国立国語研究所学術情報リポジトリ

〈共同研究プロジェクト紹介〉独創・発展型：近代語コーパス設計のための文献言語研究 『明六雑誌コーパス』『太陽コーパス』から見る近代語彙

メタデータ	言語: Japanese
	出版者:
	公開日: 2015-10-30
	キーワード (Ja):
	キーワード (En):
	作成者: 田中, 牧郎, TANAKA, Makiro
	メールアドレス:
	所属:
URL	https://doi.org/10.15084/00000727

国語研プロジェクトレビュー
NINJAL Project Review　Vol.4 No.1 pp.18–27（June 2013）

〈共同研究プロジェクト紹介〉
独創・発展型：近代語コーパス設計のための文献言語研究

『明六雑誌コーパス』『太陽コーパス』から見る近代語彙

An Overview of Modern Japanese Vocabulary Based on the “Meiroku Corpus” and “Taiyo Corpus”

田中 牧郎（TANAKA Makiro）

1. はじめに

国立国語研究所では，現代日本語の標準的な言語コーパス（言葉のデータベース）である『現代日本語書き言葉均衡コーパス』を完成させ，近世までの日本語を反映させた「通時コーパス」の設計にも着手している。この二つのコーパスの間をつなぐ近代語のコーパスを設計するために必要な，近代（明治初年から昭和 20 年まで）の文献資料と言語に関わる研究と，そのコーパス化の技術開発の研究を行うプロジェクトを実施した。このプロジェクトでは，今後構築していく近代語コーパスのモデルとして明治前期の日本語を反映する『明六雑誌コーパス』を作成し公開した。また，国立国語研究所では既に明治後期から大正期の日本語を反映する『太陽コーパス』を公開しており，今回のプロジェクトでもこれを発展させた研究も行った[1]。本稿では，『明六雑誌コーパス』と『太陽コーパス』を用いることで可能になる近代語彙についての新しい研究を紹介する。

2. 近代語コーパス設計の論点―資料選定と形態素解析―

本プロジェクトで論点となったことは多様だが，なかでもコーパス化する資料の選定と近代語テキストへの形態素解析技術の実現の二点の重要性が確認された。

資料選定のために，近代の多くの文献資料をリスト化し，その実物を見て検討していったが，近代語の場合，年代・ジャンル・文体などについてコーパス全体の中でのバランスを考慮する「代表性重視」の視点と，個々の資料の言語が持つ特性をよく考慮する「独自性重視」の視点の双方が重要だという方向性が導き出された[2]。代表性と独自性の両側面から重要資料とされたものがいくつかあるが，『明六雑誌』（明治 7（1874）～8（1875）年）もその一つである。『明六雑誌』は，西周や福沢諭吉らが集った明六社が国民への学術啓蒙のために編集刊行したものである。この『明六雑誌』の全文を，今後構築していく近代語コーパスの具体像を示すモデルとしてコーパス化した[3]。

---

[1] 『太陽コーパス』は，CD-ROM 版で公開し（国立国語研究所 2005a），その活用例を国立国語研究所（2005b）で示した。将来的には，『太陽コーパス』も今後開発する近代語コーパスに組み込んでいく計画である。

[2] こうした資料選定の考え方については，田中（2012a）で論じた。

[3] 『明六雑誌コーパス』は，国立国語研究所コーパス開発センターのホームページからダウンロード公開した。http://www.ninjal.ac.jp/corpus_center/cmj/meiroku/。その仕様は，近藤・田中（2012）参照。

形態素解析とは，テキストを単語に分割し，読み・品詞・活用形・語種などの単語に関する情報を付与することであるが，この技術を近代語にも適用する研究を行い[4]，『明六雑誌コーパス』を単語情報付きコーパスとして作成した。分量は延べ語数で約 18 万語となった。

『太陽コーパス』開発当時（2005 年ごろまで）は，形態素解析技術を近代語に適用することは極めて困難な状況にあった。ところが，その後の研究の進展と今回のプロジェクトによって，このコーパスにも形態素解析を適用する目途が立ってきた。ただ，『太陽コーパス』は，『明六雑誌』に比べてはるかに多様なジャンル・文体・著者が含まれているため部分的に解析精度が悪くなることがあり，また，非常に大規模なコーパスであるため誤解析の箇所を人手で修正することも難しい。このような問題を解決するにはまだしばらく研究の段階が必要で，単語情報付きの『太陽コーパス』の公開はもう少し先になる。しかしながら，誤解析を多少含むデータであっても，語彙の総体を概観する場合など，大勢を知るためのデータとしては，すでに利用可能な段階に達している。本稿では，『明六雑誌コーパス』と『太陽コーパス』の双方の語彙の概要を，単語情報付きデータをもとに比較することにする。

3.　語種構成比率の変化

単語情報付きの『明六雑誌コーパス』と『太陽コーパス』とを比較する観点は，語種構成比率である。図 1 は，異なり語数で集計した『明六雑誌コーパス』の語種構成比率と，『太陽コーパス』の年次別のそれとを一つのグラフにまとめて示したものである[5]。

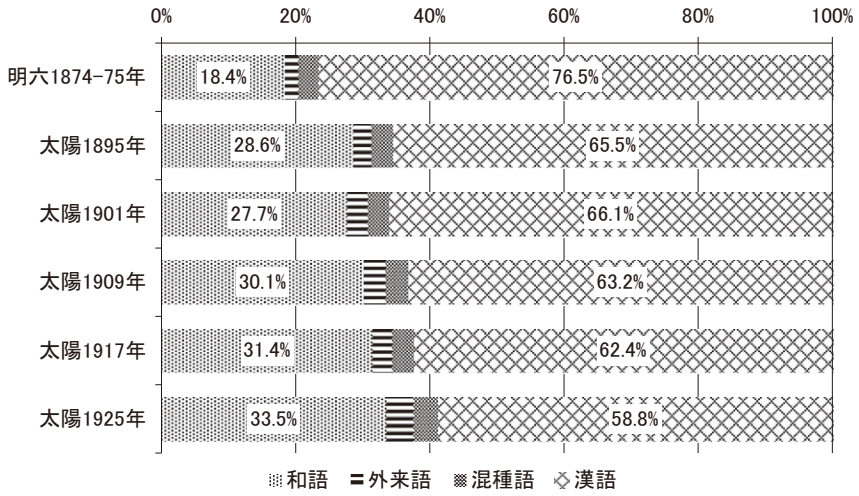

図 1　『明六雑誌コーパス』と『太陽コーパス』の語種構成比率
（異なり語数による集計）

[4] こうした近代語テキストに適用できる形態素解析技術の開発については，小木曽（2012）が論じている。
[5] 『明六雑誌コーパス』の語種構成比率は近藤（2012）に示されている。『太陽コーパス』のそれは，田中（2012b）に示したデータから固有名詞を除外して集計したものをここに提示する。

図 1 から，和語と漢語に関して次のようなことが読み取れる。

・明治前期の『明六雑誌コーパス』では，漢語の比率が極めて高い。

・明治後期から大正期の『太陽コーパス』では，『明六雑誌コーパス』に比べて漢語の比率が大幅に低くなっており，その分和語の比率が高くなっている。

・『太陽コーパス』の内部では，1901（明治 34）年以降徐々に漢語の比率が減少していき，和語の比率が上昇していく。

ここから，明治前期に非常に高かった漢語の比率は明治中期には大幅に減少し，その後も徐々に減少していくことが分かる。明治前期が漢語の氾濫の時期であったことは，これまでの近代語研究でも言われてきたことであるが，それがいつごろどのようにして落ち着いていき，現代のような和語と漢語とが均衡するように安定していったのかについては，よく分かっていなかった。このような語彙の全体的な変化は，単語情報の付いたコーパスを作ることによってはじめて明らかになったことである。

4. 『明六雑誌コーパス』と『太陽コーパス』の語彙比較

単語情報付きコーパスの利点は，3 節で述べたような語彙の全体像を俯瞰できることだけでなく，全体から細部へと焦点を絞り込んでいくような研究が可能になることもあげられる。その事例として図 1 で明らかになった，漢語の減少と和語の増加の内実を調べる研究を紹介しよう。

異なり語数において漢語の比率が減少し，和語の比率が増加したということは，使われていた漢語のある一定数が使われなくなり，使われていなかった和語のある一定数が使われるようになったということを意味している。その使われなくなった漢語と使われるようになった和語とを抽出すれば，語彙の入れ替わりの実際を明らかにすることができると予想される。

このような目的に対しては，『明六雑誌コーパス』と『太陽コーパス』のそれぞれで使われているすべての語彙のリストを作成して相互に対照し，一方にあって他方にない語彙を取り出すことで，データを整えることができる。ただ，低頻度の語彙においてはコーパスへの出現が偶然に左右されているとも考えられるので，一方のコーパスに一定の頻度以上ある語でありながら，他方のコーパスに全く使われていないものを抽出することにした。

まず，『明六雑誌コーパス』で 4 回以上使われていながら，『太陽コーパス』に 1 回も現れない漢語を抽出し，五十音順に 20 語示すと，次の通りである。括弧内にはそのおおよその意味を記した。

安栄◇（安らかに栄えること），易直◆（安らかで素直なさま），一姓◇（出身を同じくすること），遺力◆（残された力），役使◇（命令して使うこと），外顕◆（外に現れること），開交◆（自ら進んで人と交わること），外刺◆（外側から刺激すること），開物◇（人知を開かせるもの），禍鬼◆（災いをもたらすもの），火石◇（煉瓦石のこと），紈袴（がんこ）◆（白い練絹の袴），漢婦◆（中国の女性），気学◆（気象の学問），希教◆（ギリシャ正教），虧損（きそん）◇（徳や利益を損じること），強令◆（強く命令すること），偽論◆（偽りの論），空商◆（投機的な事業），君相◇（君主と宰相）

これらの漢語の歴史上の使用例を『日本国語大辞典第二版』（小学館）で調べると，江戸期までに使用例のある漢語（◇を付した語）もあれば，明治期になって使われ始めた漢語（◆を付した語）もある[6]。抽象概念を表す漢語が多いが，「火石」「紈袴」のように具体物を指し示す漢語も含まれている。いずれも，『太陽コーパス』が対象とする明治後期・大正期だけでなく，現代でも使われない漢語で，明治前期にはよく使われていたこのような漢語が，明治中期までに使われなくなっていったことが分かる。

次に，『太陽コーパス』で一定頻度以上使われていながら，『明六雑誌コーパス』に1回も現れていない和語を抽出する。『太陽コーパス』で一定頻度使われていると認める基準には66回という数値を立てた。この数値は，頻度順に語を並べて，上位の語からその使用頻度を累積していき，その累積度数がコーパス全体の延べ語数の何％を占めるかというカバー率を算出したとき，88％に到達する語の頻度である。先に『明六雑誌コーパス』では4回以上という基準を立てたが，やはりそこがカバー率88％に達する位置である。カバー率を指標に立てることで，規模が大きく違うコーパス同士でも各語が語彙の中で置かれている位置を，正しく比較していくことができるようになる。この基準で『太陽コーパス』でよく使われていながら『明六雑誌コーパス』で全く見られないものとして取り出された和語を，やはり五十音順に20語あげると次のようになる。

あ（感動詞），ああ（副詞），間柄，青，青い，赤い，暁（あかつき），上がり，明るい，秋風，呆れる，飽くまで，開（あ）ける，朝，浅，浅ましい，鮮やか，漁る，汗，遊ばす

これらは，日常的によく使われる語であり，いずれも江戸時代以前から現代まで使われ続けているものである。こうした日常語が『明六雑誌』に用いられていないのは，この雑誌の文章が学術的な論説文であることによるものと考えられる。これに対して『太陽』には，論説文以外にも多様なジャンルの文章がおさめられているため，このような日常語もよく用いられているのだと考えられる。『明六雑誌』に和語の比率が少ないのは，それが書かれた時代の語彙に和語が少なかったことを反映するのではなく，論説文というジャンルの性質を反映するものと見るべきだろう。

以上のことから，『明六雑誌コーパス』と『太陽コーパス』の語彙の比較によって，明治前期から明治後期への語彙の変化をとらえることができる側面（時代的変化により使われなくなった漢語の特定）と，コーパスが反映している言語の層の違いを浮かび上がらせる側面（論説文には使われない和語の特定）とがあると言うことができる。後者の側面は，通時的なコーパスを作る際の注意点であり，今後明治前期のコーパスの設計においては，日常語が反映している資料も加えていくことが求められる。資料選定においては，時間的変異だけでなく，ジャンルをはじめとした社会的変異を十分に反映させていくことの重要性を示す事例だろう。

[6] 『明六雑誌』の漢語を『日本国語大辞典第二版』と比較して，歴史上にこれを位置付ける研究は，本稿とは別の視点から，小野（2012）が展開している。

5. 『太陽コーパス』における語彙の変化

5.1　レベルから見た語種

ところで，『太陽コーパス』の年次による違いは，同一の雑誌を対象にしたコーパスであることから，それがそのまま時代的変化だと考えてほぼ問題ない。しかも，『太陽』はジャンルや著者が非常に多様であり，当時の書き言葉の実態をかなりの程度代表していると見ることができる[7]。その『太陽コーパス』で漢語が減少していく背景には，どのような事情があったのだろうか。

4 節で触れたカバー率で語彙を区画する方法に基づき，『太陽コーパス』の各年次の語彙を，五段階のレベルに区画する。頻度の高い方から順に配列した語彙を，その累積度数のカバー率が，a：78％まで，b：78～88％，c：88～94％，d：94～97％，e：97～100％の基準に沿って五つのレベルに分類した。先の『明六雑誌コーパス』の語彙との比較の際には，この a と b とを扱ったわけである。レベルが a から e へという方向で，基本的な語彙から周辺的な語彙へと進むことになる。各語について，このレベルが年次によって変わらないか変わるか，変わる場合基本的なレベルの方向に変わるか周辺的なレベルの方向に変わるかを見ることで，それぞれが語彙全体の中でどのような位置にあってどのような方向に動いているのかを把握できるようになる。五段階のレベルを組み合わせて，Ⅰ類：基本レベルで不変（約 3500 語），Ⅱ類：中間レベルで不変（約 4500 語），Ⅲ類：周辺レベルで不変（約 37500 語），Ⅳ類：基本レベル化（224 語），Ⅴ類：周辺レベル化（720 語）の五つに分類した。大半の語はⅠ～Ⅲ類であり，この時期の語彙も全体としては安定していると見ることができる[8]。この各類の語種構成比率をグラフで示すと，図 2 のようになる。

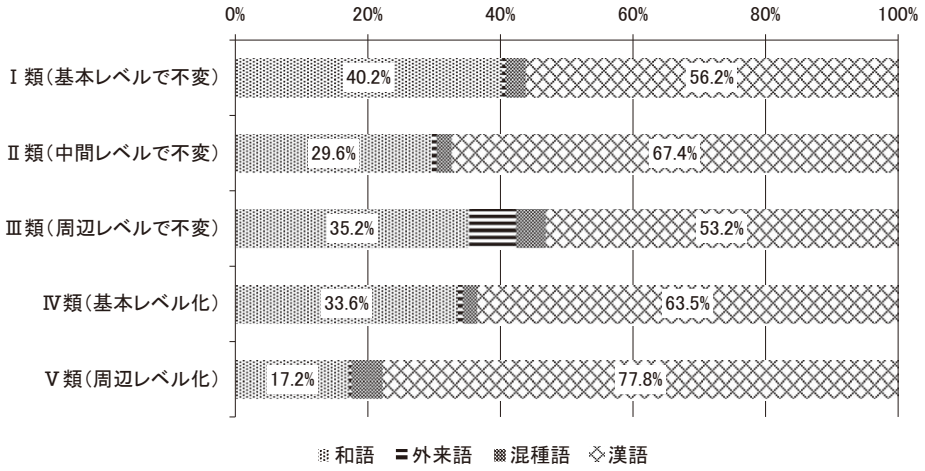

図 2　『太陽コーパス』の語彙のレベルと語種構成比率

[7] 『太陽』の持つこのような性質とこの性質を生かしたコーパス設計については，田中（2005）に述べた。

[8] 約 27700 語はどの類にも入らないが，基本レベルと中間レベル，または中間レベルと周辺レベルの間での移行にとどまるものなどである。

図 2 で漢語に着目すると，次のようなことが読み取れる。

・どの類も漢語の比率が最も高いが，類によってその比率は大きく異なっている。

・レベルが変わらないもの（Ⅰ～Ⅲ類）では，漢語の比率は中間レベル（Ⅱ類）で特に高い。

・レベルが変わるもの（Ⅳ・Ⅴ類）では，漢語の比率は周辺レベル化（Ⅴ類）で極めて高く，基本レベル化（Ⅳ類）でも高い。

これらのことから，近代の漢語は，中間レベルで安定しているものと，周辺レベルや基本レベルに向かって変化するものとに多いことが分かる。中間レベルに漢語が多いことは，現代語の語彙調査でも明らかにされていることで（国立国語研究所 1964），近代語でもそれと同様の傾向にあったことが確かめられる。また，レベルを変化させる語にも漢語が多いことは，通時的な単語情報付きコーパスを作成したことによってはじめて分かったことである。そして，周辺語化する漢語が多いことは，語彙における漢語の比重が低下していったことを意味するものであり，4 節で見た漢語の比率の低下と軌を一にする現象と言えるだろう。さらに，基本語化する漢語も多い事実が明らかになったことは，これまであまり指摘されていないことであり，重要な発見であると思われる。日本語の歴史を考える観点からは周辺語化する漢語はどのようなもので，基本語化する漢語はどのようなものなのかという点に特に注目が集まるだろう。

5.2　周辺語化する漢語

Ⅴ類の周辺語化する漢語として抽出されたものには，いくつかの種類がある。まず，「囂々（ごうごう）」「嘖々（さくさく）」「燦爛（さんらん）」「駸々（しんしん）」「紛々（ふんぷん）」「爛漫（らんまん）」「寥々（りょうりょう）」など，タリ活用形容動詞の一群がある。これらは，平安時代以来の漢文や漢文訓読文の系統で使われ続けてきたもので，明治前期の書き言葉でもよく使われていたが，書き言葉が口語体化していく流れのなかで，使われる機会が減ってきたものである。このほか，「佳なり」「緩なり」「按ずる」「詠ずる」「賀する」など，漢字一字の形容動詞やサ変動詞が周辺語化するのも，文語体書き言葉から口語体書き言葉への移行の流れに乗るものだと思われる。

一方で，「強兵」「工兵」「兵馬」「兵備」あるいは「進軍」「進撃」「要塞」「陥落」といった戦争に関わる多くの漢語が周辺語化したり，「国文」「英文」「文体」「字形」「和漢」などの言語に関わる多くの漢語など，特定の分野の語彙が周辺語化したりしていく様子もうかがえる。これらはおそらく，明治後期が日清・日露の戦争の時代であり，言文一致運動など言語改革の時代であったのに対して，大正期には戦争も言語もあまり注目されなくなっていったという話題の変化を反映するもので，日本語の語彙自体が変わっていったものではないと思われる。

以上は，周辺語化する理由が分かるものだが，なぜ周辺語化するのかが想定できないものも，「百般」「等級」「兼任」「未完」「国立」「撤去」など，極めて多く存在する。これについては，5.4 節で述べる。

5.3　基本語化する漢語

次に，Ⅳ類として抽出された基本語化する漢語はどのようなものだろうか。例えば「結構」が基本語化するのは，この語が書き言葉よりも話し言葉によく用いられるものであることから，文語体書き言葉では避けられていたものが，口語体書き言葉が一般化するにしたがって普通に使われるようになっていったものと考えられる。また，「投資」「増刊」「入院」「入党」といった語が基本語化していくのは，それぞれ経済・出版・医療・政治の分野が社会的に発達し，それぞれの活動が盛んになっていくのに伴って，これらの語が必要とされるようになっていったことによると考えられる。

ところが，基本語化する漢語の多くは，そのような口語体書き言葉の一般化や新概念の必要性ということからは説明できない。「解決」「恐縮」「目標」「手法」「興奮」などが，その例であるが，これらは決して話し言葉で使われていたものではなく，またそれぞれ「解く」「恐れる」「目じるし」「手立て」「高ぶる」など，同じような意味を持つ和語が従来からあり，特に新しい概念を表しているわけでもない。基本語化する漢語の多くは，その理由がすぐには想定できないものなのである。

5.4　新しい語彙体系の形成

周辺語化や基本語化が進む理由がはっきりしない漢語の背景には何があったのだろうか。そうした漢語の用例をコーパスから抽出して観察するうちに，ある語彙体系上の特徴を共通して持っていることが見えてきた。例えば，「あらわす」「あらわれる」の類義語である漢語（混種語を含む）には，「実現」「表現」「出現」「現出」「発現」「顕現」「現ずる」などがあるが，これらについて，『太陽コーパス』各年次の 100 万字あたりの出現率を計算してグラフに示すと，図 3 のようになる。

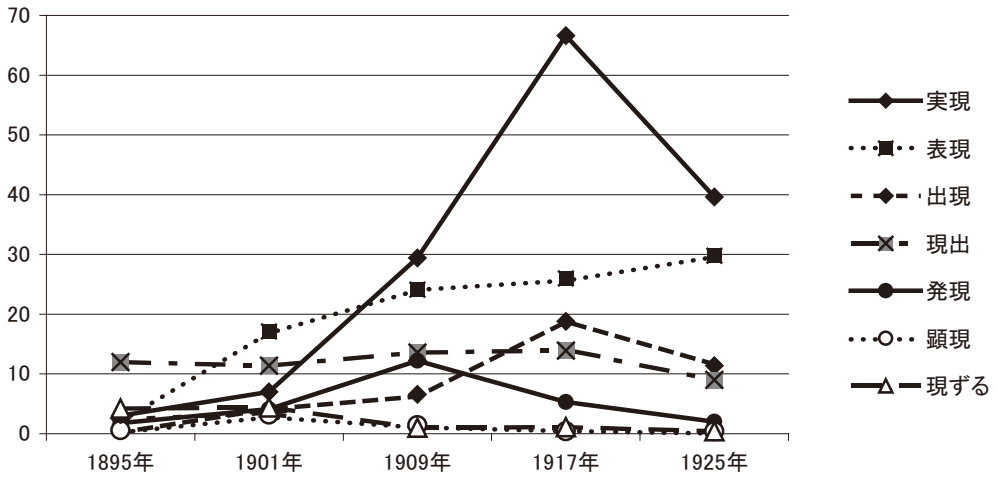

図 3　『太陽コーパス』における「あらわす」「あらわれる」の漢語類義語の出現率

「実現」「表現」は，当初は極めて低頻度であったところから次第に使用頻度を増し，基本語化の方向にあることが明らかで，「出現」もそれらに準じるものだと見てよさそうである。一方，当初よく使われていた「現出」は，最後の年次では大きく頻度をさげており，「発現」も一時増加するものの結局は減少傾向が鮮明になり，これらは周辺語化の方向にあることが見て取れる。また，「顕現」「現ずる」はずっと低頻度で，周辺語の位置に置かれたままである。

これらの語について，コーパスの用例をもとに，共起する語句の分析からその意味を考察したところ，「実現」「表現」「出現」は，その意味に変化が起こっていることが明らかになった。まず「実現」は，当初，現実化が困難なことを現実のものとするという意味だったが，現実の活動や成果として形にするという意味へと変化していく。また，「表現」は，もとは，詩歌や言語が事実を表したり，文字が言語を表したりする意味だったが，作家や画家が言葉や絵に表したり，内面にある見えないものを見えるようにしたりする意味に変わっていく。そして，「出現」は，もともと，思い描いていたことが結果として表に出ることを広く意味していたが，次第に，具体的な事物がその姿を見せる意味に限定されていき，やはり意味変化があったと言える。一方，周辺語化したり周辺語の位置にあり続ける「現出」「発現」「顕現」「現ずる」の 4 語の用例を分析すると，意味変化を起こしておらず，他の類義語との意味の違いが明確でないものが多いことが分かった。つまり，基本語化する漢語は，語彙体系の中に自らの位置を明確に定めていくように意味を変えていく動きが顕著であるのに対して，周辺語化したり周辺語にとどまったりする語は，そうした動きが見られないのである。

このように個々の漢語が語彙のレベルを変化させる動きは，その表す意味を明確にすることで語彙体系の中での位置を安定化させる動きを持つか持たないかに左右されるのである。「あらわす」「あらわれる」の類義語群以外の漢語についても，基本語化するものと周辺語化するものとは同様の意味上の特徴を持つものが多い。このような用例分析に基づく意味変化の研究にも，コーパスは非常に有用なデータを提供するのである。

6. 近代語コーパスの構築に向けて

4 節で見たように，『明六雑誌』一資料だけでは，明治前期の日本語を代表するコーパスとしては不十分であることは明らかであり，この時期については今後様々な資料を加えていく必要がある。また，『太陽』の時代に至るまでの明治中期の資料のコーパス化も必要である。さらに，『明六雑誌』『太陽』も含めた近代語コーパス全体の資料配分などの具体的設計に着手していくべき段階が来ている。

5 節で述べたように，単語情報付きの『太陽コーパス』を分析することによって，語彙の全体像を把握した上で特定の部分や変化のある部分に焦点を絞り込んでいくような研究が可能になり，さらに用例を詳細に分析することで，意味変化の方向性をとらえる研究も可能になる。これらのことについては，田中（2013 予定）に詳しく記したが，コーパスは日本語の歴史の研究にも強力なツールになるのである。

●参照文献●

国立国語研究所（1964）『現代雑誌九十種の用語用字（3）分析』東京：秀英出版.
国立国語研究所（2005a）『太陽コーパス―雑誌『太陽』日本語データベース―』東京：博文館新社.
国立国語研究所（2005b）『雑誌『太陽』による確立期現代語の研究―『太陽コーパス』研究論文集―』東京：博文館新社.
近藤明日子（2012）「『明六雑誌コーパス』の語彙量」田中・岡島ほか，144−149.
http://www.ninjal.ac.jp/corpus_center/cmj/doc/08kondo.pdf
近藤明日子・田中牧郎（2012）「『明六雑誌コーパス』の仕様」田中・岡島ほか，118−143.
http://www.ninjal.ac.jp/corpus_center/cmj/doc/07kondo.pdf
小木曽智信（2012）「近代語テキストの形態素解析」田中・岡島ほか，83−92.
http://www.ninjal.ac.jp/corpus_center/cmj/doc/05ogiso.pdf
小野正弘（2012）「文献資料内漢語の階層化―『明六雑誌』の漢語をめぐって―」田中・岡島ほか，169−180.
http://www.ninjal.ac.jp/corpus_center/cmj/doc/10ono.pdf
田中牧郎（2005）「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」国立国語研究所（2005b），1−48.
田中牧郎（2012a）「近代語コーパスにおける資料選定の考え方」田中・岡島ほか，13−26.
http://www.ninjal.ac.jp/corpus_center/cmj/doc/01Tanaka.pdf
田中牧郎（2012b）「明治後期から大正期の語彙のレベルと語種―『太陽コーパス』の形態素解析データによる―」田中・岡島ほか，153−168.
http://www.ninjal.ac.jp/corpus_center/cmj/doc/09Tanaka.pdf
田中牧郎（2013 予定）『近代書き言葉はこうしてできた』東京：岩波書店.
田中牧郎・岡島昭浩・小木曽智信・小野正弘・小島聡子・島田泰子・朱京偉・高田智和・張元哉・陳力衛・近藤明日子・須永哲矢（2012）『近代語コーパス設計のための文献言語研究 成果報告書』（国立国語研究所共同研究報告 12-03），東京：国立国語研究所.
http://www.ninjal.ac.jp/corpus_center/cmj/cmj-doc/

《要旨》 本プロジェクトでは，通時的な日本語コーパスの一部として必要な近代語のコーパスを設計するための研究を実施した。本プロジェクトで作成した『明六雑誌コーパス』は，単語に関する詳細な情報が付与されたはじめての近代語コーパスである。また，2005 年に公開した『太陽コーパス』に対しても詳細な単語情報を付与する試行を行った。明治期から大正期を対象とするこれら二つのコーパスデータを用いて，近代語彙の変化を概観する研究を行った。その結果，漢語の数が減少し，一部の漢語が基本語化していったことが明らかになった。さらにまた，基本語化した漢語は既存の基本語との間に，意味的に使い分けられることも明らかになった。これらは，明治から大正期に新しい語彙体系が形成されていったことを示している。

Abstract: This paper reports research conducted in order to design the corpus of modern Japanese that will be an essential part of the planned diachronic Japanese corpus. The “Meiroku Corpus,” which was built in our project as a model for the modern Japanese corpus, is the first modern corpus with detailed morphological annotation. We also carried out trial annotations of

the same kind for the "Taiyo Corpus," built in 2005. Using these two corpora, which cover Japanese in the Meiji and Taisho periods, we investigated changes in the modern Japanese vocabulary. The results show that the number of Sino-Japanese words decreased but that some Sino-Japanese words were gradually incorporated into the basic vocabulary. It also became clear that these newly incorporated words were semantically differentiated from basic words that already existed. These findings show that a new vocabulary system was formed in Meiji and Taisho periods.

田中 牧郎（たなか・まきろう）

国立国語研究所言語資源研究系准教授。文学修士（東北大学）。昭和女子大学講師，国立国語研究所研究員，同グループ長などを経て，2009 年 10 月より現職。

主な著書：『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―』（共編著，博文館新社，2005），『分かりやすく伝える外来語言い換え手引き』（共編著，ぎょうせい，2006），『病院の言葉を分かりやすく―工夫の提案―』（共編著，勁草書房，2009），『図解 日本の語彙』（共編著，三省堂，2011），『外来語研究の新展開』（共編著，おうふう，2012）.

社会活動：日本語学会評議員，日本語学会『日本語学大辞典』編集委員主任，言語処理学会編集委員.

独創・発展型共同研究プロジェクト「近代語コーパス設計のための文献言語研究」
プロジェクトリーダー　田中牧郎
（国立国語研究所　言語資源研究系　准教授）

プロジェクトの概要

近世までの日本語を対象とする「通時コーパス」と『現代日本語書き言葉均衡コーパス』との間をつなぐものとして，近代語のコーパスを設計する研究を行った。国立国語研究所がこれまでに作成した『太陽コーパス』などを踏まえ，近代語コーパスの原型を作り，これを使ってコーパス近代語研究の領域を開拓する方法も研究した。また，コーパスの対象にする文献の選び方を検討し，コーパス化する文献の言語に形態素解析を施す方法についても研究した。