

国立国語研究所学術情報リポジトリ

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

メタデータ	言語: Japanese 出版者: 公開日: 2020-02-06 キーワード (Ja): キーワード (En): Corpus of Historical Japanese (CHJ) 作成者: 近藤, 明日子 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002560

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—

近藤 明日子（国立国語研究所コーパス開発センター）[†]

Register Variation in Japanese Dialogue of the Meiji and Taisho Eras: Analysis with Etymological Types Ratios and Part of Speech Ratios

KONDO Asuko (National Institute for Japanese Language and Linguistics)

要旨

明治・大正期の話し言葉における語彙の位相差の実態を明らかにすることを目的として、国立国語研究所（2019）『日本語歴史コーパス 明治・大正編 I 雑誌』（短単位データ 1.2）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。その結果、漢語率は男性 > 女性、知識層 > 非知識層の位相差が認められ、知識層男性が最も高く、非知識層女性が最も低く、その中間に知識層女性・非知識層男性が位置することが明らかになった。また、品詞率では知識層男性が他の話者属性から卓立して特徴的な傾向を有し、それは知識層男性の会話文の専門性の高さ、複雑さ、論理性の高さを示す可能性があることが明らかになった。また、語種率・品詞率の通時的変化を概観した結果、非知識層男性に多くの通時的変化が見られることが明らかになった。

1. はじめに

本研究は、明治・大正期の話し言葉における語彙の位相差の実態の一部を明らかにすることを目的として行うものである。国立国語研究所（2019）『日本語歴史コーパス 明治・大正編 I 雑誌』（短単位データ 1.2）（以下、「明治・大正編 I 雑誌」と呼ぶ）から雑誌『太陽』の小説・戯曲サンプル中の口語体会話文のデータを抽出し、話し言葉を反映した資料として使用する。抽出した会話文は、独自に付与した話者情報に基づき、話者の性別（男性/女性）・社会階層（知識層/非知識層）を掛け合わせた計 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間の位相差を分析・考察する。また、雑誌の刊行年による分類も加え、明治 20 年代から大正 10 年代にかけての会話文の通時的変化についても概観する。明治・大正期の会話文の性別・社会階層による位相差を語種率・品詞率の観点から分析した先行研究は管見の限りなく、本研究により近代日本語の新たな側面が明らかになると考える。

2. 使用する資料

本研究では「明治・大正編 I 雑誌」に収録された雑誌『太陽』明治 28（1895）・明治 34（1901）・明治 42（1909）・大正 6（1917）・大正 14（1925）年刊行の計 60 号分のデータから、小説・戯曲サンプル¹中の口語体会話文を抽出し使用する²。『太陽』は当時を代表す

[†] kondo[at]ninja.ac.jp

¹ サンプルとは、コーパス収録対象として選定されたひとまとまりのテキストの範囲を指す。「明治・大正

る総合雑誌であり、論説文・報道文等の実用文を多く含む一方で、小説・戯曲といった文芸作品も相当程度含む。小説・戯曲の作品やその著者には多様性もあり（表 1 参照）、そこから抽出した会話文は当時の話し言葉の実態を相当程度反映した資料として使用することができると思われる。

「明治・大正編I雑誌」の『太陽』のデータではサンプル単位で日本十進分類法（NDC）に基づくジャンル情報が付与されており、これを利用して小説・戯曲サンプルを抽出した。ただし、その中から以下の①～③に該当するサンプルは除外した³。

- ① 外国語の原著を翻訳したサンプル
- ② 明治期より古い時代を舞台としたサンプル
- ③ 外国を舞台とし、外国人の登場人物による会話文が主に含まれるサンプル

抽出した小説・戯曲サンプルの言語量を表 1 に示す⁴。表中の「雑誌種類」とは『太陽』を刊行年順に I～V に分けて示すものである。また「作品数」とは、一作品が複数号にわたって連載され複数サンプルに分割されている場合をまとめた値である。作品同定に必要なアノテーションは「明治・大正編 I 雑誌」に付与されていないため、独自に判断しカウントした。

表 1 調査対象の小説・戯曲サンプルの言語量

雑誌種類	刊行年	サンプル数	作品数	著者数	サンプル延べ語数	
					自立語・付属語	自立語
太陽 I	1895	17	16	15	115,799	64,884
太陽 II	1901	20	18	9	115,092	61,953
太陽 III	1909	29	27	19	192,707	106,651
太陽 IV	1917	26	26	18	311,243	167,166
太陽 V	1925	40	29	23	141,442	75,491
全体		132	116	71	876,283	476,145

次に、表 1 の小説・戯曲サンプルから口語体会話文を抽出した。「明治・大正編 I 雑誌」の引用・文体に関するアノテーションを利用して、引用種別「会話」・文体「口語」のデータを抽出した。ただし、会話文中のさらなる引用部分については除外した。

抽出した会話文は、話者の性別と社会階層に基づき設定した話者属性により分類した。話者属性設定のためには、まず話者の同定を行う必要がある。「明治・大正編 I 雑誌」の

編 I 雑誌」では、雑誌に掲載された各記事を単位としてサンプルが定められており、サンプル=記事と捉えてほぼ支障はない。

² 以下、「明治・大正編 I 雑誌」の収録資料やアノテーション等の設計の詳細については、「明治・大正編 I 雑誌」のウェブページ (https://pj.ninjal.ac.jp/corpus_center/chj/meiji_taisho.html) を参照のこと。

³ ①は NDC を「912」「913」に限定することで翻訳サンプルを除外することができる。②③に関するアノテーションはコーパスにないので、独自に判断した。

⁴ 以下、言語量は国立国語研究所のデータベースに格納されたデータに拠りカウントした。データベース格納のデータと公開中の「明治・大正編 I 雑誌」では形態論情報等において若干の差異があり、本研究で算出した言語量は公開中のデータのものとは異なる可能性がある。ただし、その差異はごく僅かであり、以降の分析・考察の結果に大きな影響はないと考える。

話者名のアノテーションは、小説・戯曲の作品内で人物の区別ができるように、文脈から区別できる人には異なる話者名を付与し、同一人物には作品内で同一の話者名を付与する設計となっている。これは、一つの作品が複数号にわたって連載され複数サンプルに分割される場合も同様である。そこで、作品名と話者名のアノテーションを組み合わせることで話者の同定を行った。

同定した話者一人一人に対する性別のアノテーションは「明治・大正編 I 雑誌」には付与されていないため、独自に判断した。話者名や作品中の記述から性別判断の材料を得て「男性」「女性」を判断し、判断できない場合は「不明」とした。また、話者の社会階層のアノテーションも「明治・大正編 I 雑誌」には付与されていないため、独自に判断した。中等教育課程以上の学歴を有する者あるいはその学歴に対応する族籍・職業を有する者、およびその家族を「知識層」、それ以外の者を「非知識層」とする 2 階層を設定した。話者名や作品中の記述から社会階層を判断し、判断できない場合は「不明」とした。

このように認定した話者の性別「男性」「女性」2 種と社会階層「知識層」「非知識層」2 種とを掛け合わせた計 4 種を話者属性とし、以下の分析に使用する。表 2 に小説・戯曲サンプルから抽出した話者属性別の口語体会話文の言語量を示す。ただし、「話者数」では延べ語数（自立語・付属語）0 の会話の話者はカウントしていない。

表 2 口語体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									
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

表 2 によれば、会話文の言語量がもっとも多い話者属性は知識層男性である。これは、『太陽』の主要な読者層が知識層男性であり（永嶺 1997、pp.114-125）、同じ属性を持つ登場人物が主要人物となる作品が多く掲載された結果と考えられる。

3. 位相差の分析

3. 1. 語種率の位相差

まず、語種率から話者属性による位相差を見ていく。語種率の算出では付属語（助詞・助動詞）を除いた自立語のみを対象とした。「明治・大正編 I 雑誌」の語種は 8 種類に分類されるが、その中の主要な 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。

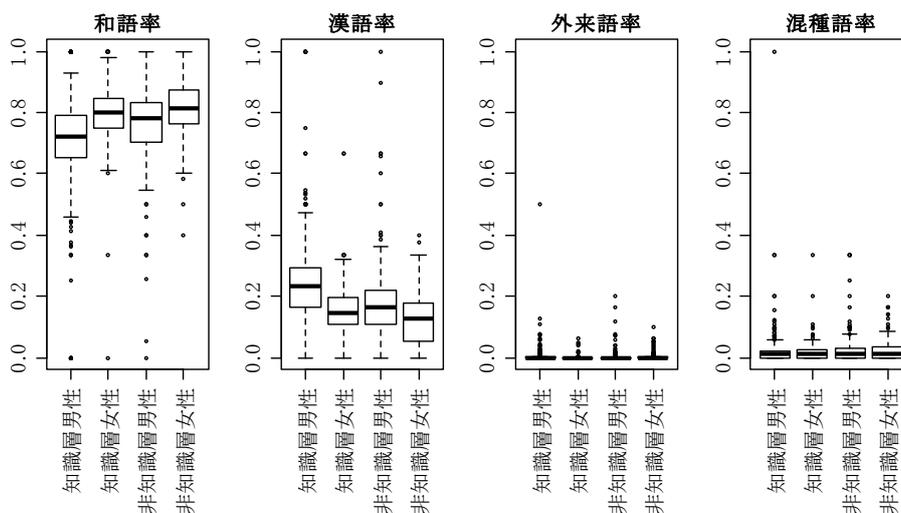


図1 語種率の分布

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

多重比較の結果、和語率では知識層男性・知識層女性間、知識層男性・非知識層男性間、知識層男性・非知識層女性間、非知識層男性・非知識層女性間が有意であった ($t(320.2) = -6.43, p < .001$; $t(302.4) = -3.09, p = .002$; $t(388.5) = -8.71, p < .001$; $t(282.1) = -3.91, p < .001$)。漢語率では知識層男性・知識層女性間、知識層男性・非知識層男性間、知識層男性・非知識層女性間、知識層女性・非知識層女性間、非知識層男性・非知識層女性間が有意であった ($t(339.3) = 6.96, p < .001$; $t(289.5) = 3.60, p < .001$; $t(431.4) = 10.75, p < .001$; $t(265.5) = 2.87, p = .004$; $t(256.9) = 4.40, p < .001$)。外来語率・混種語率ではいずれの話者属性間も非有意であった。

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⁵、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。

⁵ 明治・大正期に近い1950年代の話し言葉を対象とした野元(1959)・中野(1987)においても、女性よりも男性のほうが漢語率の高い傾向が指摘されている。

この漢語率の高低は語彙そのものの漢語率によって主にもたらされていると考えられる。そこで、各品詞の漢語率の位相差を見るために、品詞別の漢語率の分布を図 2 に、品詞別の漢語率の平均値・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。ただし、漢語を全くあるいはほとんど含まない動詞・形容詞・連体詞・接続詞・感動詞・助詞・助動詞は図表中に示していない。

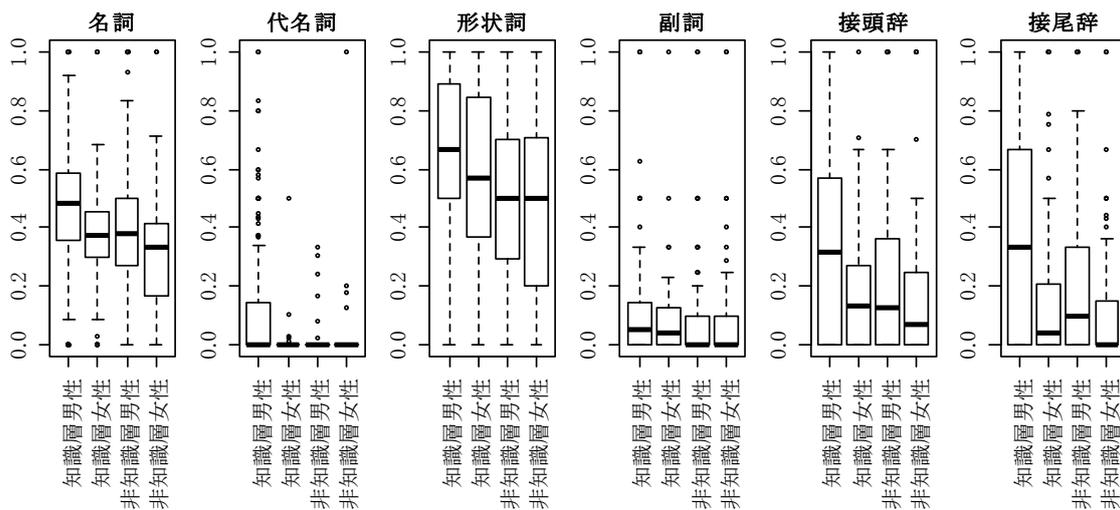


図 2 漢語率の分布

表 4 漢語率の平均値・標準偏差と多重比較の結果

品詞	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	0.464 (0.210)	0.365 (0.180)	0.402 (0.261)	0.295 (0.199)	a > b > d c > d
代名詞	0.053 (0.176)	0.048 (0.049)	0.052 (0.051)	0.046 (0.100)	a > b, c, d
形状詞	0.623 (0.320)	0.567 (0.322)	0.507 (0.348)	0.469 (0.345)	a > d
副詞	0.096 (0.152)	0.078 (0.126)	0.076 (0.163)	0.082 (0.159)	n.s.
接頭辞	0.361 (0.344)	0.175 (0.198)	0.253 (0.323)	0.163 (0.245)	a > b, d
接尾辞	0.398 (0.359)	0.147 (0.244)	0.234 (0.320)	0.109 (0.205)	a > b, c, d c > d

()内は標準偏差

多重比較の結果、名詞では知識層男性・知識層女性間、知識層男性・非知識層女性間、知識層女性・非知識層女性間、非知識層男性・非知識層女性間が有意であった ($t(267.4) = 5.09, p < .001; t(266.6) = 8.32, p < .001; t(264.7) = 3.04, p = .003; t(3.94) = 279.7, p < .000$)。代名詞では知識層男性・知識層女性間、知識層男性・非知識層男性間、知識層男性・非知識層女性間が有意であった ($t(279.7) = 3.94, p < .001; t(379.0) = 8.35, p < .001; t(382.1) = 7.96, p < .000$;

$t(333.3) = 6.17, p < .001$)。形状詞では知識層男性・非知識層女性間が有意であった ($t(144.2) = 3.61, p < .001$)。接頭辞では知識層男性・知識層女性間、知識層男性・非知識層女性間が有意であった ($t(293.4) = 6.05, p < .001; t(255.3) = 5.84, p < .001$)。接尾辞では知識層男性・知識層女性間、知識層男性・非知識層男性間、知識層男性・非知識層女性間、非知識層男性・非知識層女性間が有意であった ($t(259.0) = 7.59, p < .001; t(222.5) = 4.32, p < .001; t(352.7) = 9.86, p < .001; t(179.8) = 3.43, p = .001$)。副詞ではいずれの話者属性間も非有意であった。

このことから、名詞・代名詞・形状詞・接頭辞・接尾辞において他の話者属性よりも知識層男性は漢語率が高く、非知識層女性は漢語率が低く、知識層女性・非知識層男性はその中間に位置するとおおよそ言えるであろう。この品詞別漢語率の位相差の傾向は全体の漢語率の位相差の傾向と一致しており、語彙的な漢語率の位相差が全体の漢語率の位相差に対応していると考えられる。伝統的に漢文や漢文訓読系の文章を通して漢語を受容する機会の多かった知識層男性の漢語率が高いことは自然である。一方、女性は伝統的に漢文や漢文訓読系文章とは疎遠であったが、知識層女性は属する社会階層から漢語に接する機会が非知識層よりは多かったと想像され、そのことが同じ女性でも知識層女性のほうが非知識層女性より漢語率の高いことに繋がっていたものと考えられる。

3. 2. 品詞率の位相差

次に、品詞率から話者属性による位相差を見ていく。「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 3 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 5 に示す。多重比較は Bonferroni 補正 Welch の t 検定 ($\alpha = .05$) を行った。

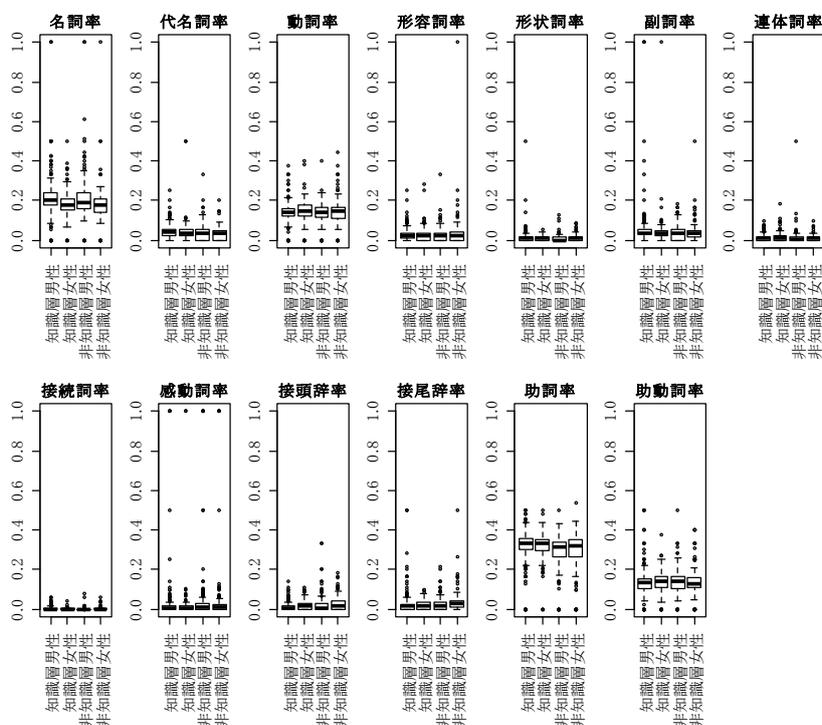


図 3 品詞率の分布

表5 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

多重比較の結果、名詞率では知識層男性・知識層女性間、知識層男性・非知識層女性間が有意であった ($t(318.2) = 3.49, p = .001$; $t(261.3) = 3.18, p = .002$)。接続詞率では知識層男性・非知識層女性間、知識層男性・非知識層男性間が有意であった ($t(428.7) = 4.28, p < .001$; $t(306.0) = 3.09, p = .002$)。接頭辞率では知識層男性・知識層女性間、知識層男性・非知識層男性間、知識層男性・非知識層女性間が有意であった ($t(197.7) = -4.19, p < .001$; $t(179.1) = -2.91, p = .004$; $t(178.8) = -4.82, p < .001$)。接尾辞率では知識層女性・非知識層女性間が有意であった ($t(203.4) = -2.68, p = .008$)。助詞率では知識層男性・非知識層男性間、知識層女性・非知識層男性間が有意であった ($t(267.2) = 3.10, p = .002$; $t(287.3) = 3.24, p = .001$)。代名詞率・動詞率・形容詞率・形状詞率・副詞率・連体詞率・感動詞率・助動詞率ではいずれの話者属性間も非有意であった。

このことから、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。名詞率の高さは談話内容の専門性の高さや複雑さと関係するとされ(小磯・小木曾・小椋・宮内 2009)、知識層男性の会話文がそのような性質を持つこと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(小磯・小木曾・小椋・宮内 2009)、知識層男性の会話文の専門性の高さは、3.1節に見た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。

4. 通時的変化

次に語種率・品詞率の通時的変化について概観する。語種率の通時的変化を図4に、品詞率の通時的変化を図5に示す。図では語種率・品詞率はそれぞれ太陽Iの語種率・品詞率を1とした場合の値によって示した。

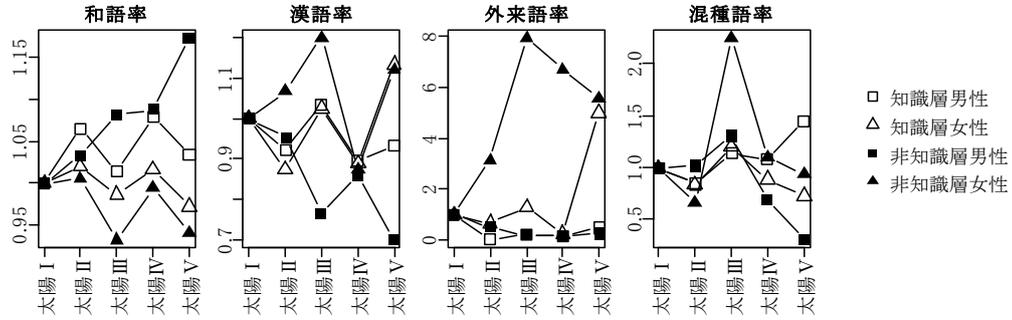


図4 語種率の通時的変化

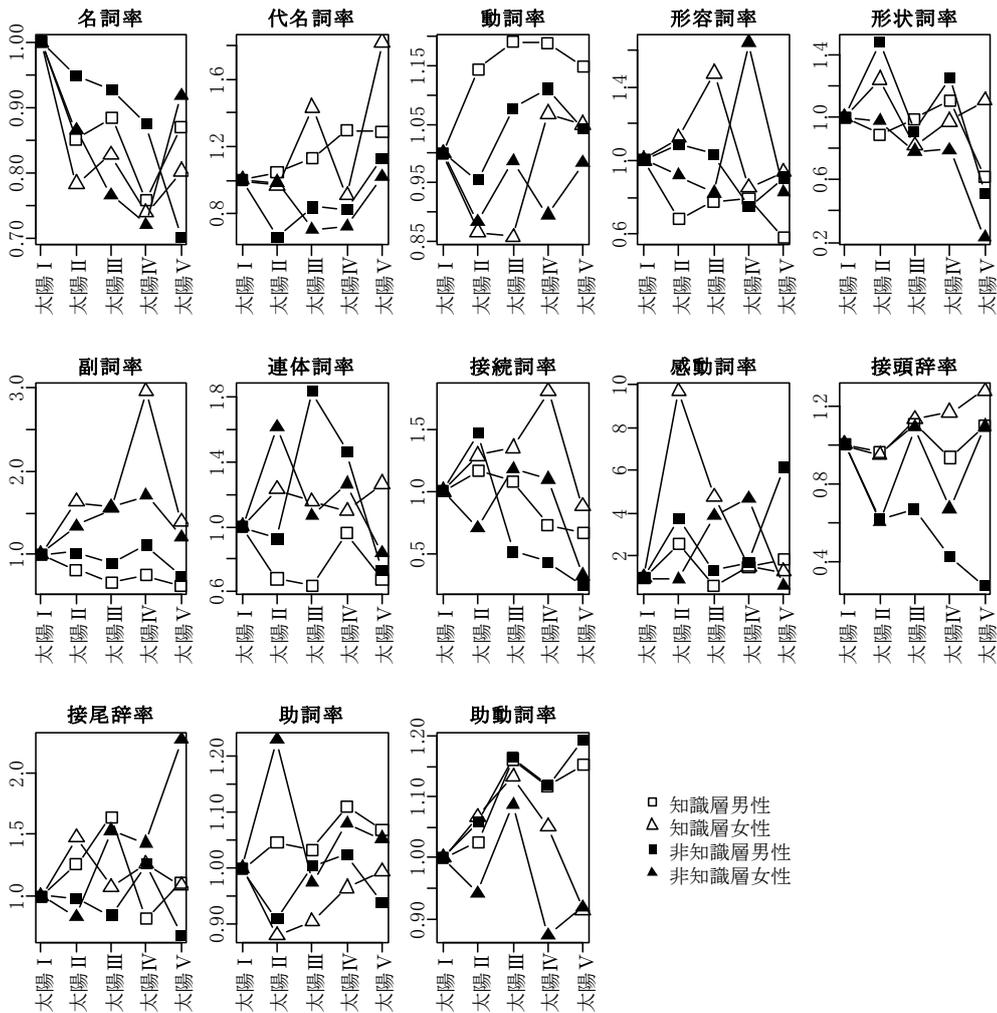


図5 品詞率の通時的変化

図4・図5から、通時的変化が多く読み取れる話者属性は非知識層男性であり、図4から和語率増加と漢語率減少、図5から名詞率減少・接頭辞率減少・助動詞率増加が見てとれる。ここでは通時的変化の指摘にとどめ、変化の生じた要因についての詳しい考察は今後の課題としたい。

5. おわりに

以上、明治・大正期の話し言葉の語彙の位相差の実態を明らかにすべく、「明治・大正編 I 雑誌」の『太陽』小説・戯曲サンプルの口語体会話文を資料として、語種率・品詞率を観点とした話者属性による位相差の分析・考察を行った。その結果、特に知識層男性が他の話者属性から卓立して特徴的な傾向を示し、漢語を多用した専門性の高い、複雑で論理的な性質を持つ知識層男性の会話文の特徴が示唆された。また、通時的変化では非知識層男性に多くの変化が見られることが明らかになった。今後は、文脈に即した考察を加え、話者属性による位相差の実態をより詳細に明らかにしていきたい。

謝辞

本研究は JSPS 科研費 16K02750 「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」および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」の成果の一部である。

参考文献

- 小磯花絵・小木曾智信・小椋秀樹・宮内佐夜香 (2009) 「コーパスに基づく多様なジャンルの文体比較 ―短単位情報に着目して―」『言語処理学会第15回年次大会発表論文集』 pp.594-597.
- 国立国語研究所 (2019) 『日本語歴史コーパス 明治・大正編 I 雑誌』(短単位データ 1.2) https://pj.ninjal.ac.jp/corpus_center/chj/meiji_taisho.html
- 中野洋 (1987) 「話しことばの語種の調査」水谷静夫教授還暦記念会 (編) 『計量国語学と日本語処理 ―理論と応用―』秋山書店, pp.137-150.
- 永嶺重敏 (1997) 『雑誌と読者の近代』日本エディタースクール出版部 (オンデマンド版 2004 に拠る)
- 野元菊雄 (1959) 「話しことばの中での漢語使用」『ことばの研究』1, 国立国語研究所論集 1, 岩波書店, pp.247-260.