

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 17 日現在

機関番号：31302

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500891

研究課題名（和文）利用者の背景を考慮した自然言語教育・学習・研究支援システムの実現

研究課題名（英文）Development of the Natural Language Education, Learning and Researching Support System taking account of the user's Background

研究代表者

坂本 泰伸（SAKAMOTO YASUNOBU）

東北学院大学・教養学部・准教授

研究者番号：60350328

研究成果の概要（和文）：

本研究では、利用者の背景を考慮した自然言語教育・学習・研究支援システム CHAPEL を開発した。CHAPEL は、従来のコーパス解析システムとは異なり、利用者間で注目している英文の共有機能を有する。この機能は外国語教育の効果を高めると考えられる部分である。そして、実際の大学に於ける英語教育の現場で、CHAPEL システムを利用した実証実験を実施し、システムに対する評価を実施した。評価は、学生及び教員に対するアンケートを通じた聞き取り調査を実施し、CHAPEL の持つ英文の共有機能の効果を明らかにした。

研究成果の概要（英文）：

CHAPEL system is taking account of user's background, and supports these users' work around Natural Language Processing such as Education, Learning and Researching. A difference between the CHAPEL and the other general corpus-based analysis systems is existence of a new function to share English sentences noticed by the system users. This function is thought as giving much effect for English education. In this research, we conducted an evaluation of the CHAPEL system in an actual English education class in University. And the effect of the new function in the CHAPEL was clarified by the evaluation of questionnaires investigation.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	600,000	180,000	780,000
2011 年度	900,000	270,000	1,170,000
2012 年度	700,000	210,000	910,000
総計	2,200,000	660,000	2,860,000

研究分野：総合領域

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：教授学習支援システム

### 1. 研究開始当初の背景

計算機可読（Computer Readable）な自然言語情報（コーパス）を使用した研究の歴史は古く、1970 年代には現在の英語コーパスを用いた自然言語研究の中心的存在となる国際機関 ICAME（International Computer Archive of Modern and Medieval English）が設立され、これまでに数多くの研究成果が

世に生み出されている。我が国でも、コーパスを用いた研究が盛んに進められており、その研究成果は自然言語の研究分野のみならず、計算機による文献検索や自動翻訳等の他の応用分野にも利用されている。また、教育支援や学習支援に関する分野では、教材開発という形でも利用されており、今後もより一層の発展に期待が持てる分野であった。

言語教育を目的とした諸外国における英語コーパスの研究では、使用目的や使用される領域を限定してその利用実態を解明し、この結果を基にして策定された語彙リストを教材として利用する試みなどがある。一方、我が国では、研究用に分類された英文に対して統計解析した結果を、主として自然言語研究者が自身の経験に基づいて教科書や教本などの材料として提供しているケースが多い。これは、実際の現場で教育をする者や受ける者が、その生徒観に見合った教材を柔軟に選択しているとは言い難い。そこで、高大の英語教育に携わる教員や、教員の指導の下で学習を進める学生、英語論文の作成が強く求められている研究者といったように、外国語の習熟レベルや利用目的を考慮した上で教材を選択し、テスト問題の作成や、長文読解の題材の抽出、レポート配布や収集等の支援機能を柔軟に提供する枠組みが必要であることに着目をした。

我が国において、コーパスを用いた研究が急速に発展してきた理由には、1990年代中頃から始まった計算機の処理性能の劇的な向上や、国内のネットワークインフラの急速な発展等の要因も非常に大きく関係している。これらの発展に伴いコーパス解析システムで扱うデータ容量も肥大化し、サービスはネットワークを通じて提供することが可能になった。情報科学の分野では、ネットワークを通じたこの様なサービスには、システムの製品寿命の延長や再利用性を可能にする Three Tier Model の考えと、リレーショナルデータベース (RDBMS) の存在が必須であると考えられていたが、自然言語の研究分野では、コーパスに蓄積される英文データそのものに対する関心が高く、ICAME 等の国際学会に於いても、コーパス解析システムの設計や性能に関する報告が少ないのが現状であった。

また、研究者が集積したコーパスは、研究者個人レベルで開発したプログラムを利用して解析されるので、研究者間でのデータの再利用性や研究で得られた成果の比較、結果の整合性の確認がスムーズに進められないという問題も存在していた。

## 2. 研究の目的

本研究では、単純な英語データベースを構築する内容とはせず、教育や学習、研究での利用を目的とした応用性の高い支援システムの開発を進める。特に、最終的には高校や大学などの実際の教育現場にて実証実験を実施することを目的とし、本研究期間における研究の中心となるキーポイントを下記の4点(順不同)に設定をした。

(1) 多様なユーザ(高大の英語教員、教育を受ける側の生徒、論文作成が必要となる研究

者等)が、本システムに対してどのような機能を要求しているのかを調査する。特に、国際学会等で学術成果を発表することが強く求められている我が国の研究者に対して、本システムがどのような機能を提供しなければならないのかを調査する。

(2) システムに対して多種多様な機能の追加が長期的に求められるので、これを可能とする一般化された構造(フレームワーク)を確立する。

(3) 一般的なコーパス解析システムでは、統計解析処理がその主たる機能となるが、本システムでは、統計解析の結果を単純に提供するものとせず、解析結果を、利用者の要求(単語力テスト、反意語や同義語の一覧表、学習補助プリント、レベル別の長文抽出、研究で必要となる統計結果の出力等)に応える形に情報変換して提供する。この情報変換の方法やタイミング、情報提供の方法に関する最適解を明らかにする。

(4) 研究用途を目的とした使用にも耐えうる、高速な解析性能や高度な解析機能の実現を目指し、詳細なデータベースのテーブル設計と性能評価を行う。

## 3. 研究の方法

本研究の目的とするシステムの特徴は、様々なモジュールを追加(add-on)することで多様な背景を持つユーザからの要求に応えるものである。この提案は、本研究の申請時に於いて、ICAME 及び CL2009 において報告を実施し、多くの会議出席者との議論を通じ、我々のアイデアが外国語教育や自然言語研究の分野に受け入れられる事を確認していた。支援システムに add-on される機能モジュールは Front End Tier に実装されるが、これらの要求は自然言語教育や学習、そして研究上からの要請である場合が多く、開発側の独善的な考えで開発を進めることは避けるべきである。そこで、自然言語教育の専門家である研究分担者の協力を得ながら、実装すべき機能を ICAME 等の国内外の学会参加を通じて情報収集をしながら明らかにする。

システムの汎用性を実現するにあたって、共通化されたインターフェイスは非常に重要となる。そこで、このシステムの中核となる Mid Tier (フレームワーク)の設計と開発を研究期間の前半に実施する。そこで、先ず RDBMS に英文を蓄積する為のテーブル設計を実施し、この英文に対する簡易(プロトタイプ的)な統計解析や検索の機能を研究期間初期に実現させる。この時点での統計解析機能は、自然言語研究の用途であることを目的とせず、可及的速やかに中期以降の開発ができる事を重視し単純な内容とする。

研究期間中期には、このプロトタイプ版に対して、学会活動を通じて調査した機能モジ

ジュールを組み込む。このモジュールの組み込み際に、統計解析結果を「単語力テスト」や「レベル別の長文抽出」と云った利用者の要求に応える形に情報変換する方法を提案する。

研究期間後期には、機能モジュールの組み込み際に、検索機能の強化を実施し、これに伴うデータベースのテーブル設計の最適化を行う。この際、既存コーパス(XML 版BNC等)との比較検討を十分に進める。データベースの構造は、提供する統計解析や検索機能から要求される内容に依存する部分が強いため、データベースの性能評価は、システムのチューニングが終了する最終年度に実施する。特に、実現したプロトタイプ版システムの性能調査を進め、その結果を公開する。公開に際しては、まず ICAME 等の国際学会において本研究の成果を発表する。この時点では、性能評価が既に終了しているため、この性能評価に用いた計算機を利用して、大学や高等学校の英語教育の現場で、実際にシステムを利用した評価を実施する。

#### 4. 研究成果

本研究の目的である自然言語教育・学習・研究支援システムの実現に向けて、研究開始年度となった 2010 年度には、システム開発を中心とした作業を進めた。まず、システムの中核となる RDBMS のテーブル設計と評価を実施し、これと並行して、システムに要求される機能の抽出を実施した。また、システム開発に必要となる Java 言語による共通化フレームワークの設計も実施した。

(1) RDBMS のテーブル設計と評価：システムの中核となる RDBMS 内部のテーブル設計を実施し、既存コーパスである BNC に採録されている英文データを蓄積する際の時間性能の評価及び、蓄積された英文に対する検索の時間性能の評価を行った。システム全体の処理時間性能は、テーブル間のリレーションの構造に大きく依存するので、非常に重要な点である。評価では、我々がこれまでに利用していたテーブルよりも大きな性能の向上が認められたが、大規模な検索が必要とされる実用化に際しては、さらなる処理性能が要求されることを明らかにした。

(2) システムに要求される機能の抽出では、自然言語学教育の専門家である研究分担者が、英語読解用の語彙選定を行いながら、システムに求められる機能の素案を提案した。自然言語教育の専門家からの要求をシステムに組み込む事で、より多くの利用者への対応が可能となるものである。

(3) 本システムでは、Java 言語によるアプリケーションが連動して利用者に対してサービスを提供する。このアプリケーション開発に際して、共通化されたフレームワークを

導入する事で開発時間の短縮を目指した。開発手法を提案し、このフレームワークの設計と開発を進め、その報告を実施した。

2010 年度から 2011 年度にかかる研究成果に関わる点として、2011 年 3 月 11 日の東日本大震災による被害で、上記 (1), (3) に関するデータディスクが破損し、一部のデータが失われている。

2011 年度は、前年度に実施した RDBMS のテーブル設計に基づいて、実際の英文データをこれに蓄積し、統計解析機能の実装を進めた。

(4) RDBMS の設計と評価に関する作業では、高い検索機能が必要とされる実用化に向けて、専用サーバを購入し、2010 年度に実施した性能評価をこのサーバ上で再度実施した。その評価の結果では、購入したサーバを使用する事で、利用者に対してシステムが必要とされる機能を提供できる事を確認し、その報告を実施した。

(5) アプリケーション開発に関する共通化されたフレームワークの中で GUI (Graphical User Interface) パッケージの開発を終了した。

(6) RDBMS に蓄積された英文データに対する統計解析機能の開発を実施した。

研究期間の最終年度にあたる 2012 年度は、過去 2 年間の研究の成果を利用して、英語教育支援アプリケーションの開発と評価を実施した。過去 2 年間では、RDBMS の設計と GUI (Graphical User Interface) パッケージの開発を終了しており、この成果に基づいて英文データを RDBMS に蓄積し、統計解析機能の実装を進めた。

また、三年間の調査を通じて、外国語学習を進める際に教員が学習者の理解度を正しく確認する事が効果的な授業を進める上で特に重要なポイントでもあることを明らかにした。とりわけ、学習者が疑問に思っている箇所や難しく感じた箇所を、教員がリアルタイムに把握することが非常に重要となり、教員が教授している箇所と学習者が注目している箇所のギャップなどを正確に把握する事が、授業の成果を大きく左右する点である。これらの課題に対する解として、利用者による英文マークの位置共有機能を提案した。

(7) 英語教育支援アプリケーション CHAPEL を開発した。CHAPEL には、リアルタイム性の利用者による英文マークの位置共有機能を始め、RDBMS 中にコーパス構築を進める際に必須となる英文整形アプリケーション (Line Formatter) や、コーパス中に含まれる英単語に対する属性情報の管理アプリケーション (Attribution Maker) を実装した。また、この他に Lemma や Lexicon の管理機能、語彙集などの管理機能も含まれる。

(8) 当初の研究計画では、CHAPEL は高等学校の英語の授業を通じて評価の実施をする予定であったが、CHAPEL の完成時期と高等学校の科目スケジュールとの関係から、大学の基礎年次教育の外国語（英語）講義での利用に変更した。評価は、学生へのアンケートの解答及び教員に対する聞き取り調査で実施し、CHAPEL が持つ、リアルタイムの利用者による英文マークの位置共有機能が教育に対して効果が高いことを明らかにした。

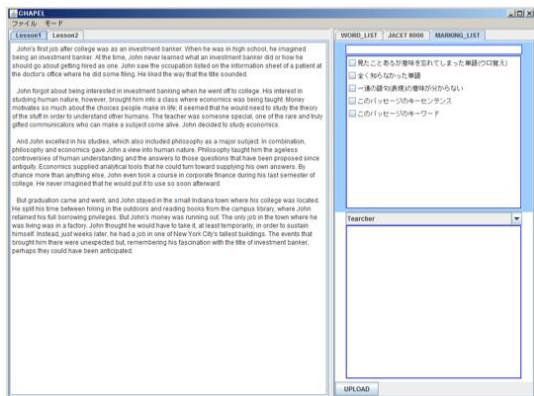


図 1：開発した CHAPEL の実行画面

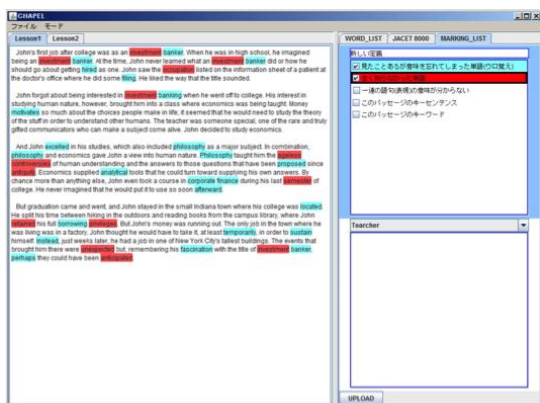


図 2：CHAPEL による英文共有の様子

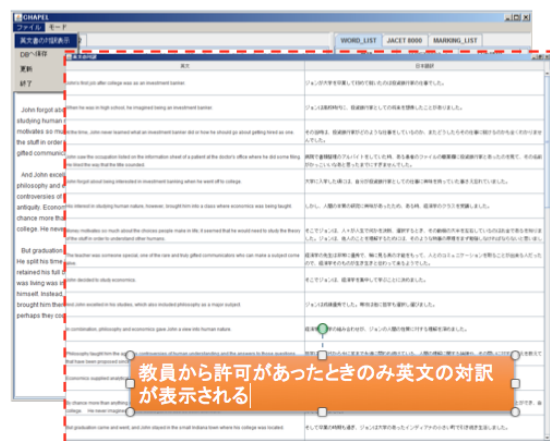


図 3：和訳表示によるサポート



図 4：語彙表示によるサポート

## 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 5 件)

- ① 岡田毅「マルチメディアコーパスの意義と特性について」、国際文化研究科論集, 東北大学国際文化研究科, 査読有り, 20 巻, 頁 145-154, 2012 年
- ② 藤野玄大, 坂本泰伸, 岡田毅「ユーザ間の情報共有に立脚した英語教育および学習支援一協働を保証する高汎用性コーパスシステムを用いて」, e-Learning 教育研究, 査読有り (研究ノート), 7 巻, 頁 11-22, 2012
- ③ 岡田毅「汎用的英語コーパス解析システムにおける入力ソース整備モジュールについて」, 国際文化研究科論集(東北大学大学院国際文化研究科), 査読有り, 第 19 号, 頁 53-64, 2011 年
- ④ 岡田毅「汎用的コーパスシステムにおけるユーザーの概念--RDBMS を中心に据えて」, e-Learning 教育研究, 査読有り, 第 6 巻, 頁 11-24, 2011 年
- ⑤ 岡田毅「東北大学全学教育における英語読解用語彙の選定について」, 東北大学高等教育開発推進センター紀要, 査読有り, 第 6 号, 頁 15-33, 2010 年

〔学会発表〕(計 8 件)

- ① 藤野玄大 (坂本泰伸) 「言語教育支援システム CHAPEL の評価実験の報告」, 2013 年電子情報通信学会総合大会, 2013 年 3 月 20 日, 岐阜大学
- ② 藤野玄大 (坂本泰伸) 「コーパスを用いた言語研究・学習・教育支援システムにおける Corpus Building Package の開発」, 2012 年電子情報通信学会 総合大会, 岡山大学 (津島キャンパス), 2012 年 3 月 20 日
- ③ 藤野玄大 (坂本泰伸) 「コーパスを利用

した研究・学習・教育支援システムの開発に向けた RDBMS の評価」, FIT2011 第 10 回情報科学技術フォーラム, 函館大学・函館短期大学, 2011 年 9 月 8 日

- ④ Takeshi Okada, " An Integrated Word Annotation System for the English Reading Classes in Japanese University " , The 44th British Association of Applied Linguistics Annual Conference, United Kingdom, Bristol City, University of the West of England, Sept. 3<sup>rd</sup>, 2011
- ⑤ 三浦宏太 (坂本泰伸) 「RDBMS を中間インタフェースとする英文書に対する属性情報付与システムの提案と開発」, 第 36 回教育システム情報学会全国大会, 広島市立大学, 2011 年 8 月 31 日
- ⑥ 三浦宏太 (坂本泰伸) 「コーパスに基づく研究・学習・教育支援システムで用いる新しい Tagger システムの提案」, 2010 年度情報処理学会東北支部第 5 回研究会, 東北学院大学, 2011 年 2 月 15 日
- ⑦ 藤野玄大 (坂本泰伸) 「コーパスに基づく研究・教育・学習支援システムで利用する英文書管理の提案と評価」, 2010 年度情報処理学会東北支部第 5 回研究会, 東北学院大学, 2011 年 2 月 15 日
- ⑧ 吉木貴紀 (坂本泰伸) 「情報教育での利用を目的とした Java アプリケーションフレームワークで用いる状態遷移機能と入出力機能の実装」, 2010 年度情報処理学会東北支部第 5 回研究会, 東北学院大学, 2011 年 2 月 15 日

## 6. 研究組織

### (1) 研究代表者

坂本 泰伸 (SAKAMOTO YASUNOBU)  
東北学院大学・教養学部・准教授  
研究者番号：60350328

### (2) 研究分担者

岡田 毅 (OKADA TAKESHI)  
東北大学・国際文化研究科・教授  
研究者番号：30185441

### (3) 連携研究者

なし