

小説会話文の文末形式の自動認定

佐藤理史

名古屋大学大学院工学研究科

sato.satoshi.g9@f.mail.nagoya-u.ac.jp

概要

小説の会話文の文末形式を自動認定するシステム Kohaku を作成した。このシステムは、形態素解析結果に対して、文末の終助詞、接続助詞、丁寧表現、特殊表現、主要素形式の 5 種類の構成要素を規則に基づいて認定し、最終的に文末形式を決定する。システム開発には約 915 万文の会話文を用い、どのような範囲の形式を文末形式に含めるか、出現形をどこまで区別し、どこから正規化するかを具体例に基づいて決定し、構成要素の認定規則集合を定めた。このシステムを使用することにより、小説の登場人物のセリフがどのように書き分けられているかを定量的に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・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、どのような終助詞を好んで使用するかには、性差が存在する (益岡・田窪 2024, 上野・定延・佐藤・野田 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, 金水 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。その理由のひとつは、コンピュータによる分析支援ツールが存在しないことにあると思われる。

既存の形態素解析器を用いて会話文を解析することは、分析の準備としては一定の価値がある。しかしながら、次節で指摘するように、分析対象の文末形式は終助詞のみに限定できないこと、終助詞の使用実態を把握する場合でも短単位の終助詞は必ずしも適切ではないこと、などの理由により、形態素解析結果をそのまま機械処理して頻度等を計測するだけでは、必ずしも有意義な調査・分析ができるとは限らない。

このような状況を打開するために、以下のような特徴を持つ文末形式認定システムを作成した。

1. システムは、与えられた会話文の末尾を調べて文末形式を認定する。
2. 文末形式を構成する要素として、終助詞、接続助詞、丁寧表現、特殊表現、主要素形式の5種類を設定し、それぞれに対する認定条件を規則によって定義する。
3. 終助詞は、長い単位で認定する。
4. いくつかの終助詞は、前方文脈の違いを考慮して区別する。

本稿では、まず2節で文末形式の自動認定における課題を整理する。次に3節で、作成した文末形式認定システムの概要を示し、続く4節と5節、および、付録でその詳細を示す。6節では、本システム開発の副産物として作成した「小説会話文文末形式リスト Kohaku-FFL」の概要を示す。7節では、本システムを利用した、『図書館戦争』シリーズにおけるセリフの書き分けの調査結果を示す。最後の8節で、本稿をまとめる。

2 文末形式の自動認定における課題

文末形式を自動認定するためには、まずもって「文末形式」を形式的に定義する必要がある。ここで、検討すべき課題は、以下の2点である。

- どのような範囲の形式(表現)を、文末形式に含めるか？
- 出現形をどこまで区別し、どこから正規化(集約)するか？

以下に示す例(1a)は、ある女性キャラクターの会話文^{*1}であるが、話し手が異なれば、例(1b)–(1d)の会話文となっても不自然ではない。

- (1) a. 出迎えてくれるなんて嬉しいわね
b. 出迎えてくれるなんて嬉しいよね
c. 出迎えてくれるなんて嬉しいよな
d. 出迎えてくれるなんて嬉しいじゃない

上記の例文(1a)–(1c)において、話し手のキャラクター性を感じさせるのは、「わよ・よね・よな」であり、これらを文末形式として採用するのは自然であろう。これらの形式を採用するのであれば、これらと交換可能である(1d)の「じゃない」も、文末形式に含めるのが妥当であろう。このように、文末形式として採用すべき形式は、短単位の終助詞ではなく、より長い単位であり、かつ、終助詞以外にも範囲を広げる必要がある。

では、どこまで広げる必要があるのだろうか。

話者のキャラクター性を考える上で考慮対象とすべき文末形式を網羅的に列挙したリストは、筆者が知る限り存在しない。この認識が正しいとすれば、このようなリストを作ること自体が、重要な研究項目となる。

このようなリストを作る際に大きな障害となるのが、いわゆる「表記ゆれ」の問題である。日本語は表記に寛容な言語であり、特に会話文では、多くの出現形が観察される。そのため、たとえば、以下に示すような「わね・わねえ・わねえ・わねー」をそれぞれ区別するか、それともいくつかの形式(たとえば、「わね」と「わねえ」)に正規化して集約するかを決める必要がある。

^{*1} 出典: 八目迷. 『小説 葬送のフリーレン 前奏』. 小学館, 2024. p134.

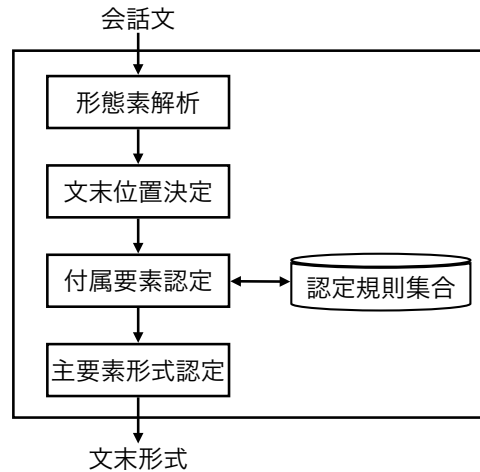


図1 文末形式認定システム Kohaku の構成

- (2) a. 出迎えてくれるなんて嬉しいわね
 b. 出迎えてくれるなんて嬉しいわねえ
 c. 出迎えてくれるなんて嬉しいわねえ
 d. 出迎えてくれるなんて嬉しいわねー

出現形の違いが話者のキャラクター性の違いに強く関与するならば、区別するのが望ましい。一方、あまり関与しないと考えられるのであれば、正規化した方が(頻度を稼げるという点で)定量的分析には都合がよい。このあたりのバランスをうまく調節する必要がある。

筆者は、これら2つの課題—認定する文末形式の範囲と正規化方法—に対しては、実際の会話文の分析を通して、試行錯誤的に決めていくしかないと考える。そのため、以下のような手続きを採用する。

1. 規則に基づいて文末形式を認定するシステムを作成する
2. これまでの分析(佐藤 2024)等に基づいて、認定規則集合(正規化規則を含む)を実装する。
3. システムを用いて、多くの会話文の文末形式を機械的に求め、それらを分析する。
4. 必要に応じて認定規則集合を修正する。
5. ステップ3に戻る。

この手続きの「規則に基づいて文末形式を認定する」という点が重要である。なぜならば、使用する認定規則集合は、対象とする文末形式の範囲と正規化方法を明示的に定義するものであるからである。

上記の手続きに従って、文末形式認定システム Kohaku を作成した。

3 文末形式認定システムの概要

文末形式認定システム Kohaku の構成を図1に示す。

3.1 文末形式の認定処理

Kohaku は、与えられた会話文^{*2}を形態素解析した後、以下の処理を経て、文末形式を決定する。形態素解析には Sudachi (Takaoka, Hisamoto, Kawahara, Sakamoto, Uchida, and Matsumoto 2018) を使用する。

1. 文末位置の決定

会話文の末尾から先頭に向かって形態素列を走査し、品詞が「補助記号」以外の形態素を文末位置 (文末の形態素) とする。Sudachi が使用する辞書では、句点、疑問符、感嘆符などに、「補助記号」という品詞が割り当てられている。すなわち、この処理は、文末のこれらの要素を無視することに相当する。

2. 付属要素の認定

認定規則集合を適用し、付属要素を認定する。付属要素は、終助詞、接続助詞、丁寧表現、特殊表現の 4 種類に分類されており、以下の出現順序制約を満たす場合に、付属要素を認定する。

$$\text{特殊表現} < \text{丁寧表現} < \text{接続助詞} < \text{終助詞} \quad (1)$$

3. 主要素形式の認定

文末の付属要素以外の最も文末に近い形態素を主要素とみなし、この形式を認定する。ただし、そのような形態素が存在しない (つまり、会話文が付属要素のみで構成されている) 場合も存在する。

この処理の中核をなすのは、ステップ 2 の認定規則集合に基づく付属要素の認定である。この詳細を、4 節と付録 A に示す。主要素形式の認定は、5 節で説明する。

3.2 システムの出力

与えられた会話文に対して、Kohaku は、以下の情報を出力する。

- F01 文末形式タイプ
- F02 文末形式 ID
- F03 終助詞 ID
- F04 接続助詞 ID
- F05 丁寧表現 ID
- F06 特殊表現 ID
- F07 主要素形式 ID

本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^{*3}。文末形式タイプ (F01) は、文末形式の大分類に相当する。具体的は、会話文に出現した付属要素の種別の最初の 1 文字を繋げたもの (ラベル) である。付属要素が全く出現しない場合、文末の形

^{*2} 会話文の末尾が仮名 (ひらがな・カタカナ) の場合は、デフォルトで句点を追加して形態素解析する。これは、文末記号の有無により形態素解析結果が異なることがあるためである。

^{*3} 文末形式が F01 と F02 の組である理由は、付録 E を参照のこと。

表 1 文末形式タイプ

	タイプ	主要素	付属要素			
1	終				終助詞	
2	接終			接続助詞	終助詞	
3	丁寧終		丁寧表現		終助詞	
4	特殊終		特殊表現		終助詞	
5	丁寧接終		丁寧表現	接続助詞	終助詞	
6	特殊接終		特殊表現	接続助詞	終助詞	
7	丁寧特終		特殊表現	丁寧表現	終助詞	
8	特殊丁寧接終		特殊表現	丁寧表現	接続助詞	終助詞
9	接			接続助詞		
10	丁寧接		丁寧表現	接続助詞		
11	特殊接		特殊表現	接続助詞		
12	丁寧特接		特殊表現	丁寧表現	接続助詞	
13	丁寧		丁寧表現			
14	丁寧特		特殊表現	丁寧表現		
15	特殊		特殊表現			
16	裸	述語				
17	-	(述語以外)				

態素が述語か否かを判定し、述語の場合はタイプ「裸」、述語以外の場合はタイプ「なし（'-」）」とする。文末形式タイプの一覧を表 1 に示す。

文末形式 ID (F02) は、構成要素の ID (F03–F07) から以下のように作成される。

1. 付属要素が存在する場合 — 存在する付属要素の ID (F03–F06) をドット（'.'）で繋いだ表現
2. 付属要素が存在しない場合
 - (a) 主要素が述語の場合 (文末形式タイプが「裸」の場合) — 主要素形式 ID (F07)
 - (b) それ以外の場合 — なし（'-」）」

システムの出力例を表 2 に示す*4。入力会話文は、すべて『図書館戦争』シリーズ (7 節で説明) の会話文である。

4 付属要素認定の設計方針

本節では、4 種類の付属要素 (終助詞、接続助詞、丁寧表現、特殊表現) の認定について、その設計方針を述べる。付属要素の認定規則の一覧は、付録 A にまとめて示す。なお、本文中で文末形式 ID や付属要素 ID を参照する場合は、'〈ID〉' と表記する。

4.1 終助詞

終助詞は、文末形式の構成要素のなかでも特に多様性に富む。終助詞の設定では、以下に示す原則を採用した。

原則 1 形態素解析結果の文末の終助詞列を、ひとまとまりの終助詞と認定する。ただし、終助詞列の頻度が低い場合、あるいは、形態素解析誤りを含む場合は、この限りではない。

*4 ここで示した出力例は、機械的に選択したもので、認定誤りを排除していない。

表 2 システムの出力例 (会話文の出典: 『図書館戦争』 (有川 2011a), 『図書館危機』 (有川 2011b))

文末形式		構成要素					会話文 (入力)
F01 タイプ	F02 ID	F07 主要素	F06 特殊	F05 丁寧	F04 接助	F03 終助	
終	わよ	V	-	-	-	わよ	ケンカ売ってんなら買うわよ！
終	のよ	V	-	-	-	のよ	協力してもいいけどめんどくさいってのがあんのよ。
終	わ	V	-	-	-	わ	あたしどうせ当たんないから任せるわ。
終	Xね	名	-	-	-	Xね	教育委員会の言いなりって感じね。
終	の	V	-	-	-	の	で、あんたはさっきから何唸ってんの
終	Xよ	名	-	-	-	Xよ	ちょっとお、こちとら仕事中よ。
終	ね	Aタ	-	-	-	ね	いやー、王子様から卒業しますはよかったね
終	よ	A	-	-	-	よ	司会進行役で駆り出されたいらしいよ。
終	Dね	だ	-	-	-	Dね	図書館側の攻略は諦めたみたいだね
終	Dよ	だ	-	-	-	Dよ	心配しなくてもあたしたちと弱い部分が違うだけだよ。
終	か	V	-	-	-	か	上官だからって口出す余地があるか
終	Bな	V	-	-	-	Bな	大したことじゃないとか言うな！
終	のか	V	-	-	-	のか	何か文句でもあるのか
終	な	A	-	-	-	な	許可は出してもいいな？
終	ぞ	V	-	-	-	ぞ	俺たちは俺たちの仕事をするぞ
接終	し.ね	だ	-	-	し	ね	こっちが情報に困ってると思われるのも癪だしね。
接終	から.ね	V	-	-	から	ね	ああ、笠原さんバネがあるからね。
接終	から.な	A	-	-	から	な	これで引込みはつかないからな。
丁終	んです.か	な	-	んです	-	か	あ、そうなんですか。
丁終	です.か	A	-	です	-	か	手、握ってもらっていいですか
丁終	です.な	名	-	です	-	な	こんな職業とはご挨拶ですな
特終	んだ. Dよ	A	んだ	-	-	Dよ	漫画に関しては良化委員会のチェックが緩いんだよ。
接	けど	A	-	-	けど	-	まあ、間違っではないけど
接	し	だ	-	-	し	-	うちのおかんってニュースはあんまり見ないクチだし。
接	から	だ	-	-	から	-	利用者がおじさんメンだから
接	って	終助か	-	-	って	-	嗅いでみるだけならオイルだけでいいかって。
丁	でしょ	名	-	でしょ	-	-	その場合でもどっちかに恋愛感情があるもんでしょ。
丁	んです	V	-	んです	-	-	あたしはあんたを超えるんです。
丁	ません	V用	-	ません	-	-	関係外の方に批判を受ける謂れはありません
丁	ます	V用	-	ます	-	-	あたしもいつか絶対取ります。
丁	です	副助だけ	-	です	-	-	実力派だなーってフツーに思ってるだけです！
特	じゃない	名	じゃない	-	-	-	俺たちは正義の味方じゃない
特	だろ	A	だろ	-	-	-	別にあり得なくはないだろ
特	んだ	V	んだ	-	-	-	やっぱあんたもそう言うんだ
特	だろう	状X	だろう	-	-	-	よし、それだけ文句が言えたら大丈夫だろう。
裸	Vテ	Vテ	-	-	-	-	この前、事務室で脳震盪起こさせちゃって……
裸	A	A	-	-	-	-	でも俺には真似できない
裸	Vタ	Vタ	-	-	-	-	笠原士長が銃撃した良化隊員は撤収した。
裸	V	V	-	-	-	-	図書特殊部隊として諸君を歓迎する。
裸	V命	V命	-	-	-	-	ついていけなかったらまたシフト戻すから言え
裸	だ	だ	-	-	-	-	でもそれを得ようとするかどうかは手塚の自由だ

原則 2 文末の「準体助詞の」、および、終助詞列の直前の「準体助詞の」は終助詞扱いとし、原則 1 に従って終助詞を認定する。

原則 3 「準体助詞の」の出現形である「の」と「ん」は、区別する。

原則 4 「終助詞もの」の出現形である「もの」と「もん」は、区別する。

これらの原則に従い、〈か〉、〈かな〉、〈かね〉、〈わ〉、〈わよ〉、〈わね〉、〈の〉、〈ん〉、〈のかな〉、〈んか〉等を、すべて異なる終助詞と認定する。

表3 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 of 終止形
R	動詞型活用 of 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」

表4 前方文脈を設定する終助詞

前方文脈	終助詞	前方文脈	終助詞	前方文脈	終助詞
D, T, X	ね	P	の	B, R, T	な
D, T, X	ねえ	P	のお	B, R, T	なあ
D, I, T, X	よ	P	のかしら	B	なや
D, I, T, X	よお	P	のかしらね	B, R, T	なよ
D, T, X	よな	P	のよ	B	なよな
D, T, X	よなあ	P	わ		
D, T, X	よね	P	わね		
D, T, X	よねえ	P	わねえ		
D, T, X	よのう	P	わよ		
		P	わよね		
		P	わよーん		
		P	ん		
		P	んかしら		
		P	んかしらね		
		P	んよ		

原則5 表3に示す前方文脈を設定し、表4に示す終助詞は、前方文脈に応じて異なる終助詞と認定する。

たとえば、「終助詞ね」は、前方文脈が「D・T・X」の場合はそれぞれ〈Dね〉、〈Tね〉、〈Xね〉と認定し、前方文脈がそれ以外の場合は、〈ね〉と認定する。

DとXは、次のような区別を可能とするため導入した。

- (3) a. それは嘘だね 〈Dね〉— 男性的
b. それは嘘ね 〈Xね〉— 女性的
c. ちゃんと書いたね 〈ね〉— 中性的

Tは、依頼を表す〈Tね〉・〈Tよ〉を、他の〈ね〉・〈よ〉と区別するために導入した。ただし、〈Tね〉は依頼を意味しない（単なる伝達の）場合もあるが、これらを形式から区別することはむずかしい。

- (4) a. きちんと書いてね 〈Tね〉— 依頼
b. ちょっとミスってね 〈Tね〉— 伝達

〈Tよ〉は、いわゆるお嬢様言葉でも出現する。この場合も依頼を意味しない。

- (5) a. きちんと書いてよ 〈Tよ〉— 依頼
b. 役割語のことがよく分かってよ 〈Tよ〉— 伝達（お嬢様言葉）

丁寧表現のテ形と結合した〈でして. Tね〉や〈まして. Tよ〉も依頼を意味しない。

- (6) a. 彼の希望でしてね 〈でして. Tね〉— 伝達
b. おふざけにも程がありましてよ 〈まして. Tよ〉— 伝達 (お嬢様言葉)

〈Tねえ〉は依頼を意味しないので、注意が必要である。

- (7) a. ちょっと足をくじいてねえ 〈Tねえ〉— 伝達
b. そんなことは、どこにも書いてねえ 〈Tねえ〉— 「書いてない」の音変化

Sudachi は、(7b) の「ねえ」を終助詞「ね」と認定するが、これは誤りである。Kohaku は、原則として形態素解析結果が正しいと仮定して文末形式を認定するため、(7b) の文末形式を〈Tねえ〉と誤認定する。

P は、いわゆるお嬢様言葉の終助詞を認定するために導入した。

- (8) a. 東京に行ったわ 〈わ〉
b. 東京に行きましたわ 〈Pわ〉— お嬢様言葉

B と R は、以下を区別するために導入した。

- (9) a. 東京へ行くな 〈Bな〉— 禁止、あるいは、確認
b. 東京へ行きな 〈Rな〉— 命令
c. その花は美しいな 〈な〉— 伝達、あるいは、確認

「終助詞な」は、〈Tな〉も区別する。

- (10) そこで少し休んでな 〈Tな〉— 命令、あるいは、伝達

Sudachi は、「よ」で終わる命令の形式に対し、句点の有無によって以下のような異なる解析結果を出力する。

1. 「書けよ。」(句点あり) — 動詞命令形+終助詞よ 〈Iよ〉
2. 「書けよ」(句点なし) — 動詞命令形 〈命ヨ〉
3. 「食べよ。」(句点あり) — 動詞連用形+終助詞よ 〈Iよ〉
4. 「食べよ」(句点なし) — 動詞意志推量形 〈V意〉

上記の 1 と 3 の場合を、他の「よ」と区別して、〈Iよ〉と認定する。

原則 6 末尾の長音化は、認定する終助詞が短単位の場合は区別し、複数の短単位から構成される場合は、区別せず正規化する。ただし、長音化がキャラクター性の違いに関与すると思われる場合は、この限りではない。

原則 6 に従い、〈な〉と〈なあ〉は区別する。「な」の長音化には「なあ・なあ・なア・なァ・な一・な〜」などがあり、これらをすべて〈なあ〉に正規化する。

一方、「かいな」は「な」が長音化しても、すべて〈かいな〉に正規化する。

原則 6 の例外には、次のようなものがある。〈よな〉と〈よなあ〉は区別する。〈よね〉と〈よねえ〉は区別する。

原則 7 複数単位の終助詞の並びを認める。

まれに、複数の (長い単位で認定する) 終助詞が連続する場合があるが、これを認める。これは、頻度が不十分、あるいは、形態素解析誤り等の理由により、ひとまとまりの終助詞としての認定を保留しているものも出力されるようにするためである。

4.2 接続助詞

接続助詞で終わる会話文の大半は、いわゆる「いいさし文」(白川 2009) である。このため、文末形式に接続助詞を含める。

接続助詞の設定では、以下の原則を採用した。

原則 1 形態素解析で接続助詞と認定される形態素を、接続助詞と認定する。ただし、いわゆるテ形とみなせる「接続助詞て」と「接続助詞で」は、接続助詞とは認定しない*5。

原則 2 假定形の形式〈たら〉、〈なら〉、〈のなら〉、〈んなら〉を接続助詞と認定する。

原則 3 〈ので〉、〈んで〉、〈のに〉を接続助詞と認定する。

原則 4 述語に接続する「副助詞って」を接続助詞〈って〉と認定する。

原則 5 末尾の長音化は元の形式に正規化する。

4.3 丁寧表現

小説によっては、ほとんどの発話を丁寧体で話すキャラクターが登場し、丁寧体の使用が、話者のキャラクター性と強く結びついている場合がある。そのため、丁寧表現を文末形式に含める。

丁寧表現の設定では、以下の原則を採用した。

原則 1 「です」のバリエーション、「ます」のバリエーションを丁寧表現と認定する。ここでのバリエーションは、おおむね、終止形・否定形・意志推量形・命令形・タ形・テ形である。

原則 2 「です」の直前が「準体助詞の」の場合は、これを含めて丁寧表現と認定する。例:〈のです〉、〈んでしょう〉

原則 3 〈ください〉、〈Tください〉、〈なさい〉を丁寧表現と認定する。

原則 4 丁寧表現の下位分類として、「ます系・でしょう系・ください系・なさい系」を設定する。

原則 5 丁寧表現として認定するのは原則として 1 単位であるが、例外的に、以下の並びの連続 (2 単位または 3 単位) を認める。

$$\left\{ \begin{array}{l} \text{ください系} \\ \text{なさい系} \end{array} \right\} + \text{ます系} \quad (2)$$

$$\text{ます系} + \text{でしょう系} \quad (3)$$

原則 5 に従い、丁寧表現〈なさい. ませ〉、〈ます. でしょ〉、〈Tください. ません. でしょう〉を認定する。

*5 テ形は付属要素を伴わない形式と認定する。

原則 6 丁寧表現の末尾の形態素は活用語であり、その出現形はバリエーションは多岐にわたる。そのため、形態素毎に正規化規則を定める。

たとえば、UniDic (unidic-cwj-3.1.0) では、「助動詞です」の終止形として、次の出現形が登録されている。

です、デス、で一す、で～す、ですー、ですう、ですゥ、す、ス、ス、でちゅ、でしゅ、でっす、でつ、でし

複数の形態素が結合した形式では、出現形のバリエーションは掛け算で増える。そのため、構成する形態素毎に正規化規則を定義し、丁寧表現全体の正規化を実現する。

4.4 特殊表現

「だろ(う)・じゃない」などは、会話文の末尾に頻出し、本来の意味を離れて、伝達や確認要求などの機能を果たすことが多い。このため、これらの表現を特殊表現として文末形式に含める。

特殊表現の設定では、以下の原則を採用した。

原則 1 「だろ(う)」は、末尾の「う」の欠落の有無を区別し、〈だろ(う)〉、〈だろ〉を特殊表現と認定する。これらの形式は、本来の意志推量の解釈も可能であるが、会話文では、伝達や確認要求を表す終助詞的用法が優勢である。

原則 2 直前に「準体助詞の」が付いた形式は、それを含めて特殊表現と認定する。例：〈のだろ(う)〉、〈んだろ〉

原則 3 〈じゃない〉、〈のじゃない〉、〈んじゃない〉を特殊表現と認定する。これらの形式は、本来の否定の解釈(「ではない」)も可能であるが、会話文では、伝達や確認要求を表す終助詞的用法が優勢である。

原則 4 〈のだ〉、〈んだ〉を特殊表現と認定する。

原則 5 〈じゃ〉、〈のじゃ〉、〈んじゃ〉は終助詞的であるが、「じゃない」の「ない」が欠落した場合もあるので、特殊表現として扱う。

原則 6 特殊表現の下位分類として、「じゃない系・だろ(う)系」を設定する。

原則 7 特殊表現として認定するのは原則として 1 単位であるが、例外的に、以下の並びの 2 単位を認める。

じゃない系 + だろ(う)系 (4)

原則 7 により、〈じゃない. だろ(う)〉や〈じゃない. んだろ〉などを特殊表現として認定する。

原則 8 特殊表現の正規化は、丁寧表現の正規化と同様に、構成する形態素毎に正規化規則を定義して実現する。

5 主要素形式認定の設計方針

文末の付属要素以外の最も文末に近い形態素を調べ、主要素形式 ID を付与する。主要素形式 ID の付与は、付属要素の有無を問わない。

主要素形式 ID を導入する主たる理由は、付属要素をまったく伴わない場合の形式を区別するためである。具体的には、

- 文末が述語とみなせるか (会話文が述語で終了する文とみなせるか)
- 述語の場合は、どんな活用形か。特に重要なのは、活用形が文の機能に直結する、命令形・意志推量形・仮定形である。

具体的には、文末形式タイプ「裸」と主要素形式 ID を以下のように付与する。ただし、文末形式タイプ「裸」を付与するのは、付属要素を伴わない場合で、かつ、主要素を述語と認定する場合のみである。

- 当該形態素が活用語の場合は、これを述語とみなし、活用型と活用形に応じた ID を付与する。ただし、「助動詞だ」と「助動詞や」を除く。
- 「助動詞だ」は、以下のように判定する。
 - － 終止形の場合は述語とみなし、出現形に応じた ID、具体的には、〈だ〉、〈だあ〉、〈ちゃ〉を付与する。
 - － タ形の場合は述語とみなし、〈だっタ〉を付与する。
 - － 連体形「な」の場合は述語とみなし、〈な〉を付与する。ただし、付属要素が存在しない場合は、形態素解析誤りであることが多い。
 - － 連用形「に」の場合は述語とみなさず副詞扱いとする。
- 「助動詞や」は述語とみなし、以下の ID を付与する。
 - － 終止形の場合は、〈や〉、または、〈やあ〉を付与する。
 - － タ形の場合は、〈やっタ〉を付与する。
- 形状詞・助動詞語幹の場合は、「助動詞だ」が省略されている述語とみなし、〈状X〉を付与する*6。
- それ以外の非活用語の場合は、述語とみなさない。

付与する主要素形式 ID の一覧を付録 B に示す。

6 小説会話文文末形式リスト Kohaku-FFL

2 節で述べたように、Kohaku システムの開発過程では、多くの会話文データに対してシステムを適用し、その出力の観察に基づいて文末形式認定規則集合を修正することを繰り返した。使用した会話文データを以下に示す。

D1 「BCCWJ 小説会話文」(山崎・宮寄・柏野 2022) から抽出した 276,576 文

D2 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・石川・夏目・小川・佐藤 2024) を文に分割した 8,879,279 文。

システム開発が完了した段階で、上記の会話文データの全てに文末形式を付与し、D1、D2 のい

*6 名詞や副詞の場合も、「助動詞だ」が省略されている可能性があるが、形式からは判定不能なため、述語とは判定しない。

いずれかで 0.1bp^{*7}以上の頻度で出現した 1,923 種類^{*8}の文末形式を収録したリスト「小説会話文文末形式リスト Kohaku-FFL」を作成した。

このリストには、以下の情報が収録されている。

- 1,065 種類の文末形式に対して、D1 における頻度 (回数と出現率)
- 933 種類に文末形式に対して、D2 における頻度 (回数と出現率)
- 806 種類の文末形式に対して、D1 と D2 の出現率の差 (対数比)
- 335 種類の文末形式に対して、話者の女性率^{*9}

このリストに収録されている 1,923 種類の文末形式は、上記 D1、D2 の会話文を 99% 以上カバーする^{*10}。

このリストのさらなる詳細は、文献 (佐藤 2025) を参照されたい。

7 『図書館戦争』シリーズにおけるセリフの書き分け

本節では、作成したシステムの利用例として、『図書館戦争』シリーズを対象に、主要登場人物のセリフがどのように書き分けられているかを調査した結果を示す。

『図書館戦争』シリーズは、アニメ化・実写化もされた有川浩の作品 (ライトノベル) である。ここでは、第 1 作の『図書館戦争』 (有川 2011a) と第 3 作の『図書館危機』 (有川 2011b) の 6 名の主要登場人物の会話文を対象とする。『図書館戦争』の角川文庫版の巻頭ページには登場人物紹介があり、6 名が次のように紹介されている。笠原郁と柴崎麻子が女性で、他の 4 名が男性である。

1. 堂上篤 (怒れるチビ) 図書特殊部隊・堂上班班長/二等図書正
冷静でいるよう努力しているが、郁の行動に雷を落とす毎日。
2. 笠原郁 (熱血バカ) 図書特殊部隊・堂上班班員/一等図書士
高三の秋に図書隊員に助けられ、彼を『王子様』と慕い、図書隊に入る。
3. 手塚光 (頑な少年) 図書特殊部隊・堂上班班員/一等図書士
真面目で努力家。堂上を慕うあまり、郁の存在が目障りとなっている。
4. 柴崎麻子 (情報屋) 図書館員/一等図書士
真の自分を隠し、八方美人に徹する。また、あらゆる情報に精通する。
5. 小牧幹久 (笑う正論) 図書特殊部隊・堂上班副班長/二等図書正
冷静沈着で、決して弱みを見せない。堂上の良き相談相手。
6. 玄田竜介 (喧嘩屋中年) 図書特殊部隊隊長/三等図書監
無茶を無茶と思わない性格。稲嶺 (関東図書基地指令) の良き補佐役。

本システムを適用する会話文データは、以下の手順で作成した。

1. 紙媒体の小説を OCR を用いてテキスト化し、誤りを修正した。
2. 小説中のそれぞれの発話に、話者の情報を人手で付与した。

^{*7} bp は万分率を表す。1bp=0.01%

^{*8} 文末形式タイプ「なし (・)」を 1 種類として含む。

^{*9} D1 に付与されている話者の性別情報を利用して算出した女性話者の比率 (%)。男性率 = 100 - 女性率

^{*10} ここでのカバー率は、会話文の文末形式が 1,923 種類のいずれかとなる文の数を、総文数で割った値である。

3. 話者毎に発話を抽出した。
4. 発話を句点で文に分割した。

表5に、本システムを用いて計測した、6名の登場人物の文末形式の使用率を示す。この表では、いずれかの人物の使用率が1.5%以上の文末形式を対象に、使用率が1.0%以上の場合は数値で、1.0%未満の場合は「-」で示した。文末形式の使用率とは、その文末形式を持つ文数をその話者の総文数で割った値である。この表の女性率は、前述の「小説会話文文末形式リスト Kohaku-FFL」に付与されている女性率である。なお、以下の説明では、一般に女性的と理解されている形式をF形式、男性的と理解されている形式をM形式と表記する。

この表から、それぞれの登場人物において、文末形式が書き分けられていることが確認できる。

- 柴崎は、F形式の〈わよ〉、〈のよ〉、〈わ〉、〈Xね〉、〈の〉、〈Xよ〉などを多用する。終助詞文末や接続助詞文末の使用率が高く、言い切りの形式(〈A〉、〈Vタ〉、〈V〉、〈だ〉)の使用率は低い。全体として、女性的な言葉遣いである。
- 笠原は、F形式を使用するが、使用率は柴崎より低い。接続助詞文末の使用率は高いが、終助詞文末の使用率はそれほど高くない。言い切り形式の使用率は低い。丁寧表現を含む文が多いのは、上官(堂上)との会話が多いためである。現代の若い女性の中性的な言葉遣いと捉えられるだろう。
- 小牧は、男性の中では終助詞文末の使用率が高い。特に、「ね」(〈ね〉、〈Dね〉)と「よ」(〈よ〉、〈Dよ〉)を多用する。接続助詞文末の使用率も、女性と同程度に高い。その一方で、言い切り形式の使用率は女性よりも高く、2名の女性が使わない〈だ〉の使用が認められる。全体として、男性にしては、柔らかい感じの言葉遣いである。
- 手塚は、〈ぞ〉、〈な〉、〈のか〉、〈Bな〉などのM形式の使用が見られる一方で、すこしマイルドは〈Dよ〉や〈よ〉も使用する。小牧が使用する〈Dね〉や〈ね〉は使用しない。言い切り形式の使用率は、小牧と同程度である。丁寧表現の使用率が高いのは、上官(堂上や小牧)に対する発話で丁寧表現を使用するからである。若い男性にみられる言葉遣いと捉えられるだろう。
- 堂上と玄田は、M形式の終助詞を多用するとともに、言い切りの形式、特に「だ」(〈だ〉)や動詞(〈V〉)で終わる文の使用率が高い。接続助詞文末の使用率が低いことも特徴の一つで、他の4名が使う〈から〉を用いず、〈から.な〉を使う。命令形(〈V命〉)が多いのは、役職(上官)のためと思われる。全体としては、男性的とみなされる言葉遣いである。
- この二人の文末形式の使用の差はそれほど顕著ではないが、玄田の方が「裸」タイプの文末形式の使用率が高く、特に〈だ〉と〈V〉の使用率に差が見られる。なお、玄田の丁寧表現の使用率が比較的高いには、対外的な折衝の場面があるためと思われる。

以上のように、『図書館戦争』シリーズでは、6名の主要登場人物のセリフが巧みに書き分けられており、その書き分けには、終助詞だけでなく、それ以外の文末形式も活用されている。本システムを利用すると、このような分析が容易となる。

表5 文末形式の使用実態 (文数以外は、パーセント)

女性率				柴崎	笠原	小牧	手塚	堂上	玄田
総文数				693	1,380	446	360	844	378
裸	終	わよ	99.5	5.3	1.2	-	-	-	-
	終	のよ	98.2	3.5	1.2	-	-	-	-
	終	わ	98.0	3.6	-	-	-	-	-
	終	Xね	94.7	2.0	-	-	-	-	-
	終	の	92.7	2.3	2.6	-	-	-	-
	丁	でしょ	92.1	2.5	-	-	-	-	-
	終	Xよ	91.9	4.3	1.2	-	-	-	-
		Vテ	78.2	-	2.5	1.8	2.2	-	-
	接	けど	76.1	-	-	2.2	-	-	-
	接	し	72.2	2.6	1.2	1.8	-	-	-
裸	接	から	71.5	1.3	2.3	4.7	1.1	-	-
	接	って	63.6	1.7	2.0	-	-	-	-
	特	じゃない	63.5	1.9	-	-	-	-	-
	丁	んです	57.6	-	1.8	-	-	-	-
	終	ね	57.4	1.7	-	3.6	-	-	-
	丁 終	んです. か	53.8	-	2.5	-	-	-	-
	接終	し. ね	52.5	-	-	2.5	-	-	-
	丁	ません	50.6	-	1.3	-	2.5	-	-
	丁	ます	46.9	1.2	2.9	-	3.6	1.3	2.4
	丁 終	です. か	43.5	-	1.6	-	2.2	-	1.1
裸	丁	です	42.5	1.7	3.4	1.1	1.7	-	2.1
	終	よ	39.9	-	1.6	6.7	1.4	-	-
	接終	から. ね	37.3	-	-	2.5	-	-	-
		A	33.9	1.3	1.7	3.6	3.3	6.8	5.0
	裸	Vタ	33.7	1.3	1.3	2.2	1.1	2.7	2.6
	特 終	んだ. Dよ	31.3	-	-	1.3	1.9	-	-
	裸	V	29.9	1.6	1.7	3.6	3.6	8.3	16.1
	終	Dね	26.9	-	-	2.0	-	-	-
	終	Dよ	24.9	-	-	1.6	1.7	-	-
	終	か	17.4	-	2.0	1.8	4.2	5.3	6.1
裸		V命	15.6	-	-	-	1.4	6.8	7.9
	特	だろ	13.6	-	-	1.3	2.5	-	-
	終	Bな	11.0	-	-	-	2.2	3.2	1.1
	終	のか	10.7	-	-	-	2.8	2.3	-
	特	んだ	10.6	-	-	-	1.4	2.7	-
	終	な	10.0	-	-	-	2.5	3.9	6.1
	終	ぞ	9.4	-	-	-	2.2	3.6	2.4
	特	だろう	8.0	-	-	-	1.9	1.7	-
	裸	だ	7.1	-	-	1.8	3.6	13.2	15.6
	接終	から. な	3.9	-	-	-	-	2.0	2.1
-	丁 終	です. な	1.0	-	-	-	-	-	1.6
	-	-	57.3	20.1	28.3	19.5	17.8	15.6	11.6
終助詞で終わる				45.9	30.9	41.7	37.2	28.4	27.5
接続助詞で終わる				14.1	12.7	13.2	8.6	4.1	2.6
丁寧表現を含む				17.0	27.0	6.5	21.9	5.3	12.4
タイプが「裸」				7.4	12.2	15.2	18.1	41.1	49.2
文末形式の平均女性率				69.4	57.2	48.1	36.6	27.1	25.0

8 まとめ

本稿では、小説会話文の文末形式を自動認定するシステムについて述べた。このシステムを用いることにより、会話文の文末形式の定量的な把握と分析が可能となる。

本システムは、形態素解析器の解析結果に基づいて文末形式を認定するため、形態素解析が正しい解析結果を出力しなかった場合は、文末形式を正しく認定できない。一般に、会話文の形態素解析は、書き言葉の形態素解析よりも精度が落ちる。特に、非標準的な表記や方言が含まれる場合は、正しく解析できないことが多い。会話文の形態素解析の向上は、本研究の範囲外ではあるが、文末形式の認定結果を調査することにより、誤解析の具体例を蓄積することができる。

本システムが認定するのは、あくまでも「形式」である。会話文の「機能」の自動認定を実現するためには、形式と機能の関係をより詳細に調査する必要がある。本システムで文末形式を付与したコーパスに、人手で機能を付与することは、その第一歩となる。

謝辞

本システムの開発に使用した「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。

参考文献

- 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 (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.

- 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)*.
- 有川浩 (2011a). 『図書館戦争』 角川文庫.
- 有川浩 (2011b). 『図書館危機』 角川文庫.
- 白川博之 (2009). 『「言いさし文」の研究』 くろしお出版.
- 山崎誠・宮崎由美・柏野和佳子 (2022). 「小説会話文への話者情報付与」 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 (2025). 「小説会話文の文末形式リストの作成」 言語資源ワークショップ 2025. 発表予定

付録 A 付属要素の認定規則

A.1 認定規則のフォーマット

認定規則のフォーマットは、すべての付属要素に対して共通である。

認定規則は、次の要素から構成されている。

規則番号

ID (付属要素の)ID

系 下位分類

M_n-M₁ 形態素パターン列

追加条件 形態素パターン列で記述した条件に加えてチェックする条件

S₁-S_m 出現形パターンのリスト

■# 規則を識別する番号である。通常の規則は、付属要素を認定する規則であるが、番号の前に‘!’が付与されている規則は、付属要素として認定しないための規則(除外規則)である。番号の前に‘?’が付与されている規則は、現時点までに適用が確認されていない規則(未使用規則)である。

■ID 認定する付属要素を識別する文字列である。

■形態素パターン列 M_n-M₁の形態素パターン列は、その規則の適用条件のうち、形態素列に対する必要条件(照合パターン)である。

M_iの形態素パターンは、1形態素に対する条件(照合パターン)を表す。多くの場合、品詞情報(品詞または品詞細分類)と語彙素の組み合わせで記述するが、活用形(f,F)や出現形(s,S)を指定することも可能である。

形態素パターンには、以下のような略記法がある。

の 「終助詞の」または「準体助詞の」

のsの 「準体助詞の」の出現形「の」

のsん 「準体助詞の」の出現形「ん」

じゃ 「助動詞だ」の連用形-融合

ない 「形容詞無い」の終止形または連用形

なし 「形容詞無い」の出現形「なし」

なかつ 「形容詞無い」の出現形「なかつ」

くださっ 「動詞下さる」の連用形-促音便

なさっ 「動詞為さる」の出現形「なさっ」

なさい 「動詞為さる」の出現形「なさい」

なさり 「動詞為さる」の出現形「なさり」

■追加条件 形態素パターン列で記述した条件に加えてチェックする条件で、形態素パターン列で記述できないような条件を記述する場合に使用する。追加条件名は、システムの内部で定義されて

いる関数名に対応し、この関数を呼ぶことにより、条件を満たすかどうかをチェックする。追加条件の一覧は、付録 C に示す。

■**出現形の標準化** 形態素パターン列と一致し、かつ、追加条件を満たした形態素列から、次で説明する出現形パターンと照合する文字列を作成するが、その際、以下の標準化を適用する。

- (JIS で定義されている) カタカナは、すべてひらがなに変換する。
- 波線「～」は長音符号「ー」に変換する。
- 長音符号の連続、および、同一の小書き文字の連続は、それぞれ 1 文字に縮退させる。

この処理により、たとえば、「よネ」は「よね」に、「くださーーい」は「くさーい」に標準化される。

■**出現形パターン** $S1-Sm$ の出現形パターンのリストは、出現形の区別と正規化を制御するための条件記述である。出現形が、(前方文脈を削除した) ID と一致する場合、あるいは、 S_i に記述されたパターンとマッチする場合、その行の規則 (ID) が採用される。

たとえば、終助詞認定規則の #41 から #44 は同一の形態素パターン列が記述されている。この形態素パターン列に合致する出現形には、「のか・のかあ・のかあ・んか・んかあ・んかー」などがある。出現形パターンに含まれる英小文字 (a, i, u, e, o) は母音の長音化を表すもので、たとえば、「a」は、「あ・あ・ー」のいずれかを表す。すなわち、この 4 行は、「のか・のかあ・んか・んかあ」を区別すること、および、出現形「のかあ・のかあ」等は〈のかあ〉に集約し、出現形「んかあ・んかー」等は、〈んかあ〉に集約することを意味している。なお、英大文字 O は、「お・お・う・う・ー」のいずれかを表す。

特別な出現パターンの記述として、「@」と「*」がある。

@ すべての出現形を、ID に集約する。

* 各構成形態素を正規化したのち、それを結合したものを ID とする。

A.2 終助詞の認定規則

終助詞の認定規則を表 6–表 10 に示す。形態素パターンで、品詞細分類が「終助詞」の場合は、品詞情報を省略した。

表 6 終助詞の認定規則 (1)

#	ID	M3	M2	M1	追加条件	S1	S2
1	い			い	is_終助い		
2	かい		か	い		かaい	かい
3	のかい	の	か	い		のかaい	
4	んかい	の	か	い		んかaい	
5	のんかい	のん	か	い			
6	もんかい	名詞もん	か	い		もんかaい	
7	ぜい		ぜ	い		ぜい	ぜーい
8	だい		助動詞だ	い	is_終助だい	でい	
9	のだい	の	助動詞だ	い			
10	んだい	の	助動詞だ	い		んでい	
11	じゃい		助動詞だsじゃ	い			
12	わい		わ	い		わaい	わい
13	ぞい			ぞい			
14	たい			たい			
15	ばい			ばい			
16	べ			べい			
17	べい			べい			
18	べえ			べい			
19	んべ		のsん	べい			
?20	んべい		のsん	べい			
?21	んべえ		のsん	べい			
22	のう			のう	is_終助のう	のう	
23	かのう		か	のう		かのu	
24	のかのう	の	か	のう		のかのu	
25	んかのう	の	か	のう		んかのu	
26	けのう		け	のう			
27	よのう		よ	のう		よのu	
?28	Dよのう		よ	のう		よのu	
?29	Tよのう		よ	のう		よのu	
30	Xよのう		よ	のう		よのu	
31	え			え			
32	かえ		か	え			
33	のかえ	の	か	え			
34	んかえ	の	か	え			
35	ぞえ		ぞ	え			
36	わえ		わ	え			
37	かあ			か	not_いうか	かa	
38	か			か	not_いうか		
39	っけか		けsっけ	か		っけかa	
40	じゃんか		じゃん	か		じゃんかa	
41	のか		の	か			
42	のかあ		の	か		のかa	
43	んか		の	か			
44	んかあ		の	か		んかa	
45	のんか		のん	か			
46	べいか		べい	か		べeか	
47	べか		べい	か			
48	ものか		名詞もの	か		ものかa	
49	もんか		名詞もん	か		もんかa	
!50	もが			もが	FALSE		
51	け			け			
52	っけ			け			
53	ちょけ		ちょ	け			
54	さあ			さ	tail	さa	
55	さ			さ	tail		

表 7 終助詞の認定規則 (2)

#	ID	M4	M3	M2	M1	追加条件	S1	S2
56	べさ			べい	さ		べさa	
?57	べいさ			べい	さ		べいさa	
58	のさ			の	さ		のさa	
?59	ものさ			もの	さ		ものさa	
60	もんさ			もの	さ		もんさa	
61	ともさ			とも	さ		ともさa	
62	わさ			わ	さ		わさa	
63	んさ			の	さ		んさa	
!64	じょ				じょ	FALSE		
!65	くさ				くさ	FALSE		
66	ぜ				ぜ			
67	ぜえ				ぜ		ぜe	
68	ぞ				ぞ		ぞO	
69	じょ				ぞ			
70	ちょ				ちょ			
71	で				で		でe	
72	ねんで			ねん	で			
!73	ど				ど	FALSE		
74	な				な			
?75	な				助動詞だsな	is_終助な		
76	Bな				な			
77	Bな				助動詞だsな			
78	Rな				な			
79	Tな				な			
80	なあ				な		なあa	
81	Bなあ				な		なあa	
82	Rなあ				な		なあa	
83	Tなあ				な		なあa	
84	らあな			ラsらあ	な			
85	いな			い	な	is_終助いな		
86	かいな		か	い	な		かいなa	
87	のかいな	の	か	い	な		のかいなa	
88	んかいな	の	か	い	な		んかいなa	
89	べな			べい	な			
?90	べいな			べい	な			
91	わいな		わ	い	な		わいなa	
92	かなあ			か	な		かなa	
93	かな			か	な			
94	っけかな		けsつけ	か	な		っけかなa	
95	のかな		の	か	な		のかなa	
96	んかな		の	か	な		んかなa	
97	けな			け	な		けなa	
98	っけな			け	な		っけなa	
99	さな			さsさ	な		さなa	
100	どな			ど	な			
101	ものな			もの	な		ものなa	
102	やな			や	な		やなa	
103	よな			よ	な			
104	Dよな			よ	な			
105	Tよな			よ	な			
106	Xよな			よ	な			
107	よなあ			よ	な		よなa	
108	Dよなあ			よ	な		よなa	
109	Tよなあ			よ	な		よなa	
110	Xよなあ			よ	な		よなa	

表 8 終助詞の認定規則 (3)

#	ID	M4	M3	M2	M1	追加条件	S1	S2
111	なよな		な	よ	な		なよなa	なaよな
112	Bなよな		な	よ	な		なよなa	なaよな
113	わな			わ	な	not_わあな	わなa	
?114	んな			の	助動詞だsな	tail		
115	てんな			てん	な		てんなa	
116	ねんな			ねん	な		ねんなa	
117	のんな			のん	な		のんなa	
118	もんな			もの	な		もんなa	
119	ね				ね	tail		
120	Dね				ね	tail		
121	Tね				ね	tail		
122	Xね				ね	tail		
123	ねえ				ね	tail	ねe	
124	Dねえ				ね	tail	ねe	
125	Tねえ				ね	tail	ねe	
126	Xねえ				ね	tail	ねe	
127	いね			い	ね			
128	かいね		か	い	ね		かいねe	
129	んかいね	の	か	い	ね			
130	のかいね	の	か	い	ね			
131	たいね			たい	ね			
132	かねえ			か	ね	not_かあねえ	かねe	
133	かね			か	ね	not_かあねえ		
134	のかね		の	か	ね	tail	のかねe	
135	んかね		の	か	ね	tail	んかねe	
136	けね			け	ね	tail	けねe	
137	っけね			け	ね	tail	っけねe	
138	さね			さs	ね	tail	さねe	
139	じゃんね			じゃん	ね	tail	じゃんねe	
140	んじゃんね		の	じゃん	ね	tail	んじゃんねe	
141	ぞね			ぞ	ね		ぞねe	
142	Rなね			な	ね	tail	なねe	
143	のね			の	ね	tail		
144	のねえ			の	ね	tail	のねe	
145	ものね			もの	ね	tail	ものねe	
146	よね			よs	ね	tail		
147	Dよね			よs	ね	tail		
148	Tよね			よs	ね	tail		
149	Xよね			よs	ね	tail		
150	よねえ			よs	ね	tail	よねe	
151	Dよねえ			よs	ね	tail	よねe	
152	Tよねえ			よs	ね	tail	よねe	
153	Xよねえ			よs	ね	tail	よねe	
154	のよね		の	よs	ね	tail	のよねe	
155	わよね		わ	よs	ね	tail	わよねe	
156	Pわよね		わ	よs	ね	tail	わよねe	
157	んよね		の	よs	ね	tail	んよねe	
158	かしらね			かしら	ね	not_か知らね	かしらねe	
159	のかしらね		の	かしら	ね	tail	のかしらねe	
?160	Pのかしらね		の	かしら	ね	tail	のかしらねe	
?161	んかしらね		の	かしら	ね	tail	んかしらねe	
?162	Pんかしらね		の	かしら	ね	tail	んかしらねe	
163	わね			わ	ね	tail		
164	Pわね			わ	ね	tail		
165	わねえ			わ	ね	tail	わねe	

表9 終助詞の認定規則 (4)

#	ID	M3	M2	M1	追加条件	S1	S2
166	Pわねえ		わ	ね	tail	わねe	
167	んね		の	ね	tail		
168	んねえ		の	ね	tail	のねe	
169	てんね		てん	ね		てんねe	
170	もんね		もの	ね	tail	もんねe	
171	の			の	tail		
172	ん			の	tail		
173	Pの			の	tail		
174	Pん			の	tail		
175	のお			の	tail	のo	
176	Pのお			の	tail	のo	
177	かいの	か	い	の		かいのo	
178	かいの		名詞かい	格助詞の	tail		
179	かの		か	の	tail	かのo	かaのo
180	のかの	の	か	の	tail	のかのo	
181	んかの	の	か	の	tail	んかのo	
182	やん		助動詞やsや	の	is_終助やん		
183	やん		や	の			
184	わあん		わ	の		わaん	
185	もの			もの			
186	もん			もの			
187	とも			とも	is_終助とも		
188	や			や	is_終助や		
189	やー			や	is_終助やー		
190	かや		か	や		かやa	
191	のかや	の	か	や			
192	んかや	の	か	や			
193	ぜや		ぜ	や	not_ぜやー		
194	ぞや		ぞ	や			
195	Bなや		な	や		なやa	なーや
196	のにや		の	格助詞にSにや		のにやa	
197	んにや		の	格助詞にSにや		んにやa	
198	ねや		ね	や	not_ねーや		
199	よ			よ	tail	よ	
200	Dよ			よ	tail	よ	
201	Tよ			よ	tail	よ	
202	Xよ			よ	tail	よ	
203	Iよ			よ	tail	よ	
204	よお			よ	tail	よO	
205	Iよお			よ	tail	よO	
206	Dよお			よ	tail	よO	
207	Tよお			よ	tail	よO	
208	Xよお			よ	tail	よO	
209	べえよ		べいsべえ	よ			
210	ぜーよ		ぜ	よ		ぜeよ	
211	かよ		か	よ		かよO	かよ
212	のかよ	の	か	よ		のかよO	
213	んかよ	の	か	よ		のかよO	
214	じゃんかよ	じゃん	か	よ		じゃんかよo	
215	ぜよ		ぜ	よ		ぜよO	
216	ぞよ		ぞ	よ		ぞよO	
217	なよ		な	よ		なよO	なaよO
218	Bなよ		な	よ		なよO	なaよO
219	Rなよ		な	よ		なよO	なaよO
220	Tなよ		な	よ		なよO	なaよO

表 10 終助詞の認定規則 (5)

#	ID	M3	M2	M1	追加条件	S1	S2	S3
221	のよ		の	よ		のよO	のよ	のよーん
222	Pのよ		の	よ		のよO	のよ	のよーん
223	べよ		べいsべ	よ		べよo		
224	ともよ		とも	よ		ともよu		
225	わよ		わ	よ		わよO	わよ	
226	Pわよ		わ	よ		わよO	わよ	
227	わよーん		わ	よ				
?228	Pわよーん		わ	よ				
229	んよ		の	よ		んよO		
?230	Pんよ		の	よ		んよO		
231	もんよ		ものsもん	よ		もんよo		
232	じゃんよ		じゃん	よ		じゃんよO		
233	らあ			ラ				
234	ら			ラ				
235	かしら			かしら	tail	か知ら		
236	かしら		かし	ラ		かしらあ		
237	かしらん			かしら	tail			
238	のかしら		の	かしら		のかしらん		
239	のかしら	の	かし	ラ		のかしらあ		
240	Pのかしら		の	かしら		のかしらん		
?241	んかしら		の	かしら				
?242	Pんかしら		の	かしら		んかしらん		
243	わ			わ	is_終助わ	わa		
244	Pわ			わ	is_終助わ	わa		
245	じゃん			じゃん		じゃaん	ちゃん	ちゃaん
246	てん			てん				
247	ねん			ねん				
248	ねーん			ねん				
249	んねん		の	ねん				
250	のねん		の	ねん				
251	よんねん		よん	ねん				
252	のん			のん				
253	よん			よん				
254	哉			哉				

A.3 接続助詞の認定規則

表 11 に、接続助詞の認定規則を示す。接続助詞は、デフォルトで末尾の長音化を元の形式に正規化するが、S1 に「!」を指定した場合は、この機能が抑制される。

表 11 接続助詞の認定規則

#	ID	M2	M1	追加条件	S1
1	ちゃあ		接続助詞では		
2	さかい		接続助詞さかい		
3	が		接続助詞が		
4	すけ		接続助詞すけ		
5	し		接続助詞し		
6	つつ		接続助詞つつ		
7	で		助動詞だsで		
8	で		接続助詞で	not_テ形	
9	って		副助詞って	is_接助って	て
10	たって		接続助詞たって		だって
11	ったって		接続助詞ったって		
12	とて		接続助詞とて		
13	ので	の	助動詞だsで		
14	んで	の	助動詞だsで		
15	と		接続助詞と		
16	ど		接続助詞ど		
17	けど		接続助詞けれど		けどu
18	けれど		接続助詞けれど		
19	に		接続助詞に		
20	きに		接続助詞きに	not_おーきに	
21	のに	の	格助詞に	not_のにゃ	
22	んに	の	格助詞に	not_のにゃ	
23	んに	の	助動詞だsに		
24	ば		接続助詞ば		
25	ては		接続助詞ては		
26	とも		接続助詞とも	is_接助とも	
27	とも		とも	is_接助とも	
?28	ちゃ		接続助詞では		!
29	ちゃ		接続助詞てsちゃ		
30	から		接続助詞から	is_接助から	かーらー
31	ながら		接続助詞ながら		
32	たら		助動詞たF假定		だら
33	なら		助動詞だsなら		
34	のなら	の	助動詞だsなら		
35	んなら	の	助動詞だsなら		
36	なり		接続助詞なり		
37	けん		接続助詞けん		
38	ばってん		接続助詞ばってん		

A.4 丁寧表現の認定規則

表 12 に、丁寧表現の認定規則を示す。形態素パターンで、品詞が助動詞の場合は、品詞情報を省略した。

形態素パターンに含まれる ‘t’ は終止形または連体形を意味する。本来は、終止形のみでよいのだが、Sudachi は終止形を連体形と誤認識するケースが多いので、連体形も認めることとしている。

すでに述べたように、S1 に「*」が指定されている規則は、文末形式を構成する各形態素毎にシステムの内部で定義されている正規化が適用される。たとえば、「動詞下さる F 命令」は、出現形「下さい・ください」を「ください」に、出現形「下さーい・くださーい」を「くださーい」に、出現形「下され・くだされ」を「くだされ」に正規化する。つまり、規則#1 は、〈ください〉・〈くださーい〉・〈くだされ〉の 3 種類の丁寧表現を認定する規則を表現している。形態素単位の正規化規則の一覧を、付録 D に示す。

表 12 丁寧表現の認定規則

#	ID	系	M4	M3	M2	M1	追加条件	S1
1	ください	ください系				動詞下さるF命令		*
2	ください	ください系				動詞下さるF連用		*
3	くださって				くださっ	接続助詞でS^て		*
4	Tください	ください系				動詞下さるF命令		*
5	Tください	ください系				動詞下さるF連用		*
6	Tくださって				くださっ	接続助詞でS^て		*
7	なさい	なさい系				動詞為さるF命令		*
8	なさい	なさい系				なさい		*
9	なさり	なさい系				なさり		*
10	なさっ					なさっ		*
11	なさって				なさっ	接続助詞で		*
12	でしょう	でしょう系				ですF意志		*
13	のでしょう	でしょう系			のsの	ですF意志		*
14	んでしょう	でしょう系			のsん	ですF意志		*
15	ましよう					ますF意志		*
16	でし					ですF連用		*
17	まし	ます系				ますF連用		*
18	で一す			格助詞で	感動詞sえ	ですsす		@
19	ま一す	ます系		感動詞sま	感動詞sあー	ですsす		@
20	ま一す	ます系		感動詞sま	補助記号s～	ですsす		@
21	ま一す	ます系			副詞まあ	ですsす		@
22	ま一す	ます系			感動詞sまア	ですsす		@
23	もうす				副詞もう	ですsす		@
24	です	です系				ですt		*
25	のです	です系			のsの	ですt		*
26	んです	です系			のsん	ですt		*
27	ます	ます系				ますt	not_すまん	*
28	やんす			やsや	のsん	ですsす		*
29	ませ	ます系				ますF未然		*
30	ませ	ます系				ますF命令		*
31	でした	でした系			ですF連用	たt		*
32	のでした	でした系		のsの	ですF連用	たt		*
33	んでした	でした系		のsん	ですF連用	たt		*
34	ませんでした	ます系	ますF未然	ずsん	ですF連用	たt		*
35	ました	ます系			ますF連用	たt		*
36	でして				ですF連用	接続助詞でS^て		*
37	まして	ます系			ますF連用	接続助詞でS^て		*
38	ませぬ	ます系			ますF未然	ずsぬ		*
39	ますれ	ます系				ますF假定		*
40	ません	ます系			ますF未然	ずsん		*

A.5 特殊表現の認定規則

表 13 に、特殊表現の認定規則を示す。形態素パターンで、品詞が助動詞の場合は、品詞情報を省略した。S1 はすべて「*」であり、形態素単位で正規化が適用される。

表 13 特殊表現の認定規則

#	ID	系	M4	M3	M2	M1	追加条件	S1
1	じゃない	じゃない系			じゃ	形容詞無いt		*
2	のじゃない	じゃない系		のsの	じゃ	形容詞無いt		*
3	んじゃない	じゃない系		のsん	じゃ	形容詞無いt		*
4	だろう	だろう系				だF意志		*
5	じゃなかったろう	じゃない系		じゃ	なかつ	たF意志		*
?6	のじゃなかったろう		のsの	じゃ	なかつ	たF意志		*
?7	んじゃないなかったろう		のsん	じゃ	なかつ	たF意志		*
8	のだろう	だろう系			のsの	だF意志		*
9	んだろう	だろう系			のsん	だF意志		*
10	たろう					たF意志		*
11	やろう	だろう系				やF意志	not_ばかやろう	*
12	のやろう	だろう系			のsの	やF意志		*
13	んやろう	だろう系			のsん	やF意志		*
14	じゃろう					じゃF意志		*
15	のじゃろう				のsの	じゃF意志		*
16	んじゃないろう				のsん	じゃF意志		*
17	じゃなし				じゃ	形容詞無いsなし		*
18	じゃなかった	じゃない系		じゃ	なかつ	たt		*
19	のじゃなかった	じゃない系	のsの	じゃ	なかつ	たt		*
20	んじゃないなかった	じゃない系	のsん	じゃ	なかつ	たt		*
21	のだった			のsの	だsだっ	たt		*
22	んだった			のsん	だsだっ	たt		*
23	のだ				のsの	だF終止		*
24	んだ				のsん	だF終止		*
25	のや				のsの	やF終止		*
26	んや				のsん	やF終止		*
27	じゃ					じゃt	not_接助ちゃ	*
28	じゃ					じゃ		*
29	のじゃ				のsの	じゃt		*
30	のじゃ				のsの	じゃ		*
31	んじゃ				のsん	じゃt		*
32	んじゃ				のsん	じゃ		*
33	っちゃ					じゃt		*

付録 B 主要素形式 ID 一覧

主要素形式 ID の一覧を表 14 に示す。

主要素形式 ID の #1 から #6 は、活用型と活用形を組み合わせた形式である。活用型は「V」(動詞型活用)、「A」(イ形容詞型活用) のいずれかである。活用形の情報は、表 15 に示す通りである。命令形は、「食べよ・(早く) せよ」の形式は「ヨ」、「書け・食べろ・(早く) しろ」の形式は「命」のように区別する。

文末の付属要素以外の最も文末に近い形態素が「助動詞た」の場合は、もうひとつ前の形態素を述語とみなし、その述語のタ形とみなす。テ形の場合も同様である。

たとえば、「食べる」の ID は〈V〉、「食べよう」の ID は〈V意〉、「食べろ」の ID は〈V命〉、「食べよ」の ID は〈Vヨ〉、「食べた」の ID は〈Vタ〉、「食べて」の ID は〈Vテ〉となる。

表 14 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	

表 15 活用形情報

#	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」)

なお、システムのデバッグを容易にするために、主要素を述語と認定しなかった場合にも、主要素形式 ID (F07) として、品詞情報が出力される。その一覧を表 16 に示す。これらの品詞情報は、文末形式の認定には関与しない。

表 16 デバッグ用の主要素形式 ID

#	ID	認定条件	備考
20	名	名詞	
21	名	接尾辞-名詞的	
22	名	形状詞-タリ	
23	代	代名詞	
24	体	連体詞	
25	副	副詞	
26	副	助動詞だ (連用形「に」)	
27	接	接続詞	
28	感	感動詞	
29	格助 + 出現形	助詞-格助詞	
30	副助 + 出現形	助詞-副助詞	
31	係助 + 出現形	助詞-係助詞	
32	接助 + 出現形	助詞-接続助詞	
33	終助 + 出現形	助詞-終助詞	
34	準助 + 出現形	助詞-準体助詞	
35	接頭	接頭辞	
36	記	記号	
37	句点	補助記号-句点	
38	読点	補助記号-読点	
39	括弧	補助記号-括弧開	
40	括弧	補助記号-括弧閉	
41	補助	補助記号	
42	空	空白	
43	タ	「助動詞た」の直前が存在しない	ほぼ、形態素解析誤り
44	乙タ	連用形以外+「助動詞た」	ほぼ、形態素解析誤り

付録 C 追加条件一覧

追加条件の一覧を表 17 に示す。追加条件は、システムの内部関数として実装されている。

表 17 追加条件

#	追加規則	意味	説明
1	tail	文末に限る	
2	FALSE	必ず偽	認定しない形式を定義する際に使用する
3	is_終助いな	〈終助詞いな〉とみなせる	「い」は小書ではない、「いなあ」ではない
4	is_接続助詞から	〈接続助詞から〉とみなせる	連語を排除する
5	is_接続助詞って	〈接続助詞って〉とみなせる	直前が活用語、接続助詞、終助詞、準体助詞のいずれか
6	is_接続助詞とも	〈接続助詞とも〉とみなせる	直前が意志推量形
7	is_終助い	〈終助詞い〉とみなせる	「ィ」「だーい」「かしい」等以外
8	is_終助だい	〈終助詞だい〉とみなせる	「だーい」等を以外
9	is_終助とも	〈終助詞とも〉とみなせる	〈終助詞とも〉以外
10	is_終助な	〈終助詞な〉とみなせる	文末で、かつ、直前が副詞か「助動詞だ」の出現形「に」
11	is_終助のう	〈終助詞のう〉とみなせる	「このう」「そのう」等以外
12	is_終助や	〈終助詞や〉とみなせる	文末で、かつ、出現形が「や」
13	is_終助やー	〈終助詞やー〉とみなせる	文末で、かつ、直前が動詞のテ形・命令形・已然形・意志推量形
14	is_終助やん	〈終助詞やん〉とみなせる	直前が終止形か連体形
15	is_終助わ	〈終助詞わ〉とみなせる	「わあな」等以外
16	not_いうか	「いうか」ではない	
17	not_おーきに	「おーきに」ではない	
18	not_かあねえ	「かあねえ」ではない	
19	not_か知らね	「か知らね」ではない	
20	not_すまん	「すまん」ではない	
21	not_ぜやー	「ぜやー」ではない	
22	not_ねーや	「ねーや」ではない	
23	not_のにゃ	「のにゃ」ではない	
24	not_ばかやろう	「ばかやろう」ではない	
25	not_わあな	「わあな」ではない	
26	not_テ形	テ形ではない	
27	not_接続助詞ちゃ	「接続助詞ちゃ」ではない	

付録 D 形態素単位の正規化規則一覧

形態素単位の正規化規則の一覧を表 18–19 に示す。この表の出現形の大半は、UniDic (unidic-cwj-3.1.0) に登録されている出現形である。形態素単位の正規化規則は、システムの内部関数として実装されている。

表 18 正規化規則 (1)

#	形態素パターン	正規形	正規化規則	出現形
1	助動詞じゃ F 意志	じゃろう		じゃろう
2	助動詞じゃ t	っちゃ	/っ/	っちゃ
3		じゃ	(2 文字)	じゃ, じゃ, ぢゃ
4		じゃあ	(3 文字)	じゃあ, じゃー, ぢゃあ
5	助動詞ず s ん	ぬ		ぬ
6	助動詞ず s ん	ん		ん
7	助動詞た F 意志	たらふ	/らふ/	たらふ, だらふ
8		たろう	/ろう/, /ろふ/	たろう, たろふ, だろう, だろふ
9		たろ	/ろ/, /ろっ/	たろっ, だろ, だろっ
10	助動詞た t	た	(1 文字)	た
11		たー	(2 文字以上)	たー, た~, たあ, たあ
12	助動詞だ F 意志	たろ		たろ
13		たろう		たろう
14		だら		だら, だらー
15		だろ		だろ, だろっ
16		だろう		だろう, だろー
17	助動詞だ F 終止	た		た
18		だん		だん
19		だ		だ
20		だあ	/~だ.+/	だあ
21		ちゃ	(終止形-融合)	ちゃ, ちゃあ
22	助動詞だ f 連用形-融合 (じゃ)	じゃ	(2 文字)	じゃ, ぢゃ
23		じゃあ	(3 文字以上)	じゃあ, じゃあ, じゃ~, じゃー, じゃア, ぢゃア
24	助動詞だ s だっ	だっ		だっ
25	助動詞です F 意志	でしょう		でしょう, でしょお, でしょー
26		でっしゃろ		でっしゃろ
27		でしよふ	/しよふ/	でしよふ, ででしよふ
28		でしよ	/でしよ/	でしよ, でしよっ
29		っしょう	/しょう/	っしょう, しょう
30		っしよ	/しよ/	っしよ, しよ
31	助動詞です F 連用	でし		でし
32		でち		でち
33	助動詞です s す	です		です
34	助動詞です t	でちゅ		でちゅ
35		でしゅ		でしゅ
36		でっす		でっす
37		でっ		でっ
38		でし		でし
39		す		す, ス
40		でーす		でーす
41		です		です
42		ですー		ですー, ですう
43		でっ	(促音便)	でっ
44		でん	(撥音便)	でん
45	助動詞ます F 假定	ますれ		ますれ, 升すれ

表 19 正規化規則 (2)

#	形態素パターン	正規形	正規化規則	出現形
46	助動詞ます F 命令	まし	/し/	まし, 升し
47		ませい	/せい/, /せー/	ませい, 升せい, ませー
48		ませ	/せ/	ませ, 升せ
49	助動詞ます F 意志	ましよう	/しょう/, /しよう/	ましよう, 升しょう, ましよお, ましよう
50		ましよ	/しよ/	ましよ, 升しよ, ましよっ, 升しよっ
51		ませふ	/せふ/	ませふ, 升せふ
52		まっしゃろ	/っしゃろ/	まっしゃろ, 升っしゃろ
53		まひよ	/ひよ/	ましよ, 升ひよ
54		みせふ	/みせふ/	みせふ
55	助動詞ます F 未然	まへ		まへ, 升へ
56		ませ	/ませ/, /升せ/	ませ, ませー, 升せ
57	助動詞ます F 連用	まし		まし, マシ, 升し
58	助動詞ます t	まちゅ		まちゅ
59		ましゅ		ましゅ
60		まつ		まつ
61		まする	/する/	まする, 升する
62		まっす	/っ/	まっす, まあーっす
63		ますー		ますー
64		まーす		まーす
65		ます		ます, 升, 升す, ㊦
66		まっ	(促音便)	まっ, 升っ
67		まん	(撥音便)	まん, 升ん
68		まさあ	(終止形-融合)	まさ, まさあ, まさア, 升さ
69	助動詞や F 意志	やらふ		やらふ, やろふ
70		やろ		やろ, やろっ
71		やろう		やろう, やろー
72	助動詞や F 終止	や		や, やあ
73	助動詞や s や	や		や
74	動詞下さる F 命令	ください		ください, 下さい
75		くださーい		くださーい, 下さーい
76		くだされ		くだされ, 下され
77	動詞下さる F 連用	ください		ください, 下さい
78		くださり		くださり, くださり
79	動詞下さる f 連用形-促音便	くださっ		くださっ
80	動詞為さる F 命令	なさい		なさい
81		なされ		なされ
82		なはれ		なはれ
83		なせえ		なせえ
84		んさい		んさい
85	形容詞無い s なかつ (なかつ)	なかつ		なかつ
86	形容詞無い s なし	なし		なし
87	形容詞無い s なさい (なさい)	なさい		なさい
88	形容詞無い s なさっ (なさっ)	なさっ		なさっ
89	形容詞無い s なさり (なさり)	なさり		なさり
90	形容詞無い t	ない		ない
91		無い		無い
92		なけん		なけん
93		なあい	/~な.+い/	なあい, なあい, なあい, なあーい, なーい
94		ねえ	/~ね/	ねえ, ね, ねえ, ねー
95	接続助詞て S~て	て		て, てー, てえ, てえ, で, でー
96	終助詞の s の (の s の)	の		の
97	終助詞の s ん (の s ん)	ん		ん

付録 E 同一 ID を持つ付属要素

同一 ID を持つ付属要素の一覧を表 20 に示す。システム出力の 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 20 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 によって区別することができる。

表 20 同一 ID の付属要素

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」

付録 F 確認されている丁寧表現 ID と特殊表現 ID

丁寧表現と特殊表現の認定規則は、正規化を形態素単位で行うので、個々の規則は、一般に複数の ID を認定する。現時点までに存在が確認されている丁寧表現 ID と特殊表現 ID を、表 21 に示す。この表の ‘#’ は、ID を認定する規則番号である。

表 21 確認されている丁寧表現 ID と特殊表現 ID

丁寧表現			特殊表現		
#	ID	# ID	# ID	# ID	# ID
1	ください	15 ましょう	26 んでっ	1 じゃあない	16 んじゃろう
1	くだされ	15 まっしやろ	26 んでっす	1 じゃあねえ	17 じゃなし
1	くださーい	15 まひょ	26 んでつ	1 じゃあ無い	18 じゃあなかった
2	ください	16 でし	26 んで一す	1 じゃなあい	18 じゃなかった
2	くださっ	17 まし	27 まさあ	1 じゃない	18 じゃなかったー
2	くださり	18 で一す	27 ます	1 じゃねえ	19 のじゃなかった
2	くだされ	19 まーす	27 まする	1 じゃ無い	20 んじゃあなかった
3	くださって	20 まーす	27 まちゅ	1 じゃ無え	20 んじゃなかった
4	Tください	21 まーす	27 まっ	2 のじゃあない	20 んじゃなかったー
4	Tくだされ	22 まーす	27 まつ	2 のじゃない	21 のだった
4	Tくだされー	23 もうす	27 まん	2 のじゃねえ	22 んだった
4	Tくださーい	24 す	27 まーす	2 のじゃ無い	22 んだったー
5	Tください	24 っす	28 やんす	3 んじゃあない	23 のた
5	Tくださっ	24 でし	29 ませ	3 んじゃあねえ	23 のだ
5	Tくださり	24 でしゅ	29 まへ	3 んじゃあ無い	23 のだあ
5	Tくだされ	24 です	30 まし	3 んじゃなあい	24 んた
6	Tくださって	24 ですー	30 ませ	3 んじゃない	24 んだ
6	Tくだすって	24 でちゅ	30 ませい	3 んじゃねえ	24 んだあ
7	なさい	24 でっ	31 でした	3 んじゃ無い	24 んちゃ
7	なされ	24 でっす	31 でしたー	4 たろ	25 のや
7	なせえ	24 でつ	32 のでした	4 たろう	26 んや
7	なはれ	24 でん	32 のでしたー	4 だろ	27 じゃ
7	んさい	24 で一す	33 んでした	4 だろう	27 じゃあ
8	なさい	25 のす	33 んでしたー	5 じゃなかったろ	27 っちゃ
9	なさり	25 のっす	34 ませんでした	5 じゃなかったろう	28 じゃ
10	なさっ	25 のでしゅ	34 ませんでしたー	8 のだろ	28 じゃあ
11	なさって	25 のです	34 まへんでした	8 のだろう	29 のじゃ
12	しょ	25 のですー	35 ました	9 んたろ	29 のじゃあ
12	っしょ	25 のでちゅ	35 ましたー	9 んだろ	30 のじゃ
12	でしょ	25 のでっ	36 でした	9 んだろう	30 のじゃあ
12	でしょう	25 のでっす	37 まして	10 たろ	31 んじゃ
12	でっしやろ	25 のでん	38 ませぬ	10 たろう	31 んじゃあ
13	のでしょ	25 ので一す	39 ますれ	11 やろ	32 んじゃ
13	のでしょう	26 んす	40 ません	11 やろう	32 んじゃあ
14	んしょ	26 んっす	40 まへん	12 のやろ	33 じゃ
14	んっしょ	26 んでし		13 んやろ	33 じゃあ
14	んでしょ	26 んでしゅ		13 んやろう	
14	んでしょう	26 んです		14 じゃろう	
14	んでっしやろ	26 んですー		14 ちゃろ	
15	ましょ	26 んでちゅ		15 のじゃろう	

付録 G 主要な文末形式

システムの開発に使用した 2 種類の会話文データのいずれかで 1bp (=0.01%) 以上出現した文末形式 (423 種類、ただし文末形式タイプ「なし (‘-’)」を含む) の一覧を表 22-25 に示す。

これらの表の各欄は、以下の内容を表す。

B 会話データ D1 における出現率 (bp)

N 会話データ D2 における出現率 (bp)

F 女性率 (%)

なお、出現率が 0.1bp 未満である場合、および、女性率が定義されていない場合は、‘-’ と表記している。

これらのデータは、「小説会話文文末形式リスト Kohaku-FFL」の一部である。

表 22 主要な文末形式 (1)

タイプ	ID	B	N	F	タイプ	ID	B	N	F
-	-	1911.01	2633.32	57.3	丁	なさい	27.37	19.57	70.9
裸	V	553.81	496.58	29.9	終	I よ	26.79	24.18	16.2
裸	A	408.78	353.23	33.9	接	けど	26.36	35.13	76.1
裸	だ	384.89	360.95	7.1	丁	でしょ	26.07	27.17	92.1
裸	Vタ	313.91	200.48	33.7	終	わね	25.74	25.47	98.7
丁	ます	270.81	308.86	46.9	裸	V用	25.53	38.95	55.1
特	んだ	253.46	127.08	10.6	終	かい	24.48	12.29	24.3
丁	です	247.56	240.85	42.5	特	じゃ	24.22	13.62	25.6
終	の	230.86	133.37	92.7	特接	んだ. から	24.19	8.86	69.7
終	か	209.24	313.83	17.4	接	が	23.03	28.09	9.2
終	よ	189.86	169.34	39.9	接終	から. な	22.74	44.55	3.9
終	わ	186.96	92.96	98.0	丁終	のです. か	22.42	29.66	44.9
丁	ました	162.49	133.30	48.9	接終	から. ね	22.20	21.96	37.3
終	な	156.27	240.25	10.0	終	のね	21.59	10.76	99.0
裸	V命	142.02	156.09	15.6	終	かね	21.55	5.30	17.8
丁	んです	131.50	31.87	57.6	特	だろ	21.37	49.23	13.6
丁	ません	130.38	98.30	50.6	接	し	21.08	37.39	72.2
終	ね	107.89	91.05	57.4	接	のに	20.97	17.96	70.0
裸	Vテ	106.12	80.09	78.2	終	なあ	20.65	27.76	27.4
終	のよ	97.44	31.38	98.2	終	Tね	20.18	17.57	67.5
終	Xよ	86.67	54.76	91.9	裸	Aテ	19.99	15.65	86.3
丁終	です. か	86.09	91.63	43.5	特	んだろう	18.66	8.71	18.2
終	のか	84.28	112.20	10.7	終	かしら	18.58	17.06	99.5
丁	でしょう	80.01	51.04	62.7	丁終	ません. よ	18.37	14.00	39.9
終	Dよ	77.70	60.92	24.9	終	のさ	17.54	9.97	29.1
終	ぞ	75.10	131.67	9.4	丁	ませんでした	15.47	8.68	53.3
裸	V意	68.88	72.56	19.9	丁	んでしょう	15.22	4.20	70.2
丁	のです	68.84	39.25	47.7	丁終	ました. よ	15.08	7.93	31.4
特	だろう	67.11	59.57	8.0	終	ねえ	14.82	15.73	50.5
裸	Aタ	64.65	35.49	43.3	接	たら	14.57	6.46	70.2
終	さ	64.32	30.32	18.3	終	のかね	14.14	3.36	6.8
特終	んだ. Dよ	59.19	47.73	31.3	丁終	でしょう. か	14.03	23.77	55.8
丁終	んです. か	58.39	44.51	53.8	丁	ください	13.81	16.66	48.4
接	から	55.68	50.53	71.5	特終	んだ. な	13.59	10.94	8.1
丁終	です. よ	53.44	46.32	32.6	丁接	です. から	13.56	11.94	57.7
丁終	です. ね	51.27	86.98	36.5	終	のかな	13.31	17.37	30.5
丁	Tください	48.96	49.63	51.0	丁終	ます. ね	13.20	24.90	35.5
特	のだ	45.95	32.11	5.6	接	ば	13.16	13.45	60.3
裸	だっタ	41.25	17.57	37.9	丁終	んです. ね	13.09	10.05	49.0
特	じゃない	40.89	26.40	63.5	特終	んだ. Dね	12.69	7.27	26.9
丁終	んです. よ	40.71	16.27	47.2	特終	じゃない. の	12.69	3.60	93.5
丁	ましょう	39.70	44.92	51.0	終	のかい	12.65	6.17	25.4
終	Xね	38.47	37.39	94.7	終	Tよ	12.58	8.59	83.7
裸	状X	38.29	45.26	79.8	接	で	11.53	9.72	47.5
接	って	37.64	28.76	63.6	裸	や	11.28	2.45	26.5
終	かな	37.60	65.34	26.4	終	のかしら	11.21	8.83	99.3
終	Bな	36.08	52.45	11.0	丁終	でしょう. ね	11.03	6.51	67.7
丁終	ます. よ	34.82	32.63	34.6	終	や	10.59	6.34	22.9
終	ぜ	34.49	32.78	4.9	丁終	ました. か	10.41	9.10	27.8
特終	じゃない. か	34.02	14.84	14.8	丁接	ます. から	10.20	8.69	59.9
丁終	ます. か	30.95	33.17	39.5	丁終	ましょう. か	10.12	12.84	63.3
丁終	ません. か	30.48	18.00	42.5	丁終	ません. ね	9.40	9.52	32.0
終	Dね	29.79	36.26	26.9	丁接	です. が	9.29	10.59	33.5
丁	でした	29.03	13.14	55.2	特接	んだ. けど	9.22	18.49	81.1
終	わよ	27.66	22.08	99.5	特終	だろう. な	8.97	10.11	5.4

表 23 主要な文末形式 (2)

タイプ	ID	B	N	F	タイプ	ID	B	N	F
丁終	です. Pわ	8.86	11.88	91.5	終	もん	4.70	3.00	74.2
丁	ませぬ	8.79	8.96	43.1	特	やろ	4.70	1.82	42.6
丁接	ます. が	8.42	9.49	19.2	終	かなあ	4.66	7.83	42.9
丁	まする	8.39	10.18	40.6	丁終	んでしょう. か	4.66	3.11	59.8
裸	な	8.24	11.60	40.7	丁終	なさい. I よ	4.59	2.67	91.4
丁終	です. な	8.10	9.74	1.0	特終	じゃない. よ	4.59	2.63	32.9
丁終	ました. ね	8.06	8.64	35.5	終	じゃん	4.52	12.62	72.5
終	んだい	8.06	3.29	22.3	終	とも	4.52	1.83	10.4
裸	A用	7.95	8.09	55.0	接終	から. さ	4.48	3.80	23.8
特	んじゃない	7.85	7.42	76.0	終	もの	4.48	2.49	91.2
丁	んでしょ	7.74	5.38	93.2	特終	のだ. Dよ	4.41	2.38	32.8
終	Bなよ	7.63	10.55	5.3	特	たろう	4.41	0.83	10.4
丁	のでしょう	7.63	8.01	58.6	特終	んだ. もの	4.30	0.85	91.2
接終	が. ね	7.48	1.78	9.4	丁終	でした. か	4.27	4.33	16.2
丁終	のです. よ	7.34	5.98	43.4	終	Xねえ	4.23	5.22	69.6
終	だい	7.34	2.89	15.5	裸	A意	4.23	4.21	13.0
丁接終	です. から. ね	7.30	6.38	24.5	丁終	んです. もの	4.01	0.39	99.2
特接	んだ. って	7.30	3.61	72.1	丁接	んです. けど	3.94	6.59	74.9
丁	まして	7.12	9.45	50.9	丁終	Tください. I よ	3.94	3.35	36.0
終	よね	6.91	16.80	62.5	終	かよ	3.76	7.49	25.4
特	んや	6.83	1.33	29.0	接	ちゃ	3.76	0.83	82.5
特終	んだ. ぞ	6.73	7.04	3.6	接	なら	3.72	4.51	60.0
丁終	ます. Pわ	6.69	7.00	97.5	丁終	です. よね	3.62	13.06	62.4
丁終	まして. Tね	6.62	1.19	21.0	特接	だろう. が	3.58	5.04	4.5
終	Tな	6.47	10.65	16.2	丁接	ました. から	3.58	2.62	63.2
特接	んだ. が	6.36	9.16	3.8	特	んじゃ	3.54	4.57	38.0
特丁終	じゃない. です. か	6.33	8.99	40.9	接終	で. な	3.54	4.49	16.8
接終	けど. ね	6.22	8.20	59.1	丁接	です. けど	3.51	5.44	83.9
特	のじゃ	6.22	7.97	32.0	丁接	ません. から	3.51	3.52	56.4
特	んだろ	6.18	10.08	30.4	特終	だろう. ね	3.51	3.21	17.5
接終	で. ね	6.18	2.33	35.5	丁終	んでしょう. ね	3.51	1.84	69.7
丁接	んです. が	6.04	4.12	33.3	特終	だろう. か	3.47	3.92	13.3
丁終	のです. ね	5.97	6.83	47.3	接	ので	3.40	11.31	49.6
特終	じゃねえ. か	5.93	6.48	2.8	接終	から. よ	3.40	1.74	86.1
丁接	んです. って	5.93	1.05	97.2	特終	んじゃない. かしら	3.40	1.33	99.0
終	Rな	5.82	4.25	43.5	終	のう	3.33	4.23	27.8
特終	んだ. ぜ	5.79	2.86	6.9	丁終	ます. な	3.29	3.94	0.0
裸	Vヨ	5.71	8.70	25.0	終	っけ	3.25	5.84	54.2
接	と	5.68	7.08	44.7	特終	んだろう. な	3.25	3.75	4.9
特終	じゃない. わ	5.53	1.26	99.7	特接終	んだ. から. な	3.25	1.86	2.5
終	で	5.46	2.60	39.7	特接終	んだ. から. ね	3.25	1.26	56.6
丁接	んです. から	5.39	2.21	54.2	丁終	ません. Pわ	3.22	3.37	99.5
接終	が. な	5.31	7.03	7.5	接終	し. ね	3.18	5.93	52.5
特終	んじゃない. か	5.28	6.99	7.6	終	い	3.07	4.52	28.1
特終	んだ. わ	5.28	0.91	99.7	丁終	ました. Pわ	3.07	2.76	96.1
特終	んじゃない. の	5.17	3.18	84.7	特終	のだ. な	3.04	3.47	2.6
特	のだろう	5.13	5.60	15.3	接	んで	3.04	2.65	23.3
丁終	のでしょう. か	5.10	8.36	57.0	接	けれど	3.04	2.38	84.5
丁接	です. って	5.10	1.78	79.5	終	よお	3.00	10.87	57.0
特終	んじゃない. かな	4.99	4.07	17.9	終	Bなあ	3.00	4.62	13.0
終	よな	4.84	11.64	8.1	終	わねえ	3.00	2.08	98.9
丁接	のです. が	4.84	9.84	36.0	丁終	です. ねえ	2.96	5.57	41.5
丁接終	ます. から. ね	4.81	4.03	35.4	終	Dねえ	2.93	5.38	53.4
終	のよね	4.74	5.75	99.0	丁接終	ません. から. ね	2.93	1.89	41.6
丁接	ません. が	4.70	4.61	29.7	終	もんか	2.93	0.92	13.2

表 24 主要な文末形式 (3)

タイプ	ID	B	N	F	タイプ	ID	B	N	F
接終	し. な	2.89	9.85	5.5	裸	だあ	1.92	5.61	48.4
特終	んじゃない. のか	2.89	2.42	2.8	終	さあ	1.88	3.47	73.9
終	ねん	2.89	1.22	39.5	特	じゃろう	1.88	3.05	37.6
丁終	ましよう. よ	2.89	1.03	86.1	丁終	ません. な	1.88	1.57	4.3
丁接終	んです. が. ね	2.86	0.48	16.0	丁接終	ました. から. ね	1.88	1.28	32.5
丁終	でした. ね	2.82	3.25	42.4	丁接	ます. と	1.88	0.89	36.6
特終	んだ. Dよね	2.78	9.17	56.3	特丁	じゃない. んです	1.88	0.32	61.2
特丁終	んじゃない. です. か	2.78	2.33	30.0	終	え	1.84	0.30	68.4
終	わい	2.78	1.36	8.4	特終	んだ. Dよな	1.81	5.15	4.4
終	Xよね	2.75	4.55	100.0	終	もんね	1.81	3.06	84.9
終	かあ	2.71	9.64	56.9	接終	って. ね	1.81	0.58	53.7
終	ん	2.68	3.34	57.8	特丁終	んじゃない. でしょう. か	1.81	0.53	46.7
丁接	ました. が	2.68	2.65	22.0	特終	のだ. ぞ	1.77	2.50	4.6
接終	けど. な	2.60	6.94	11.7	接終	けど. さ	1.74	2.56	52.1
裸	だっ	2.60	3.24	17.0	特	じゃねえ	1.74	1.46	4.6
接終	んで. ね	2.60	0.59	6.1	丁終	ませんでした. よ	1.70	0.64	37.8
特接終	んだ. が. ね	2.60	0.32	3.2	丁終	んです. かね	1.66	1.97	21.5
終	Dよね	2.57	9.52	62.6	接終	で. さ	1.66	0.91	29.9
特終	んだろう. ね	2.53	1.65	40.0	特	じゃなかった	1.66	0.71	54.6
終	わよね	2.46	3.89	100.0	丁終	です. Pわね	1.63	5.06	98.0
丁接	ます. けど	2.46	3.36	79.4	接終	から. なあ	1.63	4.43	18.8
終	んか	2.46	1.18	29.4	丁終	ました. な	1.63	1.45	0.0
接	たって	2.46	0.90	56.7	丁終	ました. Pの	1.63	1.05	100.0
特終	じゃない. のよ	2.46	0.56	98.0	丁接終	ます. が. ね	1.59	0.40	5.1
丁接	ます. ので	2.42	5.55	38.2	特接	んだ. し	1.55	1.55	72.1
丁終	でした. よ	2.42	1.18	31.4	丁終	でしょう. な	1.52	1.26	5.2
特接	のだ. が	2.39	4.03	12.9	丁終	です. とも	1.52	0.46	40.3
丁接終	んです. から. ね	2.39	0.39	31.4	終	かえ	1.52	0.16	60.0
終	のかよ	2.31	4.15	13.4	丁接終	んです. けど. ね	1.48	1.99	48.8
裸	A仮	2.31	1.37	78.5	裸	A幹	1.48	1.63	42.1
特	のや	2.31	0.16	36.9	特終	のだろう. か	1.48	1.06	15.1
丁終	です. かね	2.28	4.82	16.2	丁接	ました. けど	1.45	1.24	75.6
特接	のだ. から	2.28	1.76	61.0	終	のお	1.41	4.04	88.2
丁接	でしょう. が	2.24	1.42	49.1	丁終	Tください. な	1.41	0.70	88.2
終	べ	2.21	1.19	47.1	丁終	ます. よね	1.37	3.80	68.1
特接終	んだ. けど. ね	2.17	4.08	61.6	接終	って. さ	1.37	1.62	37.5
特終	んだ. もん	2.17	1.06	84.0	接	ては	1.37	1.04	62.1
丁終	ませんでした. か	2.13	1.05	28.8	特丁	じゃない. でしょう	1.37	0.30	73.5
丁終	でして. Tね	2.13	0.33	14.5	接終	から. ねえ	1.34	2.59	64.3
終	Dよな	2.10	6.73	14.5	丁終	のでしょう. ね	1.34	1.19	42.9
丁接	のです. から	2.10	2.18	48.9	終	かしらね	1.30	1.64	100.0
丁	ませ	2.10	1.41	82.7	特終	じゃない. かな	1.30	1.38	27.3
丁接終	です. が. ね	2.10	0.48	3.8	丁	Tくだされ	1.30	0.82	28.0
丁終	です. Pの	2.06	2.46	99.2	丁終	ます. ぞ	1.27	2.14	12.0
丁終	です. もの	2.06	1.21	98.4	終	もんな	1.27	2.14	6.2
特終	じゃ. な	2.02	3.80	19.0	丁	ましょ	1.27	1.58	93.1
接終	と. ね	2.02	2.27	62.8	丁接	でした. から	1.27	0.70	66.7
特終	じゃない. のか	2.02	1.92	0.0	接終	ば. ね	1.27	0.32	58.2
丁終	んです. Pわ	2.02	0.41	73.6	特	たろ	1.23	1.81	17.9
丁終	んです. よね	1.99	6.01	65.1	丁終	です. かな	1.23	1.55	0.0
終	Rなよ	1.99	2.12	46.0	特終	じゃない. わよ	1.23	0.72	98.7
丁終	んです. Pの	1.99	1.22	100.0	終	らあ	1.23	0.18	0.0
丁終	Tください. ね	1.95	5.73	89.3	特接終	んだ. が. な	1.19	2.23	0.0
丁	でして	1.95	1.98	16.4	丁終	ませぬ. か	1.19	1.21	52.9
特終	だろう. よ	1.95	1.45	22.3	特接終	のだ. から. な	1.19	1.12	0.0

表 25 主要な文末形式 (4)

タイプ	ID	B	N	F	タイプ	ID	B	N	F
丁終	でしょう. よ	1.19	0.28	83.2	特終	じゃねえ. よ	0.58	1.10	-
丁接終	ません. が. ね	1.19	0.24	6.6	丁終	でした. っけ	0.54	1.00	-
丁接	ません. けど	1.16	1.20	81.1	丁終	ます. ねえ	0.51	1.46	-
丁終	ましょう. ね	1.16	1.13	90.7	丁終	ません. よね	0.51	1.44	-
終	かねえ	1.16	0.99	54.1	終	X よお	0.51	1.41	-
丁接	でしょう. から	1.16	0.91	66.5	特接終	んだ. けど. さ	0.51	1.23	-
特終	んじゃない. よ	1.16	0.51	66.5	特接	のだ. けど	0.51	1.23	-
特接	じゃない. けど	1.16	0.42	70.6	特終	だろう. さ	0.51	1.12	-
丁接	ました. ので	1.12	1.35	55.3	特終	んだ. D よなあ	0.47	2.05	-
丁	まへん	1.12	-	47.9	丁	T ください. ませ	0.47	1.99	-
特接終	んだ. けど. な	1.08	2.75	13.8	丁接	ません. し	0.47	1.59	-
接終	と. な	1.08	2.69	7.2	丁終	んす. よ	0.47	1.10	-
終	T ねえ	1.08	1.94	27.3	特接終	んだ. けど. なあ	0.43	1.57	-
丁終	ます. かね	1.08	1.45	0.0	終	のかあ	0.43	1.16	-
特	じゃない. だろう	1.08	0.59	7.2	特接終	のだ. が. な	0.43	1.04	-
丁	まーす	1.05	2.17	79.3	接終	けど. ねえ	0.43	1.01	-
丁終	ました. よね	1.05	1.94	73.8	特終	んだ. D よねえ	0.36	1.89	-
丁終	ます. P の	1.05	1.73	100.0	丁接	ました. し	0.36	1.25	-
特終	じゃ. よ	1.05	0.85	15.3	丁終	っす. か	0.33	3.52	-
丁終	ません. P の	1.05	0.82	100.0	丁接	です. ので	0.33	2.01	-
丁終	です. ぜ	1.05	0.40	0.0	丁終	です. かあ	0.33	1.34	-
丁接終	です. けど. ね	1.01	1.62	42.9	丁終	ます. よお	0.33	1.32	-
接終	ので. ね	1.01	0.51	7.7	丁終	ます. P わね	0.33	1.20	-
特終	のだ. D ね	1.01	0.36	27.3	丁終	っす. ね	0.29	4.24	-
丁	なされ	1.01	0.16	62.3	丁接	でしょう. し	0.29	1.15	-
丁終	んです. な	1.01	-	0.0	丁終	です. よお	0.25	2.12	-
特	んだあ	0.98	2.01	-	丁終	です. よねえ	0.25	1.28	-
終	かの	0.98	1.99	-	終	D よなあ	0.22	1.71	-
丁接	ません. ので	0.98	1.67	-	丁終	ます. P わよ	0.22	1.30	-
終	よなあ	0.94	2.06	-	接終	って. の	0.22	1.25	-
終	D よお	0.90	3.20	-	丁終	です. P わよ	0.22	1.10	-
丁	っす	0.87	13.18	-	接終	し. なあ	0.18	1.17	-
丁接	です. し	0.83	5.10	-	終	D よねえ	0.14	1.44	-
特終	んだ. っけ	0.83	1.75	-	丁	ましたー	0.11	1.37	-
接終	ので. な	0.83	1.47	-	特接	のだ. けれど	0.11	1.08	-
接終	ば. な	0.83	1.21	-	終	I よお	0.11	1.01	-
特終	んじゃねえ. よ	0.83	1.12	-	丁終	のです. よね	-	1.04	-
丁	くだされ	0.83	1.11	-	特終	んじゃ. ね	-	1.05	-
裸	V 仮	0.80	4.06	-					
丁	す	0.80	2.34	-					
特終	んだ. D よお	0.80	1.60	-					
特終	んじゃねえ. か	0.80	1.55	-					
丁終	です. ぞ	0.80	1.46	-					
丁終	っす. よ	0.76	4.63	-					
丁接	ます. し	0.76	3.07	-					
特終	じゃない. ぞ	0.76	1.14	-					
特接	だろう. し	0.69	1.98	-					
特接	だろう. に	0.69	1.51	-					
丁	ください. ませ	0.69	1.33	-					
接終	し. さ	0.69	1.16	-					
終	かのう	0.69	1.13	-					
丁	で一す	0.69	1.01	-					
終	よねえ	0.65	1.81	-					
丁終	す. か	0.61	1.07	-					
丁終	んす. か	0.58	2.41	-					