

研究種目：若手研究（B）  
 研究期間：2007～2009  
 課題番号：19700638  
 研究課題名（和文） 第二言語作文のための有益な情報提示を目的とした多言語作文支援ツール  
 研究課題名（英文） A Multilingual Writing Support Tool for Providing Useful Information about Writing in Second Languages

研究代表者  
 安藤 一秋（ANDO KAZUAKI）  
 香川大学・工学部・准教授・  
 研究者番号：60325321

研究成果の概要（和文）：本研究では、第二言語の基本的な語彙や文法を修得したユーザに対して、第二言語での作文時に有用な情報を提示するために、Web 検索エンジンを活用した多言語作文支援ツールを開発した。本研究では、検索エンジンのヒット数を Web 上での疑似的な利用頻度とみなして利用する。ユーザが書いた文やフレーズの妥当性を検証したり、ユーザの語彙レベルに基づいて用例文を提示したりする方法を提案した。

研究成果の概要（英文）：In order to provide useful information for writing in second languages to users who learned basic vocabulary and elements of grammar, we developed a multilingual writing support tool based on a Web search engine. We assume that the total number of search results (called the hit count) is the pseudo-frequency of usage. The functions which verify given sentences and phrases by the hit count and provide example sentences based on learner's vocabulary level etc., have been proposed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,200,000	0	1,200,000
2008年度	1,200,000	360,000	1,560,000
2009年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,200,000	600,000	3,800,000

研究分野：自然言語処理，情報検索

科研費の分科・細目：科学教育・教育工学・教育工学

キーワード：作文支援，多言語，検索エンジン，Web コーパス，語学学習支援，校正支援

#### 1. 研究開始当初の背景

インターネットの普及により、国際共通語である英語だけでなく、中国語やスペイン語など母国語以外の言語に接する機会が増加してきた。それに伴い、英語以外の言語を第二言語あるいは第三言語として習得する必

要性も高くなってきた。

第二言語での作文を支援する研究は古くから行われている。近年は、電子化言語資源が整備されてきた状況や書きたい内容または構造に類似した用例の提示が作文に有効であるなどの理由から、用例に基づく作文支

援が注目を浴びている。従来の用例に基づく支援は、辞書や単言語・対訳コーパスなどの静的言語資源を利用するものが多い。静的言語資源は、基本的な語や表現（高頻度の語や表現）など普遍的な用例を検索するために有用である。しかし、用例の種類や数が制限されるため、自分が望む用例が見つからない、新しい語や新しい表現を用いた用例が含まれていないなどの問題がある。

これらの問題を解決するために、Web上のテキストデータをコーパス（Webコーパス）として利用する研究が注目されている。Webコーパスは、多言語で記述された無尽蔵の用例をもち、内容も日々更新される生きた知識源である。また、様々な分野に渡る膨大なテキストデータが蓄積されているため量的問題も解消される。一方で、記述に関する制約がないため、得られたデータが全面的に信用できる保証はない。この点には注意して利用する必要がある。

Webコーパスから用例を検索する代表的な方法として、既存の検索エンジンの利用が挙げられる。近年では主要な検索エンジンのAPIが公開され、研究に利用しやすい環境が整ってきた。また、検索エンジンのヒット数は、データの質を客観的に判定する指標になるだけでなく、用例の汎用性や妥当性を検証する指標にもなる。

以上の考えに基づき、研究代表者は検索エンジンを利用した英文作成支援ツールの構築を進めてきた。提案したツールの検索機能としては、検索要求から適切な検索式を自動的に導出する機能、ワイルドカードを利用した検索式の抽象化およびその候補指定、英文検証支援のためのn-gram分割検索、言語・トップレベルドメイン・サイト指定などを実装した。また、提示に関しては、語・表現・用例の自動抽出、ヒット数に基づく語・表現・用例のランキング、KWIC表示、学習者の語彙レベルに応じた用例ランキング、学習者が書いた英文の誤り箇所候補の表示などを実装した。これらの機能を利用することで、用例の取得だけでなく、単語や表現の比較、英文・表現の検証などが容易に行える。

検索エンジンを利用して作文支援や用例検索を行う関連研究は、いくつか存在する。代表的なものとして、山名（早稲田）らの研究とWebCorpやGoogle Fightなどが挙げられる。しかし、検証機能が十分でない、言語が英語だけに限定される、2つの異なる入力語しか比較できないなどの制約が多い。また、研究代表者の知る所、多言語での作文支援に関する研究はない。

作文支援が目的ではなく、多言語用例の検索に特化した研究としては、田中ら（東京大学）のKIWIや端言がある。しかし、KIWIや端言の機能は、用例検索特化されているため、

複数ある訳語・対訳表現候補の比較や学習者が実際に書いた文の検証など、作文に対する支援機能を備えていない。本研究において提案する多言語作文支援ツールは、用例検索の他に比較や検証に対する支援も行う。

以上より、本研究は新規性・有用性が高く、独創的な研究であるといえる。

## 2. 研究の目的

本研究では、Web検索エンジンと既存の言語処理技術・資源を有効に活用することで、用例検索・提示および産出文の検証に主眼をおいた多言語作成支援ツールを構築する。

提案する支援ツールは、自動翻訳や自動誤り訂正などシステム主導の支援が目的ではなく、あくまで学習者が主体となり、自発的・訓練的に作文が行える環境の整備が目的である。ツールが対象とする言語は、Web上で利用されている主要な言語とするが、解析技術の言語依存性を考慮して、依存・非依存で実装機能の対象言語を分類する。

提案する支援ツールの機能を以下に示す。

### (1) 用例の自動抽出・提示機能

作文を行う際、ある語・表現の用例を確認したい、自分が書きたい内容に類似した用例を調べたいなどの要望がある。そこで、検索結果から用例を自動抽出して提示する機能を実装する。特に、学習者の言語習得レベルに適応した順で用例を提示する機能やKWIC形式での提示など、作文に利用しやすい形での提示を実現する。

### (2) 柔軟な検索支援の提供

作文の際、ある語や表現の一部分なら覚えているが、その情報だけでは辞書を検索することができない、一部分を思い出しても新しい用語であるため、辞書に載っていないなどの状況が考えられる。そこで、検索エンジンがもつ特徴的な検索機能（フレーズ検索、ワイルドカードなど）を活用し、学習者からの様々な要求に柔軟に対応できる検索支援を実現する。

訳語として、どの単語・表現を使えばよいのか分からない、汎用的な語や表現が知りたい。前置詞、動詞などの品詞レベルで語や表現の使い方を比較検討したいなどの要望に答えるため、学習者が入力した比較候補の情報から検索式を自動生成して検索し、検索エンジンから得られる複数の結果を統合して、ヒット数でランキング提示する機能を実装する。

### (3) 誤り候補箇所の提示

ユーザが書いた文や表現が正しいのかどうか分からない、間違えている部分を特定したいなどの検証支援を行うために、誤り箇所

の候補を提示する機能を実装する。この機能により、どのように検証すればよいかわからない場合にも検証箇所の候補が特定できるため、第二言語の知識が乏しい学習者に対するサポートが実現できる。

#### (4) 利用しやすいインタフェースのデザイン

訳語候補の比較のように、あらかじめ利用が想定できる状況においては、学習者がキーワードやフレーズ、比較候補などを入力するだけで、検索要求の内容から適切な検索式をサジェストする機能を実装する。

#### (5) 検索ログの活用

検索ログはユーザの試行錯誤情報を含んでいると考えられる。そこで、検索ログを分析することで、作文に有用な情報を抽出する機能を実装する。

### 3. 研究の方法

多言語作文支援ツールの考案・実装にあたり、英文作成支援ツールの構築で培った知識と技術が応用できる。しかし、単言語向けの技術を多言語向けに応用するためには、支援ツールに実装する機能の言語依存性を考慮する必要がある。我々が提案した英文作成支援ツールの骨格は、パターンマッチング技術や検索技術で構成されるため、言語解析技術や辞書が十分整備されていない言語に対しても応用可能である。

以下では、具体的な研究方法について説明する。

#### (1) 多言語化に対する予備調査と新機能の考案

作文支援ツールにおいて重要な情報の1つにヒット数がある。Web上であまり利用されない言語に対しては、ヒット数が少なくなってしまう、性能があまり発揮できない。そこで、Web上で使われている主要言語を中心とした調査・整理を行う。また、代表的な言語に対する作文支援、学習支援に関する文献サーベイを行い、新たな機能の考案も行う。

#### (2) 用例の自動抽出・提示手法の考案

ツールの基本機能となる用例の自動抽出・提示機能の実装を行う。この機能は、学習者の検索要求にマッチする用例を検索結果から自動抽出し、提示するものである。提示法としては、KWIC形式のほかに、学習者の語彙習得レベルに適応した順で用例を提示する機能など利用しやすい形式を模索する。

#### (3) 柔軟な検索法の考案

主要検索エンジンがもつ特徴的な検索機能（フレーズ検索、ワイルドカードなど）を

整理する。また、検索APIの調査も行う。また、検索エンジンから得られる複数の結果を統合して、ヒット数でランキング提示する機能も実装する。

#### (4) 辞書や基礎解析ツールの調査と新機能の考案

英語での作文支援に関しては先行ツールの知識が活用できる。また、日本語解析においても、これまでの研究成果と知識が利用できるため、英日以外の言語に対する辞書や基礎解析ツールなどの調査を行う。

#### (5) 誤り候補箇所の提示法の実現

我々の先行研究である英文作成支援ツールにおいても、学習者が書いた英文の誤り箇所の候補を提示する機能を実装した。この手法の基本的な考え方は、文をそのまま検証するのではなく、文を短い単位に分割し、短単位で検証を行う。長い文を検索した場合、ヒット数が限りなく0に近づくため、短いn-gramに分割し、短単位で誤り箇所候補を特定する。多言語に対しても類似機能を実現する方法を考案する。

#### (6) 利用しやすいインタフェースの考案

あらかじめ利用が想定できる状況を分析し、学習者利便性を考慮したインタフェースを考案する。また、学習者の検索要求に対して、検索式をサジェストする機能を実現する。

#### (7) 検索ログの活用

検索ログは学習者の試行錯誤した情報を含んでいる。学習者の検索履歴を分析することで、作文に有用な情報を抽出し、活用する機能を実装する。

#### (8) 総合ツールの構築と評価

有用サイト（辞書サイトなど）へのリンクや静的言語コーパスに対する検索機能などを備えた総合インタフェースを構築し、総合評価を行う。

### 4. 研究成果

研究期間中の主な研究成果を以下に示す。

#### (1) 9ヶ国語に対応した多言語作文支援ツールの実装

Web上での利用状況や機械翻訳の対象となっている言語調査、言語資源の整備状況などから、9ヶ国の言語（日本語・英語・ドイツ語・中国語・韓国語・フランス語・イタリア語・スペイン語・ポルトガル語）での作文を支援するツールを実装した。

#### (2) 語彙レベルに応じた用例提示と日英・英日対訳表示の実現

ユーザが用例文を閲覧する際、検索エンジ

ンの検索順ではなく、各ユーザにとって可読性が高く、語学レベルに即した用例文の提示が望ましい。そこで、用例文の質を向上させるため、用例検索サイトを指定できるようにした。また、日本語と英語に対しては、提示される用例文を各ユーザの語彙レベルに即した順序で提示する機能を実装した。なお、英語の語彙レベルの判定には、JACET8000 と SIL6000 を、日本語の語彙レベル判定には JLPT 語彙リストを利用した。図 1 に用例提示の例を示す。

図 1 語彙レベルに基づく用例提示の例

第二言語で作文する際、対訳用例が有用な情報となるが、対訳情報はコストが高く、静的なコーパスでも大規模なものは存在しない。そこで、本研究では論文アブストラクトに注目した。論文は日々増加しており、多くの論文アブストラクトには日英対訳が存在する。本研究では、CiNii の論文検索 API を用いて、学習者が必要とする用例を含んでいる論文アブストラクトを検索した後、検索結果からアブストラクト対を自動抽出し、学習者に提示する機能を実装した。具体的には、学習者によって入力されたキーワードを基に検索式を自動生成し、CiNii の論文検索 API で検索を行う。そして、検索結果から論文の URL リストを取得し、日英のアブストラクトが掲載されている書誌情報を取得する。取得した書誌情報から日英のアブストラクトを抽出した後、1 文ごとに区切り、表形式で提示する。なお、学会名や論文名などを入力することで、検索対象を制限できる。対訳用例提示の例を図 2 に示す。

	日本語	英語
タイトル	反復クラスター分析による意味ネットワークに基づく作文支援システムの開発	For the Development of Composition Support System based on Semantic Network by Repeated Clustering
本文	前置詞的側面から文法的な誤りを正すためこれまで開発されてきた作文支援システムとは異なり、我々は本研究において、単語の辞書的感度だけでなく単語間の連想-連想についての意味的対訳情報を提示する。重要な観点から作文支援を行う新しいシステムを開発。さらにそのシステムをWebアプリケーション化した件を紹介する。	Unlike the composition systems to have ever been developed to usually correct the grammatical errors in the syntactical aspect, we propose here the new system to support the composition from the different respect by providing semantic information not on just the meaning of words on the dictionary, but on the associative relations between words.
図表	我々が開発したシステムは反復クラスター過程によって得られた意味ネットワークを基礎とし、学習者から入力された単語に対する連想単語を提示する形である。	For this new composition system, what we developed first is the system which is as its resource based on the semantic network obtained by the repeated clustering process and which provides as its output learners with association words for the words they input.

図 2 対訳用例提示の例

(3) 柔軟な検索法と提示法の実現

作文を行う際、複数の候補語から最適な語を選択する状況や、どのような語・表現を使えばいいのかわからない状況が考えられる。このような状況において、不明な箇所をワイルドカードで抽象化して検証できるワイルドカード検索を実装した。ワイルドカードおよびその候補は多重指定が可能である。これにより、複数のワイルドカード指定時に、それぞれのワイルドカードに対して、候補のある／なしを組み合わせて指定できる。複数のワイルドカードにそれぞれ候補を指定する場合は、@マークと+マークを用いる。+マークはワイルドカードの区切りを、@マークは候補の区切りを意味する。例えば、「敗戦の①を②する」という日本語文に対し、①の候補が「汚名」、「名誉」で、②の候補が「返上」、「挽回」である場合を考える。この場合、入力欄に「敗戦の\*を\*する」、ワイルドカードの候補欄に「汚名@名誉+挽回@返上」と入力し、検証できる。検証結果を図 3 に示す。

Rank	*の箇所該当する単語	抽出されたフレーズ	Hit count	CHECK
1	汚名 / 返上	敗戦の汚名を返上する	13	<input type="checkbox"/>
2	名誉 / 挽回	敗戦の名誉を挽回する	1	<input type="checkbox"/>
3	名誉 / 返上	敗戦の名誉を返上する	0	<input type="checkbox"/>
4	汚名 / 挽回	敗戦の汚名を挽回する	0	<input type="checkbox"/>

図 3 日本語の検証例

(4) 誤り候補箇所の提示と訂正支援

第二言語で作文する際、自分が書いた文の妥当性が分からない。そこで、本ツールでは文中から誤り候補箇所の検出を支援する機能を実装した。具体的には、入力文を n-gram 系列に分割し、その系列のヒット数により、誤り候補の場所を指摘する。

分割方法は作文の対象言語によって異なる。英語やドイツ語などの場合、空白までを 1 単語として分割、日本語の場合は Yahoo! 日本語形態素解析 Web API を利用して文節単位に分割、韓国語の場合は文節相当で分割、中国語の場合は 1 漢字単位に分割し、n-gram 系列を生成する。各 n-gram 系列を Yahoo! API で検索し、ヒット数を得た後、ヒット数に応じて色分けしたグラフでユーザに提示する。韓国語の検証結果の例を図 4 に示す。

일문말	말	줄	아세요	HitCount
일문말	말			1
	말			324000
		줄		704000
		줄	아세요	0
일문말	말	줄		341
	말	줄	아세요	

図 4 韓国語文の誤り推定の例

また、第二言語で作文する場合、誤り候補が確認できても、それに対応できない状況は少なくないため、誤り訂正を支援する機能が必要となる。そこで、本研究では、ユーザが

試行錯誤しながら n-gram 検索結果を動的に訂正できる支援を実現した。以下の6操作により、ユーザは試行錯誤しながら誤り候補を訂正できる。

- ① 任意の単語の変更・削除
- ② 列指定による単語の追加・削除
- ③ ワイルドカード (\*) による候補取得
- ④ ドラッグ&ドロップによる単語の並べ替え
- ⑤ 任意フレーズの再検索
- ⑥ 任意フレーズのログ保存・削除

図4のグラフに対して、マウスで必要な箇所を直接操作することで結果が動的に反映される。また、試行錯誤の履歴をログとしても保存できる。英文に対する訂正例を図5に示す。



図5 英文に対する訂正例

#### (5) インタフェースの改善

ツールの検索ログは、ユーザの試行錯誤の履歴と考えられる。したがって、入力時に自分自身や他のユーザが試行錯誤した履歴が参照できれば、現在の問題点に対する検証方針が立てやすくなる。そこで、入力ログを活用したサジェスト機能を実装した。図6にサジェスト例を示す。



図6 サジェスト例

サジェストはワイルドカードの候補指定にも有効である。その例を図7に示す。

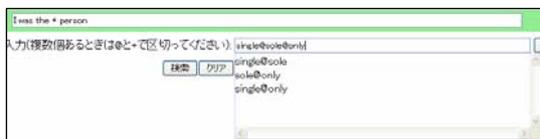


図7 ワイルドカード候補指定に対するサジェスト例

英作文時に、単語のスペルミスをすることは多々ある。そこで、Yahoo! Spelling Suggestion API を用いてスペルチェックを行う機能を実装した。具体的には、入力クエリが英語の場合はAPIを用いてスペルチェック

し、スペルミスがあった場合は訂正してユーザに提示する。

#### (6) 総合評価

日本語と英語での限定したツールの有用性評価を行った結果は、雑誌論文①などで発表した。

9ヶ国語に対応したツールの総合評価は、平成22年度に実施予定である。その成果は、雑誌論文や国際会議などで発表予定である。

#### 5. 主な発表論文等

〔雑誌論文〕(計1件)

① Kazuaki Ando, Yuichi Tsunashima, and Youhei Furukawa, An English and/or Japanese writing support tool based on a web search engine, International Journal of Computer Applications in Technology, 査読有, 2010.(Accepted, 印刷予定)

〔学会発表〕(計16件)

① 古川陽平, 綱嶋佑一, 安藤一秋, “Web検索エンジンを利用した9ヶ国語作文支援の改良”, 電子情報通信学会ET研究会, 2009年3月5日, 高知大学

② Kazuaki Ando, Youhei Furukawa, and Yuichi Tsunashima, A Writing Support Tool for Learners of Japanese as a Second Language based on a Web Search Engine, The 17th International Conference on Computers in Education (ICCE 2009), 2009年12月3日, The Hong Kong Institute of Education, 香港

③ 古川陽平, 綱嶋祐一, 安藤一秋, 検索エンジンを用いた多言語作文支援ツールの拡張, 2009年電気関係学会四国支部連合大会, 2009年9月26日, 愛媛大学

④ 古川陽平, 綱嶋祐一, 安藤一秋, Web検索を用いた多言語作文支援ツールの試作, 教育システム情報学会第34回全国大会, 2009年8月21日, 名古屋大学

⑤ Kazuaki Ando, Yuichi Tsunashima, Youhei Furukawa, and Masashi Okada, A Multilingual Writing Support Tool for Second Language Learners based on a Web Search Engine, World Conference on Educational Multimedia, Hypermedia & Telecommunications 2009 (ED-MEDIA09), 2009年6月24日, Sheraton Waikiki Hotel, ハワイ, アメリカ

⑥ 古川陽平, 綱嶋佑一, 岡田壮史, 安藤一秋, 検索エンジンを利用した9ヶ国語作文支援ツール, 電子情報通信学会ET研究会, 2009年3月7日, 香川大学

⑦ 古川陽平, 綱嶋祐一, 岡田壮史, 安藤一秋, Web検索を用いた日英独作文支援ツールの試作, 2008年電気関係学会四国支部連合大会, 2008年9月27日, 徳島大学

⑧ 古川陽平, 綱嶋祐一, 岡田壮史, 安藤一秋, 検索エンジンを利用した日英独作文支援ツール, 教育システム情報学会第33回全国大会, 2008年9月3日, 熊本大学

⑨ Kazuaki Ando, Yuichi Tsunashima, and Masashi Okada, A Writing Support Tool for Learners of English and/or Japanese as a Second Language, World Conference on Educational Multimedia, Hypermedia & Telecommunications 2008 (ED-MEDIA08), 2008年7月2日, The Vienna University of Technology, ウィーン, オーストリア

⑩ 綱嶋祐一, 岡田壮史, 安藤一秋, 検索エンジンを利用した多言語作文支援, 電子情報通信学会ET研究会, 2008年3月8日, 徳島大学

⑪ 綱嶋祐一, 安藤一秋, 岡田壮史, 検索エンジンを利用した多言語作文支援ツール, 教育システム情報学会研究会, 2007年11月25日, 香川大学

⑫ Yuichi Tsunashima, Yuta Kawasaki and Kazuaki Ando, English Composition and Learning Support System using a Search Engine, 15th International Conference on Computers in Education (ICCE2007), 2007年11月8日, グランドプリンスホテル広島, 広島

⑬ Kazuaki Ando, Yuichi Tsunashima, Yuta Kawasaki and Shoji Mizobuchi, A Writing Support Tool for Learners of English as a Second Language, The World Conference on E-Learning in Corporate, Government, Healthcare & Higher Education 2007 (E-Learn2007), 2007年10月16日, Quebec City Convention Centre, ケベックシティ, カナダ

⑭ 綱嶋祐一, 安藤一秋, 検索エンジンを利用した日英作文作成支援ツール, 2007電気関係学会四国支部連合大会, 2007年9月29日,

徳島大学

⑮ Yuichi Tsunashima, Yuta Kawasaki, and Kazuaki Ando, A Support Tool for English Composition using a Search Engine, SICE Annual Conference 2007, 2007年9月19日, 香川大学

⑯ 綱嶋祐一, 安藤一秋, 検索エンジンを利用した第二言語作文支援ツール, 教育システム情報学会第32回全国大会, 2007年9月12日, 信州大学

#### 6. 研究組織

##### (1) 研究代表者

安藤 一秋 (ANDO KAZUAKI)

香川大学・工学部・准教授

研究者番号：60325321

##### (2) 研究協力者

綱嶋 祐一 (TSUNASHIMA YUICHI)

香川大学大学院・工学研究科・学生

研究者番号なし

古川 陽平 (FURUKAWA YOUHEI)

香川大学大学院・工学研究科・学生

研究者番号なし