

レジスター別出現頻度に基づく

順接接続詞の文体差の評価

—現代日本語書き言葉均衡コーパス (BCCWJ) の用例分析から—

高野 愛子 上村 圭介

Evaluation of the stylistic level of resultative conjunctions based on the frequency by register in the BCCWJ

Aiko Takano, Keisuke Kamimura

Abstract

Conjunctions in Japanese are often defined in relation to 'stylistic' features, e.g. formality and politeness. However, it is not clear how these stylistic features are represented in the context of actual writings. As a result, the definitions of conjunctions are often ambiguous and even misleading in practice. The authors attempted to characterise, rather than define, conjunctions in relation to the registers in which they appear, thus giving both JSL learners and teachers criteria based on which they may choose appropriate conjunctions in a given context. The authors examined the use of resultative conjunctions in the Balanced Corpus of Contemporary Written Japanese (BCCWJ), collected a total of 51,121 occurrences across 11 'registers', and applied dependency tests, correspondence and clustering

analysis. They found that the conjunctions show significantly different distributions by register. Conversely, the registers can be divided into “unplanned” vs “planned” writings based on the occurrence pattern of conjunctions. They also found that the conventional definition of ‘yotte’ does not represent its function in actual writing, which demonstrates the validity of the approach proposed in this study.

1. 研究の背景

1.1 学習者・添削者にとっての問題点

レポート・論文などのアカデミック・ライティングにおいて、日本人学生・留学生（日本語学習者、以下学習者）を問わず、論説文の文体として適切ではない表現を用いていることがよく見受けられる。不適切と判断されるものには、「です・ます体」「だ・である体」の混用、会話体の使用など様々あるが、とくに接続詞・副詞における使い分けの不適切さが問題となることが多い。そのなかでも順接接続詞¹⁾は、論理展開を示すものとして重要な表現であり、類義語の種類も多く、使い分けが困難なものである。

以下は、日本語学習者から収集した作文に対して日本語教師による添削を行った誤用タグを付与した学習者作文コーパス「なたね」から、文章表現の教科書において不適切とされている順接接続詞「だから」「それで」「なので」を誤用検索した用例である。（下線は筆者。→の左側が誤用箇所とされ、→右側が訂正例である。その他は原文ママ、括弧内は、コーパスの作文 ID と学習者の母語）

- ① 目上の人には尊敬すべきだ。だから、必ず尊敬語、または尊敬の意がある言葉を使う。（作文 ID：039_a 学習者の母語：中国語）
- ② 家庭には男の子が一人もいないといけないと言うむかしからの考がえかたが基づいていた。なので→そのため、これに対して、政府は「よ

¹⁾ 本論では、「その結果」など接続詞ではないものも含め、順接としての機能をもつ接続詞・接続表現の呼称を「順接接続詞」とする。

く育った女の子1人が、10人の男の子にまさる」というキャッチフレーズを作った。（作文ID:142_c 学習者の母語：韓国語）

- ③ 失職した人々は、生活の基本保証ができなくて、社会の不安定に導くに違いない。それで→だから | したがって、いつ、どこでロボットを使うのか、しっかり考えなければならない。（作文ID：043_a 学習者の母語：中国語）

レポート・論文の順接接続詞としては「そのため」「したがって」「ゆえに」「よって」などが適切だとされているが、上記の用例から、学習者が不適切とされている語を実際に使用していることが観察される。学習者がこのような語を選択する原因は、学習者自身がそれぞれの文体的な使い分けについて認識がないこと、使い分けの必要性を認識していたとしても、類義関係にある語のなかから適切に選択できていないことにある。

一方、教師など添削者側が誤用箇所を指摘するか否か、訂正例をどう示すかは個人の語感によるところが多い。①は、許容か見落としかは不明だが指摘がなされていない。②③は、訂正例として多様な類義語のなかからそれぞれ異なる接続詞が選択されて示されている。③は、訂正例として複数の類義語が示されているうえ、不適切とされている「だから」も含まれている。

このように、学習者は教師側の指摘をそのまま受け取るか、その中からさらに選択をしなければならないことになり、適切に使い分けを判断でき運用できることは難しい。この他に学習者が言葉を選択する手段として、辞書・文法解説書・教科書の記述に頼ることになるが、その文体の記述から適切な語を選ぶことは困難なようである。以下よりその原因について述べる。

1.2 文体の記述に関する問題点

「文体」は日本語教育学会編（2005）『新版日本語教育事典』によると「個性的文体」「類型的文体」とに大別されるが、本論では「語彙・語法」「文章のジャンル」「修辞」の観点から類型的に捉えた特徴を指す「類型的文体」

の意で用いる。

類義語同士の文体的な差異に関する記述は、日本語学習者用の文章表現に関する教科書では初めの課で示される程度である。それも単語の入れ替えの一覧、短文での例示や簡単な練習が中心で、教科書によって示されている項目の種類と数も異なる。順接接続詞の場合、ほとんどの初級教科書・作文教材では「ですから」で統一されており、中級以上の文章表現を目的とする教材では「だから」「それで」は「話し言葉」で、「書き言葉」「論文やレポート」では、「したがって」「そのため」等を使うように指示している。

辞書・文法解説書においても同様で、文体差の特徴に関する記述と例文が少ないうえに、語例や用語などは統一された記述とはなっておらず、抽象的な説明で、「日常会話、会話的、やや口語的、学術的文章、硬い表現、かたい文書、公式の文書、話し言葉、書き言葉、改まった言い方、あまり使われない、よく使われる」などのように、感覚、話しことば／書きことばの対立、頻度など、さまざまな軸が混在した形で記述されている。そのうえ、「あらたまった／くだけた」「かしこまった／くだけた」というような説明は、学習者にとっては、言葉の意味をはじめその感覚がとらえにくい。

辞書、文法解説書、日本語教科書において、順接接続詞を例に文体に関する記述があるものをまとめると以下のようなになる。

表1 文体の記述による順接接続詞の分類

| 軟らかい・話しことば・口語的・日常会話・ 会 | やや硬い・書きことば | 硬い・書きことば・学術的文章・論説文・公文書・論理的な説明・ 文 |
|--|---------------------|--|
| ですから だから それで なのだ で | そのため そこで その結果 | したがって よって (判決文・演説や講演) ゆえに (数学・論理学・哲学) |

1.3 先行研究

単語の文体差に関する先行研究としては、宮島 (1972、1977)、田中

(1999)、井上(2009)、石黒(2011)がある。その文体差は、「俗語—くだけた日常語—無色透明な日常語—あらたまった日常語—文章語」(宮島1977)、「会話的—話しことば的—一般—書きことば的—文語的」(田中1999)、「卑俗体<口語体<汎用体<書記体<文語体」(井上2009)、「話し言葉専用—話し言葉・書き言葉兼用—書き言葉専用」(石黒2011)のように示され、その段階は3分類または5分類となっている。

このように、文体の名称(ラベル)は抽象的で学習者にとって実感しにくい。文体差の記述も、書き言葉(的)=硬い・フォーマル、話し言葉(的)=軟らかい・カジュアルという意味での記述が多いが、書き言葉のなかでも友人同士のメールは軟らかくカジュアルであり、話し言葉のなかでも発表は硬くフォーマルであることから、曖昧である。これについて石黒(2015)は、「話し言葉的、書き言葉的といった一本のものさしの上に並ぶ一次元的な捉え方を、二次元的に捉えなおす」ことが必要であると指摘している。さらに小林他(2016)でも、これらの用語は「汎用的で使いやすい一方で、その用語の指し示す内容がわかりにくい場合がある」と指摘し、「学習者の言語運用を支援するための文法記述は、汎用的な用語や抽象的な用語だけでなく、可視化された言語情報を基に行われる必要がある」と述べている。

接続詞の文体に関する研究には、安藤(2002)、村岡(2002)、清水(2006)、石黒(2009)がある。これらの研究では、接続詞全体を包括的に対象とし、ジャンルによる量的な出現の割合・頻度・順位からその特徴が示されているが、類義関係にある語の文体差は示されていない。接続詞の類義関係に触れているものとして、井上(2009)が「だから<したがって」のように、接続詞の文体的な隔たりの程度を<、<、⊆などの記号によって示し、石黒(2016)が社会科学専門文献の分野別特性を頻度で示している²⁾が、類義関係にある接続詞の文体差を全体的に示したものはない。

²⁾ 商学・経済学・国際政治学・法学・社会学の文献。順接の接続詞は「したがって」「ゆえに」が比較されている。

1.4 研究の目的

そこで、本研究では、類義関係にある順接接続詞を各種レジスター別に分析し、文体的な関係を示すことを目的とする。「現代書き言葉均衡コーパス(BCCWJ)」の用例を計量的に分析することを通し、類義語の一群がどのようなレジスターに特徴的に現われるかを明らかにし、類義語同士の文体的な全体像、相互の関係(近さ・遠さ)を示すことによって、学習者や添削者が適切に語を選択できる判断の資料に寄与することを試みる。

順接接続詞を対象とした理由は、先に述べた通り、学習者がよく用いる不適切なものとして、「だから」「それで」「なので」があること、その代替表現が多数存在することである。これらの中から適切な語を選択する拠り所となるよう、それぞれの語の特徴や関係を出現状況で示すことに意義があると考えた。

2. 研究の方法

2.1 対象とする接続詞の選定

本論では、文体的差異のある順接接続詞の類義表現の一群として、以下のものを調査の対象とした。

だから、ですから、それで、そこで、そのため、そのために、したがって、ゆえに、それゆえ、それゆえに、よって、その結果、結果、結果として、その結果として、このことから、このことより、これより、この結果から、これらの結果から、以上より、以上の結果から、なので、で、それだから、それなので

上記の接続詞は、白川博之監修／庵功雄他著(2001)における接続詞「順接」のうち、順接[原因・理由—帰結]型で挙げられているもの³⁾、アカデミッ

³⁾ 順接接続詞には、「だから」「そのため」等の〈理由／原因—帰結型〉と「すると」「それでは」等の〈条件—帰結型〉があるが、本論では〈理由／原因—帰結型〉の類義表現を扱う。

ク・ジャパニーズ研究会編著（2001）における「順接の接続表現」「帰結」で推奨されているもの、日本語記述文法研究会編（2009）における接続表現「論理的展開を表示する接続表現」のうち「確定条件」＝「後続部が先行部が原因で生じた結果であることを表す」に分類されるもの、石黒他（2009）の接続表現一覧にある順接の用法のもの、小学館『使い方の分かる類語例解辞典』（2004）、三省堂『新明解類語辞典』（2015）の見出語である。さらに、学習者の運用に多く見られる「なので」「で」も追加対象とした。

2.2 データ収集の方法

本論では2.1に示した接続詞の用例を現代日本語書き言葉均衡コーパス（BCCWJ）から収集した。データ収集の段階ではすべてのサブコーパスのすべてのレジスター（出版・雑誌、出版・書籍、出版・新聞、図書館・書籍、特定目的・ブログ、特定目的・ベストセラー、特定目的・教科書、特定目的・広報誌、特定目的・国会会議録、特定目的・知恵袋、特定目的・白書、特定目的・韻文、特定目的・法律）を対象としたが、後述する通り分析の段階では「韻文」と「法律」は除外した。

用例の検索には同コーパスのウェブ検索インタフェース「中納言」を用いた。検索条件の指定にあたっては、長単位における形態素分類が接続詞であるものが文頭に出現する用例を前後それぞれ40文字の文脈とともに取得した。

BCCWJでは形態素情報が短単位と長単位によって与えられているが、このうち長単位には文脈に即した品詞の認定がなされており（小椋・富士池2015）、短単位では単独の接続詞とされていない表現が、長単位では接続詞とされている（「だから」や「したがって」など）。このような理由から本研究では長単位の形態素情報を利用して用例を抽出した。

しかし、今回対象とする語句の中には「その結果」や「そこで」など、長単位においても接続詞として分類されないものがある。また、解析誤りにより接続詞として分類すべきものが他の形態素の組み合わせとして解析されて

いる場合がある⁴⁾。本研究では、機能上接続詞であるものをできる限り網羅的に収集するため、長単位で接続詞とされない表現や、接続詞以外の品詞の組み合わせと解析された用例についても収集した。

抽出された用例に対しては、さらにデータ・クリーニング、すなわち本研究の分析の対象となるもの以外を取り除く作業を行った。クリーニングに際しては、まず明らかなノイズ（「そこで」に対する「あそこで」など）や形式的観点から対象外となる用例（「それで」に対する「それでも」など）を取り除いた。その上で、用例ごとに用法を判断して、意味・機能的に本研究の対象に該当しない用例を取り除いた。例えば、「それで」には順接と添加の用法があるが、添加の用例については分析対象から削除した。

2.3 収集したデータの概要

前節の手順により抽出した順接の接続詞は 51,121 件であった（表 2）。粗頻度でもっとも多く出現したのは、「だから」（14,612 件）であり、「そこで」（10,747 件）、「したがって」（7,605 件）、「ですから」（3,690 件）、「そのため」（3,296 件）が続く。なお、調査対象とした接続詞のうち出現頻度が 0.5% 未満であったもの⁵⁾については、分析の都合上「その他」に併合した。

なお、法律レジスターでは、文と文の論理的な関係を示す順接の接続詞は確認できなかった。また、韻文は書きことばの中でも強い制約を受けたバリエーションであることから、日本語学習者が理解すべき接続詞の一般的な特徴を明らかにする上では適切でないと考え、分析の対象から外すこととした。

⁴⁾ BCCWJ の形態素解析の精度は、自動処理だけを施した「ノンコア」データで 98% 以上であり、人手による修正を施した「コア」データでは 99% 以上の精度とされている。

⁵⁾ このことから、結果として、その結果として、それだから、これより、それなので、以上より、この結果から、これらの結果から、このことより、以上の結果から

表2 レジスター別にみた接続詞の出現頻度

| | 雑誌 | 出版・書籍 | 新聞 | 図書館・書籍 | ブログ | ベストセラー | 教科書 | 広報誌 | 国会会議録 | 知恵袋 | 白書 | 合計 |
|-------|------|-------|-----|--------|------|--------|-----|-----|-------|------|-----|-------|
| だから | 784 | 3902 | 70 | 5173 | 1416 | 794 | 23 | 30 | 1061 | 1339 | 1 | 14593 |
| そこで | 488 | 3105 | 39 | 3252 | 535 | 295 | 107 | 116 | 1838 | 807 | 160 | 10742 |
| したがって | 98 | 3222 | 5 | 2464 | 115 | 134 | 171 | 14 | 888 | 244 | 249 | 7604 |
| ですから | 108 | 801 | 1 | 891 | 101 | 109 | 8 | 15 | 1144 | 509 | 1 | 3688 |
| そのため | 198 | 1363 | 13 | 951 | 153 | 94 | 110 | 77 | 32 | 200 | 105 | 3296 |
| それで | 130 | 587 | 3 | 957 | 153 | 124 | 8 | 1 | 363 | 214 | 2 | 2542 |
| その結果 | 90 | 806 | 22 | 654 | 44 | 53 | 41 | 42 | 83 | 40 | 123 | 1998 |
| なので | 20 | 20 | 0 | 4 | 596 | 0 | 0 | 0 | 1 | 801 | 0 | 1442 |
| そのために | 66 | 438 | 4 | 473 | 49 | 49 | 8 | 14 | 101 | 53 | 7 | 1262 |
| よって | 13 | 284 | 1 | 118 | 64 | 15 | 54 | 3 | 252 | 226 | 3 | 1033 |
| それゆえ | 8 | 389 | 4 | 368 | 11 | 18 | 4 | 1 | 1 | 5 | 0 | 809 |
| 結果 | 25 | 65 | 3 | 42 | 176 | 1 | 15 | 14 | 6 | 66 | 0 | 413 |
| ゆえに | 10 | 157 | 2 | 135 | 29 | 15 | 24 | 0 | 1 | 21 | 0 | 394 |
| で | 16 | 46 | 0 | 69 | 103 | 9 | 0 | 0 | 13 | 68 | 0 | 324 |
| それゆえに | 9 | 118 | 0 | 101 | 9 | 7 | 0 | 0 | 4 | 6 | 1 | 255 |
| その他 | 27 | 264 | 5 | 187 | 41 | 14 | 34 | 6 | 31 | 28 | 49 | 686 |
| 集計 | 2090 | 15567 | 172 | 15839 | 3595 | 1731 | 607 | 333 | 5819 | 4627 | 701 | 51081 |

3. 分析結果

3.1 レジスター別の出現頻度

はじめに、レジスターによる接続詞の出現頻度の偏りを確認するために、クロス集計表に対して χ^2 による独立性の検定を行った。その結果、レジスターと接続詞が独立であるとの帰無仮説は0.1%水準で棄却され(χ^2 : 16701、自由度: 150)、接続詞の出現とレジスターには関連があることが示された。文体を定義する上で接続詞は特徴的な役割をもつと考えられるが、このことが定量的にも示されたことになる。

ところで、BCCWJのレジスターには語数に大きな偏りがある。そのため、それぞれの接続詞がどのレジスターで使われやすいかを粗頻度から判断する

ことはできない。また、比率による単純な比較は、レジスターに関わらず出現しやすい接続詞であるから比率が高いのか、特定のレジスターだけに多く出現するために比率が高いのかを判断することができない。そこで、調整済標準化残差を用いて接続詞とレジスターの関係を検討する（表3）。

総頻度上位の5件の接続詞についてみると、「だから」は、雑誌、新聞、図書館・書籍、ブログ、ベストセラーで多く使われていることが分かる。「したがって」は出版・書籍、教科書、白書に多くみられる。「ですから」が多く使われているのは国会会議録と知恵袋である。「そこで」が多いのは、広報誌と国会会議録であり、それ以外のレジスターでは積極的には使われていないようである。「そのため」は、雑誌、出版・書籍、広報誌、教科書、白書のレジスターに多いと言える。

表3 調整済み標準化残差（総頻度上位5位のみ）

| | 雑誌 | 出版・書籍 | 新聞 | 図書館・書籍 | ブログ | ベストセラー | 教科書 | 広報誌 | 国会会議録 | 知恵袋 | 白書 |
|-------|--------|--------|-------|--------|--------|--------|--------|-------|--------|--------|--------|
| だから | 9.24 | -11.60 | 3.53 | 13.72 | 14.89 | 16.21 | -13.60 | -7.93 | -18.54 | 0.58 | -16.78 |
| そこで | 2.66 | -3.98 | 0.53 | -1.85 | -9.38 | -4.14 | -2.07 | 6.20 | 20.99 | -6.28 | 1.17 |
| したがって | -13.37 | 24.43 | -4.42 | 2.85 | -20.42 | -8.50 | 9.25 | -5.49 | 0.85 | -19.26 | 15.45 |
| ですから | -3.70 | -11.99 | -3.37 | -9.33 | -10.60 | -1.51 | -5.65 | -1.92 | 38.95 | 10.42 | -7.29 |
| そのため | 5.74 | 14.03 | 0.59 | -2.76 | -5.56 | -1.76 | 11.77 | 12.42 | -19.47 | -6.18 | 9.25 |

ここで、それぞれのレジスターについて接続詞の頻度と種類について検討する。接続詞の頻度が多いことが、少ない種類の接続詞が繰り返し使用されていることによるものか、あるいは、多くの種類の接続詞が数多く使用されていることによるのかを検討する。

各レジスターにおける順接接続詞全体の出現頻度⁶⁾には0.1%未満の水準

⁶⁾ 本研究での接続詞は、BCCWJの複数の長単位の組み合わせによるものが含まれるが、接続詞の比率の計算にあたっては、単独形式、複合形式にかかわらず1つの接続詞として計算した。

で偏りが見られる（ χ^2 : 3250.8、自由度: 10）。出版・書籍、教科書、国会会議録は接続詞が多く、新聞、ベストセラー、ブログ、知恵袋、白書、広報誌では接続詞が少ない。また、雑誌および図書館・書籍における接続詞の出現頻度は多いとも少ないとも言えない。

接続詞の種類分析にあたっては、レジスターごとに各接続詞の出現比率の二乗和を1から引いた値を接続詞の多様度⁷⁾と定義し、この値を頻度と比較した（図1）。教科書は接続詞の出現頻度が大きく、接続詞の多様度も高い。一方で、書籍の中でもベストセラー書籍は多様度が低い。つまり、比較的少ない接続詞が繰り返し使われていると言える。また、知恵袋では接続詞の使用は他のレジスターに比べて少ないが、接続詞の多様度は高い。

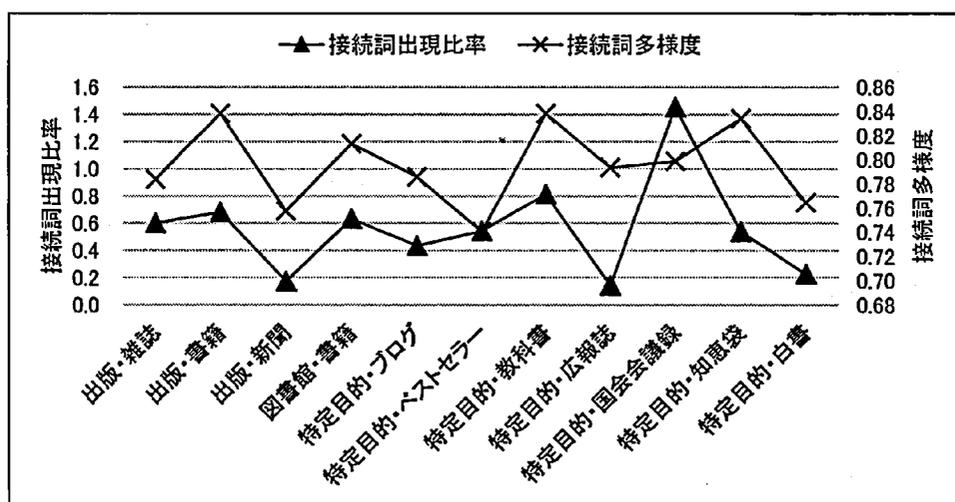


図1 接続詞出現比率と接続詞多様度

興味深いのは、国会会議録における出現状況である。国会会議録では接続詞の出現比率は大きいですが、多様度はそれほど高くない。調整済み標準化残差を見ても、「ですから」と「したがって」の残差が有意に大きい一方で、それ以外のほとんどの接続詞の残差は有意に小さい。つまり、国会会議録では「ですから」、「および」、「したがって」という特定の接続詞が接続詞の出現

⁷⁾ 同様の計算式は、言語学における Greenberg の言語多様性指数や生物学における Simpson の多様度指数に用いられている。

頻度全体を押し上げていることになる。

3.2 レジスターと接続詞の相互関係

次に、接続詞とレジスターの関係、および接続詞同士の関係を分析するために、接続詞とレジスターを二つのカテゴリカル変数とみなして対応分析を行う。

表2の集計表のデータに対して対応分析を行い、第11軸までを求めたところ各軸の寄与率（累積寄与率）は、57.88%（57.88%）、21.01%（78.90%）、12.58%（91.48%）、3.61%（95.09%）、1.95%（97.04%）、1.72%（98.76%）、0.89%（99.65%）、0.19%（99.84%）、0.13%（99.97%）、0.03%（100.00%）、0.00%（100.00%）となった。第1軸から第3軸までの累積寄与率が90%を超えることから、本研究では第1軸から第3軸までの三つの軸の解釈を検討する。第1軸から第3軸に対する行得点および列得点を図2に示す。

レジスターに与えられた列得点を見ると、第1軸では、白書、教科書、出版・書籍、広報誌、図書館・書籍、新聞、国会会議録、ベストセラー、雑誌、知恵袋、ブログが順に並んでいる。また、第2軸では、白書、教科書、広報誌、新聞、ブログ、出版・書籍、雑誌、図書館・書籍、ベストセラー、知恵袋、国会会議録の順である。第3軸では、白書、教科書、広報誌、知恵袋、出版・書籍、国会会議録、ブログ、図書館・書籍、新聞、雑誌、ベストセラーの順である。

両端と中央に注目すると、第1軸は知恵袋およびブログ、白書がそれぞれ両端に位置し、中央には、その他のレジスターが比較的集中している。第2軸は、国会会議録が突出しているが、その反対側には白書、新聞、広報誌、ブログ、教科書が集中している。また、中央には残りのレジスターが集中しているが、その中でも知恵袋は距離が隔てられている。第3軸は、教科書および白書が一端に、もう一端にはベストセラーが位置するが、この軸上の各レジスターの分布は、他の2軸に比べると偏りが少ない。

行得点、すなわち接続詞ごとの関係についてみると、第1軸では「なので」

が一端に位置し、「結果」「で」が中間に、それ以外の接続詞が反対側に集中して位置している。第2軸では「ですから」と「結果」がそれぞれ両端に位置している。第3軸における接続詞の配置は、列得点と同様に他の2軸に比べると偏りが少ない。

第1軸および第2軸については、話しことばに近いレジスター（ブログ、知恵袋、国会会議⁸⁾）や話しことばに見られる接続詞（「なので」、「結果」、「で」）が一端にあることから、この二つの軸のいずれかは、接続詞の使い分けが話しことばと書きことばの使い分けに関連しているものと考えられる。

3.3 レジスターと接続詞のグループ分け

第3軸が、レジスターを文体的にどのように特徴づけるのかを明らかにするために、対応分析の得点を用いて、レジスターに対するクラスター解析を試みた。ここでも累積寄与率が90%を超える第1軸から第3軸までの得点を使用して解析を行った(図3)。解析に際しては、第1軸、第1軸×第2軸、第1軸×第2軸×第3軸の順に一つずつ軸を追加して、クラスターの変化を見た。その結果、第1軸のみによる解析では、ブログおよび知恵袋が一つのクラスターを形成し、その他のレジスターが別の大きなクラスターを形成した。第1軸および第2軸による解析では、国会議事録が単独のクラスターとして独立し、次にブログおよび知恵袋がクラスターとして独立した。三つの軸すべての得点を使用して解析を行ったところ、特定目的サブコーパスの白書、教科書および広報誌が独立したクラスター(①)をまず形成し、ブログおよび知恵袋が第2のクラスター(②)を、国会会議録が単独で第3のクラスター(③)を形成した。そして、それ以外の書きことばによるレジスターが一つのクラスターを形成した。書きことばによるレジスターは、ベストセ

⁸⁾ 国会における議事では予め用意された答弁書や資料を読み上げることも多く、国会会議録は純粋な話しことばとは言えないが、一部の定型的な表現を除けば話しことばとして極端に不自然に感じられるものも少ない。その意味では、国会会議録は話しことばに近いと考えて差し支えないだろう。

ラー、雑誌および図書館・書籍によるクラスター（④）と、出版・書籍、新聞によるクラスター（⑤）にさらに分けられる。

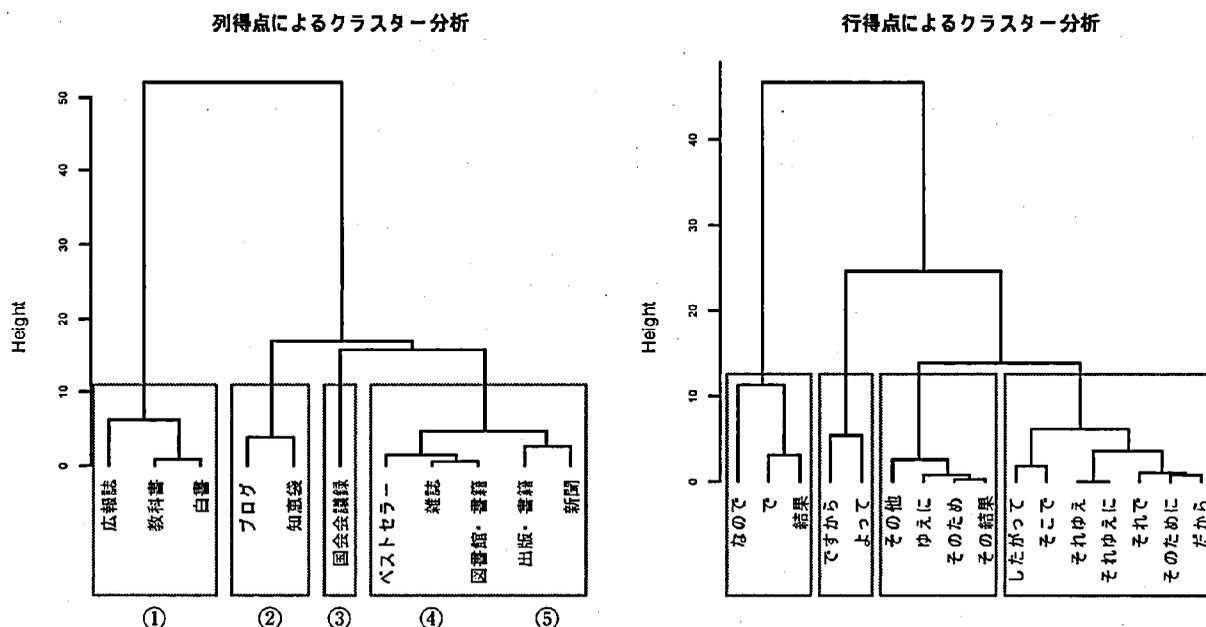


図3 クラスタ分析の結果

樹形図の構造上、第3クラスターは第2クラスターとは併合されず、第4・第5クラスターと一つのクラスターを形成する。しかし、第2クラスター、第3クラスター、および第4・第5クラスターの三つの距離は等距離と言えるほどであることから、第3クラスターと第4・第5クラスターを併合するよりも、これらの三つを独立したクラスターと考えると、それぞれの特徴を検討するほうが意味のある解釈が得られる。このように考えた場合、第2クラスターおよび第3クラスターは話しことば的なクラスターであるという特徴をもつ。これら2つは一つのクラスターにまとめることはできないものの、第4・第5クラスターに対して事前の発話計画（談話計画）がない、あるいは弱いという特徴をもつと言えるだろう。

また、樹形図の高さによる分割に従えば、白書、教科書および広報誌のクラスターを分割するより先に書きことばを二つのクラスターに分けるべきではない。しかし、白書、教科書および広報誌が他と性格の異なったクラス

ターであることを考えると、このクラスターを除き、残りのレジスターがどのようなクラスターを形成するかを検討するほうが意味のある解釈が得られることから、ここでは書きことばをあえて二つのクラスターに分ける。

第1軸および第2軸によって形成されるクラスターが、書きことばとそれ以外のレジスターのグループに分かれることから、前節で検討した通り、これらの軸は、文章の硬さ／軟らかさの次元と、話しことば／書きことばの次元を表していると考えられる。第1軸で切り出されるレジスターがブログ、知恵袋であることから、第1軸は文章の硬さ／軟らかさを、第2軸は話しことば／書きことばを表すものとする。さらに、第3軸を加えることで、特定目的コーパスの白書、教科書および広報誌が独立したクラスターを形成することから、第3軸は、一般的な書きことばではなく、接続詞の選択が固定化された特殊な書きことばを表す次元であると思われる。

同様に接続詞に与えられた行得点を利用したクラスター解析を行うと、「なので」、「で」および「結果」が一つのクラスターを形成し、「ですから」、「よって」がもう一つのクラスターを形成する。「なので」および「で」は、知恵袋とブログに特徴的に多くみられる接続詞であり、列得点によるレジスターのクラスター解析の結果と対応している。また、「ですから」と「よって」は国会会議録と知恵袋に多く見られる（ただし、「よって」は教科書にも多く見られる接続詞である）。さらに、「ゆえに」、「そのため」、「その結果」および「その他」の接続詞のグループは、特定目的サブコーパスの教科書、広報誌および白書に多く見られる接続詞と言える。

これらのことから、第1軸は文章の硬さ／軟らかさの対立を、第2軸は話しことば／書きことばの対立を、そして第3軸は表現形式が固定化された特殊なレジスターであることを、それぞれ示しているものと解釈できる。

3.4 接続詞間の共起関係

前節までの分析で、ある特定のレジスターや文体によって使われやすい接続詞と使われにくい接続詞があることが確かめられた。しかし、BCCWJの

レジスターはかなり広範なグループである。同じレジスターであっても、実際には異なる特徴をもった文章のグループに分けられるはずであり、そのグループでは接続詞の使い分けがあると考えられる。

そこで、本研究では調査対象とした接続詞が同一のサンプルに出現する場合、接続詞が共起したものとみなし、その共起関係を分析する⁹⁾。その上で、レジスター別の分布パターンによる接続詞の類似度と、実際のサンプルにおける共起の強さを比較する。レジスター別の分布パターンが似ていながら、実際のサンプルでの共起の程度が小さければ、「AとBは一緒に使えない」という意味になる。逆に、レジスター別の分布パターンは似ていないのにもかかわらず、共起の程度が強ければ「XはYと一緒に使うことが多い」という意味となる。

レジスターの出現パターンの近さは、接続詞をケースとし、各レジスターでの出現頻度を変数として、接続詞の間のコサイン類似度（1からコサイン距離を引いたもの）で表した。コサイン類似度は0から1までの値を取り、接続詞についてレジスターの出現パターンが似ていれば大きい数値に、出現パターンが違っていれば小さい数値をとる。一方、共起の強さは二つの接続詞の相互情報量によって表した。相互情報量は、二つの接続詞AとBが共起する比率 $p(A, B)$ をそれぞれの接続詞の出現比率 $p(A)$ および $p(B)$ を乗じた値で割った値 $\frac{p(A, B)}{p(A) \times p(B)}$ として定義される。この数値は、1より大きければ積極的な共起が起きていることを示し、1より小さければその二つの接続詞の共起が抑制されていることを示す。

ここでは、両者の数値に大きな開きがあれば、レジスター単位の出現パターンとサンプル単位での共起のパターンが異なると考え、それぞれの接続詞間のコサイン類似度と相互情報量を四分位で分けた上で、相互情報量が上位25%（または下位25%）の値を取りながらコサイン類似度が下位25%（ま

⁹⁾ BCCWJのサンプルは文章全体ではないため、ある文章中に出現する接続詞の共起関係の全体を見ることはできない。それでも、BCCWJの設計上の特徴からある程度の傾向は見いだせると考える。

たは上位25%)の値を取るものを抜き出した。なお、共起頻度が0または極めて小さい組み合わせについては、ここでの比較から除外した。

分析の結果、レジスター単位の出現パターンの類似度が大きいにもかかわらず相互情報量の小さい接続詞の組み合わせ、および類似度が小さいにもかかわらず相互情報量の大きい組み合わせとして表4に示すものが得られた。

表4 類似度と相互情報量が大きく異なる組み合わせ

| 類似度・大、相互情報・小 | 類似度・小、相互情報・大 |
|--------------|--------------|
| そのため/それゆえに | このことから/よって |
| そのため/ゆえに | このことから/結果 |
| そのために/それゆえに | これより/よって |
| その結果/ゆえに | だから/なので |
| | よって/で |
| | ゆえに/よって |
| | よって/結果 |

表5左列の接続詞の組み合わせは、レジスター単位では使われているように見えても、同一のサンプルには出現しにくいことを示している。

表5右列は、レジスター単位の出現パターンが似通っていないのにも関わらず、実際には同一のサンプルに出現しやすい接続詞の組み合わせである。中でも、「結果」と「よって」はレジスター単位の分析では同時に使われにくいとされるが、サンプル単位では、この二つの接続詞は共起する傾向が強い。このように、レジスター単位では共起の傾向が見えないとしても、実際の文章では同時に使われる接続詞の組み合わせがあると考えられる。

このことから、レジスターの中にも文体の異なるいくつかのクラスターが存在し、クラスター内で共起しやすくても、クラスター間では共起しにくい接続詞が存在することが窺える。接続詞の選択にはレジスターの違いが大きく働いているものの、その中にも個別の文体や内容による選択が働いている可能性がある。さらに、これらの選択がレジスターを超えて働いていることも考えられることから、今後はより条件を細かく設定した分析が必要であ

る。

4. 考察

分析の結果から、レジスターと接続詞の関係は、第1軸、第2軸および第3軸の三つの軸によって特徴付けられることが明らかとなった。学習者に接続詞の文体的な使い分けを指導する際、それぞれの接続詞ごとに与えられた従来のような記述ではなく、このようなレジスターの文体的特徴をもつ実際の文章とそこに見られる接続詞を示すことによって、母語で各レジスターに関する背景知識をもつ学習者の語感に訴えることができると考えられ、有効な手段となりそうである。

また、これまでの文体に関する記述と順接接続詞の関連の分類（表1）とは異なる実際の運用のされ方も明らかとなった。とくに、「よって」はブログや知恵袋など、インターネット上での書き言葉にも一定数現れている。これらは個人による自由度の高い書き方であることが共通している点で、また、「なので」がブログや知恵袋で顕著であることから、「よって」が必ずしも「硬い」「学術的文章」のみとは言えない側面をもつことが窺える。

そして、本研究で明らかになったように、話しことば的なクラスターは、発話計画の弱いクラスターとして、書きことば的なクラスターより先に切り出される。これは書きことばの中の接続詞の使い分けよりも、話しことばと書きことばとの間の使い分けの違いのほうが顕著であることを示している。このことを考慮するならば、接続詞の指導において学習者に対してはじめに示すべきは、レジスターによる接続詞の使い分けではなく、書きことばと話しことばの間の違いであると言える。

学習者が日本語を学ぶのは一般的に「会話」としての「話しことば」からである。「書きことば」は「作文」として、レベルに応じて段階的に日記・体験記・感想文・レポート・論文へと移行していくが、早い段階から「話しことば」「書きことば」の全体像とそれぞれの具体的なレジスターの実態を示すことによってその相違も実感できるのではないか。

それをふまえて、教室活動として「作文」を行う際には、「話すように書いてもいい自由度の高いものか」「計画性の高い型の決まったものか」など、学習者側も教師側も「書きことば」「話しことば」それぞれの特性と相違、さらに各レジスターの特徴を認識し、意識して言葉を選択することが求められる。本論で試みた順接接続詞の文体の特徴の示し方は、その意識化を促す役割が果たせるものと期待する。さらにそれぞれの典型的な用例とともに示すことができれば、より適切な言葉を選択できる判断基準となるのではないだろうか。

【引用文献】

- アカデミック・ジャパニーズ研究会編著 (2001) 『大学・大学院 留学生の日本語 4 論文作成編』 アルク
- 安藤淑子 (2002) 「上級レベルの作文指導における接続詞の扱いについて—一文系論文に用いられる接続詞語彙調査を通して—」『日本語教育』 115 号、81-89. 日本語教育学会
- 石黒圭・阿保きみ枝・佐川祥予・中村紗弥子・劉洋 (2009) 「接続表現のジャンル別出現頻度について」『一橋大学留学生センター紀要』 12, 73-85.
- 石黒圭 (2015) 「書き言葉・話し言葉と『硬さ／軟らかさ』—文脈依存性をめぐって—」『日本語学』 34(1), 14-24. 明治書院
- 石黒圭 (2016) 「社会科学専門文献の接続詞の分野別文体特性—分野ごとの論法と接続詞の選択傾向との関係」『日本語文法研究のフロンティア』 161-182. くろしお出版
- 井上次夫 (2009) 「論説文における語の文体の適切性について」『日本語教育』 141 号、57-67. 日本語教育学会
- 小椋秀樹・富士池優美 (2015) 「第 5 章 形態論情報」国立国語研究所コーパス開発センター『『現代日本語書き言葉均衡コーパス』利用の手引 第 1.1 版』 58-99.
- 小林ミナ・小西円・砂川有里子・清水由貴子・奥川育子 (2016) 「第 3 章 類義表現分析の可能性」前川喜久雄監修・砂川有里子編『講座日本語コーパス 5. コーパ

スと日本語教育』65-106. 朝倉書店

清水まさ子（2006）「談話のジャンルと接続表現との関係—新聞の報道文とコラムを比較して—」『東アジア日本語教育・日本文化研究』9, 55-75.

小学館辞書編集部『使い方の分かる類語例解辞典』（2004）小学館

白川博之監修／庵功雄・高梨信乃・中西久実子・山田敏弘（2001）『中上級を教える人のための日本語文法ハンドブック』スリーエーネットワーク

中村明編『新明解類語辞典』（2015）三省堂

日本語記述文法研究会編（2009）「第3章 接続表現」『現代日本語文法7 第12部 談話』（pp.49-141）くろしお出版

村岡貴子（2002）「農学系日本語論文の『結果および考察』における接続表現と文章展開」『専門日本語教育研究』4, 27-34. 専門日本語教育研究会

*本論文では、以下のコーパスを用いた。

国立国語研究所・現代日本語書き言葉均衡コーパス（BCCWJ）<https://chunagon.ninjal.ac.jp>

Hinoki プロジェクト・日本語学習者作文コーパス「なたね」<https://hinoki-project.org>