

# 学生のレポートにおける非論理的接続表現の検出

## Detection of Illogical Connective Expressions in Reports Written by Students

テーマ：インターネット技術とその応用

教養学部 情報科学科

指導教員：松本 章代

1057207 大友 麻実

### 1. 研究背景および目的

伝えるべき事柄を正しく、わかりやすく読み手に伝えるには、簡潔かつ一義的かつ論理的な文章を書くことが必要である。このような文章を書く能力は、学生にとってレポートや論文を書く際に必要不可欠であり、また、身につけなければならないものである。

実際に学生が書いたレポートを読むと、誤字脱字など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付く。また、根拠が十分でないまま結論にいたるケースもある。たしかに自分が書いた文章を客観的に読み、見直すことは難しく不備や不具合には気が付きにくいものである。ルールに基づいた文章を書けるようになるまでにも、完全に身に付けるまでは訓練が必要である。

そこで本研究では、技術文書推敲支援システムというウェブサービスの開発を数年前から行っており、平成22年度からは初年次教育などの授業で実際に学生に利用させている。これは提出前に学生が自分で校正・推敲をすることを支援するものであり、学生の文書作成能力の向上を目的とする教育システムである。本システムには、(1) 文章の基本ルールに沿って書かれているかどうかのチェック機能、(2) 一義的で簡潔性の高い文章に修正するための指摘をする機能、(3) 論理性を高めるための推敲を支援する「全体の流れの可視化機能」の3つが備わっている。現在はこの(3)にさらに不適切な使われ方をしている接続詞を検出する機能を追加するために検討を重ねている[1]。

学生が書いたレポートにおいて論理を導く接続表現は無理な論理展開を導きやすい傾向がある。よって本研究では、論理を導く接続表現(「したがって」「よって」)などを用いて不自然な論理展開を行っている箇所を検出する機能の追加を目指すために、実際の学生のレポートと査読付論文を比較しながら、論理的な問題が生じやすい箇所の調査・分析を行う。

### 2. 論理を導く接続表現の調査

#### 2.1 調査対象

今回調査に使用する文書は、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))である。

#### 2.2 調査対象の接続表現

学生が書いたレポートでは、「したがって」「よって」などの因果関係を表す接続詞や「すなわち」「つまり」などの言い換えの働きをする接続詞を強引に使用して論理が飛躍しているケースが多く見られる。そこで、今回の調査対象とする。本研究ではこれらをまとめて「論理を導く接続表現」と呼ぶことにする。ここで「接続詞」ではなく「接続表現」としたのは、論理を導く働きを持つものは接続詞だけではないためである。各文の文頭にある表現の中にも、それぞれ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」と定義した。接続表現は、以下の3つ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(下記)」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た。なお、接続詞は文頭に現れるもの限定している。これは今回の調査が「接続詞が使われている文より、前にある文で述べられている内容(根拠)を受ける接続詞」を調査対象としたためである。

#### 2.3 論理を導く接続表現の出現傾向

(A)(B)(C)各文書の数と総語数(a)、論理を導く接続表現の数(b)、10万語当たりの調整頻度を算出する。その結果、(A)が325.94、(B)が363.26、(C)が186.62となり、論理を導く接続表現の出現頻度は、学生の方が圧倒的に高いことがわかった。学生が書いたレポートは(C)にある各論文よりも文章量が少ない。それにも関わらず、論理を導く接続表現の頻度が高いということは、十分な根拠を示したり説明をしたりすることが無いまま論理を導く接続表現を用い、論理が飛躍している可能性があることが考えられる。

#### 2.4 論理性に関する調査

(A)(B)に使用されている「論理を導く接続表現」について前後の文脈の論理性の調査を行った。その結果、少なくとも(A)には38件、(B)には48件、非論理的であることが確認できた。86件中1位は「つまり」で16件、2位は「そのため」で15件と、1位と2位はともに全体の2割弱を占めている。

### 3. 論理を導く接続表現の頻度に関する調査

学生が書いたレポートと査読付論文に使用されている論理を導く接続表現に、どのような差や特徴があるかを確かめるために各文書からその頻度に対して主成分分析を行う。この分析によって学生が書いたレポートの特徴をとらえる。

分析を行うにあたり、(A)(B)(C)を変数、論理を導く接続表現の頻度をケースとする。分析の結果、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主成分のバイプロットを図1に示す。まず変数に着目すると、縦軸において(C)が正の方向、(A)と(B)は負の方向を向いていて、しかも近接している。このことから、(A)と(B)はかなり類似性が高いということがわかる。その一方で、(C)とは明らかに使用される接続表現に差があることが見てとれる。

次にケースに着目をする。横軸の値が大きいものほど出現頻度が高いことを表している。学生には特に「つまり」や「このように」、「よって」などが好まれ使われている傾向にあることがわかる。また、(A)(B)と(C)のちょうど中間にある「そのため」は、どの文書

にもまんべんなく頻出しているということがわかる。

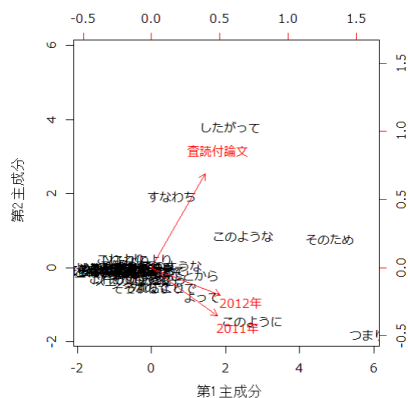


図 1. 主成分分析のバイプロット

#### 4. 接続表現「そのため」

##### 4.1 学生の「そのため」の直前に見られる特徴

主成分分析で「そのため」がどの文書にもまんべんなく頻出していることがわかった。また、2.4節で述べたとおり、(A) (B) には非論理的だと判断されたものは86件あり、少なくともそのうち15件は「そのため」となっている。この15件の「そのため」に注目したところ、「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2文前までには、「～だろう」や「～と思う」などの意見や推測が存在していた。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、学生が書いたレポートから8件抽出された。それらの文はすべて根拠が不十分で論理が飛躍しているか、意味があいまいになっていることが確認できている。

##### 4.2 「そのため」の特性

指示語を含む接続表現は「そのため」の他にもあり、「それにより」「このため」「これらによって」などが例としてあげられる。しかし、その他の指示語を含む接続表現には4.1節で述べた特徴を見ることはできない。これは、「そのため」に二つの意味があることが原因だと考えられる。ひとつは、接続詞「したがって」などとほぼ同じ意味で、前に述べた根拠を受けて結論を導く働きを持っている。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「～だろう」「～と思う」などの意見や推測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ることができる曖昧さが感じられる。

もちろん、文脈によって指示語がなにを示すか、どちらの意味で使用しているかを推測することは可能である。しかし、書き手が意図していることや伝えたいことを読み手に確実に伝えるには文章は一義的であるべきである。

#### 5. 論理を導く接続表現を含む文を構成する名詞

論理を導く接続表現を含む文を構成する名詞が、その直前にどのように出現しているか調査を行った。これは論理の飛躍がある場合、論理を導く接続表現の後に使われている名詞がその直前の文に使用されていないのではないかと推測し、それを検証するためである。なお、「つまり」や「すなわち」といった接続詞は論理を導くだけでなく、言い換えの働きをしている場合があることから、この前提に該当しないと考え、対象か

ら除外している。

この検証では、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てこれらの文から名詞を抽出し、論理を導く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した。(A)と(B)に存在する非論理的であると判断された86件と(C)とを比較する。(C)の平均値が42.3%であったのに対して、非論理的な(A)と(B)は31.6%という結果になった。(C)は10ポイント以上差をつけており、推測のとおり(C)の方が接続表現の前後の名詞に共通性が高いということがわかった。

学生には、同じ語(名詞)のくり返しを避けたいという心理が働くことが考えられる。しかし、前後の文に語のある程度共通性がなければ、新しい話題に変化してしまう。今後はこの共通性がどこにあれば良いのか、また、どの程度あれば良いのかその基準を検討する必要がある。

#### 6. 関連研究

論文の作成能力の育成を目的としたシステムは既に開発されている。稲積ら[2]は文章構成理解支援や校正推敲支援、文章見直し支援など5種のツールを開発している。文章校正理解支援ではパラグラフ・ライティングの考え方をもとに、各段落の意味・関連性、段落の中心文に伝えたい事柄を含んでいるかを確認しながら文章を作成することができる。このシステムは、本システムと一部類似している。しかし、本システムは主語・述語や品詞を色によって区別を行い、語の修飾関係や語間の距離の長さは矢印の色と長さで表す。これにより文章を視覚的に意識することができる。また、稲積らは論理展開において段落の構成に重きを置いているのに対して、本システムは接続詞や接続助詞の観点に立っている。

接続表現についても既に調査が行われている。石黒ら[3]らは、エッセイや論文、新聞の社説など、文書のタイプによって用いられる接続表現が異なることを確かめている。ただし、この「接続表現」は本研究の定義とは異なり、「いわゆる接続詞のことでおもに文頭に立ち、先行文脈をふまえて後続文脈に来る内容を予告し、読み手の理解を助ける表現の総称」としている。

#### 7. まとめ

技術文書推敲支援システムに、論理性に問題がある箇所を検出・指摘する機能を追加するため、学生が書いた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や特徴を調査した。調査した結果、査読付論文と学生が書いたレポートに現れる接続表現の傾向をつかんだ。また、学生が使う「そのため」の非論理的な場合の特徴をとらえることができていた。

今後は非論理的な接続表現の検出のため、具体的な方法・アルゴリズムの考案に着手する。さらに、本システムを実際に導入したことによる効果を検証する必要がある。

#### 参考文献

- [1] 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(2013.01).
- [2] 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, Vol.2011, CE-109, No.9 (2011.03).
- [3] 石黒圭, 阿保きみ枝, 佐川祥子, 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, Vol.12, pp.73-85 (2009.07).