

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 13 日現在

機関番号：62618

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23720242

研究課題名(和文)近代語コーパスの高度なタグ付けと形態素解析による待遇表現・人称代名詞の計量的研究

研究課題名(英文)Quantitative Study of Polite Expressions and Personal Pronouns with Advanced Tagging and Morphological Annotation of Modern Japanese Corpora

研究代表者

近藤 明日子(KONDO, ASUKO)

大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・プロジェクト非常勤研究員

研究者番号：30425722

交付決定額(研究期間全体)：(直接経費) 1,500,000円、(間接経費) 450,000円

研究成果の概要(和文)：『太陽コーパス』と『近代女性雑誌コーパス』を使って、近代日本語の待遇表現・人称代名詞の使用実態について明らかにするために、次のことを行った。まず、分析に必要な言語使用状況に関する情報と形態素解析によって得られる形態論情報をコーパスに新たに付与した。次に、形態論情報を利用してコーパスから待遇表現・人称代名詞の用例を網羅的に抽出した。そして、待遇表現・人称代名詞・言語使用状況の対応関係について計量的分析を行い、いくつかの対応関係について明らかにした。

研究成果の概要(英文)：This study explores the use of polite expressions and personal pronouns in modern Japanese through quantitative analyses of the Taiyo Corpus and Modern Women's Magazines Corpus. To obtain data for the analyses, the corpora were tagged with factors that determined registers and were annotated morphologically. Next, using the morphological annotations, polite expressions and personal pronouns were extracted exhaustively from the corpora. Finally, correspondences among the polite expressions, personal pronouns, and factors were quantitatively analyzed and identified.

研究分野：人文学

科研費の分科・細目：言語学・日本語学

キーワード：近代語 コーパス XMLタグ 形態素解析 待遇表現形式 人称代名詞 計量的分析

### 1. 研究開始当初の背景

近代日本語の研究に活用できるコーパスが普及しつつあり、代表的なものとして国立国語研究所(編)『太陽コーパス』(博文館新社、2005)と国立国語研究所『近代女性雑誌コーパス』(2006)があげられる。しかし、これらのコーパスには形態論情報(語に関する情報)が付与されておらず、多種多量の語を網羅的に取り上げる研究は事実上不可能であり、大規模コーパスの利点を十分に活用できない状況にあった。

それが、近年、近代の文語文用の形態素解析辞書「近代文語 UniDic」(<http://www2.nijjal.ac.jp/lrc/index.php?UniDic>)が開発・公開されるなど、近代語のコーパスに形態素解析(自動でテキストを単語に区切り、見出し語・品詞等の形態論情報を付与する技術)によって研究に堪える精度の形態論情報を付与できる環境が整ってきた。これは、『太陽コーパス』『近代女性雑誌コーパス』にも形態論情報を付与し、それを生かした計量的分析を行うことが可能になったことを意味する。

### 2. 研究の目的

本研究の目的は、近代語の代表的なコーパスである『太陽コーパス』『近代女性雑誌コーパス』を用いて、新たにコーパスにタグ付けする言語使用状況に関する情報と、近年実用化された近代語の形態素解析から得られる形態論情報を組み合わせ、近代語の待遇表現・人称代名詞の計量的分析を行い、言語使用状況と待遇表現・人称代名詞の間の対応関係を明らかにすることである。これにより近代語の待遇表現・人称代名詞の使用実態を一層精密に把握することが可能になる。

### 3. 研究の方法

#### (1) 形態素解析による形態論情報の付与

コーパスに形態論情報を付与するため、形態素解析を行った。文語体の記事は「近代文語 UniDic」、口語体の記事は旧仮名遣い口語文用の辞書を用いて形態素解析を行った。解析データは国立国語研究所の形態論情報データベースに格納した。

#### (2) 言語使用状況に関する情報の付与

分析に利用する言語使用状況に関する情報をコーパスのXMLタグから抽出、もしくは(1)で付与した形態論情報を利用して抽出した。抽出できないものは新たに人手で付与した。抽出・付与したデータは形態論情報とともにデータベースに格納し、データベースからXMLタグとして出力することも可能とした。主な情報の種類は次の～である。

雑誌の刊行年

「1895/1901/1909/1917/1925」の5種に分類。コーパスのXMLタグから抽出。

文体

「文語/口語」の2種に分類。コーパス

のXMLタグから抽出。

口語文体

「ございます体/ですます体/その他」の3種に分類。形態論情報を利用して抽出。

記事ジャンル

「非文学記事/文学記事」の2種、あるいは「小説・戯曲/それ以外」の2種に分類。コーパスのXMLタグから抽出。

文章種類

「地の文/会話/その他」の3種に分類。コーパスのXMLタグから抽出。

記事著者および会話話者の性別

「男/女/不明」の3種に分類。人手により付与。

会話話者の階層

「知識層/非知識層/不明」の3種に分類。人手により付与。

#### (3) 待遇表現と人称代名詞の抽出

(1)で付与した形態論情報を利用して、コーパスから待遇表現と人称代名詞を抽出した。

待遇表現は先行研究や辻村敏樹(編)『敬語の用法』(角川書店、1991)にあげられた動詞尊敬待遇表現の諸形式についてコーパスで予備調査を行い、出現頻度の高い13形式を対象に抽出を行った。さらに、命令形・非命令形の2種に分類した。

人称代名詞は一・二人称代名詞について抽出を行った。一人称代名詞はコーパス全体から、二人称代名詞はコーパスの小説・戯曲記事の口語会話部分から発話者の性別ごとに抽出した。

#### (4) 待遇表現・人称代名詞・言語使用状況の対応関係の計量的分析

(2)(3)で得られたデータを使用し、待遇表現・人称代名詞・言語使用状況の対応関係について統計量に基づいた計量的分析を行った。

### 4. 研究成果

#### (1) コーパスの言語量の解明

形態論情報の付与により、計量的分析の基礎データとなるコーパスの言語量が明らかになった。例として、各コーパスの延べ語数(助詞・助動詞を含む)を表1に示す。

表1 コーパスの延べ語数

	延べ語数
太陽	876万
近代女性雑誌	127万

また、言語使用状況に関する情報と組み合わせることにより、地の文と会話文それぞれの言語量や話者性別ごとの会話文の言語量なども明らかになった。例として、小説・戯曲記事の口語会話の話者性別別延べ語数と比率を図1に示す。

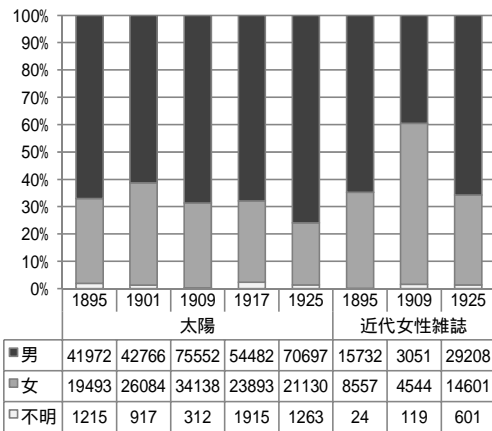


図1 口語会話の話者性別別延べ語数と比率

(2) 一・二人称代名詞の語形数と出現頻度  
形態論情報を利用して網羅的に一・二人称代名詞を抽出したことにより、一・二人称代名詞の語形とその出現頻度が明らかになった。一人称代名詞について、コーパス全体での語形数・出現頻度合計を表2に示す。

表2 一人称代名詞の語形数・出現頻度合計  
(コーパス全体)

	語形数	出現頻度合計
太陽	32	24335
近代女性雑誌	27	4896

二人称代名詞について、小説・戯曲記事の男性の口語会話での語形数・出現頻度合計を表3、小説・戯曲記事の女性の口語会話での語形数・出現頻度合計を表4に示す。

表3 二人称代名詞の語形数・出現頻度合計  
(小説・戯曲記事の男性の口語会話)

	語形数	出現頻度合計
太陽	26	2400
近代女性雑誌	20	457

表4 二人称代名詞の語形数・出現頻度合計  
(小説・戯曲記事の女性の口語会話)

	語形数	出現頻度合計
太陽	19	1158
近代女性雑誌	7	271

(3) 待遇表現・人称代名詞・言語使用状況の対応関係

待遇表現・人称代名詞・言語使用状況の対応関係について、統計量に基づく計量的分析を行ったところ、いくつかの事実が明らかになった。これにより近代語の待遇表現・人称代名詞の使用実態を一層精密に把握することができた。例として次の ~ をあげる。

一人称代名詞と文章種類との対応関係  
『太陽コーパス』『近代女性雑誌コーパス』で、一人称代名詞と文章種類(文語地の文/口語地の文/口語会話)との対応関係をコレスポネンス分析により分析したところ、両コーパスとも強い対応関係が確認され、各文章種類に対応する語形が明らかになった(表5)。

表5 一人称代名詞の語形と文章種類との対応関係

文章種類	一人称代名詞	
	太陽	近代女性雑誌
文語地の文	ごじん(吾人) よ(余) よはい(余輩)	ごじん(吾人) しょう(妾) しょうまい(小妹) それがし まい(妹) よ(余) よはい(余輩) わがはい
口語地の文	わがはい われわれ	わたくし わたし わらわ われわれ
口語会話	あたし おいら おら おれ せっしゃ(拙者) ぼく わし わたくし わたし	あたし あたし おいら おら おれ せっしゃ(拙者) てまえ ぼく わい わし わて

注: 『太陽コーパス』は出現記事数の多い14語形に限定して分析した。

二人称代名詞と動詞尊敬待遇表現形式との対応関係

『太陽コーパス』で二人称代名詞と動詞尊敬待遇表現形式との対応関係をコレスポネンス分析により分析したところ、強い対応関係が確認され、3グループに分けられることが明らかになった(表6)。

表6 二人称代名詞の語形と動詞尊敬待遇表現形式との対応関係

二人称代名詞	動詞尊敬待遇表現形式	
	非命令形	命令形
あなた	御～なさる	
	～てください	～てください
	おっしゃる	御～なさる
	御～になる	御～くださる
	なさる	おっしゃる
	～ていらっしゃる	～ていらっしゃる
	御～あそばす	御～あそばす
きみ	～たまう	～たまう
	おまえ おまえさん おめえ きさま	御～ ～なさる

動詞尊敬待遇表現形式の選択における言語使用状況の影響度  
『太陽コーパス』の小説・戯曲記事の口語会話で、動詞尊敬待遇表現形式の選択における言語使用状況（年・話者性別・話者階層・口語文体）の影響度をランダムフォレストにより分析した。その結果、特定の言語使用状況において選択されやすい傾向にある形式はあるものの、とりあげた4種の言語使用状況は形式選択に決定的な影響力を有するものではないことが明らかになった。このことは、形式選択に重要な役割を果たす言語使用状況が4種以外に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。

#### 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計3件）

近藤明日子、コーパス活用の勘所 第2回【日本語史】近代語の文法(1) 単語情報の付いた『太陽コーパス』に見る一人称代名詞、日本語学、査読無、第33巻第6号、2014、62-67

近藤明日子、近代総合雑誌記事に出現する一人称代名詞の分析 単語情報付き『太陽コーパス』を用いて、近代語研究、査読無、第17集、2013、133-154

近藤明日子、『太陽コーパス』に見る一人称代名詞「吾人（ごじん）」「余（よ）」との比較から、近代語研究、査読無、第16集、2012、63-80

〔学会発表〕（計6件）

近藤明日子、『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、第5回コーパス日本語学ワークショップ、2014年3月7日、国立国語研究所

近藤明日子、単語情報付き『太陽コーパス』に見る動詞の尊敬待遇表現、第105回国語学史研究会、2013年12月7日、天理大学杉之内キャンパス

近藤明日子、『近代女性雑誌コーパス』の小説会話部分に現れる一・二人称代名詞の計量的分析、第4回コーパス日本語学ワークショップ、2013年9月5日、国立国語研究所

近藤明日子、明治後期・大正期の小説会話部分に現れる一人称・二人称代名詞の計量的研究 『太陽コーパス』を用いて、近代語学会、2013年6月15日、白百合女子大学

近藤明日子、近代女性雑誌記事における一人称代名詞の分析 形態論情報付き『近代女性雑誌コーパス』を用いて、第3回コーパス日本語学ワークショップ、2013年3月1日、国立国語研究所

近藤明日子、単語情報付き『太陽コーパス』を用いた一人称代名詞の分析、日本語学会2012年度秋季大会、2012年11月4日、富士大学五福キャンパス

#### 6. 研究組織

##### (1) 研究代表者

近藤 明日子 (KONDO Asuko)

大学共同利用機関法人人間文化研究機構

国立国語研究所・コーパス開発センター・

プロジェクト非常勤研究員

研究者番号：30425722