

ワードファミリーを活用した 例示型英作文支援 Web アプリケーション開発

中村 亮子[†] 宮崎 佳典[‡] 中野 愛実^{*} 田中 省作^{**}

静岡大学情報学部情報科学科[†] 静岡大学大学院情報学領域[‡]

静岡大学総合科学技術研究科情報学専攻^{*} 立命館大学文学部^{**}

1. はじめに

今日の技術系の分野では英語での情報発信が一般的となっている。そして技術文書を作成する際に専門家によって書かれた技術文書を参考にするという方法がある。我々はこのような作業を支援する Web アプリケーションの構築を目指している [1]。

我々が開発中のシステム EWSS では技術文献コーパスの例文約 29 万件から検索機能を用いて学習者の入力する英文と類似する例文を提示している。学習者は提示例文を参考にして技術文書作成を行う。EWSS の問題点としては、類似する例文が提示される一方で、時に似通った例文が必要以上に上位にランクされ、多様性に富んでいなかったり、情報量に欠けたりすることがあった。これに対して我々はワードファミリーの概念を導入することでその解決策として与える試みを行い、その小実験の結果を示す。

2. 関連研究

本研究では例文提示を用いた英作文支援を行っているため例文提示形式の類似研究を紹介する。

高松ら [2] は論文執筆等の支援を目的としコーパスから文を検索、検索文や検索文数を閲覧し表現の適切性等の確認を行う用例検索システムを開発した。また陸ら [3] は第二言語での作文学習支援システムの構築を行った。このシステムでは句共起関係コーパスを利用している。

3. [1] の検索機能の概要

検索機能には類似文検索とフレーズ検索がある。本発表では類似文検索に焦点を当てその概要を説明する。まず形態素解析器 TreeTagger [4] を用いて入力文・コーパス文中に含まれる単語の品詞、原形情報を抽出する。次に英文の特徴となる要素を生成するため単語列の n-gram を生成する（本研究では $n=1, 2, 3$ とする）。全ての n-gram の中で機能語のみ

Development of a Web Application to Support Example-Based English Composition Using Word Family

[†] Ryoko Nakamura, Faculty of Informatics, Shizuoka Univ.

[‡] Yoshinori Miyazaki, College of Informatics, Shizuoka Univ.

^{*} Megumi Nakano, Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka Univ.

^{**} Shosaku Tanaka, College of Letters, Ritsumeikan Univ.

から構成される組み合わせを排除する。論文における言い換えに関する書籍からの抜粋や英語の概念辞書である WordNet [5] を利用し、単語に同義語グループ ID を割り当てる。このようにして作成した特徴集合を用いて英文間の類似度をコサイン類似度を用いて計測し、降順にソートして英文を提示する。

4. ワードファミリー導入

4.1 目的

本研究では従来の類似文検索に対し、新たにワードファミリー（見出し語とその変化形及び密接に関連する派生語からなる単語群。以下 WF）の概念を加えてアップデートを行う（以下、提案手法）。WF 導入により、従来のシステムでは考慮されなかった品詞の異なる派生表現を使用した例文に新たに加点がなされ、従来に増して多様な例文が提示されるシステムになることを目指す。

4.2 ワードファミリー導入方法

本研究では Nation [6] が提供する WF リストを利用する。事前に関リスト内の WF グループに ID を振り（以下 WF-ID）、本システムで使用するコーパス内文章と学習者による入力文に対し、各単語に同義語グループ ID 割り当てに加えて WF-ID 割り当てを行う。また、WF グループを追加することで新たに類似文として上位に認識されるよう顕在化させるため、今回は WF グループの一致に対し、適宜重み付けを行うこととした。

4.3 実装結果

WF グループを導入し、類似文検索機能において図 1 のように実装した。



図 1：提案手法実装後の検索結果画面

検索時に入力された文章が上部に表示され、同義

語グループ・WF グループ一覧が左側の同義語・派生語グループに表示される。また同義語グループ単語は黒色、WF グループ単語は青色の文字で表示される。検索結果の類似例文上位 50 位が画面中央に表示され、類似文検索 1、類似文検索 2 の部分のタブを選択することで従来の検索結果(類似文検索 1)と提案手法の検索結果(類似文検索 2)を切り替えて表示することができる。また入力文の単語と同じ同義語・WF グループに属している単語は同じ色でハイライトされる。

5. 実験

5.1 実験の概要

提案手法における有用性を調査することを目的として、某国立大学の学生 9 名を対象に実験を実施する。実験協力者は類似文検索を用いて用意した課題に解答、その後アンケートへ回答を行う。

WF 導入による効果を 2 種類の課題によって調査する。より多くの表現を用いて英文作成することが可能になるかを調査する「同義表現課題」、検索結果を参照することで、誤った品詞使用に対してより自然な英文に修正することが可能になるかを調査する「誤り訂正課題」を各 6 問(計 12 問)用意した。同義表現課題は日本語が提示され、その日本語に対応する英文を複数作成する課題で、誤り訂正は日本語とその英訳(部分的に要訂正箇所あり)が提示され、より自然な英文に訂正する課題である。

実験協力者は類似文検索を自由に使用しながら、提示例文を参考にして課題解答を行う。検索結果出力画面では①提案手法検索結果の提示、②横並びで従来の検索結果と提案手法の検索結果の両方を提示の 2 パターンを用意し、2 種類の問題を 3 問ずつ各提示パターンに割り当てた。提示パターン②を図 2 のように横並びとすることでタブ切り替えの仕様に気付かず、他方の検索結果の見落としを防ぐ。また左右の検索結果表示を問題ごとに交互に入れ替えることで、表示位置によってユーザの注目が偏ることが起こらないようにする。



図 2 : 提示パターン②

検索結果の例文中、解答作成時に参考にした例文(以下、参考例文)やその例文の中の各単語(以下、参考箇所)に対してチェックボックスにチェックすることを求めた。解答は複数作成してもよいとし、

同義表現課題に関しては複数作成を必須とした。実験解答と並行して提案手法の有用性やシステムの改善点の分析調査目的でアンケートを実施する。質問項目は各問での検索機能の評価、課題種類ごとの検索機能への評価、全体を通しての検索機能への評価となっている。

5.2 実験分析

実験時には提示パターン①②、参考例文、参考箇所、解答等の情報を自動取得し提案手法の特徴を分析する。提示パターン①から WF グループの単語を含んだ文章が参考例文として利用されるかを調査し、提案手法の有用性を分析する。提示パターン②から検索結果ごとの参考例文選択数や解答作成数、作成解答内容の比較を行い、提示方法の使い分けを分析する。また同一課題に対する複数解答に対して前後の解答を比較し参考例文・参考箇所が与える影響についての調査を行う。また用意した課題ごとの検索結果の使い分けの可否について検討する。そしてアンケートの回答から提案手法が課題解答に役立ったかを調査することで有用性を確認する。

以上の実験の結果ならびにそのデータ分析、検定結果などについては当日の発表にて報告を行う。

6. まとめ

本稿では新たにワードファミリーグループを導入し、見出し語の派生語等も含めて検索する手法を提案した。また提案手法による効果を確認するために小実験を実施した。実験結果をもとにデータ分析を行うことで同手法の有用性の検証やアプリケーションの改善に取り組んでゆきたいと考えている。

参考文献

- [1] Y. Miyazaki, S. Tanaka, Y. Koyama: Development of a Corpus-Based Web Application to Support Writing Technical Documents in English, World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (E-Learn 2014), Vol. 2014, No. 1, pp. 1371-1380 (2014).
- [2] 高松優, 水野淳太, 岡崎直観, 乾健太郎, 英作文支援のための用例検索システムの構築, 言語処理学会第 18 回年次大会発表論文集, pp. 361-364 (2012).
- [3] 陸峰, 三好康夫, 句共起関係コーパスを利用した第二言語作文支援のための用例文検索システムの構築, Technical Reports on Information and Computer Science from Kochi, Vol. 9, No. 8, pp. 1-4 (2017).
- [4] TreeTagger: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- [5] WordNet: <http://wordnet.princeton.edu>
- [6] I.S.P.Nation, "TheBNC/COCAwordfamilylists", https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/Information-on-the-BNC_COCA-word-family-lists.pdf