様式 C-19

科学研究費補助金研究成果報告書

平成 23 年 4 月 20 日現在

機関番号：３２６６５
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 Dictionary Compilation Methodology using Japanese 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

研究成果の概要（和文）：我々は、以下の四つの研究テーマをそれぞれ研究した。
（１）コーパスを用いたコロケーションの研究
コロケーションの抽出には BCCWJ よりも WWW のほうが向いていることがわかった。
（２）複合辞の抽出と整理
複合辞辞典を作成した。
（３）自他両用動詞の区分と辞書記述
動詞の用法を BCCWJ などのコーパスで調べ、区分できた。
（４）動詞の格情報とオノマトペの記述
BCCWJ からそれぞれの情報を取り出し、整理して辞書記述に役立てた。

研究成果の概要（英文）：We performed the following four research themes.
(1) research on collocations using corpora
(2) extraction and categorization of compound particles and compound auxiliary verbs
(3) classification and dictionary description of transitive-intransitive Japanese verbs
(4) description of case information of Japanese verbs and onomatopoeia

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００６年度	10,000,000	0	10,000,000
２００７年度	9,000,000	0	9,000,000
２００８年度	6,800,000	0	6,800,000
２００９年度	6,700,000	0	6,700,000
２０１０年度	5,900,000	0	5,900,000
総　計	38,400,000	0	38,400,000

研究分野：言語学
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，辞書記述，コロケーション

１．研究開始当初の背景

日本語の各種コーパスが研究に利用されるようになるとともに、この特定領域研究でも 5,000 万語のコーパスが作成されることに

なった。
そこで、これらのコーパスを利用して、国語辞典の記述や編集をすすめるとともに、その作業を通じて、どんなコーパスが望ましいか、また作成中のコーパスが本当に辞書記述に役立つのか、役立つとしたらどんなところに有用なのかを探る必要があった。

２．研究の目的

我々のチームでは、四つのグループを構成して、それぞれの課題を設定し、研究を遂行した。

第１グループでは、コーパスを用いてコロケーション辞書の概念設計を行い、一部の試作を行うことを目的とした。

第２グループでは、統語論的観点を取り入れて、コーパスから抽出したデータを用いて、複合辞の辞書記述を行うことを目的とした。

第３グループは、コーパスを利用して実態分析を行うことで、国語辞書の語義記述の改善と品詞分類の問題を追及することを目的とした。

第４グループでは、コーパスを用例集として扱い、それらの用例から動詞の格情報とオノマトペに焦点を絞り、国語辞書の記述に有用な情報を抽出し、実際の辞書の改訂に活かすことを目的にした。

３．研究の方法

第１グループでは、WWWの用例を中心に扱うことになった。ＢＣＣＷＪ をいろいろ試用してみると、コロケーション情報を抽出する基礎データとしては量が不足していることが明らかになったためである。最終的には、WWWを直接検索するのでなく、GoogleのN-gram データを利用して、研究をすすめることになった。これはＷＷＷの情報をGoogle が整理したものである。この中からコロケーションを抽出することで有意義な辞書記述ができることが確認できる。

第２グループでは、ＢＣＣＷＪから文字列の連続としてN-gram を抽出し、それらを整理することで大量の複合辞を得ることができた。それらの性質を分析するため、実際の使用例をＢＣＣＷＪから検索し、それに基づいて複合辞の分類を行った。

単純に形態素解析をしてＫＷＩＣ索引を作るだけではうまくいかないので、形態素解析を行った後、短単位が複数結合したN-gram を作成し、2 グラムから 5 グラムまでの結合をリストアップする。そして、その中で単純頻度が高いもの、ＴスコアやＭＩスコアの高いものなどを調査し、複合辞の候補リストを作成し、それらから各種のチェックリストによって手作業で抜き出した。最初はＴスコア等でかなり自動化が可能かと考えていたが、意外にむずかしく、単純頻度によるリストからの手作業の部分が大きかった。

第３グループでは、二つの研究を行った。一つは、国語辞典における形容詞の語義区分が妥当であるかどうかに関する研究であり、被調査者多数に質問調査をおこなって、辞書の語義区分と実際の用例の語義の分類を対照させた。もう一つは、動詞の自他両用動詞の研究であり、自動詞か他動詞かをめぐって、辞書の記述にゆれがある例を多数集め、その用例をＢＣＣＷＪで検索して、用例から自動詞か他動詞かを分類するようにした。

第４グループは、ＢＣＣＷＪをコーパスとして利用し、動詞を検索してそれがどのような名詞と格関係を有するかを調べた。動詞は、先行研究で問題になるような格関係を含むものとした。また、オノマトペについても同様にＢＣＣＷＪを資料として辞書記述を行った。

４．研究成果

第１グループでは、WWWの資料性に関する知見を得た。WWWにはいろいろ間違いがあるが、言語資料として十分使えることがわかった。また、ＢＣＣＷＪはコロケーション記述のためにはあまり有用ではないことがわかった。これは、コロケーション情報はきわめて個別の事象であることから、膨大な資料なしでは十分な記述ができないということを表している。ＢＣＣＷＪでは小さすぎるのである。

また、一部の語彙について、コロケーション記述の材料となるように、情報を整理した。ただし、情報の整理には思いの外、時間がかかり、十分な語数が整理できたとはいいにくい。

第２グループでは、予定通り、「ＢＣＣＷＪ複合辞辞書」(Ver.1.0)を作成した。エクセルによるデータベースとなっているが、それ以外に、複合辞辞書(印刷版)も作成した。印刷版では、小見出し 925 個を見出し語として掲出してある。特に今回の成果としたいのは、接続詞の取り扱いである。そもそも複合辞は、自立語であるいわゆる内容語（名詞・動詞等）が、付属語である機能語（助詞・助動詞）へと変化するのである。しかし、「それで」など、接続詞の多くは複数の形態素が結合した複合形式であるが、接続詞自体は、自立語である。文法論的には、接続詞は、接続助詞と副詞との中間に位置するものであり、文と文を接続する文法的機能をも持っていることは明らかである。したがって、自立語という点を重視すれば複合辞には入れるべきではないが、機能語の側面を重視すれば、複合辞としてもよい部分もある。このあたりの問題についての見通しを得ることができた。なお、「それが」(接続詞)のように、「代名詞+格助詞」が接続詞になるときには、機

能語化という面を重要視すれば、文法化が起きていると言える。それに対して「だが」(接続詞) の場合は、「・・・だが」という接続助詞 (付属語) から変化したものであるので、接続詞が自立語であるという側面を重視すれば、文法化ではなく、その逆の「脱文法化」が起きている解釈可能である。このように、接続詞の複合辞としての性格付けには複雑な問題が残っており今後の課題となる。

第３グループでは、国語辞書の中の形容詞の語義区分の問題点を明らかにした。それとともに、自他両用動詞についても、912 語について多数の用例を調べ、品詞認定をどう行うべきかについて結論を得た。

第４グループでは、動詞の格情報については、コーパスにおける使用実態を、実際に岩波国語辞典第七版に反映させた（主に例文として記載)。動詞の格情報についてもオノマトペについても、コーパスの情報を踏まえた辞書を試作した。基礎的な動詞の格情報の記述にはＢＣＣＷＪがかなり有効だが、オノマトペの記述には不十分であることが明らかになった。文法的事項と語彙的事項の差と言えるかもしれない。

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 18 件）

①丸山直子、動詞の格情報－国語辞書の記述とコーパスー、東京女子大学日本文学、査読無、107、2011、227-245

②多田知子、複合接続詞―文の文頭部分の階層性―、国文論叢（神戸大学文学部)、査読有、４２巻、2010、15-28

③矢澤真人、検証辞書・辞典　国語辞典の情報を活用する、月刊国語教育研究、査読無、４６１号、2010、28-31

④丸山直子、助詞「に」を伴う＜役割＞成分－コーパスに基づく分析－、日本語文法、査読有、10[1]、2010、71-87

⑤矢澤真人、文型と語釈、筑波日本語研究 (筑波大学)、査読無、14、2010、1-18

⑥荻野綱男、コロケーション辞書、国文学解釈と鑑賞、査読無、74[1]、2009、70-78

⑦荻野綱男、WWWをコーパスとしてみたときの間違い、言語、査読無、36[6]、2009、6-7

⑧荻野綱男、検索エンジン Google における「単語」、語文、査読無、134、2009、1-4

⑨矢澤真人、「伝統的な言語文化」と国語科教育、月刊国語教育研究(日本国語教育学会)、査読無、2009、4-9

⑩荻野綱男、WWW をコーパスとして利用する研究――文系と理系の観点から――、日本語学、査読無、27[2]、2008、4-9

⑪荻野綱男・末永絵梨・下重秋弓・三好亜萌、WWW の検索による日本語研究(2)、東京女子大学日本文学、査読無、103、2007、147-166

⑫荻野綱男、ブログにみる日本語の男女差、日本語学、査読無、26[4]、2007、58-64

⑬荻野綱男、コーパスとしてのWWW検索の活用、言語、査読無、36[7]、2007、26-33

⑭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、査読無、26[12]、2007、58-70

⑮矢澤真人、情報社会と日本語、日本言語文化（韓国日語日文学会)、査読無、10、2007、10、5-23

⑯矢澤真人、ユビキタス辞書の時代、日本語学（明治書院)、査読無、26-8、2007、58-66

⑰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、査読有、25[7]、2006、309-318

⑱荻野綱男・秋山智美・柴田雪乃、国語辞書の利用方法と利用意識、語文、査読無、126、2006、14-27

〔学会発表〕（計 8 件）

①矢澤真人、外形から引く日本語辞典への試み、平成 22 年度筑波大学国際連携プロジェクト企画国際研究フォーラム「日本語学習辞書の開発と日本語研究」、筑波大学、2010.12.12

②矢澤真人、日本語変換システムと国語辞典、語彙・辞書研究会第３８回研究発表会、新宿ＮＳビル、2010.11.20

③矢澤真人、コーパスを利用した言語教育、北京師範大学・中日の言語研究・言語教育シンポジウム、中国北京、2010.10.17

④荻野綱男、ＩＴコミュニケーションから見

る日本語の将来、日本学術会議主催公開講演会「日本語の将来」、日本学術会議、2010.9.19

⑤荻野綱男、日本語学の見地からデータ収集の過去100年と未来100年を考える、第8回韓国日本学連合会国際学術大会、2010.7.2

⑥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、学習院大学、2008.9.28

⑦丸山直子、役割のニ格－周辺的な格の扱いについてー、計量国語学会第51回大会、、日本大学文理学部、2007.9.29

⑧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、岡山県 2006.11.12

６．研究組織

(1)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

(2)研究分担者

近藤　泰弘（KONDO YASUHIRO）
青山学院大学・文学部・教授
研究者番号：20126064

矢澤　真人（YAZAWA MAKOTO）
筑波大学・人文社会科学研究科・教授
研究者番号：30182314

丸山　直子（MARUYAMA NAOKO）
東京女子大学・現代教養学部・教授
研究者番号：00199936

(3)連携研究者

なし

(4)研究協力者

多田　知子（TADA TOMOKO）
青山学院大学・大学院生