

学生レポートの自動添削システムによる学習支援

Study assistance using automatic proofreading system

共同研究メンバー

○出原至道*、中村その子*（○代表、執筆者）

Keywords : writing skill, proofreading, self-directed learning

1. 背景と目的

多摩大学の学生の基本的なリテラシー能力の中で、特に、論理的な文章構成力が低下している。また、論理構成力以前に、形式的な文章表現力を改善する必要のある学生もいる。このため、本来であれば文章の論理・構成に指導の重点を置くべきところにもかかわらず、それ以前の形式的な誤りの指導に教員の資源が割かれているのが現状である。学生は入れ替わるため、教員側から見ると、毎年新入生を迎えるたびに、同じような指導を繰り返す必要がある。一方、学生の立場から見ると、単なる表現上の誤りであっても教員からのフィードバックには時間がかかり、効率的な学習の妨げとなっている。

本研究では、多摩大学の全学生が利用可能な Google Document において、このような形式的な表現の誤りを即時・自動的に指摘するシステムを開発することによって、学生の自発的な学習効果を高めることを目的とする。

これにより、教員側では、これまで形式的な指導に費やされていた資源を、本質的な指導に振り向けることができるようになり、学生の能力向上につながることを期待される。

2. 手法

基本的な文章校正機能としては、Microsoft Word の「校閲」機能や、Google Document の "Spelling and Grammer" チェック機能がある。これらは、単純な漢字や文法上の誤りを指摘するが、論文・レポートとしての適正な形式に従っていることをチェックする能力は低い。

本研究と類似する、文章校正を目的とした代表的なシステムとして、青山学院大学日本語表現法開発プロジェクト¹による Tomarigi² がある。このシステムは、非ネットワーク環境の Windows にアプリケーションをインストールし、起動したアプリケーションに文章を貼りつけて評価することを想定している。機能の拡張は、システムの提供するインタフェースを通じ

* 多摩大学経営情報学部 School of Management and Information Sciences, Tama University

¹ <https://www.pawel.jp/>

² <https://www.pawel.jp/download/tomarigi/>

た追加で実現できるとされるが、一般には詳細が公開されていない。また、公式ページの更新が 2014 年で停止しており、対応システムが Windows 7 と記述されていることから、開発の継続性が懸念される。

これに対し、本研究では、文章構成システムとして RedPen³ を採用した。RedPen は、多摩大学の卒業生である伊藤敬彦氏によって 2014 年に開発が開始された文書フォーマットチェックツールである。日本語・英語・韓国語などの様々な言語への対応と、設定ファイルによる柔軟なルール設定などの特徴を持つ。また、GitHub 上の公開リポジトリで提供されているオープンソースソフトウェアであるため、開発の継続性や柔軟性が高い。現在も、複数の有志が参加して活発な改良が加えられている。たとえば、文書フォーマットとして、Plain Text の他、LaTeX, Wiki, Markdown などがサポートされている。実行環境には、Windows, MacOS, Linux が挙げられている他、ネットワーク上でのサーバ動作が提供されており、ユーザ側に巨大な分析ツールをインストールすることなく、校正を行うことができる。これを利用して、Atom や Visual Studio などのエディタ内から機能を利用するプログラムも作成・公開されている。

本研究では、まず、RedPen を学内サーバで動作させた。テスト用に公開されているサーバを利用することもできたが、このサーバは実行速度が遅く、また、学生のレポートを外部に送信することを避けるべきだと考えた。次に、Google Document 上で RedPen と通信して文章表現をチェックしコメントを表示するシステムを開発した。これによって、学生が簡単に校正結果を確認できるようになった。

今年度は、システムの開発とルール記述の修正、試用を行った。今後、本システムを学生に利用させ、その使用感をアンケートによって評価し、仕様前後の文章を比較することによって、学生の形式的文章表現力の向上を評価する。

3. 成果

3.1 学内サーバの設置と Google Apps Script の開発

RedPen を、学内研究室サーバに設置して(図 2)、開発環境を整えた。これによって、外部に情報を送信することなく校正を行うことができるようになった。また、改良した成果をおお目の RedPen リポジトリに提出できるようになった。

単純なサーバ動作では、ブラウザを起動してサーバに接続し、文章を貼りつけて解析を行う。しかし、これでは、レポート作成時の



図 1 学内サーバ上の RedPen

³ <https://redpen.cc/>

利便性に欠ける。このため、学生側のインタフェースとして、Google Docs でのレポート記述を対象として Google Apps Script によるサーバ接続プログラムを開発した。これによって、レポートを記述している際、その環境を離れたり文章を貼りつけたりすることなく、校正結果を確認できるようになった（図3）。現段階では、それぞれの文書にこのスクリプトを貼り付ける必要があり実用的ではないため、Google の審査を経て一般利用可能なプラグインとして提供する予定である。

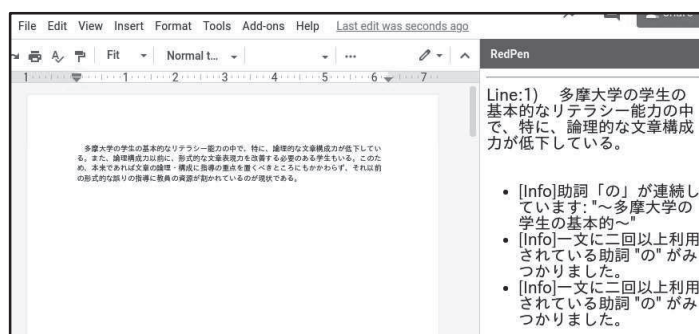


図2 Google Docs 内の校正表示

3.2 文体校正ルールの改良

RedPen の日本語文体チェックルールでは、「ですます」と「である」の混在に関して、「出現回数の多い方を正とし、少ない方にエラーを出す」という動作となっていた。一方で、レポートなどの文体では、出現頻度によらず強制的に「である」を正としたい場合がある。これは、設定ファイルのパラメータを調整するだけでは実現できない。このような機能の変更を、コード内部の修正によって実現できることが、オープンソースソフトウェアの利点である。

この機能を実現するため、RedPen 内部の処理を理解し、修正を行った（図5）。修正提案⁴を発行し、採択された。

3.3 二重否定の結果表示の改良

二重否定のチェック結果が、一般のユーザにはわかりにくい表示となっていたため、修正を行った（図6）。修正提案⁵を発行し、採択された。

```

71 75 @Override
72 76 public void validate(Sentence sentence) {
73 -     if (dearuCount > desumasuCount) {
74 +     boolean forceDearu = getBoolean("ForceDearu");
75 +     if (dearuCount > desumasuCount || forceDearu ) {
76 80         detectPattern(sentence, DESUMASU_PATTERN);
77 81     } else {
78 82         detectPattern(sentence, DEARU_PATTERN);

```

図3 文体チェックアルゴリズムの修正

3.4 試験的な学生のレポートの評価

学生から提出されたレポートを RedPen で評価し集計したところ、14 件のレポートに対して、次のように分析された（図7）。このうち、「口語的表現」には、一般に「ら抜き言葉」と呼ばれる可能表現が含まれて



図4 二重否定エラーメッセージの改良

⁴ <https://github.com/redpen-cc/redpen/pull/863>

⁵ <https://github.com/redpen-cc/redpen/pull/878>

いるが、検出精度が低く、多くの誤検出が含まれている。また、メッセージとして表示される文章が、学生に理解されにくいと懸念される。

4. 課題

本研究の成果は、(1) オープンソースソフトウェアである RedPen の開発に参加しシステムの改良に貢献したこと (2) 実用化に向けて学内サーバ環境を整備したこと (3) サーバに Google Docs から接続して結果を表示するシステムを開発したこと (4) 学生のレポートを試験的に評価したことである。

課題として、(1) 「ら抜き」言葉の検出精度の向上 (2) 学生にわかりやすい校正結果の表記 (3) 一般に利用可能なプラグインの開発 (4) 大学生のレポートに求められる新しいルールの記述 (5) システムを利用することによる教育的効果の評価がある。これらの課題については、次年度の目標として研究を継続する。

内容	人数
一文中の助詞の繰り返し	14
曖昧な修飾	12
半角アルファベット後のスペース	2
不適切な文字	1
半角アルファベット前のスペース	1
表記のゆらぎ	14
単語数が多い	14
非常用漢字	11
カタカナ語の語尾長音	8
不適切な助数詞	5
長すぎる熟語	5
単語の繰り返し	10
不適切な表現	8
多すぎる読点	5
長すぎる段落	4
段落数が多い	3
単語の繰り返し	1
長すぎる文	3
接続助詞の繰り返し	2
非正則送り仮名	1
長過ぎる () 部分	1
口語的表現	14
文末の混在	9
不適切な段落開始文字	5

図 5 校正結果