

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 29 日現在

機関番号：11301

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23531238

研究課題名(和文)日本人科学研究者向け英語学術論文執筆支援用教材システムの開発

研究課題名(英文)The development of scientific article writing support system for Japanese researchers

研究代表者

岡田 毅 (Okada, Takeshi)

東北大学・国際文化研究科・教授

研究者番号：30185441

交付決定額(研究期間全体)：(直接経費) 2,900,000円、(間接経費) 870,000円

研究成果の概要(和文)：本研究では日本人研究者が英語で執筆する科学論文の中で見られる特徴と、英語母語話者の同一使用域内での表現の特徴とを比較対照するための基礎的データ整備を行い分析を加えた。日本人科学者産出の英文特徴の抽出先として、東北大学の中で世界水準の研究成果を発信し続ける部局である金属材料研究所に焦点を当て、その主要研究分野の材料科学の領域での世界的なジャーナル7種類から、合計563編の論文を慎重に選別した。総単語数は2,307,825で、単語種類数は71,224個にのぼる。

研究成果の概要(英文)：This research revealed idiosyncratic choice of English reporting verbs and sentence structures observed in scientific research articles written by Japanese researchers in the field of materials science. To make statistical comparison between Japanese-made articles and those made by English native speakers, we obtained 563 articles from seven international journals that cover the field of materials science. The total number of running words is 2,307,825 and the number of word types is 71,224. With paying special attention to the over-use tendency of reporting verbs such as argue, assert, say, mention, suggest, predict, conclude, etc, the sentence patterns in which those verbs are used are systematically classified. By giving them a close examination, we obtained major Japanese-peculiar tendency of using reporting verbs, which would lead to be a convincing guideline for scientists who try to publish their research results as precisely and accurately as possible.

研究分野：社会科学

科研費の分科・細目：教育学・教科教育学

キーワード：教材開発 リレーショナルデータベース コーパス EFL教育 英語科学論文

1. 研究開始当初の背景

国内では Web 上から提供される日本人英語学習者産出の英文コーパスを分析対象とし、様々な教材開発や教育支援に活用する傾向が強まっている。その際に英文中の過誤認定基準に関しては世界的にも活発な議論が続いているのに加えて、特に日本の学校制度における学年別指導要領等に準拠した語彙のレベルの設定手法が一般的に採用されている。非母語話者である日本人学習者にとっての目標語彙レベルに関する一定の客観的基準として、学年別区分は評価できるが、教科書記載事項と習得事項の間に不可避免的に生じる測定不可能なギャップを処理し、説得力のある教材作成支援システムへと結びつけるという試みはなされていないのが現状である。

また、国外では学年別という範疇分けではなく、英語の使用域に関わる諸研究の成果を踏まえて、特定目的・特定領域での英語使用の実態を解明し、具体的な語彙リストの策定もなされている。日本国内でもこのような特定領域に焦点を絞った英文作成マニュアル本の類が数多く出版されているが、その大部分が当該学術分野の研究者が自身の経験に基づいて提案する極めて主観的な規範を示しているに過ぎないのも看過できない現状である。

本研究の重要な着目点として、教材開発者側からのユーザーレベル設定に対する疑問がある。指摘したような、学年別の(予測)習熟度というスケールと個々のユーザーの実際のニーズ間での乖離の問題がその根底となる。漠然としたユーザー層を想定するより、自らの作成した学術英文の、投稿先ジャーナルや当該専門分野での容認性(acceptability)を高めたいという具体的で明確な利用目的を持ったユーザー、即ち、日本人科学研究者、を前提とすることにより、コーパス解析に準拠したシステムが、よりその性能を向上させる可能性が高いということに着目している。

研究代表者は分担者と共にリレーショナルデータベースマネジメントシステム(RDBMS)を中心に据えた、ユーザー自身が定義可能な属性を柔軟に付与できるシステムの開発を推進し、これまでに国際学会でその成果を発表し、その中でシステム側の user-friendliness の向上はユーザーニーズの特定化と密接な関係にあることを明らかにしてきた。

2. 研究の目的

一定レベル以上の英語運用能力と、海外に発信すべきコンテンツを備えた日本人科学研究者にとっての有益な教材として、「ユーザー自身による学術英文データの収集・選択統合・解析全般を実現するシステム」を開発し、Web 上で提供する。当該システムにより、ユーザーの所属する学術分野及び論文の投稿先ジャーナル等でより受け入れられやす

い表現や構文特徴の客観的抽出が容易となり、特定使用域における自然で質の高い英語表現を用いた学術論文を効率よく作成する能力の養成が強力に支援されることになる。自己啓発的で自律学習を支援する e-ラーニング教材を実現し、結果的に我が国の科学研究者たちの正確で能動的な学術情報発信を一層促進することに大きく貢献することが最大の目的である。

3. 研究の方法

共同研究体制を確立させるために、研究代表者と分担者の研究室を VPN(virtual private network)システムで接続し、分担者の研究室に設置した専用サーバーマシンをシームレスに活用できる環境を整備する。これにより RDBMS の一次データとしての専門学術分野別の英語文書を高速・大量に取り入れるための手法を整備し、データベースの基本設計を完成する。

その後、試用システムを稼働し、ユーザーの要求を具体的に集約して把握する。必要機器を導入後、本システムの試験的運用を開始すると共に、研究成果の大学英語教育への応用の可能性を模索する。この際に、教養教育段階での英語教育を担当する多くの大学教員に向けて試作教材システムを公開し、システム試用後のアンケート調査等を利用して参考意見を求める。具体的なコーパスセット分析結果を国内外の学会及び学会機関誌で発表し、フィードバックを得る。

RDBMS の基礎的データとなる一次資料の収集の体系的実現、研究の最終段階での公開を目指すコーパス解析システム全般の開発に関わる共通の理解と知見の共用を保障する研究体制を確立する。研究初年度は、研究代表者と分担者の研究室間での VPN システムによるシームレスな通信環境を整備し、共同研究体制を確立する。

研究代表者がコーパス研究の立場から、特定学術分野に特徴的な語彙及び構文情報の抽出に不可欠なデータ分析手法を模索すると同時に、ユーザーである日本人研究者のニーズを柔軟に反映させるための独自品詞標識及び構文標識付与を保障する図1のような RDBMS 中のテーブル設計を推進する。

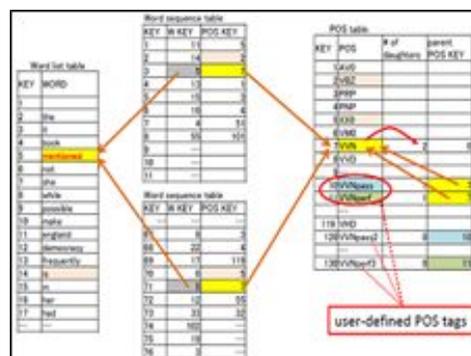


図1 RDBMS 中のテーブルの例

このプロセスで得られた知見は、研究代表者の担当する大学院の講義で活用され、コーパス言語学的な研究成果は ICAME 等の国際学会で発表される。分担者は three-tier model に準拠した RDBMS の設計を推進するが、back-end tier に相当するデータベースに大量の分析用一次データを蓄積するためのひとつの手段としての英文トークン化システムの GUI (graphic user interface) による front-end tier への移行を実現する。このトークン化システムは、処理対象の英文の入手過程 (印刷物からの OCR 取得、pdf 文書、HTML 文書等) の相違に呼応した処理系フィルタ群をユーザーが選択し、その適応順序も指定することができるという DCL (dynamic class loading) 機能を Java によって実装しているために、本研究の最大の特徴のひとつである日本人研究者自らが必要とするデータを正確に大量に取り込むことを可能にさせる。

また、研究代表者がこれまでに解明した日本人産出の英語過誤特徴に関する知見をフィルタのひとつに反映することができることから、このプログラムはで示されたトークン化システムの発展と位置付けられる。

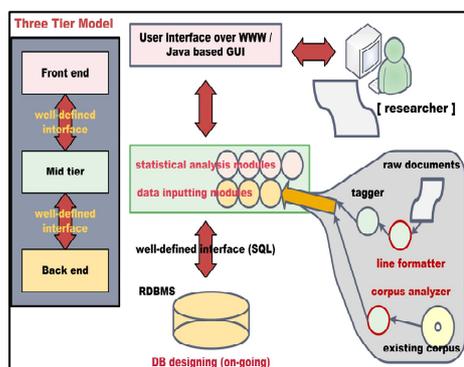


図2 three-tier model とシステム

データ収集に関わる柔軟な機能を、ユーザー側に一番近い front-end tier で提供する必要があり。大量のデータの自動収集時における初期値的なフィルタ選択及びその順序付けに必要な専門分野別の表現傾向を明らかにするために、研究代表者が遂行するコーパス分析の成果も期待される。このように、コーパス言語学が専門の研究代表者と、情報処理が専門の分担者間での密接な学際的共同研究を行う。

稼働段階まで開発を進めた一次データ収集用のシステムを最大限に活用し、目的とする、ユーザーによる定義可能な属性付与を実現するデータベーステーブルの基本設計を完成した。並行して、公刊されている日本人研究者の英語学術論文のコーパス分析と、それを通した一般的構文特徴等の抽出を行い、研究成果を国内外の学会等で発表する。パイロットシステムの試験稼働を実現し、front-end tier における GUI 機能の一層の充実を踏まえてシステムによるユーザーニ

ズパタンの集積と分析を行い、日本人研究者の具体的なニーズを的確に把握するためのアンケート調査や面談による聞き取り調査及び英語論文の校正業務を受託する翻訳支援業者等への聞き取り調査も行う。

コーパス分析の成果としての AWL (academic word list) をモデルとした特定学術分野における語彙リスト及び用法・構文リストの洗練化を試み、コーパス英語学会 39 回大会、BAAL 等の国内外学会で発表を行い、専門研究者からのフィードバックを得てシステムの改良へと繋げる。その後、試作システムの部分的公開・運用と大学英語教育教材との融合接点を模索し、教材への応用に繋げる。特に将来の研究者として、研究型大学に所属する学部学生に対する EGAP 教育への貢献を重視し、具体的な英語学術語彙集の編纂へと繋げる。

システムユーザーの利用状況をサーバー側で集積解析すると共に、各種のアンケート的な聞き取り調査を実施する。特に、基礎的な学術英語論作成技能の養成を目指す研究型大学 (研究代表者の所属する東北大学等) 所属の研究者との意見交換を活発に行う。

外国語教育関係を含む国内外学会 (BAAL 世界大会等) での発表と機関誌及び国際ジャーナルへの論文投稿を通して研究成果を公表する。以降の幅広い学際的な共同研究に向けての研究体制を一層整備し、併せて知的財産権問題処理等も慎重に行う。

4. 研究成果

学習者としてではなく、実務的に英語を用いて学術研究成果を発信する必要性を強く感じる日本人科学研究者をユーザー層とすることに本研究の大きな特色と独創性がある。想定されるユーザーのニーズは明確で、ニーズの捉えにくい一般の学習者に対するシステム開発者側からの一方的なデータや解析結果の提供というボトルネックからの脱却が図れる。ユーザー層の特定化に比例して実現の可能性が増す user friendly さと、そこから保証されるシステム自体の柔軟性が期待される。具体的には、明確なニーズを意識したユーザーである研究者自身が、データ収集や分析対象としての「コーパスセット」として必要なデータ群の統合、及びシステム機能によるアシストを介しての構文情報表示用の独自品詞標識等の付与を含めた、教材マテリアルとしてのコーパス整備を行うことを強力に支援するシステムを研究成果として期待した。

世界水準で発信すべき学術コンテンツを持つ日本人科学研究者が英語学術論文を正確に量産するために効果的な教材を作成するために必要な汎用的コーパス解析システムの全体像を各方面からの検討を加えることによって慎重に設計することができた。

システムの中核を成す RDBMS (Relational Database Management System) の基本的な構

築を、研究分担者との連携により概ね完成させ、その成果を2011年9月1日-3日に英国Bristol市のUniversity of West of Englandで開催された第44回英国応用言語学会での研究発表に繋げた。また国内学会では情報処理学会での研究発表を行った。

平成23年8月7日には、立命館大学田中省作准教授、静岡大学宮崎佳典准教授、京都大学大学院博士課程鈴木大介氏を仙台市泉区の東北学院大学泉キャンパスに招待し、(1)コーパス処理全般に係る著作権処理の問題、(2)日本人ユーザーの入力英文のベクトル解析に準拠した評価と修正候補英文の専用コーパスからの抽出と提示、等に係る研究会を実施した。これにより、既存大規模コーパスの処理とは別に、個々の研究者が独自に開発するコーパス構築に関わる著作権処理の過程を明らかにすることができた。また、日本人ユーザーの入力英文との類似性を算出したうえで、準拠すべき母語話者産出の英文をモデルとして提示するというシステムを、語彙とその連語(collocation)、連辞(colligation)のレベルで精密化できれば、特定英語ジャーナルという限定された言語使用域(register)なり、学術共同体(academic community)に最も受容されやすい表現や文体を伴ったモデル英文を検索して教材として提供するというシステムの目的に合致するということが明らかになった。

加えて、本研究で開発を目指す教材作成システムは、言語使用域・ジャンル研究というコーパス言語学の重要な分野における基礎データを副産物として生成することになり、当該研究の進展に大きく貢献するものである。また、特定学術分野に焦点を絞ったコーパス解析の成果は、特に将来の研究者として研究型大学に在籍する学部生に対する英語教育にとっての科学英語論文ライティング指導用の全く新しい大学英語教材開発とも密接に関連するものである。

本システムの中核を成すRDBMSの柔軟な設計は、ユーザー自身による解析対象としてのコーパスセットの構築や取捨選択を強力に支援することを可能にする。その結果、学術英文の分野種類数を増やし整備することにより、自然科学分野のみならず、医学や広く人文科学分野における標準的で、当該分野の学術共同体に受け入れられやすい、正確で自然な英文の高速で大量の産出に対する支援が保証される。この成果は、リアルタイムで正確な研究成果の発信能力の育成に繋がり、我が国の国際学術界における地位の一層の向上に貢献できる。

研究代表者と共同研究者の研究室間を結ぶVPNルーターは既存の装置を利用することが可能であった。一次データ収集・蓄積用サーバーマシンは分担者の研究室設置の既存装置を援用し、BNC(British National Corpus)等の市販コーパス及びコーパス関連書籍は代表者が整備を終えた。代表者と分担

者の所属大学は共に仙台市内に位置し、学部生・大学院生を含めた研究会の開催や情報交換のための打合わせ会も頻繁に開催することが可能であった。

このような綿密な共同研究体制を研究期間内の3年間に渡って維持し、研究代表者と共同研究者は恒常的に連携を保ちながら研究を推進することができた。この結果、ICAME、BAAL、英語コーパス学会その他多くの国内外学会での共同研究発表及び論文共同執筆等を行うことが可能となった。

専用Webサーバーからの試験的システム運用によりフィードバック情報を収集し、それに準拠したシステムの柔軟性及び性能向上に向けた恒常的な研究の持続を試みた。関連学会での成果公表を通じた国内外の研究者(Lancaster大学Mr Niel Miller、London大学Dr Roger Mitton、Chemnitz工科大学Dr Joseph Schmied、Dr Christoph Haase、立命館大学田中省作准教授等)との協力的な研究を推進した。また東北大学高等教育開発推進センターとの協力によって大学英語教育用教材(LincEnglish等)との接点を重要視した研究成果の実践的英語教育への応用を図った。

次頁に掲げる表は、構築した対象コーパスの典拠先データとしての、材料科学分野における世界的ジャーナルから選出した日本人研究者執筆の論文の一覧の一部である。

日本人研究者が英語で執筆する学術論文の中で見られる特徴と、英語母語話者の同一使用域内での英語表現の特徴とを比較対照するための基礎としてのデータ整備を完成させた。これを著作権問題に配慮しながら今後専用webページからの公開を検討する。日本人科学者産出の英文特徴の抽出先として、主に材料科学(materials science)の分野での世界的なジャーナルデータ7種類から合計563編の論文を慎重に選別した。総単語数は2,307,825で、単語種類数は7,122個にのぼる。この論文選別の際には研究補助者の援助を得て、個々の論文執筆者が、氏名・所属等からほぼ完全に日本人と認定できるもののみを選択しているのは大きな特徴である。これと並行する英語母語話者産出の科学技術論文の典拠先としてはPERC Corpus(<http://scn.jkn21.com/~percinfo/>)を活用した。このことにより、従来は漠然と規範的な意識からしか提案されてこなかった、日本人科学者向けの学術論文執筆のための、コーパス分析に裏付けられた客観的な指標とガイドラインを策定することが可能となったと言える。

研究機関を通して得られた成果は、(1)日本人研究者の特異な英文表現上の構文選択および伝達動詞選択の傾向の把握、(2)特に特定構文や動詞の過度使用を是正するための自律的学習用システム、(3)科学技術論文執筆の前提となる論文読解と意味把握能力の育成のためのブレンディッドeラーニング

University of West of England 大学、
2011 年 9 月 3 日
三浦 宏太、坂本 泰伸、岡田 毅、
RDBMS を中間インターフェイスとする英
文書に対する属性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 9 月 2
日

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

<http://www.elearning.he.tohoku.ac.jp/okada/>

6. 研究組織

(1) 研究代表者

岡田 毅 (OKADA, Takeshi)

東北大学・大学院国際文化研究科・教授

研究者番号：3 0 1 8 5 4 4 1

(2) 研究分担者

坂本 泰伸 (SAKAMOTO, Yasunobu)

東北学院大学・教養学部・准教授

研究者番号：6 0 3 5 0 3 2 8