様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　24 年　　5 月　15 日現在

機関番号：12603
研究種目：基盤研究（B）
研究期間：2008～2011
課題番号：20320077
研究課題名（和文）
大規模英語学習者コーパスのエラー情報を活用した言語テスト構築の基礎的研究

研究課題名（英文）　Basic Research on Language Test Design and Construction Using Error Information from Large-Scale Learner Corpora

研究代表者　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

研究成果の概要（和文）：
従来の言語テストは、専ら「宣言的知識」を測定してきたと思われる。そこで、本研究では、「手続き的知識」を測定することのできるテストの開発を試みた。このために、大規模英語学習コーパスのテキスト分析を自動で行うことにより、学習者の習得段階を明らかにし、これを反映するようなテスト方法を模索した。「テスト」という手法自体は必ずしもうまく機能しなかったものの、作文の「チェックリスト式採点」はある程度の信頼性のある結果を得ることができることわかった。

研究成果の概要（英文）：
Conventional language tests have exclusively measured “declarative knowledge”. In this study, we attempted to develop language tests which would measure “procedural knowledge”. In so doing, we searched for test methods that would reflect acquisition order derived from the analysis of large scale English learner corpora. While tests methods tired out for this research did not function very well, the checklist method for assessing writing has produced quite reliable results.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	4,200,000	1,260,000	5,460,000
2009 年度	3,600,000	1,080,000	4,680,000
2010 年度	3,500,000	1,050,000	4,550,000
2011 年度	3,900,000	1,170,000	5,070,000
年度			
総　計	15,200,000	4,560,000	19,760,000

研究分野：英語教育学
科研費の分科・細目：言語学・外国語教育
キーワード：学習者言語、言語テスト、エラー分析、習得度診断、第二言語習得、CEFR、学習者コーパス、到達度指標

１．研究開始当初の背景

英語の文法能力を測定すると言えば、伝統的な「文法テスト」が従来用いられてきていた。いわゆる、空所補充問題や並べ換え問題である。

しかし、近年の第 2 言語習得研究において

は、言語知識研究に関して、「宣言的知識」と「手続き的知識」といった分類や「明示的知識」と「暗示的知識」といった分類の必要性が強調されてきている。言語規則を明示的に説明できる能力とその言語規則を実際に使う能力は、必ずしも同一ではないというものである。

このような認識は、言語習得研究では広く広まってきたものの、言語テスト研究には大きな影響を及ぼしておらず、様々な言語テストがこれまでに開発してきているにもかかわらず、文法テストは改革されておらず、いわゆる「宣言的知識」や「明示的知識」を測るだけのものであった。そこで、これまでに測ってこなかった「手続き的知識」や「暗示的知識」を測定できる文法テストの開発が求められていた。

また、その一方で、大規模な学習者データを集積した英語学習者コーパスも開発されつつあった。こうした学習者コーパスにエラー情報を付与することで、学習者の英語の習得順序が明らかになる。

学習者コーパスを用いた英語習得順序の研究は和泉他（2004）,投野（2007）などで行われていたが、CEFR のような英語能力指標との関連はあまりなかった。しかし、English Profile Programme が学習者コーパスによる基準特性の特定という課題に取り組むようになり、Profiling Research という名称で徐々に注目されるようになった。
さらに、英語学習者コーパス分析の成果は、文法の習得順序のみならず、語彙的な知識の習得順序に関しても、有益な情報をもたらすようになっていた。

２．研究の目的
(1)英語学習者データを大量に収集電子化した「英語学習者コーパス」をもとに、学習者の語彙・文法エラーに関する情報をコーパスから系統的に抽出

(2)学習者コーパスからの基準特性の抽出を行うために、どのような言語特徴の注釈付け（annotation）が有効か、またそれから基準特性をどのような手法を用いて抽出するべきか、という 2 点に関して研究を行った。

(3)それらの学習レベル別の特性を分析

(4)分析結果をもとに言語テスト作成を行う方法論を確立

(5)テストの試作品を作成・検証

３．研究の方法
(1)学習者の語彙・文法エラーに関する情報をコーパスから系統的に抽出するために、英語学習者データを大量に収集電子化し、それらのタグ付け等を行い、「英語学習者コーパス」を構築した。

(2)和泉は投野（2007）で構築した 1 万人の中高生英作文コーパス JEFLL Corpus の生徒の書いた文と母語話者による訂正文の対応付けデータから RASP という構文解析器を用いて English Profile が行っているのと同様の構文解析を施し、その差分を特定して、それに有効なエラータグを自動付与する方法を研究した。投野は JEFLL Corpus および International Corpus of Crosslinguistic Interlanguage (ICCI) を用いて、RASP から動詞格フレームを取り出し、それが English Profile の提案する A2 レベルの基準特性を補正する A1, A2 データが有効に機能するかどうかの検証を行った。

(3)学習者の CEFR レベルごとに、「英語学習者コーパス」を分析し、学習レベル別の特性を明らかにした。

(4)従来の言語テストを振り返り、学習レベル別の基準特性に反映するであろう「手続き的知識」や「暗示的知識」を測定するような方法を模索した。

(5) その中から選ばれた手法を試行的に実施し、そのデータを分析して、想定される学習レベルを反映するものとなるかを検証した。

４．研究成果
(1)英語学習者コーパスの構築とタグ付け

日本人学習者に特有の英語ライティングにおける CEFR 各レベルの基準特性を明らかにするため、E メール、物語文、論説文の 3 つのジャンルの作文をコーパス化し、学習者コーパス分析を行なった。作文課題は、English Profile Programme のライティング課題を修正の上で利用し、2009 年度及び 2010 年度の 1 年次、2 年次生を対象として、2 学期始めの主・副専攻英語授業内で実施した。作文はすべてデジタル化された上で、文法タグ(UCREL CLAWS7)が付けられ、2 名の評価者により CEFR に準拠した課題ごとに作成された評価基準をもとに採点された。結果、のべ 5.781 名、総語数約 825,000 語のデータが分析に用いられた。

(2)学習者文と訂正文の構文解析データ比較と自動エラータグ付与

和泉は学習者文と訂正文の構文解析を RASP (Robust Accurate Statistical Parsing) で解析し、構文依存関係の抽出を行った。

例）RASP の出力結果の一例

入力文： I eat breakfast every day.

RASP によって出力される構文依存関係の解析結果：

(|I:1_PPIS1| |eat:2_VV0| |breakfast:3_NN1| |every:4_AT1| |day:5_NNT1| |.:6_.|)

(|ncsubj| |eat:2_VV0| |I:1_PPIS1| _)

(|ncmod| _ |eat:2_VV0| |day:5_NNT1|)

(|det| |day:5_NNT1| |every:4_AT1|)

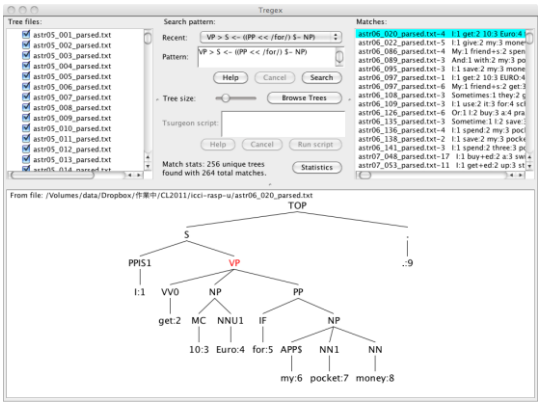
(|dobj| |eat:2_VV0| |breakfast:3_NN1|)

次に、学習者文と訂正文の構文解析から依存関係の違いを見つけることにより、両者の差分（学習者文の誤り箇所）を特定する。具体的には、学習者文と訂正文の RASP 出力結果に含まれる依存関係を先頭から順に比較し、その一致度（依存関係の種類および head、dependent として出力されている単語がどの程度一致するか）によって、学習者文内の各語が正解かどうか（訂正されていないかどうか）判定した（そのための判定ルールを 20 項目作成）。差分（誤り）が抽出された箇所には、表層的エラータグ（学習者文内のある単語が訂正文内で別の単語に置き換えられている場合は置換タグ：<rep crr= “訂正語”> 、新たに単語が加えられている場合は挿入タグ：<ins crr= “追加語”>、単語が削除されている場合は削除タグ：<del>を付与した。

最後に、手順 2 で表層的エラータグが付与された箇所の誤り語と訂正語のペアを参照し、その品詞や訂正の種類を手がかりに文法・語彙エラータグへの変換を行った。ここでは、The NICT Japanese Learner English (JLE) Corpus プロジェクトで開発された 47 種類の文法・語彙エラータグのセットを使用した。

(3)基準特性の抽出方法の検討

投野は English Profile が文法の主要な基準特性として挙げている動詞型の共起パターン（verb co-occurrence pattern）に注目し、特に A2 レベルの基準特性として挙げているパターンのいくつかは A1 レベルではないかという疑問から、A1 レベルを大量に含んだ日本人データでの再検証を試みた。RASP で構文解析済みの JEFLL Corpus および ICCI データを用いて、Tregex（構文解析データの抽出ツール）により動詞の下位範疇化パターンの頻度を、CEFR の A1～B2 に英作文の長さによって自動分類されたテキストから抽出を行った。



各レベルからの動詞下位範疇化パターンの頻度を Hierarchical Configural Frequency Analysis (HCFA, von Eye 1990; Gries 2009) によって分析し、各レベルに有意に頻度の高いものを基準特性として抽出したところ、表１のような English Profile の提示した A2 レベルの基準特性からの変更が示唆された：

Co-occurrence frames	Criterial for CEFR Level:
NP-V	A2 → A1
NP-V-PP	A2
NP-V-NP	A2 → A1
NP-V-Part-NP	A2 → B1
NP-V-NP-Part	A2 → B1
NP-V-NP-PP	A2
NP-V-NP-PP (P=for)	A2
NP-V-V (+ing)	A2 (strong predictor for B1)
NP-V-Vpinfinitaival (Subj Control)	A2 (strong predictor for B1)
NP-V-S	A2 (strong predictor for B1)

このようにA レベルのデータを増殖させて見ることにより、A1 レベルの特性として再定義できる可能性や、HCFA のような新しい統計処理を行うことで、基準特性の抽出に一定の成果があることを示した。

(4)語彙分析
分析にあたっては、JACET8000 の語彙リストとのマッチングを行なった。分析としては、共起ネットワーク分析、コレスポンデンス分析、N グラム分析が行われ、A2、A2+、B1、B1+、B2 のそれぞれのレベルにおける課題タイプごとの語彙の基準特徴が示された。

(5)句動詞テストの開発と句動詞習得順序の調査
語彙項目の 1 つとしての句動詞の習得段階を見るテストを開発し、日本人英語学習者の句動詞の習得順序を調べ、English Profile Programme における CEFR レベル判断との比較を行った。

(6)CEFR 基準特性チェックリスト方式の開発
英作文に出現した CEFR 基準特性等をもとに作成したチェックリストに、出現項目をチェックしていき、それらを正答項目として項目応答理論による項目分析を行うことで、CEFR レベルをある程度推定できることがわかった。

(7)CEFR 基準特性テスト開発
CEFR 基準特性等を数多く含む作文を選び、それらを何度か聞かせた後で、再生させるというテストを開発した。しかしながら、このテストの項目弁別力はあまり高いものではなかった。理由としては、作文の場合は、それぞれの基準特性項目を「使う」かどうかという判断は、学習者自身が行っており、それが能力の弁別に貢献していたと考えられるが、再生テストでは、基準特性項目は全員に示されており、必ずしも自らは使おうとしないような項目も産出してしまった可能性が考えられる。

今後の課題としては、テスト開発に関するものと語彙・文法習得に関するものとがある。
まず、テスト開発については、作文のチェックリスト型採点の有効性はある程度認められたが、この手法は実施と採点に手間がかかり、充分な実用性を備えているとは言いがたい。したがって、より実用性の高いテスト方法を今後開発していく必要がある。
また、語彙習得に関しては、日本人学習者に特有の基準特性の分析に、English Profile の語彙リスト等の基準リストとの比較が必要であること、また、高頻度語彙だけでなく、低頻度語彙の分析などが必要となることなどが課題として考えられる。また、文法パターンの分析も必要となるであろう。2011 年度以降も継続してデータを収集しており、現在人数の少ない B2 レベルのサンプルも増やしながら、さらなる分析を行なっていきたい。

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 25 件）

① 根岸雅史、CEFR 基準特性に基づくチェックリスト方式による英作文の採点可能性、*ARCLE REVIEW*、査読無、No.6、2012、80-89
② 根岸雅史，文法テストはこのままでよいのか，日本言語文化研究会論集、査読有、第 7 号、2011、1-15
③ Masashi Negishi, Yukio Tono, and Y. Fujita, A validation study of the CEFR levels of phrasal verbs in the English profile wordlists. *Proceedings of the 43rd Annual Meeting of the British Association for Applied Linguistics 9-11 September 2010 University of*

Aberdeen., 査読有、pp. 283-293
④ Yukio Tono, A Critical Review of the Theory of Lexicographical Functions, *LEXICON*, 査読有, 40 号, 2010, 1-26

〔学会発表〕(計 59 件)
① Masashi Negishi, The development of the CEFR-J: Where we are, where we are going, WoLSEC International Symposium 2011, 2011年3月2日, 東京外国語大学
② 和泉絵美, 知っている語彙と使える語彙―英語学習者コーパスからの検証―, 京都外国語大学メビウス研究会第 164 回研究会, 2010 年 11 月 27 日, 京都外国語大学
⑤ 工藤洋路・長沼君主・小野倫寛・高野正恵・増田斐那子, 英語ライティングにおけるコンピュータ自動採点の妥当性と CEFR ライティング評価との関連の検証, 外国語教育学会, 2010 年 11 月 14 日, 東京外国語大学
③ Masashi Negishi, Yukio Tono, and Yoshihito Fujita, A validation study of the CEFR levels of phrasal verbs in the English Profile Wordlists, British Association for Applied Linguistics, 2010年9月9日, University of Aberdeen
④ Yukio Tono, Automatic extraction of L2 criterial lexico-grammatical features across pseudo-longitudinal learner, The European Second Language Association (EUROSLA) ― Thematic colloquium: Researching vocabulary use: insights from corpus analysis 2010 年 9 月 4 日, Reggio, Emilia, Italy

〔図書〕(計 11 件)
① 投野由紀夫(監修) & Christian James(英文校閲), 東京書籍, フェイバリット英単語・熟語＜テーマ別＞コーパス 1800, 2010, 272pp.
② Weir, G. & Ishikawa, S. (eds.), University of Strathclyde Publishing, Learner Corpus Research: Some Recent Trends. In: Corpus, ICT, and Language Education, 2010, 140pp. (pp. 7-17)

６．研究組織
(1)研究代表者
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

(2)研究分担者
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

長沼　君主　(NAGANUMA NAOYUKI)
東京外国語大学・世界言語社会教育センター・講師
研究者番号：20365836

工藤　洋路　(KUDO YOJI)
東京外国語大学・世界言語社会教育センター・講師
研究者番号：60509173
和泉　絵美　(IZUMI EMI)
京都外国語大学・外国語学部英米語学科・非常勤講師
研究者番号：80450691