

他文書との比較に基づく学生レポートにおける問題点の解析

Analysis of the problems in students' papers based on comparisons with other papers

テーマ：インターネット技術とその応用

指導教員：松本 章代

教養学部 情報科学科

1157245 土生勇樹

1. 研究背景および目的

論理的な技術文書を作成する能力は、理工系の学生にとって必要不可欠である。低下傾向にある学生の文書作成能力を向上させるためには、文章を指導者に添削してもらうことが有効である。しかし、大人数の学生一人ひとりに添削を行うことは指導者の負担が大きい。また、レポートの完成度があまりに低いものである場合も、指導者は添削に大きな労力を割くことになる。

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・運用を数年前から行っており、初年次教育などの授業で実際に学生に利用させている。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、一義的で簡潔性の高い文章に修正するための指摘や、文書全体を要約して可視化する機能が備わっている。一昨年度からの研究では、これ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]。

昨年の研究 [2] では、論理を導く接続表現（「したがって」「よって」など）を用い不自然な論理展開を行うことを「強引接続」と定義し、これを検出する機能の追加に取り組んだ。学生が書いた文書・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ためである。実際に不適切な接続が指摘されれば、学生は自分が書いた文書の論理性を見直すことができる。本研究では、この研究を引き継ぎ、機能追加の前段階としてのデータをそろえ分析していく。

2. システム概要

技術文書推敲支援システムは、利便性を考慮してウェブアプリケーションとして構築されている。システムは Linux 上で動作し、開発言語には Ruby を使用する。また、日本語の係り受け解析には CaboCha を用いる。本研究による分析結果は最終的にこのシステムに組み込まれることになる。よって本研究も同様の環境で分析を行う必要がある。

3. 他文書との比較に基づく学生レポートの問題点の解析

表 1 は本研究で使用した文書一覧である。(S1) - (S4) は初年次教育の授業における学生レポートである。(M1) は電子情報通信学会論文誌に掲載され、論文賞を受賞した査読付き論文である。(M2) は平成 25 年度教育改革 ICT 戦略大会におけるレポートである。昨年度、大友は、強引接続の傾向・規則性を調査するため、(S1) と (S2) と (M1) を使用した。これに加え本

表 1. 研究で使用した文書一覧

| 記号 | 文書の種類 | 本数 | 総語数 |
|----|---------------|------|---------|
| S1 | 2011 年の学生レポート | 70 本 | 44,212 |
| S2 | 2012 年の学生レポート | 84 本 | 67,394 |
| S3 | 2013 年の学生レポート | 96 本 | 33,689 |
| S4 | 2014 年の学生レポート | 97 本 | 28,640 |
| M1 | 査読付き論文 | 12 本 | 98,975 |
| M2 | 大学教員によるレポート | 89 本 | 126,134 |

研究では、(S3) と (M2) と (S4) を使用する。(M2) は全て大学教員によって書かれている。(S1), (S2), (S3), (S4) は学生レポートであり、それと比較する材料として、(M1), (M2) を使用する。(M1) は性質上とても長い文章となっているため、学生が書いたレポートと比べる際文章量としてかなりの差がでてしまっていた。しかし、本研究で追加した (M2) は文章量としては、学生レポートとあまり差がないため、より比較対象として近いものとなっている。これらのデータを先行研究と同様に接続表現の観点から比較する他に、本研究では文末表現と形容詞の観点からも比較し分析を行う。

3.1 文書の準備

まず先行研究で使用したためすでにあつた (S1), (S2), (M1) のデータはそのまま使用させていただいたので、(S3), (M2), (S4) のデータの準備に取り掛かった。(S3), (S4) の準備にはスキャナーを使用し文字を抽出した。(M2) の準備では、PDF データがあつたので、その PDF のデータを変換ソフトを使いテキスト化して抽出した。これらのデータは抽出の際、一部文字化け等が生じるため、原本と比べながら全てのテキストデータを編集した。

3.2 文末表現、形容詞、接続表現の抽出

(S1), (S2), (S3), (S4), (M1), (M2) のそれぞれのデータから先行研究で使用したプログラムを使用し、文末表現、形容詞、接続表現の抽出を行った。このプログラムは、それぞれのデータから、文末表現、形容詞、接続表現の抽出を行うだけでなく、1つ1つのテキストの総文字数やそれぞれの抽出した接続表現等の頻出数のカウントを行う。これらの抽出データをエクセルでまとめ統計分析ツール R で分析を行っていく。

3.3 主成分分析と対応分析による比較

本研究では、含まれている語句によって、各文書間の特徴を主成分分析および対応分析の結果に基づき解

