

令和 2 年 6 月 19 日現在

機関番号：62618

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K02750

研究課題名(和文)形態論情報付きコーパスを活用した近代日本語の位相の計量的研究

研究課題名(英文) Register Variations in Modern Japanese: Quantitative Analysis of a Morphologically Annotated Corpus

研究代表者

近藤 明日子 (KONDO, Asuko)

大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・プロジェクト非常勤研究員

研究者番号：30425722

交付決定額(研究期間全体)：(直接経費) 1,200,000円

研究成果の概要(和文)：本研究では、形態論情報の付与された『日本語歴史コーパス 明治・大正編 雑誌』を用いて、近代語の位相差を明らかにするために、次のことを行った。まず、会話文の話者の性別・社会階層等の属性データを新たに作成し、コーパスの形態論情報と関連付けた。そして、品詞率・語種率について計量的分析を行うことにより、話者の性別・社会階層に基づく位相差があることを明らかにした。

研究成果の学術的意義や社会的意義

形態論情報付きの大規模近代語コーパスを利用した計量的分析という研究手法を採用することにより、語種率・品詞率といった従来の研究とは異なる視点から、近代日本語における位相差の実態に迫ることができた。また、本研究で新たに作成した使用者属性情報は、『日本語歴史コーパス 明治・大正編 雑誌』の付加情報の一つとして公開し、広く一般に利用できるようする予定である。

研究成果の概要(英文)：This study explores register variations in modern Japanese through quantitative analyses of the "Corpus of Historical Japanese Meiji-Taisho; Period Series I, Magazines," a morphologically annotated corpus of modern Japanese. Data on speakers' gender and social status were prepared and related to morphological annotations of the corpus. Quantitative analysis of the corpus based on etymological type ratios and part-of-speech ratios allowed the identification of variations determined by speakers' gender and social status.

研究分野：日本語学

キーワード：近代語 語彙 位相差 品詞率 語種率 コーパス 計量的分析

1. 研究開始当初の背景

近代日本語の研究において、著者や話者の性別・社会階層等の属性による位相差は重要な観点として従来の研究でしばしば取り上げられてきており、各種言語項目において位相差があることが先行研究によって明らかにされてきた。ただし、各研究者の主観による研究対象の限定、あるいは従来の人手作業による調査分析の限界により、位相差の全体像の把握は十分ではないという課題もあった。

このような課題を解消する方法として、バランスのとれた大規模なコーパスを利用した研究手法の採用が考えられる。しかし、これまでに公開されている代表的な近代語のコーパスである国立国語研究所(編)(2005)『太陽コーパス』、国立国語研究所(編)(2006)『近代女性雑誌コーパス』は説文から小説にいたる様々なジャンルの文章が収録されたバランスの取れたコーパスではあるものの、コーパスを利用した本格的な研究には必須の形態論情報(語に関する情報)が付与されていなかった。こうしたなか、『太陽コーパス』『近代女性雑誌コーパス』の収録記事を増補し、新たに形態論情報を付与したコーパスが構築され、国立国語研究所『日本語歴史コーパス 明治・大正編 I 雑誌』(http://pj.ninjal.ac.jp/corpus_center/chj/meiji_taisho.html) (以下、「明治・大正編 I 雑誌」と呼ぶ)の一部として公開されることとなった。この新しい形態論情報付きの近代語コーパスを利用することで、近代語の位相について俯瞰的視点から全体像を描き出すことが可能となる。

2. 研究の目的

本研究の目的は、「明治・大正編 I 雑誌」の形態論情報と、本研究で独自に付与する著者・話者の属性(性別・階層・年齢)のデータとを関連付け、従来の研究手法では調査・分析の困難であった言語項目(品詞率・語種率等)について著者や話者の属性による位相差を計量的分析によって明らかにすることである。

3. 研究の方法

(1) 著者・話者の属性に関するデータの作成

「明治・大正編 I 雑誌」には記事の著者については性別と年齢について情報が付与されている。しかし、社会階層については情報が付与されていない。また、記事中の会話部分の話者については性別・社会階層・年齢のすべてについて情報が付与されていない。そこで、これらについて本研究において独自に情報付与を行った。性別は「男性/女性」の2種、階層は「知識層/非知識層」の2種、年齢は「子/成人/老人」の3種に分類し、それぞれ情報付与を行った。そして、コーパスの形態論情報との関連付けを行った(表1)。

表1 コーパスの形態論情報と著者・話者属性情報の関連付けの一例

形態論情報								話者属性情報			
出現書字形	発音形	語彙素読み	語彙素	品詞	活用型	活用形	語種	話者	性別	階層	年齢
あれ	アレ	アレ	彼れ	代名詞			和	鷹子	女	知識層	成人
は	ワ	ハ	は	助詞-係助詞			和	鷹子	女	知識層	成人
何ん	ナン	ナニ	何	代名詞			和	鷹子	女	知識層	成人
だらう	ダロー	ダ	だ	助動詞	助動詞-ダ	意志推量形	和	鷹子	女	知識層	成人
ね	ネ	ネ	ね	助詞-終助詞			和	鷹子	女	知識層	成人

(2) 著者・話者の属性による位相差の計量的分析

(1)の形態論情報と属性情報の関連付けたデータを使用した計量的分析により、会話文の位相差について語種率・品詞率の観点から研究を行った。(1)のデータの作成の完了前は、予備的調査として語種率・品詞率を観点とした書き言葉の研究も行った。

4. 研究成果

(1) 文語文の語種率・品詞率の通時的変化・位相差

近代の非文芸ジャンルの文語文の通時的変化や読者層の性別による位相差を明らかにすることを目的として、「明治・大正編 I 雑誌」を利用して語種率・品詞率を観点とした分析・考察を行い、次の結論を得た。

- ① 名詞率の増加が見られ、文語体の使用の場が評論的・随筆的文章から報道的文章に移行したことが背景として考えられる。
- ② 男性向け雑誌では漢語率の増加が見られ、名詞率の増加の影響だけでなく、語彙自体の漢語率の増加も背景として認められる。
- ③ 女性向け雑誌では男性向け雑誌より漢語率が低い傾向が見られ、女性と和文体との強い関係が認められる。
- ④ 接頭辞率・接尾辞率の増加が見られ、近代語における字音接辞の発展という事象が数値として確認できる。

(2) 口語文の語種率・品詞率の通時的変化・位相差

近代の非文芸ジャンルの口語文の通時的変化や読者層の性別による位相差を明らかにすることを目的として、「明治・大正編 I 雑誌」を利用して語種率・品詞率を観点とした分析・考察を行い、次の結論を得た。

- ①通時的変化として(a)名詞率の増加、(b)漢語率の増加、(c)接尾辞率の増加および接頭辞・接尾辞の漢語率の増加、があげられる。それぞれの背景として、(a)口語文の使用が論説文・随筆文中心であったものが、報道文も含めた実用文全体に広がったこと、(b)名詞率が増加し、また一部の品詞については語彙自体の漢語率が増加したこと、(c)漢語接辞が進展したこと、が指摘できる。
- ② 読者層の性別による位相差として、(a)女性向け雑誌のほうが漢語率が低く和語率が高いこと、(b)女性向け雑誌のほうが名詞率が低く形容詞率・形状詞率・副詞率が高いこと、があげられる。それぞれの背景として、(a)女性は漢語と疎遠な歴史を持ち、また敬体を多用すること、(b)女性向け雑誌に随筆文が多く報道文の少ないこと、が指摘できる。

(3) 会話文の語種率・品詞率の位相差

明治・大正期の話し言葉における語彙の位相差を明らかにすることを目的として、「明治・大正編 I 雑誌」から抽出した雑誌『太陽』の小説・戯曲中の口語体会話文を話者の性別・社会階層による計 4 種の話者属性により分類し、語種率・品詞率の観点から話者属性による位相差を分析・考察した。

分析に利用した各話者属性の口語体会話文の言語量を表 2 に示す。

表 2 会話文の話者属性別言語量

雑誌種類	知識層男性			知識層女性			非知識層男性			非知識層女性		
	話者数	延べ語数		話者数	延べ語数		話者数	延べ語数		話者数	延べ語数	
		自立語・付属語	自立語		自立語・付属語	自立語		自立語・付属語	自立語		自立語・付属語	自立語
太陽 I	54	21,623	12,211	15	7,564	4,204	25	4,703	2,620	24	13,728	7,525
太陽 II	30	19,009	10,194	15	14,333	7,367	24	11,097	5,806	18	10,519	5,294
太陽 III	102	61,201	33,100	41	11,785	6,101	35	10,076	5,450	34	11,767	6,367
太陽 IV	95	48,145	25,472	37	20,020	10,291	58	10,874	5,835	60	9,251	4,915
太陽 V	70	24,113	12,463	26	6,996	3,518	17	2,184	1,172	12	2,205	1,177
全体	351	174,091	93,440	134	60,698	31,481	159	38,934	20,883	148	47,470	25,278

最初に語種率から話者属性による位相差について考察した。主要な 4 語種（漢語・和語・外来語・混種語）について話者単位の語種率の分布を図 1 に、話者属性別の語種率の平均値・標準偏差と平均値の話者属性間の多重比較の結果を表 3 に示す。多重比較は Bonferroni 補正 Welch の t 検定 ($\alpha = .05$) を行った。

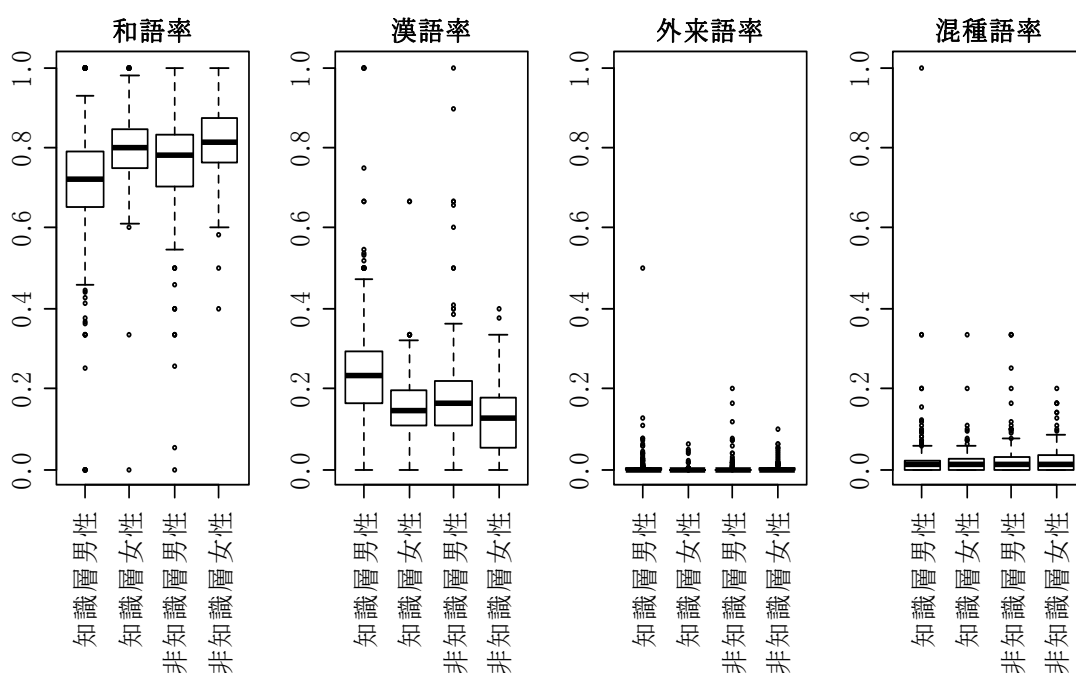


図 1 語種率の分布

表3 語種率の平均値・標準偏差と多重比較の結果

語種率	a. 知識層男性	b. 知識層女性	c. 非知識層男性	d. 非知識層女性	多重比較の結果
和語率	0.708 (0.163)	0.796 (0.122)	0.757 (0.164)	0.820 (0.114)	b, c, d > a d > c
漢語率	0.238 (0.146)	0.156 (0.103)	0.186 (0.155)	0.123 (0.090)	a > b, c > d
外来語率	0.006 (0.030)	0.002 (0.009)	0.007 (0.025)	0.006 (0.015)	n.s.
混種語率	0.021 (0.063)	0.021 (0.038)	0.029 (0.056)	0.026 (0.039)	n.s.

()内は標準偏差

以上の結果から、同じ社会階層間で比較すると女性より男性のほうが漢語率が高く、同じ性別間で比較すると非知識層より知識層のほうが漢語率が高いことが明らかになった。つまり、知識層男性は漢語率が最も高く、逆に非知識層女性は漢語率が最も低く、その中間に知識層女性と非知識層男性が位置する。

次に、品詞率から話者属性による位相差について考察した。「明治・大正編 I 雑誌」の品詞大分類 13 種のすべてについて話者単位の品詞率の分布を図 2 に、話者属性別の各品詞率の平均値・標準偏差と平均値の属性間の多重比較の結果を表 4 に示す。多重比較は Bonferroni 補正 Welch の t 検定 ($\alpha = .05$) を行った。

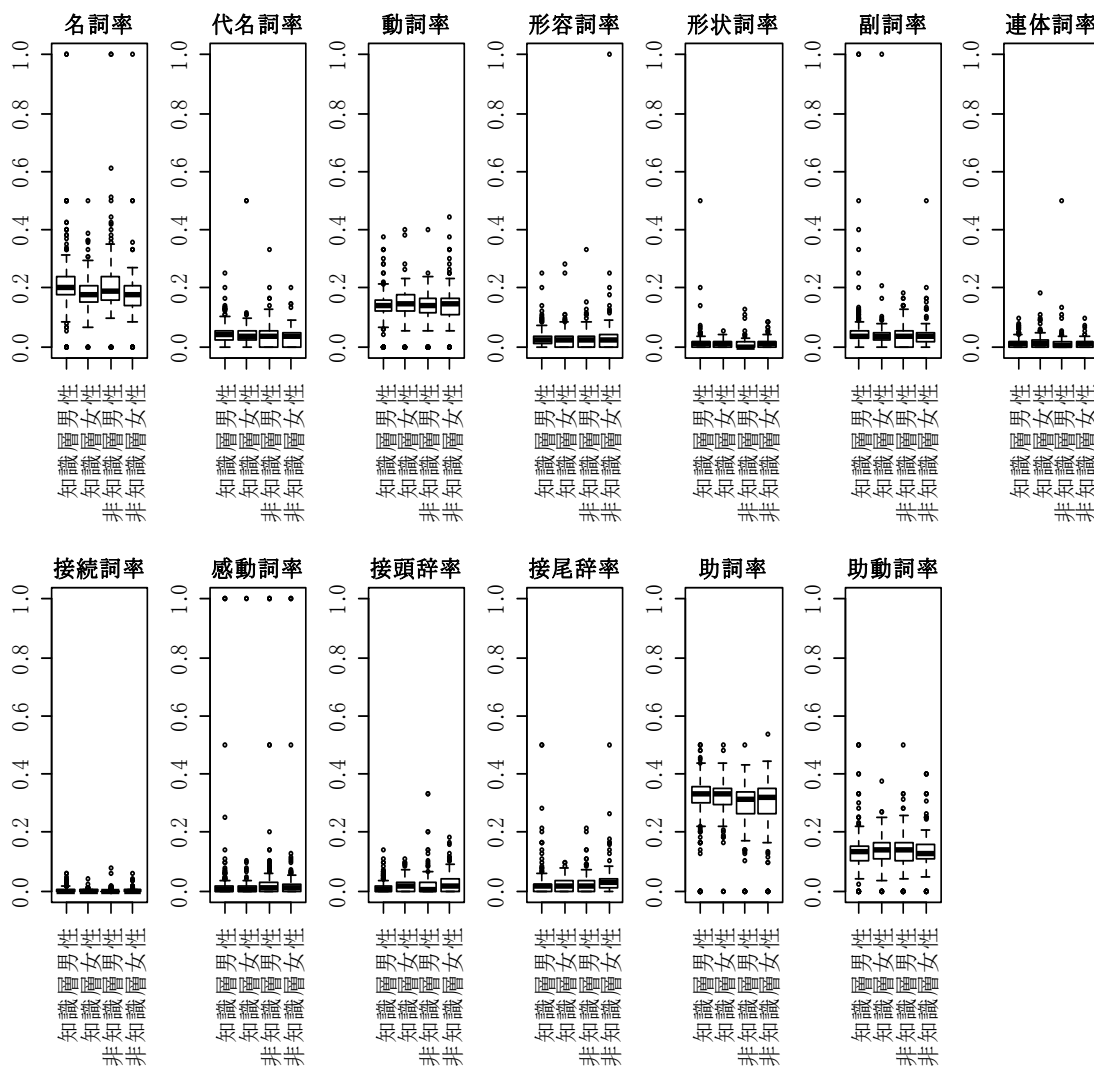


図2 品詞率の分布

表4 品詞率の平均値・標準偏差と多重比較の結果

語種率	a. 知識層男性	b. 知識層女性	c. 非知識層男性	d. 非知識層女性	多重比較の結果
名詞率	0.213 (0.097)	0.185 (0.073)	0.213 (0.127)	0.181 (0.103)	a > b, d
代名詞率	0.053 (0.032)	0.048 (0.062)	0.052 (0.043)	0.046 (0.032)	n.s.
動詞率	0.136 (0.054)	0.143 (0.061)	0.134 (0.062)	0.145 (0.069)	n.s.
形容詞率	0.026 (0.027)	0.029 (0.038)	0.025 (0.037)	0.038 (0.088)	n.s.
形状詞率	0.013 (0.031)	0.011 (0.010)	0.009 (0.017)	0.012 (0.017)	n.s.
副詞率	0.049 (0.085)	0.044 (0.088)	0.039 (0.039)	0.040 (0.051)	n.s.
連体詞率	0.012 (0.014)	0.013 (0.024)	0.013 (0.043)	0.012 (0.015)	n.s.
接続詞率	0.005 (0.008)	0.002 (0.004)	0.002 (0.008)	0.003 (0.008)	a > b, c
感動詞率	0.021 (0.097)	0.026 (0.122)	0.051 (0.165)	0.039 (0.146)	n.s.
接頭辞率	0.012 (0.019)	0.022 (0.024)	0.024 (0.049)	0.028 (0.037)	b, c, d > a
接尾辞率	0.024 (0.048)	0.024 (0.023)	0.026 (0.034)	0.037 (0.055)	d > b
助詞率	0.315 (0.084)	0.320 (0.074)	0.287 (0.098)	0.293 (0.104)	a, b > c
助動詞率	0.134 (0.064)	0.135 (0.062)	0.137 (0.070)	0.137 (0.073)	n.s.

()内は標準偏差

以上の結果から、知識層男性は名詞率・接続詞率が高く、接頭辞率が低い点においてその他の話者属性から卓立し特徴づけられることが明らかになった。名詞率の高さは、知識層男性の会話文の専門性の高さや複雑さを示すものと考えられる。専門性の高低は漢語率の高低にも現れるとされており、知識層男性の会話文の専門性の高さは、知識層男性の漢語率の高さによっても裏付けられる。また、接続詞率の高さからは、知識層男性が前後文脈の関係を明示する論理的な述べ方をすると考えることが可能である。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 0件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 近藤明日子	4. 巻 15
2. 論文標題 語種率・品詞率からみる近代文語文の通時的変化	5. 発行年 2019年
3. 雑誌名 日本語学論集	6. 最初と最後の頁 97 - 83
掲載論文のDOI（デジタルオブジェクト識別子） 10.15083/00076855	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 近藤明日子	4. 巻 21
2. 論文標題 語種率・品詞率から見る明治・大正期の口語体実用文	5. 発行年 2019年
3. 雑誌名 近代語研究	6. 最初と最後の頁 151 - 169
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 近藤明日子
2. 発表標題 近代口語文の語種率・品詞率の通時的変化
3. 学会等名 「通時コーパス」シンポジウム2018
4. 発表年 2018年

1. 発表者名 近藤明日子
2. 発表標題 近代文語文の通時的変化の分析 語種率・品詞率に着目して
3. 学会等名 言語資源活用ワークショップ2016
4. 発表年 2017年

1. 発表者名 近藤明日子
2. 発表標題 明治・大正期の口語体会話文の位相差 語種率・品詞率を観点として
3. 学会等名 言語資源活用ワークショップ2019
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考