

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 22 日現在

機関番号：37402

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520596

研究課題名(和文) 日本語文章校正ツール「推敲」技術を用いた日本語教師用例文検索システムの開発

研究課題名(英文) Development of a Japanese similar syntax sentence search function using critique tool SUIKOU

研究代表者

堤 豊 (TSUTSUMI, Yutaka)

熊本学園大学・商学部・教授

研究者番号：80227446

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：本研究では、「日本語文章校正ツール『推敲』技術を用いた日本語教師用例文検索システムの開発」をテーマに研究を行った。この研究では単に例文検索の機能実現だけでなく、ネットワークを利用した協調学習環境を語学学習において実現するためのサーバ機能の洗い出しや、クライアントソフトにおいて必要な機能の実現にも注目し研究を進めた。ターゲットとするデータベースは当初ニュースサイトを想定していたが、著作権の問題等から著作権が切れた小説などを保管するサイトに変更した。開発したシステムについては機能は満足できるが実効速度が遅いため、今後も継続して改善する必要がある事が分かった。

研究成果の概要(英文)：In this research project, we aimed at developing a similar syntax function on e-learning environment that based on general word processing software such as MS-Word. The target domain of this system was changed from online news site such as NY times, TBS news, etc. to AOZORA-bunko, which holds a number of novels and poems written in Japanese expired license and become public domain documents.

Research result shows that the turn-around-time of this system does not suit for a classroom usage and need to be continued to update the system.

研究分野：情報工学

キーワード：類似文検索 日本語教育

## 1. 研究開始当初の背景

語学教育で共通の問題として教材作成の困難さ、複雑さがある。語学教育においては、マルチメディアやインターネットの活用は、効果的な学習ができることが容易に想像される。例えば、ある構文を教えた時にホットな話題を題材にしたホームページの記事があれば、学習者の注意を惹きつけることができるだろう。

一方、一般の教員はプログラミングの専門家ではないため、自分で教材ソフトウェアを開発することは難しい。既存のソフトウェアでは、プログラミングの手間は少ないが「お仕着せ」の教材となるため、教師が補助教材として利用するためには、カリキュラムをソフトウェアに合わせる必要がある。

これらの問題点を解決するために、本研究班では、従来研究で一貫して教材作成を簡単にするためのソフトウェアの開発を行ってきた。例えば、WEB上で問題を作成できるようなツールを提供したり、ワープロ感覚で問題が作成できるツールの開発などが挙げられる。これらの研究では、英語を母語としない学生(EFL)の英語教育を対象に研究してきた。

また一方で筆者らは、九州大学牛島研究室で開発された日本語文章校正ツール「推敲」のソースコードの保守を行っている。「推敲」では、日本語文章のパターンを正規表現で切り分けることで高速に助詞、助動詞の分類を行うことができる。さらに、筆者らは類似文検索について、従来から研究を進めてきた。本研究では、この「推敲」の技術と今まで培ってきた英語教育のための教材作成ツールの技術を基盤にして日本語教師用の例文検索システムを開発することで日本語教師および日本語学習者への新しい語学学習環境の提供ができるのではないかと考えた。

## 2. 研究の目的

本研究では、1で述べたように、新しい語学学習環境の整備が目的であるが、特に日本語の類似文検索を対象としている。

教材テキスト中で注目すべき表現が出てきたときに、簡単に似た例文をインターネットの記事から検索し、利用できるようにする。その際、構文的に似た文か、意味的(語彙的)に似た文かを選択する。

教材サーバは夜のうちにインターネットの指定されたサイトから記事を記事DBにダウンロードしたのち、構文解析し、品詞情報および語彙情報とともにタグ付けされ、早見表DBに格納される。教員用クライアントを用いてMS-Wordから用例検索要求があると、VSTOアドインソフトとして開発するクライアントソフトから教材サーバにパターン・マッチング指令が出され、類似度の計算をしてアドインソフト経由でMS-Wordに表示する。

本研究の特色は、日本語学習において、「似た構文を探し」「文の難易度を計算し」「ワープロ画面に表示する」ことにより、日本語教師が教材を簡単に検索し、利用できることである。従来から新聞記事等を検索しまとめて表示するシステムは存在するが、キーワードやその組み合わせで検索するものがほとんどである。本研究では、主体は構文検索である点が大きく異なっている。また、構文と語彙情報により文の難易度を計算する点も従来のシステムにはなかった点である。さらに、ワープロ画面から簡単に検索でき、また結果もワープロ画面に出力するため、日本語教師が簡単に操作できる点も特色である。

本研究では、この類似文検索を日本語ワープロ上から利用できるように、サーバ環境とクライアント環境を整備し、かつクライアント環境からサーバの機能を呼び出すためのインタフェースプログラムを開発すること、および自動でデータベースを更新するための仕組みを構築し、実用的なターンアラウンドタイムで実現できるか実現可能性を探ることが目的である。また、それぞれに必要な要素技術の確立も重要な開発の目的となる。

## 3. 研究の方法

本研究では、「日本語文章校正ツール推敲技術を用いた日本語教師用例文検索システムの開発」をテーマに研究を開始した。この研究では単に例文検索の機能実現だけではなく、ネットワークを利用した教師と学生間、および学生同士の協調学習環境を語学学習において実現するためのサーバ機能の洗い出しや、クライアントソフトにおいて必要な機能の実現にも注目し研究を進めた。今回開発した主に、夜間に自動でインターネット上から例文を取り込み保存するサーバの開発と、例文を推敲技術により形態素解析し早見表を作成し類似検索を行う技術の開発と、クライアント機において語学学習用の協調学習環境を整備し、その中で提示された文を引数にサーバから似た文を引き出し画面に表示する部分を追加する、という3つを並行して研究してきた。このうち、語学用協調学習環境については、本研究班の先行研究である、「作業用共用データベースを利用した英専門書読解支援ソフトウェアの開発に関する研究(文部科学省平成17~18年度科学研究費補助金基盤研究C課題番号17520391:研究代表者 松野了二、研究分担者 堤豊)の研究成果を元に構築を図った。

## 4. 研究成果

### 【サーバ関連の研究成果】

例文検索のためのサーバ構築については、例文をインターネットから自動で検索しダウンロードし形態素解析を行い品詞でタグ付

けてデータベースに保管するシステムの作成を行った。ターゲットとする分野を当初は新聞データベース等に置いていたが、2つの問題に直面した。1つめは研究開発当初は新聞データベースについて比較的インターネット上で充実し無料でニュースを公開する方針であったが、徐々に閉鎖的になり、ニュース記事を全文投稿する会社が減ったことである。2つめはデータベースを自由に利用させることを嫌う会社がネット上に公開するときもJavaScript等を利用し暗号化し、簡単には生データが取り出しにくくなったことである。これらのことを考慮し、また今後ニュース記事について有料提供という方針になることも危惧されるため、当初の方針を変更し、データベースとしてはニュース記事ではなく著作権が切れた小説や詩等をデータベース化している青空文庫を選択することとした。本研究では、この青空文庫のサイトから自動で文書をダウンロードし、htmlタグを取り除き、生のテキストデータとして利用するためのソフトウェアを開発した。この成果を利用して、学習環境を構築したものが発表論文1である。ここでは、google formsを利用して小テストを実現するために選択問題、並べ替え問題、マッチング問題等のバラエティを持たせるための機構や、図やハイパーリンクを挿入するための機能を実現するために google forms の表現力をアップさせる機構を導入し、サーバ上で自動的にhtmlを解析し変換できるようにした。この中で本研究により実現した html 変換のアルゴリズムの成果が利用されている。

#### 【学習環境関連の研究成果】

本研究の構成要素のうち、クライアント環境は、Microsoft 社の Word 環境において自由に類似文検索をはじめとする語学学習支援環境を実現することである。研究論文2は、EFL 学習者を対象とした英語学習システムの実現についての論文である。これは、筆者らが長年研究してきた、語学学習を簡単に行えるシステムの実現のために、ワープロソフトを基盤に置いて開発したものである。この中で、単語の頻度情報やスペルチェック、グラマーチェック、慣用句の頻度情報、KWIC など様々な機能を実現している。これらの機能の大半は VSTO により実現されており、本研究のクライアントソフトとして利用している。

#### 【システム統合と問題点】

システム統合とは、サーバとクライアントソフトを接続し、類似用例検索をクライアントソフトから利用できるようにすることである。ここで必要な技術は、サーバ側で行っている形態素解析し品詞によりタグ付する作業と同じ作業をクライアント側で行う必要がある。このクライアント側での形態素解析

には推敲での技術を利用し VSTO プログラム環境で行った。また、同様に VSTO を利用して、クライアント機からサーバに用例検索要求を出し、結果をワープロ画面に表示する機能を実現した。これらの成果は発表論文3に記述した。ただし、問題点がないわけではない。(1)現状では応答速度が遅い。用例データベースの大きさに対し  $n \log(n)$  の計算量ではあるが、類似検索においては、パターンマッチする回数にも比例して計算量が増えるため、データベースの大きさ、類似検索におけるパターンマッチの抽象化段階の設定などについて研究を深めていく必要がある。(2)対象データベース。現在は著作権の問題を回避するため、青空文庫を対象にしているが、語学学習においては、文学作品よりむしろ時事問題を扱い、写真などのマルチメディア情報も付加されたコンテンツの方が学習者の興味を惹きやすいと考えられる。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

Matsuno, R., Matsuo, K., Gilbert, R., and Tsutsumi, Y.

MiWIT-Plus: Integrated Text Tools for Creating EFL/ESL Reading Materials Integrated Text Analysis Tools, SNS and Cloud Storage within the MS Word Platform 12th International Conference on Information Technology Based Higher Education and Training (ITHET2013)., ISBN978-1-4799-0086-2/13/, 6pages, 査読有, (2013).

〔学会発表〕(計 2 件)

Tsutsumi, Y.

A Japanese Educational Tool Featuring Similar Syntax Search Function: Practical Usage of Japanese Corpus Proceedings of e-learn 2014 (AACE) pp.1947-1949, 査読有 (2014)

堤 豊

Google Form を利用したオンライン小テストシステム

私立大学情報教育協会 平成 24 年度教育改革 ICT 戦略大会講演論文集, pp212-213, 査読有 (2012)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況（計 0 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

#### 6. 研究組織

(1)研究代表者 堤 豊(TSUTSUMI, Yutaka)  
(熊本学園大学商学部・教授)

研究者番号：80227446

(2)研究分担者 松野 了二(MATSUNO, Ryoji)  
(熊本県立大学総合管理学部・教授)  
(平成 26 年 3 月削除)

研究者番号：60044351

(3)連携研究者  
( )

研究者番号：