

可読性向上のための結束性指摘システムの提案

児玉 哲哉[†]田嶋 克向[†]澤野 弘明[†][†]愛知工業大学

1 はじめに

執筆経験が乏しい学生が書く論文は、誤字脱字や表記揺れといった文章作成の基本的なルールが守られていない文章や可読性が低い文章が多い [1]. 誤字脱字は、文章作成ソフト付属の文章校正機能により添削される。この文章校正機能では、単文の指摘に留まり、複数の文から構成される文章の指摘には対応していない。文章の表記揺れ、例えば「サーバ」・「サーバー」などの指摘に対応した自動文書検査ツール RedPen [2] が伊藤によって提案されている。RedPen による指摘箇所を修正すれば、文章作成の基本的なルールが守られた文章となるが、文章の本質的な可読性自体は向上されない。さて、文章の可読性の指標の一つに、二つの文のつながりを示す結束性 [3] と呼ばれる評価指標が用いられる。結束性のある二文が多い文章は可読性が高く、結束性のある二文が少ない文章は可読性が低い。

そこで本研究では、可読性を向上させるために、結束性がない二文を指摘するシステムを提案する。提案システムでは、まず入力された文章内の前後二文を比較して結束性の有無を判定する。つぎに結束性がないと判定された二文を指摘し、改善を促すことで、可読性の向上を支援する。本稿では、結束性判定手法および提案システムの概要と実験について述べる。

2 結束性判定手法

2.1 結束性

結束性には、文法上の結束性と語彙上の結束性がある。文法上の結束性は、前後二文で誤解釈のない代名詞が使用されているときに意味的につながる性質である。日本語では代名詞が省略されることもあり、代名詞による誤解釈の有無を正確には判定することができない。そのため、本研究では文法上の結束性は結束性判定の対象外とする。語彙上の結束性は、前後二文で共通の単語を持つときに意味的につながる性質である。語彙上の結束性がある二文とない二文の例を図 1 に示す。図 1 の一文目と二文目は、共通の単語を持たないため結束性がない。二文目と三文目は、共通の単語「み

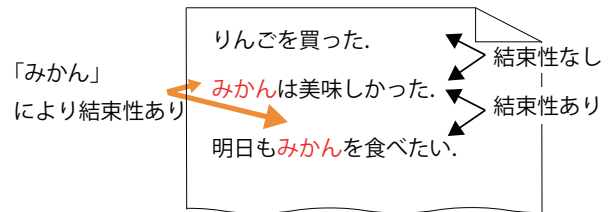


図 1: 語彙上の結束性がある二文とない二文の例

かん」を持つため結束性がある。このような語彙上の結束性の判定ならば、前後二文が持つ単語の比較により機械的に判定することができる。そこで本研究では、語彙上の結束性に着目して、前後二文が持つ単語を抽出し、抽出した単語を比較して結束性の有無を判定する。以降、本稿における結束性は語彙上の結束性を指す。

2.2 結束性判定手法

以下に結束性を判定するアルゴリズムを示す。

1. 連続した文章を入力する。
2. 文章を形態素解析により形態素に分割する。
3. 文章から得られた形態素のうち、名詞と動詞の単語のみを抽出する。このとき動詞の単語を終止形に整形する。
4. 前後二文の各単語を比較する。同一単語が存在した場合に結束性ありと判定する。

2.3 学術論文における結束性の調査

結束性判定手法が学術論文の可読性の評価に有効であるかを確かめるために、学術論文における結束性を調査した。調査には、A 大学卒業論文の 2 ページ概要書を用いた。添削による結束性の差を明らかにするために、卒業論文の概要書には指導教官による添削前後の概要書をそれぞれ 10 本用意した。また、比較対象として情報処理学会論文誌 (原著論文) 10 本も用意した。提案手法を用いて結束性の有無を測定した結果を表 1 に示す。表 1 における節タイトルを筆者の主観により、それぞれの節の内容から四つに分類した。表 1 に示すように、添削前より添削後の学生の概要書の方が結束性が高い傾向にあった。また、学生の添削後概要書の結束性は原著論文と同様に概ね 87% であった。これは、指導教官の添削により結束性が高まったためである。このような結果から、結束性判定手法が学術論文の可読

Proposal of a Cohesion Indication System for Readability Improvement
[†]Tetsuya Kodama [†]Katsuhisa Tashima [†]Hiroaki Sawano
 Aichi Institute of Technology

表 1: 学生論文と原著論文の結束性の比較

節	添削前(学生)	添削後(学生)	原著論文
はじめに	0.92(121/132)	0.92(122/133)	0.90(0,179/0,198)
提案手法	0.89(134/151)	0.90(122/136)	0.87(0,362/0,415)
実験	0.76(119/157)	0.81(134/165)	0.85(0,501/0,592)
おわりに	0.83(033/044)	0.92(034/037)	0.90(0,081/0,090)
合計	0.84(407/484)	0.87(412/471)	0.87(1,123/1,295)

結束性がある二文の割合 (結束性がある二文の数/検証数)

性を評価する手法として有効であることが示唆された。

3 結束性指摘システム

結束性判定手法が、可読性の評価に有効なことが示唆されたため、結束性がない二文を指摘するシステム(結束性指摘システム)を提案する。結束性指摘システムには、クラウド上でファイルを共有できる Dropbox を利用する。結束性指摘システムの概要を図2に示す。まず、図2に示すように、システム利用者が Dropbox フォルダにファイルを追加する。ファイルの追加後に、システムが結束性判定を実行するために、Dropbox フォルダへのファイルの追加を通知させる。ここで、Webhook の API である、Dropbox API v2 により Dropbox フォルダへのファイルの追加を通知し、あらかじめ指定した URL に HTTP リクエストを送る。HTTP リクエストを Web サーバが受け取ると、結束性判定を実行する。結束性判定の結果、結束性がないとされた二文に対する指摘をファイルの文末に書き込む。ファイルへ指摘内容を書き込んだ後、Web サーバは指摘済みのファイルを Dropbox フィルダに追加する。結束性指摘システムによる結束性がない二文の指摘の例を下記に示す。

「背景」節の一文目と二文目に結束性がありません。
一文目: リンゴを買った。
二文目: みかんは美味しかった。

上記のように、「背景」といった学術論文の節タイトル、節タイトル内の文の順序、指摘対象の二文を提示する。

4 システムのユーザビリティ評価実験

論文執筆中の A 大学の学生 6 人に対して結束性指摘システムによるユーザビリティ評価を実施した。評価には、定量的にユーザビリティ評価を行える SUS [4] を用いた。SUS では、5 段階評価による 10 項目の質問を得点化して評価する。SUS の標準平均点 (68 点) を基準の C ランクとして、10 項目の質問の合計得点 (100 点満点) により、システムが A ランクから F ランクに分類される。SUS 評価の結果、被験者 6 人の平均点は

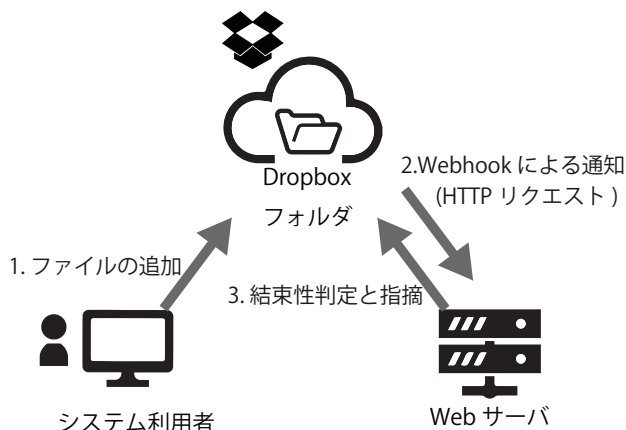


図 2: 結束性指摘システムの概要

75.4 点となり、B ランクの評価が得られた。この結果から、結束性指摘システムは評価者にとって基準より高いシステムであることが示された。また、本システムは結束性のない二文の指摘に留まっており、二文の修正を促すのみであるため、システム利用者は二文の修正方法を考える必要がある。そのため今後の課題として、結束性がない二文への修正案の提示が挙げられる。

5 おわりに

本稿では、可読性の向上を支援するために結束性がない二文を指摘する手法を提案した。まず、結束性判定手法を用いて、学術論文における結束性の傾向を調査した。調査の結果、添削前より添削後の学生概要書の結束性が高く、添削後の概要書と原著論文の結束性が同様の傾向を持つことが確認された。結束性判定手法に基づき、結束性指摘システムを構築し、システムの実験を実施した。SUS によるユーザビリティ評価実験の結果、標準平均点を上回る 75.4 点となり、B ランクの評価が得られた。今後の課題には結束性を生むような二文への修正案の提示手法の検討が挙げられる。

参考文献

[1] 中島利勝, 塚本真也: “知的な科学・技術文章の書き方”, 株式会社コロナ社, pp. i (1996)

[2] 伊藤敬彦: “自動文書検査ツール RedPen”, 信学技報, Vol. 114, No. 211, pp. 69-74 (2014)

[3] 池上嘉彦: “テキストとテキストの構造”, 国立国語研究所, 談話の研究と教育 1, pp. 7-42, (1983)

[4] J. Break: “SUS: A Retrospective”, *J. of Usability Studies*, Vol. 8, Issue. 2, pp. 29-40 (2013)