

『太陽コーパス』の口語会話部分に現れる 動詞の待遇表現の計量的分析 —聞き手に対する尊敬待遇表現を中心に—

近藤 明日子 (国立国語研究所コーパス開発センター) [†]

A Quantitative Analysis of Respectful Verbs to Addressees of Conversations in *Taiyo Corpus*

KONDO Asuko (National Institute for Japanese Language and Linguistics)

1. はじめに

近代の話し言葉における動詞の待遇表現形式は現代語以上に多様性があり、先行研究において形式間の用法の差異の解明が進められてきた¹。形式差は多くの場合位相差に基づくものと考えられ、使用年代や話し手の性別、話し手の社会的階層、話し手と聞き手との関係といったものが考察の観点として取り上げられてきた。

このように複数の形式と複数の位相との関係性の考察に重点がおかれる待遇表現研究は、多数の変数の分析に長けたコーパス言語学的手法と親和性が高いと考えられる。本稿はその一つの試みとして、代表的な近代語のコーパスである『太陽コーパス』(国立国語研究所(編)2005)中の小説・戯曲に含まれる口語体の会話部分から、主要な動詞の尊敬待遇表現の形式を抽出し、テキスト分類に用いられるランダムフォレストを用いて、形式の選択における位相の重要度を分析する。

2. 調査対象テキスト

調査に利用するコーパスは、1895(明治28)年から1928(昭和3)年まで刊行された総合雑誌『太陽』をコーパス化した『太陽コーパス』である。『太陽コーパス』には、1895・1901・1909・1917・1925年の5カ年分、各年12冊、計60冊の全文テキストと、XMLタグによる記事・引用・文字・表記等に関する情報が収録されている(田中2005)。このコーパスに対して、近代文語文を対象とする形態素解析辞書「近代文語 UniDic」(小木曾2009)と旧仮名遣いの口語文を対象とする形態素解析辞書(小木曾2012)を用いて形態素解析を行い、形態論情報を付与したデータが国立国語研究所の形態論情報データベース(小木曾・中村2011)に格納されている。本稿ではこのデータベースの2013年12月時点のデータを用いた²。

調査対象はコーパス中の小説・戯曲に含まれる口語体の会話部分である。これは、コーパス中で最も話し言葉的性質の強い文章であり、当時の話し言葉での実態を明らかにするのに最適なテキストと考えられる。ただし、近代の日本を舞台とする作品に限り、時代物や翻訳作品等は除外した。対象の抽出方法は次のとおりである。まずXML記事タグのジャンル属性を利用し、属性値のNDC番号が「912」(日本文学/戯曲)または「913」(日本文学/小説・物語)の作品を抽出し、そのなかから手作業で近代を舞台とするものを選別した。そこから、XML引用タグの種別属性値が「会話」、記事タグまたは引用タグの文体属性が「口語」の部分抽出し調査対象テキストとした。

[†] kondo@ninjal.ac.jp

¹ 近代のある程度の年代にわたる複数の資料を対象に、複数の動詞の待遇表現形式について考察を行った先行研究として、山西(1972)・辻村(1974)・近藤(2005)等がある。また、命令表現に焦点をあてたものとして陳(2005a)(2005b)(2006a)(2006a)等の一連の研究がある。

² データは整備途中のものであり、今後データの変更に伴い、本稿に挙げた数値も変動する場合がある。

調査対象テキストの延べ語数(助詞・助動詞を含む)と口語会話を含む作品数³とその著者数(異なり)を刊行年別に示したものが表1である。

表1 調査対象テキストの言語量

	1895	1901	1909	1917	1925	通年
延べ語数	41704	53124	96412	76562	31678	299480
作品数	15	13	21	22	22	93
著者数	14	8	17	17	21	67

3. 動詞の尊敬待遇表現形式の抽出

調査対象とする動詞の尊敬待遇表現形式の抽出手順は次のとおりである。まず、先行研究や辻村(編)(1991)にあげられた諸形式についてコーパスでの出現頻度の予備調査を行い、出現頻度の多い13形式を対象に選択した。ただし、尊敬の助動詞「れる・られる」による形式については「れる・られる」の意味分類が容易でないため、本稿では調査対象外とした。コーパスからの抽出は形態論情報に拠った。各形式とその検索条件を示したものが表2である。形式中の「～」には動詞連用形あるいはサ変動詞語幹が入り⁴、「御」は「～」に入る語によって「お」または「ご」になることを意味する。

表2 調査対象とする動詞の尊敬待遇表現形式

形式	検索条件	
御～あそばす	語彙素「遊ぶ」	語彙素「御覧」 ／「御免」
いらっしゃる	語彙素「いらっしゃる」	
～ていらっしゃる		
おっしゃる	語彙素「仰る」	
御～くださる	語彙素「下さる」	
～てくださる		
御～だ	語彙素「御」＋（1短単位）＋語彙素「だ」／「の」／「です」	
～たまう	動詞連用形＋語彙素「給う」	
御～なさる	語彙素「為さる」	
～なさる		
なさる		
御～になる	語彙素「御」＋（1短単位）＋出現書字形「に」＋語彙素「成る」	
御～	語彙素「御」＋動詞連用形	

この検索条件にマッチしたレコードから手作業により調査対象である形式を選別した。各形式の粗頻度と1万語あたりの相対頻度(括弧内)を示したものが表3・表4である。命令形⁵をとる場合とそれ以外では出現する形式に偏りがあるため、表3に命令形以外をとる場合、表4に命令形をとる場合を分けて示した。

³複数回にわたる連載作品はまとめて1作品とした。

⁴「～ていらっしゃる」「～てくださる」「～たまう」の「～」には動詞連用形のみが入り、「御～」には動詞連用形と一字漢語サ変動詞語幹「覧」「免」のみが入る。

⁵本稿でいう「命令形」とは、①該当形式の命令形(例「～てください」「御～なさい」)、②該当形式に助動詞「ます」の命令形が付くもの(例「～てくださいませ」「御～なさいませ」)、③「御～」を指す。

表 3 動詞の尊敬待遇表現形式の頻度(命令形以外)

	1895	1901	1909	1917	1925	通年
御～なさる	48 (11.5)	57 (10.7)	65 (6.7)	47 (6.1)	9 (2.8)	226 (7.5)
御～だ	59 (14.1)	69 (13.0)	41 (4.3)	35 (4.6)	18 (5.7)	222 (7.4)
おっしゃる	24 (5.8)	38 (7.2)	31 (3.2)	46 (6.0)	20 (6.3)	159 (5.3)
～てくださる	29 (7.0)	33 (6.2)	32 (3.3)	31 (4.0)	18 (5.7)	143 (4.8)
御～になる	22 (5.3)	14 (2.6)	25 (2.6)	41 (5.4)	35 (11.0)	137 (4.6)
～なさる	18 (4.3)	23 (4.3)	30 (3.1)	22 (2.9)	4 (1.3)	97 (3.2)
～ていらっしゃる	6 (1.4)	12 (2.3)	11 (1.1)	36 (4.7)	4 (1.3)	69 (2.3)
なさる	7 (1.7)	4 (0.8)	22 (2.3)	15 (2.0)	6 (1.9)	54 (1.8)
いらっしゃる	2 (0.5)	2 (0.4)	20 (2.1)	20 (2.6)	6 (1.9)	50 (1.7)
御～あそばす	10 (2.4)	4 (0.8)	12 (1.2)	7 (0.9)	0 (0.0)	33 (1.1)
御～くださる	2 (0.5)	4 (0.8)	15 (1.6)	6 (0.8)	5 (1.6)	32 (1.1)
～たまう	7 (1.7)	6 (1.1)	6 (0.6)	1 (0.1)	0 (0.0)	20 (0.7)
計	234 (56.1)	266 (50.1)	310 (32.2)	307 (40.1)	125 (39.5)	1242 (41.5)

表 4 動詞の尊敬待遇表現形式の頻度(命令形)

	1895	1901	1909	1917	1925	通年
～てくださる	33 (7.9)	21 (4.0)	82 (8.5)	53 (6.9)	20 (6.3)	209 (7.0)
御～なさる	30 (7.2)	27 (5.1)	62 (6.4)	33 (4.3)	24 (7.6)	176 (5.9)
～たまう	23 (5.5)	9 (1.7)	51 (5.3)	63 (8.2)	17 (5.4)	163 (5.4)
御～	31 (7.4)	37 (7.0)	32 (3.3)	21 (2.7)	18 (5.7)	139 (4.6)
～なさる	6 (1.4)	6 (1.1)	21 (2.2)	7 (0.9)	2 (0.6)	42 (1.4)
御～くださる	7 (1.7)	1 (0.2)	18 (1.9)	5 (0.7)	7 (2.2)	38 (1.3)
いらっしゃる	0 (0.0)	0 (0.0)	5 (0.5)	7 (0.9)	3 (0.9)	15 (0.5)
～ていらっしゃる	1 (0.2)	0 (0.0)	1 (0.1)	9 (1.2)	4 (1.3)	15 (0.5)
なさる	2 (0.5)	2 (0.4)	2 (0.2)	2 (0.3)	1 (0.3)	9 (0.3)
おっしゃる	3 (0.7)	0 (0.0)	3 (0.3)	1 (0.1)	2 (0.6)	9 (0.3)
御～あそばす	5 (1.2)	0 (0.0)	2 (0.2)	1 (0.1)	0 (0.0)	8 (0.3)
計	141 (33.8)	103 (19.4)	279 (28.9)	202 (26.4)	98 (30.9)	823 (27.5)

4. 形式選択における位相の重要度の分析

表3・表4にあげた形式のうち、命令形以外と命令形のそれぞれ頻度上位6形式について、形式選択における位相の重要度を分析する。

使用する分析手法はランダムフォレスト (random forest) である。ランダムフォレストはカテゴリー化されたデータを用いて分類のルールを作成する機械学習方法の一つで、テキストマイニングの分野でも用いられる (金 2009)。ランダムフォレストは分類ルール作成の過程で、副次的に分類に用いた説明変数の重要度を計算する。ここでは、形式を目的変数、位相を説明変数としてランダムフォレストを行い、算出される説明変数の重要度を用いて形式選択に対する位相の重要度を測ってみたい。

説明変数とする位相は次の①～④の4種類である。

①年

該当形式が出現する雑誌の刊行年である。コーパスのXML雑誌タグの年属性値に拠る。「1895・1901・1909・1917・1925」の5種。

②話者性別

該当形式が出現する会話の話者の性別である。コーパスには話者性別に関する情報は付いていないため、作品内容から手作業により判別した。「男・女・不明」の3種に分類し、「男・女」の2種に該当するレコードを分析対象とした。

③話者階層

該当形式が出現する会話の話者の社会的階層である。コーパスには話者階層に関

する情報は付いていないため、作品内容から手作業により判別した。「知識層・非知識層・不明」の3種に分類し、そのうち「知識層・非知識層」の2種に該当するレコードを分析対象とした。

④口語文体

該当形式が出現する口語会話の文体である。コーパスに付与した形態論情報に拠り、「ございます体・ですます体・その他」の3種に分類した。「ございます体」は該当会話中に「ござい(り)ます」が1例以上ある場合、「ですます体」は「ござい(り)ます」がなく「です」または「ます」が1例以上ある場合、「その他」は「ござい(り)ます」「です」「ます」の出現がない場合を示す。

口語文体は話し手の聞き手に対する丁寧さに関する待遇度、ひいては話し手と聞き手との関係性の一端を示すものであるとも言える。

また、本稿では聞き手を待遇対象とする形式のみを分析対象とした。聞き手以外の第三者を待遇対象とする形式を加えた場合、説明変数に「聞き手と待遇対象の関係」を加える必要が出てくるが、その分類が容易でないためである。また、東京語でない方言による会話に出現する形式も分析対象外とした。

ランダムフォレストの実行には統計分析ソフト R⁶の randomForest パッケージの randomForest()関数を用いた⁷。

4. 1. 命令形以外

命令形以外での頻度上位 6 形式「御～なさる」「御～だ」「おっしゃる」「～てくださる」「御～になる」「～なさる」についてランダムフォレストを行った。使用する 482 レコード中、382 レコードを学習用、100 レコードを評価用とした。

ランダムフォレストによって得られた分類ルールによる形式の分類実験の結果を表 5、説明変数(位相)の重要度を図 1 に示す。

表 5 分類実験の結果(命令形以外)

	実験結果					
	～てくださる	～なさる	おっしゃる	御～だ	御～なさる	御～になる
～てくださる	1	0	2	10	7	0
～なさる	0	5	0	2	3	0
おっしゃる	1	1	4	2	7	0
御～だ	0	3	2	9	5	0
御～なさる	2	3	3	0	16	0
御～になる	0	1	1	7	3	0

⁶ <http://www.r-project.org/>

⁷ スクリプト作成では金(2009)、石田・小林(2013)等を参照した。

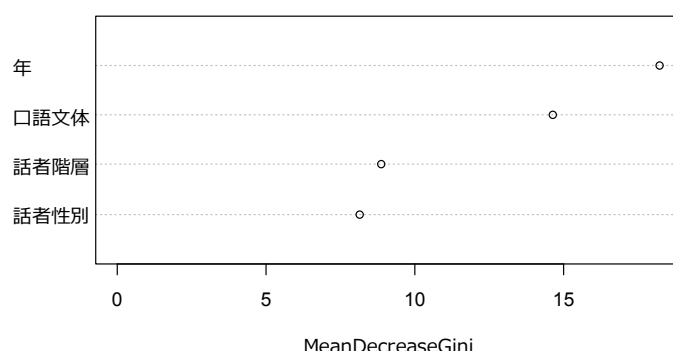


図1 説明変数の重要度(命令形以外)

表5の分類実験の正解率は35%と非常に低く、位相による形式の分類ルール作成は事実上失敗している。これは、6形式の選択では4種の位相の影響力は限定的であることを意味する。ただし、こうした分類ルール作成の過程においては、図1から4種の位相のうち年と口語文体が他の位相と比べて重要度が高いという結果が得られた。重要度上位2位の各位相でどの形式が選択されやすいのか見るため、年別の形式比率を図2に、口語文体別の形式比率を図3に示す。

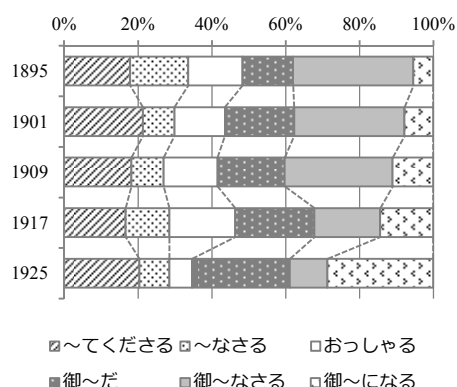


図2 年別の形式比率(命令形以外)

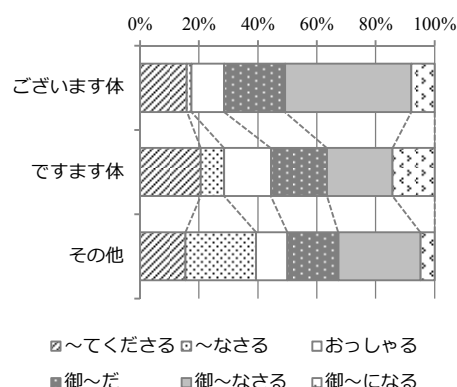


図3 口語文体別の形式比率(命令形以外)

例えば、図2から1895・1901・1909年では「御へなさる」が、1925年は「御へになる」「御へだ」が選択されやすく⁸、図3からは「ございます体」では「御へなさる」が選択されやすいことなどがわかる。

4. 2. 命令形

命令形での頻度上位6形式「～てください」「御へなさる」「～たもう」「御へ」「～なさる」「御へください」についてランダムフォレストを行った。分析に使用する575レコード中、475レコードを学習用、100レコードを評価用とした。ランダムフォレストによって得られた分類ルールによる形式の分類実験の結果を表6、説明変数(位相)の重要度を図4に示す。

⁸ 当時の動詞の尊敬待遇表現形式の主な経年変化として、「御へなさる」の頻度の減少と「御へになる」の頻度の増加が指摘されており(辻村1974、近藤2005等)、それと一致する結果と言える。

表 6 分類実験の結果(命令形)

	実験結果					
	～たまう	～てくださる	～なさる	御～	御～くださる	御～なさる
～たまう	26	0	0	0	0	0
～てくださる	7	16	0	1	0	4
～なさる	3	2	1	2	0	1
御～	4	0	0	7	0	0
御～くださる	0	6	0	0	0	0
御～なさる	5	7	0	2	1	5

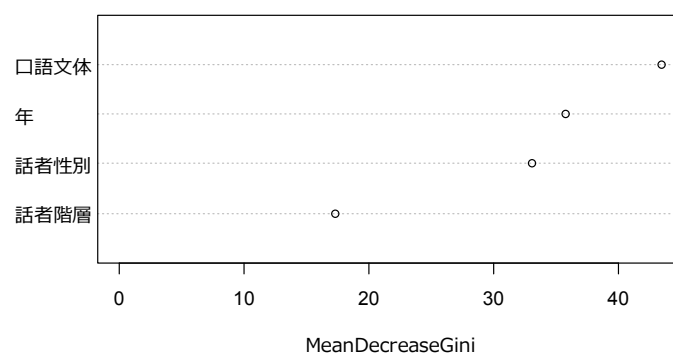


図 4 説明変数の重要度(命令形)

表 6 の分類実験の正解率は 55%で、命令形以外よりは高い値ではあるものの、分類ルールの作成に成功しているとは言い難い。命令形以外ほどではないにせよ、命令形においても、6 形式の選択に 4 種の位相はそれほど大きな影響力を持たないことがわかる。

また、図 4 から、4 種の位相のうち、重要度の高いものは第一に口語文体、次いで年と話者性別という結果が得られた。命令形以外の場合とは順位が異なることは注意される。重要度上位 3 位の各位相でどの形式が選択されやすいのか見るため、口語文体別の形式比率を図 5 に、年別の形式比率を図 6 に、話者性別別の形式比率を図 7 に示す。

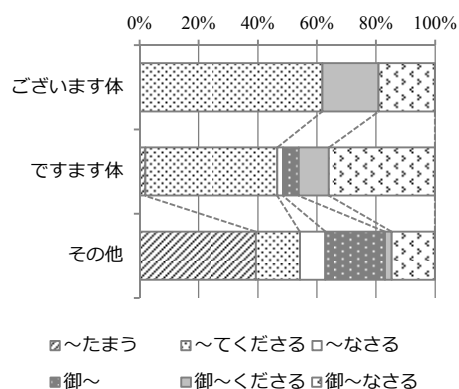


図 5 口語文体別の形式比率(命令形)

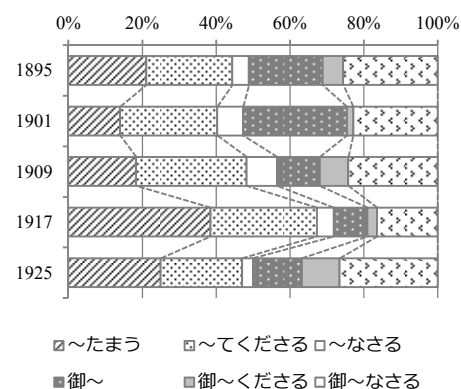


図 6 年別の形式比率(命令形)

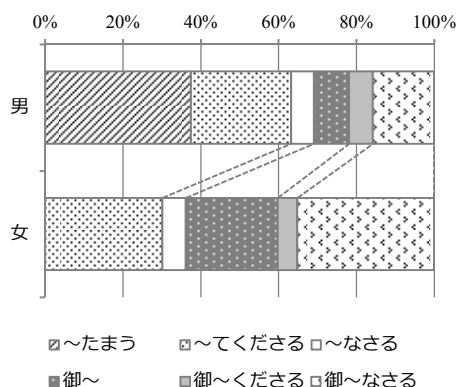


図7 話者性別別の形式比率(命令形)

例えば、図5から「ございます体」では「~てくださいる」が、「ですます体」では「~てくださいる」「御~なさる」が、「その他」では「~たまう」が選択されやすく、図6からは1917年では「~たまう」が選択されやすく、図7からは「男」では「~たまう」が、「女」では「御~なさる」「~てくださいる」が選択されやすいことなどがわかる。

5. おわりに

本稿ではランダムフォレストというコーパス言語学的な手法を用いて、動詞の尊敬待遇表現形式の選択における位相の重要度を分析することを試みた。その結果、特定の位相において選択されやすい傾向にある形式はあるものの、位相は形式選択に決定的な影響力を有することはないことが明らかになった。ただし、命令形においては命令形以外の場合よりも位相の影響度が高く、位相の重要度の順位も命令形とそれ以外では異なることも明らかになった。

取り上げた4種の位相が形式選択の決め手になり得ていないことは、これ以外に形式選択に重要な役割を果たす要素が存在する可能性があること、あるいは形式選択にはそもそも曖昧性が伴うことを示唆する。今後、より詳細な分析を行い、コーパス言語学的見地から新たな知見を得ることを目指したい。

付 記

本稿は、日本学術振興会科研費(23720242)および国立国語研究所共同研究プロジェクト「通時コーパスの設計」による研究成果である。

文 献

- 石田基広・小林雄一郎(2013)「第10章 方言データのテキストマイニング 「茸」のアクセントパターンを例に」『Rで学ぶ日本語テキストマイニング』ひつじ書房、pp.173-190
- 金明哲(2009)「第14章 テキストの分類」『テキストデータの統計科学入門』岩波書店、pp.177-202
- 国立国語研究所(編)(2005)『太陽コーパス—雑誌『太陽』日本語データベース』博文館新社
- 近藤明日子(2005)「尊敬待遇表現—動作性の名詞や動詞連用形に付く形式について—」『国立国語研究所報告122 雑誌『太陽』による確立期現代語の研究 『太陽コーパス』研究論文集』博文館新社、pp.227-250
- 田中牧郎(2005)「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」『雑誌『太陽』による確立期現代語の研究—『太陽コーパス』研究論文集—』博文館新社、pp.1-48

- 陳慧玲 (2005a) 「明治前期東京語の直接命令表現の考察—位相との関連を視点として—」
『文学研究論集』 22、pp.1-20
- 陳慧玲 (2005b) 「明治前期東京語における士族・知識層男性の命令表現の考察」『文化継承
学論集』 1、pp.27-43
- 陳慧玲 (2006a) 「明治期東京語における直接命令表現の諸相」『明治大学日本文学』 32、
pp.77(19)-64(32)
- 陳慧玲 (2006b) 「明治期東京語における下層男性の命令表現の考察」『文学研究論集』 25、
pp.21-40
- 辻村敏樹 (1974) 「明治大正時代の敬語概観」『敬語講座 5 明治大正時代の敬語』 明治書
院、pp.7-33
- 辻村敏樹 (編) (1991) 『敬語の用法』 角川書店
- 山西正子 (1972) 「「いらっしゃる」考」『国語学』 88、pp.56-78