

令和元年6月12日現在

機関番号：13103

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K13439

研究課題名(和文)文章を構成する語彙の相互関係に関する定量的研究

研究課題名(英文)Quantitative Study on the Correlation of the Vocabulary in Connection with Textual Structures

研究代表者

鯨井 綾希(Kujirai, Ayaki)

上越教育大学・大学院学校教育研究科・講師

研究者番号：10757850

交付決定額(研究期間全体):(直接経費) 1,300,000円

研究成果の概要(和文):本研究では、コーパス言語学が得意とする、大規模データとコンピュータを用いた計量分析を行い、日本語の文章中で使われる語彙の相互関係を分析した。一年目は、いくつかの語彙的要素のうち、形容動詞を中心としたコロケーションや語彙分類に関する研究を行い、計量分析の際に有効な方法論を検討した。二年目は、文章として実現されたコーパス内の語彙のうち、品詞の量的増減に関する相互関係と、そのテキストへの影響を分析対象とし、大規模データによる探索的な計量分析によりその実態を詳らかにした。以上の調査・研究を通して、文章の構造に関わる語彙の相互関係を明らかにし、語彙論と文章論にまたがる横断的な方法論を確立できた。

研究成果の学術的意義や社会的意義

本研究は、日本語の文章の中で、語彙という存在がどのような振る舞いを見せているのかという問題を詳らかにしたものである。本研究を通して、文章と語彙の相互関係が具体的に示されるとともに、その相互関係の解明に向けた研究の方向性を示すことができたと言える。また、本研究では、文章中に現れる語彙の相互関係という特定言語に依存しない視座に基づいて分析を進めた。このような視座からの研究は、本研究で対象とした日本語の文章および語彙の関係性の解明に役立つだけでなく、他言語を含めた通言語的な分析視座、特に語や文を越えた文章単位での対照研究の方法論に関する新たな可能性を提示できる可能性を持つと考えられる。

研究成果の概要(英文):This study intends to analyze quantitatively the correlation of the vocabulary in Japanese texts from the point of view of corpus linguistics. The first year, a part of classification of vocabulary and regularity of word collocation became clear by the research related to adjective verbs. In addition, effective methodology in quantitative analysis was examined. In the second year, the correlation about quantitative changes in parts of speech was analyzed. Furthermore, the influence of quantitative changes in parts of speech upon textual structures became clear. The correlation of the vocabulary in connection with textual structures became clear through this study. Moreover, the methodology based on both lexicology and text linguistics was established.

研究分野：言語学

キーワード：計量語彙論 語彙分類 品詞構成比率 語彙のネットワーク コロケーション コーパス 文章・文体

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

### 1. 研究開始当初の背景

コーパス言語学は、英語圏において 60 年代から、言語学のうちの一つの研究分野として位置づけられてきた (McEnery&Andrew,1996. Corpus Linguistic.)。同時期の日本でも、国立国語研究所によって電子データや大量の用例カードに基づく言語研究が行われていたが、データの公開や共有、再利用はほとんど行われず、コーパス言語学という発想は定着しなかった(丸山 (2013)「日本語コーパスの発展」『講座日本語コーパス 1 コーパス入門』)。2000 年代に入り、『日本語話し言葉コーパス』が国立国語研究所と情報通信研究機構から公開され、日本語を対象としてコーパス言語学的研究を行う素地ができた。さらに、2011 年に国立国語研究所が『現代日本語書き言葉均衡コーパス』を公開したことをひとつの契機として、日本語の研究にコーパスが利用される機会が大幅に増加した。中でも、計量的・統計的分析による言語上の傾向の発見が積極的に試みられており、日本語研究の中でそうした定量的研究の成果が確実にあがっていた。

### 2. 研究の目的

研究開始当初の段階で、コーパスと計量分析は様々な言語単位の中で試みられていた。その中でも、文章中の語ごとの出現位置や相互関係に基づく語彙の分類・構造の解明は、さらに研究を進める必要があった。また、それらは、大規模な言語データであるコーパスと、それを処理できるコンピュータを使用した計量分析を組み合わせることで、その他の方法論では見出せない言語的事実を発見できる可能性がある点で、将来性も有していた。さらに、文章と語彙との相互関係の分析は、有意義な文章を形成するための手段を明らかにする上でも、語彙が現実にとどのような集合体として形成されているかを明らかにする上でも、重要であると言える。

以上を踏まえ、本研究では、日本語の文章の中で使用された語同士の相互関係を分析し、文章と語彙の関係性を定量的な観点から解明することを目的として設定した。

### 3. 研究の方法

文章の中で語彙の相互関係を詳らかにするために、以下の観点・方法で研究を行った。

- (1) 文章中で使われる形容動詞相当の語群を対象に、個々の語が実際に使用された際の後接要素との共起関係を集計し、多変量解析を用いて語群の分類を行うことで、形容動詞の語彙的な広がりを明らかにする。
- (2) 文章中で使われる類義語の用法差の分析の一例として、類義語である「美しい」と「きれいだ」を取り上げ、それぞれが修飾する名詞を収集し、共起語の語彙ネットワークを構築した後、共起する名詞の種類の違いや相互の類似度を分析することで、語彙の文章中での使われ方を明らかにする。
- (3) 「樺島の法則」と呼ばれる、延べ語数における品詞構成比率の変動パターンを踏まえ、資料と品詞の種類に関する量的な充実を図りつつ再度の検討を行い、文章中における品詞構成比率の変動のあり方を明らかにする。

上記の観点・分析を通して、文章中に見られる語彙のあり方を多面的に考察し、語彙論と文章論にまたがる研究へと発展させることとした。

### 4. 研究成果

研究方法として設定した三つの観点に基づき、以下のことを明らかにした。なお、いずれも分析資料には国立国語研究所 (2015)『現代日本語書き言葉均衡コーパス』(Version 1.1 DVD版)を用いた。

- (4) 形容動詞相当の語群を、後接要素との関係で分析すると、以下の後接要素の特徴に基づいて形容動詞を分類できることが分かった。
  - a. 形容動詞の中心的後接要素「な」「の」 装定用法が特徴的な形容動詞
  - b. 特定の形容動詞に偏る形容詞・動詞、助動詞「に」 副詞的用法が特徴的な形容動詞
  - c. 特定の形容動詞に偏る係助詞・「の」以外の格助詞 名詞的用法が特徴的な形容動詞
  - d. 特定の形容動詞に偏る「である」「です」など 述定用法が特徴的な形容動詞また、使用頻度の高い 149 語の形容動詞を対象としてクラスター分析を行い、個々の語が上記四種類と「その他」のそれぞれに分類される様子を明らかにした。
- (5) 文章中での「美しい」と「きれいだ」の共起語およびそれらが使われる文章ジャンルを調査したところ、以下の結果が得られた。
  - a. 「美しい」は【自然、姿、娘、景観、言葉、風景】、「きれいだ」は【写真、手、俣、ピンク、空気、空、形、音、肌、状態、女の子、水、子、御花、方、御姉さん、体、川】を被修飾語に取る。それぞれの名詞の性質に基づく、既存の国語辞書の記述に近い差異(「美しい」は全体的/観念的な対象を指し、「きれい」は個別的/実体的な対象を指す)が見出せる。
  - b. 「美しい」と「きれいだ」の語彙的な近さは、共起語の相関係数で 0.423 であり、統計的には正の相関があるとみなせる類似度を持った位置にある。
  - c. 「美しい」は、書籍や白書に偏って使用され、非即時的な過程を経た書き言葉的な文章においてよく使用される。「きれいだ」は、ブログや知恵袋に偏って使用され、即時的な過程を経た話し言葉的な文章においてよく使用される。

これらの結果により、「美しい」と「きれいだ」の類義語としての具体的関係を見出すことができた。

(6)文書ごとの品詞構成比率を計算して並べ、名詞の増加に伴う品詞構成比率の変動を調査した。それにより、従来の「樺島の法則」と異なる以下の結果が得られた。

a. 動詞・形容詞・副詞(・助動詞・代名詞): 一定の割合で減少

b. 形状詞・連体詞(・助詞): 増加 減少

c. 接続詞: 減少 緩やかな増加ないし維持 減少

d. 感動詞: 急激な減少

また、上記の変動と文章のジャンルとの対応関係を調べたところ、名詞の増加に社会科学、自然科学、技術・工学の増加が対応し、動詞等の減少に文学と哲学の減少が対応し、形状詞等の増減に歴史と芸術・美術の増減が対応することが分かった。

以上に述べてきた三つの研究を通して、文章中における語彙の相互関係の一端が明らかになったと言える。

## 5. 主な発表論文等

〔雑誌論文〕(計 1 件)

鯨井綾希、形容動詞に属する語彙の分布と分類-語幹相当語に後接する要素の計量分析を通して-、上越教育大学国語研究 32、査読無、2018 年、pp.1-12

〔学会発表〕(計 2 件)

鯨井綾希、類義語における共起語ネットワークの差異-「美しい」と「きれいだ」を例として-、上越教育大学国語教育学会第 73 回例会、2018 年 2 月

鯨井綾希、品詞構成比率の変動と相互連関-コーパスを用いた「樺島の法則」の再検討-、日本語学会 2018 年度秋季大会、2018 年 10 月

〔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

## 6. 研究組織

(1)研究分担者

研究分担者氏名:

ローマ字氏名:

所属研究機関名:

部局名:

職名:

研究者番号(8桁):

(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。