

# 国立国語研究所学術情報リポジトリ

## A Corpus-Based Study of Sino-Japanese Words in Modern Magazines: A Comparison with BCCWJ

メタデータ	言語: jpn 出版者: 公開日: 2017-07-11 キーワード (Ja): キーワード (En): 作成者: 間淵, 洋子, MABUCHI, Yoko メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00001376">https://doi.org/10.15084/00001376</a>

## 近代雑誌コーパスにおける漢語語彙の特徴

—BCCWJ との比較から—

間淵洋子

明治大学 大学院生 / 日本学術振興会 特別研究員 (DC) / 国立国語研究所 共同研究員

### 要旨

近代語と現代語の形態論情報付きコーパス『日本語歴史コーパス 明治大正編I雑誌』と『現代日本語書き言葉均衡コーパス』を用いて、近代と現代との漢語語彙の比較を試みた。コーパスから網羅的に漢語を抽出・調査した結果、近代と現代との漢語語彙の差異および変化について以下の実態が明らかになった。

① 近代・現代を通じて、異なりで約 50,000 語の漢語が用いられており、うち、近代・現代で共通して使用の見られた漢語はほぼ半数の約 22,400 語、近代のみで使用の見られた漢語が約 19,400 語 (「発兌」「状師」「英人」「征清」「邦国」など)、現代のみで使用の見られた漢語が約 7,700 語 (「支援」「介護」「強化」「本格」「展示」など) で、近代では現代に比して豊富な漢語語彙を有していた。

② 使用頻度の有意差検定結果では、近代・現代で共通して用いられていた漢語のうち、約 42% の語 (「問題」「世界」「以上」「研究」「経済」など) では、両時代を通して有意な差が見られなかったが、約 31% の語 (「今日」「吾人」「憲政」「列国」「世人」など) は現代での使用が近代より減少、約 27% の語 (「時間」「意味」「可能」「企業」「利用」など) は増加しており、時代により使用が減衰・増大する漢語が見られた。これらの変化・不変化には、近代における漢語の使用頻度に基づく語彙レベルとの相関が見られた。

③ 近現代における漢語語彙の相違・変遷 (語の消失・減少や増加・出現) は、時代背景・社会情勢の影響が大きいもの、周辺語彙との勢力変動等によるもの、文体などの表現類型の変遷に関わるもの、言説・トピックの時代差によるものなど、様々な要因が見られた\*。

**キーワード:** 漢語, 近代語, 形態論情報付きコーパス, 頻度調査, 語彙の史的変化

### 1. はじめに

日本史における近代は、政治、国際環境、科学、哲学、文化、全てにおいて最も激しく揺れ動き、劇的な変化がもたらされた重大な転換期であった (坂野 2012)。日本語に関してもまた、近代は、それまでの漢文・文語中心のスタイルから今日の書き言葉に繋がる口語文へと移り変わる最大の変革期であった。この変革期の日本語 (= 近代語) は、模索的であり、流動的であり、現代語の確立に至るまでに、多くの「ゆれ」や「変化」が見られたことが、これまで多く指摘されてきた (今野 2012)。

\* 本稿は、日本学術振興会特別研究員奨励費 16J08872 「コーパスを利用した近現代漢語の表記・語法の多様性に関する計量的・通時的的研究」 (代表: 間淵洋子)、および、国立国語研究所機関拠点型基幹研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」 (プロジェクトリーダー: 小木曾智信) による成果の一部である。

なお、本稿は、機関拠点型基幹研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」近世・近代グループ第 2 回研究会 (2016 年 9 月 13 日, 明治大学) で口頭発表した内容に加筆・修正したものである。席上で多くの貴重な御意見・御助言を賜った。また、プロジェクトリーダーの小木曾智信准教授には、本稿を完成させるまでに、多くの有益な御指摘・御助言と御鞭撻をいただいた。ここに記してお礼申し上げる。

特に、近代化に伴う新しい物質・事象・概念の大量の流入により、飛躍的に語彙が増えた「漢語」<sup>1</sup>については、個々の語の出自や定着、変化、淘汰など、様々な事象について、これまで多くの研究がなされてきたが、その全容が明らかにされたとは言い難い。それは、漢語を含めた当時の語彙の実態、そして変化に目を向ける際に対照となる現代の語彙の実態というものの把握が極めて困難であるために他ならない。

しかし、近年、大規模な日本語のデータベースが整備されつつある。ことに、語彙を対象とした研究には欠かせない単語情報の付与されたコーパスにより、これまで一部の・局所的にしか明らかにされてこなかった近代、そして現代の語彙の全体像を実証的に捉えることが可能となっている。そこで、本研究では、日本語の漢語語彙の変化について、コーパスの計量的な調査に基づき、大局的な観点からの分析を試みたい。近代の漢語が、どのような特徴を持ち、どのような背景で変化していくのか、現代語との比較から明らかにすることを、研究の目的とする。

## 2. 先行研究

本研究は、現代語との比較において浮き彫りになる近代漢語語彙の特徴を明らかにすることを目的とするが、これまで現代語との比較という観点で近代漢語を論じたものの多くは、表記や語法の変化に重点が置かれていた。表記については、同音語の表記バリエーションを対象として特定の資料を元に併用から統合への過程を通時的に示した研究（武部 1981 など）や、明治・大正期において同語異表記が多く見られる状況の指摘とどのような語に異表記が見られるかについての調査（今野 2012 など）、語法については、近代から現代にかけて見られる品詞・用法の変化に関する指摘・報告（池上 1953, 鈴木 1998, 永澤 2010, 鳴海 2015 など）が数多くなされてきた。これらは、漢語の変化の実態や傾向を明らかにした点で極めて重要であるが、当時の漢語全体の中でそのような変化の見られる語が、どのような位置付けであり、漢語全体はどのように変化したか、あるいはしなかったかについて明確に答えを示すものではない。

一方、個々の語についてではなく、日本語の語彙の一部として漢語を捉え、日本語の中での漢語の位置付けや実態を明らかにしたものはそれほど多くないが、これらの研究では「語種」という観点から、漢語使用比率の多寡や変化、和語・外来語との関連が論じられてきた。精緻な語彙調査に基づき日本語の語彙の変化を示した国立国語研究所（1964, 1987, 2005a）では、大正～昭和～平成と時代が下るにつれ漢語が減少していることが指摘されているが、その一方で、基本語彙に占める漢語の比率が次第に増加しているという報告もあり（飛田 1966, 宮島 1967 など）、必ずしも日本語における漢語語彙の全体像が明らかになっているとは言い難い。

しかし、近年のコーパス構築とそれをういたコーパス言語学の興隆を受け、大規模なデータから帰納的にその実態を明らかにするアプローチが増えている。近代語・現代語の大規模なコーパスに基づき、日本語の語彙全体を見渡し、語種の比率や個々の語の使用実態の変遷を詳細に追った田中（2009, 2011, 2016）では、語の使用頻度の累積度数による「カバー率」（対象データの

<sup>1</sup> 本稿では、「漢語」を、成立や出自を問わず、日本漢字音で読まれる漢字表記語全ての意で用いる。

延べ語数に対して累積頻度が占める比率) という指標を用い、個々の語を「語彙レベル」により層別し、基本的な語～周辺の語のように、語彙の広がりの中に個々の語を位置付けた。さらに、このレベルの経年変化を追うことで、語の定着や交代の過程を明らかにした点で、語彙史的な成果をも得ている。そこで、本研究では、田中(2011)にならい、近代語と現代語の漢語語彙の使用実態、特に語の出現頻度を基準とした層化を通して、その共通点と相違点を明らかにし、比較により見えてくる近代漢語の特徴と、使用実態の相違の背景について検討することとする。

### 3. 研究方法

#### 3.1 調査対象コーパス

近現代における漢語の実態を把握するために、現在公開されている近代語と現代語のコーパスを用いた調査を行う<sup>2</sup>。対象とするコーパスの概要を表1に示す。

表1 調査対象コーパスの語彙量

時代	資料	出版年	語数(万)	
近代	明六雑誌	1874-5	18	
	国民之友	1887-8	101	
	太陽		1895	202
			1901	197
			1909	187
			1917	180
			1925	203
	女学雑誌	1894-5	59	
	女学世界	1909	52	
	婦人倶楽部	1925	54	
全体	1874-1925	1,253		
現代	BCCWJ (出版 SC)	2001, 2005	1,234	

現代語のコーパスとしては、2011年公開の『現代日本語書き言葉均衡コーパス(以下、「BCCWJ」と表記)』の出版サブコーパス(以下、サブコーパスを「SC」と表記)のうち2001年、2005年の発行分の可変長サンプルを用いる(記号等を除く収録語数約1,234万語、助詞・助動詞を除く自立語数約751万語)。BCCWJは、日本初となる、また唯一の、大規模バランストコーパスであるが、その中でも出版SCは、2001-2005年に出版された書籍・雑誌・新聞について、極めて精密な文字量調査に基づき推定された母集団に対して、厳密・精緻なランダムサンプリングにより、その縮図となるよう設計されたコーパスであり(前川2008)、日本語の一般的な書き言葉の

<sup>2</sup> いずれのコーパスにおいても、データには、全体に渡って人手での確認・修正作業を経て高精度の形態論情報が確保されている「コア」と呼ばれるものと、一部にのみ人手修正が加えられ、解析の誤りが含まれる「非コア」と呼ばれるものが含まれる。「非コア」のうち誤解析の生じている語については、必要に応じて、分析対象から除外する等の処理を施す。

サンプルとしてまさに好適なデータである。BCCWJには、出版SCのほかに、現代における言葉の流通実態を捉えるのに適した図書館SC(書籍のみ)、個別の研究目的に沿うデータを集めた特定目的SCがあるが、本研究では、比較する近代のコーパスが後述する通り雑誌のみであるため、逐次性の観点から共通性の高い資料として、雑誌や新聞を含む出版SCを対象とする。出版SCより2001年分と2005年分の2カ年のみを用いたのは、以下に述べる近代語コーパスと語数を概ね同様になるように調整するためである。

近代語のコーパスとしては、2016年10月に全体が公開された『日本語歴史コーパス 明治大正時代編I雑誌(以下、「CHJ 明治大正雑誌」または「CHJ」と略記)』(収録語数約1,253万語、自立語数約754万語)を用いる。CHJ 明治大正雑誌は、国立国語研究所が既に公開していた複数の雑誌コーパスを統合・再整備し、新たにWeb検索アプリケーション「中納言」により提供されたもので、明治・大正期の書き言葉の実態および変遷を見渡すことができるよう、各時代を代表する雑誌から一定の年数ごとの刊行分の全文を収録したコーパスである。BCCWJと異なりバランストコーパスではなく、またサンプルコーパスではないが、これほど大規模な形態論情報付きの近代語コーパスが構築されたのは初めてのことであり、収録資料はいずれも当時の日本語の書き言葉を知る上で欠くことのできない代表的な重要資料であり、研究価値の高いデータである。また、今回現代語の資料として用いるBCCWJ出版SCは出版間隔が5年と、ほぼピンポイントで時代のある1点を切り取ったものになっているのに対して、CHJ 明治大正雑誌は、約50年という長いスパンで劇的に変動する近代語の様々な様相を含んだコーパスになっている点でも、比較対象としての等質性に欠ける部分がある。しかし、本研究は、混沌とした近代の漢語使用状況そのものを主なターゲットとして、その期間の変化をも視野に入れて、現代のある時点からそれを見た時にどのような差異があるかという点を明らかにすることを目的とするものであるため、CHJ 明治大正雑誌の全体をひとまとまりとして用いることとした。

なお、これらのデータを調査対象資料とした最も大きな理由は、資料の重要性や代表性にもまして、形態素解析の結果付与される単語に関する情報(形態論情報)が得られることにある。この形態論情報によって、調査対象語を既存の語彙表や辞書等から選定するのではなく、データそのものからほぼ漏れなく抽出し実態を把握することが可能になるからである。上記二種のコーパスは、共通の枠組みによる形態素解析に基づき「短単位」と呼ばれる斉一な単位により単語の区切り目が認定されるほか、語種情報、品詞情報が付与されており、特に漢語については、漢字2文字を1単位とする単純明快なルールを用いているため、本研究で対象とする「二字漢語」(3.2節)を抽出するには好適のデータである。

図1に、「詩歌」という漢語を例に短単位情報の階層構造を示す。

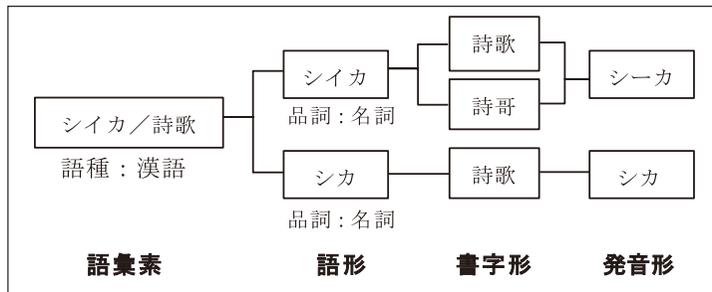


図1 短単位情報の階層構造

「語彙素」とは、辞書見出しに相当するものであり、語の様々なバリエーションを統合的に位置付けるために設定されている最上位の階層である。標準的な読みと標準的な表記によって規定される。語彙素は、その下位に「語形」という階層を持つ。ここには、読み方・発音のバリエーションが含まれている。図1に挙げた例「詩歌」については、一般的な読み「シイカ」の他に、「シカ」という読みが含まれる。形態は異なるが、同語と判別されたものが兄弟関係として位置付けられる。それぞれの語形は「書字形」という下位階層を持つ。これは、実際に現れる表記形であり、文字種・字体・用字などのバリエーションが兄弟関係として位置付けられる。つまり、この「語彙素」の情報を用いることで、語の様々なバリエーションを一括して検索することが可能であり、バリエーションを考慮することなく網羅的に語彙を調査することが可能となっている。

### 3.2 調査対象語

近代と現代との漢語語彙の差異とその特徴を明らかにするために、本研究では、調査対象コーパスに現れる二字漢語のうち固有名詞と数詞を除いたものを調査対象とする。

ここで、漢語の中から二字漢語に絞って調査対象とするのは、主に以下4点の理由による。

1. 二字漢語が、日本語の漢語語彙の種類や頻度において中心的な位置を占めていること。
2. 三字以上の漢語のほとんどは、二字漢語が語基になって成り立っていること。
3. 一字漢語は、和語の複合要素（動詞や助詞・助動詞）との結合が固く、他の二字以上の漢語と文法的振る舞いが大きく異なること。
4. 上記2, 3を反映し、調査対象コーパスの単語認定において、一字漢語や三字以上の漢語は、二字漢語と大きく異なること（一字漢語は動詞「する」と結合した「愛する」「関する」等を1語の混種語とみなす。三字以上の漢語の多くは二字をひとまとまりとして「日本／語」のように分割される）。

よって、一字漢語や三字以上の漢語は、本研究での二字漢語の実態を確認した後に別途検討することとし、今回の分析対象からは外し、今後の課題としたい。

### 3.3 調査対象語の抽出

調査対象語を抽出するため、コーパス検索アプリケーション「中納言」を用い、以下の条件に合致する用例を抽出した。

- ・短単位検索「キー」条件：語種が「漢」（漢語）であり、「語彙素」の表記が漢字2字からなるもの（検索文字列「[一 - 龠 ■ ][一 - 龠々 ■ ]」）
- ・検索対象：BCCWJ（非 NumTrans 版）は出版・書籍，出版・雑誌，出版・新聞のコア・非コア 2001 年・2005 年（検索結果から可変長データのみ抽出），CHJ は明治大正のコア・非コア

検索結果は、語彙素の表記を用いて集約した。BCCWJ, CHJ 両コーパスでは、同一表記であっても読み（語形）が異なるものを、意味の差異等を考慮し別の語彙素としている場合がある（例：異形—イギョウ/イケイ，実体—ジッター/ジッテイ，心中—シンジュウ/シンチュウ，上品—ジョウヒン/ジョウボン，など）<sup>3</sup>。また、同一表記・同一読みであっても異品詞のものを別の語彙素とする場合がある（例：段々—名詞/副詞，是非—名詞/副詞，相当—名詞/形状詞/副詞，など）。しかし、今回、近代と現代と時代を跨いで漢語の比較を行うにあたっては、漢語の読みや品詞が時代により流動的で異なることがあることを考慮し、これらの読みや品詞の差を、語の意味や用法の変化に付随する同語のバリエーションとみなして扱うこととした。同語とみなしたのものの中には、意味や用法の差が明確で同語として扱うことが不適切なものも含まれる可能性があるが、これらの検討や見直しについては、今後の課題としたい。

また、抽出された語の中から、数詞は分析対象から排除した。その結果、CHJ 明治大正雑誌と BCCWJ 出版 SC (2001 年, 2005 年出版分) のいずれかに出現した漢語の異なり語数は約 5 万語、延べ語数は約 365 万語となった。

### 3.4 分析の指標

抽出した各語について、語彙の特徴を検討するために、以下の指標を用いることとし、それぞれラベルを与えた。

1. 使用の有無…BCCWJ 出版 SC と CHJ 明治大正雑誌の両コーパスにおいて使用例が見られるか否かにより、a. **近代専用**（CHJ のみに出現し BCCWJ で用例が得られない語）、b. **近現代両用**（CHJ, BCCWJ のいずれにも 1 例以上の用例が見られる語）、c. **現代専用**（BCCWJ のみに出現し CHJ で用例が得られない語）、に 3 分類した。
2. 語の変化パターン…近代と現代の両時代で共に用いられる語については、出現頻度に統計的な有意差が見られるかどうかの検定（統計分析ソフト R によるフィッシャーの直接確率検定  $p < 0.05$  による）を行い、近代で有意に高頻度の語は b. 〈減少〉、有意差が認められない

<sup>3</sup> 意味の異なりが認められない以下のような語形のバリエーションは、原則として同一の語彙素となるため、ここで行う集約処理は不要である（小椋他 2011）。

- ・清音・濁音・半濁音の差異，連声，語末長音の短呼形，語末以外の位置の長音の有無（例：連中…レンチュウ/レンジュウ，安穩…アンノン/アンオン，女房…ニョウボウ/ニョウボ，詩歌…シカ/シイカ）
- ・漢字音の呉音・漢音・慣用音等の差異（例：重複…チョウフク/ジュウフク）

語を c. 〈不変〉, 現代で有意に高頻度の語を d. 〈増加〉とし, 加えて, 指標 1 で見た近代専用の語を a. 〈消失〉, 現代専用の語を e. 〈出現〉, として 5 分類した。

3. 使用頻度…田中 (2011) の「語彙レベル」の枠組みを参考に, 語のカバー率 (累積使用率) により以下の A ~ G に 7 分類した (表 2, 3) <sup>4</sup>。A を 60%, B を 80%, C を 90%, D を 95%, E を 99%, F を 100%, G は頻度 0 のものを基準として各レベルの頻度範囲を決定した。上位のレベルほど, コーパスを構成する語彙の中で中心的な語であり, 使用の深さを表す指標となる。頻度範囲は, 各コーパスに出現する語 (助詞・助動詞, 記号類を除く。以下同様) を出現頻度の降順にリストし, 出現頻度に対する累積頻度のカバー率が, A ~ E の各レベルの基準値 (上述) を超えた頻度を範囲の下限として決定した。当該頻度の語が複数ある場合は, その頻度の語を全て取りきるまでをカバー率の範囲下限とした。なお, 各コーパスにおける区分の妥当性を検討するため, 100 万語あたりの調整頻度 (PMW) を合わせて掲載する。調整頻度範囲を見ると, それぞれのコーパスにおけるレベルは, 概ね同様の出現頻度の語が所属するようになっていいることが分かる。

表 2 CHJ 明治大正雑誌 語彙レベル

レベル	頻度範囲	調整頻度範囲	カバー率	語数
A	1,125-358,407	150.51-47,951.51	60%	180
B	214-1,124	28.63-150.11	80%	1,223
C	60-213	8.03-28.50	90%	3,125
D	23-59	3.08-7.89	95%	4,845
E	4-22	0.54-2.94	99%	14,921
F	1-3	0.13-0.40	100%	17,521
G	0	0	100%	12,659

表 3 BCCWJ 出版 SC 語彙レベル

レベル	頻度範囲	調整頻度範囲	カバー率	語数
A	999-330,800	131.64-43,588.71	60%	375
B	204-998	26.88-131.50	80%	1,597
C	59-203	7.77-26.75	90%	2,975
D	22-58	2.90-7.64	95%	3,985
E	4-21	0.53-2.77	99%	10,153
F	1-3	0.13-0.40	100%	11,024
G	0	0	100%	19,411

<sup>4</sup> 田中 (2011) では, 語彙レベルを a (~ 78%), b (~ 88%), c (~ 94%), d (~ 97%), e (~ 100%) の 5 分類としているが, これは, 主に基本語彙等の抽出を目的として, 阪本語彙 (阪本 1984) を参考に各レベルの所属語数を定めるといった方法で分類した結果導き出された累積使用率の目安であると思われる。本研究は基本語彙抽出に主眼はなく, 語の出現頻度を元に語のレベル分けを試みるものであり, 手法として田中 (2011) のカバー率の枠組みを用いる。際立って頻度の高い語などをカテゴリーとして取り出したいため, レベル A の累積使用率を 60% レベルとして 1 段階増やし, また, 頻度 0 に一つのカテゴリーを当てるといふようにより細かいレベル設定を行った。

4. 使用記事数…単純な出現頻度(粗頻度)では、同一記事内で多用される語の影響が大きいため、参考として使用記事数を計測した。記事数の計測では、同一記事内で同一の語が何度出現しても、頻度は「1」とする。出現する記事数がコーパスに含まれる全ての記事に占める割合(記事率)により、6分類した。Aは10%以上、Bは5%以上、Cは1%以上、Dは0.1%以上、Eは0.03%以上、Fは1記事のものである。上位のレベルほど幅広い記事で用いられているということであり、語の使用の広さを表す指標となる。

表4 記事率レベル

レベル	記事率	近代		現代	
		記事数範囲	所属語数	記事数範囲	所属語数
A	10%～	574-2,682	157	523-2,634	264
B	5%～	287-573	310	262-522	465
C	1%～	58-286	2,672	53-261	2,588
D	0.1%～	6-57	14,357	5-52	11,057
E	0.03%～	2-5	12,781	2-4	7,338
F	0.01%～	1	11,538	1	8,397

#### 4. 調査結果

##### 4.1 使用の有無による近代・現代の漢語語彙概観

3.4節に示したそれぞれの指標に基づき分布を示す。

表5 使用の有無による漢語の層別

カテゴリー	全体		近代			現代		
	異なり語数	延べ語数	異なり語数	延べ語数	延/異	異なり語数	延べ語数	延/異
近現代両用	22,404	3,448,229	22,404	1,517,121	67.72	22,404	1,931,108	86.19
近代専用	19,411	113,988	<b>19,411</b>	<b>113,988</b>	<b>5.87</b>	0	0	0
現代専用	7,705	83,204	0	0	0	<b>7,705</b>	<b>83,204</b>	<b>10.80</b>
全体	49,520	3,645,421	41,815	1,631,109	39.01	30,109	2,014,312	66.90

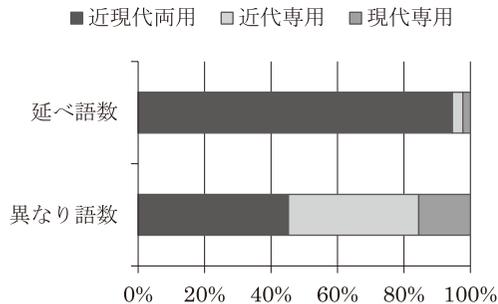


図2 使用の有無による漢語の類別

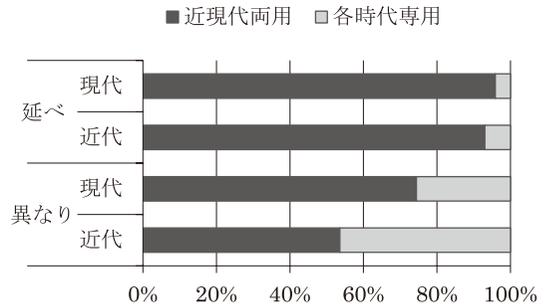


図3 共通語彙の割合

表5は、使用の有無によって漢語を層別した結果、図2は異なり語数、延べ語数別に、各カテゴリーの割合を示したグラフ、図3は異なり語数、延べ語数別に、各時代における近現代両用の語と各時代専用の語の割合を示したグラフである。これらの結果から、異なり語数では、近代で用いられる漢語の約45%が、現代でも共通して用いられている漢語（近現代両用＝共通漢語）であることが分かる。残りは近代にのみ出現のある「近代専用」が約40%、現代にのみ出現のある「現代専用」が約15%で、その割合は近代専用が多い。時代別に漢語語彙における共通漢語の割合を見ると、近代では約半数、現代では約3/4で、現代では大多数が近代から用いられている漢語が占めているものの、近代ではその時代特有の漢語が半数近くあることが分かる。ただし、延べ語数で見ると、両時代共に、実際に使用されている漢語の95%程度は共通漢語であり、使用実態はほぼ共通語彙で占められていることが分かる。

さらに、表5に見る通り、延べ語数は近代で約160万に対して現代で約200万と圧倒的に多いのに対して、異なり語数では近代が約4万と、現代の約3万を大きく上回っている点は注目すべき点である。延べ語数／異なり語数の値（一つの漢語を用いる平均使用数）でも、近現代両用の語、時代専用の語のいずれにおいても、現代で明らかに高い。つまり、近代においては、現代よりも数多くの漢語を幅広く用いており1語の使用頻度は少ない、一方の現代は、限られた漢語を用いているため1語の使用頻度が高くなっている、という結果が読み取れる。漢語においては、近代では、現代より豊富な語彙を有していたことが分かる。

#### 4.2 語彙レベルによる漢語分布とその変化

次に、各時代の語彙レベルと、変化パターンの集計結果を確認する。表6に、近代（表中CHJ）と現代（表中BCCWJ）の語彙レベルの分割表を示す。

表6 語彙レベルによる漢語の層別 (単位：語数)

		BCCWJ							合計
		A	B	C	D	E	F	G	
C	A	96	74	5	3	1	1		180
	B	154	494	316	141	94	15	9	1,223
	C	75	464	813	708	699	231	135	3,125
	D	31	229	647	879	1,518	790	751	4,845
H	E	14	222	704	1,182	3,413	3,004	6,382	14,921
J	F	3	80	264	560	1,959	2,521	12,134	17,521
	G	2	34	226	512	2,469	4,462		7,705
	合計	375	1,597	2,975	3,985	10,153	11,024	19,411	49,520

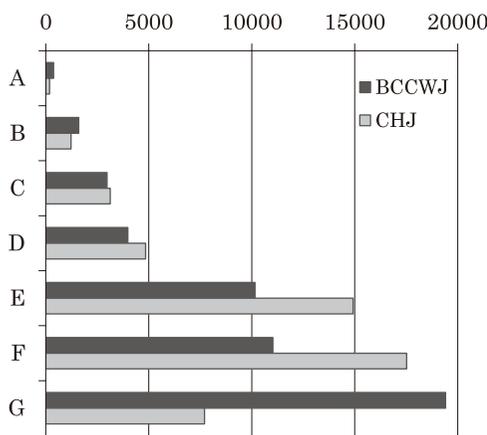


図4 時代別語彙レベル所属漢語数

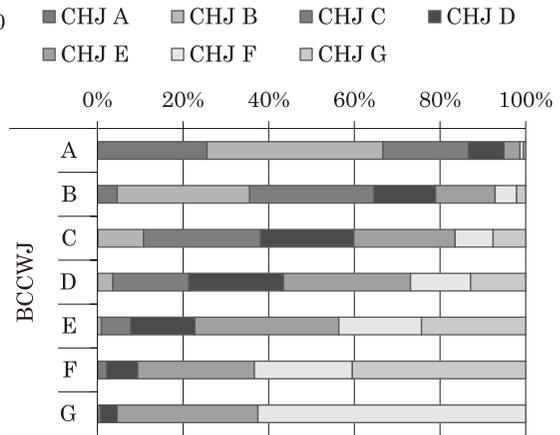


図5 語彙レベルの相関

図4は時代別に各語彙レベルに所属する漢語数を示したグラフである。両時代に共通して中頻度～低頻度の語の多いことが目立ち、基本的・中心語的な語は多くない。時代による差異に目を向けると、現代では高頻度語（A, B）が近代より多い傾向にあり、近代では中・低頻度の語（D, E, F）が現代に比べて多い傾向にある。現代における漢語使用の実態が、限られた漢語を高頻度で用いるものであるのに対して、近代は中間層がより厚く、豊富な漢語を幅広く用いるものであることが分かり、有する語の数の面でも、使用の面でも充実したものであることが分かる。次に、図5により現代における語彙レベルと、近代における語彙レベルとの相関を見ると、現代において高頻度の語は、近代においても高頻度で使われていたものが多く（現代のレベルAでは6割以上が近代でもレベルA・Bの高頻度語）、同様に現代において頻度が低い、もしくは使用が見られない語は、近代でも低頻度であるといった相関が見られる。この結果からは、語の使用頻度が語の定着に大きく関わっていることが見て取れる。

### 4.3 変化パターンによる漢語分布

次に、変化パターンにより漢語を分類し、その分布実態を見てみたい。

表7 変化パターン

変化パターン	語数
〈消失〉	19,411
〈減少〉	6,746
〈不変〉	9,480
〈増加〉	6,178
〈出現〉	7,705

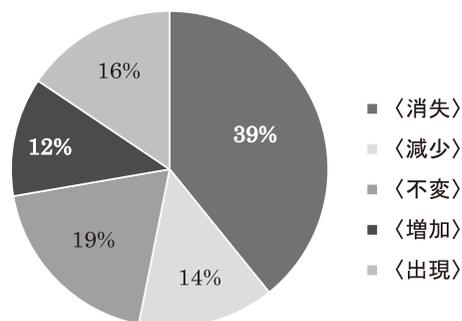


図6 変化パターンの比率

表7, 図6は, 変化パターンによる漢語の分類結果を示したものである。比率を見ると, 〈消失〉(近代から現代にかけて消失あるいは衰退した漢語)が最も多く40%程度ある。次いで, 使用頻度に有意差のない〈不変〉が約20%, 〈出現〉(近代から現代にかけて出現あるいは興隆した漢語)が16%程度, 〈減少〉(近代から現代にかけて使用頻度が有意に減少している漢語)が14%と続く。〈消失〉, 〈減少〉を合わせて, 近代での使用から勢力を減じた漢語が半数以上を占めるが, 一方では勢力を伸ばした漢語(〈増加〉, 〈出現〉)も多く見られることが分かる。近現代共用の語(近現代の共通語彙)約22,400語に着目すると, 使用頻度に有意差のない〈不変〉が約9,500語(42%)とやや多く, 使用頻度に差のある〈減少〉〈増加〉がそれぞれ約6,700語(31%), 約6,200語(27%)と同程度あることも分かる。

先に4.2節において各時代の語彙レベルが相関することを見た。同様に, この変化パターンには, 近代における出現状況(語彙レベル)が関与している可能性が高い。そこで, 漢語の変化パターンと近代の語彙レベルとをクロス集計した結果を以下に示す。

図7より, 近代において低頻度の語(E, F)ほど〈消失〉が多く, 高頻度の語ほど〈消失〉が少ない傾向が目立ち, 頻度が高い語ほど〈増加〉の割合が多いという点で頻度と変化パターンに相関はある。ただし, どのレベルにおいても〈消失〉または〈減少〉といった勢力を減ずる変化の見られる語が6割以上ある点や, 最も高頻度のレベルAより下のレベルBでより〈増加〉の割合が高く〈減少〉の割合が低い点は興味深い。頻度が高ければ高いほど語の勢力を伸ばすとは言えず, レベルAのように既に十分に基本語・中心語として機能していた漢語でありながら, 使用頻度の減少する語が7割以上存在する背景の検討が必要である。

表8 変化パターンと近代の語彙レベル  
(数字は語数)

	〈消失〉	〈減少〉	〈不変〉	〈増加〉
A		123	14	43
B	9	730	115	369
C	135	1,830	409	751
D	751	2,264	883	9547
E	6,382	1,799	4,514	2,226
F	12,134		3,545	1,842

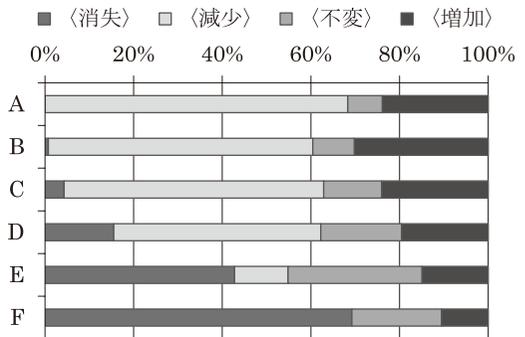


図7 変化パターンと近代の語彙レベルの相関

## 5. 考察—使用分布変化の背景

### 5.1 所属語彙一覧

次に, 前節までで見た漢語語彙の分布とその変化の背景を探るため, 各分類指標に基づく各カテゴリーに, 実際にどのような漢語が含まれるのかを確認してみよう。以下の表9~13では, 変化パターンごとに, 横に現代語における語彙レベル, 縦に近代語における語彙レベルを配した

分割表に、所属語彙（〈消失〉〈減少〉〈不変〉は近代語での出現状況<sup>5</sup> 上位 10 位までの語、〈増加〉〈出現〉は現代語での出現状況<sup>6</sup> 上位 10 位までの語）を示す。網掛けは所属語彙のないことを表す。

これらのうち、4.2 節および 4.3 節で見てきた、近代での出現頻度と語の変化パターンに相関の見られるもの、すなわち、近代で頻度が低いゆえに消失・減少していくものについては、語の固有の性質等ではなく、明治大正期を通じて当該の語が定着できなかったことによって、現代においても消失・衰退したものと捉えられるので、ここでは問題としない。むしろ、頻度が高く近代において十分定着し中心的な語として機能していたのにもかかわらず、現代にかけて衰退した語、あるいは、近代において十分に定着していなかったと思われるのにもかかわらず、現代にかけて興隆している語について、所属語彙の特徴を見出すことができれば、漢語の使用状況が変化する背景を、語の性質の面から解明することに繋がるものと思われる。そこで、以下の節では、注目する変化パターンごとに、語彙レベルの変動が大きいものを中心に所属語彙について特徴の記述を試みる。

表9 変化パターンと各時代の語彙レベルに基づく漢語の層別語例  
〈消失〉（近代での資料数>出現記事数>出現頻度上位 10 語まで）

近	現	G
A		
B		論者, 毎年, 僅々, 国人, 英人, 者流, 宇内, 足下, 征清
C		千古, 草木, 這般, 頃日, 洋行, 教徒, 才子, 小人, 際会, 英独
D		噴々, 千里, 沈淪, 講述, 寸毫, 虚名, 懷抱, 振作, 寥々, 適例
E		格外, 近着, 喧囂, 宿望, 随感, 局促, 錦繡, 今来, 宣明, 點頭
F		愛神, 惡婦, 安撫, 安福, 暗紅, 暗路, 偉器, 威能, 威烈, 意況

<sup>5</sup> 表1に示した資料別に出現した資料数の降順, 出現した記事数の降順, 出現した粗頻度の降順の順で並べ替えたもの。

<sup>6</sup> 出版SCに含まれる媒体(新聞・雑誌・書籍)の別に出現した媒体数の降順, 出現した記事数の降順, 出現した粗頻度の降順の順で並べ替えたもの。

表 10 変化パターンと各時代の語彙レベルに基づく漢語の層別語例  
 (減少) (近代での資料数>出現記事数>出現頻度上位 10 語まで)

現 近	A	B	C	D	E	F
A	結果, 世界, 勿論, 社会, 目的, 自然, 当時, 注意, 経済, 非常	今日, 一時, 容易, 次第, 多少, 以来, 種々, 進歩, 一種, 多数	到底, 一大, 東洋, 演説, 諸君	公使, 政友, 憲政	世人	余輩
B		大抵, 熱心, 都合, 機會, 有名, 依然, 他人, 幸福, 一面, 趣味	一々, 内外, 天地, 古来, 云々, 維新, 明白, 簡年, 年々, 交際	遺憾, 往々, 一事, 目下, 大勢, 苦心, 形勢, 面目, 多年, 近時	一言, 近来, 他日, 一朝, 一点, 諸氏, 識者, 政略, 高尚, 大家	騰貴, 区々, 一斑, 帝室, 三位, 泰西, 居士, 去月, 日英, 旭日
C			一拳, 目撃, 経歴, 明言, 確固, 完備, 氣力, 昨今, 名義, 精密	一読, 思慮, 一句, 大差, 見聞, 口実, 一概, 整頓, 見識, 鼓舞	精細, 紙上, 早晚, 諸般, 近傍, 名家, 趣意, 嘆息, 敬服, 体面	至大, 一分, 主意, 多事, 杞憂, 斯道, 恩沢, 一道, 聖賢, 暑中
D				清涼, 帰依, 叱咤, 死生, 通則, 傾倒, 未練, 帰途, 畏敬, 罌粟	一塊, 起居, 揚々, 決然, 追懷, 粗暴, 柔弱, 天神, 和氣, 蘇生	大息, 大言, 確然, 訛述, 愛兒, 西海, 要路, 茶話, 創痍, 一品
E					氣骨, 左官, 惡癖, 無碍, 早世, 端麗, 陽春, 要訣, 帷幄, 九死	倉卒, 霧中, 退校, 痴情, 唯々, 苦節, 後学, 幸甚, 腦病, 後嗣

表 11 変化パターンと各時代の語彙レベルに基づく漢語の層別語例  
 (不変) (近代での資料数>出現記事数>出現頻度上位 10 語まで)

現 近	A	B	C	D	E	F
A	以上, 問題, 一般, 研究, 以前, 普通, 組織, 今回, 中央, 今後					
B	一定, 科学	無理, 左右, 生命, 攻撃, 失敗, 海外, 交通, 結構, 得意, 解釈	器具, 木材, 主観, 分間, 電圧			
C		特有, 南部, 失礼, 死後, 悪魔, 国境, 証書, 急激, 感想, 信念	懸念, 有害, 夢中, 親戚, 用心, 肝心, 簡易, 誠実, 到来, 慣習	明快, 催促, 拮抗, 丁重, 身邊, 先月, 退治, 借用, 神仏, 仙人		
D			一撃, 詳述, 四肢, 未定, 華麗, 屈曲, 退学, 浸潤, 胸部, 嘔吐	一遍, 無実, 鼓動, 華奢, 正気, 称呼, 応変, 内緒, 口上, 先入	高見, 家宅, 至急, 專任, 縁日, 一喝, 管弦, 内密, 数理, 呵責	
E				親近, 年表, 学識, 一徹, 公職, 山賊, 代弁, 躍動, 巾着, 懇願	洋書, 開場, 督促, 秋冬, 固辞, 生国, 迷妄, 口述, 双眼, 潔癖	簡約, 寬厚, 水脈, 素読, 古刹, 朝命, 芳名, 暗愚, 神体, 好人
F					曲目, 中座, 余熱, 投石, 課外, 豪壯, 美名, 邦楽, 非番, 霧散	会者, 鉄柱, 文案, 某国, 法服, 開府, 丈六, 易学, 休校, 旧領

表 12 変化パターンと各時代の語彙レベルに基づく漢語の層別語例  
 (増加) (BCCWJ)における媒体数>出現記事数>出現頻度の上位 10 語まで)

現 近	A	B	C	D	E
A	自分, 必要, 関係, 意味, 人間, 一方, 方法, 生活, 最初, 理由				
B	時間, 可能, 時代, 存在, 実際, 本当, 以外, 十分, 中心, 最後	途中, 所為, 気分, 共通, 日常, 周囲, 突然, 積極, 週間, 半分			
C	現在, 部分, 状態, 全体, 同時, 基本, 自身, 同様, 確認, 対象	本来, 全部, 最終, 大事, 前後, 姿勢, 明確, 事情, 従来, 背景	最中, 皮肉, 幸運, 意地, 応援, 外出, 来年, 明記, 経由, 帰宅		
D	重要, 状況, 情報, 代表, 条件, 現実, 環境, 絶対, 文化, 自体	関心, 確保, 疑問, 魅力, 過程, 話題, 自信, 調整, 同士, 一瞬	贅沢, 悲鳴, 前日, 的確, 再度, 主催, 大胆, 実情, 初頭, 精々	啞然, 健在, 牽引, 熱気, 必至, 熱狂, 発散, 異変, 宝庫, 平静	
E	対応, 段階, 実現, 検討, 基準, 実施, 課題, 原則, 行政, 事項	事態, 周辺, 複数, 伝統, 強調, 体制, 限定, 把握, 現状, 世代	直面, 予感, 豪華, 順調, 膨大, 敏感, 情勢, 克服, 一環, 厳密	前身, 途轍, 判別, 長身, 全盛, 激減, 自負, 報復, 白髪, 殺到	大黒, 紆余, 貫祿, 辞令, 一報, 談義, 一倍, 同室, 難航, 秘話
F	具体, 導入, 画像	時点, 全員, 独自, 重視, 前提, 要因, 運営, 事例, 観点, 素材	抜群, 冒頭, 独占, 中核, 多彩, 優位, 原点, 市販, 共感, 定番	重厚, 愕然, 緊迫, 扨拭, 待望, 随所, 倍增, 激怒, 断続, 制覇	常套, 上々, 無骨, 疑心, 保身, 重鎮, 自責, 両論, 景品, 力点

表 13 変化パターンと各時代の語彙レベルに基づく漢語の層別語例  
 (出現) (BCCWJ)における媒体数>出現記事数>出現頻度の上位 10 語まで)

現 近	A	B	C	D	E	F
G	支援, 介護	高校, 強化, 本格, 実態, 移行, 発想, 対処, 最適, 現地, 連携	反論, 後述, 強引, 体調, 特集, 変身, 態勢, 週末, 迫力, 食材	弱体, 拍車, 後遺, 軌跡, 放映, 重度, 内包, 配送, 外食, 先週	一翼, 競演, 猛攻, 完敗, 交信, 廢材, 各駅, 急逝, 代役, 基軸	魚眼, 局側, 打順, 片翼, 暗喩, 異業, 温風, 仮題, 佳酒, 科技

## 5.2 消失する語の特徴

近代にのみ使用の見られる語のうち、頻度の高いものから 30 位の語を出現頻度（粗頻度）と共に示すと、以下の通りである。

足下 (603), 論者 (453), 毎年 (338), 国人 (328), 英人 (296), 僅々 (263), 征清 (255), 者流 (245), 宇内 (239), 這般 (202), 韓人 (200), 産額 (196), 教徒 (178), 洋行 (177), 英独 (156), 草木 (151), 南京 (146), 千古 (144), 小人 (132), 船中 (129), 高商 (128), 発兌 (128), 邦国 (127), 露清 (124), 宮相 (123), 顛覆 (121), 才子 (120), 官長 (117), 頃日 (117), 天理 (117)

このうち、「毎年」は、現代語においては「マイトシ」の読みが標準的であり、現代語コーパスでは統一的に「マイトシ」（コーパスの語種情報では漢語「毎」＋和語「年」で「混種語」）の読みを与え、漢語「マイネン」として扱っていない。一方、近代語においては、「毎年」に対する「マイトシ」「マイネン」のように、同一の表記が複数の語と結びつく時、その読みが振り仮名等により明確にならない場合は、原則として漢語と扱っている（国立国語研究所コーパス開発センター編 2016）。そのため、同一の表記に対して、認定される語彙素にずれが生じており、これが原因となって使用状況に差異が見られるものである。同様の背景を持つ語が、調査対象語には複数存在しており（「今年」「今日」「明日」「昨日」「大人」「海辺」等）、これらが語種の調査などにおいて近代の漢語比率を上げる原因にもなっている。コーパスの形態論情報を用いて調査・研究を行う際には留意すべき点であるが、これら同表記異語について、実際に、どちらの読み・語形を選択するのが適切であるかを見極めるには、データの原典資料を詳細に検討する必要があることはもちろんのこと、それ以外に、読みを確定しうる資料に基づく調査・研究が別途必要となるものであり、本研究ではこれ以上踏み込むことをしない。

また「論者」「者流」については、現代語において「無神論者」「学者流」のように、複合語として用いられることがほとんどで、この場合、語構成に基づき「無神 | 論 | 者」(×無神 | 論者)「学者 | 流」(「 | 」は形態素の切れ目を表す。以下引用例においても同様)のように解析されるため、現代語コーパスにおいては、「論者」「者流」の表記に対して一律に「論 | 者」「者 | 流」という解析結果が与えられている（いずれも非コア）。しかし、実際には近代に見られる例 (1) (2) と同様に、現代においても (3) (4) のような二字漢語としての用法が見られる（以下、用例提示においては、言及対象の漢語を【 】で強調し表示する）。やはりコーパスの形態論情報を用いる際の留意点であり、実態としてこれらは〈消失〉した語ではない。ただし、中納言の文字列検索で「論者」「者流」を検索し再分析した結果、BCCWJ 出版 SC (2001, 2005 年) で二字漢語として用いられている例は、「論者」15 例（レベル E）、「者流」1 例（レベル F）であり、〈消失〉には当てはまらないものの、使用状況において大きく変化のあった語であることは疑いがないことを指摘しておく（近代でのレベル B からいずれもレベルを大きく下げている）。

- (1) 横濱港が、東京築港の爲に衰微するであらうと憂ふる者があるが、【論者】は京濱兩港の各特質を備へ、兩立すべきことを思はざる僻見である。

(CHJ, 『太陽』1917, 坪谷水哉「東京築港實行の時期」)

- (2) 此かる【者流】は、今や後藤伯を崇拜し、後藤伯の神輿を擔がんとす、

(CHJ, 『国民之友』1888, 竹越與三郎「政治問題の唱首」)

- (3) 総体としてどのように評価するか、という点において、【論者】の立場が著しくことになっており、

(BCCWJ, 出版 SC・書籍 2001, 島田次郎「荘園制と中世村落」)

- (4) そんな分別は人間を全体として捉へることのでき強靱【者流】のたはごとに外ならぬのだ (BCCWJ, 出版 SC・書籍 2005, 出久根達郎『養生のお手本：あの人このかた 72 例』)

消失した語例の中で、「英人」や「英国」のように、外国地名を漢字一字で示し、一字漢語と

結合させて用いる二字漢語は、近代で多く作成され用いられたが、現代では一部の語（「英語」など）を除き外来語に取って代わられている。「南京」等の漢字圏の地名であっても、この傾向は同様である。

また、「国人」「邦国」など、国家や国民を指す語には、近代に見られた漢語が衰退・消失したものが多く。これらは、近代において「国人」と「国民」、「邦国」と「国家」「国（くに/コク）」のように、同じ概念を表す語が複数用いられている状態であったが、次第に「国民」「国家」に固定化され明治後期にはほぼこれに集約された。世界の中での日本が意識された明治初期において、「国」やそれに関連する概念を表す語が多く生まれたが、模索的であるゆえに多様だったものが、概念の定着に伴い語としても固定化され、類義語が淘汰集約されるという流れが見て取れる。



図8 「国人」と「国民」の使用状況推移  
(数字は語数)



図9 「邦国」と「国家」の使用状況推移  
(数字は語数)

(5) 我が【國人】の特質は一面に勇武に、一面に優美なりといふを得べし。

(CHJ, 『太陽』1895, 水谷不倒「戀愛小説」)

(6) 吾が國牀の特殊と吾が【國民】の特質とを外にしては、

(CHJ, 『女学雑誌』1894, 布川静淵「日本風景論を讀む」)

(7) 盖し大勢は【邦國】を勃興せしめ又【邦國】を滅亡せしむ、【邦國】の興亡する所以唯夫れ大勢を制すると否とに歸するのみ、

(CHJ, 『太陽』1895, 吉村銀次郎「國民の元氣」)

(8) 【國家】の興亡は必ずしも兵力の強弱に因るものにあらず、

(CHJ, 『太陽』1895, \*「〔社交案内〕」)

なお、「邦」は、近代において「国」を表す語素として、比較的高い造語力を持って用いられていたが、現代では「邦国」を含め半分以上の語が消失している。現代で新たに出現した漢語は5語のみであり、「邦」は漢語の語構成要素としての勢力を失っている。それゆえ「邦」を構成要素に持つ語それ自体も、同義で優勢である「国」を構成要素に持つ類語に浸食され衰退したも

のと考えられる。

表 14 「邦」を語構成要素に持つ漢語

類別	語数	語例 (近代粗頻度 / 現代粗頻度)
近現代両用	17	本邦 (557/28), 連邦 (231/338), 邦人 (180/14), 東邦 (98/8), 諸邦 (83/3), 他邦 (58/1), 異邦 (49/21), 各邦 (31/2), 小邦 (16/5), 邦語 (13/3), 邦訳 (9/22), 邦画 (5/4), 邦俗 (5/1), 万邦 (4/9), 領邦 (3/30), 邦楽 (3/7), 大邦 (1/1)
近代専用	31	邦国 (127), 邦家 (55), 友邦 (45), 隣邦 (31), 属邦 (10), 外邦 (8), 邦文 (6), 盟邦 (6), 列邦 (6), 弊邦 (5), 邦政 (5), 邦船 (5), 邦内 (5), 雄邦 (5), 旧邦 (3), 強邦 (2), 敵邦 (2), 邦教 (2), 邦言 (2), 邦民 (2), 自邦 (1), 邦音 (1), 邦外 (1), 邦境 (1), 邦訓 (1), 邦権 (1), 邦字 (1), 邦制 (1), 邦務 (1), 北邦 (1), 累邦 (1)
現代専用	5	邦銀 (9), 邦貨 (4), 邦題 (4), 邦貸 (1), 邦暦 (1)

その他、一般語において近代で多く使われていながら、現代では全く用いられなくなった語に「発兌」がある。「発兌」は 1895 年頃までは多数の記事に幅広く用いられ、『国民之友』や『女学雑誌』など、資料によっては優位な語であったが、その後定着することなく衰退する。これに取って代わるのが類語「発行」である。図 10 に、CHJ 明治大正雑誌の資料・年次別に「発兌」と「発行」の使用状況の推移を示す。「発兌」も前述「国人」「邦国」と同様、類義語が淘汰・集約された例と言える。

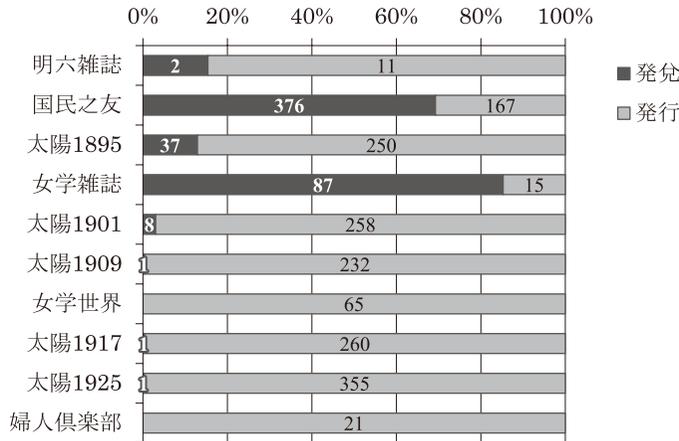


図 10 「発兌」と「発行」の使用状況推移 (数字は語数)

(9) 通俗論理談は朝夷六郎氏の譯にして、後凋閣より【発兌】せり、

(CHJ, 『国民之友』1887, \* 「諸近著の批評」)

(10) 反省會雑誌は佛法信徒中の有志が禁酒節酒を目的として【発行】せり、

(CHJ, 『国民之友』1888, \* 「刊雜書」)

一方、1894 年の日清戦争を受けて 1895 年の『太陽』の記事に偏って多用された「征清」などは、歴史や社会の流れの中で一時用いられた語であり、別語に取って代わられ淘汰される語の使用状況変化とは性質・背景を異にしている。

### 5.3 減少する語の特徴

ここでは、減少する語のうち、先述の通り近代において高頻度で基本語・中心語の位置を占めながら、現代において使用頻度が減少している語に着目する。

このうち、近代で高頻度に用いられる中心語でありながら、現代では低頻度語になっている語を、近代における調整頻度/現代における調整頻度と共に示す。

近代 A- 現代 F (1 語) : 余輩 (93.66/0.08)

近代 A- 現代 E (1 語) : 世人 (98.61/1.62)

近代 B- 現代 F (14 語) : 騰貴 (50.14/0.24), 居士 (32.9/0.24), 泰西 (26.99/0.24), 一斑 (24.35/0.24),

統監 (23.56/0.16), 旭日 (23.16/0.24), 日英 (22.92/0.24), 去月 (22.04/0.24), 三位 (21.08/0.08),

区々 (19/0.16), 協商 (18.44/0.24), 在野 (18.21/0.16), 帝室 (17.97/0.08), 対支 (17.57/0.16)

近代 B- 現代 E (94 語) : 列国 (88.23/0.41), 一言 (75.3/1.13), 近来 (70.99/0.57), 智識 (68.19/0.49),

諸氏 (63.4/1.3), 政略 (54.14/0.97), 職工 (49.19/1.38), 子爵 (45.99/0.49), 他日 (45.43/0.49),

書生 (42.88/1.7), 一朝 (42/1.54), 政事 (40.72/0.57), 器械 (37.53/1.22), 細君 (37.29/0.73),

一点 (37.05/1.46), 高尚 (36.97/1.22), 華族 (35.93/0.81), 改進 (35.61/1.3), 識者 (34.81/0.97),

婦朝 (34.33/0.89), 年来 (33.78/0.65), 畢竟 (33.62/0.81), 大家 (32.5/0.41), 世上 (31.94/0.41),

士族 (31.62/1.3), 砲兵 (31.54/0.49), 一夫 (31.46/0.41), 正貨 (31.38/0.32), 償金 (29.94/0.32),

開化 (29.46/1.46) ※上位 30 語

これらは、消失する語と同様、語として衰退していく傾向にある語だが、「書生」「華族」「士族」のように明治大正期特有の存在をさす語、「統監」のように現在にはない当時の役職に用いられた語などは、使用減少の背景が明らかでありここでは特段取り上げない。

一方で、ごく身近な概念として「去月」「近来」「他日」「一朝」といった時を表す語、「余輩」「世人」「諸氏」といった人称や人を表す語が多く含まれていることが注目される。これらは、いずれも近代よりも前の時代に用いられた古くからの漢文脈の流れを汲む漢語であり、そのため文語的・雅語的位相を持つ語として、現代においては使用が減少し、位相的にニュートラルな別語に移行している。近代では高頻度で用いられていたものが、現代で低頻度となっている語には、このような語の位相が関わるものが多くあるように思われる。

なお、近代 B- 現代 E に分類される「智識」は、現代語における「知識」と同語とみなせるもので、本来、表記のバリエーションとして語彙素「知識」と統合して扱うべきものであり、語そのものの使用の変化とは異なるものと思われる。同様に「政事」なども近代においては「政治」の表記バリエーションとして統合して扱うことが妥当であろう。3.3 節に示した通り、本研究では近代と現代と時代を跨いで漢語の比較を行うため、漢語の読み、表記、品詞等が時代により流動的で異なっている点を踏まえて、漢語の様々な層においてのバリエーションを同化して扱う立場に立つ。そのような立場からは、コーパスの語彙素の情報を、読み・品詞において同化処理したのと同様に、「智識」と「知識」、「政事」と「政治」などについても、意味等の観点から改めて分析用に基準を定めて用いる必要があるが、本稿においては、意味の分析を十分に行うことが

できなかつたため、意味による語彙素の統合は行わなかつた。これらの扱いについては、今後の課題としたい。

次に、現代でも高頻度で用いられていながら、相対的に近代より使用が減少している語を、近代における調整頻度 / 現代における調整頻度と共に示す（いずれも上位 10 語）。

近代 A- 現代 A (96 語) : 政府 (693.16/183.71), 政治 (626.33/164.1), 社会 (619.78/481.93), 国民 (588.64/156.73), 教育 (555.5/326.58), 世界 (465.52/416.53), 主義 (457.37/232.17), 結果 (382.47/343.68), 地方 (345.98/169.04), 国家 (342.39/126.58)

近代 A- 現代 B (74 語) : 今日 (733.01/20.02), 内閣 (385.19/52.35), 思想 (262.54/68.8), 選挙 (257.27/69.69), 帝国 (252.48/36.47), 鉄道 (249.61/28.04), 政党 (247.37/25.85), 進歩 (239.39/25.04), 宗教 (233.08/56.89), 勢力 (230.2/40.76)

これらの語の多くに共通するのは、政治的（「政府」「政治」「国民」「主義」「国家」「地方」「内閣」「選挙」「帝国」「政党」「勢力」等）・社会的（「社会」「教育」「思想」「宗教」等）性質の強い抽象概念を表す語であるという点である。これらの語の減少は、元となるデータの資料性を反映した結果であるように思われる。CHJ 明治大正雑誌は複数の雑誌のデータを元にしての、表 1 に語彙量を示した通り、その多くは雑誌『太陽』のデータからなる。『太陽』は幅広い読者を持つ総合雑誌として、当時の資料としてバランスの取れた構成にはなっているが（国立国語研究所 2005b）、近代という時代の新しい社会への志向性を反映し、論説、演説など社会的・政治的色合いの強い記事を多く含んでいる。次いでデータ量が豊富な『国民之友』（近藤 2014）についても、やはり啓蒙的な性質が強いため同様の性質を持つ。一方の BCCWJ 出版 SC は、現代において出版された書き言葉の縮図となるようバランスを取ってサンプリングされており、媒体も複数あり、記事のトピックもバリエーションに富む。そのため、CHJ 明治大正雑誌に見える時代背景に基づく社会性の強い語彙への使用の偏りは大きくなく、これが、使用頻度の差異となって現れているものと思われる。

#### 5.4 増加する語

現代にかけて使用が増加する語のうち、近代では低頻度であるが、現代においては対極的に高頻度で用いられるようになるものに着目し、近代における調整頻度 / 現代における調整頻度と共に示す。

近代 F- 現代 A (3 語) : 具体 (0.24/132.50), 導入 (0.24/91.25), 画像 (0.08/84.04)

近代 E- 現代 A (14 語) : 対応 (2.91/193.84), 基準 (2.65/167.26), 実施 (2.78/146.59), 住宅 (0.66/126.98), 段階 (0.79/122.28), 実現 (2.25/114.74), 検討 (1.06/108.75), 入力 (1.06/108.42), 原則 (1.98/102.67), 画面 (0.66/99.91), 課題 (1.98/92.13), 行政 (1.85/91.32), 文書 (0.66/87.60), 事項 (2.65/86.54)

近代 D- 現代 A (31 語) : 情報 (0/412.08), 重要 (3.84/264.51), 状況 (3.58/246.27), 環境 (5.43/236.31),

文化(7.69/196.6), 条件(6.1/184.04), 作成(6.76/162.32), 医療(4.11/153.49), 代表(4.77/149.03), 現実(3.98/137.36), 男性(7.29/126.91), 学習(5.3/125.2), 反応(7.16/115.64), 法人(5.83/113.37), 福祉(3.84/112.72), 特定(5.17/111.83), 当該(7.16/107.86), 分析(3.31/105.67), 業務(3.45/105.11), 絶対(5.83/104.78), 高齢(6.1/102.76), 団体(5.7/98.46), 取得(4.11/96.76), 対策(7.29/92.14), 操作(5.17/85.58), 自体(0/84.52), 現代(6.76/83.96), 努力(7.82/83.96), 登録(5.04/82.82), 再生(6.89/82.66), 体験(5.7/82.58)

近代 F- 現代 B (229 語): 素材(0.4/80.31), 部門(0.4/76.26), 要因(0.27/74.64), 譲渡(0.13/66.37), 番号(0.13/64.26), 全員(0.4/63.94), 時点(0/63.86), 運営(0.4/60.62), 事例(0.4/53.97), 重視(0.27/53.48), 該当(0.13/52.92), 前提(0.13/50.57), 活性(0.13/49.76), 独自(0.27/49.68), 個別(0.4/47.97), 駐車(0.13/47.33), 依存(0.27/47.16), 従業(0.27/46.43), 観点(0.27/43.76), 財務(0.13/41.98), 分布(0.4/41.25), 部品(0/37.68), 検索(0.27/37.52), 限界(0.13/37.52), 上記(0.4/37.03), 賠償(0.13/35.82), 融資(0.4/34.85), 基地(0.4/34.04), 弁済(0.13/33.71), 意欲(0.13/31.93) ※上位 30 語

近代 E- 現代 B (222 語): 体制(1.46/79.82), 事故(2.39/77.31), 伝統(0.8/70.58), 住民(1.33/67.67), 規制(1.19/67.26), 世代(1.06/66.13), 周辺(2.39/61.99), 概念(0.8/61.35), 複数(0.53/61.02), 分類(1.33/60.62), 自治(0.93/60.13), 達成(2.78/59.32), 給付(0.93/57.94), 事態(1.19/57.62), 実践(2.65/57.13), 構築(1.06/56.89), 把握(1.46/56.48), 優先(1.06/56.32), 合併(1.33/55.92), 限定(0.53/55.84), 視点(0.66/55.19), 取材(1.19/54.7), 動産(0.66/52.67), 防止(0.93/50.57), 充実(1.33/50.41), 投与(0.93/49.6), 強調(1.33/48.46), 水準(2.52/48.38), 需要(2.92/47.49), 現状(2.12/47.16) ※上位 30 語

これらの語に特徴的なのは、「導入」「対応」「実施」「実現」「入力」「検討」「努力」「対策」「反応」「作成」等に見られるように、サ変動詞用法を持つ語（波下線の語）が多い点である。前節では近代から現代にかけて減少する語に、政治的・社会的な語彙が多いことを指摘した。これらの語は、専ら名詞として抽象概念を表す用法のみを持つが、増加する語にサ変動詞用法を持つ漢語が多いのは、現代においては、政治・社会に関する大局的な概念に言及するのではなく、人間の動作や行動、営みに言論の視点が移行しているという背景を示しているように見える。さらに、同じ政治・社会分野の語であっても、「行政」「法人」「財務」「体制」「住民」「自治」といった語や、生活に関する「環境」「医療」「福祉」等が増えていることから、近代社会システムの成立期である近代から、それらが定着・発展する現代へと移り変わってきたことを背景として、言論の視点が、政治や社会の大局的な概念から、より細やかな社会システムや市民生活へと移行していることを指摘できる。

## 6. おわりに

本稿では、近代漢語語彙の特徴を明らかにすることを目的として、近代語および現代語の大規模コーパスを用い、漢語の網羅的な抽出と各種指標に基づく層化を試み、層別された漢語の特徴

の一端を示した。調査結果から明らかになったことをまとめると、概略以下の通りである。

1. 近代語コーパスと現代語コーパスのいずれかで用いられた漢語の総体は約5万語あり、そのうち45%の約22,400語は、両時代に共通して用いられる語である。それぞれの時代における漢語の総体は近代で約4万語、現代で3万語程度であり、近代では半数程度、現代では3/4程度が両時代に共通する語彙、残りが各時代に特徴的な時代専用の語である。漢語語彙は明治大正期から平成にかけて、半数程度が入れ替わっていると言える。語彙の入れ替えが起こったもののうち、明治大正期に用いられていたものが消失した語について、漢語の使用頻度に基づく語彙レベルとの相関を見ると、語彙レベルが低いものほど、消失する割合が高いことが分かった。
2. 語彙が入れ替わったもの以外に、両時代に共通する語彙の中でも、時代により使用状況に変化が見られる語が存在する。共通語彙のうち、①特に使用状況の変化が見られないものが42%、②使用が減少しているものが31%、③増加しているものが27%あり、これらの変化・不変化は、近代における漢語の使用頻度に基づく語彙レベルとの相関が見られた。近代で使用頻度の低いものは、現代語に引き継がれても周辺の語として残る傾向にあり、使用頻度の高いものは増加し語彙レベルを上げて定着する傾向にある。
3. 近代から現代にかけて消失、減少した漢語の中には、近代で既にその衰退傾向が明らかなのがあった。これらは、近代初期に模索的に生み出され、取り入れられたことにより多様であった漢語が、概念自体の定着と共に、用語が定着・固定化され、類義語が集約されたことによるものと考えられる。また、近代で多用され、現代では衰退している語の中には、古くから用いられた漢文脈の流れを組む漢語があり、これらは現代語において「硬い」「文語的な」文体を持つ語と認識されて使用が減少している可能性がある。
4. 使用頻度との相関でその語の変化を説明できるものの他に、意味的な語彙カテゴリーと語の使用状況の変化に関わりを認められるものがあった。近代での頻度が極めて高いにもかかわらず、現代において使用が有意に減少する漢語には、政治的・社会的な抽象概念を示すものが多く含まれていた。これは、近代社会における言論において、これらの概念が重要視され取り上げられていたのに対し、現代では、厳密なサンプリングに基づくコーパス設計という背景とも相まって、話題が分散し偏りが大きくないことが想定される。また、逆に、近代では頻度が低く周辺の語にとどまっていた漢語が、現代で使用を増やした語には、サ変動詞として用いられる漢語が多く含まれていた。この結果は、現代においては、人間の動作や行動、営みに言論の視点が移行していることを示唆しているのではないか。

なお、本研究の分析過程で問題を把握しながら今後解決すべき課題として残した事柄についても整理しておく。

1. 語の認定基準について。今回は、同表記の異語形を統一的に同語とみなして扱ったが(3.3節)、個々の語によって検討が必要なものもあろう。これは、同表記の異語(語種の異なる語。5.2節)や、同語異表記とみなせる語(5.3節)についても同様であり、今後、近代における

漢語、近代語彙を正面から扱おうとする時には、避けて通ることができない課題と認識している。検討を行い、指針を定めて調査を進めたい。

2. コーパスに付与された情報の適切性について。扱うデータは巨大で、その全てに目を通すことはできないが、分析において問題となる箇所や、重点的に扱う語については、分析過程で判明した誤解析や、現代語と近代語における情報付与基準の差異を整理した上で情報を改めて分析を進める必要がある。
3. 変化の詳細な様相について。本稿では、近代と現代の対照において異なりを明らかにすることを主眼として分析を行ったが、本稿の調査により変化したことが判明した語については、その変化の過程を丁寧に追う必要がある。近代から現代にかけて徐々に変化する語もあれば、ある一時期に一気に変化へと向かうものもあり（5.2節「征清」など）、近代における変化の様相が、その後の語の歴史・将来を方向づける可能性が高い。また、その変化の背景には、用法や意味の変化等が関与する可能性がある。変化の様相と語の歴史の関わりを明らかにすることで、言語変化のメカニズムの一端を明らかにすることが可能になると考える。引き続き、詳細な分析を行いたい。
4. 語の出自と定着の相関について。近代に用いられた漢語が現代に引き継がれ用いられたり、淘汰されたりする背景に、語の出自が関わるかという点に言及した研究は多くない。5.3節に若干触れた通り、その語がどのような文体・場面で用いられる語であったかという位相的側面、また、近代化の流れで新たに用いられた新漢語なのか、また和製漢語か、あるいは古来から変わらず用いられてきた漢語なのかといった語史の側面は、近代語から現代語への漢語語彙の変化の実態と関連が小さくないと考えている。検討を続けたい。

## 参考文献

- 坂野潤治（2012）『日本近代史』（ちくま新書）東京：筑摩書房。
- 飛田良文（1966）「明治以後の語彙の変遷」『言語生活』182: 16-24。
- 池上禎造（1953）「近代日本語と漢語語彙」金田一博士古稀記念論文集刊行会（編）『民族論叢：金田一博士古稀記念言語』187-206。東京：三省堂。
- 国立国語研究所（1964）『現代雑誌九十種の用語用字 III 分析』（国立国語研究所報告 25）東京：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）東京：秀英出版。
- 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）東京：国立国語研究所。
- 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—』（国立国語研究所報告 122）東京：博文館新社。
- 国立国語研究所コーパス開発センター（近藤明日子）編（2016）『近代文語 UniDic 短単位規程集 Ver.1.1』（[http://pj.ninjal.ac.jp/corpus\\_center/chj/doc/unicdic-MLJ\\_rulebook\\_v1\\_1.pdf](http://pj.ninjal.ac.jp/corpus_center/chj/doc/unicdic-MLJ_rulebook_v1_1.pdf)）。
- 近藤明日子（2014）『『国民之友コーパス』解説書 第 1.1 版』（[http://www.ninjal.ac.jp/corpus\\_center/cmj/doc/kokumin\\_manual\\_v1\\_1.pdf](http://www.ninjal.ac.jp/corpus_center/cmj/doc/kokumin_manual_v1_1.pdf)）。
- 今野真二（2012）『百年前の日本語—書きことばが揺れた時代』（岩波新書）東京：岩波書店。
- 前川喜久雄（2008）「KOTONOHA『現代日本語書き言葉均衡コーパス』の開発（〈特集〉資料研究の現在）」『日本語の研究』4(1): 82-95。
- 宮島達夫（1967）「近代語彙の形成」『ことばの研究』3: 1-50。東京：秀英出版。
- 永澤済（2010）「変化パターンからみる近現代漢語の品詞用法」『東京大学言語学論集』30: 115-168。東京大学文学部言語学研究室。

- 鳴海伸一 (2015) 『日本語における漢語の変容の研究：副詞化を中心として』(ひつじ研究叢書〈言語編〉125) 東京：ひつじ書房.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011) 『『現代日本語書き言葉均衡コーパス』形態論情報規程集第4版(下)』(特定領域研究「日本語コーパス」平成22年度研究成果報告書)([http://pj.ninjal.ac.jp/corpus\\_center/bccwj/doc/report/JC-D-10-05-02.pdf](http://pj.ninjal.ac.jp/corpus_center/bccwj/doc/report/JC-D-10-05-02.pdf)).
- 阪本一郎 (1984) 『新教育基本語彙』東京：学芸図書.
- 鈴木丹士郎 (1998) 「明治期漢語の品詞性と語形についての一考察」東京大学国語研究室創設百周年記念国語研究論集編集委員会(編)『東京大学国語研究室創設百周年記念国語研究論集』728-750. 東京：汲古書院.
- 武部良明 (1981) 『日本語表記法の課題』東京：三省堂.
- 田中牧郎 (2009) 「言語政策に役立つ、コーパスを用いた語彙表・漢字表などの作成と活用(〈特集〉日本語コーパス)」『人工知能学会誌』24(5): 665-672.
- 田中牧郎 (2011) 「語彙レベルに基づく重要語彙リストの作成—国語政策・国語教育での活用のために—」田中牧郎・相澤正夫・斎藤達哉・棚橋尚子・近藤明日子・河内昭浩・鈴木一史・平山允子『言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用』(特定領域研究「日本語コーパス」言語政策班報告書) 77-87. 東京：国立国語研究所.
- 田中牧郎 (2016) 「第8章 語種」斎藤倫明(編)『日本語語彙論 I』(講座 言語研究の革新と継承1) 241-274. 東京：ひつじ書房.

## 関連 Web サイト

- 国立国語研究所『現代日本語書き言葉均衡コーパス』[http://pj.ninjal.ac.jp/corpus\\_center/bccwj/](http://pj.ninjal.ac.jp/corpus_center/bccwj/) (中納言 2.2.0 (非 NumTrans 版), 2017年2月確認)
- 国立国語研究所 (2016) 『日本語歴史コーパス』<https://chunagon.ninjal.ac.jp/> (中納言 2.2.1 データバージョン 2016.10, 2016年12月確認)

## A Corpus-Based Study of Sino-Japanese Words in Modern Magazines: A Comparison with BCCWJ

MABUCHI Yoko

Ph.D Student at Meiji University / JSPS Research Fellow (DC) /  
Project Collaborator, NINJAL

### Abstract

Using morphologically annotated corpora of modern and contemporary Japanese (“Corpus of Historical Japanese” [CHJ] Meiji-Taishō Series I – Magazines and “Balanced Corpus of Contemporary Written Japanese” [BCCWJ]), I conducted a survey and analysis of the differences between Sino-Japanese words. By examining the frequency and coverage of individual words, the following three points were clarified regarding the 55,000 words utilized in the modern and contemporary eras:

- 1) Altogether, 50,000 words are used in CHJ and BCCWJ, and 22,400 words are employed in both corpora. About 19,400 words (e.g. “発兌,” “状師,” “英人,” “征清,” “邦国”) are used only in the modern era, and 7,700 words (e.g. “支援,” “介護,” “強化,” “本格,” “展示”) are used only in the contemporary era, indicating that Sino-Japanese vocabulary is larger in the modern era than the contemporary era.
- 2) A total of 42% of the common words (e.g. “問題,” “世界,” “以上,” “研究,” “経済”) in CHJ and BCCWJ have no significant differences. However, the other 31% (e.g. “今日,” “吾人,” “憲政,” “列国,” “世人”) decreased, and 27% (e.g. “時間,” “意味,” “可能,” “企業,” “利用”) increased significantly. Thus, Sino-Japanese vocabulary changed from the modern era to the contemporary era.
- 3) These diachronic changes are caused by historical context, consolidation of synonyms, stylistic changes, and changes in the themes of texts.

**Key words:** Sino-Japanese words, modern Japanese, morphologically annotated corpora, word frequency, diachronic changes of vocabulary