字の状況は、「例外的」とはいえ、異体仮名の機能という問題とは別に生じたものである。

活版印刷が、異体仮名の統一にどのように決定的に関わったのか、については、なお詳細な観察と記述を要するように思われる。そのためには、手書き段階のものと活字になったものとの比較などが有効であろう。矢田（1998）は「平仮名の歴史的研究に当たっては近世以前と以後とではその方法論において異なるものでなければならない」と述べている。方法論そのものの検討も必要になるだろう。

注

- (1) 同じ音節に対応する複数の異なる仮名字体を、互いに「異体仮名」と呼ぶことにする。たとえば、〔し〕の音節に対応する仮名〈し〉〈志〉は異体仮名である。「変体仮名」「字母」と同じものをさす。「同音異字」（前田1994）・「異字体」「同音異字体」「同字源由来の異字体」（玉村1994）とも同じものをさす。
- (2) 「仮名遣い」とは、「仮名の綴りについて社会的に設定される軌範」（『言語学大辞典』小松英雄記述）を意味する。
- (3) 本稿では字形は対象外とする。明治三十三年一月帝国教育会仮名調査委員「国語国字に関する決議」では、「仮名の字形に改革を施さず（活字を横広く作るなどは此限りにあらず）」（明治三十三年十月十五日「教育公報」第240号所載）とされている。
- (4) 『当世書生氣質』の「志」には語表記の固定がみられる。「しばしば」「居らつしやる」「しかし」「知る」など。また、

「～てしまう」「いらつしやる」など形態素の切れ目にも現われるが、『太陽』にはこのような例はあまりない。

- (5) 西野嘉章編『東京大学コレクション〈3〉 歴史の文字一記載・活字・活版』（1996・10 東京大学出版会）には、『葛原句当日記』（葛原重美1856・1857・1858・1864（安政3・4・5・元治元年）史料編纂所 貴重書 日記二冊）の木活字の写真が掲載されている。その平仮名の内訳は「い～す」47字および「志」および「ん」で計49字である。これらの活字がいろは順に並べられている。葛原重美はこの活字を用いて日記をつけていたわけだが、〔し〕の異体仮名をのみ使い分けていたことがわかる。

- (6) その他次のような字体の問題があった。

○〔わ〕の音節を〈王〉表記するもの

引用・割書部分に以下の例があった。

是を〈王〉りこさいにといふ（1895年2号「見聞拾葉（一）柳営元日の献立」小宮山綏介P104B26割書）

○〈ハ〉

本文中で、助詞の「は」を〈ハ〉表記している例が2例あった。ここでは「は」の異体仮名とはみなさなかった。

かの子〈ハ〉ははやなくなりたれば。（1895年8号「浮世のさが」小金井喜美子P109B19）

○〈ツ〉

本文中で、促音便の「つ」（主に動詞のタ形・テ形）が片仮名の〈ツ〉（大字）に表記されることがある。ここでは、〈つ〉の異体仮名とはみなさなかった。平仮名と片仮名の干渉については別途考察する。

「いや、名歌は姑預〈ツ〉てにおいて（1895年1号「取舵」尾崎紅葉P087A10）

- (7) 『共同印刷90年史』によると、火事によって明治初期の活字セットは残っていない。

明治37(1904)年12月21日 午前1時 博進社印刷工場火災

明治38(1905)年4月 博文館印刷会社としてスタート

大正4（1915）年5月17日 博文館印刷会社 二度目の火事

参考文献

- 宇野義方（1993）「現行の仮名字体をめぐって」（松村明先生喜寿記念会編『国語研究』625-648頁，明治書院）
玉村禎郎（1994）『『春色梅兒誉美』における仮名の用字法』（『国

- 語文字史の研究』2, 175-206頁, 和泉書院)
- 浜田啓介 (1979) 「板行の仮名字体—その収斂的傾向について—」
『国語学』118, 1-10頁, 国語学会)
- 古田東朔(1974a) 「変体がなから平仮名へ (上)」(『言語生活』
272号, 57-80頁, 筑摩書房)
- 古田東朔(1974b) 「変体がなから平仮名へ (下)」(『言語生活』
273号, 89-95頁, 筑摩書房)
- 三好行雄 (1977) 「〈文献学の恐さに無智な蛮勇〉について」
『文学』45-2, 230-239頁, 岩波書店)
- 前田富祺 (1971) 「仮名文における文字使用について—変体仮名
と漢字使用の実態—」(『東北大学教養部紀要』14, 99-134頁)
- 前田富祺 (1994) 「『たけくらべ』におけるひらがなの書体と字
体」(『国語文字史の研究』2, 207-227頁, 和泉書院)
- 安田章 (1980) 「仮名資料序」(『朝鮮資料と中世国語』414-
431頁, 笠間書院)
- 矢田勉 (1996) 「異体がな使い分けの衰退—トの仮名の場合—」
『山口明穂教授還暦記念国語学論集』439-457頁, 明治書院)
- 矢田勉 (1998) 「印刷時代における国語書記史の原理」(『東京大
学国語研究室創設百周年記念国語研究論集』567-585頁, 汲
古書院)
- 調査対象資料 (『太陽』以外)
- 飛田良文・松井栄一・境田稔信編『明治期国語辞書大系』大空社
1997~1998
- 『漢英対照いろは辞典』(高橋五郎編・長尾景弼・小林富実発行・
明治21年)
- 『ことばのはやし』(物集高見編・清水卯三郎発行・明治21年)
- 『日本小辞典』(物集高見編・吉川半七・坂上半七発行・明治11年)
- 『和漢雅俗いろは辞典』(高橋五郎編・長尾景弼・小林富実発行・
明治21年)
- 『日本辞書 言海』(大槻文彦編・発行・明治22~24年)
- 『日本大辞書』(山田美妙編・発行・明治25~26年)
- 『日本大辞林』(物集高見編・清水卯三郎発行・明治27年6月・
宮内省)
- 『和英大辞典』4版 (プリンクリー著・南條文雄編・明治30年12
月・三省堂)
- 『若菜集』(島崎藤村・明治30年8月・春陽堂・秀英舎第一工場)
- 『帝国文学』明治28年1月号~12月号 (帝国文学会・大日本図書)
- 『新小説』明治29年1号~7号 (春陽堂・秀英舎第一工場)

濁点文字使用率から見る濁音表記

——近藤 明日子

1

はじめに

現代日本語で濁音を仮名で表記する場合、該当の仮名すべてに濁点を施すのが一般的である。この表記法が定着したのは明治時代以降とされており、松村（1972）には次のようにある。

濁音符は、明治期において、必ず濁音を示す場合に使用されたわけではなく、必要な場合にだけ使用された。明治初期の漢字カタカナまじり文には特にこの傾向が強く、「ベシ」「ガ」「ズ」「バ」など前後の文脈で知られるものは、濁点を使用していないものが多い。そして、濁点をつけない習慣は、文の種類によって、のちのちまで続いた。（247頁）

では、濁音を仮名で表す際に必ず濁点付きの仮名を用いる表記法は、どのような過程を経て定着にいたったのであろうか。それを具体的なデータに基づき考察した研究は未だないようである。そこで本稿では『太陽コーパス』を用いて、1895年から1925年の雑誌『太陽』における濁音表記の様相を考察することとする。そこには、当時の活版印刷による総合雑誌において、濁音を仮名で表記する際に必ず濁点付きの仮名を用いる表記法の定着していく濁音表記史の一断面が立ち現れることが期待されるのである。

2

調査対象および考察方法

調査対象とするのは『太陽コーパス』の2004年9月段階のデータで、ルビ部分は調査対象外とする。『太陽コーパス』に基づき雑誌『太陽』で用いられる濁点付きの文字を抜き出してみると、現代でも一般的に用いられるガ・ザ・ダ・パ行の平仮名・片仮名や片仮名「ヴ」の他に、片仮名「ワ・キ・エ・ヲ」に濁点の付いたものや「ヶ」に濁点の付いたもの（1895年12号P125A15の1

例のみ)を見出すことができる。また、文字以外にも符号に濁点の付いたものがあり、濁点付きの一字点「ゞ・ゐ」と濁点付きのくの字点が見出される。このような濁点付きの文字・符号を以下「濁点文字」と総称することとする。これら雑誌『太陽』で用いられる濁点文字を表記するために『太陽コーパス』で用いられる濁点文字は「がぎぐげごぎじずぜぞだちづでどぼびぶべぼガギグゲゴザジズゼゾダヂヅデドバビブベボヴゞゐ」の計44種を数える。この内、最後の「ゐ」は雑誌『太陽』で用いられる濁点付きの「ワ・キ・エ・ヲ」と「ヶ」とくの字点を、それぞれ「ワゝ・キゝ・エゑ・ヲゝ」「ヶゝ」「ゐゐ」と表記するために用いるものである。

『太陽コーパス』の本文では、検索の便を考え、雑誌『太陽』で語形の上で濁音と考えられる仮名・符号の表記に濁点文字が用いられていない場合、濁点文字に置き換える校訂を行っている(注1)。つまり『太陽コーパス』の本文は濁音表記に関して校訂された本文なのであるが、校訂を行った箇所に分類属性値が「F濁点脱落」の注タグを付し、さらにその注タグの原文属性値に雑誌『太陽』原文での表記を情報として残し、原文の濁音表記を再現できるようになっている。よって、『太陽コーパス』から原文本文と校訂本文とをそれぞれ作成し、両者の濁音表記を比較することが可能である。そこで、この原文本文・校訂本文それぞれから上述の44種の濁点文字の個数を求め、次のような「濁点文字使用率」なる値を算出することとする。

$$\text{濁点文字使用率} = \frac{\text{原文本文での濁点文字数}}{\text{校訂本文での濁点文字数}}$$

現代語で一般的な、濁音の仮名にはすべて濁点文字を用いる表記法は、濁点文字使用率の観点から見れば、「濁点文字使用率100%」の表記法ということができるであろう。さらに、明治時代以降の濁音表記史は、初期には濁点文字使用率が100%に達しなかったものが、年を経るにつれ100%に近づいていくものであったことが予想されるのである。以下本稿では、この濁点文字使用率を用いて『太陽コーパス』の濁音表記の変遷をたどっていくこととする。

なお、濁点文字使用率は、記事中の引用部分(『太陽コーパス』で引用タグが付いている部分)は除き、記事地の文のみを対象と

して求めることとする。なぜなら、次の(1)のように地の文と引用部分とは濁点文字使用率に大きな差が見られる記事が存在するからである。

(1) 勅に曰く

(中略) 朕乃チ友邦ノ忠言ヲ容レ朕カ政府ニ命シテ三國政府ニ照覆スルニ其ノ意ヲ以テセシメタリ若夫レ半島地域ノ還附ニ關スル一切ノ措置ハ朕特ニ政府ニ命シテ清國政府ト商定スル所アラシメントス今ヤ講和條約既ニ批准交換ヲ了シ兩國ノ和親舊ニ復シテ局外ノ列國亦斯ニ交誼ノ厚キヲ加フ百僚臣庶其レ能ク朕カ旨ヲ體シ深く時勢ノ大局ニ視、微ヲ慎ミ漸ヲ戒メ邦家ノ大計ヲ誤ルコト勿キヲ期セヨ其の露、獨、佛の忠言を公表し給ふの一段、磊々落落、苟くも隱蔽の跡なし、聖圖の雄遠なる、實に測る可らざる也、嗚呼之を拜讀して感奮興起し、皇猷を翼賛せんことを誓ふもの國を擧げて皆然らざるはなき也(1895号6号「政治」P166B05～P166C16)

(用例としてあげる本文は、『太陽コーパス』の校訂本文から、濁音表記に関する箇所のみ、注タグの原文属性値に基づき原文の表記に復したものである。―は濁音表記に濁点文字を使用している箇所、〃は濁点文字を使用していない箇所を示す。以下同様)

(1) では1行目と9行目以降の記事地の文に挟まれて、詔勅が引用されているが、記事地の文では濁点文字使用率100%の表記法をとりながら、詔勅の引用部分では濁点文字使用率0%の表記法をとっている。また次の(2)(3)も同様に地の文と引用部分とで濁点文字使用率に大きな差が見られる記事である。

(2) 今左に小學校の管理と、教授とに關する訓令を示さん。

小學校管理

一 學校ノ体制ヲ確立セント欲セハ適當ナル規律ト善良ナル方便ヲ以テ學校ノ事務ヲ整理セサルヘカラス而シテ管内小學校ニ於テハ適當ニ管理ノ効果ヲ舉クルモノ少シ今後左ノ事項ニ付特ニ注意スヘシ

一 教員ハ毎日規定時間ニ出勤シ終業後翌日教授ノ準備作文習字等ノ訂正諸帳簿ノ整理ヲ終ハルニアラサレハ退出スヘカラス

(中略)

是の他に、尚條項頗多けれども、訓令の性質は、畧之に

依りて類推すべきが故に省畧す。此の如き老婆的訓令は、或る關係に於ては、少からぬ利益もあるべけれども、之に服従する義務ある教師の身に取りては、實に苦痛の至なるべし。且かゝる訓令の常として、當局者が氣がつきたることは、實に下らぬ事までも書き立て、氣がつかざることは、大關係のあることにても、一言も言ひ及ばざること多し。(1895年9号「教育」P158A28～P159A31)

- (3) 私はこの模寫帖を二百部程造つて、友人間に寄贈したが當時田中光顯伯が宮内大臣をして居られたので、同伯の手で明治天皇、照憲皇太后兩陛下と、時の皇太子殿下に献上したことがある。その歌は春と秋との歌であつた。

前のおほもろち君

年ふれはよはひは老いぬしかはあれと 花をし見れはもの思ひもなし

秋風に初かりかねそきえゆなる たか玉章をかけてきつらん

(1925年9号「我楽多雜記(上)」高橋義雄 P075B22～P075C05)

(2) では2～9行目が奈良県訓令の引用部分、(3) では5行目以降が伝記貫之自筆の色紙(の模写)から和歌を引用した部分であるが、(1) 同様、記事地の文は濁点文字使用率100%であるのに対し、引用部分は0%となっている。

このように記事地の文と引用部分とで濁点文字使用率に差のある記事が存在することから、当時、他の文献からの引用に際し、記事地の文の表記法に則って引用元の表記を改めるのではなく、引用元の表記を保持しようとする引用態度が存在していたことがわかる。つまり、引用を含む記事には、記事地の文と引用元の文献との複数の表記法が混在している可能性が高く、しかも場合によっては(3)のように『太陽コーパス』の扱う時期と一致しない時代に書かれた文献の表記法が混在することもあるわけである。『太陽コーパス』を通して当時の雑誌における濁音表記の実態を考察しようとする本稿では、考察の対象から他の文献からの引用部分は除外したい。よって以下の考察では、記事地の文を対象として濁点文字使用率を求めるものとする。

3

濁点文字使用率の通時的変化の概観

まず、『太陽コーパス』での濁点文字使用率の通時的変化を概観する。表1は、年別に原文本文の濁点文字数と校訂本文の濁点文字数（注2）さらに濁点文字使用率を百分率で示したもので、図1はその濁点文字使用率の経年変化をグラフとして示したものである。

表1 濁点文字数と濁点文字使用率

	1895	1901	1909	1917	1925	通年
原文本文濁点文字数	100656	121657	130096	148070	173163	673642
校訂本文濁点文字数	106464	123207	131308	148316	173402	682697
濁点文字使用率	94.5%	98.7%	99.1%	99.8%	99.9%	98.7%

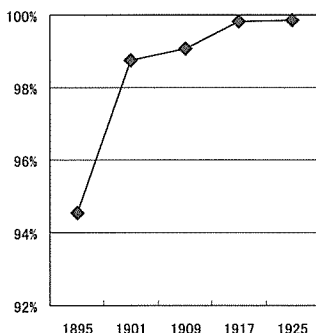


図1 濁点文字使用率の変化

ここからわかるように、濁点文字使用率は1895年が94.5%で最も低く、年を経るごとに高くなり、1925年に99.9%に達している。ただし、その上昇の度合いは一定ではなく、1895年から1901年にかけて急激な上昇の段階があり、1901年から1917年にかけては穏やかな上昇の段階、1917年・1925年にはほぼ一定という変化を見せる。また、1895年の濁点文字使用率94.5%は、それ以降の年と比べれば低い値とはいえ、既に比較的高い値であったと言える。濁点文字使用率100%の表記法が定着しつつあ

るものの未だ完全ではなかったのが1895年であり、それが完全に定着したのが99.8%とほぼ100%の値に達した1917年であると考えられる。

次に、記事の文体別に濁点文字使用率の通時的変化を見ていく。『太陽コーパス』では、記事タグの文体属性値に記事地の文の文体の情報が記されている。その内の文体属性値が「口語」と「文語」の記事それぞれの原文本文および校訂本文の濁点文字数と濁点文字使用率を示したものが表2、その濁点文字使用率の経年変化をグラフとして示したものが図2である。

表2 文体別の濁点文字数と濁点文字使用率

		1895	1901	1909	1917	1925	通年
口語	原文本文濁点文字数	10254	45514	94254	137200	170867	458089
	校訂本文濁点文字数	10443	45787	94940	137423	171102	459695
	濁点文字使用率	98.2%	99.4%	99.3%	99.8%	99.9%	99.7%
文語	原文本文濁点文字数	90402	76143	35842	10870	2296	215553
	校訂本文濁点文字数	96021	77420	36368	10893	2300	223002
	濁点文字使用率	94.1%	98.4%	98.6%	99.8%	99.8%	96.7%

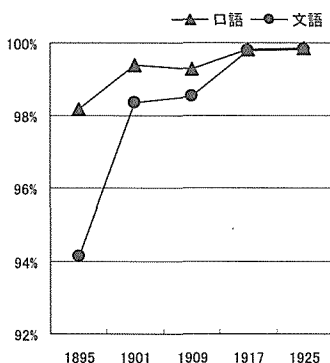


図2 文体別の濁点文字使用率の変化

ここからわかるように、口語・文語のグラフとも濁点文字使用率は1895年から1917年まで上昇し、1917年・1925年でほぼ100%の値で安定するという図1のグラフとほぼ同様の上昇の軌跡を描く。ただし、ほぼ100%に至るまでの1895年・1901年・1909年の3カ年は、文語よりも口語のほうが常に濁点文字使用

率が高い。特に差が大きいのが1895年で、文語が濁点文字使用率94.1%であるのに対し、口語は98.2%とすでに100%に近い値を示している。濁点文字の使用は、口語記事で先駆けて進んでいたと考えられるのである。

4

文体別の濁点文字使用率

次に、口語・文語それぞれについて記事単位で濁点文字使用率を見ていくことで、3.で現れてきた文体による濁点文字使用率の差に関してより詳しく考察することとする。

表3 濁点文字使用率別の口語記事数

	1895	1901	1909	1917	1925
70-80%	1				
80-90%			3		
90-95%	1		5		
95-100%	33	146	365	323	175
合 計	35	146	373	323	175

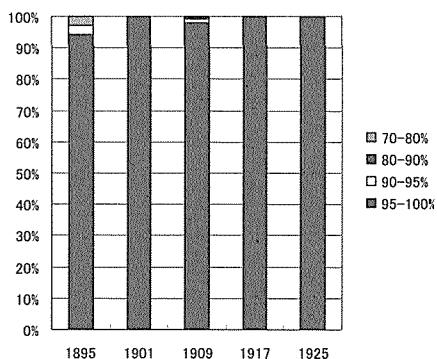


図3 濁点文字使用率別の口語記事数の構成比率

まず、口語記事について、濁点文字使用率10%幅（90～100%は5%幅）ごとの記事数を年別に示したものが表3、それをもとに年別の記事数の構成比率をグラフに示したものが図3である。ただし、濁点文字数の少ない記事の場合、例外的な濁点文字使用率となる場合があるので、校訂本文での記事地の文の濁点文字数50以上の記事に限定して集計してある。

ここからわかるように、口語記事では1895年からすでに90%以上の記事が濁点文字使用率95-100%であり、それ以降の年とそれほど差はない。これが表2で見た1895年の口語記事全体の濁点文字使用率が98.2%という既に100%に近い値を示す結果につながっていることになる。

次に、文語記事の濁点文字使用率別の記事数の分布を見ていく。文語記事について、濁点文字使用率別の記事数を年別に示したものが表4、それをもとに年別の記事数の構成比率をグラフで示したものが図4である。表3・図3と同様、校訂本文の濁点文字数50以上の記事を集計対象とした。

表4 濁点文字使用率別の文語記事数

	1895	1901	1909	1917	1925
0-10%	7		1		
10-20%	2				
20-30%	1				
30-40%	1				
40-50%	1				
50-60%	1				
60-70%	7	2			
70-80%	20	1			
80-90%	52	8	1		
90-95%	64	11	6		
95-100%	333	332	145	50	8
合 計	489	354	153	50	8

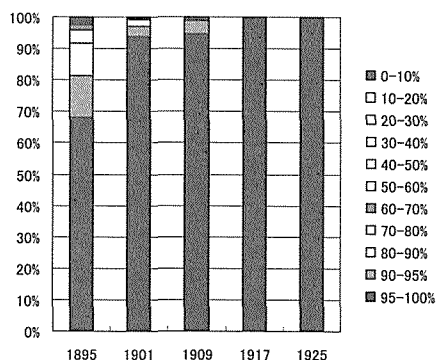


図4 濁点文字使用率別の文語記事数の構成比率

ここからわかるのは、1895年では文語記事は口語記事とは異なり、記事ごとの濁点文字使用率の値に多様性があるということであり、記事によって濁点文字使用率0-10%のものから95-100%のものまで幅広い。また、それゆえに、濁点文字使用率95-100%の記事数の構成比率が口語記事よりも低く、口語記事で構成比率が90%を超えていたものが、文語記事では70%に届いていない。これが表2で見た1895年の文語記事全体の濁点文字使用率94.1%という口語記事より低い値を示す結果につながっていることになる。以降の年は、1901年・1909年には濁点文字使用率95-100%の記事の構成比率は90%を越え、1917年・1925年にはすべての記事が濁点文字使用率95-100%になり、口語記事との差はなくなっていく。

5

濁点文字使用率の低い文語記事

ここで、表4・図4で示したように、特に1895年の文語記事に集中して現れる濁点文字使用率0-10%の記事群に注目して考察する。濁点文字使用率100%の表記法を規範として著した記事で、実際の濁点文字使用率がこれだけ低くなることはまず考えられない。これら濁点文字使用率0-10%の記事は、濁点文字使用率0%の表記法を規範として著されたものと考えるほうが自然である。1895年にはこのような濁点文字使用率0%の表記法を規範とす

る記事が特に集中して存在したことになる。それらの記事の特徴を考察するため、記事に関する情報をまとめて示したものが表5である。

表5 濁点文字使用率0-10%の文語記事

年	号	記事開始位置	記事題名	著者	濁点文字使用率
1895	4	P109B01	〈短歌〉	多	0.0%
	6	P081A01	東京感化院屏風三十六首	多	0.0%
	9	P166B04	〈短歌〉	多	0.0%
	2	P064A05	広島の形勢	野口勝一	4.3%
	6	P145A01	時	寒沢振作	4.6%
	10	P159B03	〈和歌〉	多	6.0%
	5	P072B01	〈和歌〉	多	7.8%
1909	5	P081A01	国定教科書の翻刻発行に就きて	文部省	10.3%

これを一覧してまず気付くのは、濁点文字使用率0-10%の記事群の中心を占めるのは、1895年に毎号連載された選歌の記事（著者が「多」のもの）であるということである。次の（4）にその例をあげる。

（4） 鶯有慶音 大勳位 彰仁親王

よろこひの色にいてゝもさく花のひかりをそふる鶯のこゑ
同妃 頼子御方

鶯の聲のとかにもきこゆなり君か千とせのかけやたのしき
故博恭親王妃 郁子御方

行すゑも猶なかゝれとうたふらん千世もこまれるそのゝ鶯
従一位 徳川慶喜

なれもまた君をいはふかこの春はよろこほしげに鶯のなく

(1895年4号「〈短歌〉」P113B01～P113B08)

このように、当時和歌の表記では濁点文字使用率0%が規範とされる場合があったということがわかる。ただし、同じ1895年でも記事中に引用された短歌を濁点文字を使用して表記する場合もあり、和歌であっても濁点文字使用率は著者や編者によってさまざまであったことがうかがえる。濁点文字使用率0%の表記法を規範とする意識の強い編者が、これら1895年連載の選歌の記

事を担当したのであろう。1901年以降は、そもそも地の文が和歌で占められる記事がほとんど見られなくなるのであるが、記事中に引用された当時の和歌の一部には濁点文字使用率0%の表記法が引き続き見出される（注3）。しかし、1901年以降は和歌を濁点文字使用率0%で書き表す表記法は減っていくようであり、その傾向は、例えば同じ選歌の記事であっても1909年4号の記事「和歌」（P150A01～P150C08）が、濁点文字使用率95.7%と100%に近い高い値を示していることに端的に現れている。和歌を濁点文字使用率0%の表記法で書き表すことは『太陽コーパス』の5カ年通して確かに見られるのであるが、そのような表記法は年を経るにつれ少なくなり、短歌の表記法もしだいに濁点文字使用率100%を規範とするものへ移行していったと考えられる。

次に、表5に示された記事から、唯一の1909年の記事「国定教科書の翻刻発行に就きて」を取り上げてみる。次の（5）に記事の冒頭部分をあげる。

- （5）近時新聞紙等に於て種々の攻撃非難を試み、又文部省に對して請願運動を爲すものあり。現在の制度に不完全と認むるものあらは、之か改良を講ずべきや固よりなりと雖も、其説く所正鵠を失し、往々誤解詭誣に出づるものは之を辯し、事の真相を明にすること寔に已むを得ざるなり。乃ち項を分ちて説明すること左の如し。（1909年5号「国定教科書の翻刻発行に就きて」文部省P081A04～P081A09）

この記事は著者が文部省となっていることから、公的文書に準じる文章と考えることができる。当時、詔勅・法令などの公的文書は濁点文字使用率0%の表記法をとることが一般的であった（注4）。『太陽コーパス』でも（1）（2）にあげたように、濁点文字使用率0%の当時の詔勅・法令などがそのまま引用されている例が、1925年に至るまで多数見られる。文部省によるこの記事も当時の公的文書の表記法にならったものと考えられる。

以上のように、和歌や公的文書といった一部の種類の文章では濁点文字使用率0%の表記法が規範となっていることがわかった。本稿冒頭にあげた松村（1972）の「濁点をつけない習慣は、文の種類によって、のちのちまで続いた」とは、これらの種の文章について述べたものと思われる。

ところで、表5にあげた記事で和歌や公的文書以外のものに、

1895年の2つの論説記事「広島[〰]の形勢[〰]」「時[〰]」がある。次の(6)に例として「広島[〰]の形勢[〰]」の冒頭部分をあげる。

(6) 今や廣島は其名大に内外國に顯はれ苟も時事を談するものは同地の形勢如何を知らんと欲せざるはあらず是れ征清の大師一たひ海に航せしより__大元帥陛下大轟を此に駐め大本營となし軍務を親裁し玉ふに因てなり先つ其大勢より叙述して次第に細事に及はんとす

(1895年2号「広島[〰]の形勢[〰]」野口勝一P064A11~P064A14)
論説記事は『太陽コーパス』の記事の主軸をなすものであり、『太陽コーパス』の濁音表記を語る上で論説という種類の文章での濁音表記の考察は欠かすことができないと考える。その論説記事の一部に(6)にあげたような濁点文字使用率0%の表記法を規範とするものが存在することになるが、実はそれは少数派であって、図4に示したように、70%近くの文語記事は濁点文字使用率95-100%の範囲に入り、それは論説記事においても例外ではない。

6

著者別に見る濁点文字使用率

そこで、次に著者ごとに濁点文字使用率を見ていきながら、文章の種類と濁点文字使用率との関係を引き続き考えていきたい。4.で見たように、1895年の文語記事では、記事により濁点文字使用率に多様性が見られたが、同様に著者による差も1895年の文語記事で見られることが期待される。

1895年の文語記事での濁点文字使用率の上位20位の著者を示したものが表6、下位20位の著者を示したものが表7である。著者が一個人(あるいは一組織)に特定できる記事を対象とし、かつ著者別の記事地の文の校訂本文の濁点文字数が50以上の著者のみあげた。表中の著者分野・所属は『太陽コーパス』付属の全文検索システム『ひまわり』Ver.0.9.8の著者一覧に拠っている。

ここからまず、著者により濁点文字使用率に差があったことは明らかである。そして表6に示した著者の特徴として、著者分野に小説家とある著者(表中●を付して示す)が4名おり、さらに著者分野に小説家と明記はされていないが、1895年に小説記事が掲載されており小説家に準じて扱うことのできる著者(表中▲

表6 1895年の文語記事にて濁点文字使用率の高い著者

著者	著者分野 (所属)	濁点文字 使用率
HM生		100.0%
HO生 (訳)		100.0%
TO生		100.0%
くりや女		100.0%
安田篤	(理学士)	100.0%
岸上質軒 (訳)	評論家 漢詩人	100.0%
玉露園		100.0%
桜井平吉	自由民権運動家	100.0%
修美生		100.0%
●青大通	劇作家 小説家	100.0%
●泉鏡花	小説家	100.0%
前田香雪	鑑識家 美術家	100.0%
大橋新太郎	出版人 事業家 (博文館創立者 日本工業倶楽部理事長)	100.0%
怡蝶外史 (訳)		100.0%
▲南新二		99.8%
●嵯峨の屋おむろ	小説家 詩人	99.7%
●塚原渋柿園	小説家	99.7%
木村鷹太郎	評論家 翻訳家 翻訳者	99.6%
胡蝶		99.6%
稲垣満次郎	外交官	99.6%
宕嶺居士		99.6%

表7 1895年の文語記事にて濁点文字使用率の低い著者

著者	著者分野(所属)	濁点文字 使用率
寒沢振作		42.6%
野口勝一	画家(衆議院議員)	65.1%
佐々木指月	彫刻家 詩人	65.5%
曾根俊虎	軍人(海軍少尉)	66.7%
東京商工相談会		67.9%
干河岸貫一	著作家	68.3%
天野為之	経済学者 教育家(早稲田大学教授 早稲田実業学校校長)	68.5%
S.T.生		72.7%
飯田武郷	国学者	77.3%
小西増太郎	神学者 ロシア思想研究者(露国神学士)	77.4%
KK生	(農科大学)	77.6%
国府犀東	詩人 歴史地誌学者	78.2%
捻華主人		78.9%
挹翠生(訳)		79.0%
岡倉天心	美術史家 思想家(東京美術学校校長 日本美術院創立者)	79.4%
曳尾叟	外交官 漢学者(元老院義官 貴族院議員)	79.4%
大江敬香	漢詩人	79.9%
田中従吾軒		80.0%
田岡嶺雲	文芸評論家 中国文学者	81.0%
有住斉		81.3%

を付して示す)が1名いることがあげられる。それと比べて、表7には著者分野が小説家の著者は見当たらず、論説記事の掲載される著者でほとんどが占められている。ここには表5にあげた濁点文字使用率0-10%の論説記事の著者である寒沢振作や野口勝一の名も見える。

以上のことから、同じ文語記事であっても、小説家の著す記事は濁点文字使用率が高い傾向にあると推測することができる。1895年の文語記事の校訂本文の濁点文字数50以上の著者のうち、著者分野が小説家の著者は26名を数えるが、その濁点文字使用率を示したものが表8である(表6にあげられている著者も含む)。ここから、濁点文字使用率の最も低い著者で93.0%、そ

表8 1895年の文語記事での小説家の濁点文字使用率

著者	濁点文字使用率
青大通	100.0%
泉鏡花	100.0%
嵯峨の屋おむろ	99.7%
塚原洪柿園	99.7%
樋口一葉	99.5%
広津柳浪	99.5%
渡部霞亭	99.5%
幸堂得知	99.4%
坪内逍遙	99.3%
幸田露伴	99.2%
須藤南翠	99.2%
宮崎三昧	99.2%
三宅青軒	99.2%
島崎藤村	99.0%
尾崎紅葉	99.0%
大橋乙羽	98.9%
巖谷小波	98.8%
水谷不倒	98.6%
川上眉山	98.5%
前田曙山	98.4%
小金井喜美子	98.4%
福地桜痴	98.1%
饗庭篁村	98.1%
遅塚麗水	97.5%
石橋思案	96.5%
三宅花圃	93.0%

の他の25名はすべて95%以上の高い値となっていることがわかる。小説家の著す記事は濁点文字使用率が高いという傾向は確かに存在するのである。

ここで、『太陽コーパス』の扱う年代に至るまでの明治前期の

文体・表記の様相を簡単にまとめて述べておく。まず文体は、文語が圧倒的に多く、小説と言えども地の文は文語である。ただし、小説で大きな位置を占める会話部分は口語であることも多い。また、表記に使用する仮名の種類では、公的文書や論文・翻訳書等は片仮名を用いた漢字片仮名交じり文で書かれることが多く、雑誌に掲載される論説も同様である。一方小説は、翻訳小説などではない江戸時代の戯作の流れを汲む作品では、平仮名を用いた漢字平仮名交じり文が多い。さらに本稿冒頭にあげた松村（1972）に述べられていたように、漢字片仮名交じり文では概して濁点文字使用率は低い傾向にある。一方小説は、江戸時代の戯作で既にそうであったように、濁点文字を積極的に用い、濁点文字使用率は概して高い傾向にあったと考えられる。以上のことから、特に論説と小説についての文体・表記の様相をまとめれば表9のようになる。

表9 明治前期の文体・表記

文章の種類	文 体	仮名の種類	濁点文字使用率
論説	文語	片仮名	低
小説	文語 (会話部分は口語)	平仮名	高

つまり、明治前期において、「小説＝口語文体＝漢字平仮名交じり文＝高い濁点文字使用率」という意識の繋がりがあったと考えられる。この内の特に「小説＝高い濁点文字使用率」という意識の流れを汲んで、『太陽コーパス』で1895年に既に小説家の著す文語記事の濁点文字使用率が100%近くに達していたものと一応考えることもできよう。しかし、「小説＝高い濁点文字使用率」という意識のみを強調することには慎重を要する。なぜなら、『太陽コーパス』1895年に掲載されている小説家の著す記事の内容は小説ばかりではなく論説記事も多いからである。ここからは「小説＝高い濁点文字使用率」という意識ではなく、あくまでも「小説家の著す文章＝高い濁点文字使用率」という繋がりがしか見えてはこない。さらに、表3・図3で見たように、1895年時点で口語記事の90%以上が濁点文字使用率95-100%であったが、実は、この1895年の口語記事の中に小説はほとんどなく、その多くは演説・談話の速記記事やその文体を模した論説記事なのである。これらの記事の著者（注5）が小説と関係が深かったことを

積極的に示す資料もなく、この口語記事で濁点文字使用率の高い事象も「小説＝高い濁点文字使用率」という意識からだけでは説明することは難しい。それよりも、『太陽コーパス』で口語記事の濁点使用率が高いのは、明治前期には「口語文体＝高い濁点文字使用率」という意識が存し、その流れを汲んでの事象であるからと考えるほうが説明が付きやすい。『太陽コーパス』の1895年で小説家の著す文語記事の濁点文字使用率が高いのは、小説家が小説の会話部分を通して口語文体での濁点文字使用率の高い表記法に慣れており、それを文語記事および論説記事にも適用したためと考えられ、これもまた「口語文体＝高い濁点文字使用率」という意識から説明できる事象なのである。

よって、明治前期におけるもう一方の「論説＝文語文体＝漢字片仮名交じり文＝低い濁点文字使用率」という意識の繋がり内、特に強く働いていたのは「文語文体＝低い濁点文字使用率」の意識であったと推測することができるであろう。『太陽コーパス』の1895年時点では、論説記事を含めほとんどの記事が漢字平仮名交じり文であったり、文語記事の70%近くが濁点文字使用率95-100%であったりと、既に明治前期の表記とは異なった様相を示している。その中で、文語記事の30%以上が濁点文字使用率95-100%に達しておらず、そうした濁点文字使用率の低い記事の著者の多くは口語文体と関係の深くない人物と推測される点に、明治前期の「文語文体＝低い濁点文字使用率」という意識の残存を見出すことができると考える。また、和歌や公的文書といった種類の文章が、後の時期まで濁点文字使用率0%の表記法が規範として保たれていたことも、この種の文章が文語文体で書かれるのが一般的であったことと無関係ではなく、ここにも「文語文体＝低い濁点文字使用率」という意識が見出されると考えられるのである。

7

おわりに

以上、濁点文字使用率という値に基づき、『太陽コーパス』の濁音表記について考察した。考察を通して明らかになった主な点をまとめると次のようになる。

- ① 『太陽コーパス』全体では、1895年に94.5%という比

較的高い濁点文字使用率を示すものの、年を経るに従いさらに使用率は上昇し、1917年に至って濁点文字使用率100%の表記法が定着したと考えられる。

- ② 『太陽コーパス』の口語記事や、口語文体に関係の深い小説家の著す文語記事では、1895年時点で既に濁点文字使用率100%の表記法がほぼ定着している。これは明治前期の「口語文体＝高い濁点文字使用率」という意識の流れを汲んだ結果と考えられる。
- ③ ②の傾向とは逆に、『太陽コーパス』で口語文体と関係の深くない著者による文語記事の中には、濁点文字使用率の低い記事が見出される。これは明治前期の「文語文体＝低い濁点文字使用率」という意識の流れを汲んだ結果と考えられる。
- ④ 多くの種類の文章で濁点文字使用率100%の表記法の定着した時期でも、和歌や公的文書といった一部の種類の文章では、濁点文字使用率0%の表記法が規範として保たれていた。この種の文章は文語文体で書かれるのが一般的であることから、ここにも明治前期からの「文語文体＝低い濁点文字使用率」という意識の流れを見出すことができると考えられる。

『太陽コーパス』の扱う1895年から1925年は、まさに文語文体から口語文体へという言文一致の流れが進んだ時期と一致する。その時期と前後して、表記の面でも、漢字平仮名交じり文への統一などのいくつかの大きな変化が起り、現代日本語の表記に連なる流れが進んだ。濁音表記もまたその流れの中にあって変化してゆき、濁点文字使用率100%の表記法の定着という形に結実してゆく様相が、本稿考察を通して浮かび上がってきたと言える。

濁音表記を含め明治時代以降の表記史は、未だ開拓の余地のある研究分野と思われるが、『太陽コーパス』を資料とすることで、濁音表記以外の問題についても解明されることを期待したい。

注

- (1) 『和英語林集成』『言海』といった明治前期の辞書や『日本国語大辞典 第二版』(小学館)などの現代の辞書の記述、および雑誌『太陽』での実態に鑑み、当時、語形上清濁の揺れがあったと判断した語については、『太陽コーパス』では校訂

は行わず原文のままとした。語形上の揺れと判断した主なものを次にあげる。

「あたかも／あだかも」「缺く／缺ぐ」「しけじけ／しげしげ」「招待^{たい}／招待^{だい}」「すさまじい／すさまじい」「～達^{たち}／～違^{だち}」「立ちとまる／立ちどまる」「願はくは／願はくば」「免かれる／免がれる」「招く／招ぐ」「むづかしい／むづかしい」「若しくは／若しくは」「行きつまる／行きづまる」「指さす／指さす」「ゆるかせ／ゆるがせ」

また、濁音を濁音で繰り返すの字点についても、雑誌『太陽』で「しばぐ」「わぎぐ」と濁点付きくの字点を用いる表記と「しばぐ」「わぎぐ」と濁点のないくの字点を用いる表記とが混在している実態を鑑み、『太陽コーパス』では校訂は行わず原文のままとした。

- (2) 濁点文字使用率を求めるために用いた原文本文および校訂本文の濁点文字数のデータは、小木曾智信氏（明海大学外国語学部講師・国立国語研究所非常勤研究員）によって算出されたものである。
- (3) 用例（3）であげた引用部分も濁点文字使用率0%の表記による和歌であるが、古い時代の文献（の模写）が引用元であることを考慮すると、その文献成立当時の表記における濁点文字使用率が示されたものであって、記事成立当時の和歌の表記における濁点文字使用率が示された例としてとりあげることはできない。
- (4) 千種（1955）によれば、大正14（1904）年の内閣訓令「法令形式ノ改善ニ関スル件」で「現行ノ法文ニ於テハ特殊ナル場合ノ外濁音ノ仮名ヲ用ヒザレドモ、思想表示ノ方法ヲ出来得ル限り正確ナラシメンが（本稿筆者注：千種（1955）の引用ママ）為ニハ一般ニ之ヲ用フベキ」として、法令での濁点文字使用を薦めて以来、法令に濁点を打つことが行われたそうである。
- (5) 速記記事の場合、『太陽コーパス』の記事タグの著者属性には演説者・談話者の名が記されていることもあるが、実質的な著者は速記者であると考えられる。

参考文献

松村明（1972）『国語史概説』（秀英出版）

千種達夫 (1955) 『国語シリーズ25 法令用語の改正』 (文部省)
※再録 (文化庁 『覆刻文化庁国語シリーズⅦ 表現・表記』
149-192頁 教育出版 1974)

仮名遣いについて

———小木曾 智信

1

はじめに

本稿では、『太陽コーパス』における本文中の仮名遣いの問題を取り上げる。雑誌『太陽』の表記は全体としてあまり整備されておらず、仮名遣いについても歴史的仮名遣いが徹底しない複雑な状況を示している（田中・小木曾2000）。しかし『太陽コーパス』では検索性を高めるために本文を修正して歴史的仮名遣いに統一し、その旨を注付けしている。以下はこの注の情報を元にして『太陽』の仮名遣いの実態についてまとめたものである。

2

仮名遣い注付けの方針

『太陽コーパス』の構造化テキスト本文では、今日正しいとされている歴史的仮名遣いに基づいて本文を修正し、修正箇所タグによる注を付けている。構造化テキストにおける注タグは次のような形式で表される。

<s>只今に風俗頹敗といふ聲の<注 原文="絶へ" 分類="G
仮名遣">絶え</注>ず耳朶に響くのみ。</s>
<s>余は此言を誦して<l 位置="P009A17"/>今ま新なる
を<注 原文="覺へ" 分類="G仮名遣">覺え</注>、</s>

（1895年1号「戦勝後の教育」千頭清臣）

一つの注が付けられる範囲は、語を単位としており（振り仮名に対する注の場合は、振り仮名が振られている漢字1字分が単位）、仮名遣い上の誤りを正した注は「分類」属性に「G仮名遣」という値を入れ、「原文」属性の値として元の本文の文字列が保存されている。仮名遣いの典拠は原則として『広辞苑 第五版』（岩波書店）によっている。

文語文と口語文が混在し、多彩な内容の記事が含まれる『太陽

コーパス』全体に通用するルールを設定するために、仮名遣い注(注1)による修正は、通常の仮名遣いに比較していくつもの例外的な処理が行われている。この中には当時の規範的な仮名遣いから見ても誤用と言えないものが多く含まれている。以下では分かりやすさを優先して、あえて「誤用」「誤り」という語を用いるが、「誤用」と言うよりは『太陽コーパス』の仮名遣い規準との「不一致」と表現する方が正確である。たとえば、撥音を表した「む」をすべて「ん」に改めたことによる助動詞「む」の表記の修正などは誤用と呼ぶべきでないものの代表的な例である(「5.2 撥音」参照)。

また、確かに歴史的仮名遣いに照らして誤っている場合であっても、単一の記事の内部では一貫した使い方が行われている場合が少なくない。原文の著者や編集者の誤謬と見なさなければならぬ例はさほど多くないのである。さらにいえば、そもそも原著者に仮名遣いを守ろうという意識がない場合には「誤用」という語はやはりふさわしくないといえよう。

このように、仮名遣い注には単なる誤りと見なすべきではないものがかなり含まれているが、コーパスとしての統一性・一貫性、そして検索性を高めるための便宜的な措置として多くの修正が行われている。原文を尊重するために、注タグによる修正には十分注意を払う必要があるし、注タグの原文属性は最大限に利用されなければならない。

3 仮名遣い誤用数の推移

表1は『太陽コーパス』(注2)の仮名遣い注から本文中の仮名に関するものだけを抜き出し(注3)、歴史的仮名遣いに一致しない仮名の対ごとにその数を集計したものである。振り仮名の仮名遣いについては扱っていない。表の「正:誤」は、原文属性値と本文テキストとの差分を機械的に抽出(注4)したのち、形式を整えたもので、歴史的仮名遣いに一致する仮名と原文で使われていた仮名をコロンで区切って並べている。たとえば「い:ひ」は、「い」と書くべきところを「ひ」と書いた場合で、「或いは」を「或ひは」と書いた例や「報い」を「報ひ」と書いた例などが含まれることになる。「主な例」として挙げたのは、その代表的

なもので、用例数の多いものから選んでいる。

なお、この集計にあたっては次のような前処理を行っている。

- 複数の誤りを含む注は次のように分割して数えた。

例：正・おやぢ／誤・をやじ → お：を 1 例、ぢ：じ 1 例

- 踊り字を使った畳語は 1 例として数えた。

例：正・はふ～～／誤・ほう～～ → はふ：ほう 1 例

- カタカナによって表音的に表された動植物名は集計から除いた。

表 1 仮名遣い誤用数の推移(かな別)

正：誤	1895年	1901年	1909年	1917年	1925年	合計	主な例
い：ひ	77	77	74	22	136	386	或いは・ハ行上一・副詞 つい
い：ゐ	31	21	21	62	49	184	陥いる・いらつしや る・ずいぶん
う：ふ	358	398	199	78	130	1163	ワ行上二(用うる)・形 容詞ウ音便・助動詞う
う：ゆ(注5)	148	192	74	7	10	431	ワ行上二(用うる, 率 うる)
え：へ	446	322	206	5	67	1046	ヤ行下二(見え)・感動 詞ねえ
え：ゑ	80	35	7		6	128	ワ行下二(絶え)・感動 詞ねえ
お：を	24	5	2		7	38	おろか・感動詞おい
おほ：あふ	3					3	仰せ
かう：こう	1		14	2	12	29	斯う・助動詞う(行か う)
さう：そう	14	62	107	5	17	205	助動詞さうだ・然う
さう：そふ		17	1			18	然う
じ：ぢ	14	10	12	104	61	201	じつと
ず：づ	7	7	11	8	22	55	むず痒い・ずつと
せう：しやう	32	67	61	1	19	180	でせう・ませう
せう：しょう	3	47	3	4	38	95	でせう・ませう
そう：さう	1				1	2	大層
そふ：さう					1	1	添ふ
たう：とう	12	12	28	69	43	164	到頭・本当
たう：とふ		2				2	到頭
たふ：とう	1					1	貴い

正：誤	1895年	1901年	1909年	1917年	1925年	合計	主な例
ぢ:じ	205	58	34	8	22	325	助動詞ぢや・ぢやあ
ちやう:てう		1				1	提灯
ちゆう:ちう	2	1	1	2	7	13	中(ちゆう)
づ:ず	5	12	3	7	8	35	先づ・弾む
とう:たう					3	3	到頭
なう:のう	8		1	76		85	終助詞なう
なう:のふ			2			2	終助詞なう
のふ:なふ				1		1	昨日
はう:ほう	3		3		1	7	助動詞う(貰はう)・介抱
ばう:ぼう	1	1	3		3	8	坊・副詞ばうと
は:わ	33	18	48		24	123	ハ行四段(行は、能は)・俄
はふ:ほう			1			1	這々
ひ:い	35	42	101	52	100	330	ハ行連用形(強ひ、貰ひ)
ひ:ゐ	8	17	11	2	5	43	ハ行連用形(強ひ、違ひ)
ふ:う	150	213	255	503	323	1444	ハ行四段音便形
ふ:ほ			2		1	3	倒れる・煽る
ふ:ゆ	141	90	91	26	3	351	ハ行下二(堪ふる、教ふる)
ふ:を		1				1	煽り
へ:え	192	216	191	42	23	664	ハ行下二(堪へ、教へ)
へ:ゑ	2	5			9	16	ハ行下二(押さへ、堪へ)
ほ:う	1	5	1			7	通す・催す
ぼふ:ぼふ					1	1	貧乏
ほ:ふ	16	12	30		3	61	催す・通す
ほ:を	6	1			1	8	直す
まう:もう	1	1	2		6	10	儲け
まう:もふ	1				1	2	儲け
まふ:もう					3	3	～ちまふ(てしまふ)
やう:よう	3	11	85	35	541	675	助動詞やうだ・斯様
やう:よふ		1				1	助動詞やうだ

正：誤	1895年	1901年	1909年	1917年	1925年	合計	主な例
ゆ：う	1					1	ヤ行上二(酬ゆ)
ゆ：ふ	19	6	1	3		29	ヤ行上二(覚ゆ, 酬ゆ)
よう：やう	141	239	531	350	202	1461	助動詞よう
よう：やふ			1			1	助動詞よう
らう：ろう	20	29	104	2	5	160	助動詞う(あらう, だらう)
れう：りやう	1					1	助動詞う(笑われう)
ろふ：らふ	1	1			1	3	移ろふ・繕ふ
ろふ：らう		1				1	カゲロフ
わ：は	20	8	14	5	4	51	ことわり・ざわめき
ゐ：い	16	20	11	1	7	55	用ゐる・位・参る
ゐ：ひ	290	262	149	147	275	1123	ワ行上一(用ゐ, 率ゐ)
ゑ：え	36	25	49	7	1	118	ワ行下二(植ゑ, 据ゑ)
ゑ：へ	35	60	24			119	ワ行下二(据ゑ)
を：お	56	30	23	4	5	118	をかしい・(況やへ)をや
を：ほ		1				1	香る
ん：む	1156	2377	697	197	244	4671	助動詞む・撥音便
計	3857	5039	3288	1835	2451	16470	

注の数の単純合計を元に仮名遣い誤用数の推移を見ると、1895年から1901年にかけていったん増加し、その後1917年まで減少した後、1925年には再び増加している。このような複雑な変化をもたらすことになった要因は何であろうか。

3.1 本文文字数・漢字含有率と仮名遣い誤用数

本文の文字数が多いほど仮名遣いの誤りの例が多く見られることになるのはいうまでもないが、このほかの要因として、語が仮名書きされればされるほど、また、送り仮名が送られるほど仮名遣いの誤りが生じやすくなることがあげられる。つまり仮名遣いの誤りは、記事の文字数、漢字含有率に影響されるわけである。そこで、各年の文字数合計と漢字含有率を調査したものが表2である。

表2 各年の文字数(文字種別)と漢字含有率

	ひらがな	カタカナ	漢字	記号類	合計	漢字含有率
1895年	1,422,234	50,155	1,622,497	212,325	3,307,211	49(52)%
1901年	1,408,398	58,575	1,528,316	223,129	3,218,418	47(51)%
1909年	1,404,555	56,461	1,385,510	227,168	3,073,694	45(49)%
1917年	1,386,914	46,412	1,310,873	188,125	2,932,324	45(48)%
1925年	1,666,075	92,128	1,358,269	266,947	3,383,419	40(44)%

(漢字含有率の括弧内の数字は記号類を分母から除いた場合)

文字数は、1917年分がやや少なく、1925年がやや多いが、総じて大きな差はない。ところが漢字含有率は一貫して下がっており、1917年から1925年にかけては5%ほども低下している（注6）。漢字含有率の低下は、仮名書きされる語が増えていること、送り仮名が送られる場合が増えていることによるものである。時代が下れば下るほど仮名遣いの問題が表面化しやすくなっているのである。

仮名遣いは年を追うごとに歴史的仮名遣いへと統一が進んでいくと考えられ、実際に1901年から1917年にかけて誤用は一貫して減少してゆくのであるが、1895年から1901年にかけてと、1917年から1925年にかけての二度にわたって逆に増加している。この原因の一つは、この仮名書きの増加にあるといえそうである。特に1925年にかけての増加は、仮名書きが増えたことによる誤用の増加が、仮名遣いが統一に向かうことによる減少幅を上回ってしまったことによるものだと見える。

1917年ごろに仮名遣いの統一が一段落したことは、振り仮名の字音仮名遣いの調査からも裏付けられる。『太陽コーパス』の字音仮名遣いを調査した小木曾（2003）の結果では、1909年までは全ての字音の振り仮名のうち4%以上で字音仮名遣いの誤りが見られるのに対し、1917年にはこれが1.7%に急減している。また、仮名遣いの問題を生じる字音の振り仮名に限った場合でも、1909年まで10%以上だった誤りが1917年には4.2%まで低下している（注7）。1917年ごろには仮名遣いの揺れが収まってきており、歴史的仮名遣いがかかなり定着したものと考えられるのである。

1925年の増加の具体的な理由としては、この年に「ようだ」が極端に多いことが挙げられる。これは三上於菟吉の小説で助動

詞「やうだ」が一貫して「ようだ」で書かれているためである。しかし、これも古くは「様だ」と漢字で書かれていたものが仮名書きされるようになったことに起因するのであるから、「やうだ」が多いという個別事情を汲むにしても、やはり原因は仮名書きの増加というところに帰着することになる。

なお、これには口語文記事の増加も影響していると考えられる。「やうだ」の他にも助動詞の「よう」「やうだ」「さうだ」や「ませう」などの仮名遣い誤用が増加しているが、こうした語のほとんどは口語文で用いられるものである。そもそも歴史的仮名遣いは文語文に用いられることを前提としており、口語文への適用には無理が伴う。歴史的仮名遣いの文語文においては起きえなかったような誤用が、口語文に適用されるに及んで初めて現れるようになったものと考えられる。

4 記事・著者別に見た仮名遣いの誤用

つづいて、仮名遣いの誤用について記事・著者別の観点から検討してゆくことにする。

4.1 仮名遣いの誤用の多い記事

仮名遣い注の記事別の分布を調査するため、記事ごとの文字数とそこに現れる仮名遣い注の数を取り出し（注8）、誤用の頻度を1000文字あたりの仮名遣い注の数で表して一覧を作成した。その結果得られた頻度の平均値は表3の通りである。

表3 仮名遣い誤用頻度の推移（1000文字あたりの仮名遣い注の数）

	1895年	1901年	1909年	1917年	1925年	総計
全記事平均	1.11	1.68	1.03	0.49	0.61	0.96
(1万文字以上の記事の平均)	(1.63)	(1.38)	(0.96)	(0.7)	(0.8)	(1.08)

また、表4は1万字以上の長文記事に絞って、誤用の頻度の高いものから順に30記事を挙げたものである。表の「注数」は記事中の仮名遣い注の数、「頻度」は1000文字あたりに換算したときの仮名遣い注の数である。

表4 仮名遣いの誤用の多い記事(1万字以上の記事から上位30記事)

年	号	タイトル	著者	欄名	文体	注数	文字数	頻度
1895	10	心中女	幸堂得知	小説	文語	153	15387	9.94
1901	7	姉崎嘲風に与ふる書	高山樗牛	文芸時評	文語	115	15494	7.42
1895	6	涙の媒介	条野採菊	小説	文語	100	14232	7.03
1909	4	西間島事情	村田懋麿	雑纂	口語	71	10378	6.84
1901	10	文芸時評	高山樗牛	文芸時評	文語	77	11644	6.61
1925	9	長篇小説 蛇人 (第七回)	三上於菟吉	**	口語	78	12198	6.39
1901	14	文芸時評	高山樗牛	文芸時評	文語	68	11535	5.90
1895	6	新袈裟物語	宮崎三味	小説	文語	103	18435	5.59
1901	8	文芸時評	高山樗牛	文芸時評	文語	67	12275	5.46
1895	12	滑稽 道中すごろく	南新二	小説	文語	91	18083	5.03
1901	14	外資輸入国としての 日本	近藤廉平(口述)	論説	文語	52	11015	4.72
1895	9	十三世紀に於ける 蒙古民族の雄図	田岡嶺雲	史伝	文語	57	12721	4.48
1895	6	亜歴セルカーク(下)	森田思軒	史伝	文語	55	12664	4.34
1901	5	文芸時評	大町桂月	文芸時評	文語	42	10075	4.17
1901	13	文芸時評	高山樗牛	文芸時評	文語	60	14599	4.11
1925	12	長篇小説 蛇人 (第十回)	三上於菟吉	**	口語	55	13428	4.10
1901	10	追懐談	川村純義(談)	歴史地理	口語	62	15203	4.08
1895	11	浮世新聞	嵯峨の屋おむろ	小説	文語	47	11841	3.97
1895	3	昭君怨	巖谷小波	小説	文語	53	13367	3.96
1895	9	夜の鶴(下)	福地桜痴	小説	文語	74	18781	3.94
1901	4	鎮西遊記(接前号)	久保天随	歴史地理	文語	96	24684	3.89
1895	1	フートルロー合戦 の記	戸川残花	史伝	文語	57	14775	3.86
1901	3	鎮西遊記	久保天随	歴史地理	文語	70	19290	3.63
1925	14	長篇小説 蛇人 (第十二回)	三上於菟吉	**	口語	44	12269	3.59
1895	8	夜の鶴(上)	福地桜痴	小説	文語	35	10093	3.47
1901	2	文芸時評	大町桂月	文芸時評	文語	41	11867	3.45
1917	1	失はれた原稿	里見弴	新年大附録	口語	41	12273	3.34
1909	14	実印と預金帳	柴田流星	文芸	口語	55	17881	3.08
1901	3	文芸時評	大町桂月	文芸時評	文語	37	12194	3.03
1901	4	文芸時評	大町桂月	文芸時評	文語	32	10606	3.02

表4から、仮名遣い注が多い記事にはかなりはつきりとした傾向が見てとれる。著者別に見ると、高山樗牛の文芸時評にはかなり誤りが多く、表中5回も現れている。大町桂月の文芸時評もそれにつぐ。ジャンルとしては文芸評論・小説・史伝の記事が目立つ。年ごとに見るとやはり1895年と1901年が多いが、三上於菟吉の小説は特徴的で、1925年のものであるにもかかわらず誤用が多い。

高山樗牛・大町桂月と同じ1901年の記事について見てゆくと、同じ年であっても水島鉄也・祖山鍾三・佐野善作の「商業世界」は、1号の誤用頻度が0.16 (2/12888), 2号が0.55 (7/12689), 13号が0.22 (3/13877), 14号が0.09 (1/11320) といった具合で、どの号もきわめて誤りが少ない。金子篤寿(鴨居武・塚本信治)の「工業世界」は両者の中間的なところで、1号が0.52 (7/13457), 2号が0.36 (6/16445), 13号が1.20 (24/20067), 14号が1.99 (32/16104) と1901年の記事の平均程度である。

こうしてみると、1901年ではジャンルや記事内容によって仮名遣いの誤りの数が左右されているようである。「商業世界」などの論説文では仮名書きされる語も少ないうえに、述語の形に定型の言い回しが多い。それに対して「文芸時評」などは内容的にその対極にあり、文学的な表現を駆使した文章であればあるほど仮名遣い注が多く現れている。

ただし、高山樗牛・大町桂月の場合には撥音「む」表記の問題(4.2撥音)が関わっているのが明らかであり、仔細に見るならばそれぞれの記事で固有の事情が誤用数に影響しているようである。

4.2 仮名遣いの誤用が多い著者

『太陽』では毎号、特定の欄を同一著者が担当する結果、ジャンルと著者が一致してしまう例が少なくない。そのため両者を切り離して論じることは難しいが、今度は著者の側から誤用の多いものを調査した。あわせて1万字以上の記事が収録されている著者の中から、1000文字あたりの仮名遣い注数が多いもの上位30人分をまとめたものが表5である。

表5 仮名遣いの誤用が多い著者(計1万字以上の記事を書いた著者から上位30人)

著者	1000文字あたりの仮名遣い注の数						仮名注計	総文字数
	全体	1895年	1901年	1909年	1917年	1925年		
幸堂得知	7.00	7.00	-	-	-	-	167	23845
落合直文	6.99	8.99	3.53	-	-	-	119	17027
高山樗牛	5.89	0.83	6.45	-	-	-	718	121947
条野採菊	5.72	5.72	-	-	-	-	107	18709
中内蝶二	5.64	-	5.64	-	-	-	67	11871
宮崎三味	5.59	5.59	-	-	-	-	103	18435
村田懋麿	5.42	-	-	5.42	-	-	108	19920
南新二	5.03	5.03	-	-	-	-	91	18083
塚原洪柿園	4.98	4.85	-	5.25	-	-	74	14858
近藤廉平(口述)	4.72	-	4.72	-	-	-	52	11015
田岡嶺雲	4.48	4.48	-	-	-	-	57	12721
戸川残花	4.21	4.21	-	-	-	-	101	23963
別役成義(談)	4.16	-	4.16	-	-	-	124	29816
桐生悠々	4.08	6.20	3.30	-	-	-	103	25237
川村純義(談)	4.08	-	4.08	-	-	-	62	15203
嵯峨の屋おむろ	3.97	3.97	-	-	-	-	47	11841
久保天随	3.77	-	3.77	-	-	-	166	43974
三上於菟吉	3.51	-	-	-	-	3.51	423	120495
尾崎紅葉	3.45	3.45	-	-	-	-	39	11288
大町桂月	3.41	-	3.63	2.19	0.88	3.37	596	174888
泉鏡花	3.25	11.68	-	2.07	-	-	70	21516
芥洲学人	3.21	-	-	3.21	-	-	61	18995
柴田流星	3.08	-	-	3.08	-	-	55	17881
森田思軒	3.07	3.07	-	-	-	-	106	34473
戸水寛人	3.07	-	3.07	-	-	-	52	16936
中根寿	3.01	3.01	-	-	-	-	40	13272
長谷川天溪	2.93	-	2.73	3.03	-	1.08	372	126866
手島精一	2.93	2.15	3.57	-	-	-	36	12295
草野紫二※	2.74	-	-	2.74	-	-	36	13143
吉岡芳陵	2.70	-	2.35	3.88	1.11	-	57	21095

※アナトール・フランスの翻訳

やはり1895年に文学的な内容の記事を書いている著者が目立つ結果となっており、仮名遣い注の多さは文学というジャンルに関わりが深いことがわかる。これには、文学作品には口語文が多く漢字含有率も低いといった「2. 仮名遣い誤用数の推移」で見た要因も関わっているものと思われる。

5

仮名別に見た仮名遣いの誤用

次に、仮名（音）別にみた仮名遣いの誤用の様子について見てゆくことにする。ここでは四つ仮名、撥音、ア・ハ・ワ行音について概観する。

5.1 四つ仮名

まず、四つ仮名の揺れについて見てゆくが、「じ」「ぢ」「ず」「づ」それぞれの仮名の使用頻度には大きな差があるため、仮名遣い注の数だけを見ているだけでは全体の傾向を見誤ることになる。そこで、『太陽コーパス』で用いられている「じ」「ぢ」「ず」「づ」の平仮名の総数を調査してまとめたものが表6である。

表6 仮名「じ」「ぢ」「ず」「づ」の用例数

	1895年	1901年	1909年	1917年	1925年	合計
じ	4167	3541	3234	3340	3540	17822
ぢ	523	666	1229	2501	1191	6110
ず	16946	15948	10945	7010	4694	55543
づ	2117	1733	1617	1406	1758	8631

ザ行の仮名に比較してダ行の仮名はもともとかなり数が少ない。また、「じ」や「づ」の用例数が比較的安定しているのに対し、「ず」が急激に減少してゆき、「ぢ」が1917年まで増加してゆくのが目を引く。「ず」の減少は文語文の減少に伴い打消の助動詞「ず」が大幅に減少してゆくこと、「ぢ」の増加は口語文の増加により助動詞「ぢや」などが増加してゆくことによるものと考えられる。

表7は「じ・ぢ」の仮名について、誤用の数と割合をまとめたものである（上段が誤用数、下段が誤用の割合）。ここでの誤用

の割合とは、すべての「じ」のうち「ぢ」と誤ったものの割合（またはその逆）をパーセンテージで表したものである。

表7 「じ・ぢ」の誤用の数と割合

正：誤	1895年	1901年	1909年	1917年	1925年	合計
じ：ぢ	14	10	12	104	61	201
	0.34%	0.28%	0.37%	3.11%	1.72%	1.13%
ぢ：じ	205	58	34	6	22	325
	39.20%	8.71%	2.77%	0.24%	1.85%	5.32%

「じ：ぢ」の中では、特に1917年の「じ：ぢ」が目立つが、このうち98例までが副詞「じっと」を「ぢつと」と表記した例である。『日本国語大辞典 第2版』「じっと」の項の補注に「かなづかいはいは、古い用例では『じっと』が多く、江戸期から明治にかけて『ぢつと』も多くなる」とあるが、『太陽コーパス』でも副詞「じっと」の仮名遣いは「ぢつと」で定着しており、他の語の仮名遣いの揺れが収まってきた1917年以降にも「ぢつと」の形で用いられていたために、特に目立つ結果になっているものと思われる。残りは「かぢる」「みぢめ」などの語であるが、これらはほぼ一定の割合で出現しており大きな変化は認められない。

「ぢ：じ」では1895年の誤用が突出しているが、これは小説記事の会話部分（口語文）で、助動詞「ぢや」が非常に多く現れ、なおかつそれが「じや」と表記されているためである。塚原洪柿園「他流試合」（5号）、条野採菊「涙の媒介」（6号）、宮崎三味「新袈裟物語」（6号）、福地桜痴「夜の鶴（上・下）」（8号・9号）などで軒並み30例近くの「じや」が用いられており、これが誤用の数を押し上げている。

表8は「ず・づ」の仮名について、誤用の数と割合をまとめたものである。

表8 「ず・づ」の誤用の数と割合

正：誤	1895年	1901年	1909年	1917年	1925年	合計
ず：づ	7	7	11	8	22	55
	0.04%	0.04%	0.10%	0.11%	0.47%	0.10%
づ：ず	5	12	3	7	8	35
	0.24%	0.69%	0.19%	0.50%	0.46%	0.41%

「ず・づ」に関する誤用は「じ・ぢ」に関するものと比較してかなり少なく、各年を比較してもほぼ一定している。1925年のみ誤用が多いが、これはこの年の小説に「づつと」「むづむづ」と書かれる用例が集中しているためである。「ず：づ」と「づ：ず」を比較すると、前者の誤用の割合がほぼ0.1%以下で推移しているのに対し、後者は0.4%ほどに達しておりやや誤用が目立つ。この理由は、前者の「ず」の仮名の用例が圧倒的に多く、しかも助動詞「ず」のように誤って表記する可能性がきわめて低い語が多いことによるものだと考えられる。「ず：づ」の誤用の主なものは「むずがゆい」「ずっと」「ずぶぬれ」,「づ：ず」では「先づ」「はづむ」「うづくまる」などであった。

5.2 撥音

『太陽』では撥音の表記に「む」と「ん」がともに用いられているが、どちらが使われるかは年や記事などではっきりと区分することはできない。たとえば撥音便の「飛んで」「富んで」「偲んで」などは「飛むで」「富むで」「偲むで」の形でも用いられているが、両者が混在したままでは検索において不便である。そのため、『太陽コーパス』では撥音を表した「む」はすべて「ん」に改めた。これに伴い、撥音で読んだと考えられる推量の助動詞「む」はすべて「ん」に改めることとなった。

その結果、表9に再掲するように多くの仮名遣い注が付けられている。表中の「ん」は、もともと「ん」と書かれていた仮名の総数である。1901年までは「む」表記も多いものの、「ん」で表記したものの方が各年を通して圧倒的に多いことがわかる。

表9 撥音「む」表記の数と仮名「ん」の用例数

	1895年	1901年	1909年	1917年	1925年	合計
ん:む	1156	2377	697	197	244	4671
「ん」	12745	10642	11944	11664	13337	65003

「む」による撥音表記は1895年から1901年にかけて倍増し、その後急減している。1901年に撥音を「む」で表記したものが多くのは、「む」表記の原則を貫く特定の著者が多くの記事を書いているためである。最も多いのは高山樗牛で、「文明批評家としての文学者」「王朝の絵画」「平家雑感」他5記事の中で634回、次いで大町桂月が連載中の「文芸時評」「教育時評」16記事

の中で503回使用しており、これだけでほぼ半数に達する。

「む」表記の原則を貫く著者は1925年にも存在し、上杉慎吉37例（3号「帝国憲法の過去及現状と普通選挙に対する準備」）、添田寿一65例（5号「憲政の危機と対策」、13号「日本の労働問題」）などに用例が多い。記事の口語化により助動詞「む」の出身が減るために用例数は減っていくのだが、撥音を「む」で書くという表記法は1925年に至るまで一定の割合で生き残っていたようである。

5.3 ア・ハ・ワ行の仮名遣い

つづいて、ア行・ワ行・ハ行の仮名について見てきたい。ただし、この中でも誤用が多く問題が複雑な動詞の活用については「5. 動詞の活用と仮名遣い」で扱うこととする。

「四つ仮名」の場合と同じく、問題になる平仮名全ての用例数を出したものが表10である。ただし、ここでは語頭であるか語中・語尾であるかを問わず、全ての使用仮名を分母としているので、実際には仮名遣いの問題が起きない場合も含まれていることになる。また、「を」や「は」の仮名については助詞の「を」「は」が含まれるために実際には仮名遣いの問題が起きない用例が多く含まれている。

表10 仮名「いうえお・はひふへほ・わゐゑを」の用例数

	1895年	1901年	1909年	1917年	1925年	合計
い	11189	12794	26318	30669	47669	128639
う	4731	5834	11747	14781	20400	57493
え	1814	2038	2389	1984	2762	10987
お	2855	2385	3322	3284	4427	16273
は	72755	74050	73157	69039	83284	372285
ひ	7846	5162	4206	4689	5749	27652
ふ	14089	12755	15312	14025	15586	71767
へ	9716	8264	8243	8990	11347	46560
ほ	1702	1464	1305	1437	1766	7674
わ	708	627	710	990	2161	5196
ゐ	836	968	2352	5959	13684	23799
ゑ	374	194	183	138	179	1068
を	93268	88898	72264	69145	66837	390412

表10の中では「ゐ」の仮名が急増しているのが特に目を引くが、これは動詞（補助動詞）「ゐる」が仮名書きされるようになることが大きな要因である。「い」「う」の仮名も増えているが、この理由としては口語文記事の増加によって音便形が増大したことなどが考えられる。

表10と表1の数値を元に、仮名別の誤用の数と割合を出したのが表11である。

表11 仮名「いうえお・はひふへほ・わゐゑを」の誤用の数と割合

正：誤	1895年	1901年	1909年	1917年	1925年	合計
い：ひ	77	77	74	22	136	386
	0.69%	0.60%	0.28%	0.07%	0.29%	0.30%
い：ゐ	31	21	21	62	49	184
	0.28%	0.16%	0.08%	0.20%	0.10%	0.14%
う：ふ	358	398	199	78	130	1163
	7.57%	6.82%	1.69%	0.53%	0.64%	2.02%
え：へ	446	322	206	5	67	1046
	24.59%	15.80%	8.62%	0.25%	2.43%	9.52%
え：ゑ	80	35	7		6	128
	4.41%	1.72%	0.29%	0.00%	0.22%	1.17%
お：を	24	5	2		7	38
	0.84%	0.21%	0.06%	0.00%	0.16%	0.23%
お：ほ						
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
は：わ	33	18	47		24	122
	0.05%	0.02%	0.06%	0.00%	0.03%	0.03%
ひ：い	35	42	101	52	100	330
	0.45%	0.81%	2.40%	1.11%	1.74%	1.19%
ひ：ゐ	8	17	11	2	5	43
	0.10%	0.33%	0.26%	0.04%	0.09%	0.16%
ふ：う	150	213	255	503	323	1444
	1.06%	1.67%	1.67%	3.59%	2.07%	2.01%
へ：え	192	216	190	42	23	663
	1.98%	2.61%	2.30%	0.47%	0.20%	1.42%
へ：ゑ	2	5			9	16
	0.02%	0.06%	0.00%	0.00%	0.08%	0.03%
ほ：お						
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
ほ：を	6	1			1	8
	0.35%	0.07%	0.00%	0.00%	0.06%	0.10%
わ：は	20	8	14	5	4	51
	2.82%	1.28%	1.97%	0.51%	0.19%	0.98%
ゐ：い	16	20	11	1	7	55
	1.91%	2.07%	0.47%	0.02%	0.05%	0.23%

正：誤	1895年	1901年	1909年	1917年	1925年	合計
ゐ：ひ	290	262	149	147	275	1123
	34.69%	27.07%	6.34%	2.47%	2.01%	4.72%
ゑ：え	36	25	49	7	1	118
	9.63%	12.89%	26.78%	5.07%	0.56%	11.05%
ゑ：へ	35	60	24			119
	9.36%	30.93%	13.11%	0.00%	0.00%	11.14%
を：お	56	30	23	4	5	118
	0.06%	0.03%	0.03%	0.01%	0.01%	0.03%
を：ほ		1				1
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

誤りの割合は仮名ごとにかなり大きな違いがあり、あわせて2割が誤用となる「ゑ」と、1割が誤用の「え」、5%が誤用の「ゐ」の3つは特に誤用が多い。逆に「は」「を」などは助詞としての多数の用例が誤りなく用いられることから、誤用の割合はきわめて低くなっている。

増減についてみると、「う：ふ」「え：ゑ」が急速に減少し、「い：ひ」「へ：え」「わ：は」「え：へ」をはじめとするほとんどのものが減少傾向を示す。こうした中で「ふ：う」「ひ：い」だけが増加傾向にあることが注目される。「ふ：う」は、後述するように、ハ行に活用する動詞の音便形について「言ふて」「養ふて」の形を正用とし、「言うて」「養うて」を修正としたことによるものである。「ひ：い」は、「入る」を「はいる」と仮名書きするものが増えたことが影響している。

6

動詞の活用と仮名遣い

動詞の活用語尾の仮名遣いは、用例数から見ても仮名遣いに関するもっとも大きな問題であるといえよう。ここには動詞の活用型とその歴史的変化という、仮名遣いの問題を越えたものまでが含まれており、たいへん複雑な様相を示している。以下ではこの問題について、動詞の活用型別に代表的な動詞を取り上げ、活用形別にどのような形がどれだけ現れているのかという点に絞って見てゆくことにする。

本来であれば、各用例の著者・文体・ジャンルなどのさまざまな情報を元に多角的に分析してみる必要のある問題であるが、仮

名遣いの一角として扱うにはあまりにも大きな問題であり到底ここでは扱いきれない。改めて別の機会に詳しい検討を行いたいと考えている。

なお、活用形ごとにまとめた表の用例数は、本来の活用形ではなく、語形を元にして分類した。一・二段活用の未然形・連用形、一段活用の終止形・連体形は本来なら区別しなければならないものであるが、『太陽コーパス』では活用形のタグ付けなどを行っていないため、この区別を行うにはたいへんな手間がかかる。そこで、ここでは本来の活用形ではなく簡単に比較できる語形の問題として処理した。また、以下の表に現れる用例数の中には、動詞起源の名詞や副詞も一部含まれている。

6.1 四段活用動詞

四段活用の動詞で特に仮名遣いが問題になるのは八行四段活用動詞のウ音便形である。たとえば動詞「買う」に付けられた仮名遣い注はすべて「買う」と表記されたもので、計38例ある。しかし、この中で終止形・連体形「買ふ」を誤ったものは12例にすぎず、残り26例は接続助詞「て」や助動詞「た」に続くウ音便形である。歴史的仮名遣いとしてはウ音便の形は本来「買うて」「買うた」と表記すべきであるが、『太陽コーパス』の用例では「買ふて」「買ふた」と表記されたものの方が多く、33例ある。そのため、こちらを正用と認めて、逆に「う」で書かれたもの26例を修正しているのである。「買う」の場合は用例数にさほど大きな差はないが、他の動詞では「ふ」で書かれたものの方が圧倒的に多い場合があり、たとえば「云ふ」では「云うて・云うた」69例に対して「云ふて・云ふた」298例となっている。このような状況をふまえ、「ふ」を正用と認めたものである。「ふ」で書かれたウ音便形の方が多いという傾向は他の八行四段活用動詞でも変わらないため同様の処理を行っている。

表12は『太陽コーパス』中の動詞「買う」の用例数を語形・活用形別にまとめたものである。

なお、表中の斜体で示した語形は、『太陽コーパス』で正用と認めていて仮名遣い注が付けられていないものである。

表12 「買ふ」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ハ行四段	かは 147	かひ 242	かふ 208 (かう12)		かへ 13	
		かつ 220				
		かふ 33				
		かう 26				

6.2 一段活用動詞

一段活用の動詞で仮名遣いが問題になるのはワ行上一段の動詞である。ここでは「用ゐる」と「率ゐる」を取り上げる。

表13は、動詞「用ゐる」の用例数と、それにつけられた注の数を語形・活用形別にまとめたものである。表中の「ワ行上一」といった活用型はその形を表中に位置づけるために用いたものであり、ワ行に活用していると認めたことを意味するわけではない。

表13 動詞「用ゐる」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ワ行上一	もちゐ 750	もちゐる 154		もちゐれ 11		もちゐよ
上二		もちう 3	もちうる 30	もちうれ 23		12
ハ行上一	もちひ 955	もちひる 51		もちひれ 0		もちひよ
上二		もちふ 132	もちふる 401	もちふれ 20		3
ヤ行上一	もちい 18	もちいる 0		もちいれ 0		もちいよ
上二		もちゆ 99	もちゆる 265	もちゆれ 0		0

『太陽コーパス』では表中の語形のうち、ワ行上一・上二の行以外は、すべて誤りと見なして仮名遣い注が付けられている。「用ゐる」は文語で上一段活用の語であるから、ワ行上二段の「もちう」「もちうる」「もちうれ」は本来であれば誤用ということになる。しかし『太陽コーパス』ではこれらの語形が多く用いられていることから、ワ行上二段の語「もちう」を認め、この形を「もちゆ」「もちふ」の規範形としている。これは、一段化をおこす二段活用動詞の場合に、一段・二段どちらの形も正しい形として認めたのにあわせた便宜的措置である。二段活用の一段化は通常のことであるが、一段活用の二段化したものを認めるのは同列に扱える問題ではない。歴史的に見るならば、むしろワ行上二段から転じたハ行上二段活用（ないしヤ行上二段活用）を正しい語形とすべきかもしれない。しかし、これを認めると表中のすべての語形が正しい語形ということになり、仮名遣いの統一ができなくなるのである。

また、表中のヤ行上二段「もちゆ」「もちゆる」「もちゆれ」は「モチウ」とは読まれず、本来であれば仮名遣いの問題として処理できない語形である。しかし上述の便宜的措置にあわせ、これらの語形も「もちう」の形に修正されている。

「用ゐる」の各語形の分布を活用の行別に見ると、本来のワ行が33.6%、ハ行53.4%、ヤ行13.1%となっており、ハ行がワ行を凌駕している。一段活用・二段活用に分けると、ワ行では一段活用が多く、ハ行やヤ行では二段活用が圧倒的に多い。もともとの形であるワ行一段が残存する一方で、後に変化して定着したハ行二段・ヤ行二段が広く使われながら併存している状況であるといえそうである。

表14は、ワ行上一段活用動詞のもう一つの例として「率ゐる」の用例数をまとめたものである。

各語形の分布を活用の行別に見ると、本来のワ行が67.6%、ハ行20.6%、ワ行11.8%となっており、「用ゐる」とは異なって本来の形であるワ行が多く用いられている。一段活用・二段活用に分けると、ワ行では一段活用が多いのに対し、ハ行・ヤ行では二段活用がほとんどである。ここでも、本来のワ行では一段、それ以外では二段の形が用いられるという傾向は変わりがない。

本来の形であるワ行一段が広く用いられているが、後に定着したハ行二段・ヤ行二段も併存しているという状況である。

表14 動詞「率ゐる」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ワ行上一	ひきゐ	221	ひきゐる 50		ひきゐれ 0	ひきゐよ 0
上二			ひきう 2	ひきうる 3	ひきうれ 0	
ハ行上一	ひきひ	73	ひきひる 1		ひきひれ 0	ひきひよ 3
上二			ひきふ 0	ひきふる 10	ひきふれ 0	
ヤ行上一	ひきい	11	ひきいる 0		ひきいれ 0	ひきいよ 0
上二			ひきゆ 2	ひきゆる 35	ひきゆれ 0	

6.3 二段活用動詞

つづいて、文語におけるハ・ヤ・ワ行二段活用動詞について、代表的な動詞を取り上げて、語形・活用形ごとに用例数を見てゆくことにする。ワ行一段活用動詞ほどではないが、これらの動詞もさまざまな表記がなされており、複雑な使用実態を示している。

なお、ハ行・ワ行に活用する動詞が、ヤ行二段の終止形「～ゆ」・連用形「～ゆる」・已然形「～ゆれ」の形をとったものは仮名遣いを逸脱した語形である。しかしワ行上一段活用の場合と同様、『太陽コーパス』ではこれも仮名遣い注で処理しているため、ここでもそれに準じた扱いとする。

ハ行上二段活用動詞

表15は、ハ行上二段活用動詞の例として「強ふ」の活用形別用例数をまとめたものである。ここには副詞的な「強ひて」も用例に数えているため、連用形の用例数が多くなっている。活用の行別に見ると、本来のハ行が71.6%で、ヤ行22.9%，ワ行5.6%となっている。

表15 動詞「強ふ」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ハ行下一	しひ 222		しひる 10		しひれ 0	しひよ 0
下二			しふ 2	しふる 10	しふれ 0	
ヤ行下一	しい 60		しいる 1		しいれ 0	しいよ 0
下二			しゆ 4	しゆる 13	しゆれ 0	
ワ行下一	しゐ 19		しゐる 0		しゐれ 0	しゐよ 0
下二			しう 0	しうる 0	しうれ 0	

ヤ行上二段活用動詞

表16は、ヤ行上二段活用動詞の例として「むくゆ」の活用形別用例数をまとめたものである。名詞「報ひ」は連用形「むくひ」として扱った。活用の行別に見ると、本来のヤ行が71.5%、ハ行25.4%、ワ行3.1%となっている。

表16 動詞「むくゆ」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ヤ行下一	むくい 36		むくいる 4		むくいれ 0	むくいよ 1
下二			むくゆ 7	むくゆる 45	むくゆれ 0	
ハ行下一	むくひ 30		むくひる 0		むくひれ 0	むくひよ 0
下二			むくふ 0	むくふる 3	むくふれ 0	
ワ行下一	むくゐ 3		むくゐる 0		むくゐれ 0	むくゐよ 0
下二			むくう 0	むくうる 1	むくうれ 0	

ここでとりあげた「強ふ」「むくゆ」の2語を見る限り、上二段活用動詞は7割ほどが正しい仮名遣いで用いられており、2割強がヤ・ハ行の間で誤用され、ワ行に誤るものは数パーセントとなっている。

ハ行下二段活用動詞

表17は、ハ行下二段活用動詞の例として「教ふ」の活用形別用例数をまとめたものである。活用の行別に見ると、本来のハ行が88.3%，ヤ行10.7%，ワ行1.0%となっている。

表17 動詞「教ふ」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ハ行下一	をしへ 458		をしへる 63		をしへれ 1	をしへよ 0
下二			をしふ 8	をしふる 95	をしふれ 2	
ヤ行下一	をしえ 23		をしえる 2		をしえれ 0	をしえよ 0
下二			をしゆ 6	をしゆる 43	をしゆれ 2	
ワ行下一	をしゑ 1		をしゑる 0		をしゑれ 0	をしゑよ 0
下二			をしう 1	をしうる 5	をしうれ 0	

ヤ行下二段活用動詞

表18は、ヤ行下二段活用動詞の例として「見ゆ」の活用形別用例数をまとめたものである。仮名書きの例や「観」「診」「看」など他の漢字を用いたものは含まない。ヤ行下二段活用の場合には、二段活用「みゆ」「みゆる」「みゆれ」は仮名遣いの問題を生じないので、ハ行・ワ行の表から省いてある。

活用の行別に見ると、本来のヤ行が93.5%，ハ行6.4%，ワ行0.1%となっている（下二段の形を除いて集計するとそれぞれ92.4%，7.5%，0.1%）。

表18 動詞「見ゆ」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ヤ行下一	みえ 1875		みえる 763		みえれ 1	みえよ
下二			みゆ 159	みゆる 289	みゆれ 38	0
ハ行下一	みへ 162		みへる 52		みへれ 0	みへよ 0
ワ行下一	みゑ 1		みゑる 1		みゑれ 0	みゑよ 0

なお、ワ行下一の「みゑ」「みゑる」の形と思われるものが1例ずつ認められたものの、どちらの用例も第二音節がカタカナで書かれたものである（例：「立派な金主を見付けたと見えるネ」1901年5号P087A16「投機」内田魯庵）。カタカナ以外に例がないことを考えると、これは仮名遣いというよりも字形が類似していることによる誤植と見た方がよいかもしれない。

表19は、ヤ行下二段活用動詞のもう一つの例として「絶ゆ」の活用形別用例数をまとめたものである。活用の行別に見ると、本来のヤ行が84.9%，ハ行14.7%，ワ行0.4%となっている（下二段の形を除いて集計するとそれぞれ83.9%，15.7%，0.4%）。

表19 動詞「絶ゆ」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ヤ行下一	たえ 549		たえる 13		たえれ 0	たえよ
下二			たゆ 5	たゆる 38	たゆれ 1	0
ハ行下一	たへ 105		たへる 0		たへれ 0	たへよ 0
ワ行下一	たゑ 3		たゑる 0		たゑれ 0	たゑよ 0

ここで取り上げた例を見る限り、ハ行・ヤ行の下二段活用の動

詞は総じて正しい語形で用いられている割合が高く、9割ほどが正しく用いられている。特にワ行に活用されることはほとんどなく、誤用は1%以下にすぎない。

ワ行下二段活用動詞

表20は、ワ行下二段活用動詞の例として「据う」の活用形別用例数をまとめたものである。活用の行別に見ると、本来のワ行が63.0%、ハ行16.8%、ヤ行20.2%となっている。

表20 動詞「据う」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ワ行下一	すゑ 117		すゑる 9		すゑれ 1	すゑよ 2
下二			すう 0	すうる 2	すうれ 0	
ハ行下一	すへ 33		すへる 2		すへれ 0	すへよ 0
下二			すふ 0	すふる 0	すふれ 0	
ヤ行下一	すえ 37		すえる 2		すえれ 1	すえよ 0
下二			すゆ 1	すゆる 1	すゆれ 0	

表21は、ワ行下二段活用動詞のもう一つの例として「植う」の活用形別用例数をまとめたものである。活用の行別に見ると、本来のワ行が77.1%、ハ行3.3%、ヤ行19.5%となっている。

「据う」「植う」の2語を見る限り、ワ行下二段活用動詞は、9割以上が正しく表記されていたハ行・ヤ行下二段活用動詞とは異なり、誤用例がかなり多い。特にヤ行に誤る例が多く、2割前後に上っている。

以上、動詞の活用と仮名遣いについて、問題の多い活用型について見てきた。活用する行や活用形で違いを見せるほか、個々の動詞によっても分布に差がある。これがどのような要因によるものなのかは、著者や文体・ジャンルなどを含めた今後の詳しい調

表21 動詞「植う」の語形別の用例数

	未然形	連用形	終止形	連体形	已然形	命令形
ワ行下一	うゑ 131		うゑる 13		うゑれ 1	うゑよ 2
下二			うう 5	ううる 9	ううれ 1	
ハ行下一	うへ 5		うへる 1		うへれ 0	うへよ 0
下二			うふ 1	うふる 0	うふれ 0	
ヤ行下一	うえ 31		うえる 2		うえれ 0	うえよ 0
下二			うゆ 4	うゆる 4	うゆれ 0	

査を待ちたいと思う。

7

おわりに

以上、『太陽コーパス』の仮名遣いの状況についてそのあらましを見てきた。大量のデータをもとに大まかな傾向をつかむことはできたものと思うが、コーパスのあまりにも多い情報の中には、たくさんの興味深い問題が埋没しているに違いない。個々の語や用例をより詳しく調査することにより、そうした問題をすくい上げてゆくことが望まれる。また、『太陽コーパス』以外の資料との比較も重要である。今後、周辺資料とコーパスとを併せ用いることで、規模と検索性を活かした新たな研究がなされることを期待したい。

注

- (1) 注タグ（構造化テキスト本文中における「注」要素）のうち、「分類」属性に「G仮名遣」という値を持つものを、以下「仮名遣い注」と呼ぶ。
- (2) 2004年9月時点でのデータを調査対象とした。
- (3) 振り仮名に対する注の場合には原文属性の値に「」（全角）

が含まれるので、本文中の仮名に関する注はXPath式で次のように書ける。

//注 [./@分類="G仮名遣"] [not (contains (./@原文, "[")]]

(4) 表計算ソフトの自作関数による。

(5) ヤ行に活用する動詞を仮名遣いの問題として処理することについては「5. 動詞の活用と仮名遣い」参照。

(6) 『太陽』の漢字含有率の推移を調査したものとして、土屋(1967)のサンプリング調査があるが、ほぼ同じ傾向を示している。

(7) この調査では当時正しいとされ後に修正された「水 [する]」なども誤用に含めているため、実際にはこの数値よりも正確に、一定の仮名遣いに準拠していたことになる。

(8) XPath式で次の通り (いずれも「記事」をカレントノードとしたとき)。

空白文字を除去した記事の文字数：

string-length (translate (normalize-space (.) ;',''))

仮名遣い注数：

count (//注 [./@分類="G仮名遣"] [not (contains (./@原文, "[")]])

参考文献

小木曾智信 (2003) 「『太陽コーパス』における字音仮名遣いについて」2003年3月 (『明海日本語』第8号, 117-126頁)

田中牧郎・小木曾智信 (2000) 「総合雑誌『太陽』の本文の形態と電子化テキスト」(国立国語研究所編『日本語科学』8, 141-152頁, 国書刊行会)

築島 裕 (1986) 『歴史的仮名遣い その成立と特徴』(中公新書)

土屋信一 (1967) 「雑誌『太陽』の用字の変遷」(『言語生活』193, 34-43頁, 筑摩書房)

Research on the Formative Era of Contemporary Japanese Based on the *Taiyo Corpus*

Contents

1. The Corpus Design

TANAKA Makiro

Use of Journal *Taiyo* as Material for Historical Research on the Japanese Language

YAMAGUCHI Masaya

Himawari: A Full Text Retrieval System for Structured Texts

OGISO Toshinobu

Prism and *Tanpopo* : HTML Applications to Use the *Taiyo Corpus* XML Documents

2. Analysis of Corpus Data

1. Lexicology

TANAKA Makiro

The Establishment of the Kanji Compound *yūshū* (優秀) in the Japanese Lexicon : Examining the Relation with the Expressed Subject

YOSHIKAWA Asuka

Two-Character Kanji Compounds with Reversed Orders

IDE Junko

The Shift from Kanji notation to Katakana notation of Foreign Place Names

BABA Toshiomi
On the Adversative Connectives

II. Grammar
SHIMADA Yasuko
Usages of the Conjunction *soshite* : Classification and
Analysis of Types

NAKAO Hisako
Study of the Adverb *totemo* : Historical Analysis of Usage
Change from Negative Conditional Adverb to
Unconditional Degree Adverb

KONDO Asuko
Honorific Forms in Conjunction with Action Nouns and
Infinitive Verb Forms

OGISO Toshinobu
Potential Forms of Kanji Compound Verbs : Development
of *dekiru*

III. Graphology and Orthography
TANAKA Makiro
Identification and Mapping Kanji

SASAHARA Hiroyuki
Assimilation and Collision of Kanji Forms

NAKAGAWA Miwa
On Kana Variant Forms

KONDO Asuko
Voicing Notation : Focusing on the Proportion of Voiced
Kana

OGISO Toshinobu
On Kana Orthography

あとがき

本書は、『太陽コーパス』の設計と、『太陽コーパス』を用いた確立期現代語の記述とについて、コーパスを作りながら研究してきた成果を、プロジェクトに関わった研究者の論文の形でまとめたものである。本書と同時に刊行した『太陽コーパス』CD-ROMとあわせ、コーパスを用いた日本語研究が普及していく足がかりとして、学界や社会に受け入れられれば幸いである。

われわれは『太陽コーパス』のプロジェクト進行中にも、学術誌などでいくつかの研究成果を発表してきた。プロジェクトの成果報告をまとめて記しておくことと、『太陽コーパス』による研究の広がりを見わたしたい読者の参考に供することとを目的として、その一覧を下に掲げる。『太陽コーパス』を中心的に扱っている研究に限り、『太陽コーパス』を部分的に用いた研究は除いてある。

小木曾智信 (2002) 「近代語テキストからの可能動詞の抽出—『太陽コーパス』を例に一」(『明海日本語』7, 125-135頁)

小木曾智信 (2003 a) 「『太陽コーパス』における字音仮名遣いについて」(『明海日本語』8, 117-126頁)

小木曾智信 (2003 b) 「近代日本語における「能ふ」の用法—『太陽コーパス』の用例から—」(『明海大学外国語学部論集』15, 1-11頁)

小椋秀樹・小木曾智信・近藤明日子 (2002) 「『太陽コーパス』を使った近代語表現の通時的考察—口語文体・可能表現・待遇表現について—」(『国語学会2002年度春季大会要旨集』, 177-184頁, 国語学会)

木村睦子・田中牧郎・飯島満 (1997) 『『太陽』コーパスの作成と活用』(科学研究費研究報告書・新プロ「日本語」, 国立国語研究所)

木村睦子・田中牧郎・飯島満・笹原宏之 (1999) 『『太陽』コーパス』の漢字処理—『太陽』1901の漢字調査—」(科学研究費研究報告書・新プロ「日本語」, 国立国語研究所)

- 田中牧郎 (1998) 「『太陽』コーパスの作成」 (『国立国語研究所創立50周年記念研究発表会資料集』 39-46頁, 国立国語研究所)
- 田中牧郎 (2001a) 「XMLを利用したコーパスの構築—『太陽コーパス』を中心に—」 (『日本語学』 20-13, 80-91頁, 明治書院)
- 田中牧郎 (2001b) 「『太陽コーパス』の構築による確立期現代語の研究」 (平成13年度国立国語研究所公開研究発表会予稿集, 2-11頁, 国立国語研究所)
- 田中牧郎 (2004a) 「近代語研究資料と研究 雑誌『太陽』」 (『日本語学』 23-12, 208-220頁, 明治書院)
- 田中牧郎 (2004b) 「雑誌『太陽』創刊年 (1895年) における口語文—敬体を中心に—」 (飛田良文編『国語論究11 言文一致運動』, 78-107頁, 明治書院)
- 田中牧郎 (2005印刷中) 「『敏感』の誕生と定着—『太陽コーパス』を用いて—」 (『日本近代語研究』 4, ひつじ書房)
- 田中牧郎・小木曾智信 (2000) 「総合雑誌『太陽』の本文の様態と電子化テキスト」 (国立国語研究所編『日本語科学』 8, 141-152頁, 国書刊行会)
- 中尾比早子 (2003) 「明治・大正期における程度副詞『非常に』について」 (田島毓堂・丹羽一彌編『名古屋ことばのつどい 言語科学論集』, 127-138頁)
- 山口昌也・田中牧郎 (2002) 「言語研究のための構造化テキストと検索支援システム—『太陽コーパス』を中心に—」 (『国語学会2002年度春季大会要旨集』, 169-176頁, 国語学会)
- 山口昌也・田中牧郎 (2004) 「多様な構造化テキストに対応した全文検索システム『ひまわり』」 (『日本語学会2004年度秋季大会予稿集』, 165-172頁, 日本語学会)
- 『太陽コーパス』のプロジェクトは、本書と『太陽コーパス』CD-ROMの刊行によってひと区切りとなるが、この形にまとまるまでには次のような経緯があった。
- 「本研究の目的と本書の構成」の「2.2 コーパス構築の事業へ」(ii 頁) のところでも触れたように、『太陽コーパス』の作成は、国立国語研究所の国語辞典編集室による国語辞典編集のための用例採集事業 (『日本大語誌』構想) のなかで始まった。この事業のなかで、雑誌『太陽』は、専門家による任意採集方式 (スカウ

ト式)による用例採集資料に選ばれ、延べ約80万語の用例が採集された。『太陽』本文の電子化は、採集された用例に文脈を付けることを目的として平成6年度に着手された。ここまでの作業の遂行は、木村睦子・国語辞典編集室長(当時)の指揮のもと、藤原浩史・国語辞典編集室研究員(当時)の手で進められた。平成11年度からは、用例採集事業がコーパス作成事業に移行し、『太陽』本文の電子化は、『太陽コーパス』の作成というコーパス作成事業の一環に位置づけを変えた。この移行は加藤安彦・国語辞典編集室長(当時)の指導によるところが大きい。平成13年度の国立国語研究所の独立行政法人化を機とする改組によって国語辞典編集室が研究開発部門第一領域として再編された後は、相澤正夫・研究開発部門長と山崎誠・研究開発部門第一領域長による指導や調整により、コーパスの公開と論文集の刊行に向けた準備が整えられた。

本書の執筆者には加わっていないが、『太陽コーパス』の作成や研究に関わった研究者は多い。担当した国立国語研究所の研究員・非常勤研究員・研究補佐員と、所外から協力者として参画してくださった方々は次の通りである。

飯島満、乾とね、大木一夫、大塚みさ、緒方典裕、奥村大志、小椋秀樹、貝美代子、小島聡子、柴田雅生、中山典子、服部隆、服部紀子、平澤啓、本多久美子、山田貞雄、湯浅茂雄、吉田谷幸宏

このほか、コーパスへの情報付与やデータの校正などには、非常にたくさんの作業担当者の尽力があった。コーパスの試験公開を三次にわたって行い、試験公開版に対して利用者から有益な意見や激励をいただいた。『太陽コーパス』と本書の出版、およびコーパスに含める記事の著作権に関わる調査や許諾業務では、『太陽』の刊行元である博文館を継承した株式会社博文館新社の大橋一弘社長、吉田延江取締役の理解と協力が大きかった。また、『太陽』の各記事の著作権継承者の方々からは記事を『太陽コーパス』に収録することについて許諾をいただいた。『太陽コーパス』の作成と本書の編集に御助力をいただいたすべての方々に、謹んで感謝を申し上げたい。

なお、『太陽コーパス』プロジェクトとして成果がまとまった国立国語研究所の研究事業課題は次のものである。

国立国語研究所国語辞典編集室「国語辞典編集のための用例採集」(昭和63～平成10年度)

国立国語研究所国語辞典編集室「国語辞典編集室コーパスの作成」(平成11～12年度)

独立行政法人国立国語研究所研究開発部門第一領域「現代日本語における書き言葉の実態解明と雑誌コーパスの構築」(平成13～17年度)

また、以下の科学研究費による研究課題も、『太陽コーパス』に深く関わるものである。

文部省科学研究費(創成的基礎研究費)「国際社会における日本語についての総合的研究 研究班4:情報発信のための言語資源の整備に関する研究」(略称:「新プロ」日本語)(研究代表者:水谷修,平成6～10年度)

日本学術振興会科学研究費(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」(研究代表者:田中牧郎,平成14～17年度)

本書ならびに『太陽コーパス』CD-ROMの編集製作に際しては、所内に刊行物検討委員会を設置し、内容や形態等について検討を加えた。委員は次の通りである。

相澤正夫(委員長,研究開発部門長)

横山詔一(情報資料部門第二領域長)

宇佐美洋(日本語教育部門第一領域主任研究員)

田中牧郎(研究開発部門第一領域主任研究員)

さらに、本書所収の英文題目の作成にはエリック・ロング氏(国立国語研究所非常勤研究員)の協力があつた。

国立国語研究所では、今後もさまざまなコーパスを作成し、段階的に規模を拡大し水準も向上させていくことで、現代日本語の実態を把握できる汎用コーパスの構築と研究を目指していく予定である。そのなかで『太陽コーパス』の誤りは補正し、不十分なところは改善し、本格的な日本語コーパス構築につなげていきたい。そのためにも、『太陽コーパス』や本書に対する御批正をいただくことを願っている。

田中牧郎

執筆者一覧

田中牧郎（たなか・まきろう）	国立国語研究所研究開発部門 第一領域主任研究員
笹原宏之（ささはら・ひろゆき）	国立国語研究所研究開発部門 第一領域主任研究員
山口昌也（やまぐち・まさや）	国立国語研究所研究開発部門 第一領域研究員
小木曾智信（おぎそ・としのぶ）	国立国語研究所研究開発部門 第一領域非常勤研究員・明海 大学外国語学部専任講師
近藤明日子（こんどう・あすこ）	国立国語研究所研究開発部門 第一領域非常勤研究員
井手順子（いで・じゅんこ）	元 国立国語研究所研究開発 部門第一領域非常勤研究員・ 国立国語研究所研究開発部門 研究補佐員
中尾比早子（なかお・ひさこ）	元 国立国語研究所研究開発 部門第一領域非常勤研究員・ 名古屋大学大学院文学研究科 博士課程後期在学中
中川美和（なかがわ・みわ）	元 国立国語研究所研究開発 部門第一領域非常勤研究員・ 東京都立大学人文学部助手
吉川明日香（よしかわ・あすか）	元 国立国語研究所研究開発 部門第一領域非常勤研究員

島田泰子（しまだ・やすこ） 香川大学教育学部助教授
馬場俊臣（ばば・としおみ） 北海道教育大学教育学部助教授

※島田・馬場は、『太陽コーパス』の作成と研究に外部協力者として参画した。

『太陽コーパス』に含めなかった記事の一覧

著者の著作権が未処理のため『太陽コーパス』から除外した記事は次の通りである。

- 1895年7号 「土耳其の演劇」 山田寅次郎
- 1901年3号 「梅花雜譚」 服部担風
- 1901年9号 「電気化学工業の現況」 鴨居武
- 1901年10号 「工業世界」 金子篤寿・鴨居武・塚本信治
- 1901年12号 「工業世界」 金子篤寿・鴨居武・塚本信治
- 1901年13号 「工業世界」 金子篤寿・鴨居武・塚本信治
- 1901年13号 「人種強健策」 井上豊太郎
- 1901年14号 「工業世界」 金子篤寿・鴨居武・塚本信治
- 1901年14号 「特別通信 鼓浪嶼國際居留地对我居留民の位置」 小山松寿
- 1909年1号 「韓国皇太子殿下」 栗原広太（談話）
- 1909年1号 「誕」 正宗白鳥
- 1909年1号 「故サルドウと川上（初対面の活劇）」 渋谷馬頭（抄訳）・ロイ＝フューラー
- 1909年1号 「何故に黄帝国か」 木戸解劍
- 1909年2号 「『明治評論』経営時代の鳥谷部君」 山本青城
- 1909年2号 「春汀君談片」 末広一雄
- 1909年2号 「春汀先生と開鑿事業」 小原敏磨（寄）
- 1909年2号 「最近の文学界所感」 正宗白鳥（談）
- 1909年4号 「大隈伯座談」 江森泰吉・大隈重信（談）
- 1909年4号 「をんな」 草野柴二
- 1909年4号 「煎茶の科学」 小此木忠四郎
- 1909年4号 「五十五年前に於ける日露人の会食」 山田松濤（訳）・ゴンチャロフ（記）
- 1909年4号 「海外通信」 渡辺金三
- 1909年5号 「海外通信 米国加州通信」 渡辺金三
- 1909年5号 「劇詩の新傾向とメターリンク」 川島金五郎
- 1909年5号 「故岩崎男の園芸」 北村東江
- 1909年5号 「名士の仏国観 仏国の陸軍」 金城鉄壁生（談）
- 1909年5号 「名士の仏国観 新聞雑誌記者生活」 渋谷馬頭
- 1909年6号 「佐々木高行伯」 岩崎鏡川
- 1909年8号 「世界之時局 独露の反感」 小柳夏村
- 1909年8号 「碁敵」 正宗白鳥
- 1909年8号 「親兄弟」 村山鳥逕
- 1909年8号 「きやめれおん」 荒野放浪（訳）・チエホフ（作）
- 1909年8号 「『人類の由来』梗概」 田中茂穂
- 1909年8号 「名士の露西亜観 露国の内状」 八杉貞利（談）
- 1909年8号 「名士の露西亜観 政治家を中心として観たる露国」 夏秋竜一（談）
- 1909年8号 「仏国女丈夫を訪ふ（土耳其国情談）」 渋谷馬頭
- 1909年8号 「読書断片」 中原青蕪
- 1909年8号 「外人の日本観 仏人の極東観」 ハンリ＝ラブルー
- 1909年8号 「外人の日本観 日糖事件」 ア＝モネスチエ
- 1909年8号 「外人の日本観 日清兩國国民の根本的相違」 フランツ＝ウオーアス
- 1909年10号 「海外通信 区々たる母国の言論」 植田勇
- 1909年10号 「海外通信 在米日本青年の操作問題」 吉田公重

- 1909年10号「点滴」中原青蕪（訳）・ハインツ＝トフォーテ
 1909年10号「アナトル、フランスの作風」川島風骨
 1909年10号「森の緋葉よ」内藤晨紫
 1909年10号「名士の土耳其観 土国民と海軍」光風霽月生（談）
 1909年10号「名士の土耳其観 土耳其の事情」飯島亀太郎（談）
 1909年10号「外交と温泉場」渋谷馬頭
 1909年11号「文部省留学生無用論」堀切北東（談）
 1909年11号「世界之時局 独逸宰相辞職」中原青蕪
 1909年11号「世界之時局」小柳夏村
 1909年11号「世界之時局 飛行機騒ぎ」中原青蕪
 1909年11号「海外通信 シヤトル博覧会」永山天淵
 1909年11号「名士の瑞西観 堅実なる独立心」洪沢元治（談）
 1909年12号「帰朝したる高平男」川崎巳之太郎
 1909年12号「外人の日本観 日米衝突の原因」アウエスヌ
 1909年12号「外人の日本観 土耳其青年党と日本」セフェル＝ペー
 1909年12号「世界之時局 再燃せるクリート問題」小柳夏村
 1909年12号「海外通信 布哇に於ける労働問題概見」奥田寛太郎
 1909年12号「海外通信 シヤトル博覧会雑記」永山天淵
 1909年12号「海外通信 桑港通信」渡辺金三
 1909年12号「ビュートア」草野柴二・アナトル＝フランス
 1909年13号「海外通信 南米秘露共和国の変乱」小泉有穂
 1909年13号「雷雨」正宗白鳥
 1909年13号「空中航行機の原動機関」笹本菊太郎（談）
 1909年14号「行政税制整理問題 理想的海軍行政整理」水天一碧樓主人（談）
 1909年14号「外人の日本観 対清懸案解決観」ア＝モネステエ
 1909年14号「名士の和蘭観 和蘭雑談」山川端夫（談）
 1909年14号「最近の野球界」正宗白鳥
 1909年16号「世界之時局 土耳其の将来」マームード＝シェブケト＝パシャ
 1909年16号「帝国大学派文士の長短 啓発されたことが無い」正宗白鳥（談）
 1909年16号「名士の瑞典諸威観 世界一の電話国」泉谷氏一（談）
 1917年1号「乾いた心」正宗白鳥
 1917年2号「開戦以来の伊太利」寺崎武男
 1917年3号「政戦の発展」佐々木惣一
 1917年3号「堅忍の生」与謝野寛・エレンヌ＝スガン
 1917年3号「寡婦後」尾島菊子
 1917年4号「我が貿易品に対する粗製濫造の非難と事実と救済策」記者（文責）・古谷重綱
 1917年4号「総選挙風聞録」二宮楚川
 1917年5号「露国の革命と露国の向後」蜷川新
 1917年5号「露国の革命事変」今井政吉
 1917年5号「露国の革命と外交界の影響」末広一雄
 1917年5号「船用動力界の革命」安井正太郎（筆録）・加茂正雄（口述）
 1917年5号「エンデル氏の仏蘭西観」太宰施門
 1917年5号「女子教育管見要項」城田米吉
 1917年5号「若い血の流れ」中村孤月
 1917年6号「戦後の本邦化学工業を如何せん」鴨居武
 1917年6号「米国防戦の影響」蜷川新

-
- 1917年6号 「戦後に於ける世界平和耐久策を論ず」 泉哲
1917年6号 「逐鹿異耳と新代議士」 二宮楚川
1917年8号 「蠢動」 村山勇三
1917年9号 「外観内容共に貧弱」 三木武吉
1917年9号 「露西亜の非賠償非併合主義を評す」 蟬川新
1917年9号 「大戦を予言せる独逸戦前の芸術」 寺崎武男
1917年9号 「南飛驒の溪流」 二宮楚川
1917年9号 「冬のハヶ嶽の印象」 林博太郎
1917年10号 「段の兵力と支那の政局」 寺西秀武（談）
1917年12号 「マルヌの野に立ちて」 堀口大学
1917年12号 「戦時利得税の可否」 記者（文責）・橋本圭三郎
1917年12号 「教育の形式と内容」 山田信次（寄）
1917年12号 「美術院日本画作家の感想」 横山大観
1917年12号 「美術院日本画作家の感想」 前田青邨
1917年12号 「美術院日本画作家の感想」 名取春仙
1917年12号 「美術院日本画作家の感想」 川端龍子
1917年12号 「美術院日本画作家の感想」 堅山南風
1917年13号 「奔湍翠光」 平田松堂
1917年13号 「黒髪」 簗木清方
1917年13号 「文展日本画概見」 紀星峰
1917年13号 「南半球の利用と日本」 蟬川新
1917年13号 「去勢に就きての話（上）」 角田隆
1917年13号 「祖国の残土」 与謝野寛・エミル=エ` ルアラン
1917年14号 「支那と寺内内閣」 安岡秀夫
1917年14号 「我國人は如何に化学工業を見る乎」 島田慶一
1917年14号 「動揺せる露国社会」 今井政吉
1917年14号 「海外発展と出移民」 蟬川新
1917年14号 「去勢に就きての話（下）」 角田隆
1917年14号 「独断独語」 末広一雄
1917年14号 「盆栽の新研究」 平松中陽
1925年1号 「無煙石油ランプ」 八木静雄
1925年1号 「電熱式自動煮炊器」 八木静雄
1925年1号 「天気予報の新聞に現はるゝまで」 梶間百樹（談）
1925年1号 「科学を基礎としたる『理研酒』 米を使用せず醗酵法に依らず各種成分の配合に依り製造せる新日本酒」 山本敬三
1925年1号 「完然然焼電の発明」 八木静雄
1925年1号 「欧米に於ける定期航空輸送の現状」 伊藤西夫
1925年1号 「濾過水筒の発明」 八木静雄
1925年1号 「軌近に於ける食餌療法の進歩」 高田蒔
1925年1号 「木箱に代用する段ボール紙箱の発明」 八木静雄
1925年1号 「至つて便利な瓶の口」 八木静雄
1925年1号 「缶詰仮蓋の発明」 八木静雄
1925年1号 「家庭用の缶詰器」 八木静雄
1925年1号 「毒瓦斯の話」 久村種樹（談）
1925年1号 「愈々議會を通過する普通選挙の内容解剖と其批判」 大野恭平
1925年1号 「新案 飛行服」 名木山賢一
-

- 1925年1号 「怪奇探偵小説 緑の扉」 梅原成基（訳）・ヘルマン＝ランドン（作）
 1925年1号 「欧米の発明と日本の発明」 川部佑吉
 1925年1号 「社会時評 青牛余誕」 水島爾保布
 1925年1号 「最も経済的な無線電話装置法」 田村正四郎
 1925年1号 「長篇科学小説 生ける死『第一回』」 佐野慶介（訳）・ジョン＝マルチン＝リーヒー（作）
 1925年1号 「王子製紙会社の事業と社員待遇法」 天外散史
 1925年1号 「家庭用として便利な布製組立筆筒」 八木静雄
 1925年1号 「最近に於ける外科手術の発達」 佐藤清一郎
 1925年1号 「文壇風聞記」 水上渉
 1925年1号 「水を用ゐざる新蒸気動力」 イスマー＝ギンスベルグ
 1925年1号 「英国と日本の飛行郵便」 今井富次
 1925年1号 「通俗講話 ラヂオの知識」 横尾年正
 1925年1号 「新案の御飯ふかし」 八木静雄
 1925年1号 「簡便なる可燃性瓦斯採集器」 八木静雄
 1925年1号 「我輩は牛である」 新井誠夫
 1925年1号 「紙袋開口器の新発明」 八木静雄
 1925年2号 「時事漫評 冷時冷殺」 水島爾保布
 1925年2号 「東京電燈会社の事業と社員待遇法」 天外散史
 1925年2号 「長篇科学小説 生ける死『第二回』」 佐野慶介（訳）・ジョン＝マルチン＝リーヒー（作）
 1925年2号 「肺結核の妙薬として見たる肝油の効能」 牛島保
 1925年2号 「怪奇探偵小説 緑の扉」 梅原成基（訳）・ヘルマン＝ランドン（作）
 1925年3号 「企業上より比較したる『三井』対『三菱』」 加藤良平
 1925年3号 「『理研酒』から『理研水』へ 山本敬三氏の理研酒の研究をよむ」 根本正
 1925年3号 「鮮銀と台銀社員待遇法」 天外散史
 1925年3号 「脳溢血と驟雨浴」 桑野良作
 1925年3号 「孫の顔を見ながら熱湯浴をする大塊老」 野田俊作
 1925年3号 「万病にきく驟雨浴」 井上通治
 1925年3号 「労働運動ゴシップ 二つの平行線」 村島婦之
 1925年3号 「拳法無双 鏢師の妻と少女」 赤城照二
 1925年3号 「時事漫言 白眼倒視」 水島爾保布
 1925年3号 「東京放送局の事業と将来」 新名直和
 1925年3号 「戦争廃除作用から世界協力作用へ—歴史的意義を有する第五回総会—世界開放と日本民族の将来—」 稲垣克己
 1925年3号 「密輸入物語」 渡辺悌司
 1925年3号 「長篇科学小説 生ける死『第三回』」 佐野慶介（訳）・ジョン＝マルチン＝リーヒー（作）
 1925年3号 「浪界風聞録」 南北遊客
 1925年3号 「怪奇探偵小説 緑の扉」 梅原成基（訳）・ヘルマン＝ランドン（作）
 1925年3号 「文壇風聞記」 水上渉
 1925年4号 「明治生命と帝国生命の社員待遇法」 天外散史
 1925年4号 「赤化運動と平和運動」 稲垣守克
 1925年4号 「倫敦の火—『都会の獲物』—」 田内長太郎（訳）・アーノールド＝ベンネット
 1925年4号 「拳法無双伝」 赤城照二
 1925年4号 「日本と米国の不良少年少女」 植田たまよ（談）
 1925年4号 「小鳥のうた」 さいき＝せじゅろ
 1925年4号 「社会時評 痴子春秋」 水島爾保布
 1925年4号 「最近に発達したる脳と肺の外科手術」 大池豊三郎

- 1925年4号 「長篇科学小説 生ける死『第四回』」 佐野慶介（訳）・ジョン＝マルチン＝リーヒー（作）
1925年4号 「文壇風聞記」 水上渉
1925年4号 「吉田寅二郎とロバート・ルキ・スチヴェンソン」 田内長太郎（訳）・藤沢利喜太郎（講演）
1925年4号 「世界に於ける放送無線電話の現状」 荒川大太郎
1925年4号 「ラヂオ放送と無線電信の発達」 米村嘉一郎
1925年4号 「世界無線網の現状」 加島斌
1925年4号 「最も経済的な受信回路の話」 原口猷一
1925年4号 「無線小説 ホビイとラヂオ」 クーパー＝アレク
1925年4号 「無線電話の真空管に就て」 佐野昌一
1925年4号 「——幕物——ラヂオキング」 田内長太郎（訳）・アダム＝ハル＝シヤーク（作）
1925年4号 「聾者とラヂオ」 記者・ワアーレン＝ボンド
1925年4号 「高声器の選定に就て」 田村正四郎
1925年4号 「日本の家庭とラヂオ」 赤坂東司
1925年4号 「諷刺小説『ラヂオ禁止』」＊（抄訳）・パーカー＝ルーター
1925年4号 「放送無線受話器の発達——其の形式試験を要する理由——」 宮本和一郎
1925年4号 「日本で使用し得る放送聴取用受信装置」 横山英太郎
1925年4号 「ラヂオによる印刷通信——一分間に千語以上の活字が電送される——」 秋山秀雄
1925年4号 「無線放送局の設備」 槇尾年正
1925年5号 「映画と絵画の窓の場面に就て」 畑耕一
1925年5号 「台湾第一の富豪 林本源暗闘秘史」 飯野静男
1925年5号 「我国銀行界の最大弱点」 星野行則
1925年5号 「馬來ジョホールに於ける大猛獣狩」 吉井信照
1925年5号 「最も的確なる就職案内として好評を博せる各社待遇法の解剖 十五銀行と山口銀行の社員待遇法」 天外散史
1925年5号 「将来の空中戦」 岩本周平
1925年5号 「長篇科学小説 生ける死『第五回』」 佐野慶介（訳）・ジョン＝マルチン＝リーヒー（作）
1925年5号 「藤沢博士の帝都復興論を読む 大正十三年十二月号本誌掲載『東京市の復興に就て震災直後当時滞京中のビヤード博士に送いたる覚書』参照」 スネルレン
1925年5号 「社会外の社会の一面……狂病院七年間の生活記録」 亀田健次
1925年5号 「無線電話の再放送」 槇尾年正
1925年5号 「国有鉄道の建設と改良」 伊藤霞峰
1925年5号 「文壇風聞録」 水上渉
1925年7号 「私が抜擢した人々 岩島寸三氏と山本久三郎氏 私を抜擢した人々 中上川彦次郎氏、和田豊治氏、牛場卓蔵氏、荘田平五郎氏」 西野恵之助
1925年7号 「予が世話をした人で出世した感心な人々」 安田亀一
1925年7号 「川崎造船所の社員待遇法」 天外散史
1925年7号 「大乘の偉人東郷元帥より受けたる感化」 小笠原長生
1925年7号 「『毒』利用論」 松村松年
1925年7号 「書物『デモクラシーと労働』『懺悔録』『言志録』 人物 加藤弘之、杉浦重剛」 深作安文（談）
1925年7号 「探偵二十年」 小泉惣之助
1925年7号 「新政友会総裁 田中義一男」 中野青四郎
1925年7号 「短波長による無線放送」 槇尾年正
1925年7号 「人造絹糸の現在及将来」 佐羽太三郎
1925年7号 「長篇科学小説 生ける死『第六回』」 佐野慶介（訳）・ジョン＝マルチン＝リーヒー（作）
1925年7号 「プラチナの由来」 渡辺万次郎

- 1925年7号 「時事一家言」 浜田三峰
1925年7号 「文壇風聞記」 水上渉
1925年9号 「ヒンデンブルグ出陣の意義」 松原一雄
1925年9号 「政界合同劇と其影武者—熱海会議より犬養老隠退まで—」 山岡頭巾
1925年9号 「大臣級の政治家」 南北遊客
1925年9号 「手腕家よりも円満な人格が第一」 加藤恭平
1925年9号 「スポーツマン・スピリットが出世の根本」 原邦造
1925年9号 「入露日記 モスコウにて（五月七日）」 大津六三郎
1925年9号 「鐘淵紡績会社の社員待遇法」 天外散史
1925年9号 「私の学生時代 大食のレコード」 門野重九郎
1925年9号 「長篇科学小説 生ける死『第七回』」 佐野慶介（訳）・ジョン=マルチン=リーヒー（作）
1925年9号 「大衆文芸本来の理想」 堀木克三
1925年9号 「小鳥のうた」 さいき=せじゅうろ
1925年9号 「食養法より見たる米、野菜、肉類の割合」 石塚右玄
1925年9号 「郵便飛行機に備へた無線電話」 榎尾年正
1925年9号 「文壇風聞録」 水上渉
1925年10号 「支那騒擾の思想的背景」 神田正雄
1925年10号 「時事寸言」 浜田三峰
1925年10号 「科学時代物語「一」」 水島爾保布
1925年10号 「前田侯爵家宝物の落札者とその価格」 森新八
1925年10号 「当来文芸思潮としての大地主義」 大槻憲二
1925年10号 「新らしき時代」 藤森淳三
1925年10号 「第一銀行の社員待遇法」 天外散史
1925年10号 「財界に分布する政党政派の勢力」 南北遊客
1925年10号 「京都画壇の鳥瞰図」 黒田天外
1925年10号 「南京豆の袋—逆襲の戦法—」 原奎一郎
1925年10号 「音楽と映画—その成長と今後の傾向に就て—」 永田龍雄
1925年10号 「長篇科学小説 生ける死『第八回』」 佐野慶介（訳）・ジョン=マルチン=リーヒー（作）
1925年10号 「文壇風聞録」 水上渉
1925年11号 「現代社会相の一考察」 深作安文
1925年11号 「家庭破壊の種々相と貞操観念の動揺」 山川菊栄
1925年11号 「財界の回復期は尚ほ不明」 松本重威
1925年11号 「街頭スケッチ 万年筆売り」 水島爾保布
1925年11号 「帝国電燈会社の社員待遇法」 天外散史
1925年11号 「土龍の祈り」 伊藤靖
1925年11号 「農村描写 泥の中で踊る人々」 青木純二
1925年11号 「現代世相の掃着点」 三宅驥一
1925年11号 「集団生活と静思の必要」 土方成美
1925年11号 「小鳥のうた」 さいき=せじゅうろ
1925年11号 「法廷より見たる世態人情の変遷」 宇野要三郎
1925年11号 「丸の内を中心とせる特殊犯罪」 宮沢文作
1925年11号 「不良少年少女と飛躍状態」 後藤四方吉
1925年11号 「劇場に現はれたる最近の世相」 山本久三郎
1925年11号 「婦人の求職と就職に現はれたる時代相」 宮下正男
1925年11号 「犯罪に現はれたる最近の傾向」 中村義正
1925年11号 「長篇科学小説 生ける死『第九回』」 佐野慶介（訳）・ジョン=マルチン=リーヒー（作）

1925年11号「文壇風聞記」水上渉
1925年12号「欧米印象記」賀川豊彦
1925年12号「英国のラシア商人」山県五十雄（談）
1925年12号「司法閥の過去と現在」光風霽月楼
1925年12号「太陽の黒点に絡まる悲観説—低気圧—東北地方凶作—米価騰貴—恐慌—不景氣」加藤良平
1925年12号「大飢饉と太陽黒点」田口克敏
1925年12号「太陽黒点と稲作との關係に就て（抄録）」山沢金五郎
1925年12号「騰落の分水嶺に立てる米価の前途」松村金兵衛（談）
1925年12号「刻下の失業事情と職業補導」安田亀一
1925年12号「間接費の節約と良質多産主義」雜賀良三郎
1925年12号「鐘紡主義の三大要素による経営法」福原八郎
1925年12号「職工の心理より見たる能率の増減」森誠
1925年12号「大金額の費目と複雑な費目の研究」伊吹震
1925年12号「経費節減の根本策は統制ある協力一致あるのみ」宮島清次郎
1925年12号「不景氣切抜策 直接販売制と日給制採用」松崎半三郎
1925年12号「汗の出ないところ」竹内薫兵（談）
1925年12号「明治神宮の雉」松宮春一郎（談）
1925年12号「検死」鈴木平十郎（談）
1925年12号「電燈による植物の栽培」槇尾年正
1925年12号「院展の彫刻の印象—二人の対話—」畑正吉・デルスニス
1925年12号「二科の彫刻を見て」石川確治
1925年12号「熱意の文学出でよ」木蘇穀
1925年12号「現文壇の掃趣」川崎備寛
1925年12号「或る女の陳述」湯浅真生
1925年12号「文壇風聞録」水上渉
1925年13号「労働組合法の批判とその基本觀念」三谷一二
1925年13号「〈同人会談叢〉」竹内薫兵（談）
1925年13号「労働運動挿話 新人に恋多し」村島鼎之
1925年13号「労働総同盟の人々」佐保貞二
1925年13号「外交縦談」松原一雄
1925年13号「書齋から議政壇上に送られる…貴族院の四学者議員」碓氷三郎
1925年13号「尿の學問」木内幹
1925年13号「恋の復讐 鬼火の心」三樹莊主人（訳）・ドーレヴィリー（原作）
1925年13号「抽象論の流行」大槻憲二
1925年13号「夢」間宮茂輔
1925年13号「新らしき時代のために」那珂孝平
1925年13号「人工妊娠的作家」森本巖夫
1925年13号「或る恋文のトリック」川崎備寛
1925年13号「文壇風聞録」水上渉
1925年14号「外交縦談」松原一雄
1925年14号「ソウエート露西亜印象記」八杉貞利
1925年14号「文芸批評の実用性」西村鎮彦
1925年14号「歳末時評」川崎備寛
1925年14号「遅日」間宮茂輔

国立国語研究所報告122

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

Research on the Formative Era of Contemporary
Japanese Based on the *Taiyo Corpus*

平成17年3月31日

編者 独立行政法人国立国語研究所

〒190-8561 東京都立川市緑町3591-2

電話：042-540-4300(代表)

FAX：042-540-4333(代表)

URL：http://www.kokken.go.jp/

発行者 株式会社博文館新社

代表者 大橋一弘

発行所 株式会社博文館新社

〒3811-4721 東京都文京区小石川2-14-6

電話：03-3811-4721(代表)

FAX：03-3818-1431(代表)

印刷者 凸版印刷株式会社

NDC 810.26

UDC 811.521'06

ISBN4-86115-155-4

(平16-18)