

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 14 日現在

機関番号：12603

研究種目：基盤研究(A) (一般)

研究期間：2012～2015

課題番号：24242017

研究課題名(和文) 学習者コーパスによる英語CEFRレベル基準特性の特定と活用に関する総合的研究

研究課題名(英文) the CEFR-J RLD project: Developing Grammar, Text and Error Profiles Using Textbook & Learner Corpora

研究代表者

投野 由紀夫 (Tono, Yukio)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号：10211393

交付決定額(研究期間全体)：(直接経費) 34,700,000円

研究成果の概要(和文)：本科研の目的は日本人英語学習者の英語力に関する到達指標の提案およびそのレベル別言語材料の科学的な整備である。これを行うことで、シラバス開発、教科書・教材・タスク開発、テスト開発などに資する一貫した英語資料を提供できる。指標に関してはCEFR-JというCEFR準拠の英語汎用枠に基づいた。言語材料配当は、CEFR準拠教科書コーパス・学習者コーパスを独自に構築し、500以上の文法事項、テキスト特性、エラー特性に関してコーパス解析と機械学習の手法を用いて調査し、文法、テキスト、エラーの各プロファイルとして整備した。

研究成果の概要(英文)：The present study aims to propose a common framework of reference for English language teaching and learning in Japan as well as a set of language specifications (vocabulary, grammar and texts) for each level of the framework. For the common framework, the CEFR-J, an adapted version of the Common European Framework of Reference for Languages (CEFR), is used. Two types of corpora, a corpus of CEFR-based course books and CEFR-classified learner corpora were used to extract grammatical, textual, and error features from each CEFR(-J) level. The frequency data were then analyzed using a machine learning technique called Support Vector Machine to learn how they serve to contribute to an effective classification of CEFR levels. Important features were then made into a set of inventories as reference level descriptions, called the Grammar Profile, the Text Profile and the Error Profile, which will be made publicly available for developing syllabuses, textbooks, and language tests.

研究分野：外国語教育

キーワード：英語到達度指標 CEFR 学習者コーパス コーパス言語学 第二言語習得

1. 研究開始当初の背景

現在文部科学省は GOAL2020 と称して、体系的な英語教育改革を推進している。その改革の方向性はさまざまな視点を包含しているが、本科研との関連でその方向性を大別すると以下の3点に集約される：

小学校に英語を教科として導入し、開始時期を早めることで全体的な英語力の底上げを目指す。[小学校英語導入]
 高校卒業時の大学入試を改革し、今までの読解力中心のテスト方法から、4 技能をバランスよく見るテストの導入を推奨、波及効果で中高での英語教育の方法見直しを図る。[大学入試改革]
 小中高（および大）での一貫した到達指標を CAN-DO の形式で設け、小中、中高、高大の円滑な連携を目指し、さらにその指標によって国際的な英語力基準を導入する。[CAN-DO 形式の到達指標の設定]

これは「入口」、真ん中、「出口」に対応し、英語教育改革の包括的・抜本的な改革ということができる。

このような改革は実質的な変化を～のポイントで生み出せれば、相乗効果によって所期の目的の達成が大いに期待できる反面、～のどこかが間違えば効果が出ないだけでなく、英語教育の全体構造を変えているために従来よりも状況が複雑になり混乱を来す可能性もある。学習指導要領の改訂と同時に、～のそれぞれのポイントを確実に押さえて、実効力のある改革が施行される必要がある。本研究はそのために、日本の英語教育にヨーロッパ言語共通参照枠（CEFR）を導入するための枠組みとして発表された CEFR-J（投野 2012）を活用する。

CEFR-J を具体的に言語政策決定、教科書・教材・テスト作成、教員研修などに活用するために、CEFR-J レベル別の文法事項、与えべきテキスト特徴、学習者のレベル別エラー傾向などをコーパス言語学や自然言語処理の手法を用いて科学的に調査したものである。

2. 研究の目的

本科研の目的はこのような英語教育改革の最も根本的なポイントである「英語到達度はどのようにレベル設定を行い、日本人にとってふさわしい目標をどのように設定するか」という到達指標の提案と、その具体的な実施に必要なシラバス開発、教科書・教材・タスク開発、テスト開発などに資する言語材料の包括的な整備を目的としている。これらの整備を CEFR のレベルに基づき客観的に行う手法は参照レベル記述（Reference Level Description: 以下 RLD）と呼ばれる。これにより、小学校英語導入から中高大の英語力の発達を統一

的な枠組みで目標の記述を行い、それを大学入試の改革につなげられるような道筋をつけることが具体的に可能になり、それは新指導要領の記述を側面からサポートし、また新しい枠組みに沿った教科書作成やテスト開発などの基礎資料となる。

3. 研究の方法

本科研では、CEFR-J の各レベルに対応する以下のようなプロファイリング情報を提供

Grammar Profile	レベル別に導入すべき 文法事項に関する情報
Text Profile	レベル別に使用すべき 英文テキスト特性に関する情報
Error Profile	レベル別学習者の 典型的エラーに関する情報

することで RLD の資料整備を進める：

RLD に必要な言語コーパスの整備に関しては、大別して2つのことを行った：

- (1) CEFR 準拠の英国のコースブックのデータを CEFR レベル別に収集、コーパス化した CEFR-based Coursebook Corpus
- (2) 日本人英語学習者の産出データ(話し言葉・書き言葉)を CEFR レベル別に整理した
 → JEFLL Corpus (CEFR version);
 NICT JLE Corpus (CEFR version)

その結果、以下の3つのコーパスを主要なデータソースとして使用した。

CEFR Coursebook	CEFR 準拠の英国コースブックのデータを CEFR レベル別に収集
JEFLL Corpus	日本人英語学習者 1 万件の英作文コーパス (67 万語)
NICT JLE Corpus	日本人英語学習者 1200 件の会話コーパス (200 万語)

コーパス構築は 2012-13 年度の 2 年間で費やし、その間は各自で関心のある言語特徴に関して CEFR レベルの基準特性となりうるものは何かをいろいろ事例研究を持ち寄りながら検討を重ねた。2013 年度からは東京工業大学の奥村学氏が加わり、自然言語処理の知見を加えて本格的な RLD のための言語特徴の洗い出しの方法を議論した。

コーパスの構築を受けて、言語特徴の頻度調査に関して以下のような方法を用いた：

- (1) **CEFR-J Grammar Profile:**
 主要な文法項目の出現頻度を CEFR-based Coursebook Corpus を用いて CEFR レベル別にすべてチェックし、Grammar Profile の基礎データとする。Grammar Profile の目的は「CEFR-J に基づくシラバス構築を目的としたインプットとしての文法特性とその提示イメージを提案すること」とする。
- (2) **CEFR-J Text Profile:**
 主要なテキスト特徴を CEFR-based Coursebook Corpus を用いて CEFR レベル別にすべてチェックする → これを Text Profile の基礎データとする。Text Profile の目的は「CEFR-J に基づくシラバス構築を目的としたインプットとしてのテキスト特性(流暢性、複雑性、語彙レベル、など)とその提示イメージを提案すること」とする。
- (3) **CEFR-J Error Profile:**
 (1)の Grammar Profile に関して、日本人英語学習者がどの程度当該文法項目を使えるのかをアウトプット側からの情報として、(1)に補足情報として付与する。Error Profile の目的は「CEFR-J に基づく Grammar Profile を補完する形で CEFR-J レベルごとに特徴的な学習者のエラー特性を判定し、input 基準のプロファイル情報を補うこと」とする。

4. 研究成果

- (1) 各種 Profile の構築
 CEFR 準拠教科書 (CEFR-based Coursebook Corpus)、CEFR レベル別学習者の英作文 (JEFLL Corpus) 及び会話 (NICT JLE Corpus) コーパスからの頻度抽出情報をもとに CEFR 各レベルに対応する言語特徴をどのように規定するかに関して、Grammar Profile を東工大の奥村研究室、Text Profile を大阪大学の荒瀬研究室 (研究協力者)、学習者のスピーキング・データを九州大学の廣川研究室 (研究協力者) などの自然言語処理、工学系の専門チームに検討を依頼した。
 その結果、基本的な頻度と分布の統計量を出した後、レベル別の頻度データの機械学習をサポートベクターマシンで行い、その判別に寄与した言語特徴の重みベクトルのスコアをランキング形式で抽出した。これらの結果を Grammar Profile の頻度・分布データとからめることで何らかのレベル帰属する文法項目を割り当てることを目指した。
 なお、全体の文法頻度の抽出データは石井康毅が中心となってデータ処理を行い、全員で石井康毅 (2016)「文法項目頻度リスト (version 20160108)」を共通データとして用

いた。

- (2) CEFR-J RLD 手法の公開
 これらの RLD の一連の手続きに関して、2016年3月5-6日、東京外国語大学における最終成果報告シンポジウムにおいて研究発表を行った。CEFR の開発の中心人物で第一人者である Brian North 氏 (Euro Centres Foundation)、英語における RLD の中心的プロジェクトを行っている Paula Buttery 氏 (Cambridge 大学) がシンポジウムに参加し、研究チームの明確な目的と優れた連携、科学的な RLD 手法が高く評価された。その成果は最終報告書 (投野 2016) にまとめられて、各種 profile は公開の方向で整備される。
- (3) 社会的インパクト
 CEFR-J の Grammar, Text などのプロフィール情報の整備により、CEFR-J のレベル別に語彙・文法・テキストを組み合わせてタスクを開発したり、パフォーマンステストを設計したりすることが容易になる。
 CEFR-J の CAN-DO リスト、語彙表などの一連の言語材料の資料群は研究・商用利用を問わずオープンな枠組として公開されており、教科書会社、教材制作会社、テスト会社などが具体的な商品を CEFR-J を元に開発を始めている。
 文部科学省が 2020 年を目指して改革する学習指導要領でも「指標形式の到達目標」として CAN-DO ベースの目標を組み込むことがほぼ決定しており、CEFR-J とその一連の利用環境の整備は、学習指導要領に CAN-DO をどう組み込み、どう小中高の各レベルで教員研修・指導実践に生かしていくか、について有益な情報を提供するものと期待される。

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 54 件)

Masashi Negishi & Yukio Tono, An update on the CEFR-J project and its impact on English language education in Japan, *Language assessment for Multilingualism Paperback: Proceedings of the ALTE Paris Conference, April 2014 (Studies in Language Testing)*, pp.113-133, 2016 年, 査読有

Yukio Tono, Automatic extraction of L2 criterial lexico-grammatical features across pseudo-longitudinal learner corpora: Using edi distance and variability-based neighbor clustering, *L2 Vocabulary Acquisition: Knowledge and Use: New perspectives on assessment and corpus analysis*, No.2, pp.149-176, 2013 年, 査読有

Masashi Negishi, Tomoko Takada and Yukio Tono, Developing corpus-based word lists for English language learning and teaching: A critical appraisal of the English Vocabulary Profile, A Progress report on the development of the CEFR-J, *Studies in Language Testing 36 Exploring Language Frameworks Proceedings of the ALTE Krakow Conference, July 2011*, 2013 年, pp.135-163, 査読有

Masashi Negishi, Yukio Tono and Yoshihito Fujita, A Validation Study of the CEFR Levels of Phrasal Verbs in the English Vocabulary Profile, *English Profile Journal*, Vol.3, 2012 年, 査読有,
<http://dx.doi.org/10.1017/S2041536212000037>

Yukio Tono Developing corpus-based word lists for English language learning and teaching: A critical appraisal of the English Vocabulary Profile, *Input Process and Product: Developments in Teaching and Language Corpora*, 2012 年, pp.314-328, 査読有

Yukio Tono and Masashi Negishi, The CEFR-J: Adapting the CEFR for English Language Teaching in Japan, *Framework & Language Portfolio (FLP) SIG News letter*, 2012 年, pp.5-12, 査読有

[学会発表] (計 122 件)

石井康毅・投野由紀夫, CEFR-J Grammar Profile のための文法項目頻度調査, 言語処理学会第 22 回年次大会, 2016 年 3 月 9 日, 東北大学(宮城県)

林正頼・石井康毅・高村大也・奥村学・投野由紀夫, 英語学習者の英作文から CEFR レベル別基準特性の特定, 言語処理学会第 22 回年次大会, 2016 年 3 月 9 日, 東北大学(宮城県)

水嶋海都・荒瀬由紀・内田諭, CEFR 準拠教科書における語彙・構文の特徴分析とレベル自動分類, 言語処理学会第 22 回年次大会, 2016 年 3 月 9 日, 東北大学(宮城県)

投野由紀夫・石井康毅, 英語 CEFR レベルを規定する基準特性としての文法項目の抽出とその評価, 言語処理学会第 21 回年次大会, 2016 年 3 月 9 日, 京都市

学(京都府)

Yukio Tono, Linguistic feature extraction and evaluation using machine learning to identify “criterial” grammar constructions for the CEFR levels, *Corpus Linguistics* 2015, 2015 年 7 月 22 日, Lancaster University (イギリス)

Yukio Tono, Mining Language Learners’ Production Data for Understanding of L2 Learning Systems, *PACLIC27*, 2013 年 11 月 22 日, 台湾国立政治大学(台湾)

Yukio Tono, “Criterial feature” extraction from CEFR-based corpora: Methods and techniques, *Corpus Linguistics* 2013, 2013 年 7 月 23 日, Lancaster University (イギリス)

[図書] (計 46 件)

投野由紀夫編, 「学習者コーパスによる英語 CEFR レベル基準特性の特定と活用に関する総合的研究」研究成果報告書, 2016 年, 全 459 ページ

投野由紀夫編, コーパスと英語教育(英語コーパス研究シリーズ 第 2 巻), ひつじ書房, 2015 年, 全 220 ページ

投野由紀夫, 発信力をつける新しい英語語彙指導, 三省堂, 2015 年, 全 144 ページ

投野由紀夫編著, 英語教師のためのコーパス活用ガイド, 大修館書店, 2014 年, 全 242 ページ

投野由紀夫, CAN-DO 作成・活用 英語到達度指標 CEFR-J ガイドブック, 大修館書店, 2013 年, 全 313 ページ

投野由紀夫・金子朝子・和泉絵美・杉浦正利, 英語学習者コーパス活用ハンドブック, 大修館書店, 2013 年, 全 249 ページ

[その他]

CEFR-J ホームページ

<http://www.cefr-j.org>

6. 研究組織

(1) 研究代表者

投野 由紀夫 (TONO YUKIO)

東京外国語大学・大学院総合国際学研究院・教授

研究者番号: 10211393

(2) 研究分担者

根岸 雅史 (NEGISHI MASASHI)

東京外国語大学・大学院総合国際学研究院・教授
研究者番号：50189362

相川 真佐夫 (AIKAWA MASAO)
京都外国語大学・外国語学部・教授
研究者番号：60290467

寺内 一 (TERAUCHI HAJIME)
高千穂大学・商学部・教授
研究者番号：50307146

中谷 安男 (NAKATANI YASUO)
法政大学・経済学部・教授
研究者番号：90290626

奥村 学 (OKUMURA MANABU)
東京工業大学・精密工学研究所・教授
研究者番号：60214079

金子 恵美子 (KANEKO EMIKO)
会津大学・コンピュータ理工学部・教授
研究者番号：30533624

能登原 祥之 (NOTOHARA YOSHIYUKI)
同志社大学・文学部・准教授
研究者番号：70300613

石井 康毅 (ISHII YASUTAKE)
成城大学・社会イノベーション学部・准教授
研究者番号：7053103

内田 諭 (UCHIDA SATORU)
九州大学・言語文化研究院・准教授
研究者番号：20589254

和泉 絵美 (IZUMI EMI)
同志社大学・全学共通教養教育センター・准教授
研究者番号：80450691