

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 27 日現在

機関番号：62618

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23520572

研究課題名(和文)コーパス分析に基づく辞書の位相情報の精緻化

研究課題名(英文) An Analysis of Register Information of Japanese Language Described in Dictionaries and Appearing in the Balanced Corpus of Contemporary Written Japanese

研究代表者

柏野 和佳子 (KASHINO, Wakako)

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・言語資源研究系・准教授

研究者番号：50311147

交付決定額(研究期間全体)：(直接経費) 3,800,000円、(間接経費) 1,140,000円

研究成果の概要(和文)：本研究は、その重要性にも関わらず、従来の国語辞典において記述が不十分だった位相情報を取り上げる。はじめに、国語辞典・英語辞典に注記されている位相情報を調査した。次に、位相情報のうち、「古風な語」に着目し、各種国語辞典の記述の比較調査、『現代日本語書き言葉均衡コーパス』における使用実態の調査と用例分類、辞書記述方法の作成を行った。また、日本語学習者や、母語話者である大学生が作文において誤りやすい位相情報の調査・分析を行った。話し言葉的な語がレポート類の作成時に誤って用いてしまう例を抽出した。そして、誤用を訂正する情報がコーパス分析から得られる実例を示した。

研究成果の概要(英文)：We discuss detailed description for register information of Japanese language. First, we review descriptions of register information in existing Japanese and English dictionaries. We then focus on some old-fashion words, and perform a comparative study using widely-used Japanese dictionaries and an analysis using the Balanced Corpus of Contemporary Written Japanese, to propose a method of its detailed lexical description. We also study register information that tends to be misused by Japanese learners and young native Japanese speakers, and find some typical examples where spoken styles are used in written documents. We show that our proposed description method can be utilized to avoid, detect, and correct such misuse cases.

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：辞書 コーパス 語彙 意味 位相

1. 研究開始当初の背景

これまでの国語辞典は、辞典編纂者による用例収集や内省によって記述されてきた。しかしながら、欧米におけるコーパス分析による辞典編纂の先行事例の刺激もあり、近年、日本においても、様々なコーパスが構築されるに伴い、コーパス分析を辞書記述に活かそうとする議論が活発になってきている。

本研究は、「コーパス分析に基づく辞書情報の精緻化」を進める一環として取り組むものである。中でも、その重要性にも関わらず、記述が不十分だった位相情報を取り上げ、コーパス分析に基づき、辞書の位相情報の精緻化を行う。

2. 研究の目的

国語辞典や類語辞典に注記されている「古語、雅語、文語、文章語、口語、俗語」といった位相情報は、語の理解を助け、文章作成時にはその語を用いる判断の指標になり得るものである。しかしながら、その重要性にも関わらず、分類基準の曖昧さや網羅的記述の不十分さが指摘されていた。そこで、『現代日本語書き言葉均衡コーパス』等のコーパスを活用し、コーパスから抽出可能な位相情報の分析と記述を試みる。辞書の位相情報の精緻化が目的である。本研究成果は作文教育や日本語教育、また、自然言語処理の文章生成という応用場面においての活用も期待できる。

3. 研究の方法

(1) 現行辞典の調査

国語辞典・英語辞典に注記されている位相情報を調査する。

(2) 「古風な語」の調査・分析

位相情報のうち、「古風な語」に着目し、各種国語辞典の記述調査、コーパスにおける使用実態の調査と用例分類、辞書記述方法の作成を行う。

コーパスは、主に、『現代日本語書き言葉均衡コーパス』(国立国語研究所)を用いる。ほか、『青空文庫』(<http://www.aozora.gr.jp/>)や『太陽コーパス』(国立国語研究所)なども適宜参照する。

(3) 作文において誤りやすい位相情報の調査・分析

日本語学習者が誤りやすい位相情報の調査・分析を行う。また、母語話者である大学生が誤りやすい位相情報の調査・分析を行う。

(4) 書き言葉と話し言葉の調査・分析

書き言葉と話し言葉の助詞の分布の分析を行う。

4. 研究成果

(1) 現行辞典の調査

次のとおり、合計7種の辞典について、時間的特性に関する位相情報の種類と、その注記が付与されている見出し語の比較調査を行った。

『岩波国語辞典』(第6・7版,岩波書店) ... ラベルは、〔俗〕(俗語)と〔仏〕(仏教語)の2つのみを使用。時間的特性を含めたその他の位相情報は、注記にあり。

『学研現代新国語辞典』(第4版,学習研究社) ... 〔古〕(古語)・〔文〕(文章語)・〔雅〕(雅語)を使用するほか、語釈と注記にも位相情報あり。

『明鏡国語辞典』(第2版,大修館書店) ... 〔古〕(古語)・〔文〕(文語)・〔雅〕(雅語)を使用するほか、語釈と注記にも位相情報あり。

『三省堂国語辞典』(第6版,三省堂) ... 〔古風〕・〔文〕(文語)・〔雅〕(雅語)を使用するほか、語釈にも位相情報あり。

『新選国語辞典』(第9版,小学館) ... 〔古語〕・〔文章語〕を使用するほか、語釈と注記にも位相情報あり。

『新明解国語辞典』(第6版,三省堂) ... 〔古〕(古語)・〔雅〕(雅語)を使用するほか、語釈にも位相情報あり。

『語感の辞典』(岩波書店) ... ラベルはなし。語釈中に「古風」「文語的」「古めかしい」といった位相情報が含まれる。

各辞書の注記は、例えば、次のとおり。

いかずち【雷】

『岩波国語辞典』 雅語的。
『学研現代新国語辞典』〔文〕
『明鏡国語辞典』〔雅〕
『三省堂国語辞典』〔雅〕
『新選国語辞典』〔古語〕
『新明解国語辞典』「雅語的」
『語感の辞典』「古語に近い古めかしい語」

いわお【×巖】

『岩波国語辞典』 雅語的。
『学研現代新国語辞典』〔文〕
『明鏡国語辞典』
『三省堂国語辞典』〔雅〕

『新選国語辞典』
『新明解国語辞典』
『語感の辞典』 「ぐめかしい 雅語的な和語」

また、次の英和辞典・英英辞典を対象に、位相情報の付与状況を比較分析した。

ロングマン英和辞典 / 桐原書店 (2007)
収録: 約 10 万語
ロングマン現代英英辞典 (5 訂版) / 桐原書店 (2008) 収録: 約 23 万語
ロングマン現代アメリカ英語辞典 (2 訂版) / ピアソンエデュケーション (2008)
収録: 約 18 万語
コリンズコウビルド英英辞典 (改訂新版) / 桐原書店 (1995) 収録: 約 7 万 5 千語、
ケンブリッジ英英辞典 (3 版) 収録: 約 17 万語。
ウィズダム英和辞典 (2 版) / 三省堂 (2007) 収録: 約 9 万語
ユースプログレッシブ英和辞典 / 小学館 (2004) 収録: 約 8 万 5 千

「Time」に関わるラベルには、「dated」、「old-fashioned」、「old use」、「archaic」の 4 つがある。これらは、「比較的最近まで使用され、現在高齢者に用いられている」「近現代英語では使われない」「過去の世紀に使用された」ことを示すものとして用いられている。

(2) 「古風な語」の調査・分析

『岩波国語辞典』に「古語的」と注記がついている 15 語と、「古風」と注記がついている 145 語を調査対象語に選定し、現行の国語辞典 5 種の記述を調査した。そして、『現代日本語書き言葉均衡コーパス』に収録される図書館サブコーパスの使用頻度調査と用例分析とを行った。「古風な語」の使用を、(1) 古典の引用、(2) 古い文体、(3) 時代・歴史小説、(4) 現代文脈、に 4 分類する辞書記述方法を提案し、実際に記述を行った。その結果、文体的特徴や用例を辞書に記述することは、作文や日本語学習などのさまざまな用途に対する情報量を増やし、その利用価値を高めることを明らかにした。その成果を論文誌に投稿中である。

(3) 作文において誤りやすい位相情報の調査・分析

日本語教育におけるアカデミック・ライティング指導において文体差の指導が必要となる語の分析を行った。日本語学習者にとって必要度の高い文体的特徴を持つ語を、日本語教育の修士コースの学生 4 人がアカデミックな文章として書いた合計 64 の文書から抽出した。例えば、「A. そんなに / それほど、B. どれくらい / どの程度、C. 一番 / 最も、D.

すごく / 非常に」の 4 組 (前者がいわゆる話し言葉的で修正対象となった語、後者が書き言葉的な修正候補語となった語) についてその使用実態を『現代日本語書き言葉均衡コーパス』(BCCWJ) を硬 [白書, 教科書, 広報紙, 新聞], 中 [書籍], 軟 [雑誌, Web データ] の 3 つに区分して調べた (中納言 1.0.2 を使用)。その結果、A, B は、軟らかい文体で話し言葉的な語 (そんなに、どれくらい) が多く使われ、一方 C, D は、特に硬い文体で話し言葉的な語 (一番、すごく) がほとんど使われていない、という特徴を捉えることができた。いずれも使用実態からその文体差が明確に捉えられることがわかった。また、今回の調査で、書き言葉コーパスであるにも関わらず、いわゆる話し言葉的な語が、硬、中、軟のいずれの文体の文章にも使用され、その割合もあまり低いものではないという点が確認できた。それらの成果をまとめ、ポスター発表を行った。

さらに、日本語教育の修士コースの学生 4 人がアカデミックな文章として書いた合計 72 の文書から、文体的な観点から修正が必要な例を全部で 869 件抽出した。そこから、「N の N」か「NN」か、「な」か「あるの / の」か、「ない」か「ず」か、の 3 点をとりあげ、BCCWJ を用いて分析し、その検索方法及び分析結果を報告した。

「N の N」か「NN」か
アカデミックな文章作成時においては、冗長と判断される「の」は省かれる場合が多い。「の」を省くべきと修正された事例は、おおむね「NN」の頻度が高いという結果が得られた。

「な」か「あるの / の」か
「な」か「の」かは、どちらもあり得る形の場合に正誤は言い難いが、ネイティブの日本人がより文体差、あるいは、自然さを感じる違いがあるような場合には、検索結果の頻度差としてそのことを確認することができた。

「ない」か「ず」か
BCCWJ のサブコーパスを、硬 [白書, 教科書, 広報紙, 新聞], 中 [書籍], 軟 [雑誌, Web データ] の 3 つに区分して、各表現形の 100 万語あたりに正規化した使用頻度平均を求めて分析した。その結果、文章が硬いほど「ず」が用いられ、軟らかいほど「ない」の形が用いられていることを明らかにできた。

以上のほか、「台湾人日本語学習者の誤用調査」「母語話者である大学生のレポートの語彙調査」を行った。ここから、例えば「得た」とすべきところを「もらった」とするような誤りや、「だいが、たくさん、たまに、結構、案外」などのくだけた副詞をレポートで多用するなどの問題点を抽出した。

(4) 書き言葉と話し言葉の調査・分析

本研究では、次のものを「助詞」の範囲に

定め、BCCWJ のコアを用いて調査・分析を行った。

格助詞・副助詞・係助詞・接続助詞・
終助詞・間投助詞・連体助詞・準体助詞・
並立助詞

その結果、次のことを明らかにした。

- ・全語数の約 30%が助詞である。
- ・助詞の中では格助詞が最も多い。全体語数の 15～19%。助詞内では 47%～74%。
- ・知恵袋・ブログは、他に比べて、格助詞が少なく、終助詞が多い。
- ・白書と、知恵袋・ブログは、対極にある。白書はかなり特殊である。新聞は多少白書に似た性質を持つ。知恵袋とブログは、ともに話し言葉的な性質を帯びているが、異なる性質も持つ。

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)

[学会発表](計 2 件)

木田真理・柏野和佳子「書き言葉均衡コーパスを活用した文体的特徴を持つ語の分析と記述の試み アカデミック・ライティング指導を例に」『日本語教育国際研究大会名古屋 2012 予稿集第 1 分冊 (ポスター発表)』p.217(名古屋大学, 2012 年 8 月 18 日)

木田真理・柏野和佳子「『書き言葉的』として指導する必要がある語の分析 - 『現代日本語書き言葉均衡コーパス』を利用して -」『第 5 回コーパス日本語学ワークショップ』pp.293-300 (国立国語研究所, 2014 年 3 月 7 日)

6. 研究組織

(1)研究代表者

柏野 和佳子 (KASHINO, Wakako)
国立国語研究所・言語資源研究系・准教授
研究者番号：50311147

(2)研究分担者

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

木田真理 (KIDA, Mari)
政策研究大学院大学・政策研究科・教授
研究者番号：80401727