

平成 21 年 5 月 20 日現在

研究種目： 基盤研究（B）
 研究期間：2006～2008
 課題番号： 18320093
 研究課題名（和文） 学習者プロファイリングに基づく
 日本人英語学習者音声コーパスの構築と分析
 研究課題名（英文） Utterance Data Collection from College Learners of English
 in Japan: Toward Construction of a Learners Corpus of English
 研究代表者
 原田 康也（HARADA, Yasunari）
 早稲田大学・法学大学院・教授
 研究者番号：80189711

研究成果の概要：

2006年度から2008年度にかけて早稲田大学法学部1年生3クラス75名前後を主な対象とし、毎回の授業における応答練習での発話をデジタル録音機・デジタルビデオカメラで収録した。受講生の英語学習歴などについてアンケートを実施し、学年の初め・半ば・終わりに Versant for English(口頭英語自動試験)を受験したスコアを記録した。CMS と DBMS を組み合わせた web インタフェースを構築して、特定の属性を有する発話音声にアクセスする仕組みを用意し、これを用いてアルバイト作業者が収録した音声の一部について書き起こしなどのアノテーションを付与したデータベースを構築した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	6,200,000	1,860,000	8,060,000
2007年度	5,400,000	1,620,000	7,020,000
2008年度	2,300,000	690,000	2,990,000
年度			
年度			
総計	13,900,000	4,170,000	18,070,000

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：英語教育・学習者コーパス・学習者発話コーパス・学習者プロファイリング・デジタル無圧縮録音・ハードディスクデジタルカメラ

1. 研究開始当初の背景

- (1) British National Corpus の構築と公開に伴って、外国語・第二言語の学習に関して大規模に収集された母語話者の言語資料(コーパス)を辞書・教材・シラバス・試験の開発に活用する動きが1980年代以降世界的に顕著になり、日本においても(特に英語の学習に関して)British National Corpus などの分析に基づいて学習語彙を検討するといった動きが1990年代後半から目立ってきている。こうした母語話者の産出する発話・文章を集積した言語資料の活用と合わせて、外国語・第二言語の学習者の産出する発話や文章を研究資料として収集し、学習者に特徴的な『誤り』や母語の影響を分析することを通じて、よりよい学習方法や教材の作成につなげようという動きも国内外の研究開発動向として重要な方向性を示していた。
- (2) 英語学習者のコーパスについては、本研究計画が始まった時点で公開されていたものとしては、作文など文字テキストが中心となっていた。日本人英語学習者の作成した英語作文を収集し、構築した学習者コーパスを分析して英語教育の基本的な資料とすることを目指す研究活動としては ICLE (International Corpus of Learner English) の世界的な共同研究活動に日本から昭和女子大学の研究グループなどが加わっていた。
- (3) 日本人英語学習者による発話コーパスについては本研究計画の研究分担者(2006年度・2007年度)・連携研究者(2008年度)の井佐原均が中心となって作成された NICT JLE (Japanese Learners of English) コーパス(1200人の英語発話コーパス)が公開されていたが、音声データは提供されず、書き起こしデータのみ提供となっていた。作文や書き起こしなどの文字データと比較すると、音声データや画像データは属人性が強く、プライバシーへの配慮や著作権・肖像権などの権利関係をどのように整理していくか付随的な課題も多いが、学習者コーパスを使った研究の活性化という意味では、今後ますます強く求められるものであることは明らかであった。
- (4) 従来の英語学習者の音声データの収集は、教室における授業実践を担当教員が手持ちの機材で録音したものが、実験音声学的に整えた環境で授業実践とは無

関係に録音した(読み上げなどの)音声データかに2分する傾向があったが、本研究計画においては、研究代表者の担当する授業の中で学習活動の一環として比較的自由的な発話をデジタル録音機し、学習者の外部試験のスコアや学習履歴とつき合わせて分析することが可能となるように目指した。

2. 研究の目的

- (1) 本研究は、日本の大学で英語を学ぶ学習者の英語運用能力のうち、特にスピーキングやライティングなど production に関わる技能について、どのようなレベルの学生に何ができて何ができないかを明らかにし、それぞれの学生に最適な学習素材と学習方法を提供するための基礎的な資料を構築することを最終的な目標として、英語学習者の学習履歴と運用技能の関係を明らかにすることが可能となるような学習者コーパスを構築することを目的とした。
- (2) 対象とするそれぞれの学生について、学習経歴などについてアンケート調査を行い、外部試験のスコアを記録し、授業での学習活動について、発話音声デジタル無圧縮録音するとともに、活動の状況をデジタル・ビデオに記録し、これらのデータを総合的に分析することを可能とするデータベースを構築することとした。
- (3) 発話データ収集の主な対象となる学習活動である応答練習では、受講生が3人ずつのグループに分かれ、一人が質問者、一人が回答者、一人がタイムキーパーとなって、回答者は質問者が読み上げる質問を耳で聞いて直ちに答え、その応答を質問者とタイムキーパーが採点し、一問ごとに役割を交代する。毎回の授業で授業担当者があらかじめ10の質問を用意する。質問に対しては、比較的短い時間で自発的に内容を考え英語で表現することが求められる。
- (4) この応答練習は90分の授業時間のうち25分前後を使い、残った時間の一部を使って、応答練習の際のやりとりを思い出しながら、その日のテーマについて30分ほどで400語を目標に作文をまとめるという学習活動も行い、その結果をファイルで回収している。口頭での回答と文章での回答とで時間的な制約などが異なり使用できる表現にも相違があり、比較のための資料となることも期待される。

3. 研究の方法

本研究では、上記のような発話データならびに関連する試験スコア・学習履歴アンケートならびに作文のデータを収集することを中心として研究活動を進めた。

- (1) 対象とする学生の英語学習経験についてアンケート調査を実施する。
- (2) テストスコア・アンケート情報・各種ファイル・音声データなどについて、外部組織・研究者との情報共有に協力の意向のある学生に同意書への記入を求める。
- (3) 英語リスニング・スピーキング自動試験 Versant for English の試験用紙を配布し、学生に受験を促す。
- (4) アンケート・試験用紙・同意書など紙媒体の資料をデジタル化して集積する。
- (5) 学生の学習活動の音声記録を音声収録するとともに活動状況をビデオに録画する。学生が作成したプレゼンテーション資料や英語による文書をファイルとして収集し、対応関係を記録する。
- (6) 収録した音声データの一部について、書き起こしを行う。
- (7) 国内外の関係組織を訪問し、データの分析と解釈に協力を求め、意見交換を行う。
- (8) 関連する国内・海外の学会・研究会等で研究の経過・成果について報告する。

4. 研究成果

本研究ではおおよそ当初の研究計画に沿って研究代表者が担当する早稲田大学1年次3クラスならびに2年次2クラスの英語授業において受講生の応答練習ならびに口頭発表の音声を収録するとともに、受講生の学習履歴ならびに英語リスニング・スピーキング自動試験 Versant for English のスコアを収集しデータベース化してきた。その概要ならびに特徴はおおむね以下の通りである。

- (1) これまで成果が公開されている研究での日本人英語学習者コーパスはどちらかというと書きことば(作文)コーパスが主体であり、話しことば(発話)コーパスはまだ少ない。
- (2) これまで成果が公開されている研究では、典型的な日本人大学生として海外在住経験のない学習者を想定する傾向が見られたが、近年大学の一般的な英語クラスにおいて受講生の1割から2割ぐらいがなんらかの海外在住経験を持つことを反映して、本研究ではこうした受講生のデータも積極的に取り入れている。また、少数ではあるが、日本以外からの留学生のデータも含まれている。

- (3) これまで成果が公開されている研究での日本人英語学習者の発話データはどちらかというと実験室的な環境での文の読み上げか単語の発音に関する音声データの収集とその分析が多く、本研究で進めてきたような教室での比較的自由に(質問に対する応答)長い発話(ひとつの発話が45秒から1分程度)はほとんどない。
- (4) NICT JLE コーパスでは発話の書き起こしが提供されているが、発話音声そのものにアクセスすることはできない。本研究ではCMSとデータベースを組み合わせ、webのブラウザ経由でいろいろな属性を組み合わせることで特定の発話(質問に対する応答)音声にアクセスすることが出来る。
- (5) NICT JLE コーパスでは発話に対して9段階の評定が与えられているが、本研究では発話者はVersant for Englishを学年の初め・半ば・終わりに受験し、総合点に加えて語彙・構文・発音・流暢さの4項目についてそれぞれ20点から80点までの60段階のスコアを与えられている。
- (6) 上記のスコアに加え、発話者それぞれの英語学習経歴について(たとえば英語圏での生活歴などもふくめて)アンケート調査を行った結果をデータベースに取り込んでいる。また、発話者が収集したデータについてどの範囲までの利用を認めているかもデータ化してある。
- (7) 本研究計画では、被験者から1回限りのデータ収集を行うのではなく、1学年を通じて(学期中)ほぼ毎週発話データを収集している。このため、限られた期間ではあるが、発音・流暢さ・使用する語彙・表現の豊かさについての縦断的研究を行う余地のあるデータとなっている。
- (8) 応答音声の書き起こしについては時間のかかる作業であるためすべての応答について実施することは当初の研究計画でも想定していなかったが、上記のwebインタフェースの開発が予想より早い段階で進んだため、2007年度後半から2008年度1年間をかけて2007年度の1年生1クラスについて、英語による応答のすべてについて一人の書き起こし作業員による書き起こしを付与することが出来た。複数の作業員による書き起こし間の一致度など、今後の課題も多いが、これによって作業時間の見通しを得ることが出来た。

こうした点から、本研究の成果として集積した発話データならびにこれに付随する学習到達度・学習経歴に関するデータは国内外を問わず、従来の研究では類を見ない貴重な学習者言語資源となっていることが明らかである。

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 16 件)

(1) 原田康也・前坊香菜子・河村まゆみ・前野謙二・楠元範明・鈴木陽一郎・鈴木正紀, 「大学生の英語口頭表現活動の音声ドキュメント化に向けて」, 第3回音声ドキュメント処理ワークショップ講演資料集, pp. 97-102, 豊橋技術科学大学メディア科学リサーチセンター主催, 情報処理学会音声言語情報処理研究会共催, 2009年2月27日, 査読無し.

(2) 原田康也, 「自律的学習を促す学習者主体の英語学習環境をめざして」, 人文論集, No. 47, pp. 61-84, 早稲田大学法学会, 2009年2月20日, 査読無し.

(3) 首藤佐智子・原田康也, 「統合的言語活動を促進するシラバスデザイン: 早稲田大学法学部の英語新カリキュラム」, 人文論集, No. 47, pp. 1-11, 早稲田大学法学会, 2009年2月20日, 査読無し.

(4) 原田康也・前坊香菜子・河村まゆみ・鈴木陽一郎・鈴木正紀, 「授業のデジタル化: 教員の暗黙知の共有化に向けてコンピュータでできること」, 平成20年度情報教育研究集会講演論文集, pp. 177-180, 九州工業大学主催, 国立大学情報教育センター協議会共催, 文部科学省・独立行政法人メディア教育開発センター・北九州市・福岡教育委員会・北九州市教育委員会後援, 2008年12月13日, 査読無し.

(5) 原田康也・楠元範明・辰己丈夫・前野謙二, 「英語授業におけるコンピュータ・リテラシの付随的獲得」, 平成20年度情報教育研究集会講演論文集, pp. 129-132, 九州工業大学主催, 国立大学情報教育センター協議会共催, 文部科学省・独立行政法人メディア教育開発センター・北九州市・福岡教育委員会・北九州市教育委員会後援, 2008年12月13日, 査読無し.

(6) 原田康也・前坊香菜子, 「学びあいをデザインする: 自律的相互学習のための英語授業のデザイン」, 情報コミュニケーション学会第4回研究会資料集, pp. 10-13, 情報コミュニケーション学会主催, 明治大学情報基盤本部後援, 2008年11月8日, 査読無し.

(7) 原田康也, 「コラム: 基礎体力増強のための『サーキット・トレーニング』」, 英語

教育 2008年10月号, 第57巻, 第7号, pp. 22-23, 株式会社大修館書店, 2008年10月1日, 査読無し.

(8) 原田康也, 「第二言語ライティングの産出と評価: 自律的相互学習を促す学習環境構築に関わる言語倫理的省察の試み」, 学習者コーパスに基づく英語ライティング能力の評価法に関する研究: 平成17年度~平成19年度科学研究費補助金(基盤研究(C))研究成果報告書, pp. 71-89, 2008年6月, 査読無し.

(9) Yasunari Harada, Kanako Maebo, Mayumi Kawamura, Masanori Suzuki, Yoichiro Suzuki, Noriaki Kusumoto, and Joji Maeno, "Toward Construction of a Corpus of English Learners Utterances Annotated with Speaker Proficiency Profiles: Data Collection and Sample Annotation," in T. Tokunaga and A. Ortega (Eds.): LKP 2008, Lecture Notes in Artificial Intelligence (LNAI) 4938, pp. 171-178, Springer-Verlag Berlin Heidelberg, 2008年3月3日, 査読有り.

(10) 河村まゆみ・原田康也・前坊香菜子・楠元範明・前野謙二, 「VALIS: 発話データの制限的共有と分散処理に向けて」, 情報処理学会研究報告 IPSJ SIG Technical Reports 2008-CE-93 (22), 学術刊行物情処研報 Vol. 2008, No. 13, pp. 155-162, 社団法人情報処理学会, 2008年2月16日, 査読無し.

(11) 原田康也・前坊香菜子・河村まゆみ・鈴木正紀, 「VALIS: 英語学習者のプロフィールと発話データの収集」, pp. 25-30, 信学技報 Technical Report of IEICE TL2007-11, (2007-12), 社団法人電子情報通信学会, 2007年11月16日, 査読無し.

(12) 原田康也・前坊香菜子・河村まゆみ・前野謙二・楠元範明・鈴木陽一郎・鈴木正紀, 「学習者主体の英語学習環境の構築と学習者プロフィール・発話データの収集」, 平成19年度情報教育研究集会論文集, pp. 486-489, 大阪大学主催, 文部科学省・国立大学情報教育センター協議会・独立行政法人メディア教育開発センター後援, 2007年11月9日, 査読無し.

(13) 原田康也・前坊香菜子・河村まゆみ, 「VALIS: 英語学習者発話データの書き起こし」, 情報処理学会研究報告 IPSJ SIG Technical Reports 2007-CE-90 (1), 学術刊行物情処研報 Vol. 2007, No. 69, pp. 1-8, 社団法人情報処理学会, 2007年7月7日, 査

読無し。

(14) 原田康也・前坊香菜子・河村まゆみ・前野譲二・楠元範明・鈴木陽一郎・鈴木正紀, 「VALIS: 学習者プロフィールに基づく学習者音声コーパス構築を目指して」, 情報処理学会研究報告 IPSJ SIG Technical Reports 2007-CE-88 (24), 学術刊行物 情処研報 Vol. 2007, No. 12, pp. 169-176, 社団法人 情報処理学会, 2007年2月16日, 査読無し。

(15) 原田康也, 「自律的学習を促す学習者主体の英語学習環境の構築に向けて」, 大学英語教育学会第45回全国大会要綱, pp.135-136, 大学英語教育学会, 2006年9月8日, 査読有り。

(16) 原田康也・楠元範明・前野譲二・鈴木正紀・鈴木陽一郎, 「大学英語授業でのグループ活動による自律的相互学習の効果検証を目指して」, 平成18年度大学教育・情報戦略大会抄録, pp. 244-245, 社団法人私立大学情報教育協会主催, 2006年9月6日, 査読無し。

〔学会発表〕(計9件)

(1) Yasunari Harada, "Human Use of Human Beings in Learner Utterance Data Collection: or why I may not be employing the optimally efficient data collection methods", a panel presentation in "English Learner Corpus: Global Perspectives with an Asian Focus," 2009 LTTC International Conference on English Language Teaching and Testing, The Language Training and Testing Center (LTTC), Taipei, 2009年3月6日。

(2) Yasunari Harada, "Transcriptions, Annotation Tools and Other Issues: an Interim Report on Compiling Students Impromptu Oral Responses to Questions," The 6th Korea-Japan Workshop on Linguistics and Language Processing, Kyung Hee University, Seoul, 2007年12月1日。

(3) 原田康也, 『言語学習・言語教育と言語処理・言語資源: 英語の「科学的学習法」を求めて』, 定期学術講演会, 東北学院大学英語英文学研究所主催, 2007年11月30日。

(4) 原田康也, 『ICTを活用した英語学習: 教室の現場から: 早稲田大学法学部の新しい英語教育』, グローバル・コミュニケーション・フォーラム, 『日本SGI ソリューション・

フォーラム 2008』, ウェスティンホテル東京, 2007年6月15日。

(5) Yasunari Harada, "Decentralization of Communication Channels in Class," 基調講演, JALTCALL 2007, 早稲田大学, 2007年6月2日。

(6) 原田康也, 『大学英語学習者のプロフィールと発話データの収集』, CIEC第67回研究会【e-learningシステムの開発と教育実践】ポスターセッション1【研究発表3】, 大学生協杉並会館2階201-203会議室, 2007年3月31日。

(7) 原田康也, 『英語とコンピュータ・リテラシの同時学習』, 招待講演, ティームティーチングによる二言語同時学習: 外国語教育の新たな教授形態, 京都外国語大学, 2007年2月25日。

(8) Yasunari Harada, "Do Japanese learners of English learn for themselves, by themselves and/or among themselves?: An Interim Report on the Collection of Learner Profiles and Utterance Data," 早稲田大学総合研究機構情報教育研究所主催, 早稲田大学西早稲田キャンパス8号館4階412教室, 2007年2月2日。

(9) Yasunari Harada, "Language E-Learning: How Language Technology and Language Resources can contribute to a Better Language Learning for (Asian) Students," The 5th Korea-Japan Workshop on Linguistics and Language Processing, Kyung Hee University, Seoul, 2006年12月9日。

6. 研究組織

(1) 研究代表者

原田 康也 (HARADA YASUNARI)
早稲田大学・法学大学院・教授
研究者番号: 80189711

(2) 研究分担者

以下の研究者は2006年度ならびに2007年度に研究分担者として研究活動に参加した。

中野 美知子 (NAKANO MICHIKO)
早稲田大学・教育・総合科学大学院・教授
研究者番号: 70148229

井佐原 均 (ISAHARA HITOSHI)
(独)情報通信研究機構・けいはんな研究所・
上席研究員
研究者番号：20358881

楠元 範明 (KUSUMOTO NORIAKI)
早稲田大学・教育・総合科学学術院・教授
研究者番号：60277861

北原 真冬 (KITAHARA MAFUYU)
早稲田大学・法学学術院・准教授
研究者番号：00343301

G・R Stockwell
早稲田大学・法学学術院・教授
研究者番号：90367988

以下の研究者は 2007 年度に連携研究者として
研究活動に参加した。

星井 牧子 (HOSHII MAKIKO)
早稲田大学・法学学術院・准教授
研究者番号：90339656

(3) 連携研究者

以下の研究者は 2008 年度に連携研究者として
研究活動に参加した。

中野 美知子 (NAKANO MICHIKO)
早稲田大学・教育・総合科学学術院・教授
研究者番号：70148229

井佐原 均 (ISAHARA HITOSHI)
(独)情報通信研究機構・けいはんな研究所・
上席研究員
研究者番号：20358881

楠元 範明 (KUSUMOTO NORIAKI)
早稲田大学・教育・総合科学学術院・教授
研究者番号：60277861

北原 真冬 (KITAHARA MAFUYU)
早稲田大学・法学学術院・准教授
研究者番号：00343301

G・R Stockwell
早稲田大学・法学学術院・教授
研究者番号：90367988

星井 牧子 (HOSHII MAKIKO)
早稲田大学・法学学術院・准教授
研究者番号：90339656

(4) 研究協力者

以下のものは 2006 年度より 2008 年度にかけ
て研究協力者として研究活動に参加した。

河村まゆみ (KAWAMURA MAYUMI)

前坊香菜子 (MAEBO KANAKO)
早稲田総研インターナショナル契約講師(日
本語教育研究センター設置授業担当)