

小説会話文の文末形式リストの作成

佐藤 理史 (名古屋大学大学院工学研究科)

Compilation of a Sentence Final Form List of Novel Dialogues

Satoshi Sato (Graduate School of Engineering, Nagoya University)

要旨

小説の会話文において一定の頻度が観察された 1,923 種類の文末形式のリストを作成した。このリストには、4 種類の情報が付与されている。(1) 1,065 種類に文末形式に対して「BCCWJ 小説会話文」に含まれる 276,576 文における頻度情報。(2) 933 種類の文末形式に対して「小説家になろう」に掲載されている小説から収集した会話文 8,879,279 文における頻度情報。(3) 806 種類の文末形式に対して 2 種類の会話文コーパスにおける出現率の差 (対数比)。(4) 356 種類の文末形式に対して「BCCWJ 小説会話文」を利用して算出した女性率。このリストを利用することにより、書籍体の小説とウェブ小説での文末形式の使用差や、多様な文末形式の男女の使用差を観察することができる。

1. はじめに

発話には話者の個性が投影され、それが言葉遣いに現れる。日本語の小説では、この性質を利用した「セリフの書き分け」が広く用いられている。

話者のキャラクター性 (特徴や個性) は言葉遣いの色々な側面に現れるが、日本語では、終助詞を中心とした文末形式 (文末表現) がその中心的な要素となる。よく知られているように、終助詞の使用には性差が存在する (益岡・田窪 2024, 上野ほか 2025, 小川 2006)。現代社会では、使用される終助詞の性差は縮小傾向にあるが (鈴木 2007)、小説やマンガなどでは、キャラクター性に基づく文末形式の選択的使用が積極的に利用されており、読者に登場人物の人物像を伝える一つの手段となっている。

我々日本語母語話者は、文末形式と想起されるキャラクター性の関係を直感的に理解している。その関係は、役割語研究 (金水 2003, 2014) 等により、定性的にも明らかにされている。しかしながら、話者のキャラクター性と文末形式の関連性に焦点を当てた分析は、人手による小規模なもの (遠藤 1997, 下條 2012, 朽方 2017, 安井 2024) が中心で、データによる裏付けは必ずしも十分とは言い難い。

話者のキャラクター性と会話文の文末形式の関連性を明らかにするためには、分析対象とする文末形式の範囲の決定が必要となるが、そのために必要な文末形式に関する基礎資料は不足している。このような背景により、本研究では、小説の会話文の文末形式を自動認定するシステムを作成し、これを利用して一定の頻度が観察された 1,923 種類の文末形式のリストを作成した。このリストには、4 種類の情報が付与されている。(1) 1,065 種類に文末形式に対して

表 1 文末形式リストの概要

ファイル名	KohakuFFL.20250716.tsv	
ファイル形式	TSV 形式	
行数	1,923 行	
各行のフィールド数	16 フィールド (F00–F15)	
F00	通し番号	0000 – 1922
F01	文末形式タイプ	文末形式の大分類
F02	文末形式 ID	
F03	終助詞 ID	終助詞等
F04	接続助詞 ID	接続助詞等
F05	丁寧表現 ID	「ます・です」など
F06	特殊表現 ID	「のだ・だろう・じゃない」など
F07	BCCWJ 頻度	「BCCWJ 会話文コーパス」における頻度 (回数)
F08	BCCWJ 出現率	「BCCWJ 会話文コーパス」における出現率 (万分率)
F09	なろう頻度	「なろう会話文コーパス」における頻度 (回数)
F10	なろう出現率	「なろう会話文コーパス」における出現率 (万分率)
F11	$\log_2(F10/F08)$	2 種類の会話文コーパスにおける出現率の対数比
F12	女性頻度	「BCCWJ 会話文コーパス」における女性話者の頻度 (回数)
F13	男性頻度	「BCCWJ 会話文コーパス」における男性話者の頻度 (回数)
F14	その他頻度	「BCCWJ 会話文コーパス」における女性・男性以外の頻度 (回数)
F15	女性率	「BCCWJ 会話文コーパス」に基づく女性率 (%)

「BCCWJ 小説会話文」に含まれる 276,576 文における頻度情報。(2) 933 種類の文末形式に対して「小説家になろう」に掲載されている小説から収集した会話文 8,879,279 文における頻度情報。(3) 806 種類の文末形式に対して 2 種類の会話文コーパスにおける出現率の差 (対数比)。(4) 356 種類の文末形式に対して「BCCWJ 小説会話文」を利用して算出した女性率。このリストを利用することにより、書籍体の小説とウェブ小説での文末形式の使用差や、多様な文末形式の男女の使用差を観察することができる。

2. 文末形式リストの形式

作成した文末形式リストの全体像を表 1 に示す。このデータは 1,923 行から構成されており、各行は 16 のフィールドから構成されている。16 フィールドのうち、主要なフィールドはボールド体で示した 6 つのフィールドである。作成した文末形式リストの一部を表 2 に示す。

3. 文末形式リストの作成

文末形式リストの作成には、文末形式認定システム Kohaku(佐藤 2025) を用いた。

3.1 文末形式認定システム Kohaku

文末形式認定システム Kohaku は、与えられた会話文を Sudachi (Takaoka et al. 2018) を用いて形態素解析した後、その出力に文末形式の構成要素を検出するルール群を適用し、文末形式を決定する。このシステムは、文末形式と話者のキャラクター性の関係の解明を最終目的に、その調査を容易にするためのツールとして開発した。そのため、次のような特徴を持つ。

- 終助詞は長い単位で認定する。短単位の終助詞の連続は、一定の頻度が観察される場合は、ひとまとまりで終助詞と認定する。たとえば、「わよ」、「わね」、「かい」、「だい」等

表 2 文末形式リストの一部

スペースの都合上、主要なフィールドのみ示す。値がないフィールドは「-」で表されている。

F00	F01	F02	F08	F10	F11	F15
0000	-	-	1911.01	2633.32	0.46	57.3
0001	裸	V	553.81	496.58	-0.16	29.9
0002	裸	A	408.78	353.23	-0.21	33.9
0003	裸	だ	384.89	360.95	-0.09	7.1
0004	裸	Vタ	313.91	200.48	-0.65	33.7
0005	丁	ます	270.81	308.86	0.19	46.9
0006	特	んだ	253.46	127.08	-1.00	10.6
0007	丁	です	247.56	240.85	-0.04	42.5
0008	終	の	230.86	133.37	-0.79	92.7
0009	終	か	209.24	313.83	0.58	17.4
0010	終	よ	189.86	169.34	-0.16	39.9
0011	終	わ	186.96	92.96	-1.01	98.0
0012	丁	ました	162.49	133.30	-0.29	48.9
0013	終	な	156.27	240.25	0.62	10.0
0014	裸	V命	142.02	156.09	0.14	15.6
0990	特接終	んだろう. けど. な	0.11	0.10	-0.05	-
0991	特終	じゃない. のお	0.11	0.10	-0.07	-
0992	特接終	んだ. が. よ	0.11	-	-	-
0993	接終	で. の	0.11	-	-	-
0994	特丁接	じゃない. んです. けど	0.11	-	-	-
0995	特終	んだった. かな	0.11	-	-	-
0996	丁終	のでしょう. かね	0.11	-	-	-
0997	丁	なさって	0.11	-	-	-
0998	特終	じゃねえ. ぜ	0.11	-	-	-
0999	特接終	だろう. し. なあ	0.11	-	-	-
1178	丁	でしゅ	-	0.17	-	-
1179	特	じゃ無い	-	0.16	-	-
1180	丁終	のです. Pわね	-	0.15	-	-
1181	特接終	じゃ. が. な	-	0.15	-	-
1182	終	な. の	-	0.13	-	-
1183	特接終	じゃ. から. のう	-	0.12	-	-
1184	丁終	っす. Pわ	-	0.12	-	-
1185	終	わあん	-	0.12	-	-
1186	接終	けど. よお	-	0.12	-	-
1187	特接終	んじゃ. が. な	-	0.11	-	-
1188	丁終	ましよう. かあ	-	0.11	-	-
1189	接終	のに. さあ	-	0.11	-	-
1190	終	Bな. の	-	0.10	-	-
1191	丁終	んっす. よ	-	0.10	-	-
1192	特丁終	じゃない. っす. よ	-	0.10	-	-

を、それぞれ異なる終助詞として認定する。

- 一部の終助詞は、前方の文脈付きで認定する。たとえば、「だ」に後続する「ね」は「Dね」、名詞等に後続する「ね」は「Xね」として、区別する。
- 接続助詞を文末形式の構成要素として認定する。
- 「です・ます」等の丁寧表現を文末形式の構成要素として認定する。

- 会話文の文末に頻出する「のだ」、「んだ」、「だろう」、「だろ」、「じゃない」などを文末形式の構成要素として認定する。
- 一部の異形を正規化する。たとえば、「ねえ」、「ねエ」、「ねー」は、「ねえ」に正規化する。

Kohaku が認定する文末形式は、次の 5 種類の要素から構成される。

1. 終助詞 (F03)
2. 接続助詞 (F04)
3. 丁寧表現 (F05)
4. 特殊表現 (F06)
5. 主要素形式

最後の主要素形式を除く 4 種類の要素を**付属要素**と呼ぶ。付属要素は、以下に示す出現順序制約を満たす。

特殊表現 < 丁寧表現 < 接続助詞 < 終助詞 (1)

文末形式タイプ (F01) とは、出現した付属要素の種別名の最初の 1 文字を繋げたもので、文末形式の大分類に相当する。文末に付属要素が全く出現しない場合は、文末の形態素が述語か否かを判定し、述語の場合は文末形式タイプ「裸」、述語以外の場合はタイプ「なし (‘-’)」とする。

文末形式 ID (F02) とは、付属要素を持つ場合は付属要素の ID をドット (‘.’) でつないだ文字列である。付属要素を持たない場合は、略記した述語の形式 (主要素形式 ID⁽¹⁾) ある。

文末形式認定システムが認定する**文末形式**とは、**文末形式タイプ** (F01) と**文末形式 ID** (F02) の組である⁽²⁾。

Kohaku は、形態素パターンを利用したルールベースなので、認定する文末形式の総種類数は不明である。現時点までに認定を確認した文末形式の種類は、5,246 種類である。

3.2 「BCCWJ 小説会話文」に対する文末形式の付与

「BCCWJ 小説会話文」とは、『現代日本語書き言葉均衡コーパス (BCCWJ)』の図書館サブコーパス (LB) の小説サンプルに含まれる会話文に、話者情報を付与したデータ (山崎ほか 2022) を指す。このデータから、ほぼ確実に会話文と思われる 276,576 文を抜き出し、これらの文に Kohaku を用いて文末形式を自動付与した。以降では、この会話文データを「BCCWJ 会話文コーパス (Kohaku-BCCWJ)」と呼ぶ。

表 3 に、「BCCWJ 会話文コーパス」における文末形式タイプの頻度と文末形式 ID の種類数を示す。この表に示すように、総計で 1,916 種類の文末形式が付与された。そのうち、1 万文に 1 回 (1bp) 以上出現した文末形式は 355 種類、0.1bp 以上出現した形式は 1,065 種類であった。

3.3 「小説家になろう会話文」に対する文末形式の付与

「小説家になろう会話文」とは、ウェブサイト「小説家になろう」に掲載されている小説から収集した 5,297,896 セリフ (川北ほか 2024) を文に分割した 8,879,279 文の会話文データを指

⁽¹⁾ 主要素形式 ID の詳細を付録 A に示す。

⁽²⁾ 文末形式が F01 と F02 の組である理由は、付録 B を参照のこと。

表3 「BCCWJ 会話文コーパス」の文末形式の概要

タイプ	頻度 (計)	種類数							
(頻度下限)		2,766	1,383	277	139	28	14	3	1
(出現率下限)		100bp	50bp	10bp	5bp	1bp	0.5bp	0.1bp	
終	62,006	6	12	34	41	80	98	155	194
接終	3,715	0	0	2	6	23	39	93	153
丁終	20,436	0	4	17	26	71	109	235	419
特終	8,343	0	1	5	13	38	73	183	372
丁接終	1,258	0	0	0	1	11	19	52	134
特接終	806	0	0	0	0	7	12	47	125
特丁終	519	0	0	0	1	3	7	19	72
特丁接終	5	0	0	0	0	0	0	0	5
接	7,027	0	1	9	10	17	20	25	26
丁接	3,466	0	0	2	8	25	33	71	113
特接	1,898	0	0	1	4	9	14	35	86
特丁接	22	0	0	0	0	0	0	4	12
丁	37,992	5	7	15	20	27	42	70	93
特丁	165	0	0	0	0	2	3	9	23
特	14,751	1	2	7	12	21	25	40	58
裸	61,313	6	8	13	16	21	24	27	31
-	52,854	-	-	-	-	-	-	-	-
合計	276,576	18	35	105	158	355	518	1,065	1,916

す。これらの文に Kohaku を用いて文末形式を自動付与した。以降では、この会話文データを「なろう会話文コーパス (Kohaku-Narou)」と呼ぶ。

表4に、「なろう会話文コーパス」における文末形式タイプの頻度と文末形式IDの種類数を示す。この表に示すように、総計で5,075種類の文末形式が付与された。そのうち、1bp以上出現した文末形式は373種類、0.1bp以上出現した文末形式は932種類であった。

3.4 収録する文末形式の決定

それぞれの会話文コーパスで、ある閾値よりも出現率の高い文末形式を持つ会話文が、全体の会話文のどのくらいの割合を占めるか(カバー率)を調べた。その結果を表5に示す⁽³⁾。この表に示すように、1bp以上の文末形式で、会話文のカバー率は97%以上、0.1bp以上で99%以上となる。

いずれかのコーパスで0.1bp以上出現する文末形式は、1,192種類存在する。これに、文末形式タイプが「なし(‘-’)」を含めた1,193種類を、以降では、便宜的に文末形式と呼ぶ。この1,193種類を、今回作成する文末形式リストの収録対象とし、そのそれぞれに、頻度情報(F07-F10)を付与した⁽⁴⁾。

2種類の会話文コーパスのいずれにおいても、0.1bp以上の出現する文末形式は806種類ある。806種類の各文末形式 f に対して、2種類の会話文コーパスにおける出現率の違いを、以

⁽³⁾ この表では、文末形式タイプ「なし」も1種類の文末形式として含めている。括弧内の数字は、分母と分子のいずれにおいても文末タイプ「なし」を除いた数で計算したカバー率である。

⁽⁴⁾ F07とF09が1以上の場合でも、出現率が0.1bp未満の場合は、F08とF10の値を‘-’とした。

表4 「なろう会話文コーパス」の文末形式の概要

タイプ	頻度 (計)	種類数							
(頻度下限)		88,793	44,397	8,880	4,440	888	444	89	1
(出現率下限)		100bp	50bp	10bp	5bp	1bp	0.5bp	0.1bp	
終	1,932,591	6	12	33	48	85	100	136	427
接終	148,194	0	0	2	7	24	42	80	381
丁終	683,844	0	2	16	30	79	112	218	1,073
特終	209,684	0	0	4	10	41	65	138	1,022
丁接終	28,480	0	0	0	1	6	12	41	340
特接終	29,820	0	0	0	0	9	17	47	512
特丁終	16,852	0	0	0	1	2	5	19	354
特丁接終	301	0	0	0	0	0	0	1	46
接	230,838	0	1	8	11	15	18	21	34
丁接	117,892	0	0	2	9	29	37	71	215
特接	60,157	0	0	1	4	12	16	34	205
特丁接	664	0	0	0	0	0	0	1	78
丁	1,038,269	3	5	15	20	31	38	54	158
特丁	2,326	0	0	0	0	0	2	5	90
特	330,998	1	2	7	11	18	24	39	97
裸	1,710,167	5	7	13	16	22	23	27	43
-	2,338,202	-	-	-	-	-	-	-	-
合計	8,879,279	15	29	101	168	373	511	932	5,075

表5 文末形式の頻度上位とカバー率

種類数は、文末形式タイプ「なし」を含む 括弧内の数字は、文末形式タイプを持つ会話文のみに限定した場合のカバー率					
出現率下限	BCCWJ 会話文			なろう会話文	
	種類数	カバー率		種類数	カバー率
10bp	106	88.8938	(86.2700)	102	88.0692 (83.8043)
5bp	159	92.5323	(90.7680)	169	93.0251 (90.5318)
1bp	356	97.2785	(96.6356)	374	97.5716 (96.7036)
0.5bp	519	98.3665	(97.9805)	512	98.5563 (98.0402)
0.1bp	1,066	99.6099	(99.5177)	933	99.5580 (99.4000)

下の式に従って算出した。

$$F11(f) = \log_2 \left(\frac{F10(f)}{F08(f)} \right) \quad (2)$$

F11 の算出では、BCCWJ 出現率 (F08) を分母とした。すなわち、F11 が正の値の文末形式は、「なろう会話文コーパス」の方により多く現れ、負の値の文末形式は、「BCCWJ 会話文コーパス」の方により多く現れることを意味する。

3.5 女性率の算出

「BCCWJ 会話文コーパス」の元となった「BCCWJ 小説会話文」には、話者の性別データが含まれている。これを利用して、文末形式の女性率を計算した。

「BCCWJ 会話文コーパス」における話者の性別データの分布を表6に示す。この表の「それ以外」は、性別の値が「女」または「男」以外のものを指す。

表 6 「BCCWJ 会話文コーパス」における話者の性別の分布

話者の性別	頻度	%
女	83,288	30.11
男	187,444	67.77
それ以外	5,844	2.11

表 7 2 種類のコーパスの出現率の対数比 (F11) の分布

F11		種類数			
上限	下限	10bp	1bp	0.5bp	0.1bp
	2	0	0	6	33
2	1	1	29	45	105
1	0	32	126	170	269
0	-1	56	122	171	257
-1	-2	13	46	69	102
-2		4	31	39	40
		106	354	500	806

この表に示すように、対象とする会話文の性別分布には大きな偏りがある。このため、この偏りを補正してそれぞれの文末形式 f の女性率 ($F15(f)$) を計算する。具体的には、以下の式を用いる。ここで F12 は女性頻度、F13 は男性頻度である。

$$F15(f) = \frac{F12(f) \times r}{F12(f) \times r + F13(f)} \quad (3)$$

$$r = \frac{\sum_f F13(f)}{\sum_f F12(f)} = \frac{187444}{83228} \approx 2.2506 \quad (4)$$

頻度が低い文末形式に対しては、女性率を計算しても信頼できない値となる。そのため、女性率の計算対象は、1bp 以上 (頻度 28 以上) の 356(=355+1) 種類の文末形式に限定した。

4. 2 種類の会話文コーパスにおける文末形式の使用差

作成した文末形式リストでは、806 種類の文末形式に対し、2 種類の会話文コーパスにおける出現率の対数比 (F11) が付与されている。本節では、F11 の分布と、2 種類のコーパスの出現率の差が大きい文末形式を示す。

表 7 に、F11 の値の分布を示す。この表に示すように、F11 の値が付与された 806 種類の文末形式のうち、F11 の値が 2 以上の文末形式が 33 種類、-2 以下の文末形式が 40 種類ある。これらの文末形式の一覧を表 8 に示す。なお、F11 の値が 2 の場合、「なろう会話文コーパス」の出現率が「BCCWJ 会話文コーパス」の出現率の 4 倍であることを意味する。

表 8 の左半分は、相対的に「なろう会話文コーパス」によく現れる文末形式である。目につくのは、次の文末形式である。なお、以下では文末形式を〈文末形式 ID〉の記法で示す。

- 「っす」を構成要素とする文末形式
〈っす〉、〈っす. ね〉、〈っす. か〉、〈じゃない. っす. か〉
- 長音化した「ね、よ、な、か」を内在する終助詞
〈 D よねえ 〉、〈 んだ. D よねえ 〉、〈 です. よねえ 〉、〈 し. ねえ 〉、〈 D よなあ 〉、

表 8 いずれかのコーパスに偏る文末形式

「なろう会話文コーパス」に多い					「BCCWJ 会話文コーパス」に多い				
タイプ	ID	F08	F10	F11	タイプ	ID	F08	F10	F11
丁	っす	0.87	13.18	3.93	特	のや	2.31	0.16	-3.81
丁終	っす. ね	0.29	4.24	3.87	丁終	んです. もの	4.01	0.39	-3.38
丁	ましたー	0.11	1.37	3.66	終	かえ	1.52	0.16	-3.28
丁終	っす. か	0.33	3.52	3.44	特接終	んだ. が. ね	2.60	0.32	-3.02
終	Dよねえ	0.14	1.44	3.32	特終	んじゃない. かね	0.72	0.10	-2.80
特接	のだ. けれど	0.11	1.08	3.31	終	らあ	1.23	0.18	-2.79
終	I よお	0.11	1.01	3.22	丁終	でして. T ね	2.13	0.33	-2.69
丁終	です. よお	0.25	2.12	3.07	丁	なされ	1.01	0.16	-2.68
終	D よなあ	0.22	1.71	2.98	特丁終	んじゃない. でしょう. ね	0.72	0.11	-2.65
特丁終	じゃない. っす. か	0.11	0.79	2.87	丁接終	んです. から. ね	2.39	0.39	-2.63
接終	し. なあ	0.18	1.17	2.70	終	え	1.84	0.30	-2.62
丁接終	です. し. ね	0.14	0.93	2.69	丁接終	んです. が. ね	2.86	0.48	-2.58
丁接	です. ので	0.33	2.01	2.63	特丁	じゃない. んです	1.88	0.32	-2.56
丁接	です. し	0.83	5.10	2.62	特終	んだ. わ	5.28	0.91	-2.53
丁終	っす. よ	0.76	4.63	2.61	丁接	んです. って	5.93	1.05	-2.50
丁終	ます. P わよ	0.22	1.30	2.58	丁終	まして. T ね	6.62	1.19	-2.47
接終	って. の	0.22	1.25	2.53	特	たろう	4.41	0.83	-2.41
接終	し. ねえ	0.18	0.95	2.40	特	んや	6.83	1.33	-2.36
特終	んだ. D よねえ	0.36	1.89	2.39	特終	じゃない. もの	0.58	0.11	-2.36
丁終	です. P わよ	0.22	1.10	2.35	特終	んだ. もの	4.30	0.85	-2.33
裸	V 仮	0.80	4.06	2.35	丁接終	ません. が. ね	1.19	0.24	-2.33
丁終	です. よねえ	0.25	1.28	2.33	丁終	んです. P わ	2.02	0.41	-2.32
丁	っしょ	0.18	0.86	2.25	特終	んじゃない. のかな	0.98	0.20	-2.32
特終	んじゃない. ねえ. の	0.14	0.67	2.22	丁終	でっ. か	0.51	0.11	-2.24
特	んだった	0.11	0.49	2.19	丁	なさい. ます	0.47	0.10	-2.23
丁	でっ	0.14	0.65	2.17	特丁	じゃない. でしょう	1.37	0.30	-2.21
特終	んだ. D よなあ	0.47	2.05	2.12	裸	や	11.28	2.45	-2.20
特丁	じゃない. です	0.14	0.61	2.08	特終	やろ. か	0.87	0.19	-2.19
丁	T ください. ませ	0.47	1.99	2.08	接	ちゃ	3.76	0.83	-2.18
丁終	んす. か	0.58	2.41	2.06	接終	んで. ね	2.60	0.59	-2.14
丁終	です. かあ	0.33	1.34	2.04	特終	じゃない. のよ	2.46	0.56	-2.14
丁終	ます. よお	0.33	1.32	2.02	特接終	のだ. から. ね	0.80	0.18	-2.14
丁接	ます. し	0.76	3.07	2.02	特終	じゃない. わ	5.53	1.26	-2.13
					丁接終	のです. から. ね	0.51	0.12	-2.13
					丁接終	です. が. ね	2.10	0.48	-2.12
					丁終	でしょう. よ	1.19	0.28	-2.09
					接終	が. ね	7.48	1.78	-2.08
					終	のかね	14.14	3.36	-2.07
					丁	んです	131.50	31.87	-2.04
					終	かね	21.55	5.30	-2.02

〈んだ. D よなあ〉、〈し. なあ〉、〈 I よお〉、〈です. よお〉、〈ます. よお〉、〈です. かあ〉

- 丁寧体以後続する「わよ」

〈ます. P わよ〉、〈です. P わよ〉

「なろう会話文コーパス」の元となった小説群の大半は、いわゆるウェブ小説であり、ここ数年間で執筆されたものである。これらの小説群においては、上記のような会話文の文末形式が、より多く使われているということである。

表 9 女性率 (F15) の出現率別分布

F15		出現率の下限			
上限	下限	20bp	10bp	5bp	1bp
	90	9	12	18	41
90	80	0	2	4	17
80	66.7	9	12	15	33
66.7	50	12	19	26	59
50	33.3	17	22	35	65
33.3	20	9	17	25	47
20	10	12	13	17	43
10		7	9	19	51
		75	106	159	356

表 10 女性率 (F15) のタイプ別分布

F15		文末形式タイプ														
上限	下限	終	丁終	特終	丁	丁接	接終	特	裸	接	丁接終	特接	特接終	特丁終	特丁	-
	90	16	14	7	3	1	0	0	0	0	0	0	0	0	0	0
90	80	3	4	2	1	2	1	0	1	2	0	1	0	0	0	0
80	66.7	6	5	0	3	5	0	1	3	5	0	4	0	0	1	0
66.7	50	11	8	2	10	7	7	2	2	5	0	1	2	0	1	1
50	33.3	7	19	1	8	7	3	4	7	3	4	0	0	2	0	0
33.3	20	15	8	7	1	2	2	4	3	1	3	0	0	1	0	0
20	10	14	4	7	1	1	3	6	4	0	1	1	1	0	0	0
10		8	9	12	0	0	7	4	1	1	3	2	4	0	0	0
		80	71	38	27	25	23	21	21	17	11	9	7	3	2	1

表 8 の右半分は、相対的に「BCCWJ 会話文コーパス」によく現れる文末形式である。左半分よりも特徴が明確ではないが、強いて指摘するならば、次のようになろう。

- 関西方言と思われる文末表現

〈のや〉、〈んや〉、〈んです. Pわ〉、〈でっ. か〉、〈や〉、〈やろ. か〉

- 「がね」を含む文末形式

〈が. ね〉、〈んだ. が. ね〉、〈です. が. ね〉、〈んです. が. ね〉、〈ません. が. ね〉

- 「かね」を含む文末表現

〈かね〉、〈のかね〉、〈んじゃない. かね〉

「BCCWJ 会話文コーパス」の元となった小説群は、BCCWJ の図書館サブコーパス (LB) に含まれる小説群であり、なろう会話文コーパスと比較して、より広いジャンル、かつ、より広い年代の小説が含まれている。文末形式の偏りは、これらの違いに由来すると考えられる。

5. 文末形式の使用の性差

5.1 全体的な傾向

文末形式の使用に性差があることはよく知られている。356 種類の文末形式の女性率 (F15) の頻度別分布を表 9 に、タイプ別分布を表 10 に示す。

女性率の値 (0-100) を 3 等分し、女性優位、中性的 (性差小)、男性優位のグループに分ける

表 11 性差が著しい文末形式

女性率 90% 以上					女性率 10% 未満				
タイプ	ID	F08	F10	F15	タイプ	ID	F08	F10	F15
終	Xよね	2.75	4.55	100.0	丁終	です. かな	1.23	1.55	0.0
終	わよね	2.46	3.89	100.0	丁終	です. ぜ	1.05	0.40	0.0
終	かしらね	1.30	1.64	100.0	丁終	ました. な	1.63	1.45	0.0
丁終	んです. Pの	1.99	1.22	100.0	丁終	ます. かね	1.08	1.45	0.0
丁終	ません. Pの	1.05	0.82	100.0	丁終	ます. な	3.29	3.94	0.0
丁終	ます. Pの	1.05	1.73	100.0	丁終	んです. な	1.01	-	0.0
丁終	ました. Pの	1.63	1.05	100.0	特接終	のだ. から. な	1.19	1.12	0.0
特終	んだ. わ	5.28	0.91	99.7	特接終	んだ. が. な	1.19	2.23	0.0
特終	じゃない. わ	5.53	1.26	99.7	特終	じゃない. のか	2.02	1.92	0.0
終	わよ	27.66	22.08	99.5	終	らあ	1.23	0.18	0.0
終	かしら	18.58	17.06	99.5	丁終	です. な	8.10	9.74	1.0
丁終	ません. Pわ	3.22	3.37	99.5	特接終	んだ. から. な	3.25	1.86	2.5
終	のかしら	11.21	8.83	99.3	特終	のだ. な	3.04	3.47	2.6
丁終	んです. もの	4.01	0.39	99.2	特終	じゃねえ. か	5.93	6.48	2.8
丁終	です. Pの	2.06	2.46	99.2	特終	んじゃない. のか	2.89	2.42	2.8
終	のよね	4.74	5.75	99.0	特接終	んだ. が. ね	2.60	0.32	3.2
終	のね	21.59	10.76	99.0	特終	んだ. ぞ	6.73	7.04	3.6
特終	んじゃない. かしら	3.40	1.33	99.0	丁接終	です. が. ね	2.10	0.48	3.8
終	わねえ	3.00	2.08	98.9	特接	んだ. が	6.36	9.16	3.8
終	わね	25.74	25.47	98.7	接終	から. な	22.74	44.55	3.9
特終	じゃない. わよ	1.23	0.72	98.7	丁終	ません. な	1.88	1.57	4.3
丁終	です. もの	2.06	1.21	98.4	特終	んだ. Dよな	1.81	5.15	4.4
終	のよ	97.44	31.38	98.2	特接	だろう. が	3.58	5.04	4.5
終	わ	186.96	92.96	98.0	特	じゃねえ	1.74	1.46	4.6
特終	じゃない. のよ	2.46	0.56	98.0	特終	のだ. ぞ	1.77	2.50	4.6
丁終	です. Pわね	1.63	5.06	98.0	特終	んだろう. な	3.25	3.75	4.9
丁終	ます. Pわ	6.69	7.00	97.5	終	ぜ	34.49	32.78	4.9
丁接	んです. って	5.93	1.05	97.2	丁接終	ます. が. ね	1.59	0.40	5.1
丁終	ました. Pわ	3.07	2.76	96.1	丁終	でしょう. な	1.52	1.26	5.2
終	Xね	38.47	37.39	94.7	終	Bなよ	7.63	10.55	5.3
特終	じゃない. の	12.69	3.60	93.5	特終	だろう. な	8.97	10.11	5.4
丁	んでしょ	7.74	5.38	93.2	接終	し. な	2.89	9.85	5.5
丁	ましよ	1.27	1.58	93.1	特	のだ	45.95	32.11	5.6
終	の	230.86	133.37	92.7	接終	んで. ね	2.60	0.59	6.1
丁	でしょ	26.07	27.17	92.1	終	もんな	1.27	2.14	6.2
終	Xよ	86.67	54.76	91.9	丁接終	ません. が. ね	1.19	0.24	6.6
丁終	です. Pわ	8.86	11.88	91.5	終	のかね	14.14	3.36	6.8
丁終	なさい. Iよ	4.59	2.67	91.4	特終	んだ. ぜ	5.79	2.86	6.9
終	もの	4.48	2.49	91.2	裸	だ	384.89	360.95	7.1
特終	んだ. もの	4.30	0.85	91.2	接終	と. な	1.08	2.69	7.2
丁終	ましよう. ね	1.16	1.13	90.7	特	じゃない. だろう	1.08	0.59	7.2
					接終	が. な	5.31	7.03	7.5
					特終	んじゃない. か	5.28	6.99	7.6
					接終	ので. ね	1.01	0.51	7.7
					特	だろう	67.11	59.57	8.0
					特終	んだ. な	13.59	10.94	8.1
					終	よな	4.84	11.64	8.1
					終	わい	2.78	1.36	8.4
					接	が	23.03	28.09	9.2
					接終	が. ね	7.48	1.78	9.4
					終	ぞ	75.10	131.67	9.4

表 12 終助詞の末尾母音の長音化と女性率 (F15)

標準形式			長音化形式			差分
ID	F08	F15	ID	F08	F15	
わね	25.74	98.7	わねえ	3.00	98.9	0.2
Xね	38.47	94.7	Xねえ	4.23	69.6	-25.1
Tね	20.18	67.5	Tねえ	1.08	27.3	-40.2
ね	107.89	57.4	ねえ	14.82	50.5	-6.9
よ	189.86	39.9	よお	3.00	57.0	17.1
Dね	29.79	26.9	Dねえ	2.93	53.4	26.5
かな	37.60	26.4	かなあ	4.66	42.9	16.5
さ	64.32	18.3	さあ	1.88	73.9	55.6
かね	21.55	17.8	かねえ	1.16	54.1	36.3
か	209.24	17.4	かあ	2.71	56.9	39.5
Bな	36.08	11.0	Bなあ	3.00	13.0	2.0
な	156.27	10.0	なあ	20.65	27.4	17.4

と、それぞれのグループに属する文末形式の数は、91, 124, 141 となる。使用に性差が著しい (女性率が 90% 以上または 10% 未満) の文末形式の一覧を表 11 に示す。これらの文末形式の使用に性差が著しいことは、ほぼ直感と一致する。

5.2 終助詞の末尾長音化と女性率

小説の会話文では、終助詞の末尾の母音を長音化させた異形がしばしば観察される。たとえば、終助詞「ね」に対しては、「ねえ」、「ねえ」、「ねー」などがそれに相当する。Kohaku は、これらを〈ねえ〉に集約する。

表 12 に、標準形式と長音化形式の女性率の比較表を示す。長音化形式は、標準形式より柔らかい感じとなるので、長音化形式の方が女性率が高くなる (差分がプラス) と予想したが、それに反するデータが存在する。

この表において、〈Tね〉と〈Tねえ〉の組は注意が必要である。終助詞〈Tね〉には、依頼と伝達の 2 種類の意味が存在する (佐藤 2024)。一方、〈Tねえ〉は、伝達の意味のみとなるが、形態素解析誤りにより、「テない」の音変化が混入する。差が -40.2 ポイントと大きいのは、このような理由による。

この組を除くと、注目すべき組は、以下の 2 組である。

1. 〈Xね〉と〈Xねえ〉

予想に反して、長音化形式の方が女性率が低い。おそらく、〈Xね〉はほぼ女性専用の形式であるのに対し、〈Xねえ〉は柔らかい口調の男性も用いることが、-25.1 ポイントという差が現れる理由であろう。

2. 〈さ〉と〈さあ〉

長音化形式の方が女性率が高いという点では予想通りであるが、+55.6 ポイントという差は著しい。違和感があったので調査したところ、女性の〈さあ〉の 29 回の使用のうち、18 回が特定の作品⁽⁵⁾に集中しており、この作品の存在が、大きな差をもたらしてい

⁽⁵⁾ 大城立裕の『水の盛装』において、3 名の女性登場人物が、それぞれ 9 回、5 回、4 回使用している。

表 13 終助詞の前方文脈と女性率

	前方文脈						
	(なし)	X	D	T	R	B	I
ね	57.4	94.7	26.9	67.5	-	-	-
ねえ	50.5	69.6	53.4	27.3	-	-	-
よ	39.9	91.9	24.9	83.7	-	-	16.2
よな	8.1	-	14.5	-	-	-	-
よね	62.5	100.0	62.6	-	-	-	-
な	10.0	-	-	16.2	43.5	11.0	-
なあ	27.4	-	-	-	-	13.0	-
なよ	-	-	-	-	46.0	5.3	-

ることが判明した。なお、残りの 11 回の使用は、すべて異なる作品であった。

5.3 終助詞の前方文脈と女性率

Kohaku では、いくつかの終助詞を前方文脈込みで区別する⁽⁶⁾。表 13 に、それらの終助詞における前方文脈と女性率の関係を示す。

前方文脈 X (「だ」の省略) と D (「だ」が存在) のよる女性率の違いは、Xの方が女性率が高く Dの方が低いという、よく知られた事実と一致する。しかしながら、より詳細に観察すると、以下のことに気づく。

- 〈Xね〉と〈Dね〉の組、および、〈Xよ〉と〈Dよ〉の組は、女性率の差が大きい。〈Xね〉と〈Xよ〉の女性率は 90% 以上で、ほぼ女性専用の形式とみなしてよいだろう。その一方で、〈Dね〉と〈Dよ〉の女性率は 25% 前後であり、男性が優位ではあるが、女性による使用も一定数存在する。
- 〈Xよね〉と〈Dよね〉の組は、女性率の差は比較的大きいが、〈Dよね〉の女性率は 60% を超えている。〈Dよね〉の女性率は〈よね〉の女性率とほぼ同じであり、男性優位の形式とはなっていない。
- 〈Xねえ〉と〈Dねえ〉の組は、女性率の差が比較的小さい。〈Dねえ〉の女性率は 53.4% であり、男性優位の形式とはなっていない。

このように、「BCCWJ 会話文コーパス」では、〈Dよね〉と〈Dねえ〉が中性化している実態が観察できる。

5.4 複数の付属要素から構成される文末形式の女性率

Kohaku は、従来より長い単位の文末形式を認定する。このため、これまで行われなかった分析が可能となる。

表 14 は、文末形式タイプ「接終」の文末形式の女性率を、タイプ「接」および「終」の文末形式の女性率と対比して示した表である。この表を観察すると、タイプ「接終」の女性率には、以下のものがあることに気づく。

1. 終助詞の女性率と近い値をとるもの

例：〈けど.ね〉、〈けど.な〉、〈し.ね〉、〈って.ね〉、〈ば.ね〉、〈が.な〉

⁽⁶⁾ 前方文脈を表す記号の一覧を付録 C に示す。

表 14 接続助詞と終助詞の結合と女性率

接		接終		終	
けど	76.1	けど.ね	59.1	ね	57.4
		けど.さ	52.1	さ	18.3
		けど.な	11.7	な	10.0
し	72.2	し.ね	52.5	ね	57.4
		し.な	5.5	な	10.0
から	71.5	から.よ	86.1	よ	39.9
		から.ねえ	64.3	ねえ	50.5
		から.ね	37.3	ね	57.4
		から.さ	23.8	さ	18.3
		から.なあ	18.8	なあ	27.4
		から.な	3.9	な	10.0
って	63.6	って.ね	53.7	ね	57.4
		って.さ	37.5	さ	18.3
ば	60.3	ば.ね	58.2	ね	57.4
ので	49.6	ので.ね	7.7	ね	57.4
で	47.5	で.ね	35.5	ね	57.4
		で.さ	29.9	さ	18.3
		で.な	16.8	な	10.0
と	44.7	と.ね	62.8	ね	57.4
		と.な	7.2	な	10.0
んで	23.3	んで.ね	6.1	ね	57.4
が	9.2	が.ね	9.4	ね	57.4
		が.な	7.5	な	10.0

2. 接続助詞の女性率と近い値をとるもの

例：〈が.ね〉

3. 接続助詞と終助詞の女性率の中間的な値をとるもの

例：〈けど.さ〉、〈から.ねえ〉、〈って.さ〉、〈で.さ〉

4. 独自の値をとるもの

例：〈から.よ〉、〈から.ね〉、〈ので.ね〉、〈で.ね〉、〈んで.ね〉

これらのグループのうち、最後のグループを除いた3グループの女性率は、構成要素からの合成として理解できる。しかしながら、最後のグループは、接続助詞と終助詞が結合した場合に、独自の女性率を示すものである。

最後のグループに属する5つの文末形式のうち、〈から.よ〉は、「よ」を〈Xよ〉(つまり、「からだよ」の「だ」が省略されて「からよ」の形式となっている)とみなせば、第3グループに属すると解釈できる。しかし、それ以外の文末形式は独自の女性率をとるとしか解釈できない。

表 15 に、文末形式の女性率が、分解した形式の女性率の最小値または最大値から 10 ポイント以上外れるもの (43 種類) の一覧を示す。このような文末形式の存在は、終助詞のみに着目した性差の議論が不十分であり、より長い単位での検討が必要であることを明確に示している。

5.5 「の・ん」交替と女性率

会話文では、準体助詞「の」が「ん」に交替する傾向が強く、使用においても「ん」が優位である。Kohaku は、「の」形式と「ん」形式を、異なる文末形式として区別する。

表 15 付属要素の結合と女性率

F02	F03	F15	F02	F03	F15	F02	F03	F15
丁接	です. けど	83.9	丁接終	です. けど. ね	42.9 ↓	終	ね	57.4
特	んじゃない	76.0	特丁終	んじゃない. です. か	30.0 ↓	丁終	です. か	43.5
接	から	71.5	接終	から. よ	86.1 ↑	終	よ	39.9
			接終	から. ね	37.3 ↓	終	ね	57.4
丁	なさい	70.9	丁終	なさい. I よ	91.4 ↑	終	I よ	16.2
特	じゃない	63.5	特丁	じゃない. でしょう	73.5 ↑	丁	でしょう	62.7
			特終	じゃない. のか	0.0 ↓	終	のか	10.7
丁接	ました. から	63.2	丁接終	ました. から. ね	32.5 ↓	終	ね	57.4
丁	でしょう	62.7	丁終	でしょう. よ	83.2 ↑	終	よ	39.9
特接	のだ. から	61.0	特接終	のだ. から. な	0.0 ↓	終	な	10.0
丁接	ます. から	59.9	丁接終	ます. から. ね	35.4 ↓	終	ね	57.4
丁	のでしょう	58.6	丁終	のでしょう. ね	42.9 ↓	終	ね	57.4
丁接	です. から	57.7	丁接終	です. から. ね	24.5 ↓	終	ね	57.4
丁	んです	57.6	丁接	んです. って	97.2 ↑	接	って	63.6
			丁終	んです. な	0.0 ↓	終	な	10.0
丁接	ません. から	56.4	丁接終	ません. から. ね	41.6 ↓	終	ね	57.4
丁	でした	55.2	丁終	でした. ね	42.4 ↓	終	ね	57.4
丁接	んです. から	54.2	丁接終	んです. から. ね	31.4 ↓	終	ね	57.4
丁	T ください	51.0	丁終	T ください. ね	89.3 ↑	終	ね	57.4
			丁終	T ください. な	88.2 ↑	終	な	10.0
丁	ましょう	51.0	丁終	ましょう. ね	90.7 ↑	終	ね	57.4
			丁終	ましょう. よ	86.1 ↑	終	よ	39.9
			丁終	ましょう. か	63.3 ↑	終	か	17.4
丁	まして	50.9	丁終	まして. T ね	21.0 ↓	終	T ね	67.5
丁	ません	50.6	丁終	ません. ね	32.0 ↓	終	ね	57.4
接	ので	49.6	接終	ので. ね	7.7 ↓	終	ね	57.4
丁	ました	48.9	丁終	ました. よね	73.8 ↑	終	よね	62.5
			丁終	ました. ね	35.5 ↓	終	ね	57.4
			丁終	ました. な	0.0 ↓	終	な	10.0
接	で	47.5	接終	で. ね	35.5 ↓	終	ね	57.4
丁	ます	46.9	丁終	ます. ね	35.5 ↓	終	ね	57.4
			丁終	ます. かね	0.0 ↓	終	かね	17.8
			丁終	ます. な	0.0 ↓	終	な	10.0
丁	です	42.5	丁接	です. って	79.5 ↑	接	って	63.6
			丁接終	です. から. ね	24.5 ↓	接終	から. ね	37.3
			丁終	です. かな	0.0 ↓	終	かな	26.4
丁接	です. が	33.5	丁接終	です. が. ね	3.8 ↓	終	ね	57.4
丁接	んです. が	33.3	丁接終	んです. が. ね	16.0 ↓	終	ね	57.4
丁接	ません. が	29.7	丁接終	ません. が. ね	6.6 ↓	終	ね	57.4
特	じゃ	25.6	特終	じゃ. よ	15.3 ↓	終	よ	39.9
接	んで	23.3	接終	んで. ね	6.1 ↓	終	ね	57.4
丁接	ます. が	19.2	丁接終	ます. が. ね	5.1 ↓	終	ね	57.4
特	んだ	10.6	特接終	んだ. から. ね	56.6 ↑	接終	から. ね	37.3

表 16 「の」形式と「ん」形式の女性率

F02	「の」形式		「ん」形式		差分
	F03	F16	F03	F16	
丁終	のでしょう.ね	42.9	んでしょう.ね	69.7	26.8
丁	のでしょう	58.6	んでしょう	70.2	11.6
丁	のです	47.7	んです	57.6	9.9
丁終	のです.か	44.9	んです.か	53.8	8.9
特接	のだ.から	61.0	んだ.から	69.7	8.7
特	のじゃ	32.0	んじゃ	38.0	6.0
特終	のだ.な	2.6	んだ.な	8.1	5.5
丁接	のです.から	48.9	んです.から	54.2	5.3
特	のだ	5.6	んだ	10.6	5.0
丁終	のです.よ	43.4	んです.よ	47.2	3.8
特	のだろう	15.3	んだろう	18.2	2.9
丁終	のでしょう.か	57.0	んでしょう.か	59.8	2.8
特接終	のだ.から.な	0.0	んだ.から.な	2.5	2.5
丁終	のです.ね	47.3	んです.ね	49.0	1.7
特終	のだ. Dね	27.3	んだ. Dね	26.9	-0.4
特終	のだ.ぞ	4.6	んだ.ぞ	3.6	-1.0
特終	のだ. Dよ	32.8	んだ. Dよ	31.3	-1.5
丁接	のです.が	36.0	んです.が	33.3	-2.7
特	のや	36.9	んや	29.0	-7.9
特接	のだ.が	12.9	んだ.が	3.8	-9.1

表 16 に、対応するそれぞれの形式の女性率を示す。「ん」の方が「の」よりも粗野なので、「の」形式の方が女性率が高い(表の差分がマイナス)だろうという予想は、データによって否定される。実際、「ん」形式の方が「の」形式よりも女性率が高い(差分がプラスの)組が、20 組中 14 組ある。特に、〈ん. でしょう. ね〉と〈ん. でしょう〉の 2 形式は、対応する「の」形式よりも女性率が 10 ポイント以上高い。興味深い事実である。

6. おわりに

本論文では、作成した会話文の文末形式リストについて述べた。このリストには、2 種類の会話文コーパスのいずれかで、0.1bp 以上の頻度が観察された 1,923 種類の文末形式の文末形式が収録されており、以下の 4 種類の情報が付与されている。

1. 1,065 種類の文末形式に対して、「BCCWJ 会話文コーパス」の頻度情報
2. 933 種類の文末形式に対して、「なろう会話文コーパス」の頻度情報
3. 806 種類の文末形式に対して、2 種類の会話文コーパスの出現率の差 (対数比)
4. 355 種類の文末形式に対して、話者の女性率

今回作成した文末形式リストを使用して、各種の観察と分析が可能である。本論文の 4 節では 2 種類の会話文コーパスの出現率の差の観察結果を、5 節では女性率に対するいくつかの観点での観察結果を報告した。

本リストは、名古屋大学学術機関リポジトリ (<https://doi.org/10.18999/2013067>) で公開している。

表 17 主要素形式 ID

#	ID	認定条件	備考
1	V + 活用形	動詞	
2	V + 活用形	接尾辞-動詞的	
3	V + 活用形	助動詞	語彙素の末尾が「い」以外
4	A + 活用形	形容詞	
5	A + 活用形	接尾辞-形容詞的	
6	A + 活用形	助動詞	語彙素の末尾が「い」
7	だ	助動詞だ (終止形)	
8	だあ	助動詞だ (終止形)	
9	ちゃ	助動詞だ (終止形)	
10	だっタ	助動詞だ (タ形)	
11	だっ	助動詞だ (連用形-促音便)	
12	な	助動詞だ (連体形「な」)	
13	や	助動詞や (終止形)	
14	やあ	助動詞や (終止形)	
15	やっタ	助動詞や (タ形)	
16	状X	名詞-助動詞語幹	
17	状X	形状詞-一般	
18	状X	形状詞-助動詞語幹	
19	状X	接尾辞-形状詞語幹	

謝 辞

本リストの作成で使用した「BCCWJ 小説会話文」データを提供してくださった国立国語研究所の山崎誠教授に感謝します。このデータは、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)の成果です。これらのプロジェクトに感謝します。

付 録

付録 A. 主要素形式 ID

主要素形式 ID は、付属要素を持たない会話文のうち、文末が述語と認定できるもの(文末形式タイプが「裸」)の文末形式 ID を与えるものである。

主要素形式 ID は、「助動詞だ」、「助動詞や」、および、「助動詞だ」が省略されている場合を除き、活用型と活用形の組で表現する。主要素形式 ID の一覧を表 17 に、活用形を表す記号の一覧を表 18 に示す。

付録 B. 同一 ID を持つ付属要素

同一 ID の付属要素の一覧を表 19 に示す。文末形式リストの F03-F05 の情報を使用すれば、これらを区別することができる。

本システムが認定する文末形式が、文末形式タイプ (F01) と文末形式 ID (F02) の組となっているのは、表 19 に示す付属要素が存在するためである。たとえば、文末形式 ID が〈とも〉となるものは、文末形式タイプ「終」の場合と、「接」の場合が存在する。これらは、F01 に

表 18 活用形を表す記号

#	活用形	認定条件
1	幹	語幹
2	未	未然形
3	意	意志推量形
4	用	連用形
5	(なし)	終止形
6	(なし)	連体形
7	仮	假定形
8	已	已然形
9	ヨ	命令形（出現形が「～よ」）
10	命	命令形
11	タ	タ形（「助動詞た」）
12	テ	テ形（「接続助詞て」、「接続助詞で」）

表 19 同一 ID を持つ付属要素

種別	ID	説明
終助詞	とも	「終助詞とも」。例: ほんとうだとも
接続助詞	とも	「接続助詞とも」
終助詞	で	「終助詞で」。例: 食べたで
接続助詞	で	「助動詞で」の連用形の「で」
接続助詞	ちゃ	「ては」の融合形
主要素形式	ちゃ	「助動詞だ」の終止形「ちゃ」
終助詞	な	「終助詞な」
裸	な	「助動詞だ」の連体形「な」
終助詞	や	「終助詞や」
裸	や	「助動詞や」

表 20 終助詞の前方文脈

記号	前方文脈
X	名詞・代名詞・形状詞（「助動詞だ」の省略）
D	「助動詞だ」
T	いわゆるテ形
B	動詞型活用の終止形
R	動詞型活用の連用形
I	動詞型活用（五段）の命令形、または、動詞型活用（一段）の連用形
P	「です・でした・ます・ました・ません・ませんでした」

よって区別することができる。

付録 C. 終助詞の前方文脈

表 20 に、終助詞の前方文脈を表すために使用する記号 (英大文字) の一覧を示す。

文 献

益岡隆志・田窪行則 (2024). 『基礎日本語文法 第 3 版』 くろしお出版.
上野智子・定延利之・佐藤和之・野田春美 (編) (2025). 『日本語のバラエティ』 おうふう.
小川早百合 (2006). 「話しことばの終助詞の男女差の実際と意識—日本語教育での活用に向け

- て—」 日本語ジェンダー学会 (編) 『日本語とジェンダー』, pp. 39–51. ひつじ書房.
- 鈴木睦 (2007). 「言葉の男女差と日本語教育」 日本語教育, 134, pp. 48–57.
- 金水敏 (2003). 『ヴァーチャル日本語役割語の謎』 岩波書店.
- 金水敏 (編) (2014). 『〈役割語〉小辞典』 研究社.
- 遠藤織枝 (1997). 「ドラマのことば —NHK TV「レイコさんの歯医者さん」をめぐって—」 日本語学, 16:1, pp. 67–79.
- 下條正純 (2012). 「「マリア様がみてる」における女性文末辞と人物描写」 コンテンツ文化史研究, 7, pp. 12–24.
- 朽方修一 (2017). 「ライトノベルにおける女性文末形式」 ヨーロッパ日本語教育 (21), pp. 148–153.
- 安井寿枝 (2024). 「キャラクター言語に見るジェンダー意識—宮崎駿作品の特徴とは—」 日本語学, 43:1.
- 佐藤理史 (2025). 『小説会話文の文末形式の自動認定』.
Jxiv, doi: <https://doi.org/10.51094/jxiv.1401>
- Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, Yoshitaka Uchida, and Yuji Matsumoto (2018). “Sudachi: a Japanese Tokenizer for Business.” *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- 山崎誠・宮寄由美・柏野和佳子 (2022). 「小説会話文への話者情報付与」 Technical report, 国立国語研究所. <https://www2.ninjal.ac.jp/conversation/report/report05.pdf>
- 川北雄大・石川和樹・夏目和子・小川浩平・佐藤理史 (2024). 「口調弁別評価データセットの作成と口調エンコーダの評価」 情報処理学会研究報告, Vol.2024-NL-259 No.16.
- 佐藤理史 (2024). 「小説のセリフの書き分けに使われる文末形式」 言語資源ワークショップ 2024 発表論文集, pp. 464–493.