

英作文支援のための用例検索システムの構築

高松 優[†] 水野 淳太[†] 岡崎 直観[‡] 乾 健太郎[†]

東北大学[†] 科学技術振興機構 さきがけ[‡]

{takamatsu, junta-m, okazaki, inui}@ecei.tohoku.ac.jp

1 はじめに

近年、英語の重要性がますます高まっているが、英語が母語ではない日本人にとって、英語で文章を書くことは容易ではない。そのため、英作文を支援するための研究が多くなされている。英語学習者の英作文を支援する方法の1つに、学習者自身が実際の用例から語法を発見し理解していく、帰納的な学習方法がある [1]。こうした学習方法はデータ駆動型学習 (Data-Driven Learning) と呼ばれる。DDL では帰納的学習によって、学習者の「気づき」が導かれるため、発見した語法に関する情報が学習者の記憶に残りやすい [2]。このように、英語学習者にとって DDL が有用であることは既に明らかになりつつある [2, 3, 4]。自然言語処理の分野では、機械翻訳、スペル訂正、文章校正、用例検索などの研究が英作文支援と関連が深い。

本研究では、基本的な英文法は理解している大学院生程度の学生を支援対象として想定し、論文執筆などのテクニカルライティングを支援するため、用例検索に焦点をあてる。用例検索とは、多数の英文を収録したコーパスからある表現を含む文を検索し、検索文や検索文数を閲覧することによって、その表現の適切性や用法を確認するものである。例えば、図 1 では、“approach for”を検索クエリとして入力して、大量の英文が含まれるコーパスから“approach for”を含む文を検索し、検索結果として“approach for”を含む文と、その検索文数を提示している。近年、用例検索を中心とした英作文指南本がいくつか出版されるなど、注目されている作文支援技術である [5, 6, 7]。

しかし、既存の用例検索システムは英作文における問題を十分に考慮して設計されていない。そのため、英作文における用例検索システムの有用性について、教育学的な評価を十分に行える段階にない。

本研究では、まず、英作文に関する文献に基づいて、日本人が英語を執筆する際に直面する問題に対する、用例検索システムによる支援の可能性を調査した。次に、調査結果に基づいて、英作文支援に必要な機能を検討し、それらの機能を備えた用例検索システムを設

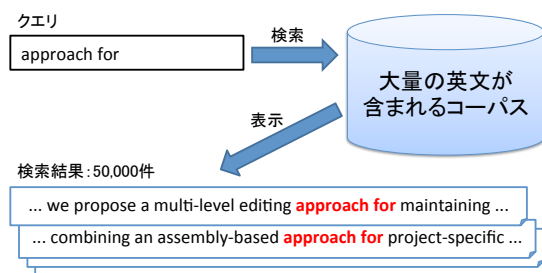


図 1: 用例検索の仕組み

計した。最後に、実際の英語教育の現場と連携して大規模かつ長期的な評価が行えるように、用例検索システムを実装した。テクニカルライティングを対象とするため、英文コーパスには ACL Anthology Reference Corpus (ACL ARC) [8] を用いた。

2 関連研究

用例検索の方法として、代表的なものに、Google などの検索エンジンを活用することがあげられる。調べたい表現を検索クエリとした場合の検索文書数やスニペットを閲覧することで、英作文に役立てることができ [5]。しかし、検索エンジンは用例検索を目的として設計されていないため、多くの用例を集約したり、品詞の指定といった複雑な検索には不向きである。そのため、用例検索に特化したツールである用例検索システムが研究されてきた。

既存の用例検索システムには様々なものがあるが、このうち、代表的な 3 つのシステムについて述べる¹。Wible ら [11] による StringNet Navigator²は、用例をそのまま表示するだけでなく、品詞別に集約することで抽象化した形で表示することも可能である。松原ら [12] による ESCORT³は、入力されたクエリ中の全ての単語が含まれており、かつその単語間に直接的な関係のある用例のみを提示する。その関係は依存構造に

¹他には、英語例文検索 EReK (<http://erek.ta2o.net/>)、フーズ検索システム SCOPE [9] (<http://scope.itc.nagoya-u.ac.jp/>)、exemplar (<http://www.springerexemplar.com/index.aspx>)、NativeChecker (<http://native-checker.com/>)、綱島ら [10] がある。

²<http://nav.stringnet.org/>

³<http://escort.itc.nagoya-u.ac.jp/>

よって捉えられる。大名 [13] は、辞書機能を組み込むことにより、学習者が適切な語を選んで検索できるように誘導している。

このように、今日、様々な用例検索システムが存在している。しかし、綱島ら [10] による小規模な評価以外に、殆どの研究では、どのような検索機能が必要であるのかといった、教育学的な評価は行なわれていない。

文章校正の研究においても、用例検索の有効性が期待されている。Chodorow ら [14] による ESL Assistant は、文章の自動訂正時に、訂正前の表現と訂正後の表現のそれぞれをクエリとして検索エンジンで検索した結果を提示する。その検索文書数やスニペットによって訂正の妥当性を示しており、これは用例検索のアプローチと類似している。

3 英作文における用例検索の有用性

日本人が英語で文章を書く際に生じる問題には様々なものがある。本節では、これらの問題に対する、用例検索による支援の可能性について述べる。

まず、日本人を対象とした英作文に関する文献 [5, 15, 16, 17, 18, 19] に基づいて、英作文の際の問題について、特に用例検索の観点から分析を行った。その結果、(1) 表現の適切性を確認したい場合、(2) 表現の組み合わせが分からない場合、(3) 表現の用法が分からない場合、の3つの場合に分類した。(1) は、クエリとして入力した表現がどの程度利用されているかに基づいて、その適切性を判断する場合であり、用例検索の最も基本的な利用方法である。(2) は、適切な表現の組み合わせを確認する場合である。(3) は、特定の表現を含む文を閲覧することにより、その表現の用法を確認する場合である。

次節以降では、それぞれの場合に、どのようなクエリが入力され、それに対してどのような用例を提示することが有効であるかについて述べる。

3.1 表現の適切性を確認したい場合

ある内容を表す表現として1つまたは複数の表現が想起できた場合、その表現が適切であるか、また、どの表現がより適切であるかを確認したいことがある。

例えば、「～をするための approach」という内容を表すために、“approach for ~”と“approach to ~”の2つの表現を思いついたとする。この時、“approach for”と“approach to”のそれぞれをクエリとして入力する。それを文字列として含む以下の用例を検索することができれば、その検索文数⁴から、“approach to ~”と表現するほうがより一般的であることが分かる。

⁴ACL ARC[8] 内で検索した場合の検索文数である。

また、“for”の後には動名詞や名詞が続き、“to”の後には動名詞や不定詞が続くことが分かる。

“approach to”で検索した結果：8,196 件
<ul style="list-style-type: none"> • ... induce a deeper <u>approach to</u> learning by means of ... • ... the mixed representation <u>approach to</u> allow the system to choose ...

“approach for”で検索した結果：1,246 件
<ul style="list-style-type: none"> • ... we propose a multi-level editing <u>approach for</u> maintaining consistency ... • ... combining an assembly-based <u>approach for</u> project-specific method construction ...

3.2 表現の組み合わせが分からない場合

表現したい内容に対して、どのように表現を組み合わせたら良いか分からない場合がある。これは、組み合わせる表現の片方を言い換えたい場合と、ある表現と組み合わせられる表現を俯瞰することにより選択したい場合に分類することができる。

3.2.1 表現を言い換えたい場合

英語のテクニカルライティングでは、同じ表現の繰り返し嫌われる傾向にある [16, 17]。しかし、学習者の語彙が少ない場合は、繰り返しを避けるために他の表現を探すのに苦労することがある。

例えば、“results show that ...”という表現の“show”を別の語に言い換えたい場合、“results [verb] that”と入力して、以下の用例を検索することができれば、“show”の代わりに“indicate”や“demonstrate”を用いることが可能であると分かる。ここで [verb] という記号は動詞を表し、あらゆる動詞が当てはまる。また、“indicate”と“demonstrate”の検索文数を比較すると、“show”と言い換える語としては“indicate”のほうがより一般的な表現であることが分かる。

“results [verb] that”で検索した結果：2,543 件
<p>“results show that”を含む用例：1,156 件</p> <ul style="list-style-type: none"> • Our <u>results show that</u> forward model adaptation alone ... • ... representation, however <u>results show that</u> system performance ... <p>“results indicate that”を含む用例：429 件</p> <ul style="list-style-type: none"> • Nevertheless, <u>results indicate that</u> our learning approach ... • ... and our own <u>results indicate that</u> the same description may ... <p>“results demonstrate that”を含む用例：75 件</p> <ul style="list-style-type: none"> • The <u>results demonstrate that</u> the sensor-system is a ... • Experimental <u>results demonstrate that</u> the use of phrase-based translation ...

3.2.2 組み合わせる表現が思いつかない場合

ある表現と組み合わせる表現として適切なものがまったく思いつかない場合について述べる。

例えば、“idea”という語と組み合わせる形容詞を俯瞰したい場合、“[adj] idea”と入力して用例を検索することにより、“main”や“key”、“novel”といった形容詞が用いられていることが分かる。

“[adj] idea”で検索した結果：7,426 件
“main idea”を含む用例：274 件
・ The <u>main idea</u> behind this heuristic is to find...
・ ... these components supported the <u>main idea</u> of the proof.
“key idea”を含む用例：165 件
・ Our first <u>key idea</u> is to maximize the target signal to ...
・ The <u>key idea</u> is that a solution of the ...
“novel idea”を含む用例：26 件
・ ... an ability to produce many <u>novel ideas</u> or solutions, a flexible approach to ...
・ We propose a <u>novel idea</u> of forming solder microbumps on the ...

3.3 表現の用法が分からない場合

ある表現の用法が分からない場合、その表現を含む文を閲覧することにより、用法を確認できる。

例えば、“precision”について述べたい場合、“precision * %”と入力して、以下の用例を検索することができれば、“a precision of ~%”や“the precision is improved to ~%”といった言い回しが用いられていることが分かる⁵。この結果から、「precision が~%である」と表現したい場合は“a precision of ~%”という表現が可能であり、「precision が改善され~%となる」と表現したい場合は“the precision is improved to ~%”という表現が可能であると分かる。

“precision * %”で検索した結果：3,353 件
・ ... lexicon was found to have a <u>precision of 77%</u> .
・ The <u>precision is improved to 62%</u> by refining ...

4 システムの設計

本節では、前節で述べた問題に対して、検索機能と検索結果の提示方法の2つに分けて、その設計方針を述べる。

4.1 検索機能

まず、英作文における問題に対して、用例検索が提供すべき検索機能について述べる。

統計表示 統計表示とは、ある表現を含む用例に対するコーパス中での出現数を表示する機能である。前

節で用いた例における検索文数がこれに相当する。この機能によって、出現数の多い表現をより一般的な表現であると判断することができる。

拡張ワイルドカード検索 拡張ワイルドカード検索とは、ワイルドカード部に任意の単語列が入ることを許して、用例を検索する機能である。拡張ワイルドカードとは、1つから複数の任意の単語に対応する記号であり、これをアスタリスク (*) を用いて表記する。一般的なワイルドカードでは対応する語数を選択できないのに対して、拡張ワイルドカードは対応する語の上限数を学習者が自由に設定できる。これにより、ワイルドカード部分に必要以上に多くの語が対応してしまうことを防ぐ。

品詞検索 品詞検索とは、特定の品詞の語を含む用例を検索する機能である。拡張ワイルドカード検索では検索文数が多すぎる場合に、目的の表現を発見するのが困難になる。品詞検索を活用して、検索文数を絞り込むことによって、この問題に対応する。3.2.1 で示した例のように、クエリに [verb] といった品詞を含めることにより品詞検索を行う。

類義語検索 類義語検索とは、特定の語を含む用例だけでなく、その類義語が含まれている用例も検索する機能である。品詞検索より更に検索結果を絞り込みたい場合には、この機能が有効である。類義語検索で用例を検索する場合は、言い換えたい語（例えば“results show that”の“show”）に類義語検索を行う記号として“+”を付与（例えば“results show+ that”）して検索するとする。

4.2 検索結果の提示方法

用例検索を利用するという観点では、検索結果の提示方法を検討することも重要である。

用例の提示方法 用例を一覧表示する際、図2に示すように、クエリと一致した部分を中心に配置し、前後に文脈を表示する。これをKWIC (Key Word in Context) 方式と呼ぶ。

用例の仕分けの方法 拡張ワイルドカード検索、品詞検索、類義語検索などで語を指定せずに用例を検索した際、拡張ワイルドカードや品詞に埋めた語で仕分けし、それぞれの統計情報を表示して欲しいことがある。例えば “[adj] idea” というクエリで検索を行った場合、図2に示すように、“[adj]”の部分に当たる“main”や“key”といった語で用例を仕分けし、それぞれの統計情報を表示する。「用例を表示」のボタンを押すことにより、その仕分けに属する用例を見ることができる。

⁵*はワイルドカードを表し、任意の単語が入る。

用例の並べ替えの方法 図2に示すように、「main idea を含む用例」や「key idea を含む用例」といった仕分けごとに、検索文数が多い順に並べ替える。また、用例は、クエリと一致した部分の前後でアルファベット順に並べ替える。このとき、前後どちらで並べ替えるか、また、昇順と降順のどちらで並べ替えるかは学習者が選択できるように設計する。

5 システムの実装

前節で述べた機能を中心に実装した、英作文支援のための用例検索システムを図2に示す。利便性のために、検索結果の提示方法を自由に選択できるように実装した。

6 おわりに

本研究では、英作文に関する文献に基づいて、英作文の際の問題を用例検索の観点から分類し、用例検索システムによる支援の可能性を調査した。また、その結果に基づき、英作文支援に必要な機能を検討し、十分な評価に耐えうる用例検索システムを設計・実装した。

今後は、実際の英語教育の現場と連携して、用例検索システムを英語の授業に取り入れることにより、更に大規模かつ長期的な評価を行う必要があると考えられる。また、評価に伴い、用例検索によってどのように英作文時の問題に対応できるかについて、更に考察し改善していく必要があると考えられる。

謝辞

本研究は、文部科学省科研費（23240018）、文部科学省科研費（23700159）、および JST 戦略的創造研究推進事業さきがけの一環として行われた。

参考文献

- [1] 齊藤俊雄, 赤野一郎, 中村純作 (編). 英語コーパス言語学—基礎と実践—. 研究社出版, 1998.
- [2] 西垣知佳子, 天野孝太郎, 吉森智大, 中條清美. 中・高生のためのコンコーダンス・ラインを利用したデータ駆動型英語学習教材の開発の試み. 千葉大学教育学部研究紀要, 第 59 巻, pp. 235–240, 2011.
- [3] 能登原祥之. 収斂型コンコーダンスの教育効果—日本人英語学習者（大学生初級・中級）の場合—. 広島大学大学院教育学研究科紀要, 第 58 巻, pp. 165–174, 2009.
- [4] 中條清美, 西垣知佳子, 内堀朝子, キャサリン・オヒガン. データ駆動型学習による効果的な英語初心者向け文法指導の試み. 日本大学生産工学部研究報告 B, 第 41 巻, pp. 15–33, 2008.
- [5] 遠田和子. Google 英文ライティング. 講談社インターナショナル株式会社, 2009.
- [6] 杉本淳一. Google 英語勉強法. 日本実業出版社, 2011.
- [7] 衣笠忠司. Google 検索による英語語法学習・研究法. 開拓社, 2010.



図 2: 用例の仕分けと並べ替え

- [8] Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In *Language Resources and Evaluation Conference (LREC 08)*, 2008.
- [9] 松原茂樹, 酒井佑太, 小澤俊介, 杉木健二. 学術論文からの英語表現集の自動生成. 第 7 回情報プロフェッショナルシンポジウム, pp. 41–45, 2010.
- [10] 綱島祐一, 川崎優太, 安藤一秋. 検索エンジンを用いた英作文支援ツール. 信学技法, 第 106 巻, pp. 87–92, 2007.
- [11] David Wible and Nai-Lung Tsao. Stringnet as a computational resource for discovering and investigating linguistic constructions. In *The NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, pp. 25–31, 2010.
- [12] 松原茂樹, 江川誠二, 加藤芳秀. 英文用例検索システム ESCORT: 論文データベースを用いた図書館サービス. 第 4 回情報プロフェッショナルシンポジウム, pp. 125–129, 2007.
- [13] 大名力. 学習者用オンライン英文検索システムの構築. 群馬大学社会情報学部研究論集, 第 7 巻, pp. 125–142, 2000.
- [14] Martin Chodorow, Michael Gamon, and Joel Tetreault. The utility of article and preposition error correction systems for English language learners: Feedback and assessment. In *Language Testing 2010*, 2010.
- [15] 中村哲三. 英文テクニカルライティング 70 の鉄則. 日経 BP 社, 2011.
- [16] Ann M. Korner. 一流の科学者が書く英語論文. 東京電機大学出版局, 2010.
- [17] 原田豊太郎. 理系のための英語論文執筆ガイド. 講談社, 2002.
- [18] 阿部一. 9 割の日本人が使い方を間違える英単語 101. ジャパンタイムズ, 2011.
- [19] James H. M. Webb. 日本人に共通する英語のミス 151. ジャパンタイムズ, 2006.