

令和 6 年 5 月 30 日現在

機関番号：35404
研究種目：基盤研究(C)（一般）
研究期間：2020～2023
課題番号：20K00745
研究課題名（和文）第二言語習得研究向け認知モデリングのための基礎的研究と探索的データ解析基盤の整備

研究課題名（英文）Fundamental research and the establishment of an exploratory data analysis platform for cognitive modeling in second language acquisition research

研究代表者
阪上 辰也（SAKAUE, Tatsuya）
広島修道大学・人文学部・准教授

研究者番号：60512621
交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：本研究は、第二言語習得研究における探索的データ分析と認知モデリングの必要性を明らかにした。先行研究を調査したところ、必要なデータの公開不足が課題として浮上したが、既存の視線計測データや独自収集したログデータの探索的分析を通じて、英語学習者についての興味深い知見が得られた。具体的には、特定の文法項目への処理困難性が観察された一方で、習熟度によっては英語母語話者と類似した処理過程を示す場合もあることが示唆された。これらの成果は、2022年度以降、国内外の複数の学会で報告した。今後は、探索的データ分析をさらに深化させ、その成果を英語学習教材の開発など、教育現場へ還元していくことを目指す。

研究成果の学術的意義や社会的意義

第二言語習得研究において、既存のデータベースのみならず、オンライン教材の利用履歴として得られるログデータなどを活用した探索的データ分析の必要性を示した。また、言語処理過程にかかわる研究を進める上で、コロケーションを中心とした認知モデリングの必要性についても示唆され、新たな分析の観点を示すことができたという点で、学術的意義のある研究であったと言える。また、学習者による困難な学習項目や自律的学習態度などを探索的に分析することにより、さらに効果的で個別最適化された外国語教育方法や教材の開発や、データに基づいた教育実践を促進できる可能性があるという点で、社会的意義のある研究であったと考えている。

研究成果の概要（英文）：This study showed that it is important to use data analysis and cognitive modeling in second language learning research. Previous studies found that not enough data was publicly available, which was a problem. However, by analyzing existing eye-tracking data and new log data, I found interesting insights about learners struggling with certain grammatical items. But I also found that, depending on their English proficiency, they could process language like native speakers. I shared these results at conferences in Japan and other countries in 2022. In the future, I plan to improve our data analysis methods and use the results to help with English education.

研究分野：第二言語習得

キーワード：第二言語習得 認知モデリング 探索的データ分析 コロケーション

1. 研究開始当初の背景

第二言語習得研究の分野において、国外では、ICLE(International Corpus of Learner English) という、異なった言語背景を持つ英語学習者の書き言葉を集めたコーパスが構築されており、データを構成する学習者の母語は 14 言語に及ぶ。規模の大きな学習者コーパスの存在は貴重であり、中間言語同士の大まかな比較は可能だが、データの収集条件がまばらで、厳密な比較を行うことが困難である。また、米国を中心として、読解実験や反応速度の計測、あるいは、視線計測などの様々な心理学的実験手法を用いた研究が数多く行われているが、データが共有される事例が少なく、その結果を再検証することや、結果を一般化するための探索的分析を行うことが難しい状況にある。

国内においては、Japanese EFL Learner (JEFLL) Corpus という日本の中学・高校生 1 万人分の英語学習者による自由英作文コーパスや、The International Corpus Network of Asian Learners of English (ICNALE) という日本人大学生を中心とした大規模な作文コーパスが構築されており、作文時間や作文テーマなどの統一が図られており、一定の統制が取られている。しかしながら、心理学的実験で行うような厳しい統制は取られておらず、第二言語習得研究に用いるための課題が今も残されている。

コーパスの分析には客観的な数値に基づく言語仕様の傾向を把握できるという利点があるが、言語処理の過程を捉えることは困難である。そこで利用できるのが、視線計測という手法であり、これには無意識レベルでの言語処理状況が把握できるという利点がある。一方で、それぞれの手法には方法論上の課題や限界点もある。また、研究者の統計解析技術や言語処理技術の不足もあり、現状では単一的な手法に頼ることが専らであり、国内外ともに、その認知過程が多角的に探られることは十分になされていなかった。

これまでの自身の研究により、日本人英語学習者が、単語という一語から成る語だけではなく、複数の語を組み合わせたコロケーションを産出しており、計量的側面からだけではなく、認知的側面からコロケーション知識の保持・運用、その処理過程の一部を明らかにしてきた。このことを踏まえ、また、上記で述べたデータや分析にかかわる課題を克服すべく、外国語教育研究のための認知モデリングにかかわる基礎的研究と探索的データ解析基盤の整備の必要性が高まっているという着想に至った。

2. 研究の目的

本研究は、第二言語習得研究向けの認知モデリングと探索的データ解析基盤の整備を目指すものであった。この課題に取り組むために、日本人英語学習者を対象とした第二言語習得研究を行うために用いるデータベースとして、いかにして条件を統制してデータを収集すべきなのか、また、いかなる内容を含んだデータが第二言語習得研究の認知モデリングを行なっていく上で有用であるのかといった必要条件を明らかにしようとした。

具体的には、視線計測によるコロケーションの読解実験を行い、コロケーションが他の語句の組み合わせよりも、速く処理されるのかを改めて検証する。その上で、学習者コーパスの調査により、コロケーションの産出頻度を算出して、どのようなコロケーションであれば速く処理されるのかを検討する。最終的には、視線計測実験とコーパスの分析を交えて両者の結果を多面的に検討し、どのような学習者であればコロケーションの認知・処理が速くなるのかについての認知モデリング実施を研究の目的とした。

3. 研究の方法

本研究の主な方法は、下記の 2 点である。

1) 認知モデリングを行うためのデータ収集条件の考察

日本人英語学習者を対象とした第二言語習得研究を行うために用いるデータベースとして、いかにして条件を統制してデータを収集すべきなのか、また、いかなる内容を含んだデータが研究上有用であるのかといった必要条件を明らかにすべく、すでに公開されている学習者コーパスや視線計測データの分析を実施した。これに加え、オンライン教材の利用ログなどのデータなども収集して大規模なデータベースの構築を行い、学習結果や習得結果を示すデータにとどまらず、学習や習得の過程を示すより多角的な分析を可能にするデータについても収集を行った。

2) 多様なデータを用いた探索的分析

コーパスを用いたコロケーション研究で用いられてきた手法のひとつである「N-gram モデル」を利用するなどして、計量的側面からコロケーションを抽出した。また、視線計測データについては、すでに公開されている海外のデータベースを探索的に分析し、母語の違いにより、どのような類似点・相違点があるのかを調査した。これらに加えて、オンライン教材の利用ログデータや学習時に得られた生体データも分析対象として加え、より多角的・探索的分析基盤を設けた。どのような学習者であれば、どのような言語表現をコロケーションとして高速に処理できるのか、また、学習や習得が進むのかについて、その条件や要因について調査した。

4. 研究成果

本研究の成果を3点に分けて述べる。

(1) 第二言語習得研究における探索的データ分析の必要性

2020年度から2021年度においては、どのようなデータを収集すべきか、探索的な分析手法として何を利用するかを検討を行うため、先行研究の調査や情報収集を中心に行なった。以前より、英文の読解データが十分に公開・利用されていないことがわかっていたが、2021年以降、The Multilingual Eye-tracking Corpus (MECO) と呼ばれるデータベースが公開されており、様々な母語話者による読解データが利用できるようになった。データとしては、さまざまな背景を持った学習者による書き言葉のコーパスデータが多数公開され、一部は無償で利用できることが分かっているが、英文読解時のデータが十分に公開されていないことが分かってきた。

また、コロナ禍により、対面式による実験実施を縮小化・中止せざるを得ない状況となる中で、オンライン式での実験方法についても調査を行なった。具体的には、PsychoPy や jsPsych といった無償で利用可能なツールを利用することができること、また、一部の調査過程をアウトソーシングすることにより、さまざまな調査対象者から効率的にデータ収集を行える手段があることなどが明らかとなった。実際に本人が実験に参加しているかどうかの本人確認や実験中の機械トラブル発生時の対応など、オンライン実験の実施やツールの運用に向けての課題も残されているが、オンラインでのデータ収集の一環として、活動量計を用いた生体データ収集を試行でき、思わぬ副産物を得た形となった。結果として、新たな探索的データ分析が可能となった。

従来型の分析方法に加え、本研究により得られたさまざまなデータを探索的に分析することによっても、英語学習者特有の行動の特徴や傾向、ならびに、英語の母語話者との類似点を明らかにすることができる可能性が示され、第二言語習得研究における探索的データ分析の必要性を確認することができた。

(2) コロケーションを中心とした認知モデリングの必要性

コロケーションの多くは、ある単語と単語の組み合わせを中心として構成されるものであり、言語習得上重要な言語表現の一種であることがこれまでの研究からも分かっているが、本研究を進める中で、コロケーションを中心とした認知モデリングの必要性がより一層高まっていることが確認された。

一例として、オープンデータの The Multilingual Eye-tracking Corpus (MECO) を用いて、英語の関係詞の *who* を読んだ際の読解時間を探索的に分析した。ドイツ語やイタリア語などを母語とする各地域の英語学習者の平均的な読解時間よりも、英語母語話者による関係詞 *who* の読解時間が比較的長くなることが分かった(図1)。従来であれば、母語話者の方がより速く読解を行うものと予想されるところであるが、それとは異なる結果が得られており、その要因の一つとして、母語話者よりも学習者の方が、先行詞と関係詞の連なりをより強固なコロケーションとして捉えている可能性が考えられる。

探索的な分析ながら、従来の研究結果やその結果から生じる新たな疑問や課題に取り組む上で、コロケーションを中心とした認知モデリングを行う必要があるという点は、興味深いものであると言える。

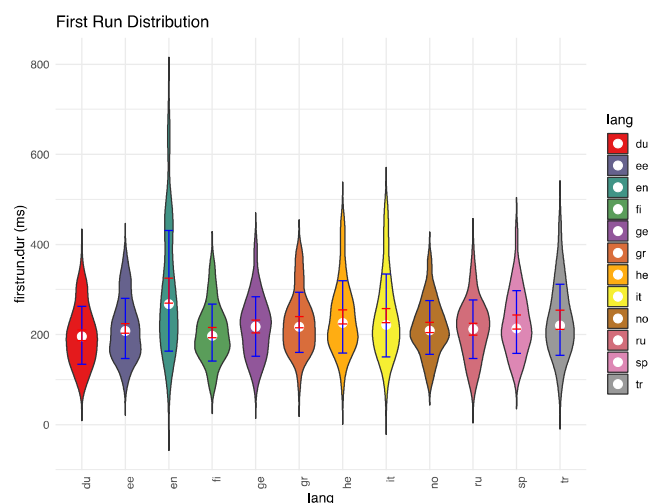


図1：視線計測データの探索的分析から得た関係詞 *who* の読解時間の分布 (enが英語母語話者、他が英語を母語としない学習者)

(3) 学会での成果報告

2022年度には、外国語教育メディア学会（LET）第99回（2022年度秋季）中部支部研究大会にて、日本人英語学習者から得たオンライン教材の利用履歴を分析し、どのような学習項目で困難さを感じるかを分析し、会話や品詞にかかわる問題への解答が困難なことを報告した。

2023年8月には、アイスランドで行われたEuroCALL 2023という国際会議にて、日本人英語学習者から得たオンライン教材の利用履歴などのデータの探索的分析を行い、学習者の自律的学習の態度にかかわる研究発表を行った。また、英語学習者による読解時の視線計測実験で得られたオープンデータが公開されていることがわかり、そのデータベースを探索的に分析した。当該データベースについては、利便性や再現性を高める試みがなされており、データ公開に際して踏襲すべき事項があることがわかった。これらの結果について、2024年2月に統計解析環境のR言語にかかわる国内の研究会にて報告した。

第二言語習得研究向けの認知モデリングに必要とされるデータとして、読解にかかわる視線計測データや語彙テストなどの解答データなどが必要となるが、研究期間全体を通して情報収集やデータ収集を計画通りに進めることが困難となる事態が生じた。そうした状況下においても、実験データや統計解析に用いられるコードなどの共有サイトなどから得られたデータを探索的に分析することにより、英語学習者が特定の文法項目における処理上の困難さが見られること、学習者の習熟度によっては、英語の母語話者と類似した処理を行い得ることを明らかにした。こうした点を踏まえ、認知モデリングにかかわる基礎的研究、また、探索的データ解析基盤の整備の必要性について意義ある報告ができたと考えている。

(4) 今後の展望

本研究は、第二言語習得研究をより発展させるための認知モデリングを実施し、探索的データ分析の基盤の整備を目指すものであったが、今後は、より多くのデータ対象として探索的に分析を行い、その成果を第二言語習得研究分野への貢献のために利用するだけでなく、成果に基づいた英語学習教材の作成など、教育現場への還元も目指したい。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 阪上 辰也、榎田 一路、山内 勝弘	4. 巻 25
2. 論文標題 大学生の生活リズムが英語学習に与える影響に関する予備調査	5. 発行年 2022年
3. 雑誌名 広島外国語教育研究	6. 最初と最後の頁 171 ~ 180
掲載論文のDOI（デジタルオブジェクト識別子） 10.15027/51968	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 阪上 辰也
2. 発表標題 オンライン英語学習教材に含まれるログの探索的データ分析とその課題
3. 学会等名 外国語教育メディア学会（LET）第99回（2022 年度秋季）中部支部研究大会
4. 発表年 2022年

1. 発表者名 Ozawa, S., Nakanishi, D., & Sakaue, T.
2. 発表標題 Promoting English Language Learners' Autonomous Online Learning before and during COVID-19
3. 学会等名 EuroCALL 2023（国際学会）
4. 発表年 2023年

1. 発表者名 阪上 辰也
2. 発表標題 Rによる The Multilingual Eye-tracking Corpus (MECO) の探索的データ分析
3. 学会等名 SappoRo.R #11
4. 発表年 2024年

〔図書〕 計1件

1. 著者名 森田光宏・榎田一路・岩崎克己・天野修一・高橋有加・草薙邦広・阪上辰也・中川篤・鬼田崇作・DAVIES Walter・FRASER Simon・TANABE Julia・SELWOOD Jaime・吉満たか子	4. 発行年 2021年
2. 出版社 溪水社	5. 総ページ数 340
3. 書名 コロナ禍の言語教育	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------