様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ９ 月 １２ 日現在

機関番号：３４３１１
研究種目：基盤研究(C)（一般）
研究期間：2014～2017
課題番号：２６３７０５１６

研究課題名（和文）大規模通時コーパスを用いた発見的研究方法の開拓

研究課題名（英文）Exploration of heuristic research methods based on large-scale diachronic corpora

研究代表者
服部　匡（Hattori, Tadasu）

同志社女子大学・表象文化学部・教授

研究者番号：４０２２８４９０

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：１　通時的研究の新手法の開拓　動詞の格表示の通時変化研究の新しい手法を提案した。多数の漢語サ変動詞について、主に国会会議録を用いて格表示の変化を探索し、ニ格からヲ格に変化したと思われるもの、その反対のものの存在を発見した。Ngramの頻度順位変化や文法的パターンの頻度比の分析に基づき、「可能性」や「～てございます」の用法変化を記述した。
２　新しい通時コーパスの利用環境の構築と利用可能性の検討　画像として公開されている「帝国議会会議録」「日本外交文書」の文字認識処理と検証を行った。特に前者は認識精度が低く厳密な統計分析には使えないが、特定表現の出現の有無や時期を検証するには使えることが分かった。

研究成果の概要（英文）：1 Explorations of new methods in corpus-based diachronic research: I proposed a new method of analyzing diachronic changes in verbal case markings. Based on the occurrence patterns of Sino-Japanese sahen verbs in the Japanese Diet minutes, many verbs show case marking shifts from dative to accusative or vice versa. I gave a description of the history of usages of the noun kanousei and the auxiliary -te gozaimasu on the basis of observed changes in the Ngram frequency ranking and the distribution of grammatical patterns.
2 Building an environment for using new diachronic corpora, and an examination of their utilizability: I extracted texts from the Imperial Diet minutes and the Documents on Japanese Foreign Policy and examined their utilizability for diachronic linguistic research. I found that, while insufficient for strict statistical analysis, they can be used for finding the period of occurrence/nonoccurrence of a specific expression.

研究分野：言語学・日本語学

キーワード：コーパス　経年変化

１．研究開始当初の背景

日本語研究では、近年、国会会議録などの新たな大規模コーパスが利用可能になったが、それらの、通時変化研究への利用の可能性やそのための手法については、なお未検討な面が多く、多様な方面からの探索試行の必要があった。

２．研究の目的

近年利用可能になった大規模な通時的コーパスを用いて通時変化を研究する新たな手法を考案し主に現代日本語のデータに適用する。具体的には、次の３つを行う。

1）「気づかれていない言語変化」をコーパスから発見する方法を探求する。

2）話者の諸属性やテキストのジャンル属性を利用した言語変化･変異の分析方法を探求する。

3）上記の目標のため必要なデータやプログラムの整備を図る。

３．研究の方法

1）コーパスとして利用可能な新たなデータの可能性の調査検討を行い、可能なものについてはその利用環境の整備を図る。

2）大規模コーパスでの用例の頻度統計を基盤として用法変化を発見探索するための分析手法を検討試行する。

3）大量の用例に基づき、語や表現の新用法の出現時期を発見する。

4）話者の諸属性やテキストのジャンル属性と表現の出現頻度との関係を分析する手法を検討試行する。

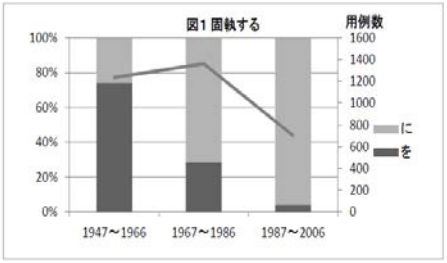
４．研究成果

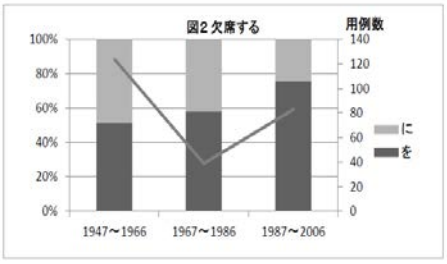
1）「気づきにくい言語変化」の発見のための分析の試行

(A)　『神戸大学新聞記事文庫』『国会会議録』などのコーパスを用い、漢語サ変動詞における格支配の交替(ニ→ヲ、ヲ→ニ)の例を抽出し、変化傾向の分析を行った。従来説とは反対に、ヲ格からニ格へと交替したように見える動詞の存在を指摘した。同一動詞におけるニ→ヲ、ヲ→ニの格表示変化について、変化の顕著な動詞を発見した。

新聞記事データベースを用い、ほぼ同様な意味を表すのにニ格とヲ格の両方の用例のある動詞の洗い出しを行い、どちらかの格の使用が増加する傾向にある動詞を発見し、先に行った国会会議録での事実と対照分析した。さらに、意味的な面を考慮して用例の傾向を分析すると、変化には何種類かの型があることが分かった。

国会会議録において格表示変化の顕著な動詞を２例挙げる。





(B)　「可能性」の意味・統語パターン変化の研究

「可能性」という語については従来から研究しているが、対象を大正・昭和前期にまで広げ、神戸大学新聞記事文庫、および、明治以前の生まれの著者の全集類を利用した調査に基づいて、この語のとる統語パターンや意味用法の上に生じた変化を 3 点発見した。

第１点は、実現性から確実性への意味変化に関わる。「可能性」の用法に関して、未実現の事柄について言う用例がおよそ戦後に一般化したものであることを発見した。

戦前期の新聞では「～シタ可能性」は１例しかなく、国会会議録での初出は 1955 年である。また、図 3 ように、現代の新聞でも、「～シタ可能性」の比率は漸増している(現代 1：1987 年 1 月～9 月、現代 2：2012 年 1 月～9 月)。

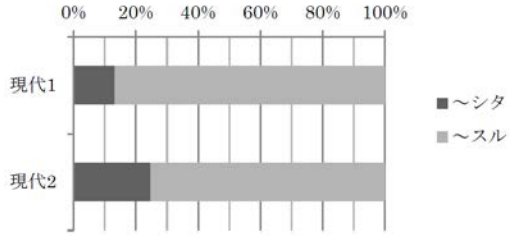


図３　「{～シタ/～スル}可能性」の比率

第２点は､程度を表す形容詞の変化である。図 4 のように、「多い」から「高い」などへの変化を指摘し、その原因が第１点と深く関わることを推測した。

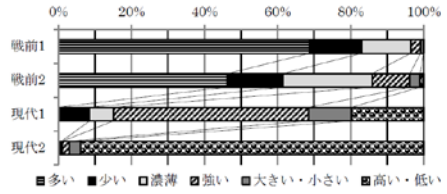


図４　「可能性」の度合を表す述語の種類と比率

第 3 点は、帰属性に関わる変化である。「可能性」を意味的な項とする述語の比率推移は図 5 のようになる(戦前１：1914～1923 年、戦前２：1934～1943 年)。所有帰属を表す述語の現象が顕著である。

以上を総合して、「可能性」の主たる意味は、「主体や状況に伏在する、ある種の事態を発現すべき因子」から「ある種の事態の成立を信じうる性質あるいは度合」へと推移したという仮説を示した。

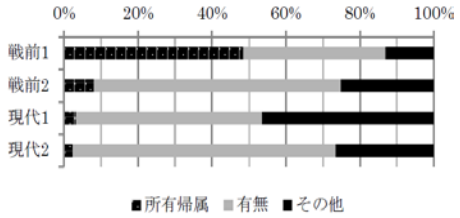


図 5 述語の種別

また「可能」「不可能」についても、新たに利用可能になった帝国議会議事速記録を利用して、明治・大正期の用例傾向の分析を行い、明白な変化傾向を発見した。

(C) 「～テございます」の使用傾向推移

帝国議会・国会の会議録を用いて、「～テございます」の使用傾向を、特に「～テいる」「～テおる」のどちらに対応するかという観点から分析した。「なってございます」のような言い方は江戸時代から連続して存在したものであるが、図 6 のように、戦後の国会では「てございます」を使用する比率が上昇している。

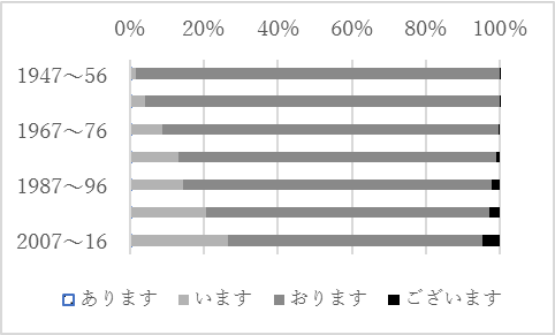


図 6 「なる」の場合の補助動詞別頻度

ほぼ「～テある」の例のない動詞のうち、「～テございます」との共起頻度比の上昇の著しいものを挙げると、図 7 のようになる。

次項に示す事実からも、官僚の国会議員への説明場面を典型例とする「低姿勢」で形式張った発話のレジスターで盛んになったようであるが、今のところ、限定されたレジスターでの現象にとどまり、広く波及する兆しはないように思われる。

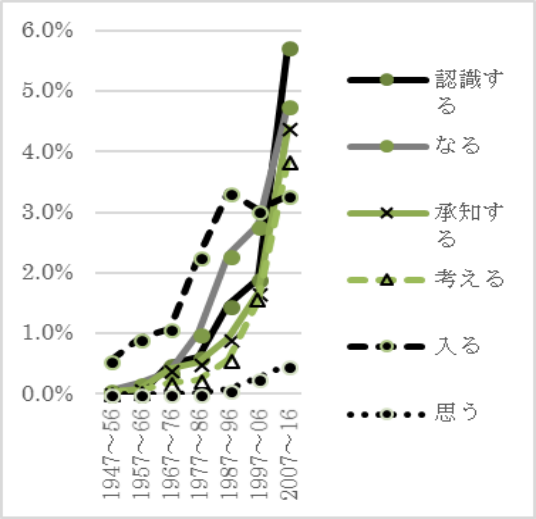


図 6 動詞別共起頻度比の推移(～テございます)

2) 言語変異分析の試行

国会会議録での発言者の出身地情報を用いた言語変異分析の試行を行った。移動動詞に対するニとへの選択に関する出身県別の相違を分析したところ、『日本語言語地図』記載の傾向とある程度の相関が見られた。

「～テございます」の使用傾向について、国会会議録に記された発言者の肩書きに基づいた分析を行ったところ、政府委員・説明員など、国会議員に説明を行う立場の者がよく使用することが分かった。

3) データやプログラムの整備

画像のみがウェブ上で公開されている、日本外交文書、帝国議会会議録の全データを取得し、独自に OCR 処理を行いテキスト化した。誤認識によるデータ誤りに対してある程度の修正処理をほどこした上、研究への利用可能性を調査検討した。その結果、特に帝国議会会議録は誤認識が多く厳密な統計分析への利用は困難であるものの、特定の表現の出現例を取得するなどの目的への利用は可能であることが分かった。これにより、方法的な制約はあるものの、研究対象となる年代を明治中期まで拡大することができた。

また、検索プログラムを開発し、コピュラ表現などに関して試験的な利用を行って、本格的な研究利用可能性の見極めを行った。

５．主な発表論文等

〔学会発表〕(計　2 件)

① 服部匡、漢語動詞における格表示変化傾向の探索 ―ヲ格とニ格― 第 8 回コーパス日本語学ワークショップ、査読有、国立国語研究所、 2015（予稿集はウェブで公開）

② 服部匡、「可能性」の意味用法の変化 ―大正から平成まで―、日本語学会春季大会、査読有、学習院大学、2016

〔図書〕（計　１件）
① 服部匡、「～テございます」の使用傾向の推移 ―「～テある」「ている」との対応関係に注目して―、和泉書院、藤田保幸・山崎誠（編）　『複合辞研究の現在』、2018、pp. 357-376

〔その他〕
ホームページ等
http://thattori.com

６．研究組織
(1)研究代表者
　服部　匡（HATTORI, Tadasu）
同志社女子大学・表象文化学部・教授
研究者番号　40228490