

令和 2 年 5 月 5 日現在

機関番号：32641

研究種目：基盤研究(B)（一般）

研究期間：2016～2019

課題番号：16H03455

研究課題名（和文）縦断的コーパスの構築と日本人高校生の英語スピーキング力の発達過程の解明

研究課題名（英文）Constructing a longitudinal learner corpus to track L2 spoken English

研究代表者

阿部 真理子（Abe, Mariko）

中央大学・理工学部・教授

研究者番号：90381425

交付決定額（研究期間全体）：（直接経費） 13,900,000円

研究成果の概要（和文）：四年間の研究期間内に、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）を完成させた。同時に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力の発達に影響を及ぼす要因を分析できるように、そのデータ整備を行った。さらには、英語スピーキング力の発達過程を解明するための記述研究と発達に寄与する要因を特定するための実証的研究を開始した。

研究成果の学術的意義や社会的意義

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（本研究では、外国語学習者が書いた、話したりした産出データ）を整備し、言語処理の技術と統計処理の知見を駆使することで、従来の第二言語習得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、これまでのSLAの仮説や理論を覆したりすることが可能になった。本研究はこのような研究の基盤となり得る縦断的英語スピーキング学習者コーパスとその学習者に関する豊富なメタ情報を整備した。これらのデータを用いることで、スピーキング力の発達に寄与する要因を特定することが可能となるため、その学術的・社会的意義は大きい。

研究成果の概要（英文）：During the four-year research period, we have completed a Longitudinal Corpus of Spoken English (LOCSE), a collection of senior high school students' utterances over a three-year period. The meta-information on learners was also collected using multiple survey methods to study the developmental process of high school students' speaking skills. We also began a descriptive study to explain the developmental process of English-speaking skills and an empirical study to identify factors contributing to their development.

研究分野：外国語教育

キーワード：高校生の英語スピーキング力 縦断的コーパス 学習者コーパス 英語スピーキングコーパス コーパス構築 メタ情報

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、研究者個人に帰属されます。

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

(1) 『英語学習者コーパス』とは、英語学習者が書いたり、話したりしたものをコンピュータで、計量的に処理できるように電子化したデータのことである。昨今では、第二言語習得の分野でも、コーパスを使った研究が盛んに行われるようになった。しかし、長期間にわたって多くの学習者を追跡した『縦断的英語学習者コーパス』は、世界的に見ても数が少ない。縦断的なデータの重要性は十分に認識されているが、その収集には、膨大な時間と労力を必要とする。話し言葉データの場合、それに加え、音声の書き起こしに大変な負担がかかる。そのため、『縦断的英語学習者コーパス』、とりわけ『発話コーパス』の構築は進んでおらず、国内外における縦断的なデータを用いた研究は、緒についたばかりである。よって、第二言語習得研究の共通基盤となる大規模なコーパスが整備されたならば、そのコーパスは広く利用されることが見込まれる。

(2) 平成 26 年度版「英語教育改善のための英語力調査事業報告」によると、公立高等学校 3 年生 (約 1.4 万人) の 89.5% が CEFR の A1 レベル (初級) と判定された。「将来は英語を使って、国際社会で活躍できるようになりたい」という高校生の意欲の有無に関わらず、スピーキング力が低いという現状は改善されるべきであろう。さらには、将来的に大学入試にスピーキングテストが導入されることを踏まえると、高校生のスピーキング力の発達過程を明らかにすることは、受験生の能力を適切に評価するための客観的な基準を示すという点においても、非常に有益である。

(3) これまで本研究課題の申請者たちは、コーパス言語学や言語処理の手法を駆使して、多種多様な言語項目に関して、9 つの習熟レベルから日本人英語学習者 1281 人の発話の特徴を記述してきた。これらの研究は、試験官が受験生と話をしながらスピーキング力を判定する対話型 **Standard Speaking Test (SST)** の音声データを書き起こした英語学習者コーパスを分析対象としていた。しかしながら、このコーパスは 1281 人の発話の集合体であるため、個人のスピーキング力の発達を観察することができない。そこで、申請者たちは、同一人物の発話を定期的に収集し続けることで、縦断的スピーキングコーパスを構築し、より詳細に発達のパターンを記述するという着想に至った。本研究で構築するコーパスは、世界に類を見ない規模のものとなり、関連する学術分野に大きな影響を与えることが期待される。

(4) 英語教育の側面からも、本研究の果たす役割は大きい。英語で話せる人材を育成するためには、まず学習者がたどるスピーキング力の発達のパターンを記述する必要がある。その後、その記述に基づいて、どのような要因がスピーキング力の発達に影響しているのかを明らかにしなければならない。そしてその要因に基づいて、スピーキング力を向上させる方法を提案する必要がある。そこで申請者たちは、縦断的なデータの収集とあわせて、自宅学習と授業に関する実態調査アンケートを行い、スピーキング力の発達に影響を及ぼしている要因を明らかにするという研究計画を立てた。これは、研究成果を教育現場に還元するために、教育実践への提案を行い、高等学校における到達目標の設定に不可欠なデータを提示することが必要であると考えたためである。

### 2. 研究の目的

本研究の目的は、高校生の英語スピーキング力を 3 年間にわたり縦断的に追跡し、その発達過程を明らかにするために不可欠なデータを整備することである。本研究が構築したデータを用いて、(1) 個人レベルでの経年変化のモデル化、(2) 習熟レベルごとの特徴抽出、(3) 内的・外的な学習者要因が発達に及ぼす影響の解明、を行う。これらの研究課題は、スピーキング力の縦断的な発達研究を促進し、第二言語習得研究のための共通基盤となるコーパスの拡充にもつながる。そして、複数の学術分野 (第二言語習得・コーパス言語学・言語テスト・英語教育) にまたがる研究者が協力・連携することで、日本人高校生のスピーキング力の発達を多層的に把握し、高等学校における教育実践と到達目標の設定、大学入試におけるスピーキングテスト導入の議論に役立てることができる。

### 3. 研究の方法

(1) 本研究は、高校 3 年間にわたり同一学習者の発話を定期的に収集する。具体的には、モノローグ型 **Telephone Standard Speaking Test (TSST)** を受験した高校生の音声データに基づいて、縦断的発話コーパスを構築する。このスピーキングテストは、電話から流れる自動音声の質問に答えるモノローグ型テストである。専門的な訓練を受け、厳しい認定率 (約 3%) の試験に合格した評価官 3 人が、1 人の受験者の習熟レベル (TSST のスコア 1~9) を判定する。またテストの質問は、受験者ごとに 5 タイプの質問群からランダムに抽出されるため、縦断的データを収集するにあたり練習効果は考慮しなくてよい。

(2) さらに、内的な学習者要因 (個人差、習熟レベル、自宅学習の時間と内容、学習動機など)

と外的な学習者要因（テスト形式、教員の指導法、授業内の英語使用率など）に関するアンケートを実施する。アンケートは、高校生だけではなく、授業担当教員と英語科主任も対象とする。そして、コーパスデータとアンケートデータを分析することにより、以下の点を明らかにする。

- ① 個人レベルでのスピーキング力の経年変化
- ② 習熟レベルと学年ごとのスピーキング力の特徴
- ③ テスト形式がスピーキング力の判定結果に及ぼす影響
- ④ 自宅学習と授業、および学習動機がスピーキング力の発達に及ぼす影響

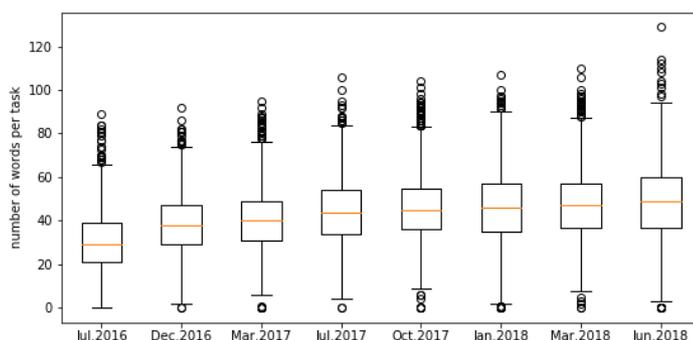
#### 4. 研究成果

(1) コーパスを用いて第二言語の発達プロセスを検証する研究は、1990年代から国内外において注目されてきたが、横断的データに基づく研究に偏っている。日本国内でも、1281人分の発話を書き起こした学習者コーパス (National Institute of Information and Communications Technology Japanese Learner English: NICT JLE Corpus) が2004年に公開されているが、このコーパスは同一学習者を継続的に追跡していないため、個人ごとのスピーキング力の発達を観察することができない。

(2) これまでの横断的データを用いたアプローチからだけでは、第二言語研究との効果的なコラボレーションができないため、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」という問題が提起されるようになった。英国(ランカスター大学・ケンブリッジ大学)と日本(神戸大学・東京外国語大学)の共同研究プロジェクトが立ち上がり、2019年の夏には、国際シンポジウムが開催され、本研究チームも招待された。具体的には、どのような「学習者コーパス」と「メタデータ」を整備すれば、より良く両者の知見を活用した研究ができるのかという議論が行われた。現在、非常に注目を浴びているこの問いかけに対して、本研究チームは「縦断的英語スピーキング学習者コーパス」と「学習者に関する豊富なメタ情報(内的・外的な学習者要因)」を併用するという方法を提示した。これら二つの異なるデータを用いて、「高校生のスピーキング力はどのように発達するのか」、また「何がその発達に影響しているのか」という問いかけに対する答えを希求し、そして最終的には「どうすればスピーキング力を伸ばすことができるのか」という研究につなげることを提案した。

(3) 上記に述べたような研究を遂行するため、同一人物の発話を数年間にわたり定期的に収集し、縦断的スピーキングコーパスを完成させた。データ収集ポイントは全8回(2016年7月、2016年12月、2017年3月、2017年7月、2017年10月、2018年1月、2018年3月、2018年6月)であり、3年間の高校生活を通して全8回全てのデータ収集に協力してくれた高校生は合計108名であった。以下の図は、一つのスピーキングタスク内の発話数を示しているのであるが、全8回のデータ収集ポイントを通して発話数が増加しているのが分かる。そして本研究は、コーパスデータの収集と同時に、複数の調査方法(実態調査・インタビュー・授業観察)を用いて学習者のメタ情報を収集し、スピーキング力の発達に影響を及ぼしている要因を特定するために必要なデータも整備した。

(4) これまでの研究では、習熟度が高い上級者の書き言葉のデータ(作文)に基づくものが多いため、学習者の英語力がどのような過程を経て発達するのかという全体像は掌握できていない。このような状況の中で、大規模スピーキングコーパス(Trinity Lancaster Corpus: TLC)が2019年に公開されたが、このコーパスも縦断的データを扱ってはならず、学習者の習熟度も中級から上級レベルに相当する。しかしながら、本研究チームが構築したLOCSEは、学習者の習熟度も初級レベルに相当するため、「初級レベルの学習者がどのようなプロセスを辿りながら中級レベルに到達するか」という習得の過程を検証することができる。よって、国内外に先駆けて大規模な縦断的スピーキングデータとその学習者のメタ情報を統合的に分析することを可能にした成果は大きい。将来的にLOCSEが公開された暁には、世界中で広く利用されることが期待される。



5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件／うち国際共著 1件／うちオープンアクセス 6件）

1. 著者名 Abe, M. & Kondo, Y.	4. 巻 29
2. 論文標題 Constructing a longitudinal learner corpus to track L2 spoken English	5. 発行年 2019年
3. 雑誌名 Journal of Modern Languages	6. 最初と最後の頁 23-44
掲載論文のDOI（デジタルオブジェクト識別子） <a href="https://doi.org/10.22452/jml.vol29no1.2">https://doi.org/10.22452/jml.vol29no1.2</a>	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 Saito, K., Dewaele J-M., Abe, M., & In'nami, Y.	4. 巻 68
2. 論文標題 Motivation, Emotion, Learning Experience, and Second Language Comprehensibility Development in Classroom Settings: A Cross-Sectional and Longitudinal Study	5. 発行年 2018年
3. 雑誌名 Language Learning	6. 最初と最後の頁 709-743
掲載論文のDOI（デジタルオブジェクト識別子） <a href="https://doi.org/10.1111/lang.12297">https://doi.org/10.1111/lang.12297</a>	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
1. 著者名 Narita, M., Kobayashi, Y., & Abe, M.	4. 巻 20
2. 論文標題 Recurrent multiword combinations in L1 and L2 argumentative essays.	5. 発行年 2018年
3. 雑誌名 Asian EFL Journal	6. 最初と最後の頁 250-275
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 Kobayashi, Y., & Abe, M.	4. 巻 20(2)
2. 論文標題 A corpus-based approach to the register awareness of Asian learners of English	5. 発行年 2016年
3. 雑誌名 Journal of Pan-Pacific Association of Applied Linguistics	6. 最初と最後の頁 1-17
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 Kobayashi, Y., & Abe, M.	4. 巻 20(1)
2. 論文標題 Automated scoring of L2 spoken English with random forests	5. 発行年 2016年
3. 雑誌名 Journal of Pan-Pacific Association of Applied Linguistics	6. 最初と最後の頁 55-73
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 Kobayashi, Y.	4. 巻 11
2. 論文標題 Heat map with hierarchical clustering: Multivariate visualization method for corpus-based language studies.	5. 発行年 2016年
3. 雑誌名 NINJAL Research Papers	6. 最初と最後の頁 25-36
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 -

〔学会発表〕 計11件 (うち招待講演 3件 / うち国際学会 8件)

1. 発表者名 Abe, M.
2. 発表標題 Creating a longitudinal corpus of L2 spoken English: Construction process and possible applications
3. 学会等名 International Learner Corpus Symposium (LCSAW 4) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Abe, M., Kondo, Y., Kobayashi, Y., Murakami, A., Fujiwara, Y.
2. 発表標題 A longitudinal study of L2 spoken English: Development of fluency and pronunciation
3. 学会等名 Learner Corpus Research (LCR) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 阿部真理子
2. 発表標題 縦断的コーパスを用いて英語スピーキング力の発達を探る：語彙の観点から
3. 学会等名 英語コーパス学会2018年度春季研究会
4. 発表年 2018年

1. 発表者名 阿部真理子
2. 発表標題 自動採点におけるエラー情報活用の可能性
3. 学会等名 早稲田大学CCDL研究所第3回シンポジウム（招待講演）
4. 発表年 2018年

1. 発表者名 Abe, M., Kondo, Y., & Kobayashi, Y.
2. 発表標題 Constructing a longitudinal learner corpus
3. 学会等名 The 12th International Free Linguistics Conference（招待講演）（国際学会）
4. 発表年 2018年

1. 発表者名 Abe, M., Kondao, Y., Kobayashi, Y., Murakami, A., & Fujiwara, Y.
2. 発表標題 Initial findings from a longitudinal learner corpus: A year-long development of L2 speaking performance
3. 学会等名 The 13th Teaching and Language Corpora Conference 2018（国際学会）
4. 発表年 2018年

1. 発表者名 Kondo, Y., Abe, M., Ishii, Y., & Kobayashi, Y.
2. 発表標題 The relationship between word error rate in automatic speech recognition and proficiency of L2 speech
3. 学会等名 The Pan-Pacific Association of Applied Linguistics ( 国際学会 )
4. 発表年 2018年

1. 発表者名 Kobayashi, Y., Kondo, Y., & Abe, M.
2. 発表標題 Predicting EFL learners' oral proficiency levels in monologue tasks
3. 学会等名 Asia Pacific Corpus Linguistics Conference 2018 ( 国際学会 )
4. 発表年 2018年

1. 発表者名 Abe, M., Fujiwara, Y., Kobayashi, Y.
2. 発表標題 Tracking L2 language development through construction of longitudinal spoken learner corpus
3. 学会等名 LCR2017 ( 国際学会 )
4. 発表年 2017年

1. 発表者名 Abe, M.
2. 発表標題 Comparing errors across L2 written and spoken error-tagged EFL learner corpus.
3. 学会等名 Paper presented at the meeting of the 12th International Conference on Teaching and Language Corpora Conference (TaLC) 2016 ( 国際学会 )
4. 発表年 2016年

〔図書〕 計1件

1. 著者名 Abe, M. Comparing errors across an L2 spoken and written error-tagged Japanese EFL learner corpus.	4. 発行年 2019年
2. 出版社 John Benjamins	5. 総ページ数 157-174
3. 書名 G. Sandra, & J. Mukherjee (Eds.), Learner Corpora and Language Teaching	

〔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
研究 分担者	小林 雄一郎  (Kobayashi Yuichiro)  (00725666)	日本大学・生産工学部・助教   (32665)	
研究 分担者	藤原 康弘  (Fujiwara Yasuhiro)  (90583427)	名城大学・外国語学部・准教授   (33919)	