

令和元年6月1日現在

機関番号：10102

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K02715

研究課題名(和文) コーパスを活用した接続表現研究の可能性に関する探索的研究

研究課題名(英文) Corpus-based exploratory research on the possibility of researching connective expressions.

研究代表者

馬場 俊臣 (BABA, Toshiomi)

北海道教育大学・教育学部・教授

研究者番号：70218668

交付決定額(研究期間全体)：(直接経費) 1,000,000円

研究成果の概要(和文)：文章・文体研究分野での接続表現研究の可能性を広げるために、コーパスを活用した探索的研究を行った。課題と成果は次の通りである。(1)『現代日本語書き言葉均衡コーパス』の形態論情報「接続詞」の解析精度の調査を行いその問題点と留意点を明らかにした。(2)文体と接続表現の使用に関する調査を行い次の3点を明らかにした。(a)学年別、教科別等の観点による教科書の接続詞の使用特徴、(b)接続詞の文体差の数値化及びその有用性、(c)助詞・助動詞で始まる複合接続表現の文体差の数値化及びそれらの文体的特徴。(3)複合接続表現の調査方法の問題点の検討を行い、文境界情報の利用の得失を明らかにした。

研究成果の学術的意義や社会的意義

日本語研究においては、近年コーパスを用いることによって、以前とは質的に異なった新たな多くの研究成果が得られてきている。本研究において、コーパスを用いた接続表現研究の方法の精緻化に関わる知見の提示や接続表現の文体研究における新たな指標の提示などを行うことによって、コーパスを用いた文章・文体研究の進展に寄与することができ、学術的意義を持つ。また、こうした知見や指標などは国語教育や日本語教育の研究や実践とも深く関わっており、教育面へ活用できるという社会的意義を持つ。

研究成果の概要(英文)：In order to expand the possibility of researching connective expressions in the field of text and style research using corpus, we conducted exploratory research. 1) We investigated and examined the accuracy of morphological information of "conjunctions" of "Balanced Corpus of Contemporary Written Japanese", and clarified the problems and notes of the information. 2) We conducted exploratory analysis on the style and the use of connective expressions, and clarified the following 3 points. a) the characteristics of the use of conjunctions in textbooks in terms of grade and subject, b) digitization of stylistic differences of conjunctions and their usefulness, c) digitization of stylistic differences of complex connective expressions starting with particles and auxiliary verbs and their stylistic features. 3) We examined the problems of the method of surveying complex connective expressions using corpus, and clarified the pros and cons of the use of sentence boundary information.

研究分野：日本語学

キーワード：接続表現 接続詞 複合接続表現 文体 コーパス

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

### 1. 研究開始当初の背景

日本語研究においては、近年、コーパスを用いた研究が一般的になってきており、文章・文体の研究分野においても接続表現研究にコーパスを有効に活用することが考えられた。コーパスを利用した接続表現研究の問題点について未検討の面が多く、その問題点を明らかにするとともに、コーパスを活用して探索的に接続表現研究を行いその可能性を広げる必要があった。

### 2. 研究の目的

コーパスを用いた文章・文体研究分野における接続表現の研究の可能性を広げるために、具体的に、次の3つの課題を行う。

- 1) コーパスを活用した接続表現研究の問題点・留意点を明らかにするために、『現代日本語書き言葉均衡コーパス(BCCWJ)』の形態論情報(品詞「接続詞」の情報)の信頼性の検討を行う。
- 2) コーパスを活用した文体と接続表現の使用に関するいくつかの探索的な調査分析を行う。
- 3) コーパスを活用した複合接続表現の調査方法の問題点の検討を行う。

### 3. 研究の方法

- 1) 『現代日本語書き言葉均衡コーパス(BCCWJ)』の形態論情報(品詞「接続詞」情報)の信頼性の検討

BCCWJの非コアデータを用いて、短単位の品詞情報「接続詞」の解析精度の調査を行う。

- 2) コーパスを活用した文体と接続表現の使用に関するいくつかの探索的な調査分析
  - (a) 教科書の接続詞の使用特徴について、学年別、教科別等の観点から量的・質的分析を行う。
  - (b) 接続詞の文体差を連続的な数値として表す試みを行い、その有用性を検証する。
  - (c) 助詞・助動詞で始まる複合接続表現の文体差を数値化し、各接続表現の特徴を分析する。
- 3) コーパスを活用した複合接続表現の調査方法の問題点の検討

助詞及び助動詞で始まる複合接続表現の用例収集において、BCCWJの文境界情報を用いる場合の得失について調査を行う。

### 4. 研究成果

- 1) 『現代日本語書き言葉均衡コーパス(BCCWJ)』の形態論情報(品詞「接続詞」情報)の信頼性の検討

『現代日本語書き言葉均衡コーパス(BCCWJ)』の短単位の品詞情報「接続詞」(非コアデータ)の解析精度を明らかにするために、使用頻度上位20位までの語を対象として調査を行った。調査の結果、短単位「接続詞」全体の解析精度は高いが、特に「で」(解析精度61.0%)、「唯」(解析精度76.0%)は極めて低く、「又」「更に」「猶」は解析精度80%台で低くなっており、長単位「接続詞」での調査結果とほぼ同様の傾向を示すことが分かった。なお、「又」については、短単位より長単位で解析精度が低く、これは連語「又は」の自動構成に失敗した用例が長単位にある程度含まれていることの影響であることを指摘した。さらに、解析精度が低い5語について、他品詞の同形の語彙素の調査を行い接続詞の使用頻度の推計を行う試みも示した。接続詞研究においてBCCWJの品詞情報の検索結果をそのまま利用する際には、本研究での指摘に充分留意する必要があることが分かった。

- 2) コーパスを活用した文体と接続表現の使用に関するいくつかの探索的な調査分析

- (a) 教科書の接続詞の使用特徴について、学年別、教科別等の観点から量的・質的分析を行う。

『現代日本語書き言葉均衡コーパス(BCCWJ)』の教科書データを利用して調査分析を行い、次の特徴があることが分かった。全体としては、小中高と学年が上がるに連れて、異なり語数、相対頻度ともに増加している。ただし、相対頻度に関しては、国語は小4-6以降は大きな変化はなく、1万語当たり約50語の接続詞が用いられている。学年段階が上がるとともに多様な接続詞が使われる。教科別では、国語が他の教科に比べ極めて多様な接続詞が用いられている。国語(特に小4-6)は他教科に比べて話し言葉的な接続詞が多く使われる。社会は、書き手の論の運び方を示す言わば叙事的な接続詞(「尤も、要するに、其れでは、然し、並びに、更に、一方、猶」など)が使われやすい。数学は、内容自体に存在する因果的な論理関係を表す接続詞(「故に、因って、但し、が、従って、即ち」など)が使われやすい。接続類型別では、大まかにみると、高校では、対比型、補足型、同列型、順接型が相対的に小4-6、中に比べて現れやすい。また、国語は転換型と逆接型、社会は添加型、理科は対比型と同列型、数学は順接型と補足型が相対的に他の教科に比べて現れやすい。国語は、一般的な文章(書籍、雑誌、新聞)に近い接続詞を用いている。一方、数学、理科、社会は一般的な文章に比べて特徴的な接続詞を用いている。特に高校数学では「因って」「故に」などが多用されている。

- (b) 接続詞の文体差を連続的な数値として表す試みを行い、その有用性を検証する。

BCCWJ「図書館サブコーパス」書籍サンプルに対する文体情報のアノテーションデータである『BCCWJ 図書館サブコーパスの文体情報』の硬度、くだけ度等の値を利用して、接続詞の文体差を数値化した。硬度付与対象使用サンプル数100以上の48語の硬度平均値は次の通りとなった。

並びに(1.893)、及び(2.024)、乃至(2.063)、然しながら(2.154)、即ち(2.176)、且つ(2.182)、

猶(2.188)、又は(2.215)、従って(2.236)、若しくは(2.275)、一方(2.314)、更に(2.32)、或いは(2.412)、但し(2.418)、因みに(2.487)、又(2.51)、然も(2.513)、然し(2.526)、尤も(2.541)、所で(2.543)、要するに(2.571)、だが(2.591)、そして(2.606)、所が(2.632)、扱(2.651)、では(2.663)、其れでは(2.671)、唯(2.672)、其れでも(2.717)、が(2.73)、だから(2.775)、其れとも(2.784)、そうして(2.791)、其れから(2.839)、けれど(2.841)、すると(2.842)、で(2.897)、ですが(2.909)、ですから(2.921)、其れで(2.931)、だったら(2.957)、なのに(2.992)、でも(3.008)、じゃ(3.034)、だけれど(3.052)、じゃあ(3.061)、だって(3.116)、けど(3.118)

専門度、客観度、硬度、くだけ度の値を利用した各平均値は相互に強い相関があることが分かった。硬度平均値の有用性を確認するために、各接続詞の先行研究における内省による文体差の段階及び『日本語話し言葉コーパス(CSJ)』の分析結果の数値と、本研究での硬度平均値との相関を調べた結果、それぞれ強い相関がありことが分かり、専門度、客観度、硬度、くだけ度の平均値を文体差の目安として利用することが可能であることが分かった。

なお、この専門度、客観度、硬度、くだけ度の平均値は、接続詞以外の単語の文体差の数値化にも活用できることも分かった。

(c) 助詞・助動詞で始まる複合接続表現の文体差を数値化し、各接続表現の特徴を分析する。

『BCCWJ 図書館サブコーパスの文体情報』の硬度の値を利用して、助詞及び助動詞で始まる複合接続表現の硬度平均値を算出した。この硬度平均値は文体差を表している。助詞及び助動詞で始まる複合接続表現の硬度平均値を、一語の接続詞の硬度平均値と比較して、助詞及び助動詞で始まる複合接続表現の文体的特徴を分析した。その結果、一語の接続詞の文体差の分布と比べると、助動詞で始まる複合接続表現は硬度平均値が高い(文体が硬くない)表現に偏る傾向があるのに対し、助詞で始まる複合接続表現は硬度平均値が低い(文体が硬い)表現から高い(文体が硬くない)表現まであり、文体的な偏りは特に見られないことが分かった。また、助詞で始まる複合接続表現のうち、硬度平均値が相対的に低い(文体が硬い)表現は、機能語としての使用率がある程度高く固定化し定着している複合辞と同じ形式の複合接続表現が多く、この複合辞としての固定化・定着が文体差に影響している可能性があることが分かった。さらに、硬度平均値が相対的に低い(文体が硬い)複合接続表現は、明治・大正期の用例が見られる傾向があることが分かった。

3) コーパスを活用した複合接続表現の調査方法の問題点の検討

助動詞で始まる複合接続表現の用例収集を BCCWJ「中納言」で行う場合に、文境界情報を用いた検索方法と、その情報を用いず空白及び句点(。?! )の情報のみを用いた検索方法とを比較し、それぞれの長短を検討した。文境界情報を用いた検索方法では、閉じ括弧類に後続する実質的用法の助動詞の用例も含まれるため、複合接続表現の用例を収集するためには効率がよくないことが分かった。また、文境界の直後に空白がある場合は効率よく検索できることが分かった。一方、空白及び句点(。?! )の情報のみを用いた検索方法では効率はよいが該当する用例が若干漏れることが分かった。効率だけを重視せずある程度網羅的に用例を収集するためには文境界情報を利用した検索方法が望ましいことが分かった。

## 5. 主な発表論文等

〔雑誌論文〕(計 9 件)

(1)馬場俊臣、BCCWJ の品詞情報の解析精度 短単位「接続詞」の場合、札幌国語研究、21、1-10、北海道教育大学国語国文学会・札幌、2016 年、査読無

(2)馬場俊臣、「か」「こそ」「といて」が付く接続詞及び接続表現 「だからか」「だからこそ」「だからといて」をめぐって、北海道教育大学紀要 人文科学・社会科学編、68-1、13-27、北海道教育大学、2017 年、査読無

(3)馬場俊臣、教科書の接続詞 BCCWJ 教科書データを用いた分析、札幌国語研究、22、1-10、北海道教育大学国語国文学会・札幌、2017 年、査読無

(4)馬場俊臣、接続詞の文体差の計量的分析の試み 『BCCWJ 図書館サブコーパスの文体情報』を用いて、北海道教育大学紀要 人文科学・社会科学編、69-1、1-14、北海道教育大学、2018 年、査読無

(5)馬場俊臣、接続詞の文体差の探索的分析 『BCCWJ 図書館サブコーパスの文体情報』5 指標を用いて、札幌国語研究、23、1-8、北海道教育大学国語国文学会・札幌、2018 年、査読無

(6)馬場俊臣、『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性、言語資源活用ワークショップ発表論文集、3、241-256、国立国語研究所、2018 年、査読有

(7)馬場俊臣、複合接続表現の文体差 助動詞で始まる複合接続表現について、語学文学、57、1-10、北海道教育大学語学文学会、2018 年、査読無

(8)馬場俊臣、BCCWJ 文体情報の各文体指標の特徴語 『BCCWJ 図書館サブコーパスの文体情報』を用いて、北海道教育大学紀要 人文科学・社会科学編、69-2、1-14、北海道教育大学、2019年、査読無

(9)馬場俊臣、助詞・助動詞で始まる複合接続表現の文体差、北海道教育大学紀要 人文科学・社会科学編、70-1、印刷中、北海道教育大学、2019年、査読無

〔学会発表〕(計 2件)

(1)馬場俊臣、教科書の接続詞の特徴 BCCWJ を用いた教科による違いの統計的分析、北海道教育大学語学文学会平成 28 年度学術研究発表会、2016 年 8 月 30 日、北海道教育大学札幌駅前サテライト、査読無

(2)馬場俊臣、『BCCWJ 図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性、言語資源活用ワークショップ 2018、2018 年 9 月 4 日、国立国語研究所、査読有

〔その他〕

ホームページ等

「接続詞関係研究文献一覧」

<http://www.sap.hokkyodai.ac.jp/baba/home/setuzokusi.htm>

## 6. 研究組織

### (1)研究代表者

研究分担者氏名：馬場 俊臣

ローマ字氏名：BABA, Toshiomi

所属研究機関名：北海道教育大学

部局名：教育学部

職名：教授

研究者番号(8桁): 70218668

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。