

文章推敲支援システムのための文章修正環境の構築

米澤 光明[†] 丸山 広[†] 中村 太一[†]

東京工科大学大学院 バイオ・情報メディア研究科[†]

1 はじめに

論理的でわかりやすい文章を書く能力(以下, 文章作成能力)を獲得するには, 作成した文章を繰り返し推敲することが有効である[1]. 自ら文章を見直すと客観的に推敲を行うことが難しく, 他方, 第三者に依頼すると, 繰り返すことは難しい. このような問題に対処する方法として, コンピュータによる推敲支援の研究が行われている[1][2][3][4].

筆者らは, 文章校正手法[2]を実装した「校正伝」と分かりにくい文の改善を支援する手法[1]を実装したシステム「推敲伝」を開発した. これらのシステムは, 利用者の環境に依存せず使えるように, Web アプリケーションとして構築している. 校正伝と推敲伝を 2 年次の学生実験に供してきた. 学生は, これらのシステムを利用し, 指摘を受けて実験レポートの文章を見直し, 修正する. 実験レポートは Word で作成することを求めているため, 学生はシステムに Word ファイルをアップロードし, 指摘を受けて Word ファイル上で文章を修正する. 修正した文章に対する指摘を得るためには, レポートを再アップロードする必要があるため, 平均 1337 文字のレポートに対し, 平均 91.9 分の時間を修正に要している.

推敲を繰り返すためには, 指摘を受けるための作業にかかる時間や手間(以下, 利用コスト)を削減することが求められる.

本研究では, 校正伝と推敲伝から指摘を得られ, それを参照しながら文章を修正し, 一操作で再度指摘を得ることのできるシステムを構築し, 利用コストを削減することを目的とする.

2 関連研究

大野らは, テキストエディタ上で推敲支援を受けながら文章を修正できる支援ツールを開発した[5]. このツールでは, 誤字脱字の指摘は表示するが, 係り受け関係に着目した指摘は, 係り受け関係を提示するのみで, 文章作成能力が低いと, 間違いに気付くことが難しい.

Correcting Sentences Environment for Polish Support System
[†]The Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

3 提案システム

3.1 文章修正フォーム

図 1 に示す文章修正フォームは, 「(1) エディットエリア」, 「(2) 指摘エリア」, 「(3) 再添削ボタン」から成る. 利用者が Word ファイルをアップロードすると, エディットエリアに Word ファイル内のテキストが挿入され, 必須修正箇所を赤文字で, 任意修正箇所を緑文字で表示する. 指摘表示エリアは, エディットエリア上の指摘がある文をクリックされた時, その文の指摘文字, 指摘理由, 改善方法を表示する. これらの機能より, 指摘理由と改善方法を閲覧しながら修正が可能である. さらに, 再添削ボタンを押すことで, 一操作で修正した文章に対する指摘を得られる.

3.2 機能構成

文章修正フォームを実現するための機能構成を図 2 に示す. サーバには以下の機能を備える.

(1) 指摘取得

校正伝と推敲伝から指摘結果を取得する.

(2) 指摘解析

(2)で得られた指摘を解析し, 図 3 の形式に統一する. 文情報は, 入力された文章の何文目かを示す文番号, その文, その文に対する複数の指摘情報を持ち, 指摘情報は, 指摘対象文字列開始位置と文字数, 必須修正か否か, 指摘理由, 改善方法を持つ.

(3) 指摘統合

複数の推敲支援システムから得られた指摘を, 文番号を元に統合する.

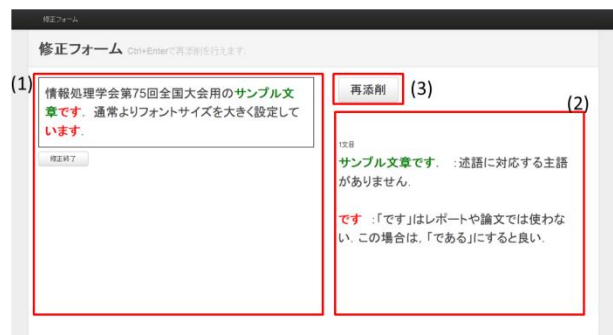


図 1 文章修正フォーム

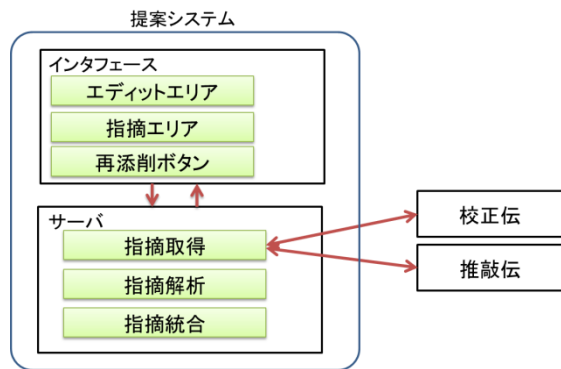


図2 機能構成

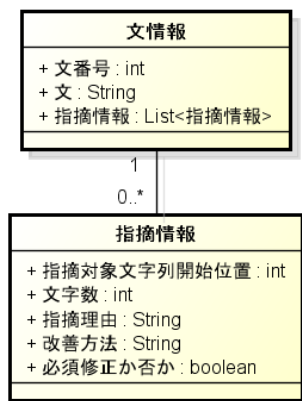


図3 指摘情報の形式

4 評価

4.1 評価方法

従来システムを用いた 2011 年度および提案システムを用いた 2012 年度の学生実験の第 1 回レポートの利用ログそれぞれ 1781 件, 2785 件を用い, 2 回以上利用した利用者の修正時間を分析し, 比較する. 次に該当するデータは修正していないとみなし, 修正時間の分析対象から除外した.

- (1) 修正時間が 3 分未満のもの
- (2) 利用間隔が 100 分以上のもの

除外後のデータ件数は, それぞれ 1711 件, 2302 件で, 利用者数は, それぞれ 114 名, 98 名である. 修正時間とは, 初めて指摘を得た時刻から最後に指摘を得た時刻までの差である.

4.2 評価結果

従来システムと提案システムの修正時間の分布を図 4 に示す. 図 4 は, 横軸に修正時間を, 縦軸に人数を示している. 修正時間の平均が 91.9 分から 19.8 分に減少している. また, t 検定より, 有意水準 0.05 で従来システムより提案システムの方が有意であり, 提案システムは有効であるといえる.

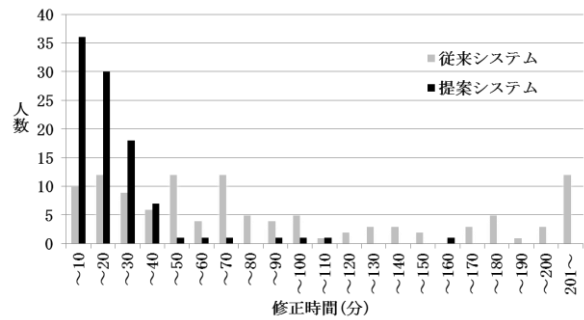


図4 修正時間の分布

5 考察

システムの利用回数が一人あたり平均 14.8 回から 23.7 回に増加した. 一操作で修正した文章に対する指摘を得られたことにより, 文章修正時間を短縮できるだけでなく, 推敲支援を受けやすくなったことで文章を推敲する回数が増え, 学生の文章作成能力の向上に寄与したと考える.

6 おわりに

本稿では, 2 つの推敲支援を受けられ, 指摘を参照しながら文章を修正できる文章修正環境を構築し, その有効性を確認した. また, 提案システムは, 推敲作業に専念でき, 即座に修正結果の確認が行なえる環境を実現できたといえる. 今後, 適合率が低い指摘および修正率が低い指摘を分析し, 指摘のアルゴリズムや指摘理由を見直すことで, 文章作成能力の向上を目指す.

参考文献

- [1] 須藤崇志, 丸山広, 中村太一, "文を分かりにくくする要因の分析と改善支援手法の提案", 電子情報通信学会技術研究報告, KBSE, 知能ソフトウェア工学 108(65), pp. 41-46 (2008-05-22)
- [2] 坂本俊介, 須藤崇志, 丸山広, 中村太一, "形態素解析を利用した文章校正手法の提案", 電気学会研究会資料. IIS, 産業システム情報化研究会 2009(50), pp. 99-104 (2009-07-30)
- [3] 林良彦, "技術文章向けの日本文推敲支援システムの実現と評価", 電子情報通信学会論文誌, D-II, 情報・システム, II-情報処理 J77-D-2(6), pp. 1124-1134 (1994-06-25)
- [4] 菅沼明, 山村広臣, 牛島和夫, "日本語文における名詞句の並列構造の推定およびその推敲支援への適用", 情報処理学会論文誌 38(7), pp. 1296-1307 (1997-07-15)
- [5] 大野博之, 稲積宏誠, "文の接続関係を利用した論理方向の可視化による技術文章作成支援", 電子情報通信学会技術研究報告, ET, 教育工学 107(48), pp. 27-32 (2007-05-12)