

シンポジウム「日常会話コーパス」I

レジスターの違いによる話し言葉の変容

山崎 誠(国立国語研究所)

概要

- 1. レジスター班紹介
- 2. データ
- 3. 頻度
- 4. 上位語
- 5. 特徴語
- 6. まとめと課題

2

1.1 レジスター班紹介

- 日常会話や独話だけでなく、シナリオや小説の会話文など書かれた話し言葉をも含む多様な言語使用域(レジスター)の話し言葉を比較し、語彙・文法・韻律などの特性を探ります。

(プロジェクトHPより)

3

1.3 研究内容

- 関連するレジスターとの比較により話し言葉の諸特性を明らかにする。
 - 実際の話し言葉と書かれた話し言葉(小説の会話文、シナリオ、国会会議録など)はどのような点が異なっているか。話し言葉らしさが具体的にどのような部分に現れるかが分かれば、話し言葉らしさの指標の開発にも役立つ。
 - 話者の属性(性別、年代)による話し言葉の変容にはレジスターによりどのような違いがあるか。
 - 話し言葉における基本的な語彙、構文とは何か。
 - etc.

4

1.4 本日のテーマ

- 現在利用できる3つのデータを比較し、どのような点が異なるか語彙的に探る。
 - 日本語話し言葉コーパス(CSJ)
 - 名大会話コーパス
 - BCCWJにおける小説の会話文

5

2.1 日本語話し言葉コーパス(CSJ)

- 日本語の自発音声を大量にあつめて多くの研究用情報を付加した話し言葉研究用のデータベース。
- 国立国語研究所・情報通信研究機構(旧通信総合研究所)・東京工業大学の共同開発。
- 2004年公開。約750万語、660時間。
以上 http://pj.ninjal.ac.jp/corpus_center/csj/より

6

2.2 CSJにおける音声のタイプ

表2 音声のタイプと時間

音声のタイプ	タイプ	話者数(異なり)	ファイル数	時間
学会講演	独話	819	987	274.4
模擬講演	独話	** 594	1,715	329.9
その他	独話	*** 16	19	24.1
学会講演インタビュー	対話	* (10)	10	2.1
模擬講演インタビュー	対話	* (16)	16	3.4
課題指向対話	対話	* (16)	16	3.1
自由対話	対話	* (16)	16	3.6
再朗読	朗読	* (16)	16	5.5
朗読	朗読	* (248)	507	15.5
	計	1,417	3,302	661.6

* ()内は全員が学会講演話者もしくは模擬講演話者としてカウントされている

** 10名は学会講演話者としてもカウントされている

*** 2名は学会講演話者としてもカウントされている

(『日本語話し言葉コーパスの構築法』,p.3)

7

2.3 CSJの語数(短単位、長単位)

表3 形態論的単位数

音声のタイプ	短単位数	長単位数	%長単位
学会講演	3,279,364	2,654,823	81.0
模擬講演	3,605,729	3,115,302	86.4
その他	282,728	239,989	84.9
朗読と再朗読	207,478	172,216	83.0
対話	149,826	131,544	87.8
全体	7,525,125	6,313,874	83.9

(『日本語話し言葉コーパスの構築法』,p.3)

8

2.4 名大会話コーパス

- 科学研究費基盤研究(B)(2)「日本語学習辞書編纂に向けた電子化コーパス利用によるコロケーション研究」(平成13年度～15年度 研究代表者 大曾美恵子)の一環として作成されたコーパス。
- 120件合計約100時間の日本語母語話者同士の雑談を文字化した会話データ。文字化テキストのみ公開。

以上 <https://nknet.ninjal.ac.jp/nknet/ndata/nuc/>より

9

2.5 名大会話コーパス:話者の属性(1)

- 性別と年代

性別 年代	女性	男性	総計
10代	13	2	15
20代	70	18	88
30代	26	1	27
40代	16	8	24
50代	18	4	22
60代	11	4	15
70代～	6		6
不詳	1		1
総計	161	37	198

- 性別:全体の約8割が女性。
- 年代:全体の約7割が40代以下。

10

2.6 名大会話コーパス:話者の属性(2)

- 出身地

出生地方	人数
北海道	11
東北	8
関東	49
中部	86
近畿	21
中国・四国	11
九州・沖縄	11
海外	1
総計	198

- 中部+近畿で半数以上を占める。

11

2.7 データ:BCCWJの小説会話文

- 現代日本語書き言葉均衡コーパス(BCCWJ)

- 現代日本語の多様な書き言葉を収録。
 - 書籍(図書館書籍、出版書籍、ベストセラー)、新聞、雑誌、教科書、広報紙、白書、韻文、法律、Yahoo!ブログ、Yahoo!知恵袋、国会会議録

- 2011年公開。約1億語(短単位)。

12

2.8 データ:BCCWJの小説会話文

- 会話の部分に対応するタグ
- <speech>: 前後に改行を伴う、「」で囲まれた文字列。
 - ほぼ会話に対応。

「兄弟って素敵？」
「さあ、どうかな、もう七年も会ってないんですよ」

- <quotation><speech><paragraph><sentence> 「兄弟って素敵？」</sentence><br type="automatic_original" /></paragraph></speech></quotation>
- <quotation><speech><paragraph><sentence> 「さあ、どうかな、もう七年も会ってないんですよ」</sentence><br type="automatic_original" /></paragraph></speech></quotation>

• Lba9_00017「中国行きのスロー・ボート」(村上春樹)

13

2.9 データ:BCCWJの小説会話文

- <quote>: 文中にある、「」で囲まれた文字列。
 - 名前、強調など、会話以外の可能性も多い。

そう、麗子には、くわすかでも、替しきみじめきは絶対似合わない。「ローマの休日」の王女はホットドッグを立食したり、町を歩いたりして喜んだが、あれは王女だったからなのだ。それができもしないし、したくないのは、女王だ。替しい、環境におかれた女王ぐらゐ。

- <sentence><quote>「ローマの休日」</quote>の王女はホットドッグを立食したり、町を歩いたりして喜んだが、あれは王女だったからなのだ。</sentence>

• Lba9_00051「死はやさしく奪う」(栗本薫)

14

2.10 データ:BCCWJの小説会話文

- <speech>,<quote> の数、サンプル数

レジスター	<speech>数	<speech>を含むサンプル数	<quote>数	<quote>を含むサンプル数
LB	173,433	5,105	165,822	8,978
OB	32,171	917	16,643	1,080
OC	0	0	0	0
OL	0	0	5,286	308
OM	22,905	159	2,885	122
OP	789	244	18,775	354
OT	0	0	0	0
OV	0	0	201	68
OW	0	0	21,966	1,352
OY	0	0	0	0
PB	108,241	3,479	173,980	8,646
PM	16,327	844	25,176	1,787
PN	1,123	199	15,062	1,455
合計	354,989	10,947	445,796	24,150

15

2.11 話者情報の付与

- BCCWJには、話者情報が付いていないため、性別と年代の情報を付与している(科研費)。
- 性別
 - 男性、女性、不明(人間以外の場合など)
- 年代
 - 若年層 10代
 - 成年層 20~59歳
 - 老年層 60歳以上

16

2.12 方法

- 短単位データをもとにして、使用頻度に基づく比較
- LLR(対数尤度比)による特徴語の比較
- 各データ内での比較とデータ間の比較

17

3. データの量的概要

18

3.1 CSJ概要

- 性別

	延べ		異なり		TTR(異なり/延べ)	
	学会講演	模擬講演	学会講演	模擬講演	学会講演	模擬講演
女性	599,507	1,853,605	11,194	26,159	0.0187	0.0141
男性	2,745,109	1,803,672	24,173	27,208	0.0088	0.0151
全体	3,344,616	3,657,277	26,897	36,284	0.0080	0.0099

19

3.2 CSJ概要

- 年代

	延べ		異なり		TTR(異なり/延べ)	
	学会講演	模擬講演	学会講演	模擬講演	学会講演	模擬講演
若年層	16,703	0	1,572	0	0.0941	
成年層	1,010,154	1,743,812	16,342	27,319	0.0162	0.0157
老年層	2,284,696	1,913,465	20,534	26,234	0.0090	0.0137
不明	33,063	0	2,261	0	0.0684	
全体	3,344,616	3,657,277	26,897	36,284	0.0080	0.0099

(注)CSJの情報を、若年層:10代、成年層:20代~50代、老年層:60代以上 としてまとめた。

20

3.3 名大会話コーパス概要

• 性別

	延べ	異なり	TTR(異なり/ 延べ)
女性	1,217,578	16,645	0.0137
男性	203,748	7,294	0.0358
全体	1,421,326	18,069	0.0127

数値は2016年6月時点でのもの(以下同じ)。

21

3.4 名大会話コーパス概要

• 年代

	延べ	異なり	TTR(異なり/ 延べ)
若年層	106,269	4,285	0.0403
成年層	1,097,443	15,918	0.0145
老年層	216,884	6,720	0.0310
不詳	730	200	0.2740
全体	1,421,326	18,069	0.0127

22

3.5 BCCWJ小説会話文概要

• 性別

	延べ	異なり	TTR(異なり/ 延べ)
女性	234,653	11,147	0.0475
男性	570,449	20,900	0.0366
不明	27,670	3,566	0.1289
(未記入)	27,608	3,441	0.1246
全体	860,380	24,525	0.0285

- 2016年8月時点での作業途中の667サンプルが対象(以下同じ)。
- 「不明」は、「不明」と判断したもの。「(未記入)」は、何も記入していないもの。

23

3.6 BCCWJ小説会話文概要

• 年代

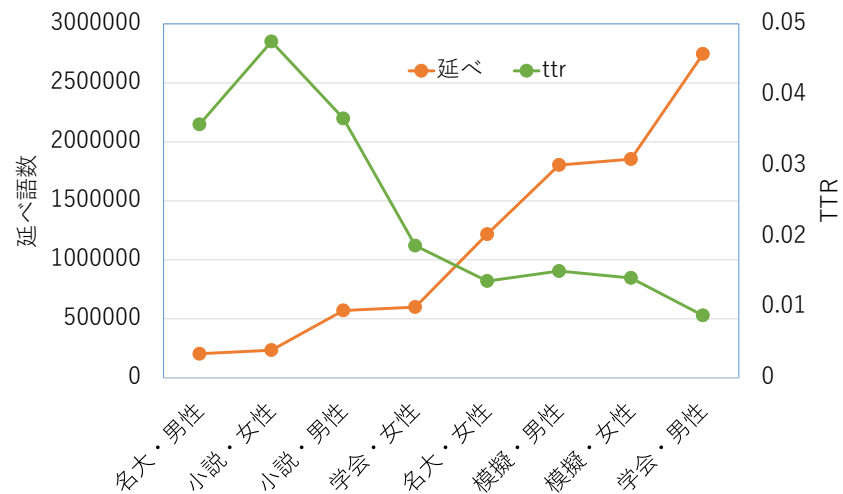
	延べ	異なり	TTR(異なり/ 延べ)
若年層	104,031	6,558	0.0630
成年層	647,690	21,534	0.0332
老年層	47,674	5,109	0.1072
(未記入)	60,985	5,736	0.0941
全体	860,380	24,525	0.0285

- 数値は2016年8月時点での作業途中の667サンプルに基づく(以下同じ)。
- 「(未記入)」は、何も記入していないもの。

24

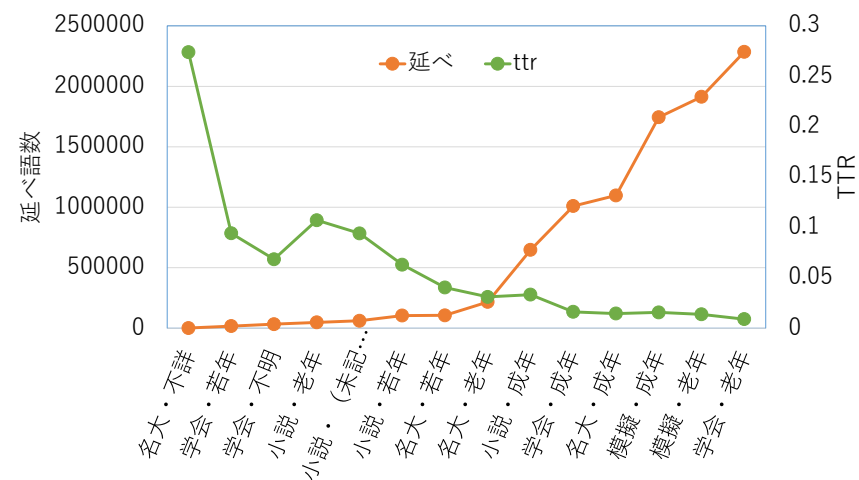
3.7 TTRの分布

・性別



3.8 TTRの分布

・年代



3.9 品詞の分布

	学会講演	模擬講演	名大	小説会話
名詞	25.77	19.86	17.18	22.45
代名詞	1.74	2.35	4.22	3.78
形容詞	0.98	1.79	3.13	2.10
形状詞	1.74	1.60	1.10	1.14
動詞	13.13	13.27	11.53	14.19
副詞	2.08	3.85	5.08	2.95
連体詞	1.41	1.13	1.30	1.31
接続詞	1.22	1.05	0.59	0.33
感動詞	6.91	5.70	7.95	0.93
接頭辞	0.60	0.56	0.55	0.88
接尾辞	3.13	2.60	1.67	2.54
助詞	29.11	32.23	33.02	33.73
助動詞	10.84	12.79	12.59	13.59
言いよどみ	1.33	1.23	0.00	0.01
その他	0.00	0.00	0.10	0.07

4.1 上位語の比較

	学会講演	模擬講演	名大	小説会話
1	の:ノ:助詞	の:ノ:助詞	だ:ダ:助動詞	だ:ダ:助動詞
2	と:ト:助詞	て:テ:助詞	うん:ウン:感動詞-一般	て:テ:助詞-接続助詞
3	で:デ:助詞	だ:ダ:助動詞	た:タ:助動詞	の:ノ:助詞-格助詞
4	だ:ダ:助動詞	と:ト:助詞	て:テ:助詞-接続助詞	は:ハ:助詞-係助詞
5	に:ニ:助詞	に:ニ:助詞	ね:ネ:助詞-終助詞	た:タ:助動詞
6	ます:マス:助動詞	が:ガ:助詞	の:ノ:助詞-準体助詞	に:ニ:助詞-格助詞
7	が:ガ:助詞	た:タ:助動詞	か:カ:助詞-副助詞	を:ヲ:助詞-格助詞
8	を:ヲ:助詞	は:ハ:助詞	と:ト:助詞-格助詞	が:ガ:助詞-格助詞
9	は:ハ:助詞	です:デス:助動詞	で:デ:助詞-格助詞	の:ノ:助詞-準体助詞
10	為る:スル:動詞	言う:イウ:動詞	の:ノ:助詞-格助詞	為る:スル:動詞-非自立可能
11	えー:エー:感動詞	も:モ:助詞	も:モ:助詞-係助詞	と:ト:助詞-格助詞
12	言う:イウ:動詞	を:ヲ:助詞	が:ガ:助詞-格助詞	も:モ:助詞-係助詞
13	た:タ:助動詞	ます:マス:助動詞	に:ニ:助詞-格助詞	です:デス:助動詞
14	です:デス:助動詞	為る:スル:動詞	は:ハ:助詞-係助詞	で:デ:助詞-格助詞
15	:::言いよどみ	か:カ:助詞	何:ナニ:代名詞	ます:マス:助動詞

4.2 上位語の比較

- 使用頻度1位
 - 「の」:CSJ学会講演、CSJ模擬講演
 - 「だ」:名大会話コーパス、BCCWJ小説会話文
- フィルター
 - CSJ学会講演に「えー」が現れる。(模擬講演20位「えー」、21位「まー」)
- 丁寧語「です」「ます」
 - 名大会話コーパス以外に上位15に登場。
 - 名大では「です」24位、「ます」56位。

29

5. 特徴語の比較

- LLR(対数尤度比)
- 特徴語を抽出する手法

30

5.1 特徴語:CSJ、名大、小説会話

•学会講演

語彙素	ヨミ	品詞	学会講演	その他	LLR
えー	エー	感動詞	80933	45279	40945.32
え	エ	感動詞	29441	9211	25892.68
ます	マス	助動詞	89311	78048	20834.31
此の	コノ	連体詞	23160	11867	13039.3
えーと	エート	感動詞	18118	8706	10959.24
的	テキ	接尾辞	14064	6049	9521.955
行なう	オコナウ	動詞	6305	652	9127.97
音声	オンセイ	名詞	4832	94	9019.45
場合	バアイ	名詞	7664	1678	8350.769
結果	ケッカ	名詞	5518	637	7740.308
用いる	モチイル	動詞	4024	66	7599.502
単語	タンゴ	名詞	3985	84	7393.45
抛る	ヨル	動詞	7300	1844	7359.228
データー	データー	名詞	3745	78	6954.646
モデル	モデル	名詞	3810	101	6931.507

31

5.2 特徴語:CSJ、名大、小説会話

•模擬講演

語彙素	ヨミ	品詞	模擬講演	その他	LLR
あの	アノ	感動詞	53090	28209	22197.53
です	デス	助動詞	85953	69826	15789.45
も	モ	助詞	69136	61037	9984.548
まー	マー	感動詞	43604	33308	9357.526
私	ワタシ	代名詞	12762	4559	8375.515
たり	タリ	助詞	8325	2782	5793.16
ん	ン	感動詞	10556	4710	5517.658
けれど	ケレド	助詞	30001	24363	5516.598
自分	ジブン	名詞	7812	2820	5078.253
矢張り	ヤハリ	副詞	9319	4059	5006.484
思う	オモウ	動詞	20723	15134	4932.581
もう	モウ	副詞	14188	9708	3874.518
迎も	トテモ	副詞	2972	447	3333.011
年	ネン	接尾辞	5504	2346	3028.74
う	ウ	助動詞	6532	3586	2604.997

32

5.3 特徴語:CSJ、名大、小説会話

・名大会話コーパス

語彙素	ヨミ	品詞	名大	その他	LLR
うん	ウン	感動詞	41016	1964	138641.1
さ	サ	助詞	8260	809	25817.28
あっ	アッ	感動詞	6557	375	21818.07
よ	ヨ	助詞	14801	13278	21122.65
ううん	ウウン	感動詞	4689	57	17000.83
ふん	フン	感動詞	3843	94	13567.35
ああ	アア	感動詞	4115	499	12448.97
零	レイ	名詞	3268	977	8010.52
じゃん	ジャン	助詞	2252	84	7756.783
然う然う	ソウソウ	感動詞	1903	30	6843.065
へえ	ヘエ	感動詞	2000	93	6776.581
彼の	アノ	連体詞	3501	2129	6380.836
まあ	マア	副詞	2246	445	6164.225
えっ	エッ	感動詞	2046	272	6093.326
はい	ハイ	感動詞	2610	1403	5067.958

33

5.4 特徴語:CSJ、名大、小説会話

・BCCWJ小説会話文

語彙素	ヨミ	品詞	名大	その他	LLR
わ	ワ	助詞	1767	288	6796.611
貴方	アナタ	代名詞	1798	1037	5031.924
俺	オレ	代名詞	1453	935	3896.962
君	キミ	代名詞	880	189	3225.76
さ	サ	助詞	753	56	3185.989
御前	オマエ	代名詞	918	679	2321.231
ああ	アア	感動詞	493	6	2281.519
様	サマ	接尾辞	951	898	2137.09
男	オトコ	名詞	670	441	1780.514
殺す	コロス	動詞	525	168	1762.593
度	ド	名詞	378	6	1737.599
女	オンナ	名詞	610	350	1710.47
年	ネン	名詞	488	186	1563.614
あっ	アッ	感動詞	351	24	1496.091
い	イ	助詞	353	32	1465.074

34

5.5 特徴語まとめ

・CSJ学会講演:

- ・「えー」「え」「えーと」などのフィラーが特徴的に表れる。
- ・「ます」

・CSJ模擬講演

- ・「あの」「まー」などのフィラーが特徴的。
- ・人称代名詞「私」、自分。
- ・助詞「も」「たり」「けれど」

・名大会話コーパス:

- ・フィラーが多い。「うん」「あっ」「ううん」「ふん」
- ・助詞「さ」「よ」「じゃん」

・BCCWJ小説会話文

- ・代名詞「あなた」「俺」「君(きみ)」「おまえ」
- ・助詞「わ」「さ」「い」

35

5.6 特徴語:名大・年代

・若年層(10代以下)

- ・LLRの値10以上112語(そのうち頻度10以上17語)。

- ・超、太い、似合う、服、目茶(副詞)、モデル
- ・プリクラ、ジーパン、不明、マミー、やや(副詞)
- ・うざい、髪型、パーマ(13)、ばい(終助詞)
- ・パイ、メイク

・ファッションに関する語

- ・「超」「うざい」などの若者語

36

5.7 特徴語:名大・年代

• 成年層(20代～50代)

• LLRの値10以上289語(すべて頻度10以上)

- か(副助詞)、ね(終助詞)、何(代名詞)、あの(代名詞)
- わけ、わたし、の(終助詞)、わ(終助詞)、あの(感動詞)
- それ(代名詞)、か(終助詞)、これ(代名詞)、お(接頭辞)
- と(格助詞)、から(接続助詞)、感じ、さ(終助詞)
- さん(接尾辞)、も(係助詞)、ほら(感動詞)、うん(感動詞)
- へ(格助詞)、なんて(副助詞)、な(終助詞)、や(助動詞)

• 一般的な機能語が上位に来る。

37

5.8 特徴語:名大・年代

• 老年層(60代以上)

• LLRの値10以上292語(そのうち頻度10以上78語)

- ていっしやる(助動詞)、俳句、いっしやる、何しろ
- コウタロウ(固有名詞)、随分、亡くなる、この頃
- 賞、種(たね)、定義、おっしやる、戦争、マップ
- メタファー、雛(ひな)、草履、ドル、孫、解釈、宅
- ササカワ(固有名詞)、デンバー、ノーベル、東急
 - ※コウタロウ=里見浩太郎(1ファイルにのみ出現)

• 敬語が上位に来る。

• 「何しろ」「随分」などの特定の副詞

38

5.9 特徴語:名大・性別

• 女性

• LLRの値10以上216語(そのうち頻度10以上212語)(空白、補助記号、記号を除く)

- うん(感動詞)、わたくし(代名詞)、です(助動詞)
- ええ(感動詞)、はい(感動詞)、とる(助動詞)
- お(接頭辞)、ず(助動詞)、わたし(代名詞)
- な(終助詞)、そう(副詞)、すごい、うん(感動詞)
- か(副助詞)、ます(助動詞)、あっ(感動詞)
- ああ(感動詞)、も(係助詞)、まあ(副詞)、もう(副詞)

• 感動詞や機能語が上位に来る。

39

5.10 特徴語:名大・性別

• 男性

• LLRの値10以上221語(そのうち頻度10以上15語)(空白、補助記号、記号を除く)

- 俺、僕、おまえ、食う、あいつ、オペラ、スペイン
- ぜ(終助詞)、要(よう)、イン、畳(じょう)、旅
- ストーリー、喉(のど)、ハンバーグ

• 「俺」「僕」「あいつ」の使用

• 「食う」「ぜ」の使用

40

5.11 特徴語: BCCWJ小説・性別

・女性

- ・LLRの値10以上141語(そのうち頻度10以上115語)(空白、補助記号、記号を除く)

- ・わ(終助詞)、の(終助詞)、わたし(代名詞)
- ・わたくし(代名詞)、かしら(終助詞)、って(副助詞)
- ・けれど(接続助詞)、あなた(代名詞)、人(ひと)
- ・もの(終助詞)、あら(感動詞)、ちゃん(接尾辞)
- ・なさる(助動詞)、ちゃう(助動詞)、ええ(感動詞)
- ・ねえ(感動詞)、パパ、ごめん、子(こ)、とても、素敵

- ・「わ」「の」「かしら」などの終助詞が上位に来る。
- ・「あら」「ええ」「ねえ」感動詞の使用。

41

5.12 特徴語: BCCWJ小説・性別

・男性

- ・LLRの値10以上239語(そのうち頻度10以上239語)(空白、補助記号、記号を除く)

- ・俺、な(終助詞)、僕、が(接続助詞)、お(接頭辞)
- ・君(きみ)、よ(終助詞)、か(終助詞)、は(係助詞)
- ・しかし(接続詞)、さん(接尾辞)、だ(助動詞)
- ・ぞ(終助詞)、おまえ、ね(終助詞)、を(格助詞)
- ・わし(代名詞)、いや(感動詞)、です(助動詞)
- ・等(ら)(接尾辞)、我々、奴(やつ)、てる(助動詞)

- ・「俺」「僕」「君」などの人称代名詞
- ・「な」「よ」「か」「ぞ」などの終助詞

42

5.13 特徴語: CSJ学会講演・性別

・女性

- ・LLRの値10以上577語(そのうち頻度10以上411語)

- ・人(じん)(接尾辞)、と(感動詞)、アクセント、
- ・リサイクル、日系、用法、天皇、乳児、女性、会話
- ・古事(こじ)、音感、アナウンス、生(なま)、フィラー
- ・フランス、母語、検査、採録、鼻(び)(接尾辞)、
- ・箇月(接尾辞)、健康、指示、感情、期(接尾辞)

- ・機能語が少ない。(LLR10以上で助詞3、助動詞2)
- ・人称代名詞がない。(LLR10以上に無し)

43

5.14 特徴語: CSJ学会講演・性別

・男性

- ・LLRの値10以上1280語(そのうち頻度10以上863語)

- ・まー(感動詞)、えー(感動詞)、居る(いる)、ね(助詞)
- ・その(感動詞)、おー(感動詞)、てる(助動詞)
- ・え(感動詞)、語(ご)(接尾辞)、あー(感動詞)
- ・者(しゃ)(接尾辞)、検索、システム、認識
- ・(言いよどみ)、モデル、うー(感動詞)、要約、刺激

- ・フィラーが上位に来る。
- ・人称代名詞が上位にない(「我々」80語、「僕」11語)

44

5.15 特徴語:CSJ模擬講演・性別

・女性

- ・LLRの値10以上869語(すべて頻度10以上)

- ・あの(感動詞)、とても、居る(いる)、わたし(代名詞)
- ・お(接頭辞)、すごい、と(感動詞)、て(助詞)、
- ・母(はは)、本当、主人、子供、た(助動詞)、も(助詞)
- ・たり(助詞)、そう(名詞)、達(接尾辞)、に(助詞)
- ・後(あと)、の(助詞)、猫、夫(おと)、って(助詞)
- ・いつ、仕舞う、思う、ハワイ、子(こ)、もう(副詞)

・「とても」の使用。

- ・講演の内容に影響を受ける。

45

5.16 特徴語:CSJ模擬講演・性別

・男性

- ・LLRの値10以上1094語(そのうち頻度10以上475語)

- ・えー(感動詞)、まー(感動詞)、おー(感動詞)、
- ・ね(助詞)、僕、わけ、非常、いー(感動詞)、ござる
- ・その(感動詞)、うー(感動詞)、これ(代名詞)
- ・つう(助動詞)、あー(感動詞)、あるいは(接続詞)
- ・です(助動詞)、え(感動詞)、言う、要する、と(助詞)
- ・野球、我々、所謂、当時、当然、やつ、選手、俺、試合

・フィラーが上位に来る。

46

5.17 コーパス内の特徴語まとめ

・名大会話コーパス

- ・若年層と老年層に特徴が見られた。

・性差

- ・名大会話コーパスとBCCWJの小説会話文とでは特徴語が異なる。(小説の終助詞「わ」「かしら」は名大では見られない)
- ・BCCWJ小説会話文はステレオタイプの。
- ・CSJ学会講演では、男性のほうがよりフィラーを使う。男女ともに人称代名詞の使用が少ない。
- ・CSJ模擬講演では、男性のほうがよりフィラーを使う。

47

6. まとめと課題

・まとめ

- ・性質の違う話し言葉の比較を語彙的な違いに基づいて概観した。

・課題

- ・データ整備(統一的に扱えるようするため)
- ・語彙的な分析を文法的な分析と結びつける
- ・レジスターの違いが言語的にどのように表れるか、現れないか。

48