

令和 6 年 6 月 11 日現在

機関番号：15301

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K00813

研究課題名（和文）縦断的コーパスを用いたスピーキング力の発達プロセスと発達要因に関する実証的研究

研究課題名（英文）Formulation and evaluation of developmental indicators for high school students' English speaking ability based on longitudinal data

研究代表者

阿部 真理子（Abe, Mariko）

岡山大学・教育学域・教授

研究者番号：90381425

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：スピーキング評価の予測に大きく寄与する項目として、発話の総語数、前置詞、名詞、冠詞、等位接続詞、動詞の現在形、動詞の過去形、強調表現、副詞、一人称代名詞などの相対頻度が挙げられることが判明した。また語彙の多様性、特に異なる単語の数がスピーキング発達の予測因子として、重要であることも明らかになった。語彙の洗練度は、学習期間と強い関連を示さなかったことから、高等学校における学習段階では、英語を話す能力は、複雑な語彙を使うことよりも、異なる単語を使うことに影響されることが分かった。これらの研究成果から、スピーキング能力の発達における言語項目の特定や学習期間との関連性について新たな知見が得られた。

研究成果の学術的意義や社会的意義

研究成果は、高等学校における教育実践、およびスピーキングテストの開発と評価にとって貴重な資料となるため、本研究の果たす役割は大きい。また、本研究チームはその研究成果を、研究者のみならず、高等学校などの英語科教員を対象とした学会・研究会において公表する活動を行っている。このような活動を通じて、さまざまな領域に関心を持つ聴衆と幅広い意見交換を行うことが可能となり、教育現場における実践的なフィードバックを得ることができるだけでなく、研究成果の普及と応用を促進することができる。

研究成果の概要（英文）：Items reported as significant predictors of speaking assessment included the total number of words in an utterance, relative frequency of prepositions, nouns, articles, coordinating conjunctions, present-tense verbs, past-tense verbs, emphatic expressions, adverbs, and first-person pronouns. Vocabulary diversity, especially the number of different words, is also an important predictor of speaking development. Vocabulary sophistication did not show a strong relationship with the length of language learning, indicating that at the senior high school level, English-speaking ability is influenced more by the use of different words than by the use of complex vocabulary. These findings provide new insights into the identification of language items in the development of speaking ability and their relationship to the length of language learning.

研究分野：第二言語スピーキング研究

キーワード：縦断的コーパス 学習者コーパス スピーキング能力 言語産出 発達指標 学習月数 流暢性 複雑性

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

情報処理技術の急速な発達に伴い、コンピュータを用いて、外国語学習者が書いたり、話したりした産出データを分析する学習者コーパス研究 (Learner Corpus Research: LCR) という分野が生まれた。この分野においては、横断的データなどを用いて、語彙や文法の習得過程の究明を目指す研究が行われてきた。また、コンピュータの技術革新以前には想定されることがなかった規模のデータを解析することで、第二言語習得 (Second Language Acquisition: SLA) における仮説や理論を実証的に再検証してきた。言い換えるならば、学習者が書いた作文と発した音声をコーパスとして整備し、言語処理の技術と統計処理の知見を駆使して分析することで、従来の SLA 研究が踏み込むことのできなかつた課題に対する答えを出したり、それまでの SLA の仮説や理論を覆したりすることができるようになったのである。そして、その成果が外国語教育へ応用されることで、研究の波及効果も認められてきた。しかしながら、これまでの横断的データを用いた LCR のアプローチからだけでは、SLA 研究との効果的なコラボレーションができない場合もあり、両分野が蓄積してきた知見をどのように融合させれば良いのかという問題が議論されるようになった。本研究プロジェクトは、縦断的英語スピーキングコーパスと学習者に関する豊富なメタ情報を併用することで、その問題に一つの解決策を提示したいという構想から開始された。

2. 研究の目的

前年度までの科学研究費採択課題において、収集が完了している二つの異なるデータを利用した分析を行うことを目的としている。具体的には、高校生の発話を3年間にわたって継続的に収集した縦断的スピーキング学習者コーパス (Longitudinal Corpus of Spoken English: LOCSE) と複数の調査方法を使って収集した学習者に関するメタ情報を用いて、高校生のスピーキング能力の発達プロセスを解明すると同時に、その発達に寄与する要因を特定することを目的としている。この研究プロジェクトにおいては、高等学校に在学中、どのような授業を受けてどのような学習を行うと、どのくらいスピーキング力が伸びるのか、あるいはそれほど伸びないのかを確認することで、効果的なスピーキングの指導と評価に対する提言を行うことを最終的に目指している。この目的達成のために、高等学校に入学してから大学の受験期を迎えるまでの23ヶ月にわたり、日本で英語教育を受けてきた英語学習者の追跡調査を行った。

3. 研究の方法

本研究が分析対象としている LOCSE データについて説明する。この学習者コーパスは、類似した環境において、英語を学習している高校生104名が23ヶ月間に8回のモノログ型スピーキングテストを受験した際の発話を書き起こしたものである。調査対象である104名の高校生のスピーキングテストのスコアは、全体的に向上する傾向にあった。8回のテストのうち、最も多かったのはレベル4であったが、レベル3の受験者数は時間の経過とともに減少し、レベル5の受験者数は4回目の受験以降、急激に増加した後一定の増加傾向を示した。23ヶ月間中、スコアが2レベル以上上昇した高校生は24名おり、集団として上昇傾向にあったと言える。発話データの概略を簡単にまとめると、延べ832件の発話サンプル(約400,000語)中、平均して1文あたり5.7語が産出されており、平均的に45秒あたり50語話している。また時間の経過とともに発話量だけでなく、発話の長さも増加している。そして流暢性が向上し、非流暢性の特徴が減少していることが確認された(近藤, 2021)。

(1) 研究事例 I: 発達指標の特定

高校生たちの発達プロセスを明らかにし、そのスピーキング能力を伸ばすにあたって、どのような言語項目(例: 語彙や文法など)がスピーキングテストの評価と関わるのかを調査した(小林, 2021; Kobayashi, Abe, & Kondo, 2022)。合計832件の発話サンプルを統計的に分析することで、スピーキング能力の発達指標となり得る言語学項目を特定するために、発話の産出状況からスピーキングスコアを予測する研究を行った(小林, 2021)。また先行研究において、スピーキング能力とスピーキングの産出語数には高い相関があることが明らかになっているため、同様の手法を用いて、産出語数をスピーキング能力の代替指標とした分析を行なった(Kobayashi, Abe, & Kondo, 2022)。これら二つの研究事例においては、分析にあたって、Multidimensional Analysis Tagger (MAT) (Nini, 2019) を用いている。MATによって集計できる言語項目の使用頻度を対象として、機械学習の手法の一つであるランダムフォレスト(Breiman, 2001) を用いた分析を行った。

(2) 研究事例 II: 語彙的複雑性と学習期間の関連

高校生たちの発達プロセスを明らかにし、そのスピーキング能力を伸ばすことを目指すために、どのような言語項目(例: 語彙や文法など)が、高校生たちの学習期間と関連するのかを調査した(阿部, 2022)。先行研究において、英語学習者の初心者は、まず語彙を向上させた後、文の構造を向上させ、再び語彙を向上させることが明らかになっている(e.g., Verspoor, Lowie & Wieling, 2021)。そこで、まずは語彙項目について着目した。分析手法としては、Lexical Complexity Analyzer (Ai & Lu, 2010; Lu, 2012) を用いた。各発話の語彙的複雑性を説明変数、データ収集地点(学習期間)を目的変数とした重回帰分析を行い、どのような語彙的複雑性の特徴(lexical density, lexical sophistication, lexical variation) がどの程度データ収集地点を予測するのに貢献して

いるかを探った。重回帰分析は、変数増加法(stepwise regression analysis)を用い、変数選択の基準として F 値が 2 以上とした。元のモデルと最終的につくられたモデルは情報量基準や検定統計量で評価した。回帰モデルに関しては、モデルがどの程度データを適切に表現できているか、いくつかの指標で確認した。

(3) 研究事例 III：発達段階の再検証

研究開始当初の背景の節で述べたように、コンピュータの技術革新に伴い、英語学習者が書いたり話したりした大量の言語データを計量的に分析することが可能になったため、従来の第二言語習得研究の仮説や理論が再検証される研究動向がある。一例を挙げると、処理可能性理論 (Pienemann, 1998) が提唱している発達段階と日本人英語学習者の発達は一致していることが、横断的コーパスを用いた検証により確認されてきている。第二言語習得研究における処理可能性理論によると、学習者の文法習得には普遍的な順序があり、語・句・節・複文の順に発達するとされている。その一方、北アリゾナ大学の研究チームは、新しいアプローチとして、一連の発達段階(節の複雑性から句の複雑性に発達するという順序)を提案している。そこで、LOCSE を用いて、句・節の複雑性という観点から、高校生の発話における文法的複雑性の時間的変化の分析を進めた。

(4) 研究事例 IV：本研究に関連して実施した国際共同研究(英国・ロンドン大学)

本研究プロジェクトでは、縦断的スピーキングコーパスを構築するためのスピーキングデータを収集するだけでなく、学習者を取り巻く学習環境やスピーキング能力の発達に関わる要因を特定するために必要な情報(高校生と教員に対するアンケートや授業観察など)を収集している。第二言語習得研究との効果的なコラボレーションを行うために、収集した学習者のメタ情報を用いて、国際共同研究を実施した。一般的に、未熟な段階にある英語学習者は自己能力を過大評価し、熟練した学習者は過小評価する傾向があるとされている。この現象について高校生 122 名の自己評価アンケートを調査した。具体的には、日本で英語を学習する高校生を対象に、外部評価と比較して、自己評価との整合性が時間とともにどのように変化するかを調査した。

4. 研究成果

(1) 研究事例 I：発達指標の特定

高校生の発話(832 件のサンプル)を用いて MAT で算出した言語項目の頻度を説明変数とし、個々の発話のスピーキングスコアを目的変数として予測するランダムフォレストを実行した結果、スコアを正しく予測できた割合は 71.03%であった(小林, 2021)。つまり 832 件中 591 件のスコアを正確に予測できたことに相当する。どの言語項目が、高校生の発話のスコアリングに寄与したのかに関しては、ランダムフォレストを実行する過程で算出される変数重要度 (Variable Importance) の上位 20 項目を確認したところ、発話の総語数 (Tokens) の寄与度が突出して高いことが明らかになった。さらに順を追って、前置詞 (PIN)、名詞 (NN)、冠詞 (DT)、because (CAUS)、等位接続詞 (CC)、動詞の現在形 (VPRT)、動詞の過去形 (VBD)、強調表現 (EMPH)、副詞 (RB)、一人称代名詞 (FPP1) の相対頻度が予測に大きく寄与することが判明した(小林, 2021)。使用頻度の増減に関しては、名詞・because・動詞の現在形・一人称代名詞 の使用頻度は減少していた。その一方、総語数・前置詞・冠詞・等位接続詞・動詞の過去形・強調表現の使用頻度は増加していることが分かった。さらには、この小林(2021)の結果を、産出語数をスピーキングスコアの代替指標として分析した Kobayashi, Abe & Kondo (2022)の結果と突き合わせたところ、スピーキングスコアを両者の研究結果には、同様の言語項目が含まれることが明らかになった(表 1)。Kobayashi, Abe & Kondo (2022)においては、変数重要度による順位づけには変動があったが、順を追って、because (CAUS)、等位接続詞 (CC)、強調表現 (EMPH)、名詞 (NN)、前置詞 (PIN)が予測に寄与することが分かった。これら二つの研究から、学習者の習熟度の推定に有効である言語項目を報告することができた。

表 1
研究成果の比較(注:数字は変数重要度の順位を示す)

小林 (2021)	Kobayashi, Abe, & Kondo (2022)
② 前置詞 (PIN)	① because (CAUS)
③ 名詞 (NN)	② 等位接続詞 (CC)
④ 冠詞 (DT)	③ 強調表現 (EMPH)
⑤ because (CAUS)	④ 名詞 (NN)
⑥ 等位接続詞 (CC)	⑤ 前置詞 (PIN)
⑨ 強調表現 (EMPH)	-

上記のような言語項目の使用頻度が評価基準となることが判明したのであるが、テストのスコアや習熟度だけではなく、産出語数も有効な評価基準として有効であることが Kobayashi, Abe & Kondo (2022)から示唆される。言語テストの習熟度尺度は、テストごとにその評価基準が異なるが、産出語数であれば、多様な状況において適用すること可能となる。

(2) 研究事例 II：語彙的複雑性と学習期間の関連

語彙の複雑さという尺度を用いて、L2 スピーキング発達の予測可能性を探究したところ、語彙の多様性、特に異なる単語の数（異語数）がスピーキング発達の強力な予測因子として、高校生たちの学習月数を予測するために重要であることが明らかになった。本研究では、語彙の洗練度は学習期間と強い関連を示さなかったことから、高等学校における学習段階では、英語を話す能力は、複雑な語彙を使うことよりも、むしろ異なる単語を使うことに影響されることが報告された。高校生の発話においては、洗練された語彙の使用が限られていると言える。

(3) 研究事例 III：発達段階の再検証

節の複雑性から、句の複雑性に発達するという順序は支持されず、句の発達が先行していることが判明した。つまり、発達順序に関する見解が一致しておらず、議論を引き起こす可能性が出てきた（近藤, 2022）。その議論については、これからの研究において順次行う。

(4) 研究事例 IV：本研究に関連して実施した国際共同研究（英国・ロンドン大学）

研究成果の結果として、自己評価と外部評価の違いには、習熟度の違いが反映されていることが明らかになった。習熟度の高い学習者は、自己評価と外部の評価を時間とともに整合させることに成功していた。一方で、習熟度の低い学習者は自己評価を過大評価していたが、時間が経つにつれてある程度の整合性を見せるようになった。さらに、自己評価の整合性には、課外での英語練習にどれだけ積極的に取り組んでいるかが関連していることが示された。つまり、自己評価が実際の実力と整合するようになるためには、課外での積極的な英語練習が重要であることが判明したのである（Saito, Trofimovich, Abe, & In'nami, 2020）。

(5) その他

科研費を使用して開催した国際研究集会として、以下を報告する。

Symposium on Longitudinal Corpus of Spoken English 2023

（LOCSE プロジェクト主催シンポジウム）

<https://sites.google.com/view/locse/symposium>

研究プロジェクトの成果を公表するシンポジウムに先立ち、プログラミング言語の一つであるPythonを使って、学習者コーパスを分析するための無料講習会を開催した。これからの時代、コーパスをもちいた学習者言語の研究がさらに発展することは想像に難くない。英語学習者が学習のある段階で産出することができる言語の特徴を計量的に捉えるために必要な知識と技術を習得することで、個々の学習者と集団の変化をより詳細に理解することができる。シンポジウムの基調講演として、最新の統計モデルを用いた縦断的データの分析手法（ロケーションスケールモデル）をバーミンガム大学の村上明氏から紹介していただいた。この講演において、CDST（Complex Dynamic Systems Theory：複雑動的体系理論）の考え方に沿ったロケーションスケールモデルの有用性の高さが具体的な研究例をもって示された。

本研究プロジェクトの研究成果は随時、以下のウェブサイトにて更新されている。

Longitudinal Corpus of Spoken English (LOCSE) Research Project

<https://sites.google.com/view/locse/>

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 2件/うち国際共著 1件/うちオープンアクセス 2件）

1. 著者名 阿部真理子	4. 巻 72(12)
2. 論文標題 LOCSE研究プロジェクト-高校生のスピーキング産出能力に対する理解を深める	5. 発行年 2024年
3. 雑誌名 英語教育	6. 最初と最後の頁 58-59
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kobayashi, Y., Abe, M., & Kondo, Y.	4. 巻 29
2. 論文標題 Exploring L2 spoken developmental measures: Which linguistic features can predict the number of words?	5. 発行年 2022年
3. 雑誌名 English Corpus Studies	6. 最初と最後の頁 1-18
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 -

1. 著者名 小林雄一郎・近藤悠介	4. 巻 70(6)
2. 論文標題 スピーキングの自動採点	5. 発行年 2021年
3. 雑誌名 英語教育	6. 最初と最後の頁 60-61
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 阿部真理子・藤原康弘	4. 巻 70(5)
2. 論文標題 学習環境と発話の発達を考慮した指導	5. 発行年 2021年
3. 雑誌名 英語教育	6. 最初と最後の頁 62-63
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 藤原康弘	4. 巻 70(4)
2. 論文標題 スピーキングにおける複雑さの発達	5. 発行年 2021年
3. 雑誌名 英語教育	6. 最初と最後の頁 62-62
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 近藤悠介	4. 巻 70(3)
2. 論文標題 スピーキングにおける流暢さの変化	5. 発行年 2021年
3. 雑誌名 英語教育	6. 最初と最後の頁 60-61
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 小林雄一郎	4. 巻 70(2)
2. 論文標題 スピーキング力と関わる言語項目の発達	5. 発行年 2021年
3. 雑誌名 英語教育	6. 最初と最後の頁 62-62
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 阿部真理子	4. 巻 70(1)
2. 論文標題 縦断的スピーキング・コースからわかること	5. 発行年 2021年
3. 雑誌名 英語教育	6. 最初と最後の頁 64-65
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Saito Kazuya, Trofimovich Pavel, Abe Mariko, In'nami Yo	4. 巻 79
2. 論文標題 Dunning-Kruger effect in second language speech learning: How does self perception align with other perception over time?	5. 発行年 2020年
3. 雑誌名 Learning and Individual Differences	6. 最初と最後の頁 101849 ~ 101849
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.lindif.2020.101849	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 該当する

〔学会発表〕 計9件 (うち招待講演 1件 / うち国際学会 2件)

1. 発表者名 阿部真理子
2. 発表標題 縦断的コーパスを利用した高校生のスピーキング能力の発達研究
3. 学会等名 中部地区英語教育学会愛知地区大会 (招待講演)
4. 発表年 2023年

1. 発表者名 阿部真理子
2. 発表標題 縦断的コーパスを利用したスピーキング力の発達研究: 言語発達の指標探究
3. 学会等名 Symposium on Longitudinal Corpus of Spoken English
4. 発表年 2023年

1. 発表者名 近藤悠介
2. 発表標題 縦断的コーパスを利用したスピーキング力の発達研究: 統語的観点から
3. 学会等名 Symposium on Longitudinal Corpus of Spoken English
4. 発表年 2023年

1. 発表者名 近藤悠介・阿部真理子
2. 発表標題 文法的に複雑になるとスピーキングの評価は上がるか？
3. 学会等名 外国語教育メディア学会(LET)第62回全国研究大会
4. 発表年 2023年

1. 発表者名 阿部真理子
2. 発表標題 縦断的スピーキングコーパス(LOCSE)の構築と活用
3. 学会等名 Symposium ON Automated Scoring (SONAS)
4. 発表年 2022年

1. 発表者名 阿部真理子
2. 発表標題 L2スピーキング能力の発達：言語的複雑性からの予測
3. 学会等名 英語コーパス学会語彙研究会
4. 発表年 2022年

1. 発表者名 近藤悠介
2. 発表標題 日本人英語学習者コーパスにおける文法的複雑性の時間的变化
3. 学会等名 英語コーパス学会語彙研究会
4. 発表年 2022年

1. 発表者名 Abe, M., Kondo, Y., Fujiwara, Y., & Kobayashi, Y.
2. 発表標題 Identifying developmental indices of L2 spoken English through statistical modelling
3. 学会等名 Corpus Linguistics (CL) 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 Abe, M., Kondo, Y., Fujiwara, Y., & Kobayashi, Y.
2. 発表標題 A longitudinal study of novice learners' development of complexity
3. 学会等名 Teaching and Language Corpora (TaLC) 2020 (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>Symposium on Longitudinal Corpus of Spoken English 2023 LOCSEプロジェクト主催シンポジウム https://sites.google.com/view/locse/symposium</p> <p>Longitudinal Corpus of Spoken English (LOCSE) Research Project https://sites.google.com/view/locse/</p>
--

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
研究分担者	小林 雄一郎 (Kobayashi Yuichiro) (00725666)	日本大学・生産工学部・准教授 (32665)	

6. 研究組織（つづき）

	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
研究分担者	近藤 悠介 (Kondo Yusuke) (80409739)	早稲田大学・グローバルエデュケーションセンター・准教授 (32689)	
研究分担者	藤原 康弘 (Fujiwara Yasuhiro) (90583427)	名城大学・外国語学部・教授 (33919)	

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1件

国際研究集会 Symposium on LOCSE 2023	開催年 2023年～2023年
-----------------------------------	--------------------

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		
英国	ロンドン大学		