

英文作成のための英文検証・用例検索支援ツールの試作

市原 加奈子[†] 多田 沙織[†] 溝渕 昭二[‡] 安藤 一秋[†]
[†]香川大学工学部 [‡]近畿大学理工学部

1 はじめに

近年、社会のグローバル化やインターネットの普及に伴い、英語で書かれた文書を読んだり、英語で作文したりする機会が増加した。一般に日本人は英文の読解に比べ、英作文が不得意であるといわれる。英語で作文する場合、少なくとも一度は、「この英語で通じるのか」、「この表現であっているのか」などの疑問、不安を覚えたことがあるだろう。

英作文が不得意な人は、日本語で書きたい内容を考え、それから英文に翻訳するステップを踏む傾向がある。その際、単語を訳すために辞書を参考にすが、単語の定義や訳だけの参照で作文することは難しい。そこで、辞書中の用例を参照し、動詞と目的語の繋がりや熟語、表現などを把握する。専門性・分野性のある場合は、一般の辞書を利用するより、英語論文や英字新聞、ビジネスレターなどの用例が適している。また、英文作成後にも、正しい英文・表現かどうかを用例と照らし合わせて検証することもある。このように、用例は英作文において必要不可欠な情報源となる。

この考えを基に、コーパスから用例を検索して提示する英文作成支援システムも数多く研究されている[1][2]。静的コーパスを対象とした用例検索では、基本語や基本表現（高頻度語や高頻度表現）の普遍的な用例の検索には有用である。しかし、母集団のデータが制限されるため、基本語（表現）以外の低頻度語（表現）や新語、新表現などを検索する場合、見つからない、参考となる具体例が存在しないなどの問題が生じる。

そこで、近年では、Web上のデータをコーパスとして利用する研究が注目を浴びている。Web上のデータは内容が日々増え続け、自動更新される動的なデータである。また、様々な分野に渡る膨大なデータが蓄積されているためデータ量の問題も解消される。一方で、質的な問題があるため得られたデータが全面的に信用できる保証はない。しかし、英文作成に利用する場合は一意の答えを求めるのではなく、あくまでも参照という立場を取れば問題はない。

Webを対象とした代表的な研究として、

GoogleFight[3]や Webcorp[4]などが挙げられる。しかし、これらは2つの異なる語のヒット件数の比較のみに限定されたり、KWIC形式の表示しかできなかったりと制約が多い。単純にサーチエンジンのみを利用して用例検索する方法[5]も有用であるが、必要な用例を見つけ出すには大量の検索結果から人手で分類、集計、抽出などの処理が必要となる。

そこで、本研究では生きた用例が存在するWeb上のデータを検索対象にして、検索結果を統合・整理し、そこから用例やその他の必要情報を抽出することで、英作文に利用しやすい形で情報提示を行うことを目的とする。本稿では、既存のサーチエンジンを利用した英文検証・用例検索ツールの試作とその評価について報告する。支援ツールの特徴としては、(1)複数回の検索が一度で行える；(2)複数の検索結果を統合し、ヒット数でランキング表示できる；(3)妥当な用例を優先的に抽出できる；(4)抽出した用例文をKWIC形式で表示できる；などが挙げられる。

2 英作文のミス調査と支援機能

英作文を行う際に、どのようなミスが多いのかを分析し、支援ツールで実装する機能の手掛りとする。ミスの分析には、成田らが実施した「日本人学生が書いた英文のエラー分析」[6]を利用する。この調査は、英文学科の学生59名を対象にして、“traveling”というトピックだけを与えて自由に英文を書かせた結果を分析したものである。成田らの調査結果を表1に示す。

表1より、英作文で最も多いミスは冠詞に関する誤りで約19%存在することがわかる。一般的に、冠詞の選択は、辞書を参照してcountableかuncountableを判断することになるが、名詞によっては両方の属性をもつ場合がある。Web上のデータを利用すれば、実例と共にヒット数が判断材料となる。続いて、名詞や動詞の語彙選択ミスが約14%を占めている。名詞と動詞の語彙選択は、互いの相性（共起）が有益な情報となるため、用例の参照が効果的である。スペリングミスに関しても、ヒット数を参考に検討できる。また、サー

表 1 自由英作文に頻出したエラー内訳 (上位 5)

エラーの内容	比率 (%)
名詞の可算性に関する誤り (不定冠詞の付け忘れ, 誤使用)	11.8
スペリングミス	10.2
動詞の補部構造に関する誤り (関係節の間違い)	8.8
名詞の語彙選択の誤り	7.3
動詞の語彙選択の誤り	7.1
限定詞 (特に冠詞) の脱落	7.1

(文献[6]より引用)

チェンジンによっては, スペルを自動補正して提示するものもある.

以上より, 作文ミスの緩和には用例が有用であるといえる. しかし, Web 上のデータは大規模であるため, 多いときには一億件以上の検索結果から適切な用例を見つけ出す必要があり, 人手では非常に困難である. そこで, 検索結果から妥当な用例を抽出して, 閲覧しやすい形で表示する機能が必要となる. さらに, 検索したい候補が複数ある場合は, その候補の数だけ検索を繰り返さなければならない. たとえば, 「She does not drink () coffee.」という文に対し, 括弧内に「a little/many/a few/much」のどれがあてはまるかわからない場合を考える. まず, それぞれの候補を当てはめた 4 つの検索式を作成し, 4 回の検索を行う必要がある¹. その後, それぞれの検索結果からヒット数や用例を参考に最終判定を行う必要があり, 情報の取捨選択に労力を要する. そこで, 1 度の検索入力で複数の検索を行い, その結果を統合したデータから妥当な用例を抽出して, ヒット数と共に表示する機能が必要となる.

3 試作ツールの実装機能

2. での分析を基に試作ツールの実装機能について説明する. 検索機能には, 一般的なサーチエンジンで利用できる AND, OR, フレーズ, ドメイン指定などの他, 「翻訳に役立つ Google 活用テクニック」[5]を参考に, 用例検索と結果表示に有用と判断した機能を実装した. 本ツールにおける用例の定義は, 完全な 1 文とした. 実装した機能を表 2 に示す.

¹ 「She does not drink (a little OR many OR a few OR much) coffee」では, 結果がまとめて表示されるため, ヒット数が参照できず, ランキング順位を基準に判定することになる.

表 2 実装機能

機能	内容
動詞の活用形の自動生成 (選択機能)	動詞が入力されると, その活用形を自動生成する.
ワイルドカード (*)	単語を抽象化して検索できる.
*の候補指定	*に対応する具体的な候補を指定できる.
*の品詞指定 (冠詞/前置詞)	*の候補を冠詞あるいは前置詞に限定して検索できる.
類似用例文の表示	検索結果から類似用例文 (完全一致, 出現順一致, AND) を抽出して表示する.
結果表示の指定 (ランキング/KWIC)	ヒット数に基づいたランキング表示と KWIC 表示が指定できる.

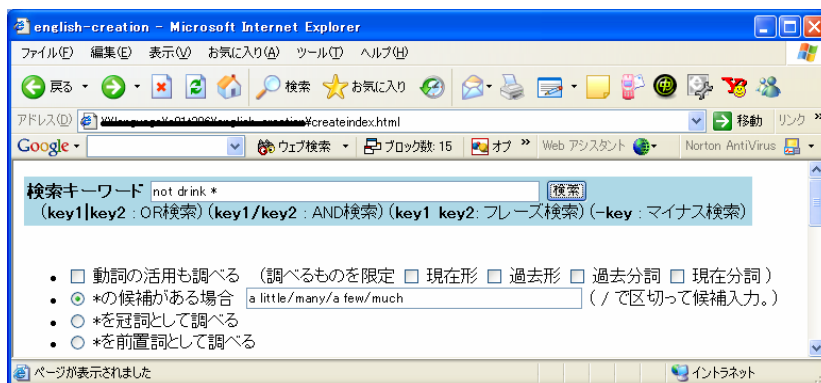
試作ツールの入出力画面を図 1 に示す. 図 1(a) は「She does not drink () coffee」という文に対し, 4 つの候補の使用用例を調べる例である. ユーザが検索キーワードに「not drink *」, *の候補に「a little/any/a few/much」を入力し, 検索した結果を図 1(b)に示す. 検索結果より, 「much」のヒット数が 3,370 件でランキング 1 位となっていることがわかる. 2 位以下とは 1 桁の差があることなどを参考に解を決定すればよい.

4 評価実験

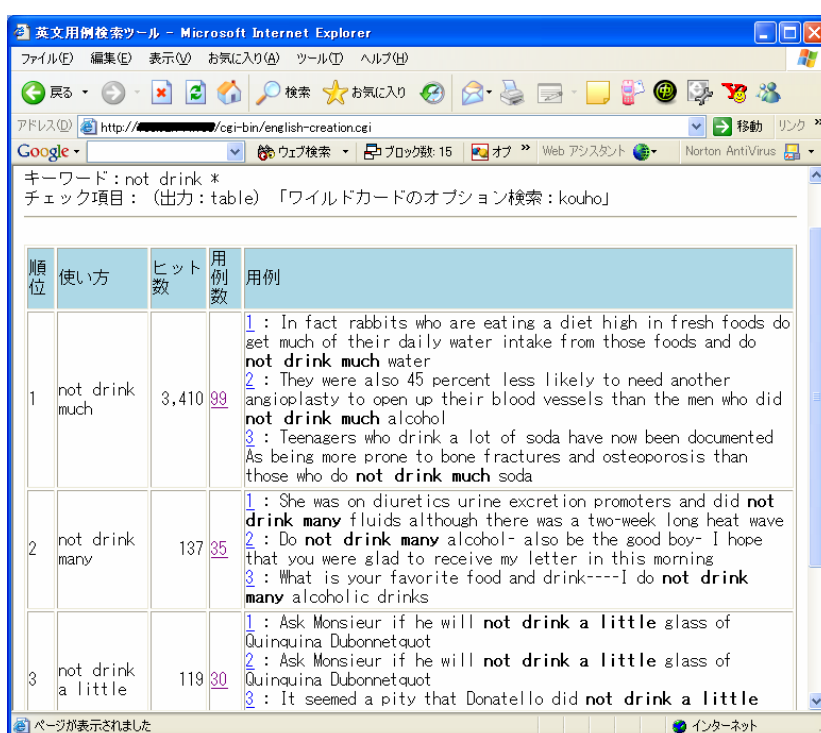
4.1 実験方法

5 名の学部生を対象に, 通常のサーチエンジンと試作ツールを利用して大学入試レベルの問題 [7][8]に解答してもらい, それぞれの正解率と解答時間, アンケートを基に試作ツールの有用性の評価を行う.

ツールの最大の特徴は, 複数回の検索を一度に実現できる検索の効率化と検索結果からの用例抽出やヒット数によるランキング, KWIC 表示などの情報提示である. そこで, 検索の効率化とヒット数に基づくランキングの有用性を確認するために 4 択穴埋め問題を設定した. また, 用例抽出や KWIC 表示など情報提示の有用性を確認するために, 英単語の並び替え問題を設定した. 実験には, レベルが等しい問題を 20 問ずつ抽出し, 10 問ずつに分類して利用した. また, ツールによる出力は, 検索時間の関係で, 用例およびヒット数共に 3 件までとした.



(a) 試作ツールの入力インターフェース



(b) 検索結果 (ランキング表示)

図 1 入出力画面

4.2 結果と考察

サーチエンジンとツールを使った場合の平均正解率と一問あたりの平均解答時間を表3に示す。表中の間1は4択穴埋め問題で、問2が並べ替え問題である。また、実験に関連するアンケート結果を表4に示す。

表3より、ツールの平均正解率が若干高いことがわかる。これは検索結果から適切な用例が抽出できたことを示すと考えられる。この点を詳細に分析するために、正しい検索式が入力された場合、間違った用例表示が行われていないかを調査し

た。その結果、正しい検索式が入力された場合は100%の確率で上位3件に解答の参考となる情報が提示されていることがわかった。解答を間違えたものは、検索式の設定がおかしく、適切な用例を検索できなかったようである。これは検索式の設定に関する支援が必要であることを意味する。

一方、解答時間に関しては、サーチエンジンとツール間でほとんど差がなく、ツールを利用することによる優位さは見られなかった。検索を実行する時間は互いにほとんど差がないため、閲覧性と検索式の設定に問題があると考えられる。しか

表 3 平均正解率と一問あたりの平均解答時間

	平均正解率 (%)	一問あたりの平均 解答時間 (分 / 問)
問 1 サーチエンジン	78	1.96
ツール	86	2
問 2 サーチエンジン	72	2.84
ツール	84	2.54

表 4 アンケート結果

(1:良い 2:ほぼ良い 3:あまり良くない 4:良くない)

アンケート内容	平均
①思い通りの結果を得られたか	2
②検索機能を理解できたか	2
③出力画面は閲覧しやすかったか	1.2
④ランク 1 位が解答だったか(4 択)	2.2
⑤表示ランクは 3 位までで十分か	2.2
⑥表示用例数は 3 文までで十分か	1.6

し、閲覧性に関しては表 4 の③が良い結果を示していることから、検索式の設定に原因があると考えられる。今回のツールは試作段階で、検索式の指定の際、一般のサーチエンジンでは指定できるが、ツールでは指定できないといった制約があった。この制約が解答時間に大きく影響していると考えられる。

表 4 のアンケート結果から、出力画面の閲覧性の評価が高いことがわかる。また、④と⑤が多少低い結果であることは、ヒット数 1 位の候補が必ずしも正解を表していないことを示す。これは、Web をコーパスとして利用する際の制約でもある。用例の抽出・ランキング方法に更なる工夫が必要である。

更に、便利な機能や欲しい機能、改善点に関するアンケートを行った。便利な機能としては、ワイルドカードの候補指定や KWIC 形式の情報表示など、ツールの特徴が挙げられた。欲しい機能としては、ワイルドカードの品詞指定の候補拡張、用例とその対訳表示の追加などがあった。また、ツールの改善点として、検索時間がワイルドカードの候補指定数に比例して長くなる点、出力画面に新規検索 BOX の作成などが挙げられた。

今回のツールは試作段階であったため、時間効率の面では有用性を示すことができなかった。しかし、複数回の検索を同時に行い、候補を統合し、ランキング付けしたり KWIC 形式で表示したりする機能は、アンケート結果より高い評価を得た。

5 おわりに

本稿では、既存のサーチエンジンを利用した英文検証・用例検索ツールの試作について報告した。本ツールの最大の特徴は、複数回の検索を一度に実現できる検索の効率化と検索結果からの用例抽出やヒット数によるランキング、KWIC 形式で表示などの情報提示である。提案手法の有用性を調べるために評価実験を行った所、複数回の検索を同時に行い、候補を統合し、ランキング付けしたり KWIC 形式で表示したりする機能は有用であることを確認できた。

今後の課題として、検索機能の改善はもちろん、初見のユーザに対して易しい入力インターフェースの実現と検索ログを利用した新たな展開などが挙げられる。また、本研究と並行して、ユーザの辞書引きに対する労力軽減と訳語選択支援を目的とした研究を行っており、それらを統合したシステムの実現を目指す。

謝辞

本研究は、文部科学省科学研究費補助金（若手研究 (B) 16700562）の補助を受けて実施した。

参考文献

- [1]武田, 古郡, “例文をもとにした英文書作成支援システム”, 情報処理学会論文誌, Vol.35, No.1, pp.53-61, 1994.
- [2]三好, 越智, 金西, 岡本, 矢野, “英作文支援における句構造情報を利用した用例検索ツール”, 日本教育工学雑誌, 27(3), pp.283-294, 2003.
- [3]Googlefight, <http://www.googlefight.com/>
- [4]WebCorp, <http://www.webcorp.org.uk/>
- [5]安藤, “翻訳に役立つ Google 活用テクニック”, 丸善株式会社, 2003.
- [6]成田, “英文作成支援環境の構築”, COE 形成基礎研究費研究成果報告『先端的言語理論の構築とその多角的な実証—ヒトの言語を組み立て演算する能力を語彙の意味概念から探る』, 2000. <http://coe-sun.kuis.ac.jp/coe/>
- [7]上垣編著, 即戦ゼミ 8 項目別征服 入試頻出英語問題基礎演習[新選改訂版], 桐原書店, 1990.
- [8]瓜生, 篠田編著, Nextage 英文法・語法問題, 桐原書店, 1999.