

令和 5 年 4 月 26 日現在

機関番号：20103

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K11991

研究課題名（和文）物語文と日常会話文における発話文体と話者属性の分析

研究課題名（英文）Analysis of Speech Style and Speaker Attributes in Narrative and Daily Conversational Sentences

研究代表者

村井 源（Murai, Hajime）

公立はこだて未来大学・システム情報科学部・教授

研究者番号：70452018

交付決定額（研究期間全体）：（直接経費） 2,000,000円

研究成果の概要（和文）：日本語の発話文では同じ内容であっても話者や聴者等に関する様々な属性の影響を受けて発話の文体が多様に変化する。そこで応募者が開発してきた日本語の物語文における発話文のコーパスと、国立国語研究所で開発された日常会話コーパスを基礎として、文体に影響を与えると考えられる様々な属性を各発話文にタグとして追加し、発話文中の文体を構成する主要な要素を対象として統計的な手法で分析した。結果として、日本語の発話文における数種類の基本的な文体のパターンを明らかにした。またこれらの文体のパターンの利用の属性による差異や物語と日常会話による差異を明らかにした。

研究成果の学術的意義や社会的意義

日本語における発話の文体と属性の関係が明らかになることで、コンピュータを用いた話者とその属性の自動推定や、より自然な対話型人工知能の開発などが可能になると期待される。またこれらの成果は国語教育や小説の執筆の補助などにおいても有効に活用可能であると考えられる。

研究成果の概要（英文）：In Japanese utterances, even if the content is the same, the style of the utterance changes diversely under the influence of various attributes of speakers and listeners. Based on the corpus of Japanese narrative sentences developed by the applicant and the daily conversation corpus developed by the National Institute for Japanese Language and Linguistics, various attributes that are thought to affect the writing style are added to each utterance sentence. They were added as tags, and the main elements that make up the style of the utterances were analyzed using statistical methods. As a result, several basic stylistic patterns in Japanese utterances were extracted. In addition, the differences between the usages of these stylistic patterns depending various attributes and the differences between stories and daily conversations were also clarified.

研究分野：Digital Humanities

キーワード：発話 文体 属性 物語 日常会話

1. 研究開始当初の背景

日本語の発話文体が話者や聴者の性別や年齢、相手との関係性などの多様な属性に影響を受けてさまざまに変化することは言語学や社会学など種々の学問分野で指摘され、例えば性別による変化などの顕著な性質に関してはこれまでも多くの研究が積み重ねられてきている。しかしながら、特定の話者の発話文体は単純に性別のような一つの属性のみで決まるわけではなく、複数の属性の相互の複雑な関係性から生成されると推測される。このような複数の属性に基づく発話文体の多様な変化のメカニズムを科学的に明らかにできれば、発話の理解と生成の双方において非常に有益であると考えられる。例えば、コンピュータを用いた話者とその属性の自動推定や、より自然な対話型人工知能の開発などが可能になると期待される。またこれらの成果は国語教育や小説の執筆の補助などにおいても有効に活用可能であると考えられる。

2. 研究の目的

(1) 発話の文体的特徴と話者や聴者の属性の関係性の総体的な特定

多数の属性の複雑な関係から決定されると推測される発話文の文体的特徴の構成要素を分解し、各属性との文体上の要素の関係性を分析する。また発話文体と複数属性の組み合わせでの頻出パターンの分析から、基本的な発話文のスタイルを特定することを目標とする。

発話文と属性の関係が明らかになればそれらに基づいて発話文から話者の特徴や感情状態、話者の置かれている状況を推測することが可能になる。また特に物語テキストなどで発話者が明示されない場合の発話者推定も文体的特徴に基づいて精度を向上できると期待される。

逆に特定の属性や状況設定を与えた場合それに応じたより自然な発話文を生成するために必要となる基礎データとしても寄与できると期待される。

(2) 物語文と日常的発話の相違分析

物語のように文字だけで発話者の意図などの種々の情報を伝えるメディアの場合、発話者の感情状態や環境などを全て文字情報として伝える必要がある。そのため読者の理解度の向上を図るために、現実の日常的な対話よりも、様々な示唆的な情報を発話文に付加する可能性が考えられる。またエンターテインメント性の強い作品群では特に登場キャラクターの個性を強調する傾向や、複数キャラクターの発話を容易に弁別できる読みやすさのために、誇張した特徴的な文体が頻繁に用いられることが先行研究などで指摘されている。しかしながらこのような物語における発話文体と日常会話文体の解離の計量的な分析は行われておらず、本研究のコーパスに基づくアプローチで従来の仮説の科学的な検証とともに、物語的・日常会話的それぞれの発話文を生成するために必要な種々のパラメータを特定することを目指す。

(3) 場の「空気」を理解する自然言語処理に向けて

発話文体での話者属性と平均的基本パターンの関係性に合わせて、発話者の意図(語用論的機能)もデータ化しかつ計量的に解析してパターンを抽出できれば、より深い発話の解釈の機械的な実現に向けて道が開かれる。これは換言すれば人間のような場の「空気」を読む言語処理の実現のための基礎データとして有益であると考えられる。

3. 研究の方法

(1) 手法の概要

物語テキストの分析対象としては、国立国語研で開発され公開されている現代日本語書き言葉均衡コーパス(BCCWJ)からランダムサンプリングした日本語の物語における発話文に対し人手で種々の属性を付与した属性付きの物語的発話コーパスを基礎とした。また現実的な日常会話の分析対象としては同じく国立国語研究所から公開されている日本語話し言葉コーパス(CSJ)に合わせて、「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」プロジェクトで構築された日常会話コーパスを用いた。

言語学、社会学等の先行研究に基づき、発話文体に対して影響を与えると考えられる属性や文脈の候補を列挙し、属性の定義表を作成した。また、属性を定義表に基づき人手で分類し、既存のコーパスに含まれないものをタグとして付与した。また、付与されたタグに対して複数分析者で一致度を測り、科学的な有効性を担保した。その後、属性タグの付与された発話文に対して種々の検定や多変量解析などの統計学的な分析を行い、相関的な関係にある属性と文体要素のペア、基本的な文体のパターン等を計量的に抽出した。また、合わせて語用論的な意図の分析とパターンの抽出と、分析によって得られた知見に基づいた大規模言語モデルを用いた発話文の生成実験も行った。

(2) 研究計画詳細

助詞・助動詞などの機能語の利用における文体的な特徴の関係性の抽出

機能語である助詞・助動詞は一般的に属性に関する文体的な変化の現れやすいと言われており、それらの出現パターンを種々の属性に照らして統計的に分析することで、主要な文体的特徴

と属性の関係性が得られると期待される。これまでに行った物語文での発話に対する予備的な分析では実際に種々の属性が特定の助詞・助動詞と相関関係にあり統計学的な有意差があることが明らかとなっている。本研究で対象となる属性の種類を増やし日常会話コーパスも対象に含めることで、より詳細で一般的な発話文体における機能語と属性の関係性の抽出を行い、また物語文体と日常会話文体の相違を特定した。

人称代名詞、主体の呼称表現における文体的特徴の抽出

機能語以外で属性による文体的な変化が表れやすいと言われているのが人称代名詞、特に話者の自称と聴者に対する呼びかけの対称である。本研究では、発話文中での話者や聴者を示す単語を特定し、それらの属性に対応した出現パターンとその属性による差異の計量的な分析を行った。

語用論的な発話意図の特徴と話者・聴者の属性の関係性の抽出

発話における単語列やその文法的な分類(平叙文か疑問文かなど)と、発話に背後にある話者の意図は単純な対一関係ではなく、種々の属性やコンテキストに依存して変化すると一般的に考えられている。将来的には計算機を用いた機械的な高精度の意図分類の実現が期待されるが、そのためにはまず、発話に対して意図や話者の属性等が付与された基礎的なデータセットの構築とその分析が必須である。本研究では先行研究を参考に発話の持つ語用論的な機能に焦点を絞り、帰納的に19のカテゴリーを作成し人手でのタグ付与を行った。また、付与された語用論的機能と話者や聴者に付与された種々の属性に対して統計的な分析を行い、属性による特徴や時系列的な意図の繊維のパターンを抽出した。

4. 研究成果

(1) 発話における基本的な文体のパターン

物語の発話文と、日常会話の発話文のそれぞれに対して、文体的特徴を構成する要素の一つである機能語(文章の内容以外の意味を示す、日本語では例えば助詞や助動詞など)に着目し、因子分析を行うことで同時に用いられやすい機能語の組み合わせのパターンを抽出した。また抽出されたパターンに対して用いられる助詞・助動詞の特徴から図1のように命名を行った。図1の6つの基本パターンは、物語の会話と日常会話の両方において共通に見られたものである。

- 第1因子:「中立」特定の属性のない一般的な機能語
- 第2因子:「方言」関西弁などの方言に対応
- 第3因子:「気さく」くだけた会話的な表現に対応
- 第4因子:「丁寧」丁寧表現に対応
- 第5因子:「女性」ステレオタイプな女性語に対応
- 第6因子:「粗野」粗暴や人物や敵対関係の表現

図1 物語での会話と日常会話に共通する基本の6文体

図2は物語の会話と日常会話のいずれにしか出現しなかったいくつかのパターンも含めた基本的な発話のパターンと機能語の関係を示したものである。

また、これらの基本パターンと発話における種々の属性との関係にいくつかの特徴がみられることも明らかになった。表1は年齢や性別と、用いられる基本的な文体のパターンにどのような関係があるかを示したものである。また表2は同様にして、話者と聴者の関係と、用いられる基本的な文体のパターンにどのような関係があるかを示したものである。表より、同じ性別でも世代によって平均的な発話の文体の傾向が大きく異なることが明らかである。また、発話の文体は人間関係によっても様々な種類の影響を受けることも実データに基づき確認された。

(2) 話者の属性と語用論的機能の特徴

話者などの属性の付与された発話に対して、語用論的な機能と属性の関係性を抽出した。例として表3には、物語の会話の場合の年齢と性別による頻出意図の相違を示している。他にも、話者と聴者の人間関係による特徴の差異、物語の会話と日常会話による特徴の差異(表4)などが同様に抽出された。

(3) 自称(一人称)・対称(二人称)の表現の特徴

同様のコーパスに基づいて、自称(一人称)・対称(二人称)を示す語彙を抽出し、分類して物語における会話文と日常会話文での差異を比較した(表5)。結果として二種類の発話で用いられる自称詞と対称詞の語彙は大きく異なり、物語において用いられる発話の方がより語彙が豊富であることが明らかになった。また発話中での出現の割合も物語での会話の方が著しく多い

という結果が得られた。しかし、特定の話者が用いる呼称の組み合わせのパターンは一部を除いて物語の会話と日常会話で概ね類似していた。

また、呼称が用いられる理由を調査するため、発話における語用論的な機能の分類と集計を行った。結果として、日常会話では人間間の摩擦を避けるような機能の選択が行われるのに対して、物語の発話は人間間の関係性のダイナミズムを誇張して描く傾向があることが明らかになった。

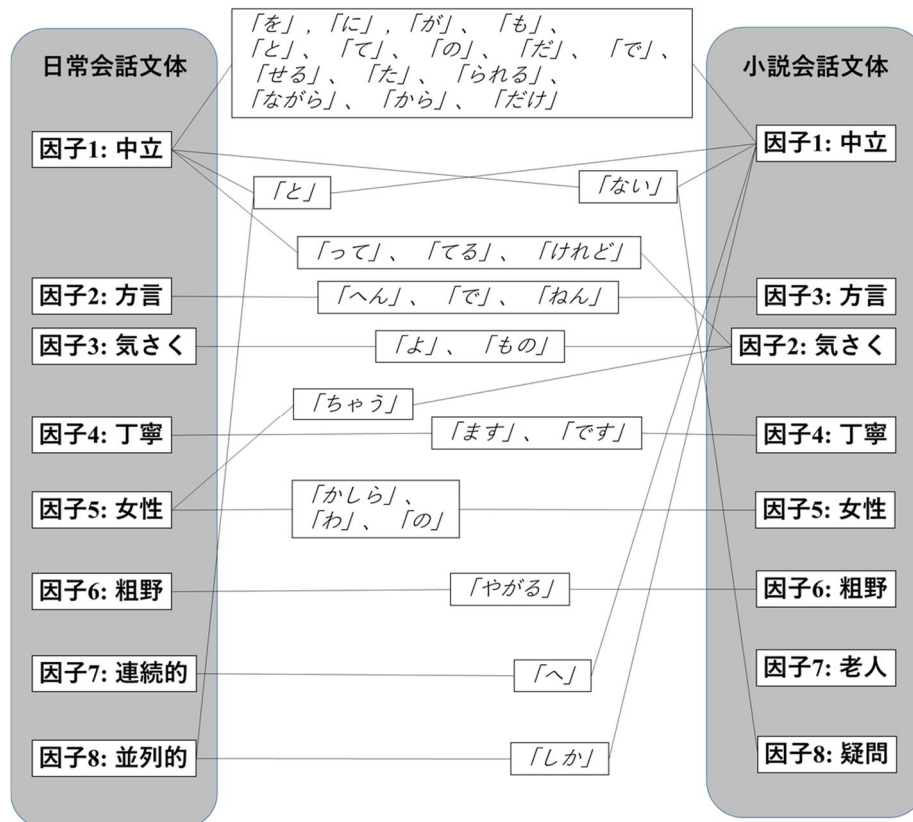


図2 物語での会話と日常会話の文体と機能語の関係性

表1 会話の基本パターンと年齢・性別の関係

	年齢	因子1	因子2	因子3	因子4	因子5	因子6
男性	10 ~ 29	0.09	0.00	0.42	0.22	-0.23	0.43
	30 ~ 49	0.12	0.39	0.06	0.11	-0.09	0.10
	50 ~ 69	-0.18	-0.09	-0.13	-0.16	-0.23	-0.11
	70 ~	-0.04	-0.07	-0.12	-0.16	-0.03	-0.07
女性	10 ~ 29	-0.21	-0.06	-0.17	-0.17	-0.32	-0.08
	30 ~ 49	0.26	-0.08	0.19	0.14	0.45	-0.02
	50 ~ 69	0.18	-0.01	0.08	0.11	0.29	-0.05
	70 ~	-0.26	-0.09	-0.27	-0.14	0.18	-0.12

表2 会話の基本パターンと話者と聴者の関係

	因子1	因子2	因子3	因子4	因子5	因子6
家族	0.00	0.04	0.24	-0.28	-0.09	0.22
家族親族	-0.32	-0.09	-0.22	-0.15	-0.02	-0.16
友人	0.07	0.05	0.08	0.11	0.15	-0.02
教師と生徒	1.03	-0.07	-0.03	1.83	0.15	0.01
ビジネス	0.36	-0.09	-0.27	0.52	-0.01	-0.07
同僚	-0.20	-0.07	-0.09	-0.05	0.00	-0.19
セールス	0.28	-0.09	0.26	0.26	0.20	-0.04

表3 物語の会話での語用論的機能と年齢・性別の関係

	男性	女性

	若年層	成年層	若年層	成年層
陳述	112 W	1583 ▲▲	132 W	637
質問	98 ▲▲	636	83 ▲	244 W
思考	55	531	35 W	290 ▲▲
依頼	40	360	43	153
応答	25	260	32	128
評価	23	225	20	124 ▲
他責	25	227	35 ▲	87
了解	14 V	270 ▲▲	27	65 W
導入	21	208	37 ▲▲	64 W
意思	16	154	11	80
提案	18	125 W	6 W	96 ▲▲
拒否	18	117 W	20	69
感情	21 ▲▲	81 W	26 ▲▲	60
配慮	7	44	7	32
自責	7	34	6	30
将来	4	49	8	15
願望	6	34	4	21
感謝	4	23	4	29
冗談	2	17	0	10

これらの発話の機能と呼称の関係を分析した結果、話し言葉と書き言葉で機能の出現割合が異なるにもかかわらず、個々の機能での呼称の用いられ方は類似していることが明らかになった。機能ごとの傾向としては、話者の知識や思考、意志や感情を示す場合には自称詞が用いられやすく、呼びかけや依頼、質問や批判など聴者に対して何らかの働きかけをする機能を用いる場合には対称詞が出現しやすいことが示された。

表4 物語と日常会話の語用論的機能の差異

	話し言葉	書き言葉
応答	2499	445
陳述	1368	2464
質問	▲▲704	1061
思考	512	911
了解	▲▲472	376
評価	▲▲307	392
提案	▲168	245
感情	▲▲138	188
依頼	W 129	596
拒否	W 76	224
他責	W 59	374
導入	W 46	330
願望	▲51	65
意思	W 40	261
将来	42	76
配慮	V 28	90
自責	27	77
感謝	23	60
冗談	12	29

	(CEJC)	(BCCWJ)
私・私達	0	423
俺・俺達	65	197
あたし・あたしたち	57	69
僕・僕達	15	79
我・我々	0	41
氏名	1	32
自分	11	11
わし	0	16
肩書・職業	0	15
家族関係	7	8
拙者	0	12
あだ名	1	5
位置関係	3	3
人数	3	3
年齢・性別	0	1
合計	163	915

表5 物語と日常会話の自称の差異

自称詞	話し言葉	書き言葉

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 村井源, 松本斉子	4. 巻 32(1)
2. 論文標題 会話文での自称詞と対称詞の出現傾向と役割 - 話し言葉と書き言葉での相違から -	5. 発行年 2022年
3. 雑誌名 情報知識学会論文誌	6. 最初と最後の頁 3-14
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2021_052	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 村井源, 石川一稀	4. 巻 31(2)
2. 論文標題 物語会話文での自称詞と対称詞の分類	5. 発行年 2021年
3. 雑誌名 情報知識学会論文誌	6. 最初と最後の頁 167-172
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2021_022	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 村井源, 松本斉子	4. 巻 60
2. 論文標題 物語テキストにおける会話文の意図の話者属性による特徴	5. 発行年 2020年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 135-143
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Hajime Murai	4. 巻 2018
2. 論文標題 Factor Analysis of Utterances in Japanese Fiction-writing Based on BCCWJ Speaker Information Corpus	5. 発行年 2018年
3. 雑誌名 Advances in Human-Computer Interaction	6. 最初と最後の頁 1-9
掲載論文のDOI (デジタルオブジェクト識別子) 10.1155/2018/5056268	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 村井源, 松本育子
2. 発表標題 日常会話コーパスにおける発話機能のアノテーション付与
3. 学会等名 第65回人工知能学会ことば工学研究会
4. 発表年 2021年

1. 発表者名 Hajime Murai
2. 発表標題 Japanese Daily Utterance Styles: A Factor Analysis based on Balanced Corpus
3. 学会等名 The 33rd Pacific Asia Conference on Language, Information and Computation (国際学会)
4. 発表年 2019年

1. 発表者名 Hajime Murai
2. 発表標題 Factor Analysis of Japanese Daily Utterance Styles
3. 学会等名 LREC 2018 Joint Workshop LB-ILR2018 and MMC2018 (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------