

令和元年8月29日現在

機関番号：33915

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16850

研究課題名(和文) 歴史コーパスに基づく中世・近世語の複合辞および連語の研究

研究課題名(英文) A Study of Compound Functional Expressions and Collocations in Medieval and Early Modern Japanese Based on the Corpus of Historical Japanese

研究代表者

市村 由貴(渡辺由貴)(Ichimura-Watanabe, Yuki)

名古屋女子大学・文学部・講師

研究者番号：10569776

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究では、中・近世日本語における複合辞・連語の使用状況を明らかにすることを目的とした。具体的には、『日本語歴史コーパス』のデータを用いて中世・近世日本語資料の単語N-gramデータを抽出した。このN-gramデータを出現頻度・出環境・文法的機能・意味の希薄化等の観点から整理し、複合辞・連語リスト作成のための分析を行った。さらに、資料ごとの複合辞・連語の出現状況から、各資料の特徴や通時の変化を検討した。また、複合形式の計量的研究手法の確立を目指し、『現代日本語書き言葉均衡コーパス(BCCWJ)』のデータを用いた複合辞の研究を行った。

研究成果の学術的意義や社会的意義

中・近世語における複合辞・連語形式については、個別形式に関する研究はなされているものの、その全体像が明らかにされているとはいえない。本研究において、統一的に形態論情報が付された『日本語歴史コーパス』のデータを用いて中・近世語における複合辞・連語の使用状況を調査した。これを計量的に検討したことにより、客観的指標に基づいた複合辞・連語研究を進めることができた。本研究で抽出・整理したN-gramデータは、日本語史分野における複合辞・連語研究の基礎的データとなりうる。また、現代語で用いられている複合辞の成り立ちや、複合辞化・連語化の傾向を検討する上でも重要な資料となるものである。

研究成果の概要(英文)：The purpose of this study is to clarify the use of compound functional expressions and collocations in medieval and early modern Japanese. N-gram data for Japanese materials from the medieval and early modern periods were extracted from “The Corpus of Historical Japanese (CHJ).” These N-gram data were organized by occurrence frequency, appearance environment, grammatical function, and dilution of meaning, and analysis was conducted to create a list of compound functional expressions and collocations. Moreover, based on the appearance of compound functional expressions and collocations, I examined the characteristics and changes over time of each material. Additionally, in order to establish a quantitative research method for compound words, I studied the compound functional expressions and collocations using data from the Balanced Corpus of Contemporary Written Japanese (BCCWJ).

研究分野：日本語学

キーワード：日本語史 複合辞 連語 コーパス 中世語 近世語 N-gram

1. 研究開始当初の背景

複合辞は、「かもしれない」「について」等、「いくつかの語が一まとまりになって、その一まとまりが固有の『付属語』(辞)的な意味を担うものとして用いられる形式(山崎他(2001, p.1))とされるものである。その認定基準として、松木(1990, pp.34-35)は「形式的にも意味的にも辞的な機能を果たしていること」「形式全体として、個々の構成要素の合計以上の独自の意味が生じていること」をあげ、さらに形式名詞・用言を中心とした複合形式については、「中心となる『詞』は実質の意味が薄れ、形式的・関係構成的に機能していること」という基準を追加している。

現代語の複合辞については多くの研究があり、また、日本語教育の分野でも重要視され、日本語学習用の文型辞典にも多くの複合辞が掲載されている。しかし、中・近世語の複合辞に関する研究は、個別の形式については進められているものの、数量的根拠に基づいて全体像を見渡すような研究はほぼなされていなかった。田中(1977)およびそれを整理した松木(2005)に、「マセナダ」「ガデキル」等、江戸～明治期にみられる複合辞を見渡した研究があり、様々な複合辞の盛衰がうかがえるものの、ここでも計量的側面からの考察はなされていない。すなわち、各研究者が複合辞と思われる形式にあたりをつけ、その形式について考察していく形の研究が主なものであった。

複合辞の認定基準は研究者により異なるものの、一時的な語の結合ではなく、ある程度固定化した形式を複合辞と呼ぶ点では共通しているといっておく、内省のききにくい中・近世語では、複合辞の認定にあたっては計量的側面からの考察が重要である。にもかかわらず、そのような研究手法がとられてこなかったのは、計量的に複合辞研究を行うために必須となる、形態論情報付きの大規模な通時的言語データがほぼ存在していなかったためであると考えられる。

連語は「風邪を引く」「恥をかく」等、「意味的な要請もなく、はっきりした理由もないのに結び付きが固定しているものを指す」(『日本語学研究事典』「連語」の項)が、このような表現に関しても、複合辞と同様の理由で、全体像を見渡すような研究が難しかったと考えられる。

そこで本研究では、連語・複合辞の選定は、連続する一定の単語数を切り出した「単語 N-gram」を利用して行うこととした(図1参照)。

文字 N-gram を利用した日本語史研究には近藤(2000等)がある。これは『古今和歌集』について3～7文字の連続した文字列を抽出し、計量的に作者の性別ごとの特徴を見出したものであるが、この文字 N-gram と比べ、単語 N-gram は無意味な文字列の連続が抽出されることがなく、取り出された全てのデータに意味がある点で優れているといえる。なお、単語 N-gram を利用した研究としては、単語 N-gram を中古日記文学テキストの類似性の指標の一つとした太刀岡(2014)がある。

近年、国立国語研究所において、『日本語歴史コーパス』が開発されている。形態論情報が付された大規模な中・近世語資料のデータは、ようやく利用できる段階に至ったところであり、そのデータについて計量的に応用するという点は、これまでにない試みといえる。『日本語歴史コーパス』のデータは統一的に単語情報が付されたものであり、これに計量的な手法を加味することにより、客観的指標に基づいた複合辞・連語研究の新たな展開が予想された。

本文	で	ござる	に	よつ	て
語	なり	御座る	に	因る	て
2 gram	でござる (なり+御座る)				
		ござるに (御座る+に)			
			による (に+因る)		
3 gram	でござるに (なり+御座る+に)				
		ござるによる (御座る+に+因る)			
			によつて (に+因る+て)		

【図1】2-gram・3-gram イメージ
(『虎明本狂言集』短単位データによる)

2. 研究の目的

本研究の目的は以下の4点である。

- (1)複合辞・連語形式抽出に適した言語データの選定・作成
- (2)単語 N-gram データの抽出
- (3)単語 N-gram データの整理・分析
- (4)得られた複合形式の研究と研究成果の公表

(1)複合辞・連語形式抽出に適した言語データの選定・作成

複合辞は「近現代日本語の分析的傾向を示す表現」(『日本語文法事典』p.529)であるとされるため、その萌芽がみられる中・近世語資料の中から本研究を行う上で有用な資料を選定する。

『日本語歴史コーパス』に収められていない資料については、『日本語歴史コーパス』に準じた形式でデータを構築する。具体的には、『日本語歴史コーパス』に収録されている『虎明本狂言集』との共通点・相違点を探るために必要な『虎寛本狂言集』を選定し検討を進める。

(2) 単語 N-gram データの抽出

『日本語歴史コーパス』のデータ及び新規に選定・作成したデータについて、各資料につき 2～7 語の連続を一律で取り出し、単語 N-gram データを抽出する。

(3) 単語 N-gram データの整理・分析

取り出した語の連続のうち、何を複合辞・連語とするかの判断は難しい。例えば「罷出-たる-者-は」のように、高頻度の表現ではあるが、同時代の資料において広く使われている複合辞・連語というよりは、狂言の名乗りの場面での定型表現とみるべきものもある。そこで、高頻度の表現について、松木(1990)をふまえ、文法的機能・意味の希薄化・個々の構成要素の合計以上の意味の有無・出現環境等の観点を用いて複合辞を認定するとともに、複数の資料に多くみられる連語や、特定の資料に頻出する定型表現の使用状況についてもあわせて分析する。

(4) 得られた複合形式の研究と研究成果の公表

(3)により得られた複合形式について、文法的機能や文体的特徴等を検討する。この成果について、関連学会において研究発表を行い、論文を執筆することにより公表する。また、(1)で新たに作成したデータについては、将来的に『日本語歴史コーパス』にあわせた形式で公開することを目指す。

このように、語の連続を一律に抽出することにより、複合辞・連語、あるいは出現頻度の高い語の連続形式を数値に基づいて客観的に提示し分析することが可能となる。さらに、公開された成果は、日本語史分野における複合辞・連語研究の基礎的データとなりうる。また、現代語で用いられている複合辞の成り立ちや、複合辞化・連語化の傾向を検討する上での資料ともなる。

3. 研究の方法

(1) テキストの選定・作成

複合辞・連語の分析に適した資料として、『虎明本狂言集』やキリシタン資料等を選定した。また、『日本語歴史コーパス』未収録の『虎寛本狂言集』のデータについては、『日本語歴史コーパス』に準じた形式でのテキスト入力および形態素解析データの整備を進めた。

(2) 複合辞・連語一覧表の作成

各資料につき 2～7 語の連続を、『日本語歴史コーパス』のデータベースから一律で取り出した。ただし、記号類は除外し、文末をまたがない範囲とした。この結果を使用頻度・共起強度等の観点から概観し、複合辞や連語と認めうるもの・それ以外の性格をもつ語の連続とに整理した。同時に、複合辞・連語の認定基準についても再検討した。

頻度	短単位	出現形
2087	と-言う	と云(1255)/といふ(539)/といひ(96)/といへ(81)/...
2057	程-に	程に(1457)/ほどに(600)/
1410	なり-御座る	でござる(578)/で御ざる(563)/で御さある(34)/...
1100	言う-て	云て(761)/いふて(245)/いひて(92)/云ふて(2)/
872	為る-て	して(872)/
825	う-て	ふと(449)/うと(376)/
816	て-御座る	てござる(372)/て御ざる(254)/てござれ(34)/...
627	と-申す	と申(542)/と申さ(54)/と申せ(29)/と申し(1)/と申す(1)/
599	事-なり	事で(443)/事に(57)/事なれ(29)/事なり(20)/...
578	なり-て	にて(578)/
...

【図2】2-gram のリスト
(『虎明本狂言集』短単位データによる)

(3) 複合辞・連語一覧表を利用した研究

上記(2)で作成・整理した一覧表をもとに、取り出した複合辞や連語個々の文法的機能や文体的特徴等を分析した。また、複数の資料の複合辞・連語一覧表を作成・比較し、資料別・時代別の特徴を検討した。

4. 研究成果

(1) 中・近世語における複合辞・連語の分析方法の検討および結果の整理

中・近世語のコーパスデータより取得した N-gram データを用いた複合辞・連語研究の検討方法およびその分析結果を学会にて発表した。

『虎明本狂言集』の N-gram のリストには、「に-因る-て・程-に」のような複合助詞、「じゃ-有る・て-有る・て-おじゃる・て-おりやる」のような複合助動詞、「重ねる-て(重ねて)・構う-て(かまいて)」のような副詞、「然-有る-ば(さあらば)・然る-ば(さらば)・然る-ながら(さりながら)」のような接続詞に相当する表現が多くみられた。ここから、辞の中の一要素となりやすい語や品詞の傾向等について分析を行った。

また、本研究以前には、日本語の通時コーパスから抽出した N-gram データを用いて複合辞・連語を検討する研究手法は確立されていなかったが、数量的な根拠をもとに分析を行うことでその手法の一例を提示することができた。同時に、分析結果を蓄積された複合辞・連語研究の成果の中に位置づけ、関連性をもたせていくことで、さらなる研究の展開を示すことができた。

今後、学会にて発表した内容をとりまとめ論文とする予定である。

(2) 『虎寛本狂言集』のデータの作成と検討

『虎明本狂言集』のデータより取得した N-gram データとの比較のため、『虎寛本狂言集』のコーパスデータを作成した。協力者を得て試行的にこのデータから N-gram データを抽出し、『虎明本狂言集』との比較を行った結果を研究会にて発表した。『虎寛本狂言集』のデータについては、追加整備を行った上で国立国語研究所のデータベースへ収録する方法等について検討を進めている。

(3) 複合辞・連語の検討と国語教育への応用的研究

コーパスデータを用いて複合辞・連語を分析する方法の確立と、その成果の教育の場への還元を目指し、現代語教科書における複合辞の使用状況を調査し、論文を執筆した。また、古文教材に使われる古典文学作品のコーパスデータより N-gram データを取得し、古典の読解において重要な連語あるいは定型的な語の連続、それぞれの古典文学作品を特徴づける表現の選定を試み、その内容について、分担執筆した書籍にて公表した。

【参考文献】

- 近藤みゆき (2000) 「n グラム統計処理を用いた文字列分析による日本古典文学の研究 『古今和歌集』の「ことば」の型と性差」 『千葉大学 人文研究』 29
- 太刀岡勇気 (2014) 「中古日記文学の計量国語学的分析と異本間の関係性の客観分析 『和泉式部日記』と『更級日記』を題材に」 『計量国語学』 29-6
- 田中章夫 (1977) 『近代語における複合辞的表現の発達』 松村明教授還暦記念 国語学と国語史』 明治書院
- 日本語文法学会編 (2014) 『日本語文法事典』 大修館書店
- 飛田良文他編 (2007) 『日本語学研究事典』 明治書院
- 松木正恵 (1990) 「複合辞の認定基準・尺度設定の試み」 『早稲田大学日本語教育センター紀要』 2
- 松木正恵 (2005) 「分析的傾向と複合辞 複合辞研究史 田中章夫の通時的研究」 『論理的な日本語表現を支える複合辞形式に関する記述的総合研究』 平成 14~16 年度日本学術振興会科学研究費補助金 基盤研究(B)(1)研究成果報告書
- 山崎誠・藤田保幸 (2001) 『現代語複合辞用例集』 国立国語研究所

5. 主な発表論文等

〔雑誌論文〕(計 2 件)

- 渡辺由貴 「BCCWJ 教科書データにおける複合辞の教科別使用状況 国語教育を視野に」 国立国語研究所論集 15 pp.195-210 2018年7月 [査読有り]
- 渡辺由貴 「BCCWJ 国語教科書データにおける複合辞の学年別使用状況 国語教育での指導の可能性」 早稲田日本語研究 26 pp.1-12 2017年3月 [査読有り]

〔学会発表〕(計 7 件)

- 北崎勇帆, 渡辺由貴, 村山実和子 「『虎寛本』狂言台本のコーパス化試論」 「通時コーパスの構築と日本語史研究の新展開」 研究発表会 2018年6月9日
- 渡辺由貴 「N-gram を利用した中世語の複合辞・連語の検討」 日本語学会 2018 年度春季大会 ワークショップ 「日本語史研究とコーパス活用 その利点と注意点」 (岡崎友子・渡辺由貴・宮内佐夜香・橋本行洋) 2018年5月20日
- 片山久留美, 渡辺由貴, 小木曾智信 「『日本語歴史コーパス 室町時代編 キリシタン資料』の公開」 日本語学会 2018 年度春季大会 2018年5月20日
- Atsuko Kawaguchi, Yuki Watanabe, Miwako Murayama “Construction and Utilisation of the Corpus of Christian Materials (Kirishitan Shiryō)” EAJIS 2017 15th International Conference of the European Association for Japanese Studies 2017年8月31日
- 鴻野知暁, 渡辺由貴, 片山久留美, 小木曾智信 「『日本語歴史コーパス 鎌倉時代編 日記・紀行』の公開」 日本語学会 2017 年度春季大会 2017年5月14日
- 渡辺由貴 「『日本語歴史コーパス 室町時代編 キリシタン資料』の現状と今後」 「通時コーパス」シンポジウム 2017 2017年3月11日
- 渡辺由貴 「BCCWJ の教科書データにおける複合辞 国語教育を視野に」 人文科学とコンピュータシンポジウム 「じんもんこん 2016」 2016年12月10日

〔図書〕(計 1 件)

- 渡辺由貴 (2019) 「コーパスデータから見た古文教材における連語」 河内昭浩編 『新しい古典・言語文化の授業 コーパスを活用した実践と研究』 朝倉書店 pp.145-152

〔産業財産権〕

出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2) 研究協力者

研究協力者氏名：

ローマ字氏名：