

# 構造的簡略化を用いた 例示型英文書作成支援 Web アプリケーションの開発と評価

天野翼<sup>†</sup> 宮崎佳典<sup>‡</sup> 田中省作<sup>\*</sup> 長谷川由美<sup>\*\*</sup>

静岡大学大学院総合科学技術研究科<sup>†</sup> 静岡大学大学院情報学領域<sup>‡</sup>

立命館大学文学部<sup>\*</sup> 近畿大学生物理工学部<sup>\*\*</sup>

## 1. はじめに

技術英文書を作成する際、学習者は適切な語や構文を選択する目的で、専門家によって書かれた技術英文を参考にするというアプローチをしばしば取る。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][2]。具体的には、大量に電子化した特定分野の英語科学論文集(コーパス)から、学習者が入力した英文に類似した技術英文や入力したフレーズを含む技術英文を複数抽出して、提示を行う。

本研究では、提案されたアプリケーション(以後、既存ツール)の有用性向上を目的とし、実験を行った。そして、実験で得られた知見に基づき、簡略化機能[3]を実装した。簡略化を行うことで、学習者の英作文作業に貢献する可能性が高くはないと考えられる技術英文中の節・句・語がそれらに対応するラベルに置き換えられ、学習者は参考になる語を把握しやすくなることが期待される。一方で、簡略化による実際の効果は現時点で調査されていないことから、再度実験を行うと共にアンケート調査を実施した。本発表では、以上のことについて報告する。

## 2. 既存ツール

既存ツールは、類似文検索とフレーズ検索の2種の検索機能から成る。

類似文検索では、学習者の入力した英文およびコーパス内の技術英文の  $n$ -gram、品詞、同義語を考慮した上で、 $\cos$  類似度を算出する。そして、 $\cos$  類似度が高い上位 50 件の技術英文を学習者に提示する。

フレーズ検索では、学習者の入力したフレーズが含まれる技術英文を提示する。入力するフレーズは品詞オプションやワイルドカードオプション等を指

定でき、例えば“is not [J]”([J]は形容詞を表す品詞オプション)と入力することで[J]が任意の形容詞の単語に置換されて検索され、“is not necessary”や“is not available”を含む技術英文を検索することができる。

## 3. 既存ツールを用いた実験

### 3.1. 実験の概要

既存ツールの有用性を調査するために、某私立大学の学生 36 名を対象に実験を行った。

実験協力者に求められるタスクは既存ツールに実装されている 2 つの検索方法(類似文検索・フレーズ検索)を用いて技術英文を検索し、各問題に解答することである。問題は、正解文探し、誤り訂正、自由英訳の 3 種類の問題を各 4 問、計 12 問用意した。それぞれの問題の詳細については下記の表 1 に示す。

表 1: 実験で用いた問題の種類

問題の種類	問題の詳細
正解文探し	コーパス内の技術英文を翻訳した日本語が与えられ、その日本語に対応する英文を探す
誤り訂正	日本語、ならびにその英訳(部分的に要訂正箇所あり)が与えられ、より自然な英文に修正する
自由英訳	日本語が与えられ、それを英訳する

検索した際に、参考になった技術英文(以後、参考文)がある場合、チェックボックスにチェックを入れて頂く。そして、実験協力者が修正の必要がないと判断した段階で、最終的な解答を提出して頂いた。

初歩的な問題である正解文探しについては、実験協力者を選定するために用いた。正解文探しの正答率が 50%未満の実験協力者については既存ツールを使うに当たって不適切であると考え、データから取り除いた。その他の問題については、複数の大学教員が採点を担当した。

### 3.2. 実験結果

参考文の数、参考文の語数、問題の得点などの基礎データを基に検討を行った。

実験で実験協力者が選択した参考文の語数の平均値は 17.2 語であった。一方、提示された技術英文

Development and Assessment of Example-Based Web Application to Support Technical Writing Using Structural Simplification

<sup>†</sup> Tsubasa Amano, Graduate School of Integrated Science and Technology, Shizuoka University

<sup>‡</sup> Yoshinori Miyazaki, College of Informatics, Shizuoka University

<sup>\*</sup> Shosaku Tanaka, College of Letters, Ritsumeikan University

<sup>\*\*</sup> Yumi Hasegawa, Biology-Oriented Science and Technology, Kindai University

の語数の平均値は22.8語であり、2つに差があるかどうかについて t 検定を行ったところ、有意差が認められた ( $p=0.000001<0.001$ )。このことから、語数が比較的少ない技術英文が学習者にとって、より参考になっているという傾向が認められた。

また、問題の点数が高い実験協力者と低い実験協力者の差は参考文献の数によるものではないかと考え、問題の合計点数が高いものから順に上位群(7名)、中位群(16名)、下位群(9名)の3群にウォード法を用いてクラスタリングした。そして、各群の参考文献の数の平均値を調査してみたところ、下記の表2のようになった。

表2：参考文献の数の平均値

誤り訂正			自由英訳		
上位群	中位群	下位群	上位群	中位群	下位群
2.1個	1.0個	0.7個	3.0個	0.8個	1.2個

各群の参考文献の数の差があるかどうかについて t 検定を行ったところ、誤り訂正では上位群・中位群、上位群・下位群に優位差が認められた ( $p=0.02<0.05$ ,  $p=0.0004<0.001$ )。同様に自由英訳についても上位群・中位群、上位群・下位群に優位差が認められた ( $p=0.0007<0.001$ ,  $p=0.02<0.05$ )。このことから、参考文献の数が英作文に影響する可能性が示唆された。

#### 4. 技術英文簡略化機能

実験結果から、技術英文を簡潔に表示する必要性が指摘された。本稿では、技術英文簡略化機能[3]を実装し、その実験およびアンケート調査した。

##### 4.1. 簡略化機能の概要

簡略化は2段階のプロセスによって行われる。

1段階目のプロセスでは、取得した技術英文の集合  $S$  において、共通して出現しない語に対し、それを優先的に対応する品詞ラベルに置換する。共通して出現しない語は学習者にとって参考となる程度が低いものであると考えられる。共通して出現するかどうかの指標は、技術英文の集合  $S$  に含まれる各語  $w$  について、以下の数式によって計量される：

$$\text{itf-isf}(w; S) = \frac{1}{f(w)} \log \frac{|S|}{sf(w)},$$

なお、ここに、 $f(w)$  は  $S$  に含まれる  $w$  の数、 $sf(w)$  は  $S$  のうち  $w$  を含む文の数、 $|S|$  は取り出された例文の総数を表す。この  $\text{itf-isf}$  は低頻度・低文数なほど大きな値を取る指標であり、この値が高い語は英作文の参考にならない可能性が高いと考えられ、品詞ラベルに置き換えられて行く。

2段階目のプロセスでは、構文構造の観点から、品詞ラベルと機能語の羅列のような冗長な箇所を節あるいは句ラベルに置き換え、語数の削減を行う。例えば、1段階目のプロセスによって技術英文が

“We propose a new solution for the <NN> of <DT> <NN> <NN>.” となったとする (<DT>, <NN> は各々、限定詞、名詞を表す品詞ラベル)。この文の末尾(下線部)は冗長な箇所であり、このプロセスの処理により “We propose a new solution for <NP>.” と語数が削減され、簡潔に表示される (<NP> は名詞句を表す句ラベル)。

##### 4.2. 簡略化機能を用いた実験およびアンケート

現時点では簡略化機能の有用性について調査されていない。そこで、既存ツールに簡略化機能を実装したツール(以後、現行ツール)を用いた実験を某私立大学の学生34名の協力の元、実施した。実験は、既存ツールを用いた際の実験結果と比較するために、実験の流れおよび問題は同様とし、計12問の問題を現行ツールを用いて解答して頂いた。簡略化機能の使用の有無については実験協力者の任意とした。ログデータは、既存ツールを用いた実験で取得したログデータに加え、簡略化機能を使用した回数等を取得した。そして、簡略化機能がどの程度使用されていたのか、簡略化によって参考文献の数に変化が生じるか、簡略化によって参考文献の語数に変化が生じるか等についてログデータを基に調査した。

また、現行ツールの更なる実用化に向け、アンケートを実施した。アンケートでは、類似文検索、フレーズ検索、簡略化機能それぞれに4段階で評価して頂き、その評価に至った理由についても聞き取りを行った。

以上の結果については、本発表にて報告する。

##### 5. まとめ

本稿では、既存ツールの有用性向上に向け、実験を行い、実験結果から簡略化機能を実装することにした。そして、簡略化機能の有効性を調べるために実験およびアンケートを実施した。実験およびアンケートの結果から、現行ツールの更なる有用性向上を目指している。

##### 参考文献

- [1] 山本昇平, 宮崎佳典, 技術文献コーパスを用いた英文書作成支援ツールの開発—類似文検索機能とパターン検索機能—, 統計数理研究所共同研究リポート 295, pp.71-95(2013)
- [2] 戸沢信晴, 宮崎佳典, 田中省作, 技術文献コーパスを用いた例文提示型英文書作成支援ツールの開発, 電子情報通信学会技術研究報告信学技報, 114(82), pp.69-72 (2014)
- [3] 天野翼, 渡部孝幸, 田中省作, 宮崎佳典, 構文情報を考慮した検索英文集合に対する汎化手法, 第14回情報科学技術フォーラム講演論文集, pp.(2)-211-214(2015)