

『近代女性雑誌コーパス』の小説会話部分に現れる 一・二人称代名詞の計量的分析

近藤 明日子 (国立国語研究所コーパス開発センター) †

First- and Second-Person Pronouns in *Modern Women's Magazines Corpus*: A Quantitative Analysis Focusing on Conversational Sentences in Novels

KONDO Asuko (National Institute for Japanese Language and Linguistics)

1. はじめに

近代日本語の一・二人称代名詞に関する研究はこれまで主に、小説・戯曲の会話部分、落語速記、口語文典などの話し言葉的性質の強い口語文を資料として、当時の話し言葉における実態の解明に焦点をあてて進められてきた。ある程度の年代にわたる複数の資料を対象に複数の語形について考察を行った先行研究として、一人称代名詞を対象とした岡田 (1998)・房 (2004)・祁 (2006a・2006b) や二人称代名詞を対象とした永田 (2006・2008a・2008b・2009) などがあげられる。

今後、近代語のコーパスの開発が進むにつれ、コーパス中の話し言葉的性質の強い口語文を利用した一・二人称代名詞の研究も広がることが予想される。コーパスを利用することで、大量のテキストからすべての一・二人称代名詞を網羅的に抽出することがこれまでよりも低コストでできるようになり、一・二人称代名詞の全体像がさらに明らかになることが期待される。また、コーパス利用によって、これまでほとんど不可能であったテキストそのものの言語量の算出やそこから見出される言語的性質の分析も可能となり、対象テキストの性質をふまえた上でのより精緻な一・二人称代名詞の分析が展開されることも期待される。

本稿ではその一つの試みとして、近代語のコーパスの一つである『近代女性雑誌コーパス』(国立国語研究所(編)、2006)に形態論情報を付与したデータを用い、コーパス中の話し言葉的性質の強い口語文の代表として小説(戯曲を含む)に含まれる口語体の会話部分に着目し、その言語量とそこから見出される言語的性質について分析・考察を行う。そしてそこに出現する一・二人称代名詞を網羅的に抽出し、分析・考察を行う。考察では適宜、もう一つの代表的な近代語のコーパスである『太陽コーパス』(国立国語研究所(編)、2005)の分析結果との比較を交え、『近代女性雑誌コーパス』の特徴について考えたい。

2. 『近代女性雑誌コーパス』の小説・戯曲の口語会話の抽出

『近代女性雑誌コーパス』は明治後期から大正期にかけて刊行された女性向け雑誌に基づくコーパスである。『女学雑誌』(1894~1895年刊行分31冊)、『女学世界』(1909年刊行分6冊)、『婦人倶楽部』(1925年刊行分3冊)の計40冊1362記事が収録されており、『太陽コーパス』と対比させながら、当時の女性が読んでいた書き言葉の実態を把握することが可能な資料として設計されている(田中、2006)。

この『近代女性雑誌コーパス』に対して、近代の文語論説文を対象とする形態素解析辞書「近代文語 UniDic」(小木曾、2009)と旧仮名遣いの口語文を対象とする形態素解析辞書(小木曾、2012)を用いて形態素解析を行ったデータが、国立国語研究所の形態論情報データベース(小木曾・中村、2011)に格納されている。本稿ではこのデータベースの2013

† kondo@ninjal.ac.jp

年7月時点のデータに基づき、分析・考察を行う。

コーパス中の小説・戯曲の区別はコーパスのXMLの記事タグのジャンル属性に基づき、属性値のNDC番号の1桁目が「9」、2桁目が「1~9またはX」、3桁目が「2~3」の記事を調査対象とする。会話部分の区別は引用タグの種別属性に基づき、属性値が「会話」のものを調査対象とする。ただし抽出された会話部分には文語体のテキストも含まれるため、記事タグ・引用タグの文体属性に基づき、口語体のテキストのみを抽出し調査対象とする。

3. 小説・戯曲の言語量

まず、コーパス全体および小説・戯曲とその口語会話部分の言語量について見る。表1は『近代女性雑誌コーパス』『太陽コーパス』についてコーパス全体、小説・戯曲、小説・戯曲の口語会話それぞれの言語量を示したものである。なお、『近代女性雑誌コーパス』の1894年分は1895年分にまとめて示す(以下同様)。

表1 コーパス全体、小説・戯曲、小説・戯曲の口語会話の言語量

		近代女性雑誌コーパス				太陽コーパス					
		1895	1909	1925	通年	1895	1901	1909	1917	1925	通年
コーパス全体	延べ語数	586665	406889	272325	1265879	2031346	1929238	1725992	1619638	1456055	8762269
	記事数	690	407	265	1362	729	635	652	504	889	3409
	号数	31	6	3	40	12	12	12	12	12	60
	1号あたりの平均延べ語数	18925	67815	90775	31647	169279	160770	143833	134970	121338	146038
小説	延べ語数	17889	30148	129620	177657	209529	187423	113857	265431	252996	1029236
	記事数	10	27	47	84	29	30	18	22	58	157
	著者数(異なり)	6	17	32	54	25	14	21	21	23	84
	1記事あたりの平均延べ語数	2982	1773	4051	3290	8381	13387	5422	12640	11000	12253
戯曲	延べ語数	29309	0	0	29309	0	0	96172	42239	0	138411
	記事数	9	0	0	9	0	0	10	4	0	14
	著者数(異なり)	2	0	0	2	0	0	6	4	0	10
	1記事あたりの平均延べ語数	14655	—	—	14655	—	—	16029	10560	—	13841
小説・戯曲	延べ語数	47198	30148	129620	206966	209529	187423	210029	307670	252996	1167647
	記事数	19	27	47	93	29	30	28	26	58	171
	著者数(異なり)	8	17	32	56	25	14	27	25	23	94
	コーパス全体に占める割合	8.0%	7.4%	47.6%	16.3%	10.3%	9.7%	12.2%	19.0%	17.4%	13.3%
小説・戯曲の口語会話	延べ語数	24313	7714	44410	76437	62680	69767	110002	80290	93090	415829
	口語会話を含む記事数	13	26	40	79	22	21	25	26	50	144
	口語会話を含む記事の著者数	5	16	28	48	10	11	20	21	23	85
	コーパス全体に占める割合	51.5%	25.6%	34.3%	36.9%	29.9%	37.2%	52.4%	26.1%	36.8%	35.6%

小説・戯曲の延べ語数は通年で206966語、『太陽コーパス』の通年1167647語の約1/6の言語量となっている。年ごとに小説・戯曲の延べ語数を見ると、もっとも多い1925年は129620語であるのに対し、もっとも少ない1909年は30148語と約4倍の違いがある。3カ年の延べ語数の変動係数は0.77で、『太陽コーパス』5カ年の延べ語数の変動係数の0.20と比べて高く、『近代女性雑誌コーパス』では年によるばらつきが顕著である。さらに、小説・戯曲中の口語会話の延べ語数においても、3カ年の変動係数が0.72で『太陽コーパス』の0.23と比べて高く、年によるばらつきが顕著である。1909年はもともと小説・戯曲の延べ語数が少ないのに加え、その中に口語会話の占める割合も25.6%と他の年と比べて低く、その結果、口語会話の延べ語数は7714語と極めて少なくなっている。

次に、1記事あたりの平均延べ語数を見ると、戯曲は通年で14655語と『太陽コーパス』の13841語と大きな違いはない一方で、小説は通年で3290語と『太陽コーパス』の12253語の約1/4である。その中でも1909年は1773語と特に少ない。この背景として、1号あたりの言語量(平均延べ語数)が少ないことや1909・1925年は読者投稿による100語前後のごく短い作品が多く掲載されていることが考えられる。

このように『近代女性雑誌コーパス』の小説・戯曲はその延べ語数や含まれる口語会話の延べ語数、1記事あたりの平均延べ語数に『太陽コーパス』と比べて大きな違いがあり、その中でも1909年は他の年と比べてその特異性が目立つ。『近代女性雑誌コーパス』の小説・戯曲の口語会話部分の利用においては、こうした性質に留意する必要がある。

4. 小説・戯曲の口語会話部分の言語量

次に、小説・戯曲の口語会話について、話者の性別と文体の観点から見ていく。表 2 に『近代女性雑誌コーパス』『太陽コーパス』の小説・戯曲の口語会話について、話者の性別ごとに話者数・延べ語数・会話数を示す。会話数は、コーパスの XML の 1 つの引用タグでマークアップされている範囲を 1 会話としてカウントした。話者の性別は、コーパスの XML の引用タグの話者属性ごとに作品内容から判断した。また、会話の文体は、文末辞「ごさい(り)ます」の出現する「ごさいます体」、「です」「ます」の出現する「ですます体」、「ごさい(り)ます」「です」「ます」の出現しない「その他」の 3 種を設定し、会話ごとに文体を決定した。

表 2 小説・戯曲の口語会話の言語量

		近代女性雑誌コーパス				太陽コーパス						
		1895	1909	1925	通年	1895	1901	1909	1917	1925	通年	
男	話者数	99	36	164	299	149	173	177	177	249	925	
	延べ語数	15732	3051	29208	47991	41972	42766	75552	54482	70697	285469	
	会話数	ごさいます体	38	3	50	91	88	18	34	26	154	320
		ですます体	246	55	378	679	214	425	927	518	910	2994
		その他	850	166	1034	2050	957	1141	2614	1984	2458	9154
計	1134	224	1462	2820	1259	1584	3503	2528	3522	12396		
女	話者数	48	62	106	216	50	47	94	117	98	406	
	延べ語数	8557	4544	14601	27702	19493	26084	34138	23893	21130	124738	
	会話数	ごさいます体	53	18	36	107	39	41	152	86	158	476
		ですます体	227	112	320	659	160	255	683	495	340	1933
		その他	305	186	580	1071	360	585	950	689	647	3231
計	585	316	936	1837	559	881	1857	1270	1145	5712		
不明	話者数	6	10	35	51	35	16	21	70	55	197	
	延べ語数	24	119	601	744	1215	917	312	1915	1263	5622	
	会話数	ごさいます体	0	0	0	0	6	2	2	0	1	11
		ですます体	2	4	14	20	28	14	0	5	20	67
		その他	7	13	40	60	54	30	30	119	72	305
計	9	17	54	80	88	46	32	124	93	383		
合計	話者数	153	108	305	566	234	236	292	364	402	1528	
	延べ語数	24313	7714	44410	76437	62680	69767	110002	80290	93090	415829	
	会話数	ごさいます体	91	21	86	198	133	61	188	112	313	807
		ですます体	475	171	712	1358	402	694	1610	1018	1270	4994
		その他	1162	365	1654	3181	1371	1756	3594	2792	3177	12690
計	1728	557	2452	4737	1906	2511	5392	3922	4760	18491		

表 2 に示した値に基づき、図 1 に会話の延べ語数の性別比率を、図 2 に会話数の性別比率を示す。

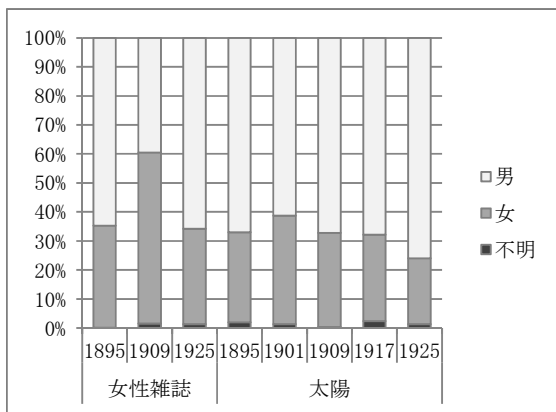


図 1 会話の延べ語数の性別比率

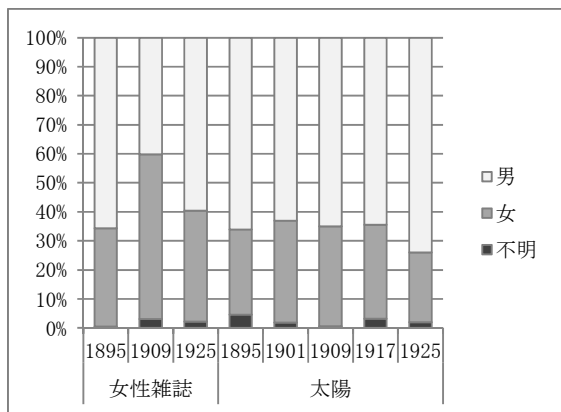


図 2 会話数の性別比率

『近代女性雑誌コーパス』において会話の性別比率は延べ語数・会話数のいずれでも 1909 年が他の年と比べて女性の比率が突出して高い。これは『太陽コーパス』と比較しても高い値である。話者の性別は一・二人称代名詞の分析の観点として重要なものであるが、『近

代女性雑誌コーパス』の利用においては1909年の特異性に留意する必要がある。

次に表2に示した値に基づき、図3に男性の会話における文体比率を、図4に女性の会話における文体比率を示す。

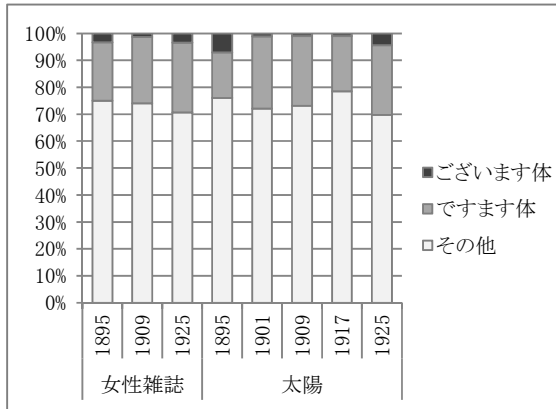


図3 男性の会話の文体比率

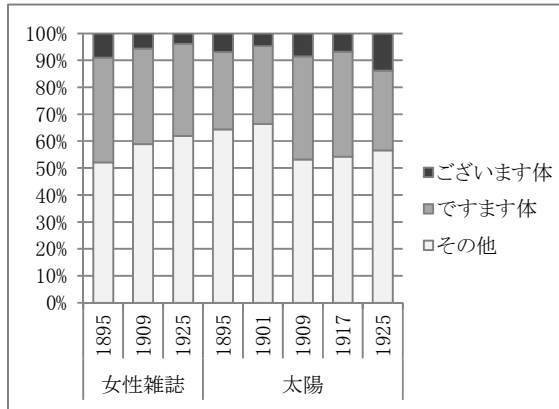


図4 女性の会話の文体比率

『近代女性雑誌コーパス』において、女性は男性より「ございます体」「ですます体」といった敬体の会話の比率が全体的に高い。これは『太陽コーパス』にも共通する傾向である。会話の文体もまた一・二人称代名詞の分析の観点として重要なものであるが、話者の性別により文体比率に差があることに留意する必要がある。

5. 男性の会話に出現する一・二人称代名詞

以上、言語量から見てきた『近代女性雑誌コーパス』の会話・戯曲とその口語会話の言語的性質を前提として、そこに出現する一・二人称代名詞を抽出・分析する。

分析対象とする一・二人称代名詞は、コーパスに付与された形態論情報の一部の人手修正を経て、形態論情報で品詞が「代名詞」の語彙素（見出し語）から、一人称代名詞・二人称代名詞いずれかの用法のみ持つものを選択した。また、二人称代名詞は接尾辞「さま」「さん」の下接するものは別語形とした。なお、本文テキストで漢字表記されるものについては、ルビ情報のあるものはそこから語形を特定し、ルビ情報のない漢字表記のものは最も一般的と考えられるよみを語形とした。これらのものは本稿中ではカタカナで表記する。ただし、ルビ情報のない漢字表記のものを語形の一つに特定することが困難と考えたものは語形未確定として、本稿中では漢字で表記する。

最初に、男性の会話に出現する一人称代名詞を見ていく。表3に男性の会話に出現する一人称代名詞の粗頻度・出現会話数・出現記事数（空欄は0を表す）を示す。

『近代女性雑誌コーパス』からは14種類の語形が抽出された。『太陽コーパス』で同様の調査をすると21種類の語形が抽出され、『近代女性雑誌コーパス』のほうが語形の種類が少ない。年別に見ると、1909年は4種類と特に種類が少ない。

次の図5は、表3で通年の粗頻度降順上位4語「ボク」「オレ」「ワタシ」「ワシ」とその他の語形をあわせて「その他」として、通年での粗頻度の比率を示し、比較のため『太陽コーパス』についても同様に示したものである。

上位4語形の比率は合計82%で『太陽コーパス』の80%と大きな差はない。ただし、4語形の中では「ボク」の比率が52%で『太陽コーパス』の33%と比べて高い。

表 3 男性の会話に出現する一人称代名詞の粗頻度・出現会話数・出現記事数

	1895			1909			1925			通年		
	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数
ボク	34	31	7	24	20	4	214	139	11	272	190	22
オレ	14	11	3	6	4	2	44	36	11	64	51	16
ワタシ	30	27	8	3	2	2	27	26	8	60	55	18
ワシ	3	2	2				33	22	5	36	24	7
セッシヤ							33	28	3	33	28	3
ワタクシ	14	14	3				11	11	2	25	25	5
ワレワレ	3	3	3				10	9	2	13	12	5
オイラ	6	6	3							6	6	3
ヨ							5	5	3	5	5	3
テマエ							4	4	2	4	4	2
私				4	2	2				4	2	2
オラ							2	2	2	2	2	2
オヌシ	1	1	1							1	1	1
ワイ							1	1	1	1	1	1

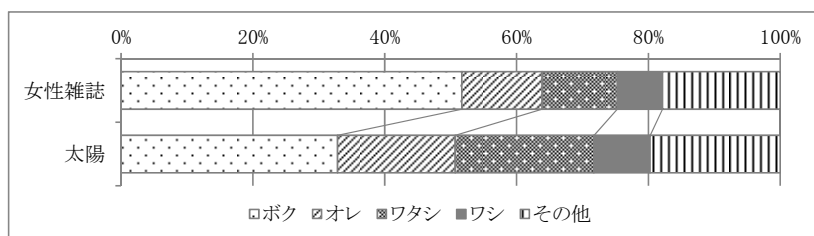


図 5 男性の会話に出現する一人称代名詞の粗頻度(通年)の比率

次に男性の会話に出現する二人称代名詞を見ていく。表 4 に男性の会話に出現する二人称代名詞の粗頻度・出現会話数・出現記事数 (0 は空欄) を示す。

表 4 男性の会話に出現する二人称代名詞の粗頻度・出現会話数・出現記事数

	1895			1909			1925			通年		
	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数
キミ	80	70	8	13	13	4	61	48	6	154	131	18
アナタ	41	35	8	5	3	3	106	75	15	152	113	26
オマエ	1	1	1	5	5	2	53	44	11	59	50	14
キコウ							19	16	3	19	16	3
キサマ							17	10	3	17	10	3
ソノホウ							10	10	3	10	10	3
ナンジ	8	5	1				1	1	1	9	6	2
オマイ	7	7	4							7	7	4
ソチ	1	1	1				5	5	1	6	6	2
オマエサン				1	1	1	3	3	2	4	4	3
アナタサマ	3	3	1							3	3	1
オメエ	1	1	1				2	2	2	3	3	3
オマイサン	2	2	1							2	2	1
キデン							2	2	1	2	2	1
オ前				2	2	1				2	2	1
ウヌ	1	1	1				1	1	1	2	2	2
オメエサマ	2	1	1							2	1	1
オメエサン	2	1	1							2	1	1
オ前サン	1	1	1							1	1	1
テメエ							1	1	1	1	1	1

『近代女性雑誌コーパス』からは 20 種類の語形が抽出された。『太陽コーパス』から抽出される 26 種類と比べると少ない。年別に見ると、一人称代名詞同様、1909 年は 5 種類と特に種類が少ない。

次の図6は表4で通年の粗頻度降順上位3語「キミ」「アナタ」「オマエ」とその他の語形をあわせて「その他」として、通年での粗頻度の比率を示し、比較のため『太陽コーパス』についても同様に示したものである。

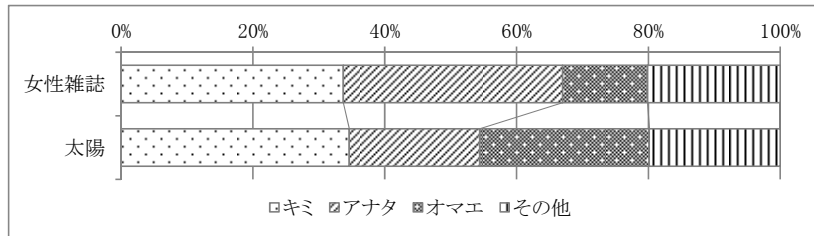


図6 男性の会話に出現する二人称代名詞の粗頻度(通年)の比率

上位3語形の比率は合計80%で『太陽コーパス』と同値である。ただし、3語形の中では「アナタ」の比率が33%で『太陽コーパス』の20%と比べて高く、一方「オマエ」が13%で『太陽コーパス』の28%と比べて低い。

以上のように、男性の会話に出現する一・二人称代名詞は『太陽コーパス』と比べて語形の種類が少なく、また、一人称代名詞は「ボク」、二人称代名詞は「キミ」「アナタ」といった特定の語形が偏って出現する傾向にあることが分かった。

最後に、男性の会話に主に出現する一・二人称代名詞について、文体ごとの出現会話率を概観しておく。出現会話率とは総会話数に対する該当語形の出現会話数の占める割合のことである。図7は一人称代名詞「ボク」「オレ」「ワタシ」の、図8は二人称代名詞「キミ」「アナタ」「オマエ」の文体ごとの出現会話率を示したものである。

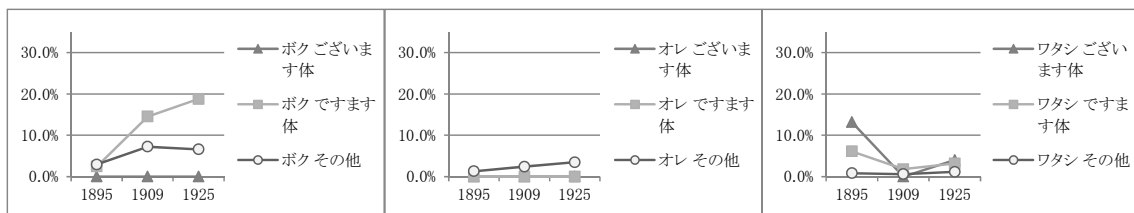


図7 男性の会話に出現する一人称代名詞の出現会話率

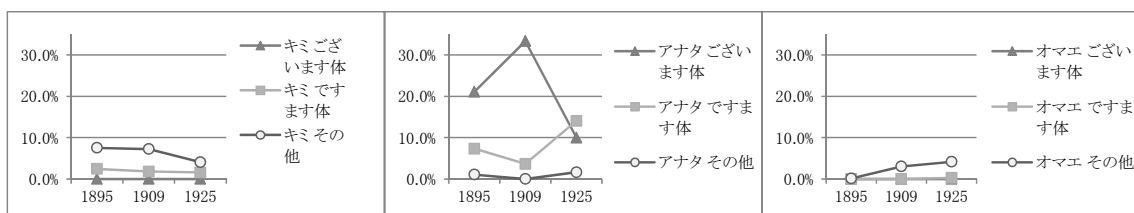


図8 男性の会話に出現する二人称代名詞の出現会話率

一人称代名詞の「ボク」は「ですます体」と「その他」に、「オレ」は「その他」に、「ワタシ」は「ございます体」「ですます体」に主に出現する。文体別に見ると、「ございます体」に「ワタシ」、「ですます体」に「ボク」「ワタシ」、「その他」に「ボク」「オレ」が主に出現する。

二人称代名詞の「キミ」は「その他」「ですます体」に、「アナタ」は「ございます体」「ですます体」に、「オマエ」は「その他」に主に出現する。文体別に見ると、「ございます体」に「アナタ」、「ですます体」に「アナタ」「キミ」、「その他」に「キミ」「オマエ」が主に出現する。

6. 女性の会話に出現する一・二人称代名詞

次に女性の会話に出現する一・二人称代名詞を見ていく。表5に女性の会話に出現する一人称代名詞の粗頻度・出現会話数・出現記事数(0は空欄)を示す。

表5 女性の会話に出現する一人称代名詞の粗頻度・出現会話数・出現記事数

	1895			1909			1925			通年		
	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数
ワタシ	122	94	9	28	25	4	186	147	16	336	266	29
ワタクシ	1	1	1	4	4	2	31	21	4	36	26	7
ワレワレ							10	6	1	10	6	1
アタシ				3	2	2	4	4	1	7	6	3
私				4	4	3	2	2	1	6	6	4
アタイ				3	3	2	1	1	1	4	4	3
ワシ							4	3	1	4	3	1
ワテ							2	2	1	2	2	1
妾				1	1	1				1	1	1
オレ							1	1	1	1	1	1
テマエ							1	1	1	1	1	1

『近代女性雑誌コーパス』からは11種類の語形が抽出された。『太陽コーパス』から抽出される16種類と比べると少ない。年別に見ると、1895年は2種類と特に少ない。

次の図9は、表5で通年の粗頻度降順上位2語「ワタシ」「ワタクシ」とその他の語形をあわせて「その他」として、通年での粗頻度の比率を示し、比較のため『太陽コーパス』についても同様に示したものである。

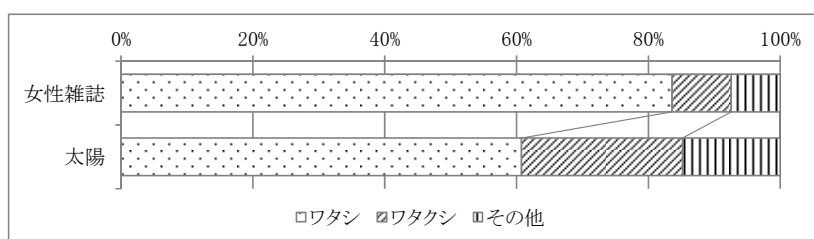


図9 女性の会話に出現する一人称代名詞の粗頻度(通年)の比率

上位2語形の比率は合計93%で『太陽コーパス』の85%より高い。また、2語形の中では「ワタシ」の比率が82%で『太陽コーパス』の61%より高く、「ワタクシ」は9%で『太陽コーパス』の24%より低い。

次に二人称代名詞を見ていく。表6に女性の会話に出現する二人称代名詞の粗頻度・出現会話数・出現記事数(0は空欄)を示す。

表6 女性の会話に出現する二人称代名詞の粗頻度・出現会話数・出現記事数

	1895			1909			1925			通年		
	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数	粗頻度	出現会話数	出現記事数
アナタ	108	95	9	9	9	6	98	88	16	215	192	31
オマイ	20	15	4							20	15	4
オマエ				2	2	1	15	14	7	17	16	8
アンタ							9	6	2	9	6	2
オマエサン				1	1	1	5	5	3	6	6	4
オ前				3	3	2				3	3	2
オマイサン	1	1	1							1	1	1

『近代女性雑誌コーパス』からは7種類の語形が抽出された。『太陽コーパス』から抽出される19種類と比べると少ない。

次の図 10 は、表 6 で通年の粗頻度降順上位 3 語「アナタ」「オマイ」「オマエ」とその他の語形をあわせて「その他」として、通年での粗頻度の比率を示し、比較のため『太陽コーパス』についても同様に示したものである。

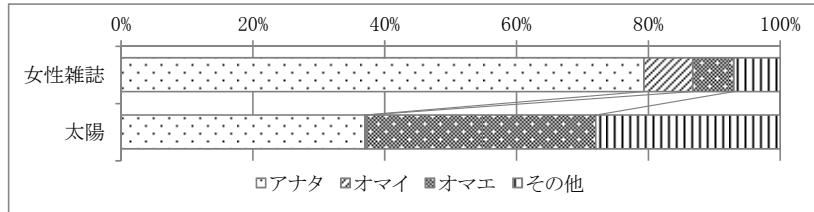


図 10 女性の会話に出現する二人称代名詞の粗頻度(通年)の比率

上位 3 語形の比率は合計 93%で『太陽コーパス』の 72%より高い。3 語形の中では「アナタ」の比率が 79%で『太陽コーパス』の 37%より高く、「オマエ」は 6%で『太陽コーパス』の 35%より低い。なお、「オマイ」は特定の 1 連載記事群のみに出現する語形で、その特殊性を考慮する必要がある。

以上のように、女性の会話に出現する一・二人称代名詞も男性の場合と同様に、語形の種類が少なく、また、一人称代名詞は「ワタシ」、二人称代名詞は「アナタ」といった特定の語形が偏って出現する傾向にあることが分かった。

最後に、女性の会話に主に出現する一・二人称代名詞について、文体ごとの出現会話率を概観しておく。図 11 は一人称代名詞「ワタシ」「ワタクシ」の、図 12 は二人称代名詞「アナタ」「オマエ」の文体ごとの出現会話率を示したものである。

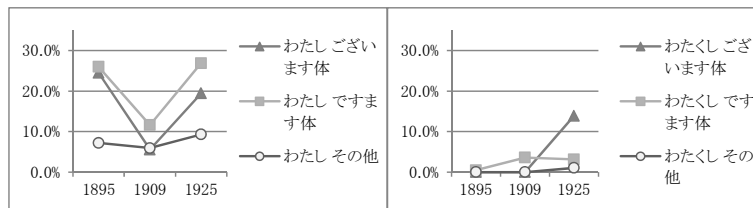


図 11 女性の会話に出現する一人称代名詞の出現会話率

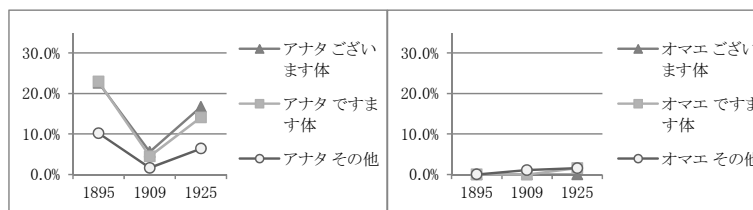


図 12 女性の会話に出現する二人称代名詞の出現会話率

一人称代名詞では「ワタシ」が、二人称代名詞では「アナタ」がどの文体でももっとも多く出現する語形となっている。

7. おわりに

以上、『近代女性雑誌コーパス』の小説・戯曲の口語会話部分に着目し、言語量から見た言語的性質をふまえた上で、そこに出現する一・二人称代名詞を抽出・分析した。

言語的性質については、1909 年が他の年や『太陽コーパス』と比べて言語量や会話の性別比率において異質であることが明らかになった。『近代女性雑誌コーパス』を利用する際には留意すべき点である。特に経年変化を分析する際、年ごとのばらつきは支障となる可

能性がある。

一・二人称代名詞については、特定の語形が偏って出現する傾向にあることが明らかになった。この背景として、会話の話し手・聞き手の社会階層や話し手と聞き手の関係といった一・二人称代名詞の選択に関わる場面要素に偏りがあることが想定される。それは単純に小説・戯曲の口語会話の言語量が多くないために生じたことなのか、それとも女性雑誌ゆえに小説・戯曲の題材が限定されてのことなのか、今後のさらなる調査・分析をまちたいが、いずれにせよ、『近代女性雑誌コーパス』の小説・戯曲の口語会話部分からは当時の話し言葉における一・二人称代名詞の或る一面しか見て取れないことは明らかである。『近代女性雑誌コーパス』の利用にあたっては『太陽コーパス』など他のコーパスと組み合わせるなど工夫が必要であろう。

また、本稿では一・二人称代名詞の分析の観点として話し手の性別と文体をとりあげたが、その他に話し手・聞き手の社会階層や話し手と聞き手の関係など、研究において重要な観点がまだ残されており、それらすべてを組み合わせる分析を行う必要がある。今後の課題としたい。

付 記

本稿は、日本学術振興会科研費(23720242)および国立国語研究所共同研究プロジェクト「通時コーパスの設計」による研究成果である。

文 献

- 岡田賢二(1998)「明治期の東京語における人称代名詞の研究—明治・大正期の落語の速記本にあらわれた一・二人称代名詞—」『埼玉大学国語教育論叢』2、pp.34-58
- 小木曾智信(2009)『科学研究費補助金研究成果報告書 近代文語文を対象とした形態素解析のための電子化辞書の作成とその活用』(<http://www2.ninjal.ac.jp/lrc/index.php?UniDic> よりダウンロード可能)
- 小木曾智信(2012)「旧仮名遣いの口語文を対象とした形態素解析辞書」『じんもんこん 2012 論文集』2012:7、pp.25-32
- 小木曾智信・中村壮範(2011)『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 『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装 改訂版』(http://www.ninjal.ac.jp/corpus_center/bccwj/doc/report/JC-U-10-01.pdf よりダウンロード可能)
- 祁福鼎(2006a)「明治時代語における自称詞の使用実態と使用規範について」『文学研究論集』24、pp.45-61
- 祁福鼎(2006b)「明治時代語における自称詞の推移と位相について」『明治大学日本文学』32、pp.95(1)-78(18)
- 国立国語研究所(編)(2005)『太陽コーパス—雑誌『太陽』日本語データベース』博文館新社
- 国立国語研究所(編)(2006)『近代女性雑誌コーパス』(http://www.ninjal.ac.jp/corpus_center/cmj/woman-mag/よりダウンロード可能)
- 田中牧郎(2005)「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」『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—』博文館新社、pp.1-48
- 田中牧郎(2006)『近代女性雑誌コーパス』の概要』『日本学術振興会科学研究費補助金研究成果報告書 基盤研究(B)「20世紀初期総合雑誌コーパス」の構築による確立期現代語の高精度な記述』pp.55-62 (http://www.ninjal.ac.jp/corpus_center/cmj/doc/19w-mag-summary.pdf よりダウンロード可能)
- 永田高志(2006)「明治前期東京語の対称詞—散切物を通じて」『国語国文』75:6、p.16-33
- 永田高志(2008a)「国定教科書の対称詞」『国語と国文学』85:3、pp.56-68

永田高志 (2008b) 「明治後期・大正期東京語の対称詞」『日本文化の鉾脈—茫洋と閃光と』
風媒社、pp.95-110

永田高志 (2009) 「総合雑誌『太陽』に見る対称詞」『国語と国文学』86:9、pp.56-70

房極哲 (2004) 「近代語における一、二人称代名詞の変遷について」『日本文化學報』21、p
p.1-15