

逆接の接続詞から見る明治・大正期の書き言葉の文体の通時的変化

著者	近藤 明日子
雑誌名	言語資源活用ワークショップ発表論文集
巻	5
ページ	238-248
発行年	2020
URL	http://doi.org/10.15084/00003163

逆接の接続詞から見る明治・大正期の書き言葉の文体の通時的変化

近藤 明日子（国立国語研究所コーパス開発センター）[†]

Diachronic Variation in Written Japanese Styles of the Meiji and Taisho Eras: Analysis with Adversative Conjunctions

KONDO Asuko (National Institute for Japanese Language and Linguistics)

要旨

明治・大正期における口語体書き言葉の定着の過程において、文語体書き言葉との関係性を探索するために、『日本語歴史コーパス 明治・大正編 I 雑誌』を資料として、逆接の接続詞の語形を例に、文章種類と語形との対応関係および、文語体・口語体ごとの語形の通時的変化を分析した。その結果、次のことが明らかになった。(1)文章種類と逆接の接続詞の語形との間には強い対応関係がある。(2)文語体と敬体多の口語体では、書き言葉的性質の強い語形の使用が減少し、話し言葉的性質の強い語形の使用が増加する通時的変化が見られる。(3)口語体ではその萌芽期から文語体で広く使用される語形の一部が取り入れられており、文語体の影響を受けながら口語体が成立した様相をうかがうことができる。

1. はじめに

明治・大正期の文章史・文体史は口語体の成立・定着によって特徴付けられる。論説文・報道文等の実用文においては、明治期は文語体が主であり、大正期に入ってから口語体が定着する。その口語体の成立・定着には文語体との関わりがあったことが指摘されているが（森岡（編著）1991:p.43、遠藤 2007）、その具体的実態については明らかになっていないことも多い。そこで本研究では、逆接の接続詞に焦点をあて、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）（以下、「明治・大正編 I 雑誌」と呼ぶ）を用い、明治・大正期の雑誌の書き言葉の文体とそこで使用される逆接の接続詞の語形との対応関係とその通時的変化を計量的に分析し、文語体と口語体との関係性について探索を試みる。

2. 使用するコーパスとテキスト

本研究では、明治・大正期の雑誌のコーパスである「明治・大正編 I 雑誌」に収録された雑誌のうち、『明六雑誌』『東洋学芸雑誌』『国民之友』『太陽』の4誌を利用する¹。その4誌のコーパス収録刊行年と、雑誌名と刊行年による区分（以下、「雑誌種類」と呼ぶ）を表1に示す。

[†] kondo@ninjal.ac.jp

¹ 「明治・大正編 I 雑誌」には、他に『女学雑誌』『女学世界』『婦人倶楽部』も収録されているが、この3誌は女性知識層を主な読者層とする雑誌であり、男性知識層を主な読者層とする『明六雑誌』『東洋学芸雑誌』『国民之友』『太陽』とは位相差があると予想され、本研究では研究対象外とした。

表1 本研究で使用する雑誌種類

雑誌種類	雑誌名	刊行年
明六	明六雑誌	1874-75
東洋	東洋学芸雑誌	1881-82
国民	国民之友	1887-88
太陽Ⅰ	太陽	1895
太陽Ⅱ	太陽	1901
太陽Ⅲ	太陽	1909
太陽Ⅳ	太陽	1917
太陽Ⅴ	太陽	1925

最初に、これらの雑誌種類のコーパスデータから、書き言葉を代表するテキストとして非文芸ジャンルのサンプルを抽出した²。ただし、①明治期より以前に書かれたサンプル³、②外国語の原典を翻訳したサンプル⁴、③雑誌1冊ごとの構造要素をまとめたサンプル⁵は、当時の書き言葉を代表するテキストからは外れると考え、抽出対象外とした。さらに、サンプル中に引用された会話文や別資料の文章も当時の書き言葉を代表するテキストから外れると考え、それらを除いた地の文のみを抽出した⁶。このようにして抽出した非文芸ジャンルサンプルの地の文からさらに文体が「文語」「口語」のテキストをそれぞれ抽出した⁷。以下、この2種のテキストをそれぞれ「文語体地の文」「口語体地の文」と呼ぶ。

加えて、口語体地の文については下位分類として「口語文体」を設定した。口語体地の文内で使用される敬体の文末辞「です」「ます」「であります」「ござる」「ござり(い)ます」と常体の文末辞「である」「だ(意志推量形・終止形のみ)」の頻度をサンプル単位でカウントし、文末辞の頻度合計に対する敬体の文末辞の頻度合計の割合が0.5以上のサンプルの口語文体を「敬体多」、0.5未満のサンプルの口語文体を「常体多」とした。

次に、書き言葉の対照資料として、当時の日常会話を反映していると考えられる小説・戯曲サンプル⁸の口語体会話文⁹を抽出した。ただし、①明治期より以前に書かれたサンプル¹⁰、

² テキストをコーパスに収録する際にテキストを一定の範囲で分割する必要があり、その各範囲をサンプルと呼ぶ。「明治・大正編Ⅰ雑誌」のサンプルは記事を単位としており、1サンプル=1記事と考えてほぼ支障はない。コーパスではサンプル単位で「文芸」「非文芸」の2種に分類するジャンル情報が付与されており、このジャンル情報が「非文芸」のサンプルを抽出した。以下、「明治・大正編Ⅰ雑誌」の設計の詳細については、間淵・近藤・服部・南雲(2019)を参照のこと。

³ コーパスに付与されているサンプル著者の生没年の情報から判断した。なお、著者の没年は「明治・大正編Ⅰ雑誌」では公開されていない情報であり、国立国語研究所にある非公開データを利用した。

⁴ 翻訳のサンプルはコーパスの著者情報において訳者名が表示されることから判断した。

⁵ 該当サンプルはサンプルIDの末尾が「000」であることから判断した。

⁶ コーパスでは「地の文」「会話」「典拠」等の本文種別情報が付与されており、これを利用して「地の文」を抽出した。

⁷ コーパスにおいて各短単位に付与されている文体情報から判断した。

⁸ 小説・戯曲サンプルであることの判断は、まずコーパスに付与されているサンプルのジャンル情報が「文芸」であることに拠った。ただし、「文芸」には小説・戯曲以外に韻文も含まれる。そこで、ジャンル情報にさらにNDC(日本十進分類法)の第1次区分から第3次区分までを表す3桁の数字が付与されている『太陽』については、NDC番号が「912(文学/日本文学/戯曲)」「913(文学/日本文学/小説・物語)」であることをもって小説・戯曲であることを判断した。『明六雑誌』『東洋学芸雑誌』『国民之友』はジャンル情報にNDC番号が付与されていないため、「文芸」のサンプルの内容から独自に小説・戯曲であることを判断した。

⁹ コーパスの本文種別情報が「会話」、文体情報が「口語」であることを利用して判断した。

¹⁰ 注3参照。

②外国語の原典を翻訳したサンプル¹¹、③明治期より以前の時代を舞台とするサンプル¹²、④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¹³、の口語体会話文は、当時の日常会話を代表するテキストからは外れると考え、抽出対象外とした。以下、このテキストを「会話文」と呼ぶ。

次の表 2 に、3 種のテキスト「文語体地の文」「口語体地の文」「会話文」（以下、このテキストの分類を「文章種類」と呼ぶ）の延べ語数¹⁴（上段）と該当文章種類のテキストが出現するサンプル数（下段）を示す。

表 2 テキストの延べ語数・サンプル数

文章種類	明六	東洋	国民	太陽 I	太陽 II	太陽 III	太陽 IV	太陽 V	計
文語体地の文	143,105	164,869	788,411	1,285,604	1,225,779	573,535	209,910	4,135	4,395,348
	133	113	1,171	585	399	255	87	15	2,758
口語体地の文	8,164	0	6,891	93,344	288,728	795,462	1,084,883	1,451,860	3,729,332
	4	0	4	33	148	367	366	882	1,804
敬体多	8,164	0	6,877	73,345	94,087	27,534	45,195	51,247	306,449
	4	0	3	20	47	11	18	43	146
常体多	0	0	14	19,999	194,641	767,928	1,039,688	1,400,613	3,422,883
	0	0	1	13	101	356	348	843	1,662
会話文	0	0	0	36,731	56,410	101,331	95,553	39,920	329,945
	0	0	0	15	15	28	26	40	124

表 2 から文体の延べ語数比率の通時的変化を概観しておく。まず、文語体地の文と口語体地の文との比率を見ると、「太陽 I」の時期までは文語体が主たる文体であり口語体はごく僅かだったものが、「太陽 II」以降口語体の比率が増加し、「太陽 IV」の時期に至り口語体が主たる文体となる通時的変化が見出される。次に、敬体多と常体多との比率を見ると、口語体地の文の内部では「太陽 I」の時期までは敬体多が主たる口語文体であったものが、それ以降常体多の比率が増加し、「太陽 III」の時期に至り常体多が主たる口語文体となる通時的変化が見出される。当初の口語体が敬体の文末辞を多用するのは、講義や演説の筆記あるいはそれを模して書かれた文章であったためであり（森岡 1991: pp.28-32）、講者・演者が聴衆を意識して使用した文末辞をそのまま反映したものである。敬体の文末辞の多用は当時の口語体の話し言葉的性質の強さを示すものと言える。それが、次第に常体の文末辞を多用する文体へと移行することは、口語体が話し言葉的性質を脱し書き言葉的性質を強めていく過程を象徴するものと言える。こうした過程を経て、口語体は文語体に代わって書き言葉の主たる文体の位置を獲得していくのである。

3. 短単位 n-gram による逆接の接続詞の抽出

続いて、2 節で抽出した 3 種の文章種類のテキストから、形態論情報を利用して逆接の接続詞を抽出した。「明治・大正編 I 雑誌」には短単位の形態論情報が付与されているが、接続詞は複数短単位からなるものも多く、1 短単位に注目するのみでは網羅的な抽出は難しい。そこで、テキストから短単位による n-gram（以下、「短単位 n-gram」と呼ぶ）を作成し、

¹¹ 注 4 参照。

¹² コーパスには作品の舞台に関する情報は付与されていないため、サンプルの内容から独自に判断した。

¹³ コーパスには登場人物の出身に関する情報は付与されていないため、サンプルの内容から独自に判断した。

¹⁴ 延べ語数はコーパスの語の単位である「短単位」に基づき、1 短単位=1 語として、記号類・未知語類を除いてカウントした。

そこから接続詞と考えられるものを独自に判断し抽出する手法で、網羅的な抽出を目指した。日本語のコーパスから作成した短単位 n-gram を用いて特定の意味・機能を有する短単位連続を抽出する先行研究に、近藤 (2012)、李・佐々木 (2015)、渡辺 (2018) 等があり、接続詞の抽出にも短単位 n-gram は有効な手段となると考える。

まず、テキストから 1~8gram の短単位 n-gram を作成した。ただし、接続詞ではない可能性の高い短単位 n-gram を排除し、接続詞を効率的に抽出するため、文頭¹⁵の短単位 n-gram を対象とし、かつ記号類・未知語類・固有名詞を含む短単位 n-gram は除いた。

次に、作成した短単位 n-gram から、独自の判断により接続詞を選定した。1gram からは、短単位の品詞が「接続詞」のものを接続詞として選定した。2~8gram からは、『日本国語大辞典 第 2 版』(小学館)等の辞典類や青木 (1973) を参照しつつ、そこには載っていない語形も含め、独自の判断で接続詞を選定した。ただし、当時の書き言葉である程度一般的に使用される語形に限定するため、該当接続詞の語形の使用されるサンプル数が、文語体地の文では 28 以上 (文語体地の文の全サンプル数 2,758 (表 2 参照) の 1%以上)、あるいは、口語体地の文では 19 以上 (口語体地の文の全サンプル数 1,804 (表 2 参照) の 1%以上)であることを条件に選定した¹⁶。そして、選定した接続詞を市川 (1976) の 7 分類により分類し、逆接の接続詞¹⁷に分類される語形を研究対象とした¹⁸。

加えて、書き言葉で使用される語形と比較対照するために、当時の日常会話で広く使用された接続詞として、会話文での使用サンプル数が上位 10 位までの逆接の接続詞も選定し、研究対象とした。

4. 逆接の接続詞の語形と文章種類との対応関係

次の表 3 に、3 節で選定した逆接の接続詞を五十音順に掲出し、その文章種類別の粗頻度(a)・使用サンプル数(b)・使用サンプル率(c)をあげる (空欄は値 0 を示す)。使用サンプル率とは各文章種類の全サンプル数 (表 2 参照) に対する逆接の接続詞の使用サンプル数の割合を示す値である。使用サンプル率の右列に、文語体地の文で使用サンプル数が 28 以上の語形に*、口語体地の文で使用サンプル数が 19 以上の語形に‡、会話文で使用サンプル数上位 10 語形に‡‡を示した。

¹⁵ コーパスに付与されている文境界情報を利用し、文頭を認定した。ただし、文頭に空白や括弧等の記号類がある場合は、その記号類を除いて文頭を認定した。

¹⁶ なお、接続詞として認定した短単位 n-gram の中に、接続詞として認定した他の短単位 n-gram が重複している場合 (例えば、「しかし」の中に「しかしながら」の重複する場合) があるが、使用サンプル数はその重複は除外してカウントし直した。重複を認めカウントし直した語形の詳細は注 19 を参照のこと。また、接続詞「されども」として認定された n-gram の用例中、書字形 (表記) が「然れども」「然ども」「雖然」のものは「しかれども」の用例としてカウントし直した。

¹⁷ 市川 (1976) では、接続詞を順接・逆接・添加・対比・転換・同列・補足の 7 種に分類し、逆接の接続詞は「前の内容に反する内容を導く」用法を持つものと定義する。

¹⁸ よって、研究対象として抽出した用例の中には、同語形別用法のような実際には逆接の接続詞ではないものも含まれており、逆に、実際には逆接の接続詞であるにもかかわらず、形態論情報の誤りや文頭以外の位置に出現することなどにより、研究対象から漏れてしまった用例もある。このように研究対象には多少の遺漏があるが、全体の傾向を考察する上では大きな問題にはならないと考える。

表3 逆接の接続詞とその文章種類別の粗頻度(a)・使用サンプル数(b)・使用サンプル率¹⁹

語形	文語体地の文			口語体地の文			会話文		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
が	1	1	0.0%	440	250	13.9%*	24	12	9.7%**
けれど				63	40	2.2%*	12	10	8.1%
けれども				603	282	15.6%**	48	23	18.5%**
さりながら	48	22	0.8%	34	32	1.8%*			
されど	723	225	8.2%*	89	48	2.7%*			
されども	87	49	1.8%*	5	5	0.3%			
しかし	189	143	5.2%*	2287	825	45.7%**	332	55	44.4%**
しかしながら	151	112	4.1%*	703	295	16.4%**			
しかりといえども	391	259	9.4%*	3	3	0.2%			
しかるに	1452	725	26.3%*	1381	567	31.4%**	3	3	2.4%
しかれども	3794	1255	45.5%*	132	78	4.3%**	1	1	0.8%
それでも				71	58	3.2%**	31	25	20.2%**
だが				114	74	4.1%**	62	31	25.0%**
ただ							19	11	8.9%**
だって							80	43	34.7%**
ですが							14	12	9.7%**
でも				37	26	1.4%**	117	47	37.9%**
ところが				654	324	18.0%**	26	18	14.5%**

表3に示したように、研究対象とする逆接の接続詞は、文語体地の文では7語形（*を付した語形）、口語体地の文では13語形（*を付した語形）、会話文では10語形（**を付した語形）ある。これらを合わせた18語形（異なり）について、文章種類との対応関係を見るために、表3から得られる18語形×3文章種類の使用サンプル率のクロス表を使用し、コレスポネンス分析を行った。その第1次元（寄与率72.7%）の文章種類・語形それぞれのスコアを表4・表5に示す。

表4 コレスポネンス分析による文章種類の第1次元スコア

文章種類	スコア
文語体地の文	1.812
口語体地の文	0.079
会話文	-0.835

¹⁹ カウントの重複を避けるためカウントをし直した語形は「けれど」（「けれども」を除外）、「しかし」（「しかしながら」を除外）、「ただ」（「ただども」を除外）である。

表 5 コレスポネンス分析による語形の第 1 次元スコア

語形	スコア	グループ
しかりといえども	2.145	A
しかれども	1.952	
されども	1.899	
されど	1.668	
しかるに	0.962	B
さりながら	0.742	
しかしながら	0.510	
しかし	-0.304	C
が	-0.354	
ところが	-0.397	
けれども	-0.502	
けれど	-0.768	D
だが	-0.851	
それでも	-0.854	
でも	-0.965	
だけど	-1.006	
だって	-1.006	
ですが	-1.006	

表 4 の文章種類のスコアの高低と表 5 の語形のスコアの高低は対応関係にあり、スコアの低い文章種類にはスコアの低い語形が使用されやすく、スコアの低い文章種類にはスコアの低い語形が使用されやすいことを示している。3 種の文章種類の中で、文語体地の文は最も書き言葉的性質が強く、逆に会話文は最も話し言葉的性質が強く、口語体地の文はその中間の性質を持つと考えると、コレスポネンス分析による語形の第 1 次元スコアは、各語形の書き言葉的性質あるいは話し言葉的性質の強弱を示す指標となっていることになる。そこで、語形のスコアおよび使用サンプル率の比率から勘案し、18 語形を「A 主に文語体地の文で使用される」4 語形、「B 口語体地の文を中心に文語体地の文でも使用される」3 語形、「C 口語体地の文を中心に会話文でも使用される」4 語形、「D 主に会話文で使用される」7 語形の 4 グループに分け、そのグループ分けを表 5 に示した。この A~D グループの書き言葉的性質の強弱は A>B>C>D の順となり、話し言葉的性質の強弱は D>C>B>A の順となる。この A~D のグループ分けと語形の第 1 次元スコアを 5 節の通時的変化の分析・考察での観点として使用する。

5. 逆接の接続詞の語形の通時的変化

次に、地の文における逆接の接続詞の使用語形の通時的変化の分析・考察を行う。4 節で明らかになったように、同じ非文芸ジャンルの地の文であっても文語体と口語体では使用される語形が異なるため、通時的変化も文体ごとに見ていく。

5. 1. 文語体地の文における通時的変化

最初に、文語体地の文での使用語形の通時的変化を見ていく。表 3 で*を付した 7 語形のうち、1 つ以上の雑誌種類（表 1 参照）で使用サンプル率が 20%以上の 4 語形を文語体地の文での主要な語形と見なした。次の表 6 に、その主要 4 語形の使用サンプル数（上段）と使用サンプル率（下段）を雑誌種類ごとに示す。逆接の接続詞が使用されない「太陽 V」は

表中に示していない。なお、ここでいう「使用サンプル率」は、表3で*を付した7語形が使用される全サンプル数(表6の最下行「全体」の値)に占める割合を指して言う(表7も同様)²⁰。

表6 文語体地の文における主要語形の使用サンプル数(上段)・使用サンプル率(下段)

グループ	語形	スコア	明六	東洋	国民	太陽I	太陽II	太陽III	太陽IV
A	しかりといえども	2.145	17 14.4%	13 14.3%	128 20.5%	53 14.0%	39 13.5%	6 3.9%	3 11.5%
	しかれども	1.952	94 79.7%	83 91.2%	449 71.8%	307 81.0%	212 73.6%	95 62.5%	15 57.7%
	されど	1.668	1 0.8%	3 3.3%	24 3.8%	63 16.6%	93 32.3%	37 24.3%	4 15.4%
B	しかるに	0.962	67 56.8%	62 68.1%	179 28.6%	162 42.7%	145 50.3%	86 56.6%	24 92.3%
	全体		118 100.0%	91 100.0%	625 100.0%	379 100.0%	288 100.0%	152 100.0%	26 100.0%

表6では通時的変化の概略をつかむため、各文章種類での使用サンプル率上位3語形のセルに色を付けた。ここからまず、すべての文章種類で1位・2位に位置する語形は「しかれども」「しかるに」であり、当時の文語体地の文で広く使用された語形であることが分かる。また、「しかりといえども」は刊行年の古い雑誌種類に広く使用されるのに対し、「されど」は刊行年の新しい雑誌種類に広く使用されることも分かる。

このような語形と雑誌種類との対応関係をより詳しく見るために、表6から得られる4語形×7雑誌種類の使用サンプル率のクロス表を用いてコレスポネンス分析を行った。分析結果から作成した第1次元(寄与率61.2%)と第2次元(寄与率33.9%)のスコアの散布図を文章種類・語形ごとに図1・図2に示す。

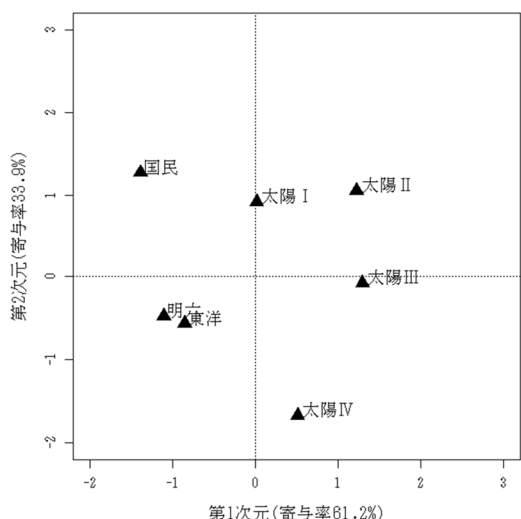


図1 文章種類のスコア散布図

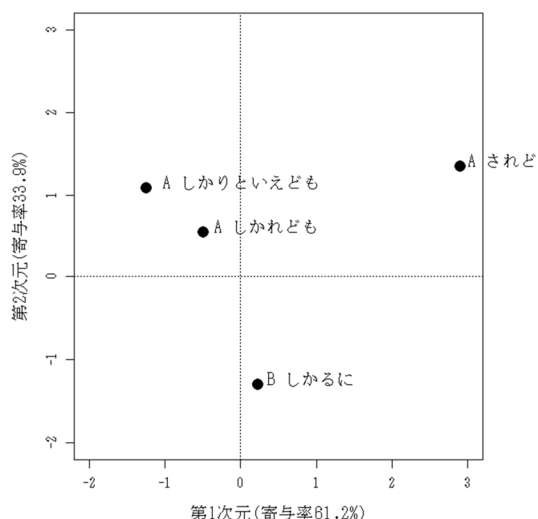


図2 語形のスコア散布図

²⁰ ここで明らかにしたいことは、逆接の接続詞内部での勢力関係の通時的変化であるから、調査対象の全サンプル数に占める比率ではなく、逆接の接続詞の使用されるサンプル数に占める比率に基づき分析を行う。

図 1 において語形の使用パターンの近い雑誌種類は近くに布置され、使用パターンの異なる雑誌種類は遠くに布置される。また、図 2 において、各文章種類における使用パターンの近い語形同士は近くに布置され、使用パターンの異なる語形同士は遠くに布置され、図 1 における文章種類の分布と図 2 における語形の分布が対応する。

図 1 を見ると、雑誌種類は第 3 象限→第 2 象限→第 1 象限→第 4 象限の順に、刊行年の古いものから新しいものへと布置されている。つまり、「明六」から「太陽Ⅳ」にかけて語形の使用パターンに通時的変化があることが示されている。さらに図 1 と図 2 の対応から、各雑誌種類を特徴づける語形が分かる。刊行年の古い「明六」「東洋」「国民」「太陽Ⅰ」は A グループの「しかりといえども」「しかれども」という書き言葉的性質の強い語形によって特徴づけられる。これに対し、刊行年の新しい「太陽Ⅱ」「太陽Ⅲ」は A グループのなかでも書き言葉的性質の弱い「されど」によって特徴づけられる。そして、最も刊行年の新しい「太陽Ⅳ」は B グループの「しかるに」によって特徴づけられる。

このように文語体地の文において、逆接の接続詞の語形は、書き言葉的性質の強い A グループの語形から書き言葉的性質の弱い B グループの語形へと推移していく通時的変化があることが明らかになった。

5. 2. 口語体地の文における通時的変化

次に、口語体地の文での使用語形の通時的変化を見ていく。表 2 で見たように、口語体はその内部で敬体多から常体多へと口語文体が移行する変化があるため、口語文体も分析の観点として使用する。

表 3 で \ddagger を付した 13 語形のうち、口語体地の文全体で 1 つ以上の文章種類での使用サンプル率が 20%以上の 7 語形を口語体地の文における主要な語形と見なした。次の表 7 にその 7 語形の口語体地の文での使用サンプル数（上段）と使用サンプル率（下段）を口語文体・雑誌種類ごとに示した。口語体地の文のテキストのない「東洋」と常体多の「明六」、逆接の接続詞が使用されない常体多の「国民」は表中に示していない。

表 7 口語体地の文における主要語形の使用サンプル数（上段）・使用サンプル率（下段）

グループ	語形	スコア	敬体多					常体多							
			明六	国民	太陽Ⅰ	太陽Ⅱ	太陽Ⅲ	太陽Ⅳ	太陽Ⅴ	太陽Ⅰ	太陽Ⅱ	太陽Ⅲ	太陽Ⅳ	太陽Ⅴ	
A	しかれども	1.952			1 5.6%	12 27.3%						11 15.7%	18 6.0%	24 9.2%	12 2.6%
B	しかるに	0.962	1 33.3%	1 100.0%	12 66.7%	24 54.5%	1 11.1%	5 27.8%	6 37.5%		2 18.2%	40 57.1%	152 50.5%	170 64.9%	153 32.6%
	しかしながら	0.510			8 44.4%	10 22.7%	3 33.3%	1 5.6%	3 18.8%		1 9.1%	11 15.7%	64 21.3%	96 36.6%	98 20.9%
	しかし	-0.304	2 66.7%		11 61.1%	23 52.3%	6 66.7%	16 88.9%	12 75.0%		8 72.7%	46 65.7%	213 70.8%	188 71.8%	300 64.0%
C	が	-0.354			2 11.1%	3 6.8%		1 5.6%	2 12.5%		4 36.4%	7 10.0%	48 15.9%	42 16.0%	141 30.1%
	ところが	-0.397		1 100.0%	4 22.2%	6 13.6%	3 33.3%	4 22.2%	4 25.0%		5 45.5%	19 27.1%	74 24.6%	54 20.6%	150 32.0%
	けれども	-0.502		1 100.0%	3 16.7%	7 15.9%	3 33.3%	3 16.7%	8 50.0%		2 18.2%	14 20.0%	99 32.9%	34 13.0%	108 23.0%
	全体		3 100.0%	1 100.0%	18 100.0%	44 100.0%	9 100.0%	18 100.0%	16 100.0%		11 100.0%	70 100.0%	301 100.0%	262 100.0%	469 100.0%

表7では通時的变化の概略をつかむため、各口語文体・文章種類（サンプル数の少ない敬体多の「明六」「国民」「太陽Ⅲ」、常体多の「太陽Ⅰ」を除く）での使用サンプル率上位3語形のセルに色を付けた。そこから、口語体地の文ではすべての口語文体・文章種類で「しかるに」と「しかし」が広く使用されることがわかる。また、敬体多においては刊行年の古い「太陽Ⅰ」「太陽Ⅱ」では「しかれども」「しかしながら」が広く使用されるのに対し、刊行年の新しい「太陽Ⅳ」「太陽Ⅴ」では「ところが」「けれども」が広く使用されることも分かる。

このような語形と雑誌種類との対応関係をより詳しく見るために、表7から得られる7語形×8口語文体・雑誌種類（敬体多の「明六」「国民」「太陽Ⅲ」、常体多の「太陽Ⅰ」を除く）の使用サンプル率のクロス表を用いてコレスポネンス分析を行った。分析結果から作成した第1次元（寄与率52.4%）と第2次元（寄与率21.2%）のスコアの散布図を文章種類・語形ごとに図3・図4に示す。

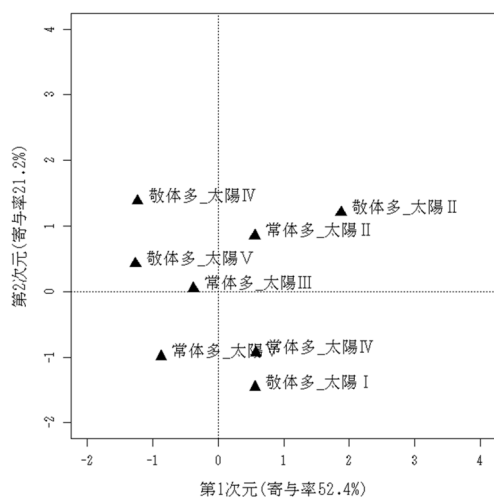


図3 口語文体・文章種類のスコア散布図

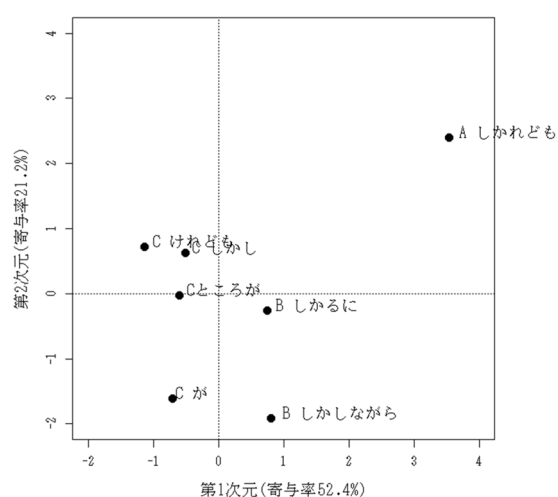


図4 語形のスコア散布図

図3においてまず敬体多の雑誌種類を見ると、「太陽Ⅰ」「太陽Ⅱ」が第1次元の正の値をとり、「太陽Ⅲ」「太陽Ⅳ」が第1次元の負の値をとることが分かる。この第1次元は図4では、より書き言葉的性質の強いA・Bグループの語形が正の値をとり、より話し言葉的性質の強いCグループの語形が負の値をとる。このことから、敬体多の文章において、逆接の接続詞の語形は書き言葉的性質の強いものから話し言葉的性質の強いものへ移行するという通時的变化が見られることが分かる。一方、常体多の雑誌種類については、図3・図4からは通時的变化は特に見出されなかった。

ここで注目したいのは、刊行年の古い「太陽Ⅰ」「太陽Ⅱ」の口語体では当初から「しかし」のようなCグループの語形だけではなく、Bグループ「しかるに」のような、文語体でも広く使用される書き言葉的性質の強い語形を使用している点である。2節において、当時の口語体は講義・演説といった話し言葉に基づいた文体であり、その意味で話し言葉的性質が強いと述べたが、逆接の接続詞に関しては当時の日常会話とは大きく異なる語形が使用されていたことになる。このことから、書き言葉として萌芽期にある当時の口語体は、話し言葉を土台としつつも、書き言葉的性質の強い語形を一部取り込んだ文体であったと言え

る。

そして、常体多の口語体が書き言葉の標準として確立していく「太陽Ⅲ」～「太陽Ⅴ」の時期、常体多の口語体では B グループの語形が引き続き使用された。常体多の口語体を書き言葉に相応な文体たらしめるため、書き言葉的性質の強い語形を使用し続けることが必要であったものと考えられる。反対に、敬体多の口語体では B グループの語形の使用が減少し、C グループの語形の使用が広がった。これは書き言葉として傍流となった敬体多の口語体において、書き言葉的性質を醸成する語形を維持する必要がなくなったことを示していると考えられる。

6. おわりに

以上、「明治・大正編Ⅰ雑誌」を資料として、逆接の接続詞の語形と文章種類との対応関係および文体ごとの使用語形の通時的変化やその関係について分析・考察した。まず、文章種類と語形との間には強い対応関係があることが明らかになった。また、文語体や敬体多の口語体では、書き言葉的性質の強い語形の使用が減少し、より話し言葉的性質の強い語形の使用が増加するという通時的変化が見られた。また、口語体はその萌芽期から文語体で広く使用されていた語形を取り入れていた実態も明らかになった。このように、口語体書き言葉はその発生・定着の過程において、既に書き言葉として確立していた文語体から語彙の面で影響を受けつつ、新たな書き言葉としての地位を獲得していったものと考えられる。

今後は、他の言語項目についても研究を蓄積し、口語体書き言葉の発生・定着の過程における文語体書き言葉の関わりの全体像を明らかにしていきたい。

謝 辞

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」の研究成果の一部である。

文 献

- 青木玲子（1973）「接続詞および接続詞的語彙一覧」『接続詞・感動詞』品詞別日本文法講座 6, 明治書院, pp.210-253
- 市川孝（1976）「副用語」『文法Ⅰ』岩波講座 日本語 6, 岩波書店, pp.219-258
- 遠藤好英（2007）「普通文」飛田良文（編集主幹）, 遠藤好英・加藤正信・佐藤武義・蜂谷清人・前田富祺（編）『日本語学研究事典』明治書院, p.502
- 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#zasshi
- 小林雄一郎（2010）「コレスポンデンス分析：データ間の構造を整理する」石川信一郎・前田忠彦・山崎誠（編）『言語研究のための統計入門』pp.245-264
- 小林雄一郎（2017）『Rによるやさしいテキストマイニング』オーム社
- 近藤泰弘（2012）「BCCWJ 複合辞リストについて」『青山語文』42, pp.10-15
- 間淵洋子・近藤明日子・服部紀子・南雲千香子（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） テキストの凡例と『中納言』表示項目について」https://pj.ninjal.ac.jp/corpus_center/chj/doc/abstract-zasshi-201903.pdf
- 森岡健二（編著）（1991）『近代語の成立—文体編—』明治書院
- 渡辺由貴（2018）「N-gram を利用した中世語の複合辞・連語の検討」『日本語学会 2018 年

度春季大会予稿集』 pp.182-185

李在鎬・佐々木馨（2015）「教科書コーパスを利用した難易度別コロケーション辞書の提案」『第8回コーパス日本語学ワークショップ予稿集』国立国語研究所言語資源研究系・コーパス開発センター, pp.273-278