

平成 21 年 6 月 26 日現在

研究種目：若手研究(B)
 研究期間：2007～2008
 課題番号：19720145
 研究課題名（和文） 英語学習者の音読における韻律の研究：項目応答理論による評価の分析と音響学的測定
 研究課題名（英文） A study of prosody of second language learner of English in Reading aloud: Educational measurement by Item Response Theory and acoustic analysis
 研究代表者
 近藤 悠介（KONDO YUSUKE）
 立命館大学・言語教育センター・講師
 研究者番号：80409739

研究成果の概要：

本研究の最終目標は英語学習者の発話自動評価システムを構築することである。このシステムは、熟練した語学教師の主観的評価をコンピュータで計算できる客観的な学習者の発話の特徴から再現することである。主観的評価値を予測する学習者の発話の特徴を検討するためにアジア人英語学習者から発話データを収集し、そのデータから発話の特徴を抽出し、主観的評価値との関係を重回帰分析によって明らかにした。本研究での客観的測定値と主観的評価値との重相関係数は.69であった。また発話データに音素アライメント、学習者に関する情報、主観的評価値を加え、Websiteで公開した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,300,000	0	1,300,000
2008年度	500,000	150,000	650,000
総計	1,800,000	150,000	1,950,000

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：学習者データベース、発話評価

1. 研究開始当初の背景

英語学習者の音読における発音および韻律に関する評価および測定は、英語教育、音響学、教育評価・測定の分野で行われている。それぞれの分野の研究で得られた知見、実験計画および分析の手法を総合的に検討し、取り入れることによって、各分野での研究を補完することになり、測定および評価の妥当性、信頼性を向上させることが可能である。本研究では、英語学習者の音読における発音およ

び韻律に関して、主観的評価の妥当性を検証するとともに、主観的評価と客観的測定値の関係を探り、英語学習者の音読における発音および韻律に関して妥当な客観的指標を提示する。本研究は、妥当性および信頼性が検討された客観的測定値および評価値が付与された、英語学習者の音声データベースを作成するための基礎研究である。

近藤、前木、白井、匂坂(2003)、前木、内藤、近藤、白井、匂坂(2003)、Muto, Sagisaka,

Naito, Maeki, Kondo, & Shirai (2003)において、匂坂らは、日本人英語学習者に英語の単文を音読させ、韻律という観点から評定者に学習者の音声を評価させた。学習者が音読した文に関しては、発話の時間的長さを測定した。客観的測定値として母音および子音の長さ、文の長さ、強勢母音の長さ、ポーズの長さなどを測定し、母語話者のモデル音声との差異を算出した。これら客観的測定値と評定者による主観評価がどの程度関連しているか、また、客観的測定値はどの程度正確に主観を予測するかということを検証するために、主観評価値を従属変数、客観的測定値を独立変数として重回帰分析を行った。結果は、客観的測定値はかなり正確に主観的評価値を予測することが分かった（決定係数0.71）。この結果は、匂坂らの研究において、用いられた学習者の音声に関する客観的指標の妥当性が高いことを示している。つまり、評定者による主観的評価は、客観的測定値を指標として表せるということである。さらに、用いた客観的測定値は、主観的評価と相関が高く、妥当性の高い測定値と言える。匂坂らの研究は、韻律の自動評価システムの開発を目的としたため、韻律のみに焦点を当てたものである。従来、英語教育の研究において、客観的な指標を用いて韻律という観点から学習者の発話を測定することは、困難であったが、匂坂らが用いた客観的測定値を用いることによって、学習者の発話の客観的指標を得ることができた。

学習者の発話を評定者が評価する場合、分析の対象が、学習者、評定者、評価項目と3相になり、従来の信頼性の指標の利用、因子分析を用いて評価項目の精選する場合は、評定者内で中央値、平均値など算出し、評定者の相を圧縮し、2層で分析することが多かった。また、段階的な評価を項目応答理論により分析する際には、正解あるいは不正解といった2値に情報を圧縮して分析を行うことが多かった。このような問題を解決するために、Linacre (1989)によって相という概念が導入され、項目の困難度、評定者の厳しさを考慮して、受験者の能力を推定する多値型の項目応答理論で分析可能な FACETS が開発され、学習者に対する評価を分析することが可能となった。また、評定者の人数、評価項目の数に関しても、適切さが求められるが、この問題に関しては、一般化可能性理論を用いて、評定者の人数、評価項目の数に関する信頼性の指標を算出することが可能となった。

一般に評価項目を設けて評定者が評価する場合は、5段階や7段階の段階評価を行うことが多いが（7が「良い」、1が「悪い」の7段階評価など）、学習者の発話を測定する際には、評定者個人が、学習者の発話に関してモデルを持っていなければ、段階の境界

や段階が示す対象が曖昧になる場合がある。例えば、5段階で文のリズムに関して評定者が学習者の発話を評価する場合、評定者は、母語話者のリズムと同じリズムで発話されたものに最高の評価を与えるのか、母語話者のリズムとは若干異なるが、かなり近い発話には最高の評価を与えるべきか、それとも2段階目の評価を与えるかは、評定者個人に委ねられている場合が多い。このような問題を解決するためには、多段階で評価する際に、それぞれの段階のカテゴリを評定者に明確に示し、訓練する必要がある。Council of Europe (2001)が提案する Common European Framework of References (CEF)は広範囲にわたって、外国語教育においてシラバスや教科書の作成に用いる学習者がどのような言語に関する知識や運用能力を持つべきかを記述し、参照の基準となっている。CEFは、英語母語話者を最終的な目標と考えるのではなく、非母語話者同士のコミュニケーションを念頭において学習者を評価することも可能であるため、CEFを評価の参照基準とすることは、「世界の英語たち」という考えに合致し、現在の英語教育の参照基準として妥当である。CEFによれば、学習者は6段階に分けられる。このレベル分けされた学習者の発話は、映像として Language Policy Division, Eurocentres, (2003)や North & Hughes (2003)が提供している。この学習者の発話場面を映した映像を評定者に視聴させ、6段階それぞれのカテゴリに入る学習者の特徴を捉えさせる訓練をした、本応募者が共同研究で行った評価に関する実験では、信頼性の高い評価が得られている。また、全体的な熟達度を下位分類した、発音や韻律などの評価においても信頼性および妥当性の高い結果が得られた（中野他 2006）。このようにCEFを参照して学習者の発話を段階に分けることで、多段階での評価における段階間の曖昧さはある程度解消された。

2. 研究の目的

本研究では、英語学習者の音読における発音・韻律に関して、主観的評価の妥当性を検証するとともに、主観的評価と客観的測定値の関係を探り、英語学習者の音読における発音・韻律に関して妥当な客観的指標を提示する。本研究は、妥当性および信頼性が検討された客観的測定値および評価値が付与された、英語学習者の音声データベースを作成するための基礎研究である。

本研究では、音響学、教育評価・測定、英語教育の知見を取り入れ、主観的な評価と客観的に測定できる特徴量の両面から英語学習者の音読における発音・韻律を捉え、それらの値を付与した英語学習者の音声データベースを作成するための基礎的な研究をおこ

なった。音声データは、定量的分析を行うため、安定したデータが収集できる文章の音読を対象とした。2年間の研究期間において、これらの学習者から発話データを収集し、CEFを参照して評定者の訓練を行い、評定者の主観的評価と関係の高い客観的測定値を抽出する。小規模な英語学習者の音声データベースを作成し、公開した。

異なる母語を背景として持つ英語学習者の客観的指標には一定のものがなく、本研究により提示される客観的指標が、英語学習者の発音・韻律を測定する際の客観的指標として利用されることが望まれる。

3. 研究の方法

評定は、音読データが日本語を母語とする学習者が主であるため、日本語を母語とする英語の英語教育の研究歴と教育歴を持つ者が行った。また、中野、近藤、上田、筒井、大和田(2006)のデータを用いて、一般化可能性理論を用いて、評定者の人数、項目の数を検討する。項目の精選は、生データを検証するとともに、項目応答理論、一般化可能性理論に基づいて行った。

これらの主観評価値との相関関係に基づいて、Ano, 2001; Foster and Skehan, 1996; Iwashita, McNamara, and Elder, 2001; Yuan and Ellis, 2003; Munro and Thomson, 2004; Trofimovich and Baker, 2007; Riggenbach, 1991; Towell, Hawkins, and Bazergui, 1996の外国語教育の分野で使用されている学習者の発話の指標となる客観的測定値を重回帰分析を用いて検証した。

4. 研究成果

本研究では、評定者の訓練の効果を検証した。評定者の訓練を行う前と後で評定者はアジア人英語学習者の自己紹介のスピーチを評価し、その評価の信頼性を一般化可能性理論に基づいて検証した。評定者の訓練はヨーロッパ言語共通参照枠(CEFR)を参照し、評定者がCEFRが設定する6レベルの学習者の特徴に関して議論を行い、評価の統一性を図った。評価項目は八代、荒木、樋口、山本、コミサロフ(2001)を参考に評定者間の議論により、精選し、評定者の利便性を図り評価用のWebsiteを作成した。一般化可能性理論に基づいた分析はGENOVA(Crick and Brennan, 1984)を用いて行った。結果は評定者の訓練の前と後で評価項目の分散推定値が約6分の1ほど小さくなることが分かった。この結果から、評定者の訓練によって評定者間で評価項目の理解が統一されたと解釈することができ、評定者の訓練に意義があることが示唆される。

また、評価における一貫性が低い評定者、信頼性の低い評価項目を精選するために

Multifaceted Rasch Analysisに基づいて評価項目、評定者の分析を行った。分析にはFACETS(Linacre, 2006)が用いられた。本研究の評定者には、McNamara, 1996が提案する基準を逸脱する評定者はおらず、評価項目に関しては、基準を超える2項目が削除された。本研究が扱う音読データに関しても同様の手順で評定者および評価項目の精選が行われた。

以上の手順で信頼性が検証された評定値と音読に現れる学習者のスピーチの特徴の関係を検証する相関研究が行われた。読む速さ、ポーズの長さ、ポーズの数、ポーズの位置、母音の質、リズムに関する指標、読み間違い、平均発話長(音節)などが評定値の予測変数として検討されたが、読む速さに関する指標、およびリズムに関する指標が統計的に有意な予測変数であることが重回帰分析により検証された($F_{(2, 98)} = 44.57, p < .01, \text{adjusted } R^2 = .47$)。本研究で行った主観評価値と客観的測定値の相関研究では、重相関係数が.69であった。

上記の分析を踏まえ、アジア人英語学習者の音声データベースを作成し、その音声データに信頼性が検証された評価値、音素アライメントの情報を付与し、公開することである。作成された音声データベースの情報は以下に示すものである。

- 読み上げ音声(wav形式 モノラル 16kHz 16bit)
- 読み上げ文のスクリプト
- 音素アライメントの情報
- 読み上げ文に対する評価
- 学習者の情報(年齢、母語、英語の学習期間など)

これらの情報を研究成果としてWebsiteで公開した。

5. 主な発表論文等

[雑誌論文](計2件)

Kondo, Y., Tsutsui, E., Nakano, M. (2008). Fundamental research of automatic speech evaluation of L2 English. *Proceedings of the 13th Conference of Pan-Pacific Association of Applied Linguistics*. 85-86. 査読有

Nakano, M., Kondo, Y., & Tsutsui, E. (2008). Fundamental Research on Automatic Speech Evaluation. *9th APRU Distance Learning and the Internet Conference--New Directions for Inter-institutional Collaboration: Assessment & Evaluation in Cyber Learning*. 207-212. 査読無

〔その他〕

ホームページ等

<http://www.yusukekondo.com/data/html>

6．研究組織

(1)研究代表者

近藤 悠介 (KONDO YUSUKE)

立命館大学・言語教育センター・講師

研究者番号：80409739