

医学系論文抄録の日英対訳コーパスを活用した 医学英語教育支援システム開発

Development of a Medical English Education Support System Using Parallel Corpora of Medical Research Article Abstracts

中野 愛実^{*1}, 宮崎 佳典^{*2}, 浅野 元子^{*3}, 藤枝 美穂^{*3}

Megumi Nakano^{*1}, Yoshinori Miyazaki^{*2}, Motoko Asano^{*3}, Miho Fujieda^{*3}

^{*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} 大阪医科薬科大学医学部

^{*3} Faculty of Medicine, Osaka Medical and Pharmaceutical University

Email: nakano.megumi.17@shizuoka.ac.jp

あらまし：医学英語教育分野において、医学系論文抄録の読解、執筆能力を身に付けることは、論文本体に取り組む前段階として特に肝要である。国際医学誌に公表された研究論文の英文抄録と公式日本語訳データを活用し、医学英語教育支援システムのプロトタイプを開発した。本研究では、医学生が実際に英語の授業で使用した学習履歴からシステムの改善点について検討し、新機能の提案・実装を行う。

キーワード：医学英語、対訳コーパス、Web アプリケーション、英文読解・作文支援、学習履歴

1. はじめに

医療の最新情報を得るために英語論文を読む、外国人の患者を診る機会があるなど、医師には英語力が不可欠であり⁽¹⁾、医師・医学専門教員を対象とした英語能力のニーズ調査から、学生に求めるスキルとして論文読解へのニーズが高いことが分かっている⁽²⁾。医学論文読解には ESP (English for Specific Purposes, 特定の目的のための英語) の中でも特に医学分野に特化した EMP (English for Medical Purposes) の学習が必要である。これらの背景から、本研究では提供された医学英語論文 (タイトルならびに概要部) の対訳付きコーパスに対し、医学英語学習者向けに英文読解・英作文学習サポートを目的としたシステムを試作した。本発表では、システムのプロトタイプを医学生が授業で使用した履歴を分析することでシステムの改善点について検討を行い、その結果行った機能の追加・利用状況を調査・報告する。

2. 関連研究

関連研究として JECPRESE⁽³⁾や AWSuM⁽⁴⁾などが挙げられる。⁽³⁾は単語・フレーズを検索するコンコーダンサ (検索語を中央にしてコーパスから表現を抽出する KWIC (keyword in context) 表示機能を有するソフト) である。口頭発表をテキスト化したデータを対象とし、話者の目的別に英語と日本語の両方から文を検索する。⁽⁴⁾はムーブ (表現意図) を用いた英文作成支援システムであり、論文のセクションとムーブごとに高頻度の語連鎖を提示し、コンコーダンサの機能等も備わっている検索システムであるが、日本語の対訳表示はない。

3. 医学英語学習システム

3.1 医学英語論文コーパス

本システムで使用する医学英語論文 (タイトルならびに概要部) の対訳付きコーパスのデータ構造を例と共に以下の図 1 に示す。なお、このタイトル・概要部の 1 編を英文書と定義する (ここに英文書数 1,469, 英文数 19,581。また各英文は日本語対訳を有する)。英文書はタイトル, 背景, 方法, 結果, 結論 (研究情報) の 6 セクションより成る。図 1 の例は、ID=1695 の英文が英文書 2 の 14 文目、結論部分で使用されていることを示す。

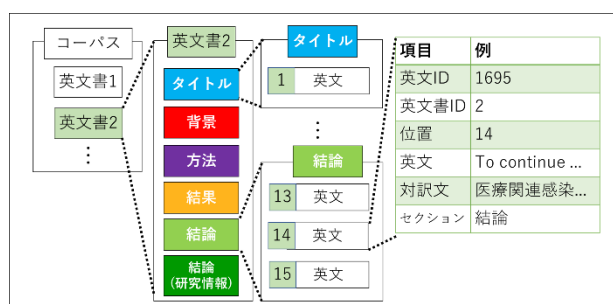


図 1：データ構造，データ例

3.2 システムの概要

まず、提案システムを図 2 に示す。検索には検索窓 1 つに入力する単一検索、検索窓を 4 つまで追加して入力可能な複合検索があり、1 つの検索窓に 2 語以上入力された時 (例: high risk, adherence to), そのひとまとまりが使用された文を検索する。セクション指定や活用表現等の詳細設定機能も備えるほか、対象言語を検索窓単位で指定可能なため、複合検索

時に日本語と英語を組み合わせたクロス検索を実現している。検索エンジンは Elasticsearch⁽⁵⁾を使用し、ユーザの入力から検索クエリを自動生成する。出力結果はヒットした英文単位でKWIC表示されたのち、英文・対訳文双方の文全体を表示したり、ヒットした英文が属する英文書全体を、対訳文と共に文全体表示が可能である。各種機能操作時には学習者の操作ログを取得し、学習分析 (LA, Learning Analytics) に結びつけることを念頭に置いている。

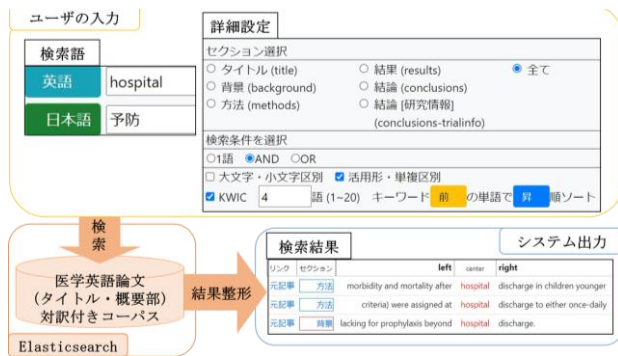


図 2：システムインターフェース

4. 評価実験・分析

医学生向け授業の使用履歴を分析し、システムの改善点・追加項目を検討した。実験は国内医科薬科大学の1年生75名、3年生84名に対して各学年2回の計4回実施した。授業内容は、1年生は語彙レポート、サマリーライティング、3年生はパラフレーズ2回(それぞれ異なる語を対象)を行った。4回の実験を通じた検索総数(1,597件)について、英語を入力した検索(1,053件)に着目し、検索種別、検索結果件数で分析した結果を以下の表1に示す。

表 1：英語検索分析(検索種別・検索結果件数)

検索種別	検索条件	検索結果件数	
		0件	1件以上
単一		339	630
複合	AND 検索	42	38
	OR 検索	0	4

結果、36.2% (381件) で何もヒットせず、内訳は、各検索窓ではコーパスに存在するものの、組み合わせが存在しなかった複合検索6.8% (26件)、言語設定のミス10.5% (40件)、誤字18.4% (70件)、コーパスに存在しない語・表現64.3% (245件)であった。なお、入力と言語設定が一致していない場合は言語設定のミスと判断、それ以外の入力については1語ずつ誤字を判定した。今回の判定基準に際してはPython3.6.9のAutocorrect⁽⁶⁾を利用した。

以上の結果より、新機能として、検索結果が得られなかった時コーパス内部から単語をサジェストする機能、検索時言語を自動設定する機能を実装した。

5. サジェスト機能

サジェスト機能は、検索結果が得られなかったとき、入力語とレーベンシュタイン距離に近い英単語5件を単語データから提示する(図3)。



図 3：サジェスト機能実行例

単語データは、コーパス内英文の形態素解析結果から単語単位で抽出した単語リストの中から、数字のみ、1文字、数字から始まる語を除いたデータ(4,938件)を使用する。

6. 入力言語自動検出

入力語から検索クエリを作成する際に、言語検出を追加する。現在は、正規表現を用いることで、英字のみで構成されている場合は英語検索、それ以外は日本語検索と判断している。自動で検出するかどうかはユーザが設定する仕様とした(図4)。



図 4：言語選択インターフェース

7. まとめ

本研究では、実際の授業で使用された履歴より検討した追加機能として、サジェスト機能・言語自動検出機能を実装した。機能実装後の実験は6月に実施予定、発表時に結果を報告する予定である。

参考文献

- (1) 坂田直樹, 田中英理, 藤枝美穂, 鈴木幸平, 中村仁紀, Can-Do Statements を利用した医学英語教育ニーズの分析: 医学部教員へのアンケート結果について, Journal of Medical English Education, 14 (1), pp. 15-24 (2015).
- (2) Naruenatwatana, N. & Vijchulata, B. A study of the needs of medical students in the use of academic English perceived by three groups: Medical students, teachers of English and subject teachers, Studies in Language and Language Teaching, 10, pp. 1-23 (2001).
- (3) Kunioishi, N., Noguchi, J., Hayashi, H. & Tojo, J. An online support site for preparation of oral presentations in science and engineering, European Journal of Engineering Education, 37:6, 600-608 (2012).
- (4) 水本篤, 浜谷佐和子, 今尾康裕, ムーブと語連鎖を融合させたアプローチによる応用言語学論文の分析—英語学術論文執筆支援ツール開発に向けて—, 英語コーパス研究 23, pp. 21-32 (2017).
- (5) Elasticsearch, <https://www.elastic.co/jp/>, (参照 2022-5-22).
- (6) Filip Sondej, 2021, Autocorrect, <https://github.com/filyp/autocorrect>, (参照 2022-5-22).