

# 近現代日本語の意味変化分析のための単語データセット構築の試み

間淵 洋子      小木曾 智信  
 国立国語研究所 言語変化研究領域  
 {mabuchi, togiso}@ninjal.ac.jp

## 1 はじめに

近年、日本語のコーパスが相次いで構築・公開されたことにより、コーパスを用いた日本語の実証的研究が盛んに行われている。中でも、国立国語研究所構築の『現代日本語書き言葉均衡コーパス』(以下、BCCWJ) や『日本語歴史コーパス』(以下、CHJ)「明治・大正編」を用いた近現代日本語を対象とした研究は、データ量や文章バリエーションの豊富さを背景に、数多く行われており、BCCWJ, CHJ「明治・大正編」を複合的に利用することで、近代から現代にかけての日本語の通時的変遷を見出そうとする研究も見られる(間淵 2017[1], 高橋・東泉 2018[2], 市村 2019[3], 馬場 2019[4]等)。

一方、自然言語処理の分野では昨今、言語変化を検出する手法が多く提案されており、英語・ドイツ語・ロシア語などでは、提案手法と共に通時的変化を生じた単語のリストが共有されている(Kulkarni et al. 2015[5], Schlechtweg et al. 2018[6], Rodina et al. 2020[7])。これらのリストは、新たな手法の評価等に利用することができ、言語変化の検出という自然言語処理の一課題に対して有用なリソースとして機能している。

そこで本研究では、両研究領域にわたって近現代日本語の意味変化分析に利用可能な単語データセットを構築することを目的として、意味変化が想定されるいくつかの語を対象に、コーパスでの使用実態調査を行う。コーパス調査に基づき、これらの語の通時的変化の様相を明らかにした上で、言語変化の生じた単語のリストを試作する。

## 2 研究方法

### 2.1 調査対象語

本研究では、日本語学および自然言語学の両分野において利用可能な単語データセットの構築を目指すことから、データセットを構成する単語の条件と

して、(1)使用頻度の高さ、(2)近代から現代にかけて(特に戦前・戦後間)の意味変化の大きさ、の2点を定めた上で、この条件を満たす近現代日本語を探索するために、コーパス調査の対象とする候補語リストを作成した。

調査対象語のリストは、CHJ「明治・大正編」とBCCWJの網羅的な調査に基づき近現代漢語の表記・品詞用法の実態把握を行った間淵 2018[8]の結果を元に、近代から現代にかけて品詞用法の使用割合に大きな変化が見られた漢語を中心として、意味変化の認められる語を選定し、ジャパンナレッジ版『日本国語大辞典』「語誌」欄で戦前・戦後間での意味変化に触れている語などを適宜追加した。

【和語】 旨い、心持ち、衣、こだわる、痺れる、逆も、増埒

【漢語】 愛人、管制、貴族、教授、教養、結構、広告、故障、自然、住居、渋滞、柔軟、主婦、障害、情報、女性、尋常、精々、設備、全然、団塊、端末、適当、澁刺、非常、風俗、婦人、普段、普通、明細、免許、遊撃、優勝、要領、了解

【外来語】 カフェ、ケース、スーパー、ポイント、ボタン、モデル

### 2.2 コーパス

前節に示した調査対象語の使用実態を把握するために使用するコーパスの概要を表1に示す。

表 1 使用コーパスの概要

コーパス	収録年次	時代区分	語数 (記号類除く)
CHJ明治・大正編   雑誌	1874-5, 1881-2, 1887-8, 1894-5	1明治前期	4,008,418
	1901, 1909	2明治後期	4,377,329
	1917, 1925	3大正	4,367,844
SHC雑誌	1933, 1941	4昭和前期	5,838,258
	1949, 1957, 1965	5昭和中期	6,194,199
	1973, 1981	6昭和後期	5,010,046
	1989-1997	7平成前期	5,325,362
BCCWJ雑誌	2001-2005	8平成後期	4,444,492

明治・大正期における使用実態調査には CHJ「明治・大正編 I 雑誌」を用いる。明治・大正期に続く昭和期以降の使用実態調査には、現在筆者らが構築を進めている『昭和平成書き言葉コーパス』(以下 SHC, 近藤他 2020[9]) の雑誌データの一部を用いる。このコーパスは、既存の近代語のコーパスである CHJ「明治・大正編」と、現代語のコーパスである BCCWJ の間にある時間的空隙を埋め、さらに BCCWJ 以降現在までの期間のデータを追補することを目的として設計したコーパスである。現況で整備が進んでいる 1997 年までの雑誌データを利用する。さらに、現代の使用実態を BCCWJ 出版サブコーパス・雑誌により確認する。これらのコーパスを複合的に用いることで、近代語から現代語にかけて生じている意味変化の実証的な分析が可能となる。

なお、表 1 に示した「時代区分」は、言語変化が時代背景や社会情勢に大きく影響されることを鑑み、日本史上の区分に大きく分けた上で、コーパスの収録語数に大きな偏りが生じないよう考慮した任意の区分である。

## 2.3 意味分類

次に、コーパスで取得した用例の分析方法について述べる。

本研究では、コーパスの調査に基づき、通時的な意味変化の様相を明らかにしようとするものであるため、まず必要となるのは、各用例がどの意味で用いられているかを判断することである。ここで問題になるのは、各語の意味をどのように、またいくつに弁別するかという点である。

例えば、和語「痺れる」を例に、辞書による意味の捉え方・分類のし方を確認すると、辞書の性質によって差異があることが分かる<sup>i</sup>。

### ■デジタル大辞泉

- 1 からだの一部または全体の感覚が失われ、自由がきかなくなる。「正座して足が一・れる」
- 2 電気などを感じてびりびりふるえる。「感電して一・れた」
- 3 心を奪われてうっとりとする。強烈な刺激を受けて陶酔する。「ジャズ演奏に一・れる」

<sup>i</sup> いずれもジャパンナレッジ版より引用。

<sup>ii</sup> 以下、用例掲出においては、コーパス本文に続けて出典を〈サンプル ID, 開始位置〉の形式で示す

### ■プログレッシブ和英中辞典

I 〔麻痺(まひ)する〕become numb; 〔電気で〕receive an electric shock

II 〔陶酔する〕be enraptured, be bewitched, be fascinated ((by))

本研究では、近代から現代にかけての変化が大きく捉えられる点を重視するため、できるだけ大まかに意味を捉えることとし、可能な限り、原義的(古い用法, 近代または戦前に多い用法)か、転義的(新しい用法, 現代または戦後に増えている用法)かの観点で意味分類を行うこととする。すなわち、上記「痺れる」の例では、『プログレッシブ和英中辞典』と同様、肉体的な作用(原義, 例 1)と心理的な作用(転義, 例 2)の 2 分類で用例分類を行う<sup>ii</sup>。

(例 1) 急に降りようとして立上ると、復た先日重かつた足がしびれて利かないので、暫く擦つて居るとなほつた。

〈60M 太陽 1925\_03058, 42480〉

(例 2) 赤と黒のコントラストにしびれちゃった。

〈PM11\_00842, 31800〉

## 3 調査結果

### 3.1 頻度変化と意味変化

2 節に示した調査対象語のコーパス調査から、まず頻度の変化に着目して結果を確認する。付録として、調査対象語のコーパスの収録語数(空白、補助記号、記号を除く)100 万語あたりの調整頻度(PMW)を掲載し、以下の議論ではこれを用いるが、紙幅の都合上、和語・外来語を中心に述べる。

**和語** 和語の候補語のうち、「痺れる」「坩堝」は、低頻度で推移しているため、データセットとしての条件を満たさない。頻度の下降が見られる「心持ち」は、戦前(昭和前期)まで[気立て・気持ち]の意味での用法が見られるが、戦後はこれが衰退し、程度がごく僅かであることを表す副詞としての用法で残存するものと思われる。

(オンライン検索アプリケーション『中納言』でんの「位置検索」により当該用例の再検索が可能。)

同様に下降する「衣」は、原義である[衣服・着物]の意が衰退し、[揚げ衣]の意での用法が保たれる。頻度の上昇の見られる「こだわる」は、従来の[執着する]意がマイナスの意味を持って用いられる一方、プラスの転義(意味の上昇)として、[妥協せず追求する]意を獲得することで頻度を上昇させていると考えられる。[味覚的な良さ]に加えて、より抽象的な[程度の高さ]の意を獲得した「旨い」も同様である。

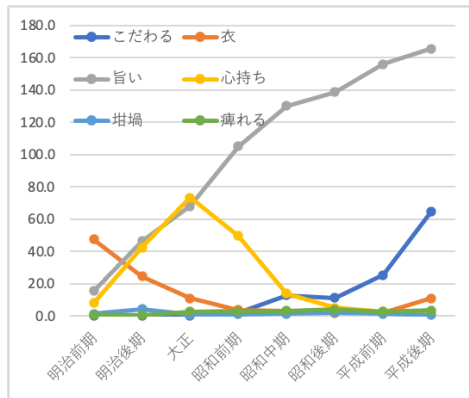


図 1 和語の頻度推移 (PMW)

**外来語** 外来語語彙は、いずれも明らかに頻度の上昇が見られた。「ボタン」「ケース」「ポイント」「モデル」等、意味の一般化・抽象化による意味の拡張・多義語化が、頻度上昇の要因となっているものと考えられる。

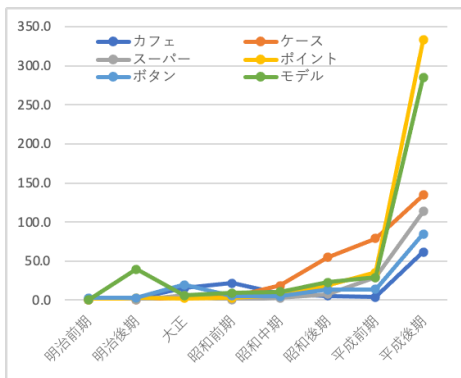


図 2 外来語の頻度推移 (PMW)

### 3.2 データセットの選定

前節までの議論から、候補語のうち、頻度条件を満たさないもの、意味分類が困難なもの、意味分布の通時的な差が大きくないものを除き、意味変化分析に適した単語データセット試作として、以下のリストを提示する。

【和語】 旨い, 心持ち, 衣, こだわる, 逆も

【漢語】 管制, 貴族, 教授, 教養, 結構, 故障, 自然, 住居, 渋滞, 柔軟, 障害, 情報, 女性, 尋常, 精々, 設備, 全然, 適当, 非常, 風俗, 普通, 免許, 優勝, 要領, 了解

【外来語】 ケース, ポイント, ボタン, モデル

## 4 おわりに

日本語研究, 自然言語処理の両領域にわたって, 近現代日本語の意味変化分析に利用可能な単語データセットを構築することを目的として, コーパスでの使用実態調査に基づき, 通時的な意味変化を生じた単語データセットを試作した。

本研究では, 意味変化をできるだけシンプルに捉えるために, 原則として原義・転義の2分類に意味分類し意味分布とその推移を確認したが, そのためにやや強引な用例分類を行った面もある。今後, この単語データセットを意味分析の数理的研究や, 言語変化の検出モデルの構築などに活用することで, データセットの問題点の把握, 見直しなどを行っていく。

## 付記

本研究は, 国立国語研究所共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」, 同「現代語の意味の変化に対する計量的・統計学的アプローチ」, JSPS 科研費 19H00531, JSPS 科研費 20K13060 による研究成果の一部である。

## 参考文献

- [1] 間淵洋子. 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—. 国立国語研究所論集, 13, 2017, 143-166.
- [2] 高橋圭子・東泉裕子. 近現代語コーパスにおける「(さ)せていただく」の用法. コミュニケーション文化, 12, 2018, 177-192.
- [3] 市村太郎. 副詞「ほんとうに」の展開と「じつに」「まことに」—近代語から現代語へ—. 国文学研究, 188, 早稲田大学国文学会, 2019, 1-15.

- [4]馬場俊臣. 助詞・助動詞で始まる複合接続表現の文体差. 北海道教育大学紀要(人文科学・社会科学編), 70 (1), 2019, 1-11.
- [5]Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. Republic and Canton of Geneva: CHE, In Proceedings of the 24th International Conference on World Wide Web, WWW '15, 2015, 625–635. International World Wide Web Conferences Steering Committee.
- [6]Dominik Schlechtweg, Sabine Schulte im Walde, Stefanie Eckmann. Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), 2018, 169-174. Association for Computational Linguistics.
- [7]Julia Rodina, Andrey Kutuzov. RuSemShift: a dataset of historical lexical semantic change in Russian. Proceedings of the 28th International Conference on Computational Linguistics, 2020, 1037–1047. International Committee on Computational Linguistics.
- [8]間淵洋子. 近代漢語における表記・語法の多様性とその変化に関する計量的研究: 現代語確立期にみる言語変化の様相と背景. 明治大学, 2018.
- [9]近藤明日子, 小木曾智信, 高橋雄太, 田中牧郎, 間淵洋子. 「昭和・平成書き言葉コーパス」の設計. 日本語学会 2020 年度秋季大会予稿集, 日本語学会, 2020.

## 付録 近現代日本語意味変化分析のための単語データセット頻度表(PMW)

語種	語彙素	原義	転義	明治前期	明治後期	大正	昭和前期	昭和中期	昭和後期	平成前期	平成後期	変化類型
和	こだわる	執着	無妥協	0.0	0.0	0.5	1.9	12.8	11.2	25.4	64.8	大幅な上昇
和	衣	着物	揚げ衣	47.6	24.7	11.0	3.8	3.1	2.0	2.3	11.0	大幅な下降
和	旨い	味覚	程度・高	15.5	46.4	68.0	105.2	130.1	138.7	156.2	165.8	大幅な上昇
和	心持ち	気持ち	程度・低	8.0	42.5	73.5	49.7	14.0	5.2	2.8	2.2	大幅な下降
和	坩堝	溶解炉	混沌・熱狂	1.5	4.3	0.9	1.2	1.5	2.2	1.3	0.7	低く推移
和	痺れる	身体作用	心理作用	1.0	0.7	2.7	3.1	3.4	4.0	2.8	3.6	低く推移
和	込も	どうしても	程度・高	36.9	50.0	60.7	54.5	66.4	124.0	150.2	195.7	大幅な上昇
漢	愛人	愛する人	不貞相手	2.7	1.4	8.7	5.1	6.6	9.2	18.6	9.7	上昇後下降
漢	管制	管理統制	飛行管制	0.0	0.2	2.1	2.7	6.0	11.0	4.7	2.2	低く推移
漢	貴族	身分	優雅な状況	84.6	185.0	112.4	42.1	25.5	31.1	15.4	8.5	大幅な下降
漢	教授	教える	職位	117.5	122.0	138.5	321.2	170.5	243.7	245.6	106.0	高く推移
漢	教養	教育する	「教養のある・ない」	14.0	5.0	13.3	23.0	26.6	12.0	22.5	4.5	下降
漢	結構	構造・優れた様	程度・高	43.2	42.5	47.6	36.3	39.7	43.9	57.1	66.1	上昇
漢	故障	支障	機械身体	8.2	14.8	19.5	6.3	8.6	13.0	9.4	12.8	低く推移
漢	広告	報知	商業広告	28.9	19.9	26.3	36.8	63.3	38.1	48.1	59.6	上昇
漢	自然	Nature	自ずから	265.7	361.0	278.9	238.9	167.1	144.9	145.5	234.4	高く推移
漢	主婦	家を仕切る女性	結婚した(無職)女性	8.0	20.3	17.9	11.5	20.0	30.1	31.4	40.5	上昇
漢	住居	住む	家	20.5	8.9	3.4	3.4	5.8	16.4	11.1	11.0	下降
漢	柔軟	触覚	触覚+精神	5.0	5.7	1.8	2.4	8.1	11.2	13.7	19.6	上昇
漢	渋滞	滞り	交通渋滞	6.5	5.3	4.8	2.9	1.8	2.6	5.8	8.8	低く推移
漢	女性	女性の性質	女性	36.9	20.6	96.8	53.3	105.6	250.1	348.3	404.3	大幅な上昇
漢	障害	差し支え	身体・精神障害	20.2	23.5	28.8	30.1	28.3	33.7	44.5	75.6	上昇
漢	情報	報知		0.0	1.6	4.1	23.0	82.3	172.7	333.3	306.2	大幅な上昇
漢	尋常	通常	「尋常でない」	89.8	29.7	19.2	10.1	8.4	4.4	6.0	2.7	大幅な下降
漢	精々	精一杯	たかだか	1.5	4.1	7.6	12.3	25.0	25.5	19.2	9.0	上昇後下降
漢	設備	備え	機械・施設	4.5	86.1	75.1	43.2	71.4	36.5	31.9	29.2	上昇後下降
漢	全然	まったく	(否定呼応)	15.0	95.7	109.9	76.1	95.4	59.1	54.3	53.5	上昇後下降
漢	端末	はし・末端	端末機	1.0	1.8	0.0	0.0	0.2	2.0	5.8	15.3	上昇
漢	団塊	塊	「団塊の世代」	0.5	4.1	0.0	0.0	0.0	0.4	13.3	3.1	低く推移
漢	適当	適切・妥当	いい加減	91.1	131.1	105.3	63.4	44.9	32.3	16.1	14.6	大幅な下降
漢	澆刺	魚の勢い良い様	人の元気の良い様	1.7	1.8	8.5	6.9	4.4	2.6	3.2	1.3	低く推移
漢	非常	いつもと異なる	程度・高	136.5	318.5	406.4	341.9	365.7	198.0	169.2	121.0	上昇後下降
漢	婦人	成人女性	既婚女性	357.0	250.6	265.8	70.7	72.8	79.6	33.8	24.7	大幅な下降
漢	普段	常に	日常	2.0	9.1	17.9	36.8	23.7	24.8	33.4	83.9	大幅な上昇
漢	普通	一般に通ずる	通常・並	137.2	175.9	232.6	104.5	95.3	100.4	143.5	131.4	高く推移
漢	風俗	風習・習俗	風俗店	92.1	108.7	19.7	17.3	28.6	12.4	11.5	9.4	大幅な下降
漢	明細	詳細	会計明細書	2.5	2.1	0.9	0.5	0.8	0.4	2.3	1.1	低く推移
漢	免許	許す・認める	運転・資格免許	12.7	5.5	4.3	3.9	10.7	10.2	20.1	29.7	上昇
漢	優勝	優れ勝る	一位	9.2	9.8	8.0	7.0	10.5	24.0	46.9	107.3	大幅な上昇
漢	遊撃	戦況に応じ攻撃する	野球の遊撃手	6.7	0.7	1.4	5.0	2.6	0.4	1.7	0.4	低く推移
漢	要領	要点	「要領が良い」	13.0	21.5	18.3	7.7	10.2	7.4	7.3	10.1	低く推移
漢	了解	理解	承知	19.7	35.2	48.5	36.1	27.3	23.2	19.0	5.6	上昇後下降
外	カフェ	女給のいる店	喫茶店	0.2	1.8	15.6	21.4	8.2	5.2	3.6	61.4	大幅な上昇
外	ケース	容器	場合	0.2	0.0	3.9	4.1	18.7	54.7	78.7	134.5	大幅な上昇
外	スーパー	超	店	0.7	0.2	5.7	0.2	2.7	7.4	28.9	113.6	大幅な上昇
外	ポイント	場所	得点要点	0.2	2.5	2.3	2.6	6.1	18.8	35.5	332.8	大幅な上昇
外	ボタン	釦	装置	3.0	2.7	19.7	5.3	5.0	13.2	13.9	84.6	大幅な上昇
外	モデル	原型	手本・型式・定式	0.2	39.5	6.0	9.1	10.5	22.8	28.9	285.3	大幅な上昇