

チャンク情報を考慮した例示型英文書作成支援Web アプリケーションの開発

メタデータ	言語: ja 出版者: 静岡大学 公開日: 2019-05-09 キーワード (Ja): キーワード (En): 作成者: 宮崎, 佳典 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00026470

平成 30 年 6 月 18 日現在

機関番号：13801

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K01064

研究課題名(和文) チャンク情報を考慮した例示型英文書作成支援Webアプリケーションの開発

研究課題名(英文) Web Application Supporting Writing Technical Documents in English: In Consideration of Chunk Information

研究代表者

宮崎 佳典 (Miyazaki, Yoshinori)

静岡大学・情報学部・教授

研究者番号：00308701

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究の目的は、技術系の人間に必要な、英語による技術文書作成を支援するアプリケーションを開発することである。具体的には、ユーザが英文を入力すると、技術論文を集めたコーパスの中から類似した文を抽出し例文として提示し、必要に応じてその例文を簡略化して表示する。開発Webアプリケーションには各ユーザの詳細ログデータを取得するなどの機能も追加・さらにインタフェースの改良を加え、「類似文検索機能」ならびに「フレーズ検索機能」を実装した。2年連続してユーザ実験ならびにアンケート調査を実施し、本Webアプリケーションの有用性について評価を行った。また、ユーザのログデータを集計、分析した。

研究成果の概要(英文)：This paper discusses the development and brief assessment of a Web application that supports non-native speakers of English in writing technical documents in English. It is assumed that referring to technical documents in English written by professionals is highly likely to occur on such occasions, for the purpose of selecting appropriate wordings and sentence structures. One problem is that the sample sentences selected by the application are usually too long to refer to for learners, considering the characteristics of such sentences being a part of technical documents. This fact was revealed from the experiment conducted last year, and the authors decided to add a new function to make original sentences simplified structurally. The result of another experiment will be shown, with this function implemented into the system to show its effectiveness.

研究分野：教育工学, 数値解析, e-Learning

キーワード：技術英文作成支援 技術文献コーパス コサイン類似度 パターンマッチング 簡略化

1. 研究開始当初の背景

日本は技術立国でありながら、現状ではその技術を表現する英語力が不足し、世界に充分発信できていない。従って技術者の英語による情報発信をバックアップする体制作りは急務である。

専門的な技術文書に類するものは英語で書かれる場合が多い。技術文書には日常では使用されない単語や表現が頻繁に現れることから、特に EFL(English as a Foreign Language)環境下にある大学生などの英語学習者にとってアカデミックな英文の文書作成は簡単な作業ではない。このような背景から、英語学習者向けに技術文書の書き方を指南する本も多く出版されている。しかしながら、これらの本は技術文書に多く見られる基本的な単語や文法の紹介に終始しており、それだけでは英語で技術文書を書くことはまず不可能である。また、一般的な英文書作成支援ツールは今までもいくつか開発されているが、技術に特化した英文書作成支援ツールにフォーカスをあてたものは見当たらない。さらに一般英文書用ツールでも、その多くは日本語を入力するシステムであり、日本人にとって利用しやすいシステムである反面、機械翻訳の不正確さが原因で不適切・無関係な例文が提示されることも少なくない。それはまたユーザを日本語の母語話者とする者に限定してしまうことにもなる。

2. 研究の目的

以上のような先行研究の状況に対して、本研究では英語学習者の実際の学習プロセスに可能な限り符合し、学習者のニーズにあった学習効率の高い技術文書作成支援システムを目指している。その意味で、本システムは先行研究の英文書作成支援ツールとはその立場を異にしている。

英語学習者の英文作成のプロセスに即して本アプリケーションの意義を考察する。学習者は英文技術文書作成の際、(Point 1)当該分野のオーソライズされた技術文書から自分が書きたいものに似た英文を探し、それを手本にしようとする。しかし、書きかけの英文中の単語では類似した英文を探すことは容易ではない。そこで、(Point 2) 随時類似した単語や表現に置き換えて検索することが必要になる。

さらに学習効率を考えれば、コーパスから類似英文の検索時に (Point 3) 技術文書に特有の、当該分野毎の特徴表現やコロケーション(語の組み合わせ)に重み付けすることで他の一般的な表現とは差別化することが必要である。さらに実際に検索される例文は構造が複雑で難解なことが多く、(Point 4) 文構造が明確化されるような簡略化機能(文内の部分的な構造を省いたり記号化することで簡略に見せる機能)を付加することによってそれが実現する。

以上の全項目を実装することで実用的な

英文書作成の支援が可能となると考える。この理念に基づき、本研究グループは例文提示型支援システム構築のための、初期段階のプロトタイプ作成を終えている。しかし現段階では上記(Point 1)~(Point 4)の動作を実現するには至っておらず、理論に基づく仕様策定および実用に耐えられるレベルの実装が今後必要となる。一般公開を想定した実用性の高い Web アプリケーションを構築するには、詳細な開発工程に裏付けされた動作確認やデバッグ作業、ならびに開発後の複数の実験を通じた有用性の実証が不可欠であり、本申請はこれを目的とするものである。また、本システムは入力には完結した英文のみならず、フレーズ単位や単語入力だけでもそれに応じた適切な出力を返す特性を持ち、ユーザに負担を感じさせないシステムとすることが特長に挙げられる。加えて出力表示用インタフェースには対応語が特定しやすいよう、目的に応じた多段階のハイライト機能を装備する予定である。

3. 研究の方法

以上目的について述べたが、実用化に際し、この英文書作成支援システムには解決しなければならない問題点も多い。それは計算速度と精度の問題に大別される。

本システムの計算速度に関しては、特に入力文に類似した文を大量の用例(コーパス)から検索する際に行われる大規模なベクトル計算時間に依るところが大きい。その計算の抑制のために、すでにデータ構造に工夫を重ねている。また、入力に対して参考になる例文提示の精度に関しては、多面的アプローチが必要と考えている。コーパスを拡充すれば当然、入力に対するユーザが満足する適切な例文が含まれる可能性は高くなる。これには技術ジャーナルから著作権を購入することで質・量ともに充実化を図る。入力文内の単語を適宜置き換えるための同義語グルーピングの精度向上も重要なポイントである。特に技術用語の同義語群はコーパスに依存し、複数のシソーラスなど参照しながらの生成は大量の計算を要し、動的に対応するため新規導入するサーバで即時処理を行う予定である。(Point 2)。分野別の特徴表現やコロケーションを取得することにも重点を置き、文中の n-gram(連続する語 n 組)を採用することで、単語間の修飾関係を擬似的に扱うものとする。(Point 3)。最後にかつ最重要課題はユーザのパーソナライゼーションである。どの文を類似度が高い・参考になるとみなすかはユーザの研究分野や英文の嗜好性に依存し、また提示すべき例文の簡略化のレベルも同様であろう。本テーマは精度やユーザの満足度に大きな影響を及ぼすことは間違いなく、ユーザの各種学習履歴情報(閲覧したコーパス文、実際に推敲に使用された表現、入力英文、過去に選んだ汎化レベルなど)を戦略的に保存・利用していき、必要に応じて因

子分析法などの統計的手法を駆使する。特に簡略化のレベルは現時点においてユーザが指定する形式になっており、システム使用の都度設定する手間を考えると、パーソナライゼーションの必要性は高い (Point4)。

以上のように、本研究で開発する英文書作成支援アプリケーションには実用性のみならず、簡略化機能、パーソナライゼーションに代表される学術的な意義も有する。対象を技術文書作成とし、固有の特徴表現や n-gram を用いて類似度を計算する点も本研究の独創的な点である。

4. 研究成果

(1) Web アプリケーション全体の枠組みが大方完成しているのを鑑み、理論的なアプローチを精査した。特に句構造やチャンク化処理ならびにその高速化、パーソナライゼーション、そしてユーザビリティを考慮したインタフェース開発について検討を行った。研究代表者ならびに研究分担者間でミーティングを行い、以下を討議した：・システム処理の高速化 (類似文の検索法の改善、処理の並列化など) ・句構造、チャンク構造を勘案した類似度計算のアルゴリズム立案 ・簡略化機能による文コンパクト化 ・パーソナライゼーションへのアプローチの決定。また、データの流れを精査しながら、システムの実現可能性 / 可用性についても議論した。

(2) 次に、Web アプリケーションの全体の枠組みに沿って実装を進めた。現時点でアプリケーション (言語) が動作しており、さらなる追加コンポーネントを制作した。コーパスには、名古屋工業大学の研究グループによって編纂されたコーパスの使用許可をすでに得ており、総計で既に 30 万文を超えている (Association for Computational Linguistics, Nature, Scientific American, Biology 系のテキストで構成される) が、実用性を考え、また過去の研究成果からも学習者の研究分野に近い分野のコーパスが他のコーパスに比べ有用であることを踏まえ、さらなるコーパスデータの拡充を策定した。さらにインタフェースの改良を加え、もともと 2 つの機能だった「類似文検索機能」ならびに「フレーズ検索機能」を一手に引き受ける Web アプリケーションを現在開発した。

(3) さらに、初年度に詳細に行った Web アプリケーションの全体の枠組みに沿って実装を推し進め、ユーザの詳細ログデータを取得するなどの大方の機能も追加され、概形が整った。さらにインタフェースの改良を加え、もともと 2 つの機能だった「類似文検索機能」ならびに「フレーズ検索機能」を一手に引き受ける Web アプリケーションも試作した。英語教員である研究分担者の勤務先大学において 2 年連続してユーザ実験ならびにアンケート調査を実施し、本 Web アプリケーション

(特にその中の簡略化機能) の有用性について評価を行った。また、教師権限で閲覧可能なユーザのログデータを集計、分析することで、英語修得レベルの推測を行った。

(4) 開発された Web アプリケーションの実用性向上に向け 2 度実験を行い、それより得られた知見を元に、簡略化という機能の追加実装を行った。1 度目の実験で「学習者が参考にする英文は技術英文の中では比較的簡易なものである」、「多くの英文を参考にすることができる学習者はより良い英作文を行える」という事実が観察された。これらの事実から、英文を簡潔に表示する必要性が指摘された。そこで、英文を簡潔に提示するための手法である簡略化機能を現行ツールに実装した。簡略化により、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と考えられる。

(5) 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、2 度目の実験で簡略化機能によって、参考文の数、参考文の語数に生じる変化を調査した。実際、簡略化機能によって参考文の数が上昇するという実験結果を得ることができた。また、簡略化機能によって、学習者は比較的複雑な英文も参考にすることができるという知見が得られた。これらの結果から、簡略化機能は英作文をする上で、有用であるということが示唆された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 3 件)

天野 翼, 宮崎 佳典, 田中 省作, 長谷川 由美, 簡略化を用いた例示型英文書作成支援ツールの開発と検証, 統計数理研究所共同研究リポート 397, pp. 13-24 (2018), 査読無

天野 翼, 宮崎 佳典, 田中 省作, 長谷川 由美, コーパスを用いた技術英文書作成援用ツールの開発とその評価 (その 2), 統計数理研究所共同研究リポート 382, pp. 41-52 (2017), 査読無

宮崎 佳典, 戸沢 信晴, 田中 省作, コーパスを用いた技術英文書作成援用ツールの開発とその評価, 統計数理研究所共同研究リポート 356, pp. 1-16 (2016), 査読無

〔学会発表〕(計 8 件)

天野 翼, 宮崎 佳典, 田中 省作, 長谷

川 由美, 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, 情報処理学会第 80 回全国大会【学生奨励賞受賞】, 2018 年

天野 翼, 宮崎 佳典, 田中 省作, 長谷川 由美, 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, 2017 年度 JSiSE 学生研究発表会(東海地区)【優秀発表賞受賞】, 2018 年

Y. Hasegawa, Y. Miyazaki, T. Amano, A Web Application to Support Technical Writing for Non-native (EFL) English speakers, The 31st Annual Hawaii Association of Language Teachers (HALT) Conference, 2017

宮崎 佳典, コーパスを用いた技術英文書作成支援ツールを用いた実験とその評価, 言語研究と統計 2017(統計数理研究所発表会), 2017 年

宮崎 佳典, 田中 省作, コーパスを用いた技術英文書作成支援ツールの開発とその評価, 言語研究と統計 2016(統計数理研究所発表会), 2016 年

天野 翼, 渡部 孝幸, 田中 省作, 宮崎 佳典, 共起関係ならびに構文情報を考慮した英文汎化と英作文支援, 2015 年度 JSiSE 学生研究発表会(東海地区), 2016 年

戸沢 信晴, 宮崎 佳典, 長谷川 由美, 田中 省作, コーパスを用いた技術英文書作成支援ツールの開発とその評価, 日本 e-Learning 学会 2015 年度学術講演会, 一般セッション, 2015 年

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(FIT)講演論文集, 2015 年

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

取得状況(計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

〔その他〕
ホームページ等
<http://mya-lab1.cs.inf.shizuoka.ac.jp/~tozawa/ewss/web/index.php>

6. 研究組織

(1) 研究代表者

宮崎 佳典 (MIYAZAKI, Yoshinori)
静岡大学・情報学部・教授
研究者番号 : 00308701

(2) 研究分担者

小山 由紀江 (KOYAMA, Yukie)
順天堂大学・医学部・非常勤講師
研究者番号 : 20293251

田中 省作 (TANAKA, Shosaku)
立命館大学・文学部・教授
研究者番号 : 00325549

長谷川 由美 (HASEGAWA, Yumi)
近畿大学・生物理工学部・准教授
研究者番号 : 40585220

(3) 連携研究者

()

研究者番号 :

(4) 研究協力者

()