

令和元年6月19日現在

機関番号：62618

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K13228

研究課題名（和文）大規模コーパスに基づく日本語機能語の基礎研究と機能語検索ツールへの応用

研究課題名（英文）Large-scale corpus based research on functional words in Japanese and its application to the development of a search tool

研究代表者

PARDESHI P.V. (Pardeshi, Prashant)

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・理論・対照研究領域・教授

研究者番号：00374984

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：本研究では機能表現の振る舞いを提示する新たなレキシカルプロファイリング型ツールの開発・公開するために、機能表現のプロファイリングに関する基礎研究を行い、機能表現の範囲を明確にした上で、機能表現の品詞・タイプごとに共起する品詞の種類や文法形式のパターンを整理し、コーパスから機能表現を抽出するための文法パターンを作成した。このパターンファイルを利用して、BCCWJから機能語のパターンを網羅的に抽出し、機能語の振る舞いを記録したデータベースを作成した。最後に、このデータベースを検索するための機能語のプロファイリングツールを一般公開した。<http://nlb.ninjal.ac.jp/fep/>

研究成果の学術的意義や社会的意義

日本語のコーパス検索ツールを無料一般公開することは、言語調査のための基盤を広く一般に提供することにつながる。まず、機能表現が検索できるツールNLBが公開することにより、これまで、内容語の研究にしか利用できなかったこのツールが機能表現の研究にも利用され、その研究基盤としての有用性をさらに高めることになる。また、NLBは公開以来、国内外の大学での日本語教育など、教育分野でも広く利用されてきた実績があり、機能表現を学習する授業での活用や、機能表現の教材開発など、幅広い利用が期待される。さらに、このツールがいわば機能表現辞典としての機能を果たし、一般社会にも大きく寄与できるものと期待される。

研究成果の概要（英文）：In this project we conducted research on functional words in Japanese using a large scale corpus (BCCWJ) with a view to develop a search tool dedicated to functional words. First, we identified the functional words and classified them according to their part-of-speech and types. Further we prepared grammatical patterns to extract the words that collocate with the functional words. Using these patterns, we extracted data from the BCCWJ corpus and created a database. Finally, we developed a search tool that allows the user to search functional words and extract information on their collocates in order to understand their usage in context. The tool can be accessed for free from the following link: <http://nlb.ninjal.ac.jp/fep/>

研究分野：コーパス言語学、日本語学

キーワード：機能表現 レキシカルプロファイリング型ツール

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

コーパス検索ツールとして長年利用されてきたのがコンコーダンスである。コンコーダンスでは、検索語句を縦一列に中央に表示し、その左右に文脈を表示する。前後の文脈が簡単に確認できるため、検索語がどのような語と共起し、どのような文法形式で用いられるのかが一目で分かるようになっている。しかし、ウェブ上のテキストデータをコーパスとして利用し始めた1990年代の後半には、コーパスの大規模化が進み、このような1件ずつ文脈を確認するコンコーダンスのみによる分析は限界に達し、代わって、検索語と共起する語や文法形式などをあらかじめ統計的に要約するレキシカルプロファイリングツール Word Sketch が生まれた (Adam Kilgarriff, David Tugwell 2001)。このツールは、特に EFL 辞書の編集の分野で大いに活用され、現在では、EFL 辞書の編集には欠かせない存在となっている。従来、このレキシカルプロファイリングでは、内容語の振る舞いにのみ焦点が当てられてきたため、研究開始時点では、機能語などの機能表現のためのレキシカルプロファイリングツールは国内外を含めて存在しなかった。

### 2. 研究の目的

近年、大規模コーパスに基づいた日本語の実証的な言語研究が盛んである。コーパス基盤の研究環境として、大規模コーパスとともに重要な位置を占めるのが、コーパス検索の専用ツールである。研究代表者がリーダーを務める研究チームでは、『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)のレキシカルプロファイリング型の検索ツール NINJAL-LWP for BCCWJ (以下、NLB)を開発し、2012年6月に国立国語研究所より一般無料公開した。以下の図は NLB による動詞「売る」のプロファイリング結果である。



公開から3年半足らずの間に、国内外の6万人以上(2015年10月現在)のユーザに利用されるツールにまで成長した。NLBを含めて、従来のレキシカルプロファイリングツールでは、動詞や名詞などの内容語の振る舞いのみを対象としてきたが、機能表現を対象としたレキシカルプロファイリングツールがないため、大規模コーパスに基づいた機能表現の実証的な研究が難題である。本研究では、言語研究のテーマとして関心の高いモダリティも念頭に置き、機能表現

の振る舞いを提示する新たなレキシカルプロファイリング型ツールの基礎研究と開発を行い、ツールを一般公開することを目的とした。

### 3. 研究の方法

本研究では、以下の3つの課題に取り組んだ。

(あ) まず、ツール開発の準備段階として、機能表現のプロファイリングに関する基礎研究を行う。複合辞も含めて、機能表現の範囲を明確にした上で、機能表現の品詞・タイプごとに共起する品詞の種類や文法形式のパターンを整理し、コーパスから機能表現を抽出するための文法パターンを作成する。同時に、モダリティを抽出するためのパターンも作成する。

(い) 次に、作成したパターンファイルを利用して、BCCWJ から機能語のパターンを網羅的に抽出し、機能語の振る舞いを記録したデータベースを作成する。

(う) 最後に、このデータベースを検索するための機能語のプロファイリングツールを NLB に追加して一般公開する。

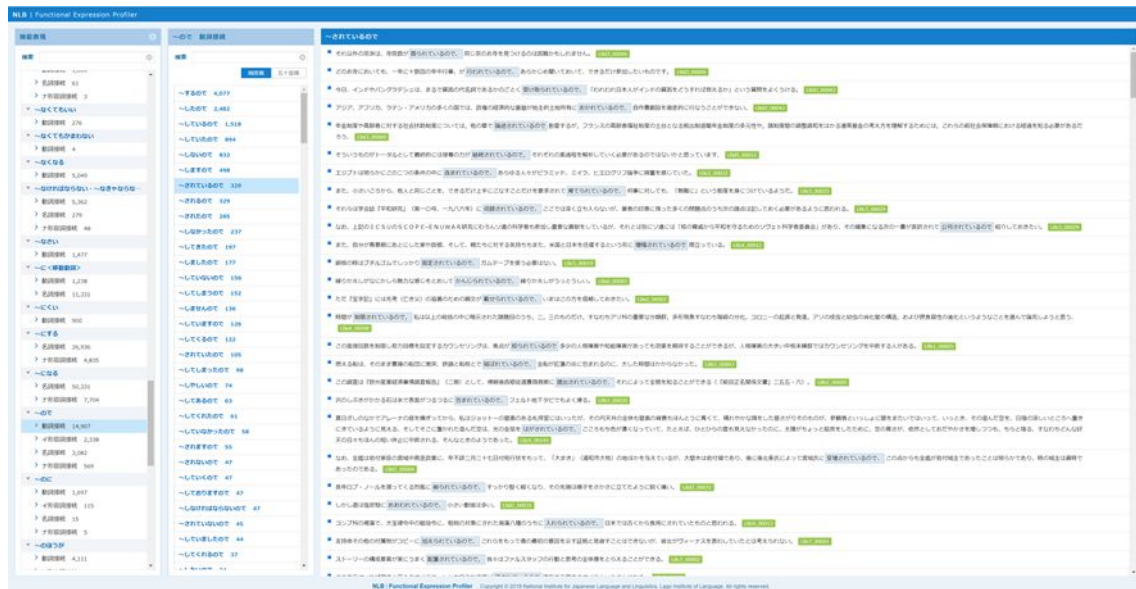
### 4. 研究成果

(あ) まず、ツール開発の準備段階として、機能表現のプロファイリングに関する基礎研究を行う。複合辞も含めて、機能表現の範囲を明確にした上で、機能表現の品詞・タイプごとに共起する品詞の種類や文法形式のパターンを整理し、コーパスから機能表現を抽出するための文法パターンを作成した。同時に、モダリティを抽出するためのパターンも作成した。

(い) 次に、作成したパターンファイルを利用して、BCCWJ から機能語のパターンを網羅的に抽出し、機能語の振る舞いを記録したデータベースを作成した。

(う) 最後に、このデータベースを検索するための機能語のプロファイリングツールを NLB に追加して一般公開した。以下のリンクから無償で利用できる。

<http://nlb.ninjal.ac.jp/fep/>



日本語のコーパス検索ツールを無料一般公開することは、言語調査のための基盤を広く一般に提供することにつながる。まず、機能表現が検索できる NLB が公開できたため、これまで、名詞、動詞、形容詞、副詞などの内容語の研究にしか利用できなかったこのツールが機能表現の研究にも利用され、研究基盤としての NLB の有用性をさらに高めることになる。また、NLB は公開以来、国内外の大学での日本語教育など、教育分野でも広く利用されてきた実績があり、機能表現を学習する授業での活用や、機能表現の教材開発など、幅広い利用が期待される。さらに、NLB は、ライターや翻訳者などの一般のユーザの利用も多いため、このツールがいわば機能表現辞典としての機能を果たし、一般社会にも大きく寄与できるものと期待される。

## 5. 主な発表論文等

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 0 件)

〔図書〕(計 0 件)

〔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  
<http://nlb.ninjal.ac.jp/fep/>

## 6. 研究組織

### (1) 研究分担者

研究分担者氏名：砂川 有里子  
ローマ字氏名：SUNAKAWA Yuriko  
所属研究機関名：筑波大学  
部局名：人文社会系（名誉教授）  
職名：名誉教授  
研究者番号（8桁）：40179289

研究分担者氏名：今井 新悟  
ローマ字氏名：IMAI Shingo  
所属研究機関名：筑波大学  
部局名：人文社会系  
職名：教授  
研究者番号（8桁）：50346582

### (2) 研究協力者

研究協力者氏名：  
ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。