国立国語研究所学術情報リポジトリ

「太陽コーパス」の構築による確立期現代語の研究

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 田中, 牧郎 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002938

「太陽コーパス」の構築による確立期現代語の研究

田中牧郎（研究開発部門第一領域）
mtanaka@kokken.go.jp

1．はじめに

日本語研究においても、コーパスを用いて記述をおこなう試みが普及しつつあるが、その方法論は、今のところ試行錯誤の段階にあるように見受けられる。コーパスの構築やこれを利用した記述を、計画的・体系的に進めることを通して、この方法の有効性を多角的・実践的に検討していくことが望まれる。

コーパスの構築は、大規模な予算や人手、組織的な管理を必要とする点で、個人の研究者の手には余るところがあり、学会や研究機関などがプロジェクトを組んで進めるのが現実的である。そうした見地から、国立国語研究所でも、いくつかのコーパスを構築することを計画し実践している。本報告では、研究開発部門第一領域で構築を進め、来年度完成を目指している「太陽コーパス」について、構築方法の概要と研究例の一端を紹介する。なお、「太陽コーパス」は現状では未完成であり、以下に示すデータの形式や用例、数値などは、暫定的なものであり、完成時には変更になる場合がある。

2．「太陽コーパス」の構築

2.1　目的・対象

「太陽コーパス」は、現代日本語の記述研究に資することを目的に、特に、現代語の確立期（「確立期」という呼び名は、東京語に関する松村 1957 から借用）である 20 世紀初期の日本語コーパスとして、雑誌『太陽』を対象に構築を進めているものである。当初は、国語辞典編集のために採集した用例に文脈を付与する目的で作業に着手したが、コーパス構築を直接の目的とする研究対象に転換した（木村・加藤ほか 1999）。

『太陽』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で、幅広い読者に多く読まれ、記事のジャンルが広範囲であり、当時の日本人に多大な影響を与えたメディアである（鈴木 2001、永嶺 1997）という点で、20 世紀初期の書き言葉を代表させることのできる、価値の高い資料である。20 世紀の始点である 1901 年を起点に 8 年刻みで、1909、1917、1925 年の各 12 冊と、『太陽』の創刊年である 1895 年 12 冊と終刊年の 1928 年（2 号で終刊）2 冊を加え、全 62 冊を対象にしている。 全体で 1500 万字を越える規模になる。

2.2 電子テキストの作成

文献資料を対象にして電子テキストを作成し、コーパスを構築するためには、対象とする文献の言語構造を把握し、そこに含まれる言語情報を抽出する、という作業が必要になる（文献学と情報工学の両面から電子テキスト作成の問題を多角的に扱ったものとして、安永 1998 が有益である）。まず文献の言語構造の把握については、原資料『太陽』を文献学的に研究することが前提になる。『太陽』の文献学的研究には、土屋

（1966・1967）に先例があり、われわれも研究を進めてきた（田中 1998、木村・田中ほか 1999、田中ほか 2000·2001）。『太陽』は、現在の雑誌とはかなり異なる特徴を多く持っており、それを簡単にまとめると、次の通りである。

（1）句読法の変異が大きく、段落や文などの単位が整っていない。

（2）ルビを豊富に持つ。

（3）字体の変異が大きい。

（4）誤植が多い。また、規範的な用法からの逸脱と思われる箇所も多い。

（5）著者、ジャンル、文体などが、多様である。

こうした特徴に応じて、文献に含まれる言語情報を的確に引き出すことができるようデータの形式を整えることが求められる。上の五点に対する具体的な方針は、次の通りである。

（1）記事、句読区切を、本文の構造上の基本単位とする。

（2）情報価値の高いルビを埋め込む。

（3）字体の包摂規準を策定し字体の統合の規準を明示し、必要に応じて字体情報を埋め込む。

（4）誤植は修正し、修正箇所に注記を埋め込む。また規範から逸脱する箇所には、ママ注記を埋め込む。なお、清濁・仮名遣い等、ゆれが著しい現象は、規範的な用法に統一し、修正箇所には修正注記を埋め込む。これらは、本文批判の記録として整備する。

（5）ジャンル、著者、文体などの情報を埋め込む。

この五つの方針を実現させる手だてとして、文書記述言語の国際規格としてさまざまな利点をもつＸＭＬを活用する（小木曽 2001、田中 2001）。（1）は、XML で本文を構造化する際の単位であり、（2）～（5）で埋め込む情報は、XML によるマークアップにおいてタグに属性を加えて記述する。図１は、XML で記述した電子テキストの例である。次頁の表１は、主なタグセットである。

図１　ＸＭＬによる電子テキスト


```
<記事 題名="老壮士" 著者="内田魯庵" 著作権="〇" 保護期限="1979" 欄名="小説雑俎" 文体="口語" ジャンル
="913">
<s>　老壮士　<l id="P109A05"/><br/></s>
<s>　内田魯庵　<l id="P109A06"/><br/></s>
<s>　<r rt="わし">俺</r>の<r rt="らい">來</r><r rt="れき">歴</r>を<r rt="き">聞</r>きたい。</s>
<s>ふンふツ……<r rt="が">斯</r>うして<r rt="やま">山</r>の<r rt="なか">中</r>に<r rt="ひつ">引</r><r rt="こ
">込</r><l id="P109A07"/>んで、</s>
<s><r rt="つくえ">机</r><注 原文="机［つくゑ］" 分類="G 仮名遣" />に<r rt="むか">向</r>つて<r rt="あを">青
</r><r rt="べう">表</r><r rt="し">紙</r>を<r rt="ひら">繙</r>いてゐると、</s>
<s><r rt="たけ">竹</r><r rt="なか">中</r><r rt="はん">半</r><r rt="べ">兵</r><r rt="ゑ">衛</r>の<r rt="かく">
隠</r>れ<l id="P109A08"/><r rt="が">家</r><r rt="ぜん">然</r>として<r rt="こん">此様</r>な<r rt="へ">平
</r><r rt="ぼ">凡</r>な<r rt="つら">面</r><r rt="つき">貌</r>が<r rt="ひと">一</r><r rt="くせ">癖</r>ありげに
<r rt="み">見</r>えるさうで、<l id="P109A09"/></s>
<s><r rt="とき">時</r><r rt="〜">々</r><r rt="みち">道</r>に<r rt="まよ">迷</r>つて<r rt="たづ">尋</r>ね"
```

要素名	説明	属性（括弧内は任意）	
太陽	雑誌１冊	年・号・Version	箱形要素
記事	各記事	題名・著者・著者生年・著作権・保護期限・欄名・ジャンル・文体（・文体備考）	
引用	記事中の引用部分	種別・話者（・ジャンル・文体）	
s	句読点等による区切り		
r	ルビ	rt（･j･s･lrt･lj）	行内要素
外字	字体情報	（文字鏡・記号）	
注	本文批判の記録	（原文｜規範形）・種類（・情報）	
l	原文誌面の頁行番号	id（･元位置）	

表１　主なタグセット

2.3 電子テキストからデータベースへ

文献資料の構造に即して XML により記述した電子テキストを作成しておけば、文献言語が含む様々な情報を、目的に応じて自在に取り出すことが可能になる。XML 本文に対しては、XSLT という変換言語を用いることができ、「太陽コーパス」でも、これを活用している。その主な用途は、次の通りである。

ブラウザ表示用：HTML 形式に変換

印刷用：pLaTeX 形式に変換

テキスト生成：タグを除去したプレーンテキストを生成

情報抽出：記事、引用、熟字訓、外字、注記等の各一覧を生成

次に、HTML に変換してブラウザで表示した結果（図２）、記事一覧を生成した結果を示す（図３）。

図２　XSLT によるブラウザ表示への変換結果

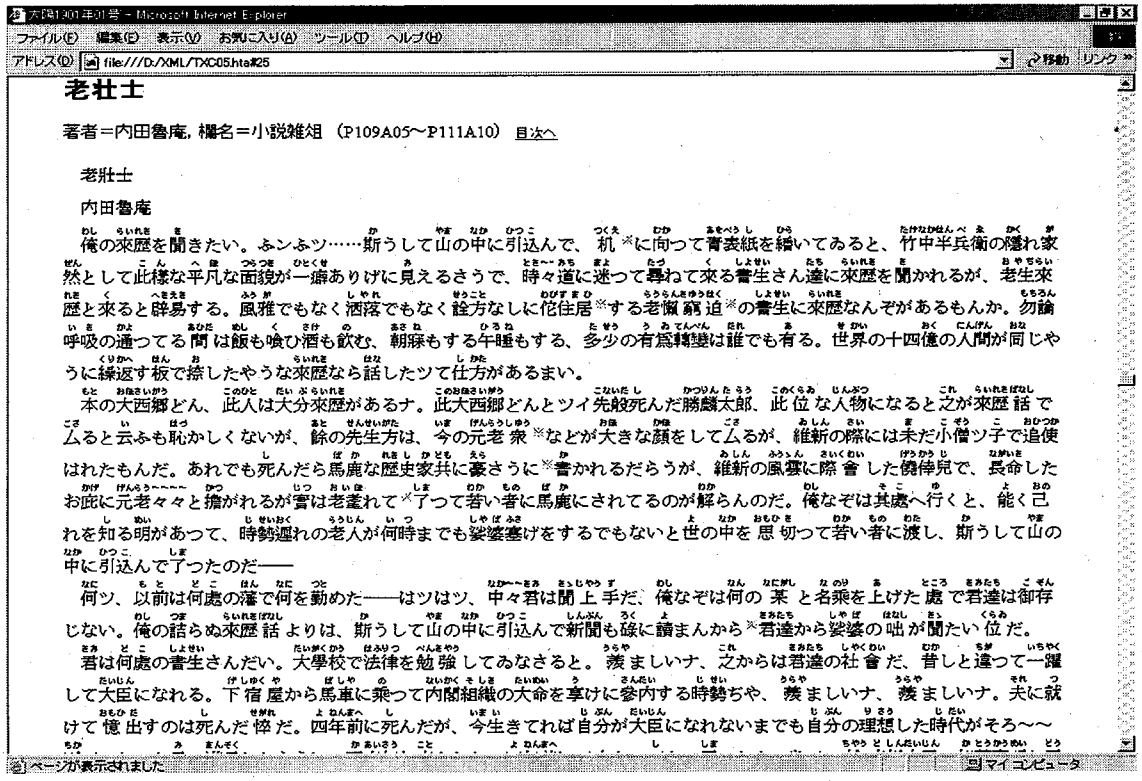
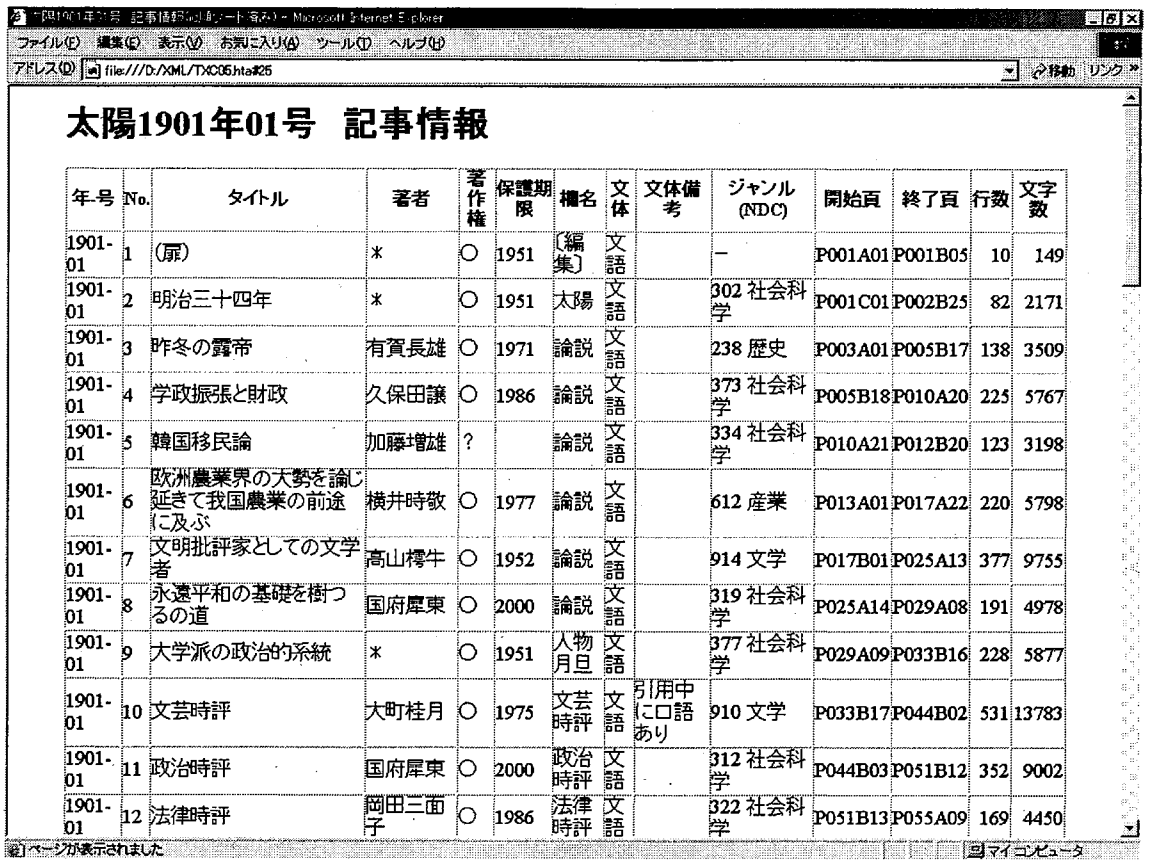


図３　XSLT による記事一覧の生成結果



太陽1901年01号　記事情報

年-号	No.	タイトル	著者	著作権	保護期限	欄名	文体	文体備考	ジャンル (NDC)	開始頁	終了頁	行数	文字数
1901-01	1	(扉)	*	○	1951	〔編集〕	文語		―	P001A01	P001B05	10	149
1901-01	2	明治三十四年	*	○	1951	太陽	文語		302 社会科学	P001C01	P002B25	82	2171
1901-01	3	昨冬の露帝	有賀長雄	○	1971	論説	文語		238 歴史	P003A01	P005B17	138	3509
1901-01	4	学政振張と財政	久保田譲	○	1986	論説	文語		373 社会科学	P005B18	P010A20	225	5767
1901-01	5	韓国移民論	加藤増雄	?		論説	文語		334 社会科学	P010A21	P012B20	123	3198
1901-01	6	欧洲農業界の大勢を論じ延きて我国農業の前途に及ぶ	横井時敬	○	1977	論説	文語		612 産業	P013A01	P017A22	220	5798
1901-01	7	文明批評家としての文学者	高山樗牛	○	1952	論説	文語		914 文学	P017B01	P025A13	377	9755
1901-01	8	永遠平和の基礎を樹つるの道	国府犀東	○	2000	論説	文語		319 社会科学	P025A14	P029A08	191	4978
1901-01	9	大学派の政治的系統	*	○	1951	人物月旦	文語		377 社会科学	P029A09	P033B16	228	5877
1901-01	10	文芸時評	大町桂月	○	1975	文芸時評	文語	引用中に口語あり	910 文学	P033B17	P044B02	531	13783
1901-01	11	政治時評	国府犀東	○	2000	政治時評	文語		312 社会科学	P044B03	P051B12	352	9002
1901-01	12	法律時評	岡田三面子	○	1986	法律時評	文語		322 社会科学	P051B13	P055A09	169	4450

このように XML で記述することにより、文献言語がもつ様々な情報を電子テキストに埋め込んでおき、目的に応じて必要な情報を自在に取り出すことが可能になるのである。

これらの構造化された豊富な言語情報は、電子テキストの検索にも生かすことが期待される。「太陽コーパス」では、これを実現する検索システムを二種類用意している。

（１）HTA による検索システム

XML テキストを直接検索する HTML アプリケーション。XML 自体を扱うことができ、データ量も少なくて済むのが利点。検索速度が遅いのが難点。

（２）suffix array による検索システム

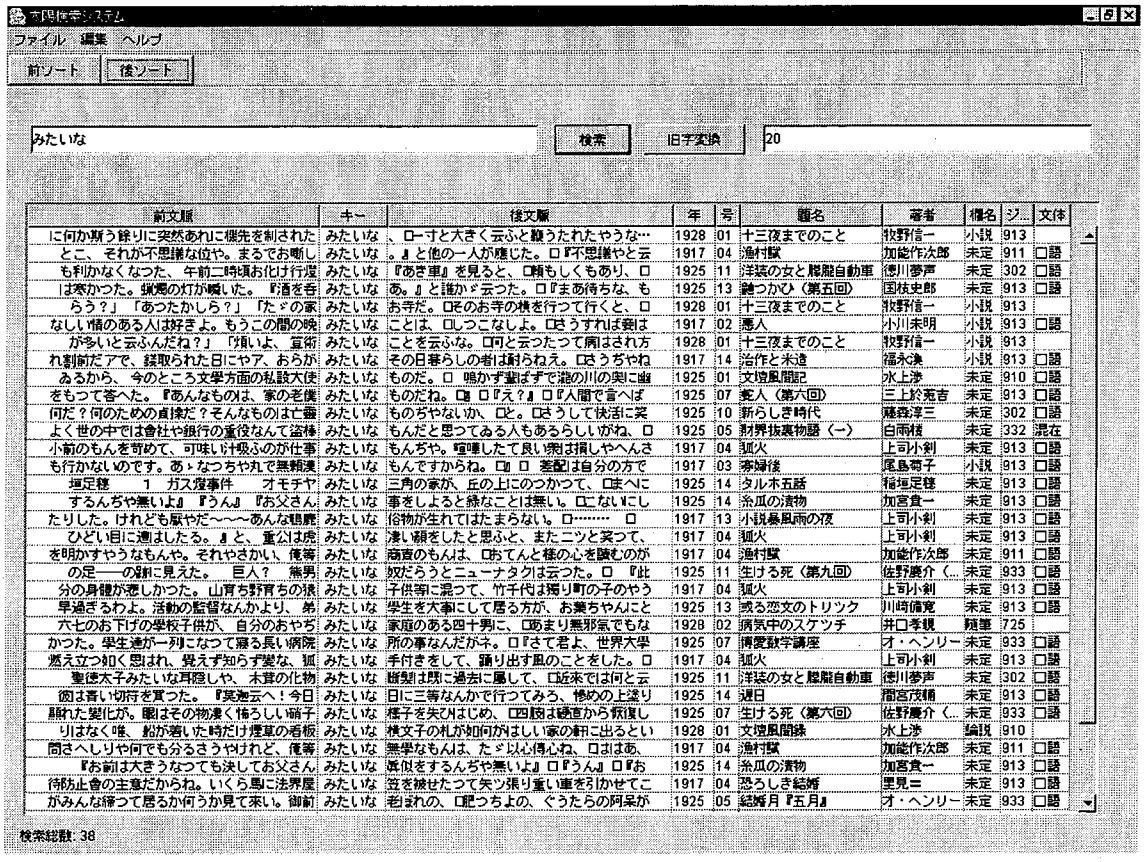
XML テキストからあらかじめ各種のテキストを生成しておき、これにインデックスを付与したテキストデータベース。非常に高速であることが利点。データ量が多くなるのが難点。suffix array については近藤（2001）に紹介がある。

このうち（２）の検索結果画面を、図４（次頁）に示す。

質の高い構造化電子テキストを基盤に置き、変換・検索のツールを整備することで、情報を自在に引き出せるデータベースへと、機能を高めることができるのである。

以上のようにして構築された「太陽コーパス」は、確立期現代語の記述研究に対して、多方面で貢献を果たすことが期待できる。次節では、表記・語法・語彙の各領域から具体的な言語現象を取り上げ、「太陽コーパス」によって可能になる、特色ある研究の一端を紹介したい。

図４　「太陽コーパス」の検索結果



前文脈	キー	後文脈	年	号	題名	著者	欄名	ジ…	文体
に何か斯う錆りに突然あれに棲先を刺された	みたいな	、ロ一寸と大きく云ふと難うたれたやうな…	1928	01	十三夜までのこと	牧野信一	小説	913	
とこ、それが不思議な位や。まるでお嘶し	みたいな	。』と他の一人が應じた。ロ『不思議やと云	1917	04	漁村獄	加能作次郎	未定	911	□語
も利かなくなつた、午前二時頃お化け行燈	みたいな	『あき車』を見ると、ロ頼もしくもあり、ロ	1925	11	洋装の女と朧朧自動車	徳川夢声	未定	302	□語
は寒かつた。縞燈の灯が瞬いた。『酒を呑	みたいな	あ。』と誰かゞ云つた。ロ『まあ待ちな、も	1925	13	鍾つかひ（第五回）	国枝史郎	未定	913	□語
らう？」「あつたかしら？」「たゞの家	みたいな	お寺だ。ロそのお寺の横を行つて行くと、ロ	1928	01	十三夜までのこと	牧野信一	小説	913	
なしい情のある人は好きよ。もうこの間の晩	みたいな	ことは、ロしつこなしよ。ロさうすれば要は	1917	02	愚人	小川未明	小説	913	□語
が多いと云ふんだね？」「煩いよ、宣術	みたいな	ことを云ふな。ロ可と云つたつて病はされ方	1928	01	十三夜までのこと	牧野信一	小説	913	
れ割前だアで、採取られた日にヤア、おらが	みたいな	その日暮らしの者は耐らねえ。ロさうぢやね	1917	14	治作と米造	福永渙	小説	913	□語
ゐるから、今のところ文學方面の私設大使	みたいな	ものだ。ロ 鳴かず飛ばずで濁の川の奥に幽	1925	01	文壇風聞記	水上滝	未定	910	□語
をもつて答へた。『あんなものは、家の老僕	みたいな	ものだね。ロ ロ『え？』ロ『人間で言へば	1925	07	愛人（第六回）	三上於菟吉	未定	913	□語
何だ？何のための直接だ？そんなものはと番	みたいな	ものぢやないか、ロと。ロさうして快活に笑	1925	10	新らしき時代	藤森淳三	未定	302	□語
よく世の中では會社や銀行の重役なんて盗様	みたいな	もんだと思つてゐる人もあるらしいがね、ロ	1925	05	財界抜裏物語（一）	白雨楼	未定	332	混在
小前のもんを苛めて、可味い汁吸ふのが仕事	みたいな	もんぢや。喧嘩したて良い肉は損しやへんさ	1917	04	狐火	上司小剣	未定	913	□語
も行かないのです。あゝなつちやれて無頼漢	みたいな	もんですからね。ロ ロ 差配は自分の方で	1917	03	寒婦後	尾島菊子	小説	913	□語
垣足穂　1　ガス燈事件　オモチヤ	みたいな	三角の家が、丘の上にのつかつて、ロまへに	1925	14	タルホ五話	稲垣足穂	未定	913	□語
するんぢや無いよ』『うん』『お父さん	みたいな	事をしようと緑なことは無い。ロこないにし	1925	14	糸瓜の漬物	加宮貫一	未定	913	□語
たりした。けれども厭やだ～～～あんな魑魅	みたいな	俗物が生れてはたまらない。ロ……　ロ	1917	13	小説暴風雨の夜	上司小剣	未定	913	□語
ひどい目に遭はしたる。』と、重公は虎	みたいな	凄い顔をしたと思ふと、またニツと笑つて、	1917	04	狐火	上司小剣	未定	913	□語
を明かすやうなもんや。それやさかい、俺等	みたいな	商賣のもんは、ロおてんと様の心を験むのが	1917	04	漁村獄	加能作次郎	未定	911	□語
の足――の鋼に見えた。　巨人？　熊男	みたいな	奴だらうとニューナタクは云つた。ロ『此	1925	11	生ける死（第九回）	佐野廣介（…	未定	933	□語
分の身體が悪しかつた。山育ち野育ちの狼	みたいな	子供等に混つて、竹千代は猫り町の子のやう	1917	04	狐火	上司小剣	未定	913	□語
早過ぎるわよ。活動の監督なんかより、弟	みたいな	学生を大事にして居る方が、お兼ちやんこと	1925	13	或る恋文のトリック	川崎備寛	未定	913	□語
六七のお下げの學校子供が、自分のおやぢ	みたいな	家庭のある四十男に、ロあまり無邪氣でもな	1928	02	病気中のスケツチ	井口孝親	随筆	725	
かつた。學生達が一列になつて寝る長い病院	みたいな	所の事なんだがネ。ロ『さて君よ、世界大學	1925	07	博愛教学講座	オ・ヘンリー	未定	933	□語
燃え立つ如く思はれ、覚えず知らず變な、狐	みたいな	手付きをして、踊り出す風のことをした。ロ	1917	04	狐火	上司小剣	未定	913	□語
聖徳太子みたいな耳隠しや、木茸の化物	みたいな	縦横は隙に過去に屈して、ロ近来では何と云	1925	11	洋装の女と朧朧自動車	徳川夢声	未定	302	□語
図は青い切符を買つた。『英迪云へ！今日	みたいな	日に三等なんかで行つてみう、惨めの上塗り	1925	14	遅日	宮宮茂輔	未定	913	□語
顛れた變化が。眼はその物凄く怖ろしい硝子	みたいな	格子を失ひはじめ、ロ四肢は硬直から恢復し	1925	07	生ける死（第六回）	佐野廣介（…	未定	933	□語
りはなく唯、船が着いた時だけ煙草の看板	みたいな	横文字のれが如何がはしい家の軒に出るとい	1928	01	文壇風聞録	水上滝	論説	910	
問さへしりや何でも分るさうやけれど、俺等	みたいな	無學なもんは、たゞ以心傳心ね、ロまはあ、	1917	04	漁村獄	加能作次郎	未定	911	□語
『お前は大きうなつても決してお父さん	みたいな	無似をするんぢや無いよ』ロ『うん』ロ『お	1925	14	糸瓜の漬物	加宮貫一	未定	913	□語
待防止會の主意だからね。いくら馬に法界屋	みたいな	笠を被せたつて矢ツ張り重い車を引かせてこ	1917	04	恐ろしき結婚	里見ニ	未定	913	□語
がみんな締つて居るか何うか見て來い。御前	みたいな	老ぼれの、ロ肥つちよの、ぐうたらの阿呆が	1925	05	結婚月『五月』	オ・ヘンリー	未定	933	□語

３．「太陽コーパス」を利用した確立期現代語の記述

3.1 表記　―ハヤワ行下二段（下一段）動詞語尾の仮名遣い―

明治時代の仮名遣いは、大きく見れば、歴史的仮名遣いに次第に統一されていく流れがあり、大体、明治末年ごろまでにはその流れが完結したのではないか、とされている（築島 1986）。ただ、その実態調査は十分行われておらず、具体的なことになると、不明な部分が多い。『太陽』についても、初期の年次を中心に仮名遣いのゆれは相当に大きく、本文批判にあたり、相応の処理が要求された。2.2(4)で述べた「修正注記」に記録された「仮名遣い」の情報を活用すれば、「太陽コーパス」によって、仮名遣いの実態調査を行うことができる。ここでは、ハ行・ヤ行・ワ行に活用する下二段動詞（口語法で下一段化した例を含む、以下同）の語尾「ふ」「ゆ」「う」、「へ」「え」「ゑ」について実態調査を行った結果を報告する。

まず、ハヤワ各行の下二段動詞のリストのうち、同じ行の四段活用動詞（口語法で五段化した例を含む。以下同）があるもの（「叶ふ（下二）」に対する「叶ふ（四段）」）や、漢字の読み分け作業を要するもの（「捕ふ〔つかまふ・とらふ〕」）を除外する。これは、検索結果に別語が混入することを避けるためである。この処理によりリストに残った語を検索し、「太陽コーパス」の 1895･1901･1909･1917･1925 の５年分で 100 例以上あると見なされた、次の語を調査対象とした。

ハ行下二段動詞　23 語

*与ふ、*訴ふ、*愁ふ、*終ふ、*押さふ、*衰ふ、換ふ、抱ふ、*数ふ、*考ふ、*加ふ、拵ふ、*答ふ、支ふ、*備ふ、堪ふ、*貯ふ、携ふ、*譬ふ、控ふ、交ふ、*迎ふ、教ふ

ヤ行下二段動詞　10語

覚ゆ、消ゆ、*聞こゆ、越ゆ、栄ゆ、聳ゆ、絶ゆ、*増ゆ、見ゆ、燃ゆ

ワ行下二段動詞　2語

植う、据う

調査の結果、歴史的仮名遣いに合致しない例が、全体の 1%に満たないものが、ハ行 14 語、ヤ行 2 語あった（上の一覧で＊を付けた語）。これらは、『太陽』において仮名遣いが安定していると考えられるものである。ハ行動詞に、仮名遣いの安定している語が多く、ヤ・ワ各行の動詞は、仮名遣いのゆれている語が多いことがわかる。

次に、無印の語（仮名遣いがゆれている語）について、歴史的仮名遣いに合致しない例が、その語の全用例の中で占める百分率を年次ごとに示すと、表 2 のようになる。ハ行・ヤ行・ワ行それぞれの語群のなかでは、「太陽コーパス」5 年分で歴史的仮名遣いに合致しない率の高いものから順に配列した。

この表から、全体を通して、1901 年までは、仮名遣いのゆれが大きく、1909 年でいくつかの語のゆれがなくなり、1917 年に至って、一部の語を除き、ほぼ全体が安定する、という事実を知ることができる。ハ行動詞が比較的ゆれが小さく、安定するのも早く、ヤ行動詞がこれに次ぎ、ワ行動詞のゆれがもっとも大きかった、という行による差異も発見できるのである。築島（1986）の所説を大筋で実証できるとともに、変化の過程を細部まで明らかにできるのである。本文批判の記録をコーパスに埋め込んだことで可能になる記述例のひとつである。

		明治			大正		総数
		1895年	1901年	1909年	1917年	1925年	
ハ行	堪ふ	10.3%	26.3%	26.2%	2.7%	3.4%	1318
	控ふ	12.1%	14.8%	16.0%	0.0%	0.0%	176
	拵ふ	0.0%	21.1%	0.0%	0.0%	0.0%	268
	交ふ	3.4%	1.8%	5.3%	0.0%	0.0%	201
	教ふ	5.0%	6.1%	0.0%	0.0%	0.0%	695
	抱ふ	0.0%	10.0%	0.0%	0.0%	2.4%	133
	換ふ	3.4%	5.3%	1.3%	0.0%	0.0%	842
	携ふ	0.0%	3.1%	7.0%	0.0%	0.0%	259
	支ふ	2.0%	2.8%	0.0%	0.0%	0.0%	300
ヤ行	栄ゆ	12.5%	31.3%	9.5%	0.0%	0.0%	103
	絶ゆ	5.3%	23.2%	11.0%	0.0%	0.0%	712
	越ゆ	7.0%	23.9%	0.0%	0.0%	0.0%	445
	覚ゆ	7.9%	2.9%	8.3%	1.8%	2.9%	967
	見ゆ	3.2%	3.6%	4.9%	0.0%	0.7%	3303
	聳ゆ	5.6%	4.3%	0.0%	0.0%	0.0%	124
	消ゆ	8.5%	0.0%	1.5%	0.0%	0.0%	379
	燃ゆ	3.4%	6.1%	2.8%	0.0%	0.0%	247
ワ行	据う	36.7%	38.7%	55.1%	0.0%	0.0%	200
	植う	14.3%	18.2%	27.9%	0.0%	0.0%	232

表 2　歴史的仮名遣いに合致しない例の百分率

3.2 語法　―比況の助動詞「みたやうだ」「みたいだ」―

比況の助動詞「みたいだ」は明治期から用いられるようになり、江戸時代から使わ

れていた「みたやうだ」が変化して生まれたものである（原口 1974・松村 1977）。この「みたやうだ」から「みたいだ」への交替の様相は、「太陽コーパス」を用いることで、かなり明瞭にとらえることができる。

		1895年	1901年	1909年	1917年	1925年	計
みたやうだ	地の文	3	3	21	8	10	45
	会話文	1	3	3	1	2	10
	計	4	6	24	9	12	55
みたいだ	地の文			2	6	20	28
	会話文		7	13	23	32	75
	計		7	15	29	52	103
合　計		4	13	39	38	64	158

表３　「みたやうだ」「みたいだ」の地の文・会話文別の用例数

表３は、この二語の出現回数を、地の文・会話文別にまとめたものである。1909 年までは「みたやうだ」の方が多いが、1917 年以後は「みたいだ」の優勢が明らかになる。「みたやうだ」は、地の文・会話文ともに用いられていたところから、次第に地の文に限定されていく方向にある。これに対して「みたいだ」は、会話文に出現することから始まり、次第に地の文へも進出していく方向が見えている。衰退する語形と台頭する語形との対照が際立つ分布を見せている。

表４は、この二語の出現回数を、次のような用法分類別にまとめたものである。

述定法

・お役人の商賣は武士の商法見たやうで（会話 1925 年 2 号　原胤昭「死刑囚の最期」）

・この洞穴は竈みたいだ。（会話　1925 年 05 号　佐野慶介訳「生ける死」）

連用法

・初心の碁打ち見たやうに、いよ～～解散となつて見なければ、大局が見えず（地 1925 年 13 号　鬼谷庵「政界鬼語」）

・僕はまるで蟲けら見たいにこの年まで生きて來たです。（会話 1925 年 13 号　下村千秋「蟋蟀」）

連体法

・小説の斷篇見たやうなものを綴つたりして（地　1925 年 11 号　坪内逍遙「学生時代の追憶」）

・聖徳太子みたいな耳隠しや（地　1925 年 11 号　徳川夢声「洋装の女と朦朧自動車」）

		1895年	1901年	1909年	1917年	1925年	計
みたやうだ	述定法					1	1
	連用法	2	1	8		2	13
	連体法	2	5	16	9	9	41
	計	4	6	24	9	12	55
みたいだ	述定法			3	5	3	11
	連用法			4	5	22	31
	連体法		7	8	19	27	61
	計	0	7	15	29	52	103
合　計		4	13	39	38	64	158

表４　「みたやうだ」「みたいだ」の用法別の用例数

「みたやうだ」は、ほぼ連用法と連体法に限られ、次第に連体法を中心とするものに用法が固定していく方向が認められる。一方「みたいだ」は、連体法から始まり、次第に連用法や述定法に用法

が拡大していく方向が見て取れる。「みたいだ」が述定法にも拡張する先には、「昭和期に一般化」（松村 1977）する、「僕らは根の土を水で洗われてしまったみたいですからね」（横光利一『旅愁』）のような、活用語に下接する助動詞らしい用法への発展に、つながるものであると考えられる。

新旧の語形の交替をこのようになめらかに描くことは、コーパスによらない従来の研究では簡単に実現できなかったものである。

3.3 語彙　―<重要>を意味する漢語形容動詞の文体的価値―

明治・大正期の総合雑誌の文体は、はじめ文語文が中心であったところから、次第に口語文が増え、やがては口語文ばかりになる（見坊 1957、国立国語研究所 1987）。文体の大きな変革と他の言語現象との関わりを探ることは、確立期現代語を対象とする研究テーマとして重要なものである。「太陽コーパス」は、記事ごとに文末辞を指標に文語か口語かの認定を行い、この情報を、検索結果で取得できるようにしている。ここでは、〈重要〉を意味する漢語形容動詞を例にとって、語の文体的価値について、探索してみたい。

国立国語研究所『分類語彙表』の分類番号「3.37 経済」の最初のグループにある語彙、「大切」「大事」「要」「肝要」「肝腎」「緊要」「重要」「枢要」「主要」「必要」のうち、やや性質の異なる「要」「必要」を除いた９語を調査対象にする。この９語につき、「太陽コーパス」の 1895～1925 年の５年分を検索し、形容動詞の単独用法として用いられた例のみにつき、文語記事・口語記事いずれに出現しているかを数え、文語記事に用いられた比率をまとめると、表５のようになる。表の最下段には、各年ごとの全記事数に対して文語記事数が占める比率を掲げたが、この数値を仮の期待値として、文語記事への出現率が期待値を上回るものをＡ群、下回るものをＢ群とした（年により期待値との上下が入れ替わる語は、総合的に判断した）。相対的に、Ａ群は文語的な価値、Ｂ群は口語的な価値、とい

		1895年	1901年	1909年	1917年	1925年	総語数
A群	緊要	100.0%	97.3%	80.0%	41.2%	6.7%	141
	重大	100.0%	93.0%	66.4%	23.4%	0.0%	608
	主要	94.2%	92.2%	53.3%	24.0%	0.0%	144
	重要	100.0%	89.0%	55.5%	17.9%	0.0%	604
	枢要	100.0%	73.3%	66.7%	0.0%	0.0%	60
B群	肝要	80.4%	68.5%	40.9%	7.4%	0.0%	173
	肝腎	61.9%	27.3%	11.8%	7.7%	0.0%	92
	大事	55.9%	36.4%	6.5%	0.0%	0.0%	149
	大切	80.7%	33.3%	19.1%	4.3%	0.6%	461
文語記事率		96.3%	73.0%	37.8%	18.8%	2.4%	

表５　〈重要〉を意味する漢語形容動詞の文語記事への出現率

		日本国語大辞典第二版初出	和英語林三版	言海
A群	緊要	大学垂加先生講義(1679)	○	×
	重大	日本外史(1827)	×	○
	主要	小学化学書(1874)	×	×
	重要	西国立志編(1870-71)	×	×
	枢要	続日本紀(790)	○	○
B群	肝要	江談抄(1111 頃)	○	○
	肝腎	三代実録(859)	○	○
	大事	十七箇条憲法(604)	○	○
	大切	殿暦(1107)	○	○

表６　〈重要〉を意味する漢語形容動詞の文献出現状況

う差異をもつと見てよいものと考えられる。

これらの語について、『日本国語大辞典第二版』（小学館）の初出文献、『和英語林集成第三版』（1886 年）『言海』（1889-91 年）の見出しへの掲出の有無をまとめると、表 6（前頁）のようになる。A 群の語には、初出が新しく、明治中期までの辞書の見出しになっていない語が目立つのに対して、B 群の語は、古代からあり、明治の辞書にも見出しになっているものばかりである。明治になって新たに使われ始めた漢語が文語的であり、古くから用いられてきた漢語が口語的であるという大勢を具体的に確認することができよう。

		1895年	1901年	1909年	1917年	1925年
A群	緊要	100.0%	100.0%	100.0%	75.0%	50.0%
	重大	100.0%	96.8%	84.1%	59.7%	45.5%
	主要	100.0%	100.0%	100.0%	86.4%	60.9%
	重要	93.9%	92.9%	78.7%	68.1%	36.1%
B群	肝要	82.4%	100.0%	62.5%	80.0%	12.5%
	大事	14.3%	0.0%	0.0%	0.0%	0.0%
	大切	83.3%	61.5%	54.8%	25.9%	13.0%

表 7　〈重要〉を意味する漢語形容動詞の連体形における文語形「なる」の比率

形容動詞の場合、例えば連体形には、文語形「なる」に対して口語形「な」という形態の対立がある。各語の連体形におけるこの二形の出現回数を数え、文語形「なる」の出現率を求めると、表 7 の通りである。なお、連体法の形態には「の」もあるが、これには、文体以外の問題があるのでここでは除外し、「の」を用いることが一般的である「枢要」「肝要」はこの調査対象から省いた。

文語形「なる」の比率は、A 群により高く、B 群により低いという差異があることと、時代を追ってその比率が減少していくことが認められ、出現する記事の文体について調査した結果（表 5）と、連動した傾向ではある。しかし、記事のほとんどが口語体になる 1925 年においても、文語形「なる」は健在であり、文体の変化よりも語法の変化は後れて進んでいることも確かめられる。「太陽コーパス」から得られたデータは、語自体がもつ文体的価値、語が用いられる文章の文体、語の形態、これら三者の関係を考察する好材料を提供している。

語の文体的価値の認定は、従来は内省により行われることが多かったが（国立国語研究所 1972 など）、コーパスを用いることで、実用例に即した多角的な研究を進めることができるようになると思われる。

4．おわりに

以上に記した通り、「太陽コーパス」は、質の高い電子テキストを基盤に置き、そこから様々な言語情報を取り出せるデータベースとして整備する方向で構築を進めており、これを活用することで、対象とする確立期現代語の記述に、新生面を切り開くことが期待できると考えている。

参考文献

小木曽智信（2001）「雑誌「太陽」データベースの XML による表現と処理」『XML pro/con —XML で書く文献学的データ』（古典学の再構築—情報処理班主宰討論会 10 月 27 日

九州大学）発表資料
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―』科研費「新プロ日本語」成果報告書
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」『日本語科学』5（国立国語研究所）
見坊豪紀（1957）「明治時代の文語文」『言語生活』74
国立国語研究所（1972）『動詞の意味・用法の記述的研究』秀英出版
国立国語研究所（1987）『雑誌用語の変遷』秀英出版
近藤泰弘（2001）「Linuxによる言語処理―高速文字列検索を例として―」『日本語学』20-13
鈴木貞美編（2001）『雑誌『太陽』と国民文化の形成』思文閣出版
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』（国立国語研究所）
田中牧郎・小木曽智信（2000）「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（国立国語研究所）
田中牧郎・小木曽智信（2001）「総合雑誌『太陽』の資料性と電子化テキスト」第184回近代語研究会（5月18日 神戸山手大学）発表資料
田中牧郎（2001）「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
築島裕（1986）『歴史的仮名遣い』中央公論社
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
土屋信一（1967）「雑誌「太陽」の用字の変遷」『言語生活』193
永嶺重敏（1997）『雑誌と読者の近代』日本エディタースクール出版部
原口裕（1974）「『みたやうだ』から『みたいだ』へ」『静岡女子大学国文研究』7
松村明（1957）『江戸語東京語の研究』東京堂出版
松村明（1977）『近代の国語―江戸から現代へ―』桜楓社
安永尚志（1998）『国文学研究とコンピュータ』勉誠社

付記

研究室公開では、「太陽コーパス」のデモを行い、コーパスに触れていただく予定です。また、「太陽コーパス」Ver.0.6の試験公開版（1901年の著作権をクリアした記事を対象にしたもの）を、次の利用条件に承諾していただくことで、無料で提供いたします。

- 試験公開の目的は、利用者の意見を完成時の仕様に生かすことです。「太陽コーパス」への意見、要望、誤りの指摘などを寄せてください。
- 「太陽コーパス」の試用版を利用して研究成果を発表する場合は、利用した旨を明記してください。また抜刷・コピーなどを国立国語研究所に寄贈してください。
- 「太陽コーパス」の試用版の一部または全部について、複写物や加工物を流通させてはいけません。

申し込みに際して、利用条件承諾書を提出していただきます。申し込みは研究室公開の会場で受け付けます。