

令和 3 年 6 月 21 日現在

機関番号：62615

研究種目：基盤研究(C) (一般)

研究期間：2018～2020

課題番号：18K00703

研究課題名(和文) 日本語作文支援システムのための複合機能表現の獲得と用法の分析

研究課題名(英文) Acquisition and Usage Analysis of Compound Functional Expressions for a Japanese Writing Support System

研究代表者

阿辺川 武 (Abekawa, Takeshi)

国立情報学研究所・コンテンツ科学研究系・特任研究員

研究者番号：00431776

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：日本語学習者の接続表現の使用傾向の分析から、接続表現の使い方の問題点を見つけ出し、適切な指導法の提案をおこなった。分析手法として大規模コーパスから作成した接続表現リストを用いて、(1)形態的特徴、(2)指示語との結合の有無、(3)辞書記載の有無という3つの観点から接続表現を分類し、学習者論文や学習者作文を含む5つのコーパスの接続表現の使用傾向を比較した。分析の結果、学習者作文での接続表現の使用頻度が最も高く、接続表現の使用を抑制する指導法が有効であることが明らかになった。

研究成果の学術的意義や社会的意義

本研究では接続表現に焦点を当て、日本語学習者論文や学習者作文を含む5つのコーパスごとに使用傾向を分析した結果、学習者作文では日本語初級から中級で学習する語彙の使用に集中し、母語話者の用いる多様な表現習得に至っていないことを明らかにした。学習者論文と学習者作文では、接続表現の使用傾向には大きく差があり、レジスターの違いが大きな要因であると示唆された。接続表現に限らず、レジスターに焦点を当てた学習者作文の分析は単一のレジスターだけでは表出しない言語現象を見つけられる可能性があり、今回の学習者論文のように、誤用分析のために収集された作文とは異なる種類のレジスターの収集に力を入れていく必要がある。

研究成果の概要(英文)：By analyzing the tendency of Japanese learners to use conjunctive expressions, we found problems in the usage of conjunctive expressions and proposed an appropriate teaching method. Using a list of conjunctive expressions created from a large-scale corpus, we classified conjunctive expressions from three perspectives: (1) morphological characteristics, (2) whether or not they are combined with indicative words, and (3) whether or not they are described in a dictionary. We compared the usage trends of conjunctive expressions in five corpora including learner's papers and learner's essays. As a result of the analysis, it was found that the frequency of use of conjunctive expressions in learner's essays was the highest, and that the teaching method to suppress the use of conjunctive expressions was effective.

研究分野：自然言語処理

キーワード：接続表現 日本語作文支援 アカデミックライティング

1. 研究開始当初の背景

国際交流基金による調査では日本語を学ぶ学習者は2015年は365万人であり、2012年調査の398万人から減少している。これは「語学教育として日本語を教えている学校やその他の機関」の集計数であるが、学習者が減少した国数(55カ国)より増加した国数(86カ国)の方が多いことや、世界的なインターネットの普及を鑑み独学での学習者を考えると、日本語を学習している人の数はこの調査での学習者総数を大きく上回っていると推定されている。つまり、大学や日本語学などの教室で日本語教師から直接教える学習者だけでなく、海外にいながらオンラインの学習教材を用いて自発的に学ぶ学習者が増えていると考えられる。第二言語習得の四技能(読む、聞く、話す、書く)のうち、作文は学習期間後半で習得する傾向にあり、Web上で提供されるサービスを見ても他の技能に比べて作文支援を目的としたサイトは少ない。このような状況で我々は、作文で使用したい語に関する共起表現の検索と例文参照を可能にする作文支援システムや、計量的手法を取り入れることで、理系レポートなどの作文に含まれるレジスター誤りを指摘する作文推敲システムの開発をおこなってきた(<https://hinoki-project.org/>)。

レジスターとは社会的な規範の中で話し手(書き手)、聞き手(読み手)、目的、対象、手段によって発話あるいは文章の表現が異なることをいう。レジスターの例として大学でレポートを書く場合、書き手が学生で読み手が専門知識をもつ教師であることから、口語表現を用いるのではなく専門的な知識を前提とする書き言葉を用いる。レジスターは学習者にとって母語と異なる目標言語において、使い分けに関する知識の不足による誤用が社会言語学の視点からも説明できる。中・上級者でもレジスターの混同から誤用を誘発することが多い。上記のような場面にあった日本語表現をするためには表現豊かな機能語を用いることが鍵といえる。たとえば前述の結果を受けて結論を導くとき、「結果として」や「以上のことから」のような接続表現を使えばより自然な日本語になる。しかし従来の日本語学習者向け教材では、このような複合語から構成される機能表現(以下、複合機能表現と呼ぶ)についてほとんど掲載されていない。さらに辞書を引いても該当する表現に行き当たることができない。そこで我々は文献[1]において、複合機能表現のうち接続表現を網羅的に獲得する手法を提示し、そこから学習者が一連の文字列を辞書項目として選択できる可能性を示した。

2. 研究の目的

本研究の目的は3つの副目的から構成される。(a)日本語学習者にとって習得が困難である接続表現について、仁科らの手法[1]を拡張して網羅的に獲得する手法を確立させること。(b)獲得した接続表現のレジスター別の使用状況を比較することにより、日本語学習者だけでなく日本語母語話者にとっても意識していなかったレジスター特有の使用傾向をあぶり出すこと。(c) (a)で得られた接続表現の網羅的な文字列を辞書項目として整理し、(b)で得られた使用傾向を日本語作文支援システムに組み込み、日本語学習者にとって有用な情報を提示すること、である。

3. 研究の方法

(1) 分析に使用したコーパス

接続表現の出現頻度の集計と分析において、「現代日本語書き言葉均衡コーパス」の一部(BCCWJ*)、「人文・社会科学論文」、「科学技術論文」、「学習者論文」、「学習者作文」の5つのコーパスを使用した。

(2) 本章で対象とする接続表現

一般に接続詞は「要素と要素をつなぐ言葉」と定義され、文と文、句と句、語と語をつなぐ機能を持つ。その中で本稿では、おもに文頭におかれ、前文脈を踏まえて後文脈に来る内容を予告し、読み手の理解を助ける機能を持つ表現に限定する。接続詞は品詞論上の概念であり、他の隣接する品詞(例えば副詞)との境界を定義するには厳密な議論が必要なため、我々は石黒他[2]と同様にこれを「接続表現」と呼称する。接続表現には「しかし」「まず」といった誰もが接続詞と認める表現の他に、「結果として」「だからこそ」のような複数の形態素から構成される複合表現も含める。

(3) 接続表現リストの構築

接続表現を網羅した辞書は存在していないため、我々は独自に文章中で頻出する接続表現を収集した。その収集方法について概説する。(1)で示したコーパスのうち「BCCWJ*」および「科学技術」に対し、電子化辞書UniDicと形態素解析ツールMeCabを用いて形態素解析したものをデータベースとし、次の(ア)から(エ)の作業をおこなう。

(ア)文において文頭から読点までの表現のうち、UniDicの定める語の単位の5単位までの表現を抽出する。(イ)接続表現の対象に該当しない記号、数字、アルファベットなどを排除する。

(ウ)(イ)で示した不要なものを排除後、表現ごとに集計し頻度順に並べ替え上位1,000番目までのリストを作成する。(エ)得られた表現のリストを対象に接続表現に該当するか否かを1項

目ずつ研究代表者ら4名で検討する。検討においては、石黒[3]など接続表現を扱った各種資料を参照するとともに、「人称代名詞」「時を表す表現」「場所を表す表現」「感情を表す表現」「感動詞」「フィラー」「副詞のみの機能と判断される表現」に該当する表現は排除する。

その結果、1,000項目の候補リストから523項目の接続表現リストが得られた。この作業により既存辞書を用いずに高頻度で出現する接続表現を網羅的に収集することができた。

4. 研究成果

(1) 3つの観点による接続表現の分類

日本語学習者が接続表現を使用する際の問題点を分析するにあたって、構築した接続表現リストに対し3つの観点から分類した。

接続表現の形態論的分類

接続表現リストの中には「結果」「結果は」「その結果」「結果として」「その結果として」のように接続表現の中核をなす語基「結果」によってグループを構成する一連の表現がある。このようなグループはニュアンスが微妙に異なるもののグループ内でお互いに置換可能であるものが多い。そこで、形態論的に由来が同じであり、意味が同等な接続表現をグループ化して分類した。

指示語系・非指示語系

接続表現の中には「そうして」「こうして」「そのため」「この結果」のように指示語を含む表現が存在する。指示語単独では接続詞の機能を持たないため、必然的に指示語に助詞、補助動詞、形式名詞など品詞が後接して接続表現を構成する。指示語を含む接続表現において学習者の使用傾向が母語話者の傾向と異なることがわかっていたため、得られた接続表現リストを指示語の有無によって分類する。指示語を含む表現を指示語系接続表現、含まないものを非指示語系接続表現と呼称する。分類の結果、指示語系は242項目、非指示語系は281項目となった。

辞書記載の有無

日本語学習者が未知の表現に出会ったとき、辞書を引きその意味を調べることができるが、今回取得した接続表現は辞書に含まれていない複合接続表現も多いと思われる。複合接続表現については、その表現が広く社会に認知されているかどうかで記載の有無が決まり、認知の度合いによる判定は出版社や辞書編纂者によって異なるため、見出し語の立項は辞書により違う。

そこで、作成した接続表現リストがどの程度辞書に記載されているかを複数の紙の辞典(4辞典)と形態素解析で使用される電子化辞書(3辞書)を用いて調査する。紙の辞典では見出し語として立項されているものだけでなく、語釈文中の追い込みで記述されている表現についても記載有として数え上げる。なお、接続表現の中には接続詞と副詞の境界がはっきりしない表現や、名詞として品詞付けされている表現もあるが、辞書に併記されている品詞は接続詞であるかを考慮せず、単純に見出し語として立項されているかどうかで記載の有無を決定する。

(2) コーパス別の接続表現の使用傾向

5つのコーパスに対しそれぞれ接続表現の出現頻度を集計し、学習者と日本語母語話者の接続表現の使用傾向を比較した。

接続表現全体の使用傾向

使用項目数は基本的にコーパスサイズが大きく内容が多様であるほど大きくなる指標で、「BCCWJ*」が最も多く、次いで「人文社会」が多かった。コーパスサイズでは1.5倍近くある「科学技術」が「人文社会」の項目数よりも少ないということは、明らかに「科学技術」における接続表現の語彙は小さいことがわかる。ただし学習者論文、学習者作文についてはコーパスサイズが小さいため、語彙が小さいとはいえない。

辞書記載の有無

辞書記載数別の各コーパスにおける項目数と出現頻度を集計し、さらに1項目あたりの出現頻度を計算した。どのコーパスでも全ての辞書に記載されている接続表現(記載数7)は特に高頻度であることがわかった。「学習者作文」に着目すると、記載数7だけでなく記載数が少ない表現でも他のコーパスより1項目あたりの頻度が高い。我々の予想では記載数7の基本的な表現のみを多用し、記載数の少ない表現ではむしろ他のコーパスより頻度が低いと想定していたが、記載数の多寡にかかわらず多用していることがわかった。

指示語系接続表現

指示語系接続表現を構成する指示語系列はコソアドのうちソ系とコ系の2系列が大勢を占める。本研究でもソ系とコ系に着目し、コーパス別の指示語系列の項目数と、合計頻度をカウントした。結果、どのコーパスでもソ系の項目数がコ系の項目数より多いことがわかった。一方、合計出現頻度を見ると、「学習者作文」のソ系の頻度が際立って高いことがわかり、「学習者作文」ほどではないが「学習者論文」もソ系がコ系よりも出現頻度が高い傾向にあることがわかった。また、学習者コーパス以外で際立つ特徴として「科学技術」ではソ系よりもコ系の出現頻度が高いことがわかった。

(3) 学習者の問題点と指導法

(2)の分析結果から、学習者コーパスが母語話者コーパスと比較して使用傾向に大きく差があることがわかった。ここではその問題点をさぐり、指導法を提案する。

学習者の問題点

・接続表現の使用頻度が高い

(2)の分析結果より「学習者作文」の使用頻度が他のコーパスより高いが、中でもソ系の接続表現の頻度が高く、さらに接続表現のもつ機能別に調べると「整理・並列」の頻度が高いことがわかる。「整理・並列」の接続表現には頻度順に「そして(548ppm)」「また(340ppm)」「それに(153ppm)」「それから(102ppm)」(ppmは百万字あたりの出現回数)が含まれており、従来の学習者作文の分析と同様の結果である。「整理・並列」の機能は共通点、類似点のある事柄を並べて書くもので、母語話者は接続表現を用いる以外の方法で文の接続関係を示せるが、学習者は安易に接続表現に頼ってしまうことが接続表現の多用につながるのではないかと考えられる。

また、学習者が文章の組み立て方の整理ができず、読み手に接続関係を把握してもらえないか自信がないために接続表現を使用してしまうことも考えられる。文頭の接続表現を多用しない方法の1つとして、短文をつなげ文中で使用できる書き方がある。たとえば、「～である。そして～」という2文を「～であり、かつ～」と1文にできる。実際、「人文社会」は平均文長が約67文字、「学習者作文」は約35文字で「人文社会」はほぼ2倍の文長を持つ。本研究では文頭の接続表現のみを対象としたため現象として捉えられなかったが、文中の接続表現も勘案した上で接続表現の多寡を分析する必要があるだろう。

・接続表現の使用域が狭い

学習者作文コーパスは規模が小さく著者の人数も少ないため、出現する接続表現の項目数は少ない。機能別にみると「理解・換言」のなかでは「つまり(116ppm)」が多く「すなわち(23ppm)」「要するに(20ppm)」が少ない。同様に、「整理・並列」では「そして(548ppm)」「また(340ppm)」「それに(153ppm)」は多いが、他のコーパスで比較的頻度の高い「同時に(7ppm)」「加えて(出現なし)」の頻度が「学習者作文」では低い。

これらのうち「そして」「それから」「それに」「また」「つまり」は、日本語教科書初級から中級までの多くの教科書に出現している表現であり、学習者には馴染みのある使いやすいものと考えられる。学習者は、初級から中級までの日本語教科書の影響が極めて強く、多くの場合、アカデミック・ライティングでの使用語彙を特段に学んでいないということが問題といえる。

・レジスターの影響による使用傾向の違い

「学習者論文」と「学習者作文」では日本語能力レベルがほぼ同等であるにもかかわらず接続表現の使用傾向が異なり、「学習者作文」の方が接続表現を多用する傾向にあることがわかる。その原因には論理的に結論を導き出す論文と自分の意見を記述する作文というレジスターの違いがまず挙げられる。次いで、論文は公開が前提にあり教員のチェックが2、3回入っていることに対し、作文は収集の目的が主に誤用分析にあるため学習者以外のチェックが入っていない、ありのまま文章であることも使用傾向の異なる原因であると考えられる。

学習者への指導法

・接続表現分類の活用

(2)から学習者の語彙が母語話者に比べ少ないことがわかった。より豊富な接続表現の使用のために(1)で用いた分類を使用することが有効である。基本・派生接続表現の分類を利用すれば、意味を調べようとしても辞書に記載のない「派生」に対し、辞書記載率が高い「基本」へ誘導することができる。この接続表現リストを積極的に活用することは、学習者の語彙の広がりにつながるのと期待できるため、現在、我々がインターネット上に公開している共起語検索システム「なつめ」[4]に接続表現リストを組み込むことが考えられる。このシステム内にはコンコーダンス機能もあり、学習者は豊富な例文を検討することで適切な接続表現の使用につながるであろう。

・レジスターに適した用法の確認

アカデミック・ライティングといっても(2)で見てきたように「人文社会」と「科学技術」では使用傾向に大きな差がある。このようにレジスターによる使用傾向の異なりがあることを考えると、学習者がある接続表現を用いようと思うとき、コンコーダンスなどでレジスター別の頻度を比較し、執筆しようとする分野で許容されるかを確認できれば適切な表現が選択できる。また、入力した文章について、指定したレジスターで使用できるかを自動的に計算し指摘するシステムも有用であろう。

・ソ系とコ系の使用分布の差異を用いた指導法

(2)で示したように「科学技術」ではコ系がソ系より高頻度であり、中でも「ここで」が145ppmと突出して多用されていることがその要因の1つである(他のコーパスでは「BCCWJ*」「人文社会」「学習者論文」「学習者作文」の順で10ppm、19ppm、8ppm、14ppm)。また、合計出現頻度においても「科学技術」ではコ系がソ系より多く、「人文社会」でも他のコーパスに比べるとコ系の使用割合が高い。これは、アカデミック・コーパスの特徴である可能性が考えられ、前のコンテキストを後ろにつなぐときに用いる接続表現は、対象が心理的に近いものを指す方法としてコ系が選ばれると思われる。

そこで、アカデミック・ライティングを目標とする学習計画では、上記のことを実現する学習シラバスが必要となる。そのためには文章全体を視野においた接続表現の使用を訓練しなければならない。その方法としては、文献[5]のように具体例を挙げながら、どのようときにソ系

あるいはコ系が使えるかを示していくことも有効である。そのためには「こうして」「そうして」のようなお互いに対応するソ系とコ系のペアの作成と、ペア間の出現傾向の違いをレジスターごとに分析しておく必要がある。

・レジスターを意識させる

学習者論文と学習者作文で接続表現の使用傾向が大きく異なるのは、指導教官のチェックが入っていることが1つの要因ではあるが、必ずしもそれがすべてであるとは思えない。作文のテーマが主に自分の経験や意見であり心に浮かんだ内容を逐次的に記述することが多いのに対し、論文では自分の研究成果を客観的に見つけ、推敲を重ね、論理的な構造で他人にわかりやすく記述することを心がける。

さらに、接続表現の項目数が作文より論文の方が多きことから、執筆者自身がよりフォーマルな文体や語彙を用いようとする意識(態度)が見られる。こうしたアカデミック・ライティングへの意識が接続表現の使用頻度の差異に表出しているのではないだろうか。このような意識を持つことが接続表現の過度な使用の抑制につながると考えられる。レジスターの違いが学習者の文章にどのような影響を及ぼすかについて、大規模な分析をおこなった研究はなく、今後、分析を深めることが期待される。

参考文献

- [1] 仁科喜久子、八木豊、ホドシチェック・ボル、阿辺川武。「作文学習支援システムのための接続表現辞典構築」計量国語学 31(2): pp.160-176 (2017)
- [2] 石黒圭、阿保きみ枝、佐川祥予、中村紗弥子、劉洋。「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: pp.73-85 (2009)
- [3] 石黒圭。「文章は接続詞で決まる」光文社文庫、光文社 (2008)
- [4] 仁科喜久子、吉橋健治、曹紅荃。「作文支援システム『なつめ』における共起表現表示機能と評価」日本語教育方法研究会誌 14(1): pp.44-45 (2007)
- [5] 三枝令子。「文脈指示の「コ」と「ソ」の使い分け」一橋大学留学生センター紀要 1: pp.53-66 (1998)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 阿辺川武, 仁科喜久子, 八木豊, ホドシチェク・ボル	4. 巻 32巻7号
2. 論文標題 日本語接続表現の計量的分析に基づく指導法の提案	5. 発行年 2020年
3. 雑誌名 計量国語学	6. 最初と最後の頁 387-402
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件／うち国際学会 2件）

1. 発表者名 ホドシチェク ボル, 阿辺川 武, 仁科 喜久子, 八木 豊, ベケシュ アンドレイ
2. 発表標題 アカデミックライティング指導のための学術論文コーパスと学習者コーパスの接続表現に見られる特徴分析
3. 学会等名 第11回 日本語実用言語学国際会議（国際学会）
4. 発表年 2020年

1. 発表者名 ホドシチェク ボル
2. 発表標題 日本語共起表現検索システムNatsumeにおけるジャンル別分散表現
3. 学会等名 第31回日本語教育連絡会議
4. 発表年 2018年

1. 発表者名 八木 豊, ホドシチェク ボル, 阿辺川 武
2. 発表標題 多言語母語の日本語学習者横断コーパス（IJAS）のストーリーライティングタスクを用いた接続表現リストの評価
3. 学会等名 国際シンポジウム「新世代の日本語学習」（クロアチア・プーラ大学）（国際学会）
4. 発表年 2018年

〔図書〕 計2件

1. 著者名 Hodoscek Bor, Nishina Kikuko, Yagi Yutaka, Abekawa Takeshi (分担執筆)	4. 発行年 2020年
2. 出版社 Znanstvena založba Filozofske fakultete Univerze v Ljubljani	5. 総ページ数 p169-p192 (241p)
3. 書名 The Japanese Language from an Empirical Perspective: Corpus-based studies and studies on discourse	

1. 著者名 仁科喜久子/ ボル・ホドシチェク/八木豊/阿辺川武(分担執筆)	4. 発行年 2018年
2. 出版社 くろしお出版	5. 総ページ数 p75-p96 (223p)
3. 書名 「コーパスを活用した文章の言語学的分析」(『大学と社会をつなぐライティング教育』村岡貴子/鎌田美千子/仁科喜久子編)	

〔産業財産権〕

〔その他〕

Japanese Language Learning Systems https://hinoki-project.org/

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	ホドシチェク ボル (Hodoscek Bor) (10748768)	大阪大学・言語文化研究科(言語文化専攻)・准教授 (14401)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------