

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

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

天野 翼^{*1}, 宮崎 佳典^{*2}, 田中 省作^{*3}, 長谷川 由美^{*4}

Tsubasa Amano^{*1}, Yoshinori Miyazaki^{*2}, Shosaku Tanaka^{*3}, Yumi Hasegawa^{*4}

^{*1} 静岡大学大学院総合科学技術研究科

^{*1} Graduate School of Integrated Science and Technology, Shizuoka University

^{*2} 静岡大学大学院情報学領域

^{*2} College of Informatics, Shizuoka University

^{*3} 立命館大学文学部

^{*3} College of Letters, Ritsumeikan University

^{*4} 近畿大学生物理工学部

^{*4} Biology-Oriented Science and Technology, Kindai University

Email: gs16001@s.inf.shizuoka.ac.jp

あらまし：技術英文書を作成する際、書き手は適切な語や構文を選択する目的で、実際に論文で使われた文を参照することがある。我々は、このような作業を支援する Web アプリケーションを構築している。本研究では、過去の予備実験から、参照する技術英文の複雑さが問題となっており、それらを構造的に簡略化する機能を追加した。また、アプリケーションの評価のため、追加機能など利用者の使用機能の履歴を集め、分析した。本稿では、その実験と結果を示す。

キーワード：英作文支援、簡略化、Web アプリケーション

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⁾。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（論文データベース）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⁽²⁾を実装した（以後、現行ツール）。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。本発表では、簡略化による実際の効果に関する実験とともに実施したアンケート調査について報告する。

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、コサイン類似度を算出する。そして、コサイン類似度が高い上位 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“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。

この既存ツールは提示された技術英文を参考にし英作文を行うことが想定されている。しかし、実際のところ既存ツールによって提示される技術英文は、実際の技術文書という性質上、一般に語数が多いものが多く、複雑かつ長文が故に参考にしづらいたことが考えられた。そこで、本研究では大学生 36 名の協力のもと実験を行い、学習者が参考にする技術英文（以後、参考文）の傾向を調査した。結果、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。

3. 現行ツール

既存ツールを用いた実験により、技術英文を簡潔に表示する必要性が指摘された。そこで、既存ツールに簡略化機能⁽²⁾を実装し、ツールの有用性を調査するために実験およびアンケートを実施した。

3.1 現行ツールの簡略化機能

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、検索で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（＝特定のいくつかの文にしか出現しない語）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：

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

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 $itf-isf$ は低頻度・低文数なほど大きな値を取る指標である。したがって、この値が高い語は所与の検索という観点で合致しない断片であることから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ゆく。

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)。

3.2 現行ツールを用いた実験

実験は前回の実験と同じ大学の学生 34 名 (学生は前回とは異なる) の協力のもと実施した。実験協力者に求められるタスクは現行ツールに実装されている 2 つの検索方法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。問題は表 1 にある 3 種類の問題を各 8 問、計 24 問用意した (うち半数の 3 種類の各 4 問は前回実験と同一)。

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

検索した際に参考文献がある場合、実験協力者はチェックボックスにチェックを入れる。なお、ここでの簡略化機能の使用は任意とした。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する。また、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%未満の実験協力者のデータは分析対象から除いた (該当者 4 名)。

3.3 実験結果

簡略化機能の使用は任意としたため、簡略化機能を 1 回以上使用した実験協力者は 12 名 (A~L と表記) であった。ここで、簡略化機能が使用された解答を使用群、簡略化機能が使用されなかった解答群を未使用群として各問題の参考文献の数の平均値を比較した (表 2)。誤り訂正において使用群、未使用群の 2 つに差があるかどうか t 検定を行ったところ、

有意差が認められた ($p < 0.001$)。また、自由英訳についても有意差が認められた ($p < 0.001$)。

表 2 使用群・未使用群の参考文献の数

問題	使用群	未使用群
誤り訂正	1.9個	0.5個
自由英訳	2.0個	0.5個

同様に、使用群と前回実験の解答 (以後、前回実験群) との比較も行った (表 3)。なお、ここでの使用群は前回実験群の問題と同一の問題のみにしている。結果、誤り訂正および自由英訳において使用群、前回実験群間に有意差が認められた ($p < 0.05$)。

表 3 使用群・前実験群の参考文献の数

問題	使用群	前回実験群
誤り訂正	2.1個	1.1個
自由英訳	2.1個	1.4個

今回、簡略化機能を 1 回以上使用した実験協力者 12 名が単に多くの参考文献を選ぶ傾向だった可能性もあるため、当該実験協力者の簡略化機能の使用/未使用間の差について調査した。各問題において簡略化機能使用時と未使用時の参考文献の数の平均値は以下の表 4 の通りである。

表 4 使用時・未使用時の参考文献の数

実験協力者	A	B	C	D	E	F
使用時	2.0	2.0	1.7	2.0	1.5	3.0
未使用時	0.2	0.3	0.9	0.3	0.5	0.6
実験協力者	G	H	I	J	K	L
使用時	2.0	2.0	2.0	2.0	1.0	2.0
未使用時	0.6	0.6	0.9	0.5	0.5	0.6

使用時と未使用時に差があるかどうか t 検定を行ったところ、有意差が認められた ($p < 0.001$)。これは、今まで一瞥するのみだった英文が、簡略化機能によって参考文献を選ぶ際の材料となったのではないかと推察される。その他の実験結果およびアンケートについての詳細については、発表時に報告する。

4. おわりに

本研究では既存ツールを用いた実験を実施し、語数が少ない技術英文が参考になりやすいという知見が得られた。そして、複雑かつ長文をコンパクトに一旦提示する簡略化機能を実装し、その有用性を調査したところ、英作文時に参考となるコーパス文の数が上昇することが統計的に有意に観察された。

参考文献

- (1) 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)
- (2) 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 FIT 講演論文集, pp. (2)-211-214 (2015)