

令和 4 年 6 月 17 日現在

機関番号：14303

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K00849

研究課題名(和文) 日本の大学生によるCBT英語スピーキングテスト回答音声データに基づくコーパス構築

研究課題名(英文) Development of a corpus based on speech responses by Japanese university students to a computer-based English speaking test

研究代表者

神澤 克徳 (Katsunori, Kanzawa)

京都工芸繊維大学・基盤科学系・助教

研究者番号：00747024

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)：本研究では、京都工芸繊維大学において学部1年次生全員を対象に実施したCBT英語スピーキングテスト(KIT Speaking Test)の解答音声データに基づいてコーパスを構築し、KIT Speaking Test Corpusとして専用のウェブサイト(<https://kitstcorpus.jp>)を通して無償公開した。また、そのコーパスを用いて基礎的な分析を行い、日本語を母語とする英語学習者のスピーキング能力の特徴の一端を明らかにした。

研究成果の学術的意義や社会的意義

現在の日本の英語教育は生徒や学生のスピーキング能力やコミュニケーション能力の向上に主眼がおかれている。スピーキングの指導や評価のためには、日本語を母語とする英語学習者のスピーキング能力の現状や特徴を把握することが不可欠であるが、そのためのデータが不足している。本研究では、日本の大学生を対象に実施したスピーキングテストの解答音声を用いたコーパスを構築することで、その問題の解決に貢献できると考える。今後、このコーパスを用いた分析が行われることで、日本語を母語とする英語学習者のスピーキング能力の特徴が解明され、スピーキング指導やスピーキングテストの開発・改良など場面での応用が期待される。

研究成果の概要(英文)：In this project, we have developed a corpus based on the speech responses of a computer-based English speaking test (KIT Speaking Test) administered to all first-year undergraduate students at the Kyoto Institute of Technology. We have named the resource “the KIT Speaking Test Corpus” and made it available on our website (<https://kitstcorpus.jp>). We have also conducted some fundamental analysis and revealed some of the characteristics of the English ability of learners whose native language is Japanese.

研究分野：言語学、外国語教育

キーワード：学習者コーパス 話し言葉コーパス CBT英語スピーキングテスト 日本語を母語とする英語学習者 外国語教育

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

急速に進展するグローバル化に対応するため、日本の英語教育ではスピーキング能力やコミュニケーション能力の育成が急務とされている。このことを受けて、中高の学習指導要領は生徒のコミュニケーション能力の向上に主眼をおいた内容となっている。また、2020年度から新たに導入される「大学入学共通テスト」の英語試験において、スピーキングテストが課されることも発表された。このような現状において、生徒や学生にスピーキング指導を行ったり、彼らのスピーキング能力を適切に評価したりするためには、日本語を母語とする英語学習者のスピーキング能力の現状やその特徴を把握することが不可欠である。しかし、この分野の研究は発展途上段階にある。その一因として、元となるデータが圧倒的に不足していることが挙げられる。現時点で日本語を母語とする英語学習者を対象とする話し言葉コーパスはNICT JLE CorpusやICNALEなど数えるほどしか存在しない。

一方、研究代表者は京都工芸繊維大学において、2014年度より学部1年次生全員を対象に、CBT英語スピーキングテスト(KIT Speaking Test)を定期実施しており、受験者から公開の承諾を得た解答音声データを保有している。このデータを書き起こし、コーパス化することによって、日本語を母語とする英語学習者のスピーキング能力の解明につながるデータを提供できると考えた。

2. 研究の目的

この研究の目的は、前述のKIT Speaking Test1回分(574名×9問分)の解答音声データに基づいてコーパス(KIT Speaking Test Corpus)を構築し、公開することである。また、構築したコーパスを用いて分析を行い、日本語を母語とする英語学習者(大学生)のスピーキング能力の特徴を明らかにし、スピーキング指導やスピーキングテストの開発・改良等に貢献できるデータを提供することである。

3. 研究の方法

本研究は、2019年度から2021年度までの期間に、(1)データの書き起こしとタグ付け、(2)ローカル版コーパスの構築とデータ分析、(3)オンライン上でのコーパス公開、の3段階に分けて行う。2019年度、2020年度は主に(1)と(2)を行い、最終年度である2021年度には主に(3)を行う。

(1)については研究代表者が中心となり、京都近郊に住む英語教育や言語学関係の大学院生に依頼して行う。作業はMicrosoftのVideo Indexerで自動書き起こししたものを手作業で修正し、タグ付けを行う。書き起こし・タグ付けの方法は先行のコーパスであるNICT JLE Corpusに準拠する。そのことで、両者の対応分析が可能となる。作業の質を担保するため、作業マニュアルを作成する、関係者間でいつでも連絡が取り合えるようにする、定期的にミーティングを開催する、書き起こしの一致率を確認するなどモニタリングを行う、などの対策を講じる。

(2)は研究代表者と分担者のうち、データ分析班(神澤、小林、李、光永、田中)が中心となって行う。まずは先行研究(投野 2004)で行われているような基礎的な分析を行い、データの全体的な傾向を明らかにする。

(3)はコーパス公開班(神澤、森)が中心となって行う。専用のウェブサイトを作成し、誰もが無償でコーパスをダウンロードし、利用できる仕組みを構築する。

4. 研究成果

当初の計画通り、2021年度末にコーパスの構築を完了し、KIT Speaking Test Corpusとして専用のウェブサイト(<https://kitstcorpus.jp>)において一般公開した(図1)。ユーザーは自身の簡単な情報を登録することでコーパスを無償でダウンロードし、使用できる。海外研究者の使用も想定し、サイトの言語は英語とした。今後、国内外においてこのコーパスを用いた研究が行われることによって、日本語を母語とする英語学習者のスピーキング能力の特徴の解明や他言語を母語とする英語学習者との比較等がなされ、スピーキング指導やスピーキングテストの開発・改良などの場面での応用が期待される。

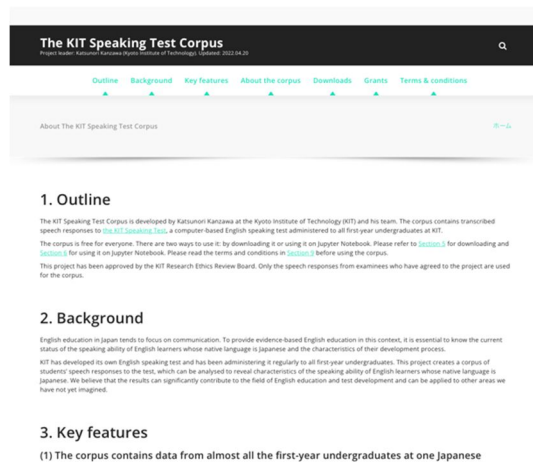


図1 KIT Speaking Test Corpus 公開用ウェブサイト

また、研究チームは、コーパスの一般公開に先立ってコーパスを用いて基礎的な分析を行った。神澤、小林、田中(2020)は、スピーキングテストのスコアに基づいて学生の習熟度を4レベルに分け、各レベルの発話におけるワードリスト、3-gram リスト、キーワードリストを確認した。キーワードリストの分析(図2)では、投野(2004)などで指摘されているような、日本語を母語とする英語学習者のスピーキングの一般的な特徴が見られた。具体的には、習熟度の低い学生には、フィラー、繰り返し、日本語などが特徴的だった一方、習熟度の高い学生には、一般的に中級以上に見られる3人称代名詞、副詞、現在形以外の時制、接続詞や前置詞による句や節の発達、that 節の増加などが確認できた。その一方で、ワードリストの分析(図3)では、一般的な傾向とは異なる傾向も見られた。一般的に、日本語を母語とする英語学習者は、冠詞の脱落エラーが多いとされるが、今回のデータでは、どのレベルにおいても冠詞が比較的多く用いられていることがわかった。

本研究メンバーは、本科研プロジェクトに続く研究として、基盤(C)22K00736(2022年度~2024年度)を獲得した。今後はスピーキングテストの重要な評価観点である Task Achievement の構成要素の解明を中心として、本コーパスの分析をさらに進めていく。

	-39	40-49	50-59	-60
1	fff	fff	game	maybe
2	rrr	osaka	mary	and
3	nvs	study		or
4	jp			it
5	handball			mean
6	singing			like
7	mountain			to
8	freedom			research
9	snowboarding			that
10	susan			quite
11				will
12				experiment
13				kind
14				chengdu
15				husband
16				taro
17				year
18				on
19				about

キーワード抽出は、対数尤度比検定を使用(閾値は、 $p < 0.05$)

図2 キーワードリストの分析

	-39	40-49	50-59	-60
1	fff	fff	fff	fff
2	rrr	rrr	sc	and
3	i	sc	rrr	to
4	sc	i	i	i
5	is	is	is	sc
6	to	and	and	rrr
7	and	to	to	the
8	the	the	the	is
9	a	so	so	so
10	so	a	a	a
11	s	he	he	he
12	because	in	in	it
13	he	s	but	in
14	want	she	for	my
15	but	think	want	think
16	in	but	she	that
17	my	want	because	s
18	she	because	think	of
19	very	for	s	very
20	jp	my	very	because

図3 ワードリストの分析

参考文献

神澤克徳、小林雄一郎、田中悠介. 2020. 「大学生を対象とする英語スピーキングテストの回答音声に基づくコーパス構築」英語コーパス学会第46回大会.
 投野由紀夫. 2004. 「The NICT JLE Corpus に見る英語学習者の発表語彙の使用状況」、和泉絵美、内元清貴、井佐原均『日本人1200人の英語スピーキングコーパス』96-112. 東京: アルク.

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 5件／うち国際共著 0件／うちオープンアクセス 4件）

1. 著者名 光永悠彦、羽藤由美、神澤克徳	4. 巻 -
2. 論文標題 英語スピーキングテストにおけるテストデザインの違いがパラメタ推定値に及ぼす影響	5. 発行年 2020年
3. 雑誌名 日本教育心理学会第62回総会発表論文集	6. 最初と最後の頁 387-387
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 神澤克徳、羽藤由美	4. 巻 23
2. 論文標題 CBTスピーキングテストの舞台裏、どこがどう難しいのか？KIT Speaking Testの実践より	5. 発行年 2021年
3. 雑誌名 JACET Kansai Journal	6. 最初と最後の頁 96-120
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 光永悠彦、神澤克徳	4. 巻 23
2. 論文標題 KIT Speaking Testにおける共通尺度化の方法について	5. 発行年 2021年
3. 雑誌名 JACET Kansai Journal	6. 最初と最後の頁 121-129
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 Katsunori Kanzawa, Haruhiko Mitsunaga, Glen Edmonds, Yumi Hato, Yasushi Tsubota, Masayuki Mori, Yuko Shimizu	4. 巻 14
2. 論文標題 Development and administration of a Skype-based English speaking test in a Japanese high school	5. 発行年 2022年
3. 雑誌名 Bulletin of Kyoto Institute of Technology	6. 最初と最後の頁 27-47
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 神澤 克徳, 田中 悠介, 井上 優大	4. 巻 -
2. 論文標題 第二言語の発話における自己訂正と学習者の習熟度の関連性	5. 発行年 2019年
3. 雑誌名 言語処理学会第25回年次大会発表論文集	6. 最初と最後の頁 982-985
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 神澤 克徳, 森 真幸, 羽藤 由美	4. 巻 -
2. 論文標題 インターネットが利用できるIBT英語ライティングテストの実現に向けた予備調査	5. 発行年 2019年
3. 雑誌名 日本教育工学会2019年秋季全国大会講演論文集	6. 最初と最後の頁 483-484
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 光永 悠彦, 羽藤 由美, 神澤 克徳	4. 巻 -
2. 論文標題 英語スピーキングテストにおけるスコアの共通尺度化法 京都工芸繊維大学における事例から	5. 発行年 2019年
3. 雑誌名 日本テスト学会第17回大会発表論文抄録集	6. 最初と最後の頁 188-189
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 光永 悠彦, 羽藤 由美, 神澤 克徳	4. 巻 -
2. 論文標題 少数カテゴリーの併合がパラメタ推定に及ぼす影響 パフォーマンステストデータに多相ラッシュモデルを適用した場合	5. 発行年 2019年
3. 雑誌名 日本教育心理学会第61回総会発表論文集	6. 最初と最後の頁 477-477
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9件（うち招待講演 2件 / うち国際学会 0件）

1. 発表者名 神澤克徳、森真幸
2. 発表標題 ライティングテストにおけるインターネットの使用が受験者の回答に与える影響
3. 学会等名 日本教育工学会2020年秋季全国大会
4. 発表年 2020年

1. 発表者名 光永悠彦、羽藤由美、神澤克徳、Glen Edmonds
2. 発表標題 英語スピーキングテストにおけるテストデザインの違いがパラメタ推定値に及ぼす影響
3. 学会等名 日本教育心理学会第62回総会
4. 発表年 2020年

1. 発表者名 神澤克徳、小林雄一郎、田中悠介
2. 発表標題 大学生を対象とする英語スピーキングテストの回答音声に基づくコーパス構築
3. 学会等名 英語コーパス学会第46回大会
4. 発表年 2020年

1. 発表者名 神澤 克徳, 田中 悠介, 井上 優大
2. 発表標題 第二言語の発話における自己訂正と学習者の習熟度の関連性
3. 学会等名 言語処理学会第25回年次大会 (NLP2019)
4. 発表年 2019年

1. 発表者名 光永 悠彦, 羽藤 由美, 神澤 克徳
2. 発表標題 英語スピーキングテストにおけるスコアの共通尺度化法 京都工芸繊維大学における事例から
3. 学会等名 日本テスト学会第17回大会
4. 発表年 2019年

1. 発表者名 神澤 克徳, 森 真幸, 羽藤由美
2. 発表標題 インターネットが利用できるIBT英語ライティングテストの実現に向けた予備調査
3. 学会等名 日本教育工学会 2019年秋季全国大会
4. 発表年 2019年

1. 発表者名 羽藤 由美, 神澤 克徳, 光永 悠彦
2. 発表標題 大規模スピーキングテストの舞台裏、どこがどう難しいのか？ 京都工芸繊維大学の実践より
3. 学会等名 2019年度JACET関西支部大会（招待講演）
4. 発表年 2019年

1. 発表者名 森 真幸, 神澤 克徳
2. 発表標題 IBT英語ライティングテストを公平・公正に実施するためのシステム環境の構築
3. 学会等名 FLEXICT Expo 2019
4. 発表年 2020年

1. 発表者名 森真幸, 神澤克徳
2. 発表標題 情報インフラから見たKIT英語スピーキングテストプロジェクト運営の裏側
3. 学会等名 情報処理学会教育学習支援情報システム (CLE) 研究会 第31回研究会 (招待講演)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>The KIT Speaking Test Corpus https://kitstcorpus.jp</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	李 在鎬 (Lee Jaeho) (20450695)	早稲田大学・国際学術院 (日本語教育研究科)・教授 (32689)	
研究分担者	小林 雄一郎 (Kobayashi Yuichiro) (00725666)	日本大学・生産工学部・講師 (32665)	
研究分担者	森 真幸 (Mori Masayuki) (90528267)	京都工芸繊維大学・情報工学・人間科学系・助教 (14303)	

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	光永 悠彦 (Mitsunaga Haruhiko) (70742295)	名古屋大学・教育発達科学研究科・准教授 (13901)	
研究分担者	田中 悠介 (Tanaka Yusuke) (10908825)	福岡大学・公私立大学の部局等・外国語講師 (37111)	

	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
研究協力者	近 大志 (Chika Taishi)		
研究協力者	加藤 満幸 (Kato Mitsuyuki)		
研究協力者	北原 匠 (Kitahara Takumi)		
研究協力者	谷脇 聡史 (Taniwaki Satoshi)		
研究協力者	本澤 拓 (Motozawa Taku)		

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
---------	---------