

学生レポートにおける接続表現に関する調査

大友 麻実 松本 章代

東北学院大学 教養学部 情報科学科

1 はじめに

我々が所属する学科では、日本語の「科学的文章」の書き方を学ぶ授業が1年次の必修科目となっている。ここでいう「科学的文章」とは、科学的な事柄についての文章を意味するのではなく、「伝えるべき事柄を正しくわかりやすく読み手に伝える文章」のことである。科学的文章は、事実と意見を区別して正確に伝達するために、簡潔かつ一義的かつ論理的であることが求められる。

学生の文章を添削すると、一度読み直せば気が付きそうな不注意によるミスや、ルールを知ってさえいれば防げる誤りが非常に多いことに気がつく。確かに、自分自身が書いた文章を客観的に見直すことは難しく不備・不具合に気が付きにくいものである。ルールを守った文章が書けるようになるのにも、完全に身に付くまである程度トレーニングが必要である。また、学生の多くは他人に読んでもらうための文章を書くという意識が希薄である、ということも問題の一因である。そこで「自分自身が書いた文章を客観的に見直す手助けをする」「科学的文章のルールを身に付けさせる」ための機能を有し、「推敲・校正の必要性を理解させ、他者に提出する文章は十分に吟味する癖をつけること」を目的とした、「学生自身による推敲・校正を支援する教育システム」を数年前から構築している。本システムは、(1)「校正」を支援する「科学的文章のルールチェック機能」、(2)簡潔性・一義性の観点から「推敲」を支援する「わかりにくい文の指摘・可視化機能」、(3)論理性的の観点から「推敲」を支援する「全体の流れの可視化機能」、の3つの機能から構成される[1]。

現在、(3)に、論理的に問題のある箇所を自動検出し指摘する機能を追加することを目指している[2]。本稿では、文をつなぐ接続表現に着目し、実際の学生レポートと査読付き論文とを比較しながら、論理的な問題が生じやすい箇所を調査する。この結果に基づき、非論理的な接続表現を検出・指摘するための戦略を検討する。

2 論理を導く接続表現の調査

2.1 調査対象の接続表現

学生の書いたレポートには、「よって」「したがって」といった因果関係を表す接続詞や「つまり」「すなわち」などの言い換えを行う接続詞を強引に使用し、論理の飛躍が起こっているケースが散見される。そこで、「学生レポートにおける論理の飛躍は、因果関係を表す接続詞や言い換えを行う接続詞を伴って起こりやすい」と

いう仮説を立て、これらの出現傾向を調査・分析する。本稿では、これらの接続詞をまとめて「論理を導く接続表現」と呼び、調査対象とする。「接続表現」としたのは、文を接続する働きをするのは接続詞だけではなく、文頭の表現によっては、接続詞ではなくても文と文をつなぐ機能を持っている場合がある。そのような機能を持つものを「接続表現」と定義する。接続表現は、以下の条件を満たすものとする。

- 文頭から「名詞、形容詞、形容動詞、記号、人称代名詞のいずれかを含む文節」の直前まで
- 名詞のうち「通り」、「上記(下記)」、「上述(下述)」は例外とする
- 最長3文節

接続表現のうち、各文書で使用されている論理を導く表現を調査の対象として加える。なお、接続詞も文頭に現れるものに限定する。これは今回の調査が「接続詞の直前にある文で述べられている内容を受ける接続詞」を調査対象としたためである。

2.2 調査対象の文書

コーパスとして、学部1年生が書いたレポート154本(以下2011年分を(A)、2012年分を(B))と、情報学系の学会論文誌に掲載され、論文賞を受賞した査読付論文12本(以下(C))を用いる。なお、これらに著者の重複はない。

2.3 論理を導く接続表現の出現頻度

(A)(B)(C)それぞれの総語数(a)、論理を導く接続表現の数(b)、10万語あたりの調整頻度($b/a \times 100,000$)を調査する。その結果、調整頻度は(A)が325.94、(B)が363.26、(C)が186.62となり、学生レポートの方が明らかに多いことがわかった。説明が不十分な状態で結論を急ぎ、論理の飛躍が起こっている可能性があると考えられる。

続いて、各接続表現の頻度を調査し、学生レポートと査読付き論文に用いられている論理を導く接続表現にどのような差があるかを確認する。(A)(B)(C)を変数、論理を導く各接続表現の頻度をケースとして、主成分分析を行う。その結果、固有値は第1主成分が2.20、第2主成分が0.63、寄与率は第1主成分が73.36%、第2主成分が20.96%となった。第2主成分と合わせると累積寄与率は94.33%になることから、この2つの成分で元のデータの多くが説明できているといえる。

第1主成分と第2主成分のパイプロットを図1に示す。まず、変数に着目すると、縦軸において(C)は正の方向、(A)(B)は負の方向を向いており、しかも近接している。このことから、(A)(B)にはかなりの類似性が認められる一方、(C)とは明らかに使用される接続表現に差があることがわかる。次に、ケースに着目すると、横軸の値が大きいものほど出現頻度が高く、「つま

Research on Connective Expressions in Reports Written by Students

Asami OTOMO, Akiyo MATSUMOTO

Department of Information Science, Tohoku Gakuin University

り」や「このように」「よって」などが特に学生に好んで使われる傾向にあることがわかる。

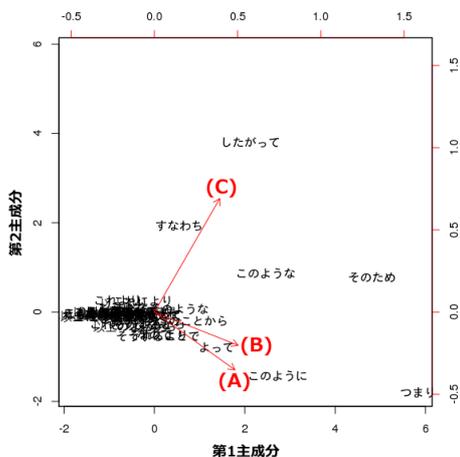


図 1: 主成分分析のバイプロット

2.4 論理性に関する調査

(A)(B)に含まれる各「論理を導く接続表現」についてその前後の文脈の論理性を確認したところ、少なくとも(A)は38件、(B)は48件が非論理的であると認められた。接続表現別に集計すると、86件中、1位が「つまり」で16件、2位が「そのため」で15件を占めた。

今回は「そのため」に着目する。「そのため」の出現箇所について詳細に調査したところ、「そのため」が使われている文の直前には不要な文(要素)が入っていた。さらに、「そのため」を含む文の2文前までには、「～だろう」や「～と思う」といった意見の文が存在していた。これらの特徴をもとに条件を設定して抽出を行うと、それに該当する文が学生レポートから8件抽出された。この8件はすべて上述の調査において、根拠が不十分で論理が飛躍している、あるいは意味が曖昧という理由で、「非論理的」と判断されたものであった。

指示語を含む接続表現は「そのため」の他にもある。しかし、その他の指示語を含む接続表現には前述の特徴を見ることはできない。その理由について考察する。これは、「そのため」に2つの意味があることに起因すると考えられる。ひとつは、接続詞「したがって」などとほぼ同じ意味で、結論を導く働きを持つ。もうひとつは、「～をするために」という意味である。目標や目的を達成するための方法や手段を表現している。「そのため」の2文前までに意見の文が含まれている場合には、このどちらの意味にも捉えられる曖昧さが感じられる。もちろん文脈によって指示語が何を示すかを推測することは可能である。しかし、科学的文章は一義的であるべきである。

2.5 論理を導く接続表現を含む文を構成する語

論理を導く接続表現を含む文を構成する語が、その直前にどのように出現しているかを調査する。論理の飛躍がある場合、論理を導く接続表現の後に続く語句が、その直前に用いられていないのではないかと推測したため、これを検証する。ただし、「つまり」「すなわち」といった言い換えの接続詞は、この前提に該当しないと考え対象から除外する。

今回の検証方法について述べる。論理を導く接続表

現を含む文とその直前の3文前までを対象とし、これらから名詞を抽出する。論理を導く接続表現を含む文を構成する名詞の中で直前の文に出現する語の割合を算出する。査読付き論文内の論理を導く接続表現と、2.4節で述べた非論理的箇所86件とを比較する。

それぞれの平均値を算出したところ、査読付き論文が42.3%であったのに対し、非論理的箇所は31.6%という結果であった。予想通り、査読付き論文の方が明らかに接続表現の前後の語に共通性が高い、という結果となった。今後、両者の違いについてさらに詳細に分析し、論理の飛躍の自動検出につながるような特徴を見つけ出すことを目指す。

3 関連研究

技術文書を対象とした推敲支援ツールは既にいくつか開発されている。菅沼ら[3]は、マニュアルの執筆を想定し、読み手に誤解される文の検出を行っている。我々のシステムが機械学習を用いて「意図が伝わりにくい文」を統計的に判断するのに対し、菅沼らはヒューリスティックな理論に基づき判断を行う仕組みを提案している。

また、稲積ら[4]は大学生の日本語文章力の育成を目的として、校正推敲支援や文章構造理解支援など5種類の支援ツールを開発している。これらのツールに備わっている、技術文書を書く上で順守すべきルールを指摘できる機能や、長文について係り受けの確認と修正を支援する機能は、我々が構築しているシステムの一部と類似している。ただし、本システムは品詞や主語・述語を色・形によって区別し、文章をより視覚的に意識させることができる。また、論理性の支援においては、稲積らのシステムがパラグラフライティングの観点から行われるのに対し、本システムは接続詞や接続助詞の見直しに重点を置いている。

4 まとめ

「学生自身による推敲・校正を支援する教育システム」に、論理的に問題のある箇所を自動検出し指摘する機能を付加するため、学生レポートと査読付論文を対象として論理を導く接続表現の出現傾向を調査した。今後は、非論理的な接続表現の具体的な検出アルゴリズムを考案し、実装する。

謝辞

本研究はJSPS 科研費(若手B, 課題番号24700906)の助成を受けている。

参考文献

- [1] 松本章代: 科学的文章の推敲・校正を支援する教育システムの構築, 東北学院大学教養学部論集, No. 167 (2014).
- [2] 青木大輔, 松本章代, 高橋光一: 学生レポートにおいて不適切な使われ方をしている接続詞の検出, 教育システム情報学会2012年度第5回研究会 (2013).
- [3] 菅沼明, 小野貴博: 文章推敲支援における読み手に誤解される文の抽出, 情処研報2007-DD-61, Vol. 2007, No. 50, pp. 31-38 (2007).
- [4] 稲積宏誠, 大野博之, 竹内純人, 大久保麻里子, 又平恵美子: ICTを活用した日本語文章力育成への取り組み, 情処研報2011-CE-109, Vol. 2011, No. 9, pp. 1-10 (2011).