Vocabulary Diversity in Conversation as Seen in Different Corpus Registers

レジスター・位相の違いによる会話文の語彙的多様性

山崎 誠（国立国語研究所研究系言語変化研究領域）†

Vocabulary Diversity in Conversation as Seen in Different Corpus Registers

Makoto Yamazaki (National Institute for Japanese Language and Linguistics)

要旨

本研究は，以下の 5 つのコーパスを用いて，日本語の会話文の多様性をレジスターや位相（話者の性別，年代）の観点から語彙的に分析するものである。使用したコーパスは，『日本語話し言葉コーパス』（CSJ）の学会講演及び模擬講演，『日常会話コーパス』（CEJC・構築途中のもの），『名大会話コーパス』『女性の言葉・男性の言葉（職場編）』，『現代日本語書き言葉均衡コーパス』（BCCWJ）中の小説会話文である。分析の方法は，品詞構成比，上位語，対数尤度比による特徴語の比較である。特徴語はコーパス間の比較に加えて，性別と年代による比較も行った。品詞構成比では，名詞，副詞は，CSJ 学会講演と日常会話・名大とが対照的な分布を示し，また，終助詞，感動詞-フィラーは，CSJ 学会講演・CSJ 模擬講演と日常会話・名大とが対照的な分布を示すことが分かった。特徴語では，コーパス間で感動詞（一般，フィラー），終助詞，人称代名詞の分布に違いが見られた。また，これらの語の使用において，性差の違いの方が年齢層の違いよりも特徴的な語数が多いことが観察された。

１．はじめに

本研究は，話し言葉の多様性をとらえるため，日常会話，学会講演等の改まった話し言葉，職場での会話，書かれた話し言葉（小説の中の会話）といった多様な話し言葉を語彙的に比較し，話し言葉の広がりを記述することを目的とする。

話し言葉の計量的記述研究は国立国語研究所（1955）をはじめとして多くの研究があるが，異なるレジスターの話し言葉を比較したものは少ない。高崎（1981），大石（1987）は，話し言葉と小説の会話文を比較した研究であり，示唆に富むものであるが，データ的には少量のものにとどまっている。本研究では，利用可能な話し言葉のデータが多くなってきたことから，それを比較してみようとするものである。比較の観点は，品詞構成比，上位語，対数尤度比による特徴語である。

２．データ

2.1　データの概要

本研究で使用するデータは以下のとおりである。（　）内は本研究で使う略称である。

『日本語話し言葉コーパス』（CSJ[1]）
　学会講演（CSJ_APS）
　模擬講演（CSJ_SPS）
『日常会話コーパス』（日常会話，CEJC[2]）
『名大会話コーパス』（名大，NUC[3]）
『女性の言葉・男性の言葉（職場編）』（職場会話，WP[3]）

† yamazaki@ninjal.ac.jp

1 『中納言』に収録されているデータ。品詞の情報が小分類まで付いている。

2 2017 年 4 月 14 日時点でのデータ

3 2017 年 4 月 4 日時点でのデータ。

『現代日本語書き言葉均衡コーパス』（BCCWJ）に収録されている小説のサンプルの会話文（小説会話，B_Novels[4]）

CEJC は構築途中のデータの一部を先行的に利用させていただいた。CEJC は，今後語数が増えてくるため，本研究で行う分析結果は，あくまでも現時点でのものである。小説会話は，主に BCCWJ の図書館サブコーパスに収録されている小説の中の会話部分に話者情報を付けたデータを利用した。なお，本研究で利用する言語単位はすべて短単位である。

表 1~4 は，本研究で利用するデータに関する語数情報である。

表 1　各コーパスの語数（補助記号等除外前）

	模擬講演	学会講演	名大	職場会話	日常会話	小説会話
Token	3,640,759	3,322,869	1,419,729	250,997	110,096	2,591,654
Type[5]	32,820	23,376	18,134	7,260	5,371	40,770

表 2　各コーパスの語数（補助記号等除外後）

	模擬講演	学会講演	名大	職場会話	日常会話	小説会話
Token	3,591,810	3,244,296	1,128,735	185,845	106,756	2,005,791
Type[5]	32,643	22,951	18,010	7,190	5,326	40,603

表 2 の語数は，表 1 で示した語数から以下の語を差し引いて集計している。

(a)品詞の大分類が「補助記号」「空白」「記号」「言いよどみ」「処理保留」「形態論情報付与対象外」「解釈不明」「未知語」「新規未知語」「伏せ字」「伏字化人名」「方言」「カタカナ文」となっているもの[6]。

(b)語彙素が空文字列（null）であるもの[7]。

2.2 性別と年代

表 3 は性別，表 4 は年代別の語数の情報である。年代は小説会話を除いては発話者の年

表 3 各コーパスにおける性別による語数

性別		模擬講演	学会講演	名大	職場会話	日常会話	小説会話
女性	Token	1,822,102	582,810	951,031	87,122	8,1767	531,308
	Type	24,175	10,356	16,505	4,633	4,483	20,353
男性	Token	1,769,708	2,661,486	157,899	94,754	21,472	1,385,712
	Type	25,056	20,676	7,175	5,182	2,276	35,959
不明	Token	-	-	19,805	3,969	3,517	88,771
	Type	-	-	1,253	618	675	8,313

[4] 2017 年 7 月 4 日時点でのデータ。

[5] Type は，語彙素，語彙素読み，品詞，語彙素細分類の 4 つが一致したものを同じ語とみなして集計したが，小説会話のみ語彙素，語彙素読み，品詞の 3 つで集計している。したがって，厳密な意味での比較という点では問題があるが，今回のマクロな分析にあたっては，その違いは無視できる範囲と判断した。

[6] 一つのコーパスにこれらすべてが現れるわけではない。

[7] 上記(a)の条件の多くは，語彙素が null になっているため，冗長であるが，品詞の大分類が名詞や感動詞で語彙素が null のものが若干あり，それを排除するため。

表 4 各コーパスにおける年代別による語数

年代		模擬講演	学会講演	名大	職場会話	日常会話	小説会話
若年層	Token	-	16,317	80,625	371	5,387	196,699
	Type	-	1,522	4,168	157	834	11,311
成年層	Token	1,667,288	978,470	858,861	174,965	75,473	1,622,596
	Type	25,168	14,889	15,800	6,999	4,294	37,460
老年層	Token	1,924,522	2,217,240	168,909	4,816	22,379	125,404
	Type	24,213	17,563	6,582	786	2,222	10,298
不明	Token	-	32,269	20,340	5,693	3,517	61,092
	Type	-	2,155	1,315	832	675	6,319

齢あるいは，年齢にもとづく 5 歳刻みの細かい年代情報がついているが，ここでは，いちばん刻みの粗い小説会話に合わせた。なお，若年層は 19 歳以下，成年層は 20 歳以上 59 歳未満，老年層は 60 歳以上に区分した。性別，年代ともに「不明」があるが，これらの中には，性別が「男＆女」(男女の同時発話) となっているような不明ではないものも含まれている。

3. 延べ語数と異なり語数の関係

図 1，2 は，延べ語数（Token）と異なり語数（Type）との関係を示したものである。通常，延べ語数が増えると異なり語数も増えるが，図 1 の CSJ_APS（学会講演）では異なり語数の相対的下降が見られる。この下降は直前の小説会話の異なり語数が相対的に多いことによってより強調されている。図 2 に語彙密度の指標とされる R(Guirad's Index)[8]と C[9]（Herdan）（石川,2012:143-144）の値を示したが，小説会話がその前後に比べて高い値を示していることが分かる。R の値からは，日常会話，職場会話，名大，学会講演，模擬講演の値と小説会話とは値が大きく異なることが見て取れる。

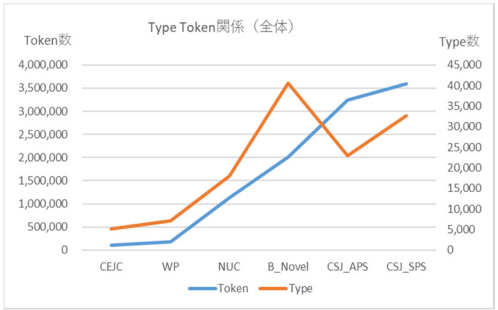

図 1 Type Token 関係（全体）

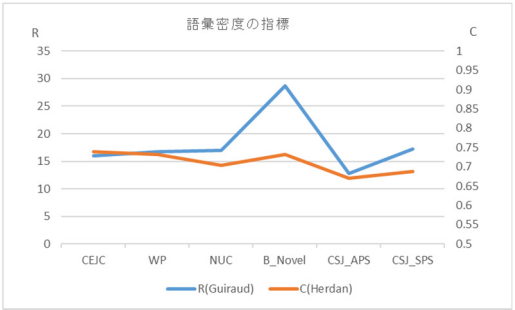

図 2 語彙密度の指標

4. 品詞構成比の比較

表 5 は，品詞の大分類による各コーパスの品詞構成比を示すものである。個別の品詞を見ると次のようなことが分かる。各品詞の構成比の平均の 1.3 倍以上を（相対的に）多い，0.7 倍以下を（相対的に）少ないとした。

・名詞は，学会講演で多い。

[8] R=Type/√Token で計算される。

[9] C=log$_e$Type/log$_e$Token で計算される。

・代名詞は，学会講演で少ない。
・形容詞は，日常会話，名大で多く，学会講演で少ない。
・副詞は，日常会話で多く，学会講演で少ない。
・連体詞は，日常会話で少ない。
・接続詞は，学会講演で多く，小説会話で少ない。
・感動詞は，日常会話で多く，小説会話で少ない。
・接頭辞は，小説会話で多い。
・形状詞，動詞，助詞，助動詞，接尾辞は，コーパスによって目立って多い少ないという特徴はない。

表 5 品詞構成比（赤字は構成比が相対的に多いもの，青字は相対的に少ないもの）

	模擬講演	学会講演	名大	職場会話	日常会話	小説会話
名詞	0.217	0.277	0.172	0.198	0.177	0.224
代名詞	0.024	0.018	0.042	0.035	0.038	0.037
形容詞	0.018	0.010	0.031	0.025	0.032	0.022
形状詞	0.012	0.013	0.011	0.008	0.010	0.011
動詞	0.134	0.133	0.115	0.116	0.110	0.143
副詞	0.037	0.020	0.051	0.046	0.061	0.030
連体詞	0.011	0.014	0.014	0.011	0.008	0.013
接続詞	0.011	0.012	0.007	0.008	0.008	0.003
感動詞	0.058	0.070	0.078	0.070	0.101	0.009
助詞	0.326	0.295	0.332	0.320	0.306	0.338
助動詞	0.127	0.108	0.127	0.137	0.126	0.136
接頭辞	0.006	0.006	0.005	0.005	0.007	0.009
接尾辞	0.019	0.024	0.016	0.019	0.017	0.025

表 5 は品詞の大分類による構成比であるが，分類を細かくしていくと，別の様相が見えてくる。図 3 は，話し言葉に特徴的である終助詞および感動詞の中分類（感動詞-一般，感動詞-フィラー）の構成比を示したものである。この 3 つの品詞の割合は，学会講演と模擬講演がよく似ており，また，名大，職場会話，日常会話の 3 つが似ていることが分かる。小説会話はさらに別のグループとも考えられる。

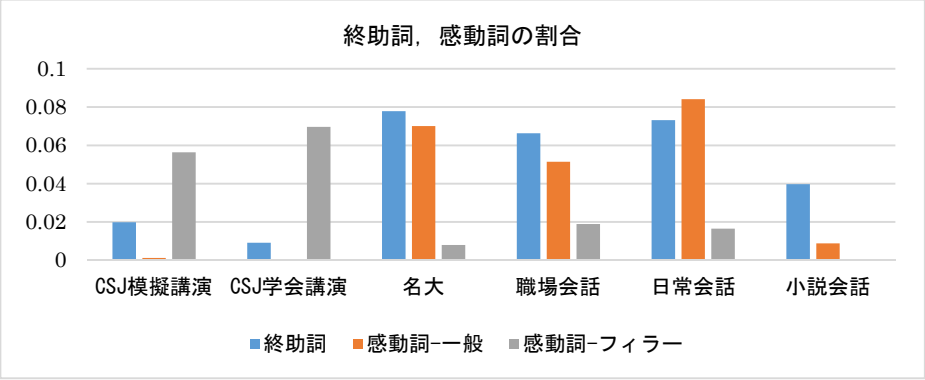

図 3 終助詞，感動詞の割合

5. 上位語の比較

表 6 は各コーパスにおける出現頻度の高い語 15 語をそれぞれ挙げたものである。いずれも助詞，助動詞等の機能語が上位に来ていることは変わりないが，いくつか特徴を指摘することができる。学会講演の第 4 位の「えー」というフィラーは，他のコーパスよりも上位に位置している。日常会話と名大には「うん」「ね」「そう」「か」が現れ，この 2 つのコーパスの近さを表している。また，職場会話では「です」が他のコーパスより上位に来ている。書き言葉では使用頻度がもっとも高い語は，一般的に格助詞の「の」であるが，表 6 からは「の」が 1 位に来ているのは学会講演のみで，助動詞の「だ」が 1 位になっているものが 4 つ，接続助詞の「て」が 1 位になっているものが 1 つとなっている。

表 6 各コーパスにおける頻度上位 15 語

模擬講演			学会講演		
出現頻度	語彙素	品詞	出現頻度	語彙素	品詞
135265	て	助詞-接続助詞	139151	の	助詞-格助詞
123755	だ	助動詞	111755	と	助詞-格助詞
107384	の	助詞-格助詞	111054	て	助詞-接続助詞
102697	と	助詞-格助詞	110374	えー[10]	感動詞-フィラー
99858	に	助詞-格助詞	95845	だ	助動詞
93626	の	助詞-準体助詞	93037	に	助詞-格助詞
92212	た	助動詞	89314	ます	助動詞
86136	は	助詞-係助詞	85464	を	助詞-格助詞
85973	です	助動詞	81746	は	助詞-係助詞
82990	が	助詞-格助詞	80815	為る	動詞-非自立可能
81389	言う	動詞-一般	74568	が	助詞-格助詞
69133	も	助詞-係助詞	74302	言う	動詞-一般
68427	を	助詞-格助詞	55159	た	助動詞
67183	ます	助動詞	49312	です	助動詞
57926	為る	動詞-非自立可能	43576	の	助詞-準体助詞

表 6 各コーパスにおける頻度上位 15 語（続き）

名大			職場会話		
出現頻度	語彙素	品詞	出現頻度	語彙素	品詞
53610	だ	助動詞	7841	だ	助動詞
41013	うん	感動詞-一般	5189	の	助詞-準体助詞
30103	た	助動詞	4717	て	助詞-接続助詞
29663	て	助詞-接続助詞	4624	ね	助詞-終助詞
29568	ね	助詞-終助詞	4501	です	助動詞
26343	の	助詞-準体助詞	4221	の	助詞-格助詞
23665	か	助詞-副助詞	4135	た	助動詞
21052	と	助詞-格助詞	4002	は	助詞-係助詞

[10] 赤字は本文中で言及した語。

20747	の	助詞-格助詞	3442	に	助詞-格助詞
20033	も	助詞-係助詞	3404	と	助詞-格助詞
19691	で	助詞-格助詞	3162	が	助詞-格助詞
19569	が	助詞-格助詞	3044	で	助詞-格助詞
19496	に	助詞-格助詞	2902	も	助詞-係助詞
18696	は	助詞-係助詞	2850	言う	動詞-一般
18259	そう	副詞	2475	よ	助詞-終助詞

表 6 各コーパスにおける頻度上位 15 語（続き）

日常会話			小説会話		
出現頻度	語彙素	品詞	出現頻度	語彙素	品詞
5356	だ	助動詞	86944	だ	助動詞
4408	うん	感動詞-一般	72791	て	助詞-接続助詞
2924	ね	助詞-終助詞	69974	の	助詞-格助詞
2715	た	助動詞	68994	は	助詞-係助詞
2544	て	助詞-接続助詞	65263	た	助動詞
2391	の	助詞-準体助詞	62306	に	助詞-格助詞
2242	そう	副詞	47905	を	助詞-格助詞
1932	か	助詞-副助詞	45161	が	助詞-格助詞
1766	で	助詞-格助詞	43853	の	助詞-準体助詞
1760	も	助詞-係助詞	38085	為る	動詞-非自立可能
1750	と	助詞-格助詞	37645	と	助詞-格助詞
1712	に	助詞-格助詞	33841	も	助詞-係助詞
1707	の	助詞-格助詞	29584	です	助動詞
1685	が	助詞-格助詞	25711	で	助詞-格助詞
1595	は	助詞-係助詞	24488	ます	助動詞

6. 特徴語による比較

6. 1 コーパス全体

特徴語は対数尤度比（LLR）を利用して抽出した。計算式は以下のとおりである。金愛蘭ほか（2008：201）から引用する。

LLR=2(alna+blnb+clnc+dlnd-(a+b)ln(a+b)-(a+c)ln(a+c)-(b+d)ln(b+d)
-(c+d)ln(c+d)+(a+b+c+d)ln(a+b+c+d))

a：Aでの単語 W の頻度 b：Bでの単語 W の頻度
c：Aの延べ語数-a d：Bの延べ語数-b ※ln(x)=x の自然対数
ただし、単語 W のAでの使用率がBでの使用率より低い場合、×(-1)の補正を行う。

表 7～12 はそれぞれのコーパスの特徴語上位 20 語を抜き出したものである。模擬講演では「あの」「まー」「んー」などのフィラーが上位に来ている。一方，学会講演では同じフィラーでも「えー」「えーと」が特徴語となっている。名大は感動詞「うん」と「ね」「さ」な

どの終助詞が上位に現れる。職場会話では，フィラーは現れず，「はい」「あっ」「ええ」などの感動詞が上位に来ている。また，挨拶の「おはよう」が出ていることも職場会話の特徴である。日常会話では「うん」「ううん」「そう」などの応答に使う語が上位に来ている。小説会話はまったく様相が異なり，「貴方（あなた)」「俺」「私」「君」などの代名詞が特徴的である。また，「わ」「よ」「ぞ」という役割語的な終助詞が上位に来ていることも注目される。

表 7 模擬講演の特徴語

LLR	頻度(対象 C)[11]	頻度(参照 C)[12]	語彙素	品詞
28733.05	53096	30137	あの	感動詞-フィラー
14928.70	43603	33560	まー	感動詞-フィラー
11992.63	85973	95387	です	助動詞
6850.22	93626	121352	の	助詞-準体助詞
6164.77	10557	5588	んー	感動詞-フィラー
5901.83	30001	29616	けれど	助詞-接続助詞
5654.08	8325	3820	たり	助詞-副助詞
5386.62	69133	88383	も	助詞-係助詞
5046.30	9319	5256	矢張り	副詞
4044.79	20926	20777	思う	動詞-一般
4040.79	7812	4569	自分	名詞-普通名詞-一般
3380.82	81389	115900	言う	動詞-一般
3333.90	14188	13046	もう	副詞
3008.42	2972	811	迚も	副詞
2734.89	11734	10824	その	感動詞-フィラー
2488.66	12769	12632	私	代名詞
2402.21	28568	35874	で	接続詞
2120.99	1083	16	無人	名詞-普通名詞-一般
2114.00	30963	40556	って	助詞-副助詞
1956.72	17514	20563	行く	動詞-非自立可能

表 8 学会講演の特徴語

LLR	頻度(対象 C)	頻度(参照 C)	語彙素	品詞
86506.55	110374	55493	えー	感動詞-フィラー
21489.76	89314	97724	ます	助動詞
13645.67	18119	9352	えーと	感動詞-フィラー
12585.05	139151	204033	の	助詞-格助詞
12270.15	23178	16331	此の	連体詞
10451.88	14068	7359	的	接尾辞-形状詞的
10214.37	4830	102	音声	名詞-普通名詞-一般

[11] 表 7 では，着目している模擬講演での頻度。表 8～12 も同様に，特徴語を取り出す対象となるコーパスの頻度。なお，「対象 C」「参照 C」は，それぞれ対象コーパス，参照コーパスの意味。

[12] 対象コーパス以外の 5 つのコーパスにおける頻度を合算したもの。表 8～12 も同様。

9851.37	6403	940	行う	動詞-一般
9282.57	111755	166548	と	助詞-格助詞
9047.85	7664	2119	場合	名詞-普通名詞-副詞可能
8979.66	80815	112679	為る	動詞-非自立可能
8429.11	5518	828	結果	名詞-普通名詞-副詞可能
8415.55	4028	100	用いる	動詞-一般
8379.32	3985	91	単語	名詞-普通名詞-一般
7901.67	4140	232	データ	名詞-普通名詞-一般
7830.39	3960	171	モデル	名詞-普通名詞-一般
7559.69	85464	125536	を	助詞-格助詞
7148.51	6372	1929	語	名詞-普通名詞-一般
7118.53	7298	2747	因る	動詞-一般
6971.27	3520	150	実験	名詞-普通名詞-サ変可能

表 9 名大の特徴語

LLR	頻度(対象 C)	頻度(参照 C)	語彙素	品詞
136756.95	41013	9169	うん	感動詞-一般
27064.77	29568	67956	ね	助詞-終助詞
26413.50	9120	3425	さ	助詞-終助詞
20351.51	17633	32807	何	代名詞
18403.87	18259	38536	そう	副詞
16628.97	6518	3515	あっ	感動詞-一般
16217.55	14788	28869	よ	助詞-終助詞
16131.34	23665	67120	か	助詞-副助詞
13880.91	6934	6232	の	助詞-終助詞
13560.54	4689	1763	ううん	感動詞-一般
13095.18	3844	742	ふん	感動詞-一般
11411.53	18033	53486	って	助詞-副助詞
9591.85	4115	2740	ああ	感動詞-一般
9553.45	10955	25988	から	助詞-接続助詞
9015.26	3259	1405	零	名詞-数詞
7907.60	14832	48486	てる	助動詞
7509.48	2253	484	じゃん	助詞-終助詞
7193.08	1903	186	然う然う	感動詞-一般
6372.79	4073	5323	彼の	連体詞
6082.09	1988	613	へえ	感動詞-一般

表 10 職場会話の特徴語

LLR	頻度(対象 C)	頻度(参照 C)	語彙素	品詞
5096.35	1517	5411	はい	感動詞-一般
3317.35	4624	92900	ね	助詞-終助詞
2363.18	2475	41182	よ	助詞-終助詞
2310.20	1124	8909	あっ	感動詞-一般

1999.53	728	3765	ええ	感動詞-一般
1653.77	660	3954	まあ	副詞
1516.23	752	6103	ああ	感動詞-一般
1512.64	2286	47896	うん	感動詞-一般
1395.44	2415	54380	そう	副詞
1256.13	662	5790	ううん	感動詞-一般
1092.64	406	2170	ゼロ	名詞-数詞
1058.64	1655	35288	から	助詞-接続助詞
946.11	646	7486	ちゃう	助動詞
880.00	2269	61049	てる	助動詞
727.44	754	12412	の	助詞-終助詞
636.12	1155	26585	一	名詞-数詞
630.57	1186	27772	良い	形容詞-非自立可能
618.05	112	94	御早う	感動詞-一般
582.02	564	8832	彼の	連体詞
581.35	383	4281	否	感動詞-一般

表 11 日常会話の特徴語

LLR	頻度(対象 C)	頻度(参照 C)	語彙素	品詞
11495.03	4408	45774	うん	感動詞-一般
3001.85	877	5575	ううん	感動詞-一般
2752.92	2242	54553	そう	副詞
2444.82	2924	94600	ね	助詞-終助詞
2050.22	830	9203	あっ	感動詞-一般
1484.79	1482	42175	よ	助詞-終助詞
1481.86	586	6269	ああ	感動詞-一般
1375.46	564	6364	はい	感動詞-一般
1053.65	660	12506	の	助詞-終助詞
1022.20	1401	49039	何	代名詞
966.84	1130	35813	から	助詞-接続助詞
947.21	383	4231	まあ	副詞
810.90	1932	88853	か	助詞-副助詞
726.72	1453	61865	てる	助動詞
647.89	835	28123	良い	形容詞-非自立可能
546.07	1458	70061	って	助詞-副助詞
525.96	5356	367995	だ	助動詞
516.29	239	3183	えっ	感動詞-一般
484.84	209	2528	じゃん	助詞-終助詞
472.76	120	563	んっ	感動詞-一般

表 12 小説会話の特徴語

LLR	頻度(対象 C)	頻度(参照 C)	語彙素	品詞
11552.04	5163	1091	貴方	代名詞
9931.15	5111	1637	わ	助詞-終助詞
9373.44	17303	26354	よ	助詞-終助詞
8591.20	4102	1070	俺	代名詞
8311.00	10690	12557	さん	接尾辞-名詞的-一般
8291.54	11191	13671	ず	助動詞
7483.98	5438	3387	私	代名詞
7466.77	68994	192175	は	助詞-係助詞
6678.80	2445	195	君	代名詞
6597.05	65263	184324	た	助動詞
5949.64	21965	46839	居る	動詞-非自立可能
5624.25	2697	712	御前	代名詞
4862.98	2817	1180	様	接尾辞-名詞的-一般
4242.04	11558	21667	御	接頭辞
4236.73	1840	350	ぞ	助詞-終助詞
3721.07	2006	718	男	名詞-普通名詞-一般
3567.53	1428	192	殺す	動詞-一般
3426.09	3756	3820	へ	助詞-格助詞
3308.26	86944	286407	だ	助動詞
3239.79	16680	39802	無い	形容詞-非自立可能

6．2　性別と年代

表 13 に性別による各コーパスの特徴語を，表 14 に年代による各コーパスの特徴語をそれぞれまとめた。いずれも各コーパス内における特徴語である[13]。ここでは，特徴語のうち，話題に左右されることが少ない語として，位相による差が出やすいと思われる，人称代名詞，終助詞，フィラー[14]，感動詞[15]を採り上げた。表 13，表 14 ともに，対数尤度比（LLR）の値が 10.83 以上[16]でかつ対象コーパスでの出現頻度 10 以上という条件で取り出された上位 25 語以内に現れた範囲での集計である。特徴語の数は，性別では延べ 99 個，年代では 88 個で性別のほうが多い。また，特徴語が現れたセルの数は性別で 48 個中 33 個（=0.69）であるのに対して，年代では 68 個中 39 個（=0.57）であり，性別の方が年代よりも特徴語が現れやすいことが分かる。品詞別で見ると，人称代名詞が性別では 24 個（延べ）現れているが，年代では 14 個（延べ）であり，人称代名詞は年代差より性差のほうが出やすい。

フィラーと感動詞を比べると，表 13（性別）では模擬講演，学会講演ではフィラーに大きな性差が現れているが[17]，名大，職場（女性），日常会話，小説会話では，フィラーはほとんど現れず，感動詞の差の方が目立つ。これは独話と対話の違いを表していると見ることができるだろう。同じ傾向は表 14（年代）にも見られ，模擬講演と学会講演ではフィラーが

13 例えば，模擬講演の女性であれば，対照コーパスは模擬講演の女性，参照コーパスは模擬講演の男性となっている。表 13，14 とも，「不明」は含んでいない。
14 品詞情報が「感動詞-フィラー」であるもの。
15 品詞情報が「感動詞-一般」であるもの。
16 対数尤度比（LLR）が 0.1%で有意である値（高見(2003:89)）。
17 職場会話にもフィラーに大きくはないが性差が認められる。

特徴語に現れ，感動詞にはほとんど現れない。また，その他のコーパスでは，特徴語が感動詞に多く現れ，フィラーにはほとんど現れず，傾向が逆になっている。

個別の観察では，終助詞「ね」が男性の特徴語にも女性の特徴語にも現れていることが興味深い。「ね」は，模擬講演，学会講演では男性の特徴語，名大，日常会話，小説会話では女性の特徴語となっている[18]。名大と日常会話の老年層に「ほら」が現れているが，これは一見気づきにくい年代差であるのかもしれない。

表 13 性別による特徴語の出現状況

コーパス	性別	人称代名詞	終助詞	フィラー	感動詞
模擬講演	女性	わたし		あの，と	
	男性	僕，我々	ね，か	えー，まー，おー，いー，その，うー，あー	
学会講演	女性			と	
	男性		ね	えー，まー，その，おー，あー，うー，えーと，いー	
名大	女性	わたし	ね，の		うん，ううん，あっ，ふん，へえ
	男性	俺，僕，おまえ，あいつ	な，ねん		はい，ああ，いや，おお
職場	女性	わたし	わ，かしら，の	んー	うん，ふん，あっ，ううん，ねえ，んっ
	男性	僕，俺，おまえ		えー，まー，あの	
日常会話	女性	わたし	ね		ううん，えっ
	男性	俺，僕	もの	あの	いや，ねえ，んっ
小説会話	女性	わたし，わたくし，あなた	わ，の，かしら，よ，ね，もの		あら，ええ，ねえ
	男性	俺，僕，君，わし，おまえ，我々	な，か，ぞ，ぜ，い		いや，おい

表 14 年代による特徴語の出現状況

コーパス	年代	人称代名詞	終助詞	フィラー	感動詞
模擬講演	若年	わたし		と	
	成年	僕		えーと，まー，えー，と	うん
	老年	わたくし	ね	おー，あー，いー，うー，んー，その	
学会講演	成年			えーと，えー	
	老年	わたくし	ね	おー，あの，あー，うー，んー，いー	
名大	若年		さ，よ，じゃん		あっ，そうそう，ああ，えっ
	成年	俺	さ，か，な，じゃん，け		ううん，ふん，うん
	老年	わたし	ね，わ，の	あの	ええ，ほら，あら
職場	若年				
	成年				はい，ふん
	老年	僕	な，ね		ああ

[18] これには用法上の差異があるものと思われるが具体的な分析は行っていない。

日常会話	若年			あー，えーと	んっ
	成年	わたし	ね，さ，か		うん，はい，あ あ，へえ，ううん，ええ
	老年			あの	ほら
小説会話	若年	僕，わたし，俺	よ，の，さ，じゃん，もの		うん，あっ，えっ，ねえ
	成年	彼女，彼	か		
	老年	わし	ね		

7. まとめと今後の予定

本研究では話し言葉の多様なあり方を明らかにするため，語彙の量的な観点からの記述を行った。利用した 5 つのコーパスにそれぞれ違いがあることが確認され，レジスターの違いが語彙の量的な側面に現れていることが確認された。位相差では，人称代名詞，終助詞，感動詞，フィラーに性差，年代差が確認された。

今後の課題としては，本研究で用いなかった語彙的な考察を進める必要がある。例えば，ユールの K 特性値で比較する，頻度分布を比較することなどである。さらに文末表現や語順など文法的な観点からの比較を行い，話し言葉の多様性の記述の幅を広げていく必要があるだろう。

謝　辞

本研究は，国立国語研究所のプロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」（プロジェクトリーダー・小磯花絵）および日本学術振興会・科学研究費補助金「会話文への発話者情報の付与によるコーパスの拡張」（15H03212）による成果に基づいて行われている。

文　献

大石初太郎(1987) 近代・現代小説会話文の資料性『国文学解釈と鑑賞』52(7),72-79.

金愛蘭・桐生りか・近藤明日子・田中牧郎(2008)「一般向け専門用語」抽出の試みー医療用語を例にー,『日本語学会 2008 年度春季大会予稿集』199-206.
http://pj.ninjal.ac.jp/byoin/tyosa/corpus/zuhyo/corpus_siryo2.pdf

国立国語研究所(1955)『談話後の実態』東京：秀英出版

高崎みどり(1981) 小説の中の会話文について『ことば』2,86-97.

高見敏子(2003)「高級紙語」と「大衆紙語」の corpus-driven な特定法『北海道大学大学院国際広報メディア研究科言語文化部紀要』44.