

インターネット言語資源を利用した英文検証支援ツール

安藤一秋[†] 綱嶋祐一[†] 溝淵昭二[‡]

[†]香川大学工学部 [‡]近畿大学理工学部

1. はじめに

近年、電子化コーパスに含まれる用例データが英語教育や英作文の支援に利用されるようになった[1]。静的なコーパスは、普遍的な用例の利用には適しているが、収録文数や異なり語彙数など、データ数が制限される問題がある。そこで、近年、Web上の膨大なテキストデータをインターネット言語資源（Webコーパス）として利用する研究が注目されている。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データといえる。一方で、質的な問題も避けては通れないが、膨大な情報量という魅力は大きい。

Webコーパスを英作文に利用する研究[2-4]もいくつか提案されているが、最適候補が提示できない、検索式が自動生成できないため初心者向きでない、利用する人の英語力が問われる等、検索時または支援においていくつか問題がある。

本研究では、信頼性の面からWeb上のデータだけでなく、オンライン辞書や新聞・論文アーカイブなど電子化された言語資源をフル活用して、英文作成のための有益な情報提供を目的とした英文作成支援システム[5]を構築中である。本稿では、そのシステムの内、Google Web APIs (Google API)[6]を利用したWeb・用例検索モジュールについて説明する。本モジュールは、語彙選択や英文・表現の検証支援、用例提示などを行うものであり、Perl CGIとして実装した。

2. Web・用例検索モジュールの概要

現在開発中の英文作成支援ツールは、Webアプリケーションであり、英文作成、Web・用例検索、マイニングの3つのモジュールで構成される[5]。英文作成モジュールは、入力された和文・表現を形態素解析し、各種の電子化言語資源またはそこから抽出された情報を利用して、対訳候補をランキング提示する機能や有用サイトへのリンク、英文テンプレートなどを提供する。マイニングモジュールは、検索ログをマイニングすることで、苦手な単語や表現、構文などを分析する。また、単語・表現帳などを自動生成することで、英語の学習支援を実現する。図1に、Web・用例検索を示すが、本稿では、Web検索部を中心に説明する。

図1中の入力インタフェースを通じてユーザが

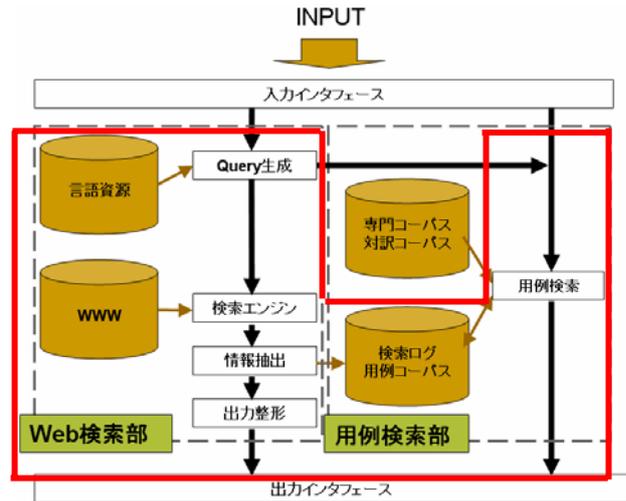


図1 Web・用例検索モジュールの構成

検証したい表現や文と共に要望に応じたオプションをチェックして検索を実行すると、Query生成処理により、検索式が自動生成される。英文の検証機能は、大鹿ら[4]と同様に「翻訳に役立つGoogle活用テクニック[7]」から検索式が自動生成できるものを実装した[5]。このテクニックの基本は、検索結果から得られるヒット数やsnippetを利用して妥当性を検証するという考えである。実装機能としては、自動フレーズ検索、ワイルドカード及びその候補指定、電子化辞書を利用した類義語や活用形の展開、言語・ドメイン・サイト指定などがあり、これらを活用して検索・検証が行える。

自動生成された検索式は、Google APIを通じてGoogleから検索結果を取得する。次に、検索結果から有用な情報（ヒット数や用例など）を抽出すると共に検索式や用例などは検索ログとして蓄積される。このログは、従来手法のようにキャッシュとしても利用されるが、用例検索あるいはデータマイニングのデータとして利用される点が従来システム[2-4]と大きく異なる。

通常、複数の検索式で実行した場合、複数のウィンドウで結果が表示されるが、本モジュールでは、検索結果を統合し、ユーザが閲覧しやすい形（ヒット数でランキングされたテーブル表示、KWIC表示、ドメイン分類、n-gramヒット数表示など）で提示される。

3. 動作例

前置詞の選択問題（He opened the door (by / in / on / with) a key.）を使って動作を説明する。

A Support Tool for Verifying English Sentences using Language Resources on the Internet

[†] Kazuaki Ando, Yuuichi Tsunashima · Kagawa University

[‡] Shoji Mizobuchi · Kinki University

まず、図 2 に示すように、入力インタフェースに文の一部（あるいは全文）、表現などを入力し、オプションの設定を行った後、検索を実行する。

複数の検索式で検索を行う場合、通常は複数のウィンドウ（またはタブ）に表示される内容と比較しながら検証を行う必要がある。本ツールでは、図 3 に示すように、ヒット数順のテーブルの形で提示される。この結果を見ると、解として、“with”の妥当性が高いことがわかる。ここで、図 3 中の“opened the door by a key”をクリックすると、図 4 のようにドメイン毎に分類された用例文が閲覧できる。また、KWIC オプションを設定していれば、図 5 のように、KWIC 形式でも閲覧できる。

また、図 6 に示すように、単語列を n-gram に分割し、そのヒット数を利用して妥当性を検証する WebLEAP[2]型の表示もオプションで選択できるように実装した。更に、誤った箇所を特定しやすくする工夫として、図 7 に示すような接続スコアに基づく要約出力も作成した。この出力法は、まだ改良の余地があるが、WebLEAP にない機能である。

複数語検索したい場合は、検索語を*で区切って入力してください(ワイルドカードは*で記入)
 入力: opened the door * a key

◎入力を単語、フレーズとして検証を行う
 *の候補の入力(※で区切ってください) by:with|with

検索対象 ◎英語圏(米、英)のページだけを検索 ◎すべてのページを検索
 Webから検索する ログから検索する

表示: ◎ヒット数、用例文、リンクを表示する ○ヒット数のみを表示 ○KWIC表示

検索 クリア

図 2 入力インタフェースの例（一部）

Rank	Key	Hit count
1	opened the door with a key	152
2	opened the door by a key	4
3	opened the door on a key	0
4	opened the door in a key	0
ALL COUNT		156

図 3 表形式の検索結果表示

Domain = info	
Title	snippet
The Blossom and the Fruit 02	Fleta opened the door by a key that was attached to a chain hanging from her waist, and she closed it behind her. Hilary asked no questions, for she seemed ...

Domain = uk	
Title	snippet
The Proceedings of The Old Bailey, London 1674 to 1834	Q upon your solemn oath, did You mean it was Mc PHERSON - a I did mean Mc PHERSON, he opened The door by a key. Q he had The key by The authority of

図 4 ドメイン分類された用例文（抜粋）

Domain = com	
(3d) The student is falling. We now turn to opened the door with a key the formalization of a proposition involving a three-place predicate. (4) John	The major opened the door with a key and walked in. I followed. There was a small table against the wall and two chairs nothing else.
After Tyrone opened the door with a key , he repeated his demand that Perkins get off the property, stepped inside, and let th door (which had an automatic	

図 5 KWIC 表示された用例文（抜粋）

opened	the	door	by	a	key	HitCount
the	door	by				26100
door	by	a				3320
by	a	key				12600
opened	the	door	by	a	key	728
the	door	by	a			1780
door	by	a	key			16

opened	the	door	with	a	key	HitCount
the	door	with				119000
door	with	a				53600
with	a	key				69100
opened	the	door	with	a	key	2420
the	door	with	a			20100
door	with	a	key			926

図 6 n-gram ヒット数表示（抜粋）

Number 4 + 5		
Keywords		score
opened the door by	the door by a	0.38938
opened the door with	the door with a	2.00000

Number 4 + 6		
Keywords		score
opened the door by	door by a key	0.31811
opened the door with	door with a key	2.00000

Number 5 + 6		
Keywords		score
the door by a	door by a key	0.10584
the door with a	door with a key	2.00000

図 7 接続スコアに基づく要約出力（抜粋）

以上のような操作を行うことで、英作文に有用な情報の取得と作成した英文の検証が行える。

4. おわりに

本稿では、インターネット言語資源を利用して、作成した英文あるいは表現の検証を支援するツールを提案した。今後は、使いやすいインタフェースの考案および評価実験を行うことで、本ツールの有用性を評価すると共に、Web サービスとしての公開を目指す。また、英語教育に携わっている教員や研究者の意見を参考に更なる機能拡張を施し、最終的には、日本語を始め多言語の作文支援ツールを構築する予定である。

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] 齋藤 他：改訂新版英語コーパス言語学，研究社，2005。
- [2] Yamanoue et al. : “Learning Usage of English KWICly with WebLEAP/DSP”, Proc. of the 2nd International Conference on Information Technology for Application, pp.380-385, 2004.
- [3] 相原 他： “英作文支援システムの構築”，情報処理学会第 67 回全国大会論文集，5P-7, 2005 年 3 月。
- [4] 大鹿 他： “Google を活用した英作文支援システムの構築”，DEWS2005, 4B-i8, 2005.
- [5] 安藤 他： “言語資源と検索エンジンを利用した英文作成支援ツール”，教育システム情報学会第 30 回全国大会論文集，p.251-252, 2005 年 8 月。
- [6] Google Web APIs : <http://www.google.com/apis/>
- [7] 安藤進： “翻訳に役立つ Google 活用テクニック”，丸善，2003。