

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 9 日現在

機関番号：32641

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24520631

研究課題名(和文) 言語処理技術と多変量解析を用いた中間言語の全体像の解明 新しい研究手法の確立

研究課題名(英文) Using multivariate statistical techniques to analyze development of speaking and writing proficiency level

研究代表者

阿部 真理子 (Abe, Mariko)

中央大学・理工学部・准教授

研究者番号：90381425

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：日本人英語学習者のスピーキング能力とライティング能力の発達に関する研究である。7つの異なる習熟度グループに属する1200名以上の日本人英語学習者(初級・中級・上級)の話し言葉から構成される大規模コーパス(NICT JLE)を分析対象とし、58項目にわたる多種多様な言語項目(語彙・品詞・統語構造・談話構造など)の使用頻度の情報をもとに、多変量解析の手法を用いて、発達の指標となる言語項目の特定を試みた。結果、日本人英語学習者の話し言葉と書き言葉に関するさまざまな特性を明らかにすることができた。

研究成果の概要(英文)：This project investigated the overall patterns of variation across seven proficiency levels of 1,263 Japanese EFL learners and native English speakers. The methodological approach combined a learner corpus, language processing techniques, and multivariate statistical analyses to identify patterns of language use. The largest spoken learner corpus in Japan, NICT JLE Corpus, was used for the analysis. The study found interesting rising, flat, and falling frequency patterns in how several linguistic features are used in different oral proficiency levels. Some linguistic features (e.g., phrasal coordination and nouns) were frequently used by the learners at a low level. The frequencies of some others (e.g., contraction, pronoun it, and emphatics) increased as oral proficiency increased. The study identified a set of linguistic features that differentiate among second language proficiency groups as well as between non-native and native speakers of English.

研究分野：コーパス言語学

キーワード：Learner Corpus

### 1. 研究開始当初の背景

(1) 英語学習者が書いたり、話したりした産出データを大量に収集した学習者コーパスを利用して、習得のプロセスを検証する研究は、近年注目を浴びてきており、computational SLA という新たな枠組みを与えられている。海外では、異なる母国語をもつ英語学習者のデータから構築された International Corpus of Learner English (ICLE) を用いて、中間言語の多様性、およびその要因としての母語の影響をさぐる研究が盛んである。

(2) しかし、中間言語の発達をたどる研究の成果は少ない (Myles, 2005; Pendar & Chapelle, 2008)。さらには、発達の過程を多面的にたどる研究はほとんど行われていない。そのため、中間言語の全体像が見えてこないという問題がある。従来の中間言語の研究は、特定の言語項目— 語彙、あるいは文法構造のいずれかに焦点が置かれており、対象となる言語項目の数が圧倒的に少ないのである (Biber, Conrad, & Reppen, 1998)。

(3) そこでこれまであらかじめ学習者の習熟度情報がついている The National Institute of Information and Communications Technology Japanese Learner English (NICT JLE) コーパスに、31 項目のエラー情報を付与し、中間言語の発達をエラーの観点から、多面的に明らかにする研究を行ってきており、一定の成果をあげた (Abe, 2007a, 2007b)。しかし Abe (2007a) において、エラーではなく、学習者の正しい言語使用に着目した方が、より中間言語の発達の特徴を明らかにできる場合があるという知見も得た。

(4) これを受けて、予備研究 (阿部, 2010) では、Biber (1988) がネイティブの英語を分析するために開発した研究手法 — 『言語処理技術』と『多変量解析』を用いたコーパスの分析— を中間言語に応用し、学習者の正しい言語使用を 57 項目から分析した。その結果、ネイティブの話し言葉と書き言葉の違いを多次的に明らかにした手法が、学習者言語にも利用できることがわかった。また、(a) 代名詞、(b) 否定表現、(c) 等位構造、(d) 置き換え、(e) 会話の話題、(f) 語用論的な洗練度、などの項目が習熟レベルの違いを判別できることが明らかになった。

(5) 本研究は、これまでの研究成果を踏まえ、新しい研究手法を用いて、日本人英語学習者の習熟レベルを一定の精度で判別できる項目を特定する。そしてそれらの項目を通して、中間言語の発達を包括的に記述することで、全体像を解明することを目的とする。よって、第二言語習得の分野において、新しい研究手法による、新しい知見の提供を目指す先駆的な研究として、位置づけられる。

### 2. 研究の目的

(1) 本研究の目的は、中間言語の発達の全体像を解明する研究が不足しているという問題を解決し、さらに語学教育の応用研究へとその成果が展開できる新しい手法を確立することである。そのために、まず習熟レベルの異なる学習者がさまざまな言語項目 (語彙・統語構造・談話構造・品詞など) をどれくらいの頻度で使用しているかという情報を学習者コーパスから言語処理技術によって抽出し、多変量解析の処理を行い、一定の精度で習熟レベルを判別できる項目を特定する。そしてそれらの項目を通して、中間言語の発達を包括的に記述し、全体像を明らかにする。

### 3. 研究の方法

(1) まず予備研究で扱った 57 項目以外から、分析対象とする項目を選定し、話し言葉の習熟レベルを判別できる項目を特定する。さらには以下のような言語項目を追加することもできないかということについて検討を行う。

- コロケーション (例: make / take / get / have + 目的語)
- バイグラム (例: of the, you know) 品詞連鎖 (例: 前置詞 + 名詞、副詞 + 動詞)
- 日本人にとって習得が困難な項目 (例: 冠詞、時制と相)
- 検定教科書において扱われる項目 (例: 間接疑問文、直接疑問文、仮定法)

(2) 次に以下の分析手順 (A ~ C) により、1200 人の話し言葉コーパス (NICT JLE Corpus) から、習熟レベルを判別できる項目を特定する。

A) NICT JLE コーパスは、ACTFL OPI に準拠したスピーキング・テストを受験した学習者の発話を書き起こしたデータを使用している。そのため、専門の評価官が、一定の基準により判定した 9 段階の習熟レベルが全データについている。そこで、プログラミング言語 Perl を用いて、各言語項目の使用頻度を習熟レベルごとに計算させるプログラムを作成する。

B) (習熟レベル) × (言語項目) のクロス表を完成させる。

C) 多変量解析 (対応分析とクラスター分析) を用いて、変数間の類似・差異関係を調べ、一定の精度で、習熟レベルを判別できる項目を特定する。

(3) 次に学習者と同じタスクを行ったネイティブのデータを利用することで、両者を判別する項目を特定すると同時にその類似点・差異点を明らかにする。

(4) 最後に書き言葉の習熟レベルを判別できる項目を特定したのち、話し言葉と書き言

葉を比較する。

(5) 総括を行い、研究手法を確立させ、今後の課題をさぐる。

#### 4. 研究成果

(1) 日本人英語学習者のスピーキング能力とライティング能力の発達に関する研究である。2012度は、7つの異なる習熟度グループに属する1200名以上の日本人英語学習者(初級・中級・上級)の話し言葉から構成される大規模コーパス(NICT JLE)を分析対象とし、58項目にわたる多種多様な言語項目(語彙・品詞・統語構造・談話構造など)の使用頻度の情報をもとに、多変量解析の手法を用いて、発達の指標となる言語項目の特定を試みた。

(2) その結果、日本人英語学習者の話し言葉と書き言葉に関するさまざまな特性を明らかにすることができた。初級レベルの英語学習者は、(a) WH questions, (b) synthetic negation, (c) phrasal coordination を多用する。上級レベルの学習者、および英語母語話者は、(a) analytic negation, (b) hedges, (c) emphatics, (d) demonstrative pronouns, (e) indefinite pronouns を特徴的に用いる。さらに英語母語話者は、(a) pronoun it, (b) pro-verb do を学習者より多く使用している。

(3) さらに話し言葉だけではなく書き言葉に研究対象としているため、東アジアにおける英語学習者の作文コーパス(アジア圏英語学習者コーパス、International Corpus Network of Asian Learners of English: ICNALE) (参考資料)を分析した。まずはICNALEから、香港、台湾、韓国、および日本の英語学習者(大学生)が産出した英文エッセイのデータを抽出し、それぞれのグループにどのような言語的使用の特徴があるかを探った。

(4) 次に、学習者の母語と英語習熟度レベルのちがいによって、言語使用にどのような類似点や相違点が見られるのかという課題についても、研究を進めた。そして、英語学習者と同じ課題で、英文エッセイを英語母語話者に作成してもらったデータを用いて、英語母語話者による言語使用の特徴と学習者との言語使用の差異を分析した。

(5) その結果、言語使用の特徴によって、3つのグループに分かれた。日本人英語学習者の多くは、東アジアにおける他の英語学習者とは一線を画し、1つのグループを構成することが明らかになった。そして日本人英語学習者の言語的特徴も明らかになった。具体的には、日本人英語学習者が偏用しがちな(過剰使用する)項目は「1人称代名詞」、「現在形」といった話し言葉に特徴的なものであることも明らかになった。またあまり使用し

ない(過少使用する)項目は「限定形容詞」であった。

(6) そしてさらには、英文エッセイのトピックが使用言語にどのような影響を与えるのかを分析を行った。結果、英語学習者が産出する英文エッセイは、トピックの影響を受けることが明らかになった。

#### <引用文献>

Abe, M. (2007a). A corpus-based investigation of errors across proficiency levels in L2 spoken production. *JACET Journal*, 44, 1-14.

Abe, M. (2007b). Grammatical errors across proficiency levels in L2 spoken and written English. *The Economic Journal of Takasaki City University of Economics*, 49, 117-129.

Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4), 373-391.

Pendar, N., & Chapelle, C. (2008). Investigating the promise of learner corpora: Methodological issues. *CALICO Journal*, 25(2), 189-206.

阿部真理子 (2010)「コーパスを用いた日本人英語学習者の習熟度の分析」第36回英語コーパス学会 全国大会 研究発表 於東京大学

#### 5. 主な発表論文等

[雑誌論文](計 5件)

Abe, M. (2014). Frequency change patterns across proficiency levels in Japanese EFL learner speech. *Journal of Applied Language Studies*, Special issue on "Learner language and learner corpora", 8(3), 85-96. (査読あり)

Kobayashi, Y. & Abe, M. (2014). A machine learning approach to the effects of writing task prompts. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world*. Vol. 2. (pp. 163-175). (査読あり)

Abe, M., Kobayashi, Y., & Narita, M. (2013). Using multivariate statistical techniques to analyze the writing of East Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world*. Vol. 1. (pp. 55-65). (査読あり)

小林雄一郎・阿部真理子・成田真澄 (2013)「書き手の習熟度と母語が第2言語ライティングに与える影響」『人文科学

とコンピュータシンポジウム論文集 人文科学とコンピュータの新たなパラダイム』東京: 情報処理学会 (pp. 89-96) (査読あり)

阿部真理子(2012). 「多変量解析を用いた日本人英語学習者の発達指標の特定」『人文科学とコンピュータシンポジウム論文集 つながるデジタルアーカイブ』(pp.55-60). (査読あり)

[学会発表](計 9件)

阿部真理子 (2015). 科学研究費補助金基盤(B)「第二言語ライティング研究の現代的課題と解決のための将来構想 - 東アジアからの発信 - 」中間報告、コーパス班の取り組みから 第 10 回千葉大学英語教育学会、千葉大学(2015年3月15日)

Kobayashi, Y. & Abe, M. (27th September, 2014). “The similarity and difference between human scoring and automated scoring.” Applied Linguistics Association of Korea (ALAK) 2014 (Sangmyung University, Korea)

Kobayashi, Y. & Abe, M. (22nd July, 2014). “How do L1 and proficiency level influence L2 written production?” Teaching and Language Corpora (TaLC) 11 (Lancaster University, UK)

Abe, M. & Kobayashi, Y. (19th March, 2014). “Profiling developmental patterns of learner language.” Digital Humanities Australasia 2014 (University of Western Australia, Australia)

Kobayashi, Y. & Abe, M. (7th March, 2014). “Automated scoring of L2 spoken English with random forests.” Asia Pacific Corpus Linguistics Conference (APCLC2014) (Hong Kong Polytechnic University, Hong Kong)

Abe, M., Kobayashi, Y., & Narita, M. (28th September, 2013). “Linguistic features discriminating between native English speakers and East Asian learner groups with different proficiency levels.” Learner Corpus Research 2013 (Bergen, Norway)

Abe, M. (16th March, 2013). “Using multivariate statistical techniques to analyze development of oral proficiency.” American Association for Applied Linguistics 2013 (Dallas, USA)

Abe, M. (October, 2012). “Learner

speech variation across proficiency levels of Japanese EFL learners Learner” Language, Learner Corpora 2012 (University of Oulu, Finland)

Abe, M. (July, 2012). “Variations across Proficiency Levels in L2 Spoken English” Teaching and Language Corpora 10 (University of Warsaw, Poland)

## 6 . 研究組織

### ( 1 ) 研究代表者

阿部真理子 (Abe, Mariko)

中央大学・理工学部・准教授

研究者番号 : 9 0 3 8 1 4 2 5