

研究種目：特定領域研究
研究期間：2006～2010
課題番号：18061006
研究課題名（和文）コーパスを利用した国語辞典編集法の研究

研究課題名（英文）A Study of Compilation Methods of Japanese Language Dictionaries using Large-scale Corpora

研究代表者
荻野綱男 (OGINO TSUNAO)
日本大学・文理学部・教授
研究者番号：00111443

研究分野：日本語学
科研費の分科・細目：日本語学
キーワード：コーパス 国語辞典 現代日本語 BCCWJ

1. 研究計画の概要

本研究では、代表者・分担者それぞれで4つの課題を追求するとともに、全体でコーパスと辞書の関連を研究し、辞書記述に応用できるコーパスのあり方を考えるものである。

- (1) コロケーション辞書の作成
- (2) 複合辞辞書の作成
- (3) コーパスを利用した現行国語辞典の評価法の開発
- (4) 動詞の格情報と、オノマトペの意味・用法を中心に、コーパスを利用した辞書の記述

2. 研究の進捗状況

全般に、研究は順調に進んでいる。以下、4グループごとに述べる。

- (1) 現在まででWWWおよび現代日本語話し言葉均衡コーパス(BCCWJ)からコロケーションを取り出し、どのようにコロケーションが記述できるか、またデータとその網羅性などを検討した。合わせて、コーパス中の間違いが辞書記述に与える影響についても研究した。
- (2) 現在までの段階で、20年度までの試行コーパスを用いて複合辞候補を抽出し終わったところである。それぞれの意味の解析およびデータの整理を行っている。また、複合辞全体の体系についての考察も徐々に進めている。
- (3) 18年から20年にかけて、利用者の観点から国語辞典の項目間の揺れを測定し、それを国語辞典の項目分けや国語辞典の記述に適用することを示した。21年からは、コーパスを用いて、辞典における動詞の自他認定の基準を策定する作業に入っている。
- (4) 18年度には、辞書に必要な情報の調査と

して辞書編纂者からの聞き取りを行い、それを報告書にまとめた。また、既存のコーパス（主に新聞）の収集を行った。19年度には、「ト格要素をとる体言の用例収集と分析」及び「役割の二格を中心にした周辺の格情報の分析」を行い、辞書に格情報を記述する方法を探った。20年度には、辞書の例文にどのように格情報を盛り込むかの検討を行った（約1,400語）。領域内公開コーパス(BCCWJ)がどの程度利用可能かの調査を行った。合わせてオノマトペの調査も行った。

3. 現在までの達成度

全体として、「②おおむね順調に進展している」に該当する。以下、4グループごとにおおよその達成度の目安を示す。

- (1) 「③やや遅れている」状態である。コロケーションの記述についていえば、BCCWJでは、十分な量のデータが含まれず、有意義な辞書記述が行えないことがわかった。したがって、当初の目的としていたBCCWJを利用してコロケーション辞書を作るという方針は5年間で達成できない見込みである。現在は、当初の方針を変更してWWWをコーパスとして利用して、コロケーション辞書の作成を試みているが、こちらはコーパスが大きすぎるとのことと、各種の間違いが大量に含まれるため、別の意味で辞書記述にふさわしくないという問題があることがわかった。
- (2) 目標に対して40%ほどの達成であり、「②おおむね順調に進展している」状態である。予定していた辞書形式の複合辞データベースの基礎的な形は完成した。まだ語彙のリストをつくった段階であるが、今後は、用例の増補などを行っていく予定である。

(3)「②おおむね順調に進展している」といえる。当初の目標に対してほぼ80%の達成段階である。国語辞典の意味区分の問題点などについては、ほぼ解明した段階であり、今後残された問題があるものの、それぞれ担当者を決め、それぞれのテーマとして追求している体制になっている。

(4)「②おおむね順調に進展している」といえる。BCCWJを活用して辞書記述のいろいろな側面について順次検討しており、その成果は論文・学会発表などで公表してきた。その進展を考えると、当初の目標に対してほぼ60%を達成していると考ええる。

4. 今後の研究の推進方策

4グループがそれぞれに研究を進めるとともに、全体会議で研究の進捗状況を確認しあい、お互いにコメントする中で共同研究の実をあげる予定である。

以下、4グループごとに記す。

(1)WWWのデータを活用しつつコロケーションの研究を進める予定である。Google社が販売しているNグラムデータ集を利用して、コロケーション記述の効率化を図る予定である。

(2)今後は、コーパスの完成版による複合辞の増補作業や、複合辞辞書をデータベースの形にすることなどを計画している。

(3)20年度までの研究成果を元に、形容詞を対象に注釈付き見本辞書を作成するとともに、漢語の自他両用動詞の使用実態を記述した見本辞書を作成する。

(4)21年度は、20年度に調査した約1,400語の動詞から、語義と格情報との関わりに特徴のあるものを抽出し、領域内公開コーパスの情報をもとに記述を開始する。オノマトペについては、領域内公開コーパスにとどまらず、他のコーパスも収集しながら、語義分析、主体・述語の調査を進める。22年度は、動詞について、他のコーパスの情報も参照しながら記述を点検する。オノマトペの記述も完成させる。コーパスの利用法と評価をまとめる。

5. 代表的な研究成果

(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12件)

①荻野綱男「コロケーション辞書」国文学解釈と鑑賞 Vol.74, No.1, pp.70-78,(2009),査読なし

②荻野綱男・荻野孝野「日本語のコロケーション研究の歴史——計量言語学、自然言語処理などを中心に——」日本語学 Vol.26, No.12, pp.58-70,(2007),査読なし

③荻野綱男「形容動詞連体形における「な／の」選択について」計量国語学 Vol.25, No.7, pp.309-318,(2006),査読あり

④近藤泰弘「『敬語の指針』と敬語理論」日本語学 27巻7号,(2008),査読なし

⑤矢澤真人「ユビキタス辞書の時代」日本語学 26巻8号, pp.58-66,(2007),査読なし

[学会発表] (計6件)

①荻野綱男「WWWを使ったコーパス研究の現在と、その問題点——日本語研究の観点から」ドイツ語情報処理学会,(2008.9.28)

②荻野綱男「WWWによる単語の文体差の研究」日本語学会2006年度秋季大会,(2006.11)

③近藤泰弘「日本語の照応表現をめぐる」日本語文法学会第9回大会シンポジウム,(2008.10)

④矢澤真人・橋本修・楊ソルラン・下村健司「利用者から見た国語辞典のランチ分けの妥当性について」筑波大学日本語日本文学会,(2008.9.20)

⑤丸山直子「役割の二格—周辺の格の扱いについて—」計量国語学会第51回大会発表(計量国語学26[3], pp.106-108), (2007.9)

[図書] (計0件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]