

国立国語研究所学術情報リポジトリ

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

| | |
|-------|---|
| メタデータ | 言語: jpn 出版者: 公開日: 2020-03-18 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属: |
| URL | https://doi.org/10.15084/00002770 |

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 牧郎 (国立国語研究所言語資源研究系)¹

1. はじめに

明治期に多くの漢語が日本語に取り入れられたが、大正・昭和・平成と進むにつれて漢語は減少していくことが明らかにされている(国立国語研究所 1964、国立国語研究所 1987、国立国語研究所 2005a)。一方、基本語彙の中に占める漢語の比率は、次第に増加していくという報告もある(飛田 1966、宮島 1967)。また、明治前期から徐々に取り入れられた外来語は、大正期から増加傾向が顕著になり、第二次大戦中にいったん減少した後、昭和時代の終わりまで増加の一途をたどる(橋本 2010)。このように、近代から現代にかけての語彙の歴史は、語種の観点からみたとき、大きな変化があることが明らかにされている。従来の研究は、漢語あるいは外来語という、ある語種に光をあててその歴史的特徴を明らかにしてきたが、和語をも含めた語種全体を見わたした歴史については、十分に解明できていないところがある。和語を含めて全体的な視点をもつことによって、漢語や外来語の歴史についても、新たな視点からその特徴を見直していくことができるのではないかと思われる。

しかし、そのような研究を行う前提となる、語彙全体の実態把握を行うことは容易でなかった。現代語については、国立国語研究所による語彙調査のデータはあったが、ある時代は雑誌、別の時代は教科書、さらに別の時代にはテレビ放送といったように特定の媒体の調査であり、日本語全体として語彙がどのように変容したのかについて考えるほどにはデータが示されてはいなかった²。まして、近代語については、ごく一部の資料にしか調査データが存在しておらず(国立国語研究所 1959、1985-1997 など)、その語彙の変容を記述することは難しかった。

ところが、本プロジェクトなどで、明治以後の近代日本語のコーパス構築と関連技術の整備に着手したことにより、語彙全体を射程に入れた近代語彙史の体系的な記述を行える状況に近づきつつある。本稿では、公開済みの『太陽コーパス』に対して、本プロジェクトなどで整備中の形態素解析辞書「近代文語 UniDic」を用いて形態素解析を施すことで、明治後期から大正期の語彙の体系的な変化を語種の視点からとらえる研究例を示したい。

2. 『太陽コーパス』への「近代文語 UniDic」の適用

2.1 『太陽コーパス』

『太陽コーパス』(国立国語研究所 2005a)は、博文館から刊行された総合雑誌『太陽』(1985~1928年)を対象としたコーパスである。1895(明治28)年、1901(明治34)年、1909(明治42)年、1917(大正6)年、1925(大正14)年の5年分の全文(著作権処理ができなかった記事を除く)を対象にしている(田中 2005)。この『太陽』は、分量の多さ、ジャンルの広さ、執筆陣の多彩さ、読者層の厚さなどの点で、当時の文献資料としては格別の価値を持っていることから、何か一つの資料で当時期の書き言葉を代表させるとしたら、おそらく筆頭に挙げてよい資料の一つである。

¹ mtanaka@ninjal.ac.jp

² 国立国語研究所(1987)は、雑誌という一媒体に限られるが、時間軸による語彙の変化をとらえようとしていて点で特徴的である。調査データの量は少ないが、現代語彙を通時的に扱った最初の研究として価値が高い。

2.2 「近代文語 UniDic」

分かち書きがされない日本語は単語認定の複雑さが大量のデータに対する語彙調査の制約になっていたが、近年、国立国語研究所が中心に開発を進めている形態素解析辞書「UniDic」は、従来の人手による語彙調査で実績のある言語単位に基づく斉一な単位での解析を可能にしたことと、階層構造を持たせることで品質管理や同語異語判別等の便宜を向上させたことの2点が特筆される（伝ほか 2007）。この UniDic をもとに近代語資料に対する形態素解析を可能にしたものに「近代文語 UniDic」があり（小木曾 2009 など）、本報告書におさめられている小木曾論文、須永・近藤論文³にも記載のある通り、実用化に見通しが立ちつつある。

2.3 『太陽コーパス』に対する「近代文語 UniDic」による自動形態素解析

この「近代文語 UniDic」によって、『太陽コーパス』に自動形態素解析を施す研究の現状は、口語体の部分では誤解析が少なくないこと、文語体の部分でも語や表記によっては誤解析が生じる場合が残されているなど、自動形態素解析の結果をそのまま無条件に利用できる段階には至っていない。

しかしながら、高精度の解析が実現されるまで待たないと、『太陽コーパス』の形態論情報を利用した研究は行えないと考えるよりも、データの完成度が低い段階でも、データに誤りが含まれる可能性には十分留意しつつも、形態論情報を使うことで可能になる新たな研究領域を開拓していくべきだと考えることの方が、建設的だろう。そこで、本稿では、『太陽コーパス』に対して、「近代文語 UniDic」による自動形態素解析を施し、その結果を用いた研究を試みることにする。具体的には、解析結果データをもとに、年次別の語彙頻度表を作成し、語彙頻度によって語彙をレベルに分け、そのレベルを指標として、語種の観点からみた語彙の変化の実態を把握することを試みる。

3. 『太陽コーパス』の語種比率

「近代文語 UniDic」は、直接的には文語文を対象とするものであるが、ここでは、口語文も含めた『太陽コーパス』の全体を対象とした。文語文に比べて口語文は解析精度が悪くなるものの、決定的に劣るというわけではなく、口語文においても大部分は正しく解析できる。口語文を対象から外すと、新しい年次（1909 年以後）の分量が、かなり少なくなってしまう、経年的比較が難しくなってしまう。それよりも、多少精度が低くても全体を扱って、そのようなデータでも活用が可能な研究を展開するのがよいと考えた。

『太陽コーパス』全体に対して、「近代文語 UniDic Ver.1.2」(MeCab 版)を用いて自動形態素解析を実施した。UniDic が規定する品詞体系のうち、記号・付属語・未知語は除外した。また、UniDic の語種情報は、和語・漢語・外来語・混種語・固有名詞・記号の六種に分かれるが、このうち「記号」はアルファベット略語の類が分類されており、これは「外来語」にまとめた。その五種類の語種の年次別の語数について、延べ語数、異なり語数を集計したものが、表 1・表 2 である。表 1・表 2 をもとに、語種比率を見るためにグラフ化したものが、それぞれ図 1・図 2 である。

図 1・図 2 を一見すると、『太陽コーパス』において、語種別の比率は年次によって大きな変動はないように見える。語種から見た語彙のありようは、明治後期から大正期にかけて、大きな変化はなかったというように見ることもできそうである。しかし、よく見ていくと、わずかずつではあるが一定の方向での変化も見られ、それはこの時期の語彙の歴史として重要な側面を浮かび上がらせているのではないかと考えられる。

³ 小木曾智信「近代語テキストの形態素解析」、須永哲矢・近藤明日子「近代語コーパスのための形態論情報付与規程の整備」(いずれも本報告書所収)。

表1 『太陽コーパス』の年次別・語種別語数(延べ語数)

| 語種 | 1895年 | 1901年 | 1909年 | 1917年 | 1925年 | 全体 |
|------|--------|--------|--------|--------|--------|---------|
| 和語 | 639896 | 574523 | 518387 | 484725 | 452507 | 2670038 |
| 漢語 | 566709 | 530110 | 453738 | 421473 | 353699 | 2325729 |
| 外来語 | 5499 | 6545 | 4755 | 4452 | 7027 | 28278 |
| 混種語 | 32553 | 30537 | 24295 | 21839 | 16937 | 126161 |
| 固有名詞 | 66852 | 47491 | 43237 | 40191 | 36902 | 234673 |

表2 『太陽コーパス』の年次別・語種別語数(異なり語数)

| | 1895年 | 1901年 | 1909年 | 1917年 | 1925年 | 全体 |
|------|-------|-------|-------|-------|-------|-------|
| 和語 | 11543 | 10026 | 9781 | 9818 | 10761 | 17878 |
| 漢語 | 26456 | 23947 | 20526 | 19485 | 18883 | 35023 |
| 外来語 | 1128 | 1133 | 1095 | 947 | 1321 | 2886 |
| 混種語 | 1297 | 1152 | 1092 | 1009 | 1167 | 2177 |
| 固有名詞 | 9349 | 6791 | 5889 | 5128 | 6089 | 16125 |
| 計 | 49773 | 43049 | 38383 | 36387 | 38221 | 74089 |

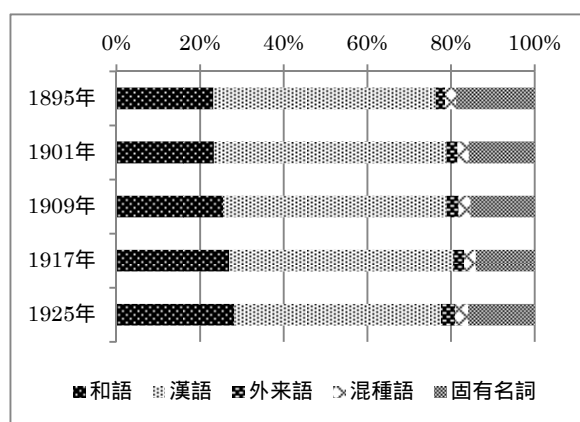
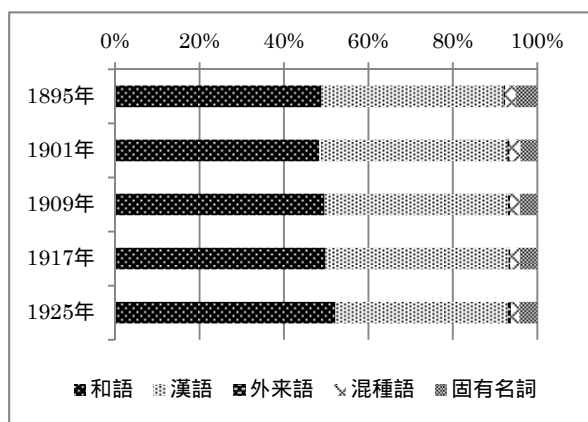


図1 『太陽コーパス』年次別語種比率(延べ語数) 図2 『太陽コーパス』年次別語種比率(異なり語数)

まず、図1で延べ語数における語種比率を見ると、和語の比率は各年次50%前後であるものの、よく見ると、その比率が年次を追って少しずつ増加していることに気づく。一方、漢語を見ると、各年次40%数%であるが、わずかながら減少していることが分かる。外来語・混種語・固有名詞は、いずれも非常に少なく、年次による変化もとらえにくい。つまり、延べ語数では、語種構成に大きな変化はないが、年次を追って少しずつ、和語が増加しその分漢語が減少していった様子が見て取れるのである。

次に、図2で異なり語数における語種比率を見ると、やはり和語の増加と漢語の減少を確かに見て取ることができ、その増減の幅は延べ語数の場合よりもやや大きいことが分かる。また、外来語が1925年で比率を高めていることもとらえることができる。

このように、『太陽コーパス』における語種構成には、年次による大きな変化はないものの、和語の増加とその反面である漢語の減少が、確かな変化として認められ、大正後期に

は外来語の増加も見え始めるのである。

4. 『太陽コーパス』の語彙のレベル分け

4.1 語彙のレベルの考え方

国立国語研究所(1964)は、語彙を使用頻度によって階級に分けた場合に、階級によって語種構成比に違いがあることを明らかにしている。昭和中期(1956年)に刊行された雑誌90種の語彙調査で得られたデータをもとに、使用頻度別に七つの階級に分け、各階級の語種構成比率を示す、次のグラフを掲載している。

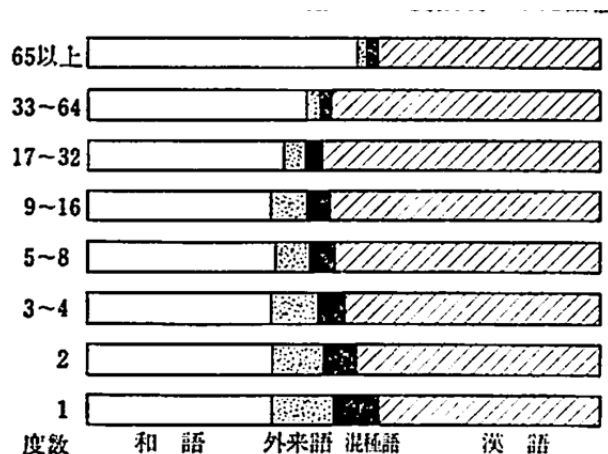


図3 雑誌90種調査における度数階級別の語種構成比(国立国語研究所1964から転載)

図3によると、最も使用頻度の高い階級では和語の比率が高く、漢語は和語よりも比率が低く、外来語・混種語の比率は極めて低い。使用頻度の低い階級に進んでいくほど、和語の比率は徐々に低くなるが、4番目の階級から後は、比率が不変となる。これに対して、漢語は、最も使用頻度の高い階級では和語よりも低い比率であるが、2番目の階級では比率を大きく高めて和語を上回り、3番目の階級で最も高くなる。その先は使用頻度の低い階級に進むほど、比率を低下させていく。そして、外来語と混種語は、使用頻度の低い階級に進むにつれて、その比率を高めていき、度数1の階級に至るまでその傾向が続いていく。以上のことから、和語は高頻度の階級で多く、漢語は中頻度の階級で多く、外来語と混種語は低頻度の階級で多いという傾向があるということができよう。

このような使用頻度によって語彙を階級に分ける考え方は、語彙の特徴を研究する際に有効だと考えられ、とりわけ、語種の枠組みを用いて記述しようとする場合は、たいへんに役に立つのではないかと思われる。そこで、本稿では、使用頻度によって分ける語彙の階級を「レベル」と呼び、その概念を明確化し、この枠組みを用いて、明治後期から大正期の語彙のありようを、語種の視点で記述していきたい。

4.2 カバー率によるレベル分け

4.1で引いた、国立国語研究所(1964)の語彙頻度に基づくレベルは、ある1時点におけるひとつの調査によるものであった。ところが、『太陽コーパス』が対象とする5つの年次それぞれで語彙レベルに分けて相互に比較しようとする、この方法では問題が生じる。なぜなら、各年次で延べ語数が異なるために、レベルを区画する使用頻度をどこに設定するかという、線引きの困難さに直面するからである。

この問題に対応するために、使用度数の高いものから順に各語の度数を累積していき、その累積度数が延べ語数の何パーセントを占めるかという、カバー率(累積度数占有率)を算出し、この数値をもとに一定の基準を定めてレベル分けを行うことにした。五つの年

次を比較するには、あまりレベルの区画が多いと繁雑になると考え、5段階に分けることとした。その具体的なカバー率の基準と、各年次にこれを適用した頻度区間を示したのが、表3である。

表3 『太陽コーパス』年次別の語彙のレベルと頻度区間

| レベル | カバー率 | 1895年 | 1901年 | 1909年 | 1917年 | 1925年 | 全体 |
|-----|---------|-------|-------|-------|-------|-------|--------|
| a | -78% | -46 | -51 | -54 | -53 | -40 | 202- |
| b | 78-88% | 45-17 | 50-19 | 53-19 | 52-19 | 14-39 | 66-201 |
| c | 88-94% | 16-7 | 18-8 | 18-8 | 7-18 | 13-6 | 24-65 |
| d | 94-97% | 6-4 | 7-4 | 7-4 | 6-4 | 5-3 | 11-23 |
| e | 97-100% | 3-1 | 3-1 | 3-1 | 3-1 | 2-1 | 1-10 |

* 各年・全体の数字は使用頻度の区間を示す。

最も高頻度のレベルaの基準を、カバー率78%までと定めたところ、1895年では使用頻度46以上の語まででこの基準に達し、1901年では51以上のところでこの基準に達した。次のレベルbは78%から88%と定めると、1895年では使用頻度17以上45以下の語がここに配された。このようにして、各年次のすべての語をいずれかのレベルに配属させた。その結果の語数をまとめたものが、表4である。

表4 『太陽コーパス』年次別・レベル別の語数

| レベル | 1895年 | 1901年 | 1909年 | 1917年 | 1925年 | 全体 |
|-----|-------|-------|-------|-------|-------|-------|
| a | 3915 | 3156 | 2680 | 2531 | 2928 | 3476 |
| b | 4760 | 3839 | 3364 | 3112 | 3889 | 4835 |
| c | 8083 | 5995 | 5441 | 5652 | 5869 | 8334 |
| d | 7314 | 7389 | 6102 | 5063 | 7544 | 10177 |
| e | 25701 | 22720 | 20796 | 20029 | 17991 | 47267 |
| 計 | 49773 | 43049 | 38383 | 36387 | 38221 | 74089 |

* 各年・全体の数字は配分された語数を示す。

4.3 レベル別の語種比率

次に、各年次別にレベルごとの語種構成比率を算出し、グラフに表示してみよう(図4~8)。どの年次も、和語はレベルaで最も高く、漢語はレベルb・c・dといった中間的なレベルで高く、外来語はレベルeで最も高いという特徴を共通してもっている。これは、4.1 で見た昭和中期の特徴と同じであり、明治後期以来、この特徴は変わらなかったと考えられる。

それでは、レベル別の語種構成比率に、時代による変化は認められないのだろうか。そのことを調べるために、各年次ごとに、レベル別の和語比率、漢語比率を算出し、その経年変化が分かるように折れ線グラフを作成した(図9~10)。その際、固有名詞は除外し、和語・漢語・外来語・混種語の四つの合計の中で、和語・漢語それぞれが占める比率を算出した。

図9で和語比率の変化を見ると、どのレベルでも上昇していくが、当初低かったレベルb・c・dでの上昇の度合いがやや大きい。また、図10の漢語比率を見ると、和語の場合とは反対に低下していくが、当初高かったレベルb・c・dにおいて低下の度合いがや

や目立っている。このように、中間的なレベルの語彙ほど、漢語から和語へという移行がより目立っていると見ることができよう。

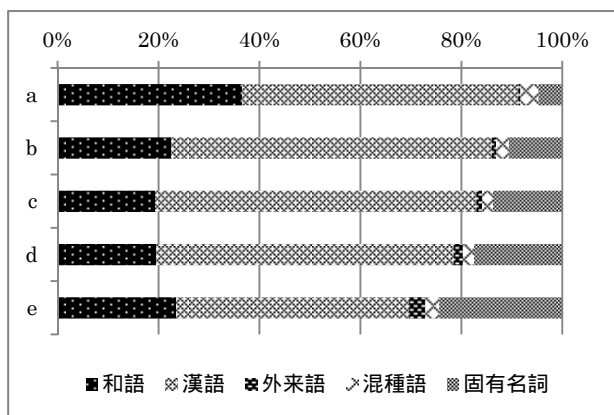


図4 1895年のレベル別語種比率

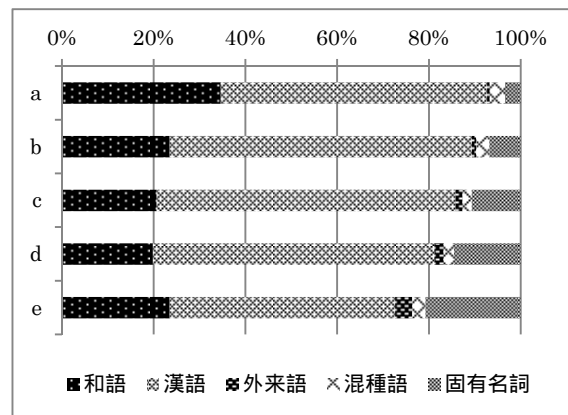


図5 1901年のレベル別語種比率

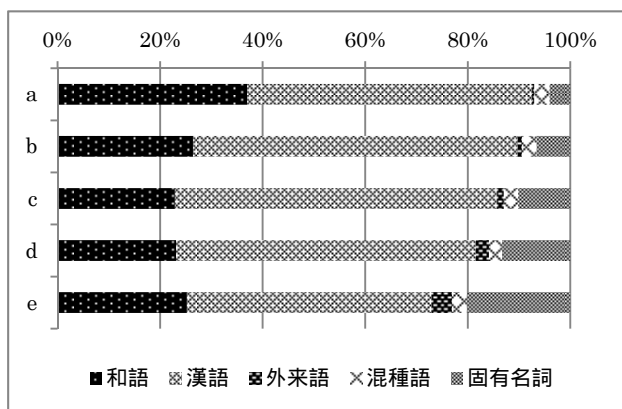


図6 1909年のレベル別語種比率

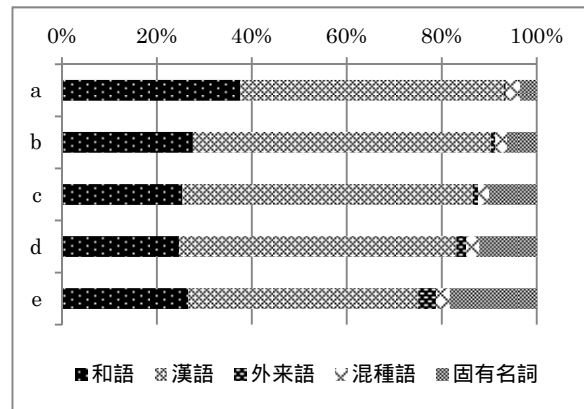


図7 1917年のレベル別語種比率

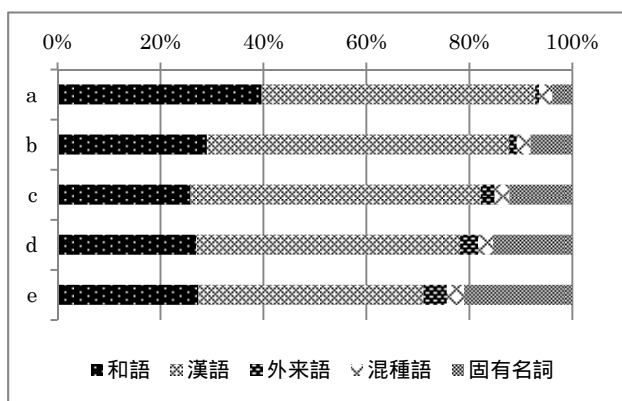


図8 1925年のレベル別語種比率

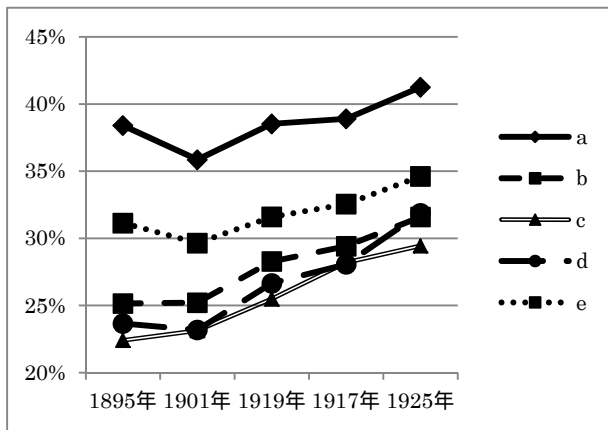


図9 レベル別の和語率の推移

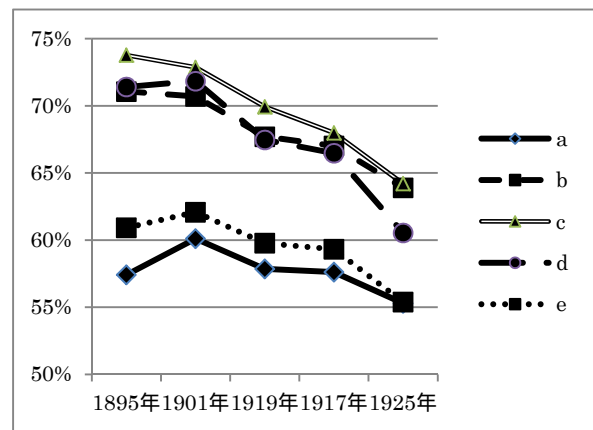


図10 レベル別の漢語率の推移

5. レベルの変動による類型化

5.1 類型化の基準

次に、個々の語のレベルが、『太陽コーパス』が対象とする5つの年次でどのように変わる(変わらない)のかという点から、語彙の類型化を考えてみよう。年次を通じて、レベルが不変な語彙もあれば、一定の方向にレベルが変化していく語彙もある。この、レベルの不変と変化の観点から、特徴的な語彙を抽出して類型化することを考えると、表5のような五つの類型にまとめるのが分かりやすいと思う。

表5 レベルの不変と変化による類型化

| 類型 | 定義 | 基準例 | 語数 |
|-----|------------|----------------------|---------------|
| 類型 | 基本的なレベルで不変 | 全年次を通じてレベルaまたはb | 3523 (4.8%) |
| 類型 | 中間的なレベルで不変 | 全年次を通じてレベルbまたはcまたはd* | 4328 (5.8%) |
| 類型 | 周辺的なレベルで不変 | 全年次を通じてレベルeまたは「なし」 | 37522 (50.6%) |
| 類型 | 基本的なレベルに変化 | 図11の濃い網掛け部分 | 231 (0.3%) |
| 類型 | 周辺的なレベルに変化 | 図12の濃い網掛け部分 | 795 (1.1%) |
| 類型外 | 特定の傾向なし | 上記以外 | 27690 (37.4%) |
| 全体 | | | 74089 (100%) |

*すべての年次でレベルbのものは、ここに入れずに類型とする。

類型 は、基本的なレベルで安定している語彙である。最も基本的なレベルのaであり続けるものだけを基準にとれば1650語になるが、ここでは少し基準を緩くして、aまたはbのいずれかのレベルにおさまっているものという基準を立てたところ3523語となった。次に、類型 として、中間的なレベルで変わらない語彙を、全年次でレベルbからdの間に入っているものという基準を立てたところ4328語となった(全年次レベルbのものは、ここに入れずに類型 に入れる)。そして、類型 として周辺的なレベルで不変な語彙を、全年次でレベルeまたは「なし」のいずれかのものという基準で算出したところ、37522語となった。「なし」とは、当該の年次には使用例がないものである。以上の三つの類型は、ある一定のレベルの範囲で変化しないものであったが、一定の方向のレベルに移行していくものもある。類型 は、より基本的なレベルへと変化するもので、図11の網掛け部分におさまるとい基準を立てると、231語となる。また、類型 は、図

12の網掛け部分におさまるという基準で、795語となった。

| | 1895年 | 1901年 | 1909年 | 1917年 | 1925年 | | 1895年 | 1901年 | 1909年 | 1917年 | 1925年 |
|----|-------|-------|-------|-------|-------|----|-------|-------|-------|-------|-------|
| a | | | | | | a | | | | | |
| b | | | | | | b | | | | | |
| c | | | | | | c | | | | | |
| d | | | | | | d | | | | | |
| e | | | | | | e | | | | | |
| なし | | | | | | なし | | | | | |

図 11 類型（基本レベル化）とするもの 図 12 類型（周辺レベル化）とするもの

前段落で示した基準によって五つの類型に分類すると、そのいずれにもあてはまらない類型外のもの27690語残る。表5の基準は絶対的なものではなく、もっと緩い基準にすれば、類型外だった語をいずれかの類型に入れることもできる。ただ、あまり特徴のはっきりしたものも類型にまとめてしまうという問題が生じるため、ここでは、表5の基準で分類した。表5を見ると、類型（周辺のレベルで不変）の語彙が最も多く、他の類型を圧倒している。多くの語は、周辺のレベルにあってその位置を変えないということが分かる。類型に比べればかなり少なくなるが、次に多いのが、類型（中間的なレベルで不変）そして、類型（基本的なレベルで不変）である。類型（基本的なレベルに移行）または類型（周辺のレベルに移行）の、レベルが一定の方向に変化していくものは、少数派である。

5.2 各類型の語種構成

こうして設定した類型ごとに語種構成比率を調査すると図13のようになる。この図から次のようなことが読み取れる。

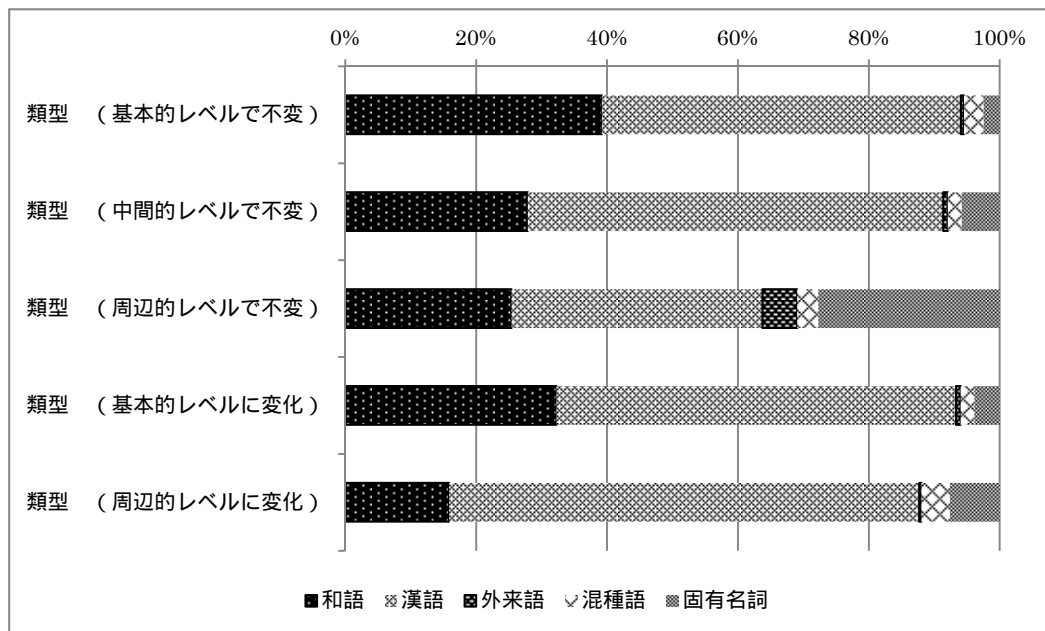


図 13 類型別の語種比率

まず、どの類型においても漢語の比率が最も高く、漢語は類型の違いにかかわらずもっとも多い語種であることが確認できる。類型別に見ていくと、まず、類型（基本レベルで不変）では、他の類型に比べて和語の比率の多さが目立ち、反対に、類型（中間レベルで不変）では、漢語の多さが目立つ。このことは、さきに4節で見た、基本的なレベルには和語が多く、中間的なレベルには漢語が多いという特徴が、通時的に見た場合も変わらずに維持されていると見ることができる。次に、類型（周辺レベルで不変）を見ると、他の類型と異なり、固有名詞や外来語が多くなっているが、和語と漢語の対比では、他のどの類型よりも漢語の比率が小さくなっていることに気づき、周辺のレベルのまま変わらない語彙には、漢語は多くないと見ることができよう。そして、類型（基本レベル化）では、類型～のいずれよりも漢語が多くなっており、基本的なレベルに進出していく語彙には、漢語がたくさんあったことが分かる。さらに、類型（周辺レベル化）においては、他のどの類型よりも漢語の占める割合が高く7割近くをも占めており、次第に周辺のレベルに追いやられていく語彙には、漢語が非常に多かったことが判明するのである。

以上のように、類型化によって見えてくる語彙の特徴は、和語と漢語に顕著にうかがえ、とりわけ漢語の動きが注目される。次節では、和語と漢語それぞれについて少し詳しく見ていこう。

6. レベルから見た和語の特徴

6.1 品詞構成

図14は、類型ごとに和語における品詞構成をまとめたものである。「近代文語 UniDic」では、精細な品詞情報が出力されるが、ここでは、下に示す6種に統合して集計した。

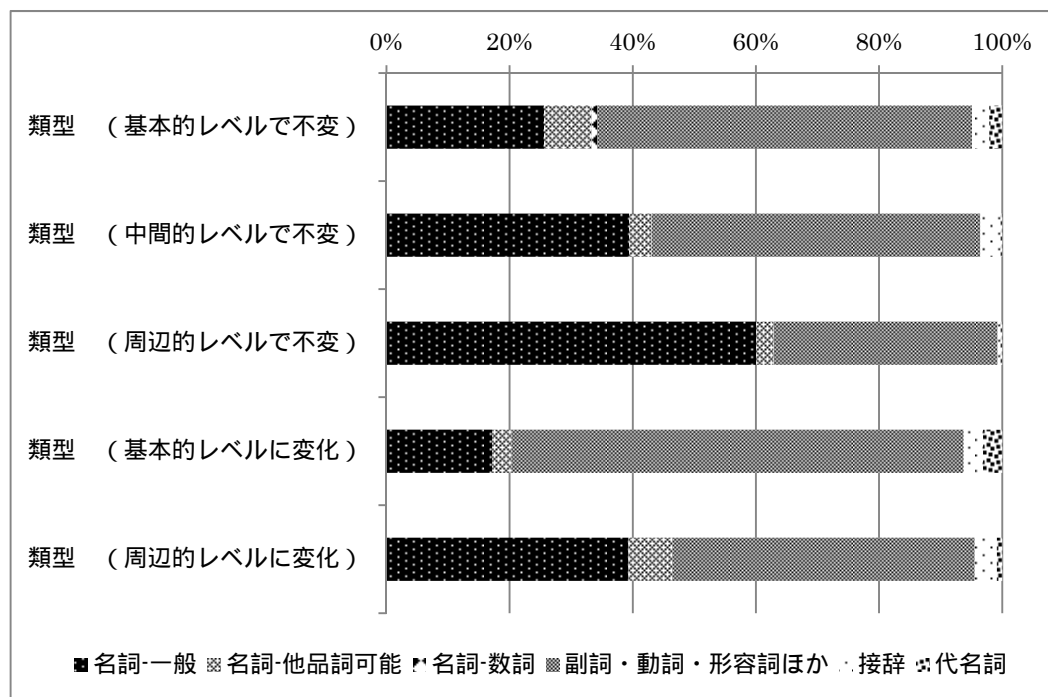


図14 類型別の和語の品詞比率

名詞一般：「名詞-普通名詞一般」と出力されるもの

名詞-他品詞可能：「名詞-普通名詞-サ変可能」「名詞-普通名詞-形状詞可能」「名詞-普通名詞-副詞可能」「名詞-普通名詞-助数詞可能」などと出力されるもの

名詞-数詞：「名詞-数詞」と出力されるもの

副詞・動詞・形容詞ほか：「副詞」「形状詞-タリ」「形状詞-助動詞語幹」「接続詞」「動

詞一般」「動詞-非自立可能」「形状詞一般」「感動詞」「代名詞」などと出力されるもの
接辞：「接頭辞」「接尾辞-名詞的-一般」「接尾辞-名詞的-サ変可能」などと出力されるもの
代名詞：「代名詞」と出力されるもの

図 14 を見ると、類型による品詞構成が大きく異なっていることが分かる。類型 では「副詞・動詞・形容詞・形状詞ほか」が 60%程度という多数を占め、ついで「名詞一般」が 25%程度となっている。これが、類型 、類型 へと進むにしたがって、「副詞・動詞・形容詞ほか」が減少していく一方で、「名詞一般」が増加していき、類型 では、これが逆転している。

6.2 和語における類型別の特徴

各類型にはどのような語彙が入っているのかを具体的に見ていきたい。その際、類似の語彙において類型間の比較を行うのが、各類型の特徴を見えやすくすると考えられるので、意味分野を同じくする一群の語彙を取り上げてみたい。国立国語研究所(2004)『分類語彙表増補改訂版』が示す分類項目のうち、「抽象的關係」の「様相」について、「体の類」「用の類」「相の類」のそれぞれの品詞に分類される語彙(具体的には、「1.13」「2.13」「3.13」の中項目番号の語彙)を事例に取り上げる。この類の語彙のうち、和語の各品詞について、類型別に一覧にすると、下の通りである。『分離語彙表増補改訂版』は多義語について語義ごとに別の意味番号を与えているが、「様相」の分類に複数箇所掲げられている場合は一つのみを掲げた。また、「様相」の分類以外にもその語が掲げられている場合は、より基本的と思われる語義の語と扱った。なお、レベル分けに用いた頻度は、多義語の場合、全語義を合わせたものになっており、「様相」に分類される語義でない例も頻度に含まれてしまっているという問題がある⁴。

名詞など(「体の類」に相当)
類型 (基本的レベルで不変)
様(さま) ありさま、傾き、趣、味、息、姿、身、素(もと)、根、常、差し支え
類型 (中間的レベルで不変)
成り行き、出来(でき)、目付き、仕掛け、仕組み、構え、国柄、傷、一通り、似合い、締まり、妨げ
類型 (周辺のレベルで不変)
死にざま、風向き、気配、あく抜け、上向き、口当たり、手触り、肌触り、めりはり、言葉付き、口前、見栄え、上っ面、見せ掛け、見てくれ、男前、手振り、体付き、足付き、目色、なり、身なり、見目、山並み、ばね仕掛け、成り立ち、骨組み、間取り、店構え、身構え、質(たち)、木口、肌合い、持ち前、得手、並並、持ち合い、兼ね合い、狂い、乱れ、ほつれ、人込み、もつれ、緩み、がら空き、がら明き、差し障り、当たり障り、がた、見事、あしざま
類型 (基本的レベルに変化)
該当なし
類型 (周辺のレベルに変化)
(該当なし)

⁴ この点は大きな問題であるが、コーパスの用例を語義別に分類することは、現在の形態素解析技術では対応できず、そこに人手を関係づけて修正する手法についても確立されていない。今後の重要な研究課題である。

動詞など（用の類に相当）

類型（基本的レベルで不変）

味わう、帯びる、装う、こしらえる、乱れる、散る、緩む、締める、絞める、張る、詰まる、込む、払う、そろう、備わる、備える、尽くす、妨げる、損なう

類型（中間的レベルで不変）

成り立つ、組む、組み立てる、似合う、向く、外す、乱す、かき回す、散らす、締まる、絞まる、詰める、そろえる、行き届く、障る、損ねる、傷付く、傷付ける、汚（よご）す、汚（よご）れる、汚（けが）す、壊す、廢れる、荒れる、荒らす、すさぶ、すさむ、

類型（周辺のレベルで不変）

身構える、そぐう、持ち合う、並外れる、狂わす、踏み外す、振り乱す、散らかる、入り組む、込み合う、立て込む、こんがらかる、こんぐらかる、こじれる、たるむ、引き締まる、締め付ける、出そろう、打ちそろう、取りそろえる、繰り返す、踏み荒らす、荒（あら）らげる、荒立つ、荒立てる、追い詰める、追い込む、切羽詰まる

類型（基本的レベルに変化）

作り上げる、片付ける

類型（周辺のレベルに変化）

（該当なし）

形容詞・形状詞・副詞など（相の類に相当）

類型（基本的レベルで不変）

珍しい、まれ、殊に、おかしい、よい、よろしい、よく、悪い、うまい、甘い、寂しい、濃い、薄い、いろいろ、美しい、清い、難しい、難い、穏やか、危うい、

類型（中間的レベルで不変）

まずい、見事、あっぱれ、目覚ましい、うらやましい、分けて、ゆゆしい、好ましい、程よい、ふさわしい、滑らか、緩い、まばら、細か、様々、華やか、醜い、汚い、かるうじて、平たい、危ない、厳しい

類型（周辺のレベルで不変）

めちゃくちゃ、むちゃくちゃ、よしなに、とりどり、色よい、あつらえ向き、似つかわしい、なまなか、すっぱり、ごたごた、耳障り、まちまち、しどけない、ぎしぎし、がたがた、きつい、ばらばら、粗い、ゆったり、だらだら、きちきち、ややこしい、しげく、すっきり、手短、はでやか、はではでしい、汚らしい、易しい、やすやす、すんなり、まどか、危なっかしい、やばい

類型（基本的レベルに変化）

すばらしい、あたり前、ちゃんと、ゆっくり、いろんな

類型（周辺のレベルに変化）

（該当なし）

上記のリストから、語数の多い類型、類型、類型の和語の特徴として、品詞を問わずに指摘できそうなことは次の通りである。まず、類型は単純語がほとんどを占め、類型は単純語と合成語の両方が同じ程度あり、類型は合成語が多くを占めているという、語構成上のはっきりした差異を見て取ることができる。このことにも関連するが、類型は、意味の抽象度が高く、幅広い意味で用いられる語が多いが、類型、類型へと進むにしたがって、具体的で限定された意味を表す語が多くなる。類型～は、レベルを変えない語彙であるが、ここに述べた特徴は、基本レベルにある語彙、周辺レベルにある語彙、その中間にある語彙の一般的性格と言ってよいものだと考えられる。

一方、類型と類型は、時代によってレベルが一定方向に変わっていく語彙であるが、語数が少ないため、<様相>の語群だけからでは、その特徴ははっきりと分からない。そこで、これらについては、その類型に属するすべての和語を一覧にしてみよう。誤解析が

原因でこの類型に入ってきたことが明らかな語は除外した。なお、『分類語彙表増補改訂版』では「その他」に分類される接続詞や感動詞の類や接辞は、ここでは便宜的に「相の類」に分類した。

まず、以下は、類型 の基本レベル化していく和語のすべてである。

名詞など（「体の類」に相当）

あそこ、あちら、いたずら、苦しみ、小春、差し引き、背中、立場、手合い、飛び、羽目、振り替え、みちのり、無駄、読み物

動詞など（「用の類」に相当）

煽る、言い替える、疑る、打ち込む、生み出す、教わる、片付ける、勝ち得る、腰掛ける、喋る、逸れる、携わる、突き込む、作り上げる、付け加える、取り去る、取り締まる、吐き出す、引っ張る、見付かる、持ち込む

形容詞・形状詞・副詞など（「相の類」に相当）

当たり前、色んな、うん、うんと、黄色い、確り、じっと、直ぐ、すっかり、素晴らしい、そこら、そっと、たらしい、小さな、ちゃんと、ちょい、っこ、はっきり、ふむ、ほっ、まあ、真っ黒、惨め、むやみ、もう、尤も、ゆっくり

上記の一覧で、まず目に付くのは、「相の類」の中の、「うん」「ふむ」「ほっ」「まあ」などの感動詞、「うんと」「すっかり」「じっと」「はっきり」「ゆっくり」など、擬態語に由来する副詞である。これらは、話し言葉によく使われる語である。「黄色い」「素晴らしい」「真っ黒」などの形容詞・形状詞、「煽る」「疑る」「片付ける」「喋る」「逸れる」「吐き出す」「引っ張る」などといった動詞、「あそこ」「あちら」「いたずら」「羽目」「無駄」等の名詞や代名詞も、話し言葉的な語である。基本レベル化していく類型 の和語のリストからは、話し言葉に特徴的な用語が、書き言葉に進出していく流れがあったことが見て取れる。

次に、類型 の周辺レベル化していく和語を一覧にしよう。

名詞など（体の類に相当）

錨、命、飢え、受け渡し、討ち死に、媪（おうな）、大麦、伯母、下ろし、蚕、水夫（かこ）、守（かみ）、鯨、厨（くりや）、褻（け）、薦（こも）、今宵、逆様、棧敷、柴、酢、簾、黄昏、盥（たらい）、費え、一日（ついたち）、鼓（つづみ）、科（とが）、仲立ち、法（のり）、僻（ひが）、広（ひろ）、誉（ほまれ）、麻呂、帝（みかど）、御代（みよ）、姪、設け、催し、諸（もろ）、諸共、諸人、刃（もんめ）、山羊、寡（やもめ）、行く末、忽（ゆるが）せ、葦（よし）、装い、読み売り、我が家、弁え

動詞など（用の類に相当）

勇む、諫める、入り来たる、失せる、おわします、おわす、帰る、肯（がえ）んずる、くずれる、寿ぐ、差しのぼる、授かる、諭す、統べる、謗る、立ち出でる、轟く、煮る、言う、阻む、侍る、払い込む、秀でる、ひしぐ、紐解く、臥せる、屠る、参らす、まします、まる、報う、愛でる、めとる、申し付ける

形容詞・形状詞・副詞など（相の類に相当）

いづくんぞ、最も、憂（う）い、うたた、おのが、思しい、買い、くちおしい、げに、さら、ただただ、力無い、つらい、中々、生臭い、干（ひ）、偏に、一入（ひとしお）、吹き、間々（まま）、見、睦まじい、易（やす）い、ゆかしい、よし無い、

このリストを見ていくと、その多くが、古典にはよく出てくる語でありながら現代の書き言葉としては、古風な語感のする語であることが分かる。名詞では「討ち死に」「媪」「守」「褻」「今宵」「黄昏」「誉れ」など、動詞では「いさめる」「おわします」「おわす」「寿ぐ」「秀でる」「愛でる」など、形容詞・形状詞・副詞では「いづくんぞ」「憂い」「うたた」「く

ちおいしい」「げに」「ゆかしい」「よし無い」などである。これらは、もともとは話し言葉でも使われていたと思われるが、明治期までに話し言葉では使われなくなり、書き言葉として継承されてきて、明治期の文語体書き言葉には受け継がれたが、口語体書き言葉が成立するとそこでは使われなくなっていったと考えられる。書き言葉に残存した「古語」が、口語体書き言葉の成立とともに姿を消していく例である。

しかし、上記のリストの中には、「錨」「大麦」「蚕」「鯨」「酢」「煮る」などがあり、これらは古語ではない。こうした具体物や具体的行為を意味する語は、雑誌で取り上げられる話題が変わってきたことを反映したものだと考えられ、言語の変化によるものではないだろう。

7. レベルから見た漢語の特徴

7.1 品詞構成

図 15 は、類型ごとの漢語について、品詞構成をまとめたものである。

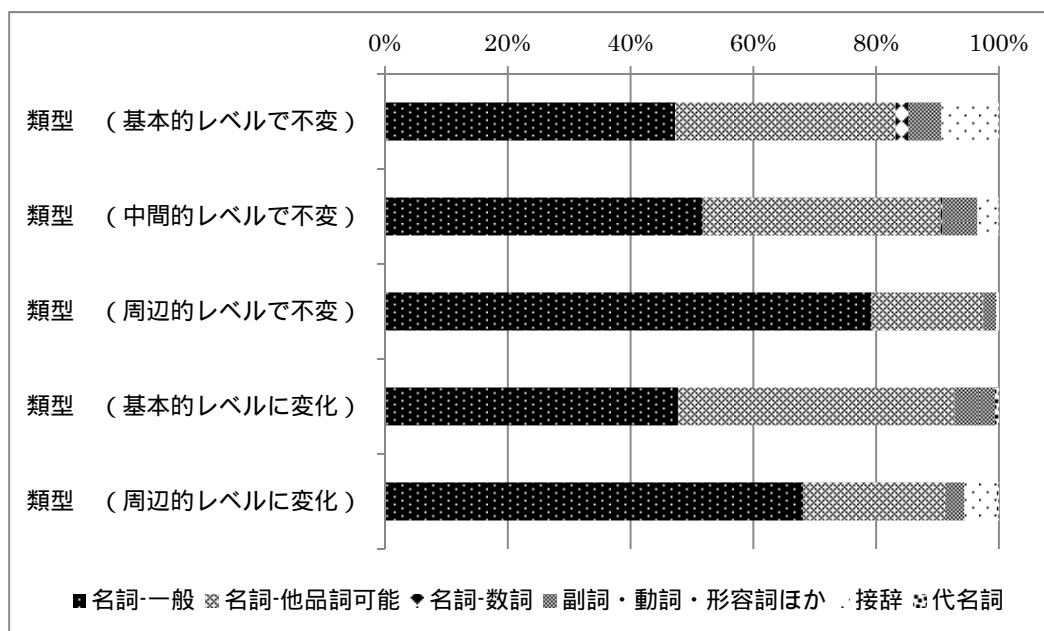


図 15 類型別の漢語の品詞比率

図 15 を見てまず目を引くこととして、類型 (周辺レベルで不変)において名詞一般の比率が極めて高いことがあげられ、類型 (周辺レベル化)においても、これに準じて名詞一般の比率が高い。これとちょうど反対に、類型 (基本レベルで不変)において名詞一般の比率がやや低くなっており、類型 (基本レベル化)がこれに次ぎ、類型 (中間レベルで不変)においても、それと大差ない。これら類型 . . . においては、名詞・サ変等可能が、他の類型に比べて多くなっている。基本的なレベルで安定していたり、基本的なレベルに向けて変化する漢語には、普通名詞以外の品詞に転成するものが多く含まれていたが、周辺レベルで不変であったり、周辺レベルに追いやられる漢語には、普通名詞が多いという傾向を指摘することができよう。

やや細かいところに着目すると、接辞は、類型 (基本レベルで不変)にはまとまった量が見られるが、類型 (基本レベル化)にはほとんど見られないという違いがある、漢語接辞は、明治中期までにすでに基本的な語彙となっていたものが多く、明治後期以後に新たに基本レベル化するものはほとんどなかったということが分かる。一方、副詞・動詞・形容詞ほかは、類型 (基本レベル化)で最も高い比率を示しており、明治後期以後に基本的な語彙となっていくものが多かったと見られるのである。

7.2 漢語における類型別の特徴

6.2 において和語について考察したのと同じ手順で漢語における類型別の特徴を概観したい。やはり「様相」の分類項目に分類される語彙のうち、漢語の各品詞について、類型別に一覧にする。

名詞（体の類に相当）

類型（基本的レベルで不変）

相（そう）、真相、状態、価値、模様、様子、状況、現状、事情、裏面、形勢、大勢、景気、傾向、趨勢、気味、時勢、消息、空気、都合、調子、気風、観、相、体、像、実、要素、構造、性、性質、質、性格、人物、人格、特色、一種、長所、欠点、弱点、難、一般、精、良、不可、秩序、波瀾、美

類型（中間的レベルで不変）

側面、現況、実情、内情、国情、常態、旧態、事態、病状、好況、盛況、風潮、安危、動静、筆致、塩梅、口調、外観、外面、美観、奇観、格好、風采、醜態、品位、素質、本質、特質、気質、天性、本性、性情、異彩、長短、短所、遜色、空前、国粹、当否、怪、支障、事故、艱難、危急

類型（周辺のレベルで不変）

諸相、体様、原状、生態、病態、別状、騰勢、靈気、妖気、鬼気、俳味、力感、調、口跡、新風、和風、欧風、唐風、相好、家相、風水、吉相、仙骨、死相、人体、温容、機構、体制、材質、音質、軟質、硬質、上質、生得、獸性、神性、知性、癩、耐水、慣性、剛性、弾性、塑性、展性、延性、乾性、磁性、悪性、異体、利点、稀覯、精髓、満点、欠格、均整、変哲、波乱、粗密、繁簡、可視、可動、便宜

類型（基本的レベルに変化）

体質、個性、特徴、特長、欠陥、最善

類型（周辺のレベルに変化）

実況、惨状、学風

名詞-サ変可能（用の類に相当）

類型（基本的レベルで不変）

呼吸、加減、構成、組織、相応、調和、一致、整理、始末、障害、故障、破壊、適當、円満

類型（中間的レベルで不変）

塩梅、適合、妥当、画一、拮抗、攪乱、混交、紊乱、紛乱、錯綜、混雑、雑踏、紛糾、緩和、整頓、整備、網羅、斟酌、配合、損傷、荒廢、退廢、難儀、重宝、邪魔

類型（周辺のレベルで不変）

活性、混線、混信、簡約、調律、緊迫

類型（基本的レベルに変化）

緊縮、調節

類型（周辺のレベルに変化）

輻輳

名詞-形状詞可能・形状詞など（相の類に相当）

類型（基本的レベルで不変）

次第、不振、風、子細、普通、非常、特殊、珍、固有、妙、可、奇、複雑、単純、種々、純粹、困難、無理、楽、便利、便宜、不便、安全、無事、完全、危険、急、立派、大抵、尋常、一応、特別、格別、特有、変、重大、良好、善良、適切、健全、大概、一樣、奇麗、密、簡単、容易

類型（中間的レベルで不変）

異常、単調、皮相、壯観、不備、凡庸、非凡、破天荒、奇異、穏当、適度、適任、不当、混沌、稠密、精巧、煩、繁雑、艶、無垢、不潔、不能、平易、簡便、平安、平穩、緊急、平凡、通例、希有、独特、奇妙、奇抜、新奇、乙、珍奇、異様、深刻、粗悪、整然、粗末、粗雑、貧弱、緻密、雑多、簡明、純、流麗、至難、簡易、軽便、平易、険悪

類型（周辺のレベルで不変）

好調、貧相、優性、中性、別段、最悪、適正、豊麗、典麗、凄艶、十全、火急、頓狂、大抵、特別、変挺、最適、冗漫、濃密、繚乱、安直

類型（基本的レベルに変化）

結構、雑然

類型（周辺のレベルに変化）

優、疎、簡、紛々、燦爛、爛漫

上記のリストで語数の多い類型、類型、類型に分類される語を見ても、それぞれ、語構成や意味の面でこれといった特徴は指摘しにくい。和語では、類型間で語構成上の明瞭な差異が見えたのに対して、漢語では、どの類型でも二字漢語が大部分を占めていて⁵、語構成上の差異を指摘することが難しい。また、和語では、意味の抽象度・具体度の点で類型間に差異があったが、漢語については、そのような差異も指摘しにくい。確かに、現代語の語感として、類型には最も基本的な語が多く、類型にはなじみの薄い語が多く、類型はその中間にある語という傾向を感じ取ることはできるが、それは現代語の語感であり、和語の場合のような、歴史を通じた確かな違いとは言えないように思われる。そのような語感に反するものの一例をあげれば、類型の「安危」「艱難」「稠密」などは、現代語では周辺の語という感じがし、一方、類型の「好調」「貧相」「最悪」「適正」「大抵」などは、現代語では基本的な語と感ぜられる。これらの語は、『太陽コーパス』の時代以降にレベルの移行があったと思われるが、『太陽コーパス』の時代では漢語のレベルはまだ安定しておらず、各類型に属する漢語の特徴もはっきりしていない面があるのだと考えられる。

類型、類型について、語数の少ない上記のリストからは、これといった特徴は見出せない。一方、語彙全体を対象としたリストを作成して分析を試みても、和語の場合のように、それぞれの類型に属する語彙の特徴を明確に指摘することは難しい。この時期に基本レベル化したり周辺レベル化したりする理由は、個々の漢語の語形を見ているだけでは不明である。基本レベル化や周辺レベル化する理由を探るには、漢語の使われ方を用例に即して分析していく必要がある。これについては、コーパスにおける用例分析を中心とする別の研究への展開が望まれるところである⁶。

8. おわりに

以上、本論文では、『太陽コーパス』に形態素解析を施し、その語彙全体を頻度調査の対象として、頻度に基づくレベル分けを実施し、このレベル情報を利用した語彙史研究を試みた。これは、近代語のテキストにも形態素解析が実現しつつある近年の研究状況を踏まえた新しい段階の研究例を示そうとしたものである。もっとも、開発途上の技術を利用しているためデータの精度には問題があることには留意しなければならない。

各語種が語彙全体の中でどのような位置を占め、それが時代によって変化する全体的な傾向を示すことができたところは、近代語彙史の研究の新しい展開だと言えるだろう。形

⁵ 本稿では UniDic の言語単位である「短単位」によっているため、3 字漢語以上は分割されることになる。「長単位」によって調査を行えば、類型による語構成の差異が見出される可能性はある。

⁶ 田中（2011、2012）などは、その端緒としてまとめた研究である。

態素解析を施したコーパスを作って研究できる環境を整えていくことは、近代語の語彙史研究にとって、大きな意義を有すると考えられる。

文 献

- 小木曾智信 (2009) 『近代文語文を対象とした形態素解析のための電子化辞書の作成とその活用』(科学研究費補助金成果報告書)
- 国立国語研究所 (1959) 『明治初期新聞の用語』(国立国語研究所報告 5)
- 国立国語研究所 (1964) 『現代雑誌九十種の用語用字 分析』(国立国語研究所報告 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 (2004) 『分類語彙表 増補改訂版』(国立国語研究所資料集 14、大日本図書)
- 国立国語研究所 (2005a) 『現代雑誌の語彙調査 1994 年発行 70 誌』(国立国語研究所報告 121)
- 国立国語研究所 (2005b) 『太陽コーパス 雑誌『太陽』日本語データベース』(CD-ROM、国立国語研究所資料集 14、博文館新社)
- 国立国語研究所 (2005c) 『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』(国立国語研究所報告 122、博文館新社)
- 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 (2005c)、pp.1-48)
- 田中牧郎 (2011) 「近代漢語の定着 『太陽コーパス』に見る」(『文学』12-3、岩波書店、pp. 136-153)
- 田中牧郎 (2012) 「新漢語定着の語彙的基盤 『太陽コーパス』の「実現」「表現」「出現」と「あらかず」「あられる」など」(『日本語の学習と研究』160、北京、pp.39-47)
- 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007) 「コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用」(『日本語科学』22、pp.101-122)
- 橋本和佳 (2010) 『現代日本語における外来語の量的推移に関する研究』(ひつじ書房)
- 飛田良文 (1966) 「明治以後の語彙の変遷」(『言語生活』182、筑摩書房、pp.16-24)
- 宮島達夫 (1967) 「近代語彙の形成」(『ことばの研究第3集』、秀英出版、pp.1-50)

付 記

本稿は、国立国語研究所国際学術集会「漢字漢語研究の新次元」(2010年7月30日、国立国語研究所)および、第104回漢字漢語研究会(2012年8月1日、早稲田大学)で発表した内容をもとに、大幅に書き改めたものである。特に国際学術集会のときには、『明六雑誌』『国民之友』のデータについても扱ったが、これらは『太陽』とは質の異なる面もあるため、今回は『太陽』のみを対象にした。また、本論文における頻度調査やレベルの設定は、今回新たに処理を行った形態素解析データに基づいている。