

機関番号：12603

研究種目：基盤研究（A）

研究期間：2007年度～2010年度

課題番号：19202015

研究課題名（和文） 多言語話しことばコーパスと学習者言語コーパスに
基づく言語運用の研究と教育への応用研究課題名（英文） Studies of language use based on multilingual and learner corpora
and their application to teaching

研究代表者 川口 裕司

KAWAGUCHI YUJI

東京外国語大学・大学院総合国際学研究院・教授

研究者番号：20204703

研究成果の概要（和文）：

本研究プロジェクトの主たる目的は、(1)「多言語による話し言葉コーパス」および「学習者言語コーパス」という2つの言語コーパスを構築すること、(2)コーパスの言語分析を行い、研究成果の言語教育への応用可能性を検討することである。このため海外の高等研究機関との協力関係を構築しつつ、話し言葉コーパス(フランス語、スペイン語、トルコ語、中国語)および書き言葉コーパス(ラオ語、クメール語(カンボジア語))を開発した。また、タスク中心の e-learning 教材に基づいた学習者言語データ収集のためのweb録音システムを開発し、学習者言語コーパス(英語、フランス語、日本語)を構築し、コーパス分析を実施した。

研究成果の概要（英文）：

The objective of the present project is to (1) construct multilingual spoken language and learner language corpora and (2) investigate the application of corpus analysis in language education. With the assistance of overseas universities, we developed both spoken language corpus (French, Spanish, and Turkish) and written language corpus (Lao and Khmer). We also used web-recording systems for collecting learner language data based on e-learning task-based materials and analyzed learner language corpus (English, French, and Japanese).

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	9,800,000	2,940,000	12,740,000
2008年度	10,400,000	3,120,000	13,520,000
2009年度	10,900,000	3,270,000	14,170,000
2010年度	4,200,000	1,260,000	5,460,000
総計	35,300,000	10,590,000	45,890,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：言語コーパス、学習者言語、話しことばコーパス、言語学、応用言語学

1. 研究開始当初の背景

1980年以降、言語研究において言語コーパスを利用したコーパス言語学が隆盛となった。とくに英語では蓄積された学習者の言語データを利用し、コーパス研究の理論と成果を言語教育に応用すべく、1990年代から学習者言語コーパスの研究が始まった。東京外国語大学大学院地域文化研究科では、2002年に本研究代表者をリーダーとする21世紀COE「言語運用を基盤とする言語情報学拠点」が採択された。21世紀COE拠点事業の中で、コーパス言語学と学習者言語コーパスの研究は採択当初から継続的に行われ、かなりの研究成果をあげてきた。なかでも多言語にわたる多数の言語研究者を擁する東京外国語大学では、様々な言語を対象としたコーパス言語学の研究を行い、その成果を公刊した。T. Takagaki et alii, *Corpus-Based Approaches to Sentence Structures*, 2005, John Benjamins 参照。2006年度に21世紀COEプログラムが終了すると、本学では続けて Global COE プログラム「コーパスに基づく言語学教育研究拠点」が採択された。Global COEにおいても言語コーパス、とりわけ話し言葉コーパスと学習者コーパスは重要な研究テーマであった。Y. Kawaguchi et alii, *Corpus Analysis and Variation in Linguistics*, 2009, John Benjamins 参照。研究代表者は Global COE の推進担当者であり、言語情報学班のまとめ役でもあったため、本科学研究費と Global COE の両方で、類似の研究（フランス語話し言葉コーパスおよびトルコ語話し言葉コーパスの構築と分析、フランス語学習者言語コーパスの構築と分析）を遂行することとなった。同様に本科研の分担者のうち、海野多枝も Global COE の推進担当者であり、高垣敏博、望月圭子、林俊成も Global COE の研究協力者であった。このため研究開始時点で、プロジェクト経費については厳密な棲み分けを行うものの、成果物である言語コーパスについては、可能な限りにおいて、本科学研究費と Global COE の両方で利用しながら研究を行っていくことにした。

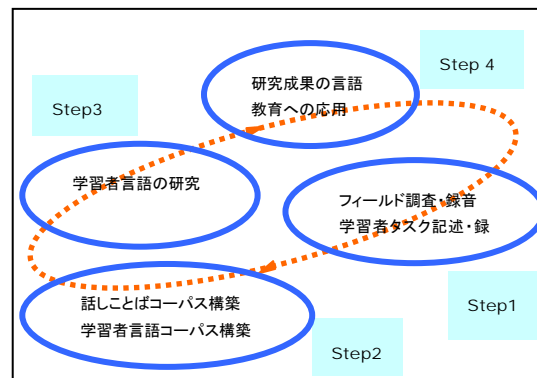
2. 研究の目的

本研究プロジェクトの目的は以下の2つである。(1)「多言語による話し言葉コーパス」と「学習者言語コーパス」という2つの言語コーパスを構築すること、(2)コーパスの言語分析を行い、研究成果の言語教育への応用可能性を検討することである。そのため21世紀COEプログラムにおいて2002年度～2006年度に構築された話し言葉コーパスをさらに拡充し、海外の大学や高等

研究機関との協力関係をさらに強化しながら、教員と大学院生によるフィールド録音を行った。また、学習者言語コーパス構築のための Web を利用した録音プログラムを新たに開発し、それを用いて構築した学習者言語コーパスについて、各学習者の学習過程を分析した。

3. 研究の方法

本研究のプロセスは4つの Step から構成される。Step1 では、現地でのフィールド調査を通じて話し言葉の録音を行い、あるいは既存テキストの電子化を行った。学習者言語については、e-learning 講義等の中で学生にタスクを与え、作文や録音を行った。Step2 では臨地録音で収集されたデータの文字化を行い、それらにメタ言語タグを付与して電子化した。学習者タスクの回答は、開発されたプログラムによって、自動的にサーバーに蓄積され、コーパス化する仕組みを構築した。Step3 では、多言語による話し言葉コーパスの言語分析を行い、言語運用の実態と言語学的特徴を研究した。他方、学習者言語コーパスのデータを用いて、学習過程の解析や学習者の様々な特徴について考察した。最後の Step4 では、言語運用の研究と学習者言語データの分析から得られた知見をどのようにして言語教育に応用できるかを検討した。



4. 研究成果

2011年3月の科学研究費補助金終了の時点において、話し言葉コーパスはフランス語で約50万語、スペイン語で約40万語、トルコ語で約20万語、台湾中国語で約100万語が構築された(21世紀COEからの積算語数)。いずれも当初の予想をはるかに超える規模のコーパスが完成したと言える。

本科学研究費補助金では、言語コーパスおよび学習者言語コーパスの作成のために、50名の院生および教員が参加し、こうした多数の研究協力者が全面的に支援することで言語コーパスが実現した。また一連の言語コーパス構築の作業を通して、東京外国語大学とエクス・マルセイユ大学(フランス語)、マドリード自治大学(スペイン語)、マルマラ

大学（トルコ語）、淡江大学（台湾中国語）の間には、緊密な研究協力体制が生まれた。このうちフランス語とスペイン語については、研究代表者の川口と研究分担者の高垣が Global COE と共同研究を行い、Tree Tagger による品詞タグをコーパスに付与し、品詞検索の可能なサイトを 2009 年に開設した。これらのコーパスを利用した院生あるいは研究分担者等による複数の研究論文が執筆され、学会等でも報告がなされた。

ラオ語とクメール語（カンボジア語）に関しては、研究分担者の鈴木と上田が既存テキストの電子化を行い、クメール語(20 万語)とラオ語(5 万語)のコーパスが完成した。これらについてはテキスト検索も可能になっている。両言語ともコーパスの選定および電子化に際しては、国立プノンペン大学とラオス国立大学の研究者の協力を得た。

学習者言語コーパスも順調に構築された。まず、研究分担者の斎藤が英語のプロソディー誤用分析のための音声データは約 110 名分を収集した。研究分担者の吉富は 94 名のナラティブおよび母語話者との会話データ（約 5 万語）を収集し、電子化が行われた。このコーパスを利用して、研究分担者の斎藤は日本人学習者にみられるアクセント配置の誤用等について、複数の国際会議で発表を行った。

上級日本語学習者の作文データは、278 名分（約 83 万語）に達した。このコーパスについては、学習者の文法・語法における誤用の分析を行い、誤用箇所を検索するためのタグを検討した。研究分担者の望月と海野が院生協力者と共に、一連の研究成果を世界日本語教育大会等の国際会議で報告した。

台湾中国語話し言葉コーパスは 2008 年 12 月に分担者の林および院生が台湾の淡江大学と協力し、約 13 万語の録音収録を行い、翌年に全文が文字化された。

フランス語学習者言語データに関しては、研究代表者の川口が Global COE プログラムと連携をしつつ、フランスの国際プロジェクト「現代フランス語の音韻論 Phonologie du français contemporain (PFC)」の推進者らと協議し、2009 年度に「現代フランス語の中間言語音韻論 InterPhonologie du français contemporain (IPFC)」プロジェクトを新たに立ち上げ、ヨーロッパのみならず、広く世界のフランス語研究者に参加を求めた。このうち IPFC 日本は、プロジェクト設立の 2009 年度より本格的に研究を開始し、反復、文章朗読、自由会話等のタスクに基づく約 110 名の音声データを収集し、研究成果を複数の国際学会で発表した。この学習者言語コーパス収集においては、言語研究と言語教育に利用可能となる音声の録音・再生とサーバーへのアップロードツールを開発し、オープンソース

の e-learning システム Moodle と併用できるようにした。

以上のように、最初に述べた Step3 までは当初の予想をはるかに上回る研究成果を達成することができたと考える。

上記の言語コーパス構築を受けて、2009 年度からは、とくに Step4 の課題である話し言葉コーパスや書き言葉コーパスのデータ解析に着手した。同時に、本科学研究費補助金の最終成果として、Step3 までの個別言語を対象とする言語分析だけでなく、類型および系統を異にする様々な言語について、共通の言語現象を通言語的に分析することも望まれる。そのため「語順」を共通テーマとして、各言語で言語コーパスの分析を行うことにした。2010 年度には英語、スペイン語、トルコ語、ラオス語、クメール語、中国語、日本語について「語順」に関する論文をまとめ、Global COE 研究報告集『コーパスに基づく言語学教育研究報告 7』（2011）に投稿した。

学習者言語コーパス分析については、2011 年 3 月 16 日・17 日に東京外国語大学において国際会議を予定していたが、東日本大震災および東京電力福島第一原子力発電所における事故の影響によってやむを得ず中止となった。しかしながら、国際会議で発表される予定であった諸研究を論文としてまとめ、John Benjamins 社から 2012 年に刊行する予定である。

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 5 件）

Tae Umino, Learning how to learn with media: A case of L2 group self-instruction, in Chan, W.M., Chin, K.N., Nagami, M., & Suthiwan, T. (Eds.) *Media in foreign language teaching and learning*, Singapore: Centre for Language Studies, National University of Singapore., 査読有, 2010, 189-216.

鈴木玲子, Serial verbs in Lao, in *Vientiane 450 years Anniversary of National University of Laos*, 42, 査読有, 2010, 49-58.

望月 圭子, 中国語を母語とする上級日本語学習者によるヴォイスの誤用分析—中国語との対照から—, 『東京外国語大学論集』第 78 号, 査読有, 2009, 85-106.

望月 圭子, 日本語・中国語の自動詞構文の対照研究—‘-e-’自動詞と‘-ar-’自動詞との対照をめぐって—, 『漢日理論言語学研究』, 2009, 査読有, 277-284.

Yuji Kawaguchi, A Corpus-Driven Analysis of -r Dropping in Spoken Turkish, in *Corpus Analysis and Variation in Linguistics*, Y. Kawaguchi, M. Minegishi, J. Durand, eds., John Benjamins, 査

読有, 2009, 281-297.

〔学会発表〕(計5件)

Yuji Kawaguchi, *Langue parlée dans les documents écrits – Cas du français et du japonais-*, *Penser les langues avec Claire Blanche-Benveniste*, 2010年12月4日, École Normale Supérieure, Ulm, Paris/France, 招待講演

鈴木玲子, *A Study of 's i? in Lao*, The Third International Conference on Lao Studies, 2010年7月15日, Khonkaen University

Sylvain Detey, Isabelle Racine, Yuji Kawaguchi, *Du lexique canadien à la phonologie suisse en FLE au Japon: ressources audiovisuelles pédagogiques et données sonores authentiques en milieu hétéroglotte*, *Variétés Variations & Formes du français*, 2010年9月24日, Ecole Polytechnique, Paris/France

望月圭子, 申 亜敏, 漢日語の主語指向型結果複合動詞, 第九屆世界華語文教學研討會, 2009年12月26日, 台北市

Hiroko Saito, *Spelling-to-sound or sound-to-spelling? Errors found among Japanese learners of English*, *Phonetic Teaching and Learning Conference*, 2009年8月7日, University College London

〔図書〕(計4件)

Yuji Kawaguchi, et al., *Corpus Analysis and Variation in Linguistics*, John Benjamins, 2009, 399p.

小野尚之, 申 亜敏, 望月圭子他, 『結果構文のタイポロジー』, ひつじ書房, 2009, 487p.

Yuji Kawaguchi et al., *Corpus-Based Perspectives in Linguistics*, John Benjamins, 2008, 439p.

斎藤弘子, 他, 『スペシャリストによる英語教育の理論と応用』, 松柏社, 2008, 253p.

〔その他〕

ホームページ等

科研 HP: <http://www.tufs.ac.jp/ts/personal/ykawa/index.php?id=171>

話し言葉フランス語品詞検索 : <http://cblle.tufs.ac.jp/tag/fr/index.php?menulang=ja>

話し言葉スペイン語品詞検索 : <http://cblle.tufs.ac.jp/tag/es/index.php?menulang=ja>

6. 研究組織

(1)研究代表者

川口 裕司(KAWAGUCHI YUJI)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 20204703

(2)研究分担者

高垣 敏博(TAKAGAKI TOSHIHIRO)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 00140070

黒澤 直俊(KUROSAWA NAOTOSHI)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 80195586

浦田 和幸(URATA KAZUYUKI)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 50468762

斎藤 弘子(SAITO HIROKO)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 10205669

望月 圭子(MOCHIZUKI KEIKO)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 90219973

鈴木 玲子(SUZUKI REIKO)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 40282777

佐野 洋(SANO HIROSHI)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 30282776

林 俊成(LIN CHUNCHEN)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 70287994

吉富 朝子(YOSHITOMI ASAKO)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号 : 40272611

海野 多枝(UMINO TAE)

東京外国語大学・大学院総合国際学研究院・准教授

研究者番号 : 00251562

上田 広美(UEDA HIROMI)

東京外国語大学・大学院総合国際学研究院・准教授

研究者番号 : 60292992

(3)連携研究者 なし