

平成 26 年 6 月 10 日現在

機関番号：32617

研究種目：研究活動スタート支援

研究期間：2012～2013

課題番号：24820041

研究課題名(和文) パラレルコーパスによる日本人英語学習者のライティングの分析

研究課題名(英文) The analysis of writing by Japanese learners of English with a parallel corpus

研究代表者

三木 望 (Miki, Nozomi)

駒澤大学・総合教育研究部・講師

研究者番号：00632100

交付決定額(研究期間全体)：(直接経費) 2,100,000円、(間接経費) 630,000円

研究成果の概要(和文)：本研究は、日本人英語学習者のエッセイ340個に異なる英語圏の母語話者62名によって、3通りの添削文を作成して(計1020個)、日本人学習者のエッセイと英語母語話者による添削文の間に文単位で1対1の対応関係を持つパラレルコーパスを作成した。そして、パラレルコーパスで英語母語話者が日本人学習者の過剰使用語をどのように添削したのかについて量的・質的に分析して、日英の発想の違いの一端を明らかにした。ライティングの教師や英語学習者のために、ウェブ上で使用が可能なENEJE (English Native Edited Japanese Essays) パラレルコーパスを開発して公開した。

研究成果の概要(英文)：In this research, 340 Japanese students' essays were corrected separately by any three of 62 native-English editors, producing 1020 corrected essays. With these data, I compiled three sets of parallel corpora showing sentence-by-sentence correspondence between the uncorrected and corrected essays. I conducted quantitative and qualitative analyses to see how the English native speakers corrected these essays and discovered a different conception of language in the two groups. The web-based parallel corpus engine, the ENEJE (English Native Edited Japanese Essays) Parallel Corpus was developed and released for the benefit of teachers and learners of English.

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：コーパス言語学 学習者コーパス 添削文 日本人学習者 ライティング

## 1. 研究開始当初の背景

### (1) 従来の学習コーパスの問題点

学習者コーパスの研究は、学習者コーパスを英語圏の大学生のエッセイのコーパス (LOCNESS) と比較して得られる過剰語と過少語やコロケーションの分析が中心である。しかし、同じ論説文でもトピックやエッセイの長さが異なる参照コーパスとの比較は、特徴語の結果に影響を与えやすい。そして、学習者コーパスから独立して産出され、直接的な対応関係がない英語母語話者の参照コーパスから、学習者が表現したかった英語を探すのは、知識と経験が必要で、対応する表現があるとは限らない。学習者コーパスの研究はライティングのフィードバックにおいて、教師がよく直面する疑問「学生の英語は、英語母語話者ならばどのように表現するのだろうか」について、必ずしも直接的な答えを提示するものではない。

### (2) 研究代表者の過去の研究と問題点

これらの問題を解決するために、研究代表者の過去の研究では、翻訳研究の平行コーパスの研究を応用して、NICE (the Nagoya Interlanguage Corpus of English) の TOEIC/TOEFL のスコアが判明している学習者のエッセイと英語母語話者による添削文から、文単位で対応関係を持つ平行コーパスを作成して、いくつかの言語特徴を分析して発表した。このパイロットスタディでは、学習者のライティングの特徴について新たに判明した点もああったが、課題も残った。学会の発表や論文の査読で、英文校正者の個人差の影響の可能性が指摘され、英文校正者に関する情報について質疑が出たが、NICE に校正者の詳細な情報が載っていない。パイロットスタディは、学習者の習熟度別の分析をする目的で、NICE から学習者の TOEIC のスコアがわかるデータだけを使用しているため、結果について統計的妥当性を得る程サイズが大きくない。また、この研究を開始したとき、NICE (Ver.2.0.2) に追加されたデータには添削文はついていなかった。

## 2. 研究の目的

本研究の目的は、日本人学習者のライティングの添削文の平行コーパスを作成することと、日本人のライティングの特徴を従来のコーパスと異なる視点から明らかにすることである。そして、添削文の平行コーパスのインターフェースを開発してインターネットで公開することである。

### (1) 添削文の平行コーパスの作成

英文校正者の透明性を高めるため、NICE の既存の添削文を使用せずに、新たに添削文をつけて、平行コーパスを作成する。英文校正者の個人差による影響を軽減するために、1つのエッセイに対して、3人の英文校正者によって3通りの添削文を作成する。

習熟度情報が統一されていない問題を解決するため、英語を母語とする英文校正者がスコアを付けて、習熟度別の分析を可能にする。

### (2) 添削文の平行コーパスの量的・質的分析

文単位で結果が得られる平行コーパスの特徴を活かして、接続詞を含む複文、特にパイロットスタディで扱った原因と結果に焦点をあて、学習者と英語母語話者間の認識の違いを明らかにする。

### (3) 添削文の平行コーパス・オンライン

研究者だけでなく英語教師や学習者にも利用しやすいシステムを開発して、インターネットで添削文の平行コーパスを公開する。

## 3. 研究の方法

### (1) 添削文の作成

NICE (Ver. 2.1.2) の英語学習者のエッセイ (340 テキスト) の添削文を英文校正会社に依頼して作成した。NICE には添削文があるが、添削の条件を統一するために、既存の添削文は敢えて使用せず、新たに添削文を作成した。校正会社への添削に関する指示は以下の通りである。

学習者のエッセイ1つに対して、異なる英文校正者3人で3通りの添削文を仕上げる。

3人の英文校正者のうち1人は異なる国籍の出身者が担当する。

校正者の情報(出身国、性別、年齢、学位、英文校正の経験)を担当するテキストの冒頭に記載する。

各文の添削で気が付いた点はすべてその文のコメント (COM) として記載する。

文法的で自然な英語であれば、学習者の英語を使用する。

英文校正者はワードなどのソフトウェアの校閲機能を使用せずに、文ごと書き直したテキストファイルを納品する(つまり赤いれではない)。

英文校正者は、CHAT形式ではなく、パラグラフ単位で保存された学習者の英文を一読してから、6段階評価で、ライティングにスコアを付ける。

文脈上、文全体を削除した方がよい場合は、文法を修正した上で<del></del>のタグで当該の文を囲む。

2文以上を接続詞で連結した方がよい場合は、%MER\_NTVで始まる行を追加して、接続詞で連結された2文を書く。

評価基準は、NICEにライティングの指示文がないことを考慮して、Criterion (ETS) の Generic Scoring Model for High School (Grade 9 to 12)を使用した。

研究代表者は、納品された添削文を全て目視で確認して、英語のケアレスミスや疑問が

ある場合、担当の英文校正者と連絡を取り確認して修正させた。学習者が日本語を使用している場合は、その意味を英語で説明した上で英文校正者に当該の表現を英語に修正させた。

## (2) 添削文の平行コーパスの整形

%MERG\_JPN と %MERG\_COM

(1) の添削文の作成の留意点 について、添削文の %MERG\_NTV に合わせて、%MERG\_JPN と %MERG\_COM の行を挿入して、接続詞で連結される前の日本人学習者の文 2 つと校正者のコメント 2 つ (主に連結の理由) を手作業でそれぞれ写してまとめた。

学習者の英文と添削文の対応関係の形成

NICE には、英文の他に年齢やコメントなど様々な情報が行単位で保存されているので (CHAT 形式)、その中から学習者の英文だけを抽出して、個別ファイルに保存した。平行コーパスで 1 文でもずれると検索ができないので、最後は各ファイルの行数を基にして手作業で対応関係の確認を行った。

スコア別の平行コーパスの編成

学習者のライティングの 3 つのスコアの中央値に基づいて、6 つのレベルを初級・中級・上級に分け、それぞれ日本人英語学習者コーパスと添削文コーパスを形成した。

## (3) 添削文の平行コーパスの分析

量的分析

学習者と添削文のコーパス及び習熟度別サブコーパスの特徴語と連語、特に 3 連語 (trigram) を尤度比検定に基づいて、学習者の語彙の特徴をまとめた。

学習者と添削文のコーパスに CLAWS (Ver.

4) で品詞タグを付与して、学習者のエッセイと添削文の接続詞の頻度の違いを分析した。

質的分析

量的分析で特に有意な差異を示した接続詞とパイロットスタディで取り上げた原因・結果の because/so を平行コーパス専用ソフトウェア ParaConc で検索して、文単位で抽出して、分析した。

## (4) 平行コーパスのオンライン検索ソフトの開発

平行コーパスのオンライン・ソフトウェアをローレンス・アントニ氏に協力を依頼して開発してもらった。研究代表者はローレンス・アントニ氏と入念な打ち合わせをして、各機能についてフィードバックを行い、議論を重ねた。

## 4. 研究成果

### (1) 添削文コーパスの完成

添削文コーパスの総語数 (token) は、3 セットを合わせて、約 340 万語で、異なり語 (type) は、約 7000 語だった。この研究に携わった英文校正者は計 62 名で出身はアメリカ 31 人 (50%)、イギリス 26 人 (42%)、カナダ 4 人

(6%)、オーストラリア 1 人 (2%) で、性別では、男性 17 人 (27%) と女性 45 人 (73%) である。

日本人英語学習者のスコア別の分類では、スコア 3 が最も多く (145 個、42.6%)、スコア 2 が続いた (112 個、32.9%)。そして、最高スコアの 6 は 1 つだった (0.3%)。実際の分析では、スコアを初級 (スコア 1, 2)、中級 (スコア 3, 4)、上級 (スコア 5, 6) に分類して行った。

これまで、他の研究者によって日英の平行コーパスが開発・公開されてきたが、学習者のライティングとその添削文から構成される平行コーパスは数少ない。更に、英語学習者の 1 つのエッセイに対して 3 通りの添削文があるデータベースは、研究代表者の知る限りない。3 通りの添削文があることによって、研究者が英語の添削が偶発的な言語使用なのか、校正者個人の嗜好なのか、疑問を持ったときに、ある程度調べることができる。

### (2) 添削文の平行コーパスの分析の結果

量的分析の結果

特徴語分析と 3 連語 (trigram) の分析によって、日本人学習者が口語的な縮約形と二人称代名詞を過剰使用しているのに対して、添削文ではフォーマルな表現の頻度が高いことがわかった。これは、パイロットスタディの結果と一致する。一方、学習者コーパスの定冠詞・不定冠詞の過剰使用は、日本人学習者にとって冠詞の習得が問題であることを示唆している。

接続詞の量的分析では、一部の接続詞 (unless, while, till) を除いて、全体的に学習者は接続詞を過剰使用していた。特に、等位接続詞 (and, but, or) が上位を占めていた。

質的分析の結果

学習者が等位接続詞を副詞のように文頭で過剰に使用し、同じ多義語を繰り返し使用するのに対して、添削者は意味が明確な表現に書き換えていたことがわかった。例えば、学習者は多義語の and を過剰に使用するが、校正者は追加 (furthermore, also, moreover) や原因・結果 (because, so, consequently) の明確な論理関係を示す談話標識に書き換えていた。

日本人学習者は接続詞的用法の so の頻度が高いが (例: I want to get a lot of money so I work hard in September.)、この接続詞の so に対応する添削文を分析すると and so や and therefore に置換されていることが多かった (例: I want to earn a lot, and therefore, I work hard in September.)。これは日本人学習者が、推測して帰結を導くべきところで、直接的な因果関係として事象を表現していることを示している (cf. Content Domain vs. Epistemic Domain (Sweetser, 1990))。量的にも CLAWS で

so\_CS とタグが付与される接続詞の so と and  
 CC so\_RR とタグが付与される副詞の so、and  
 CC therefore\_RR とタグが付与される  
 therefore は、日本人学習者のエッセイと添  
 削文の間で有意な差異を示し、質的分析の結  
 果を裏付けた。

日本人学習者の因果関係の把握の仕方が  
 異なることについて、関連学会で口頭・論文  
 発表した。今後は、接続詞だけでなく、談話  
 標識全体を分析対象にして、本研究のパラレ  
 ルコーパスで分析したい。

### (3)The ENEJE (English Native Edited Japanese Essays) Parallel Corpus の開発と公開

ENEJE パラレルコーパスは KWIC  
 Concordancer、File View、Word List の3つ  
 から構成されている。KWIC Concordancer で  
 は、ある特定の語(句)を検索すると、同時  
 に上下に二つのウィンドウズが現れ、上の  
 ウィンドウズには、学習者コーパスの検索語  
 を中央に含む行が表示され、下のウィンドウ  
 ズには対応する添削文が表示される。File  
 View では、テキスト全体が表示されるだけ  
 なく、日本人学習者の情報(ID、年齢、性、  
 専攻、英語の資格)と英文校正者3名の当該  
 のテキストの各スコア及び合計と中央値、コ  
 メントが表示される。Word List は、学習者  
 コーパスと3つの添削文のコーパスが並行し  
 て表示される。また、これら3つの主要機能  
 を使用する際に、エッセイのトピック、スコ  
 ア、校正者の国籍と学位、学習者の性と専攻  
 を指定できる。

これまで、NICE の添削文は CHAT 形式で保  
 存されて、インターフェースがないため、  
 コーパスを専門としない一般の研究者や英  
 語教師、学習者には利用しにくかった。ENEJE  
 パラレルコーパスは、利用者にわかりやすい  
 ツールとして開発され、研究だけでなく、教  
 育にも貢献することを目指している。

今後は、ENEJE パラレルコーパスが利用者  
 にとっていっそうわかりやすいツールにな  
 るように改良して、その機能に特徴語分析  
 (Keyword analysis)等を追加していきたい。  
 そして、ENEJE パラレルコーパスのタスクを  
 考案・実践して、学会で報告したい。マイ  
 ル型の ENEJE パラレルコーパスの開発も検討  
 したい。

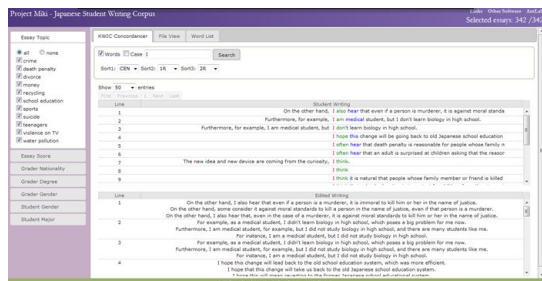


図1 The ENEJE (English Native Edited Japanese Essays) Parallel Corpus

## 5. 主な発表論文等

〔雑誌論文〕(計1件)

Miki, N. "A parallel corpus approach to Japanese learners' causality in argumentative writing," *Opening new lines of communication in applied linguistics: Proceedings of the BAAL Annual Conference 2013*. 査読有, 2014年掲載予定.

〔学会発表〕(計1件)

Miki, N. "A parallel corpus approach to Japanese learners' causality in argumentative writing," The 46<sup>th</sup> Annual meeting of the British Association of Applied Linguistics, 2013年9月5日、ヘリオット・ワット大学(英国)

〔その他〕

The ENEJE (English Native Edited Japanese Essays) Parallel Corpus  
<http://www.antlab.sci.waseda.ac.jp/eneje/>

## 6. 研究組織

### (1)研究代表者

三木 望 (MIKI, Nozomi)

駒澤大学・総合研究教育部・講師

研究者番号: 00632100