

# 「テキストマイニング分析による日本人英語学習者の 理解度別英語習得過程研究」

浅野恵子 (順天堂大), 小川敦 (一橋大大学院) 小堂俊孝 (日大)

Keiko\_asano@sakura.juntendo.ac.jp atsushi.ogawa@arcor.de Kodoh.toshitaka@nihon-u.ac.jp

## 1 はじめに

本研究では、第二言語学習者は母語話者のそれと言語獲得背景、目標言語の習得方法が異なるため、体系的に文法ルールを学ぶ必要があるという仮説に基づいて、第二言語としての日本人英語学習者がどのような文法知識を習得、構築していくかをテキストマイニングという方法で分析した。第二言語学習者の習得過程を調査するにあたり、学習者の母語の干渉から来る誤りや誤答分析は多く行われてきた[1]。しかし誤りには複雑な要因が絡み合い数量化できるものでもないという指摘もある[2]。特に選択肢から解答を選ぶテストではない自由作文においては、どこからが間違えて、どこからが正解であるか、評価する側の観点の問題も生じてくる。また、第二言語習得過程を断片的にしかとらえておらず、さまざまな習得段階における学習者の誤りの変化をとらえている研究も殆どないといわれている[3]。今回使用したテキストマイニングによる分析は単に言語習得時における誤答に固執するのではない、という点が特徴の一つである。

従来、テキストマイニングはアンケートの自由記述文やコールセンターのオペレータ記録、業務日報、さらにはブログなどの自由回答された大量の文書から情報を抽出し、解釈のばらつきが生じるそれらの情報から質的な相違を分析するために用いられてきた。

習得過程別の学習者を上級と中級の 2 グループに分け、テキストマイニング用のソフトを用いて自由英作文を実施・分析した。学習者の各グループ間で使用する単語頻度、品詞、係り受け単語、ターゲット単語を観察した。さらに各問題間で共通するであろう文法項目

も抽出し、分析を試みた。理解度別グループ間で特徴的傾向が、英文から観察された。

## 2 実験

### 2.1 言語材料

分析用英文は、60 問の短い会話文で構成されており、分析対象とした英文を日本語に直したものを提示し、英語に自由作文を指示したものを使用した。それらの会話文は文法構造を分析・網羅した英文法テスト[4]を元に行っている。回答問題数は全部で 60 問である。

### 2.2 方法

分析対象となった英文は、被験者によって自由回答された 60 問のうち、この英文法テストで比較的正答率の低かった (60% 台以下) 問題の 10 文を選択したものである。A: と B: は問題中の会話の発話者側を示している。日本語以外の会話部分の英語をそのまま残して提示したので、会話の流れを作文作成時に反映できるようにしてある。以下の 10 文が自由英作文させるために用いた日本語部分である。番号は 60 問中、今回の分析で取り上げたものを示している。

5. A: Napoleon was born in 1769.

B: ナポレオンはいつ生まれたって?

9. A: この若者は成功するだろう。

B: What makes you think so?

11. A: 貧乏人は金持ちから憐れまれるのを嫌う。

B: The rich like to pity the poor.

13. A: 多くの交通事故がここ 2, 3 か月のうちに増加している。

B: That's too bad.

24. A: The swindler killed himself.

- B:信じられない！そんなに汚い男が自ら身を断つなんて。
28. A:多くのアメリカ人が海を見たことがない。
- B:That's not quite true.
33. A:この物質はガンの原因であることをいくつかの実験が示してきている。
- B:How terrible!
39. A:メアリーは恋人いない歴8か月だ。
- B:Her last boyfriend was just awful, wasn't he?
40. A:車があると便利だね。
- B:Not really in a big city
44. A:ジャケットに合うネクタイを探しています。
- B:These are on sale today, sir.

この英文法テストの正答率は、今回の被験者と同じレベルのグループに対して 2007 年に実施した結果であり、今回の被験者には英文作成の手がかりを与える懸念があるので、このテストは実施していない。英文法テストにおいて使用された英文の解答例と正解率を Appendix で提示する。

分析に使用したテキストマイニングのソフトは TRUE TELLER 英語版 Version 6.0.0 Build 3 A[5]である。自由英作文した文の各単語の頻度数、品詞別分類、係り受けなどが分析可能な点から、言語習得に関連する調査にも応用が可能と思われる。

## 2.3 被験者

実験に参加した被験者は日本人大学生 1.2 年生の 84 名で、必修科目である英会話・リスニングを中心とした Oral English を受講している。理解度別グループの分け方は、TOEIC の点数に基づいている。上級者グループは 50 名で TOEIC 平均点が 620 点台、中級者グループは 34 名で 400 点台である。海外滞在経験の有無は今回考慮に入れていない。

## 3 分析

### 3.1 基本情報

はじめにデータから、基本情報を観察した。両グループを合わせた総分析文章数は 819 文で、中級者グループの総分析文章数は 370 文で、上級者グループは、449 文であった。文

章における最大使用単語数は 10.8 個で、最小使用単語数は 3.82 個であった。一文における平均使用単語数は 8.2 個で中級者グループでは 8 個、上級者グループでは 8.3 個であった。各グループ間の使用単語数には差がなく、両グループとも各問題の平均使用単語数にも差はなかった。

## 3.2 頻度分析

### 3.2.1 単語頻度分析

理解度グループ別に使用した単語頻度を観察した。図 1 は各問題のグループごとの単語頻度数である。分析対象としている英語の文章が問題によって異なるため、グループ間の単語頻度を示している。

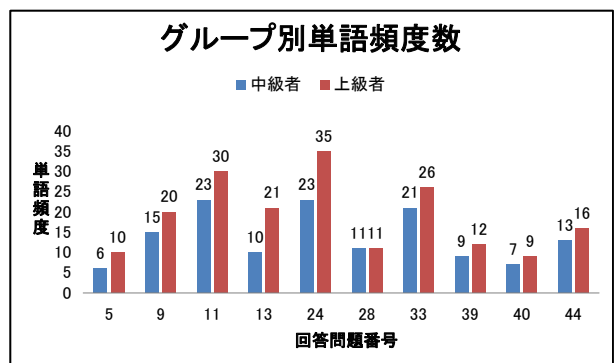


図 1. 理解度グループ別単語頻度回数

全体の傾向として、上級者グループは多様な種類の単語を用いて英文を作成していることが観測された。中級者グループからは検出できなかった単語も多くあった。例えば、問 11 において、上級者グループでしか検出されなかった単語は、mercy, sympathy, compassion, 問 24 では、dishonest, greedy, black-hearted が挙げられる。

しかし上位頻度単語（上位 3 単語）の種類は、各グループともそれぞれの問題に対して 10 問中、7 問が同じ種類の単語を使用していた。キーワードとなり得る単語は中級者グループであっても把握出来ているといえよう。例えば、問 40 の両グループの上位 3 単語は、car, convenient, useful、問 9 は young, future, man である。品詞別頻度についてはおよそ 90% が名詞で構成されている。

### 3.2.2 係り受け分析

次に、理解度グループごとの係り受け分析

を行った。一般に日本語用のテキストマイニングで行われる品詞間の関係を元に、英語の係り受けは、いわゆる修飾語あるいは主語・動詞の関係として分析した。各グループに係り受けが多く見られたのは、名詞－形容詞の繋がりであった。また、係り受けされている単語自体もグループ間によって同じものであった。上位頻度単語で検出されたものと同じ単語が多く見受けられた。

### 3.2.3 特徴分析

特徴分析とは、各英文間に付随する文法的属性で特徴的に出現する単語及びその品詞を抽出したものをいう。英文作成時に、各問題間で共通で使用されるであろう文法構造にかかわる特徴的単語に着目して分析した。英文作成に使用されるであろう対象文法項目は、完了形、that 節、関係代名詞、形式主語の it の 4 項目である。

図 2 は、完了形を用いて作文されたグループごとの比率を示している。問題番号は問 13、28、33、39 である。全体的に、上級者グループのほうが完了形を用いて英文を作成しているのが観察される。ただし、主語と動詞の一致、時制の一致等、詳細な文法構造については更なる分析が必要である。

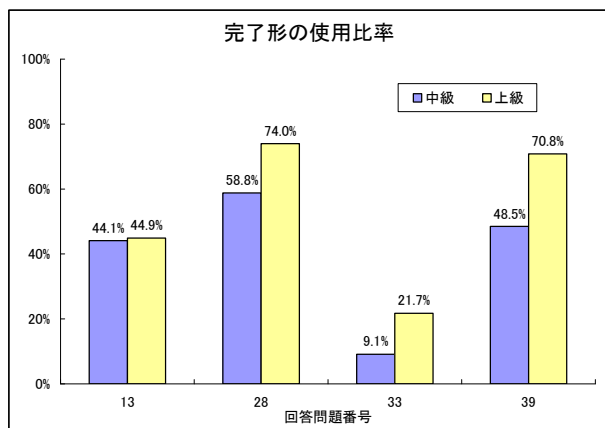


図 2. 完了形を使用している比率

図 3. は、問 44 における各グループの関係代名詞を使用した比率を示している。中級者グループは関係代名詞を 38.2%、上級者グループでは 72.3%が使用していた。中級者グループは which, who のどちらの意味にもなる that を用いて書かれている英文が多い。一方で、上級者グループは、which と who を使い

分けている人が多かった。

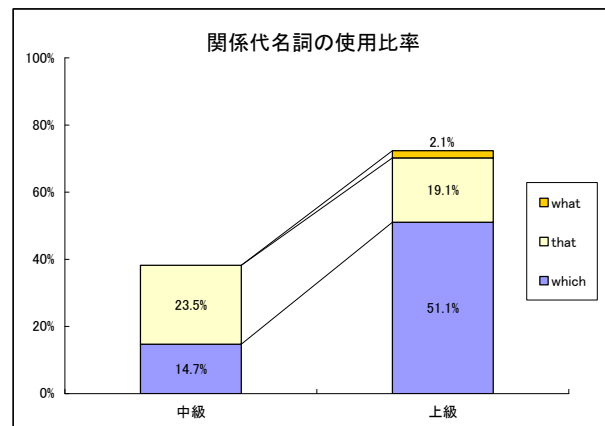


図 3. 関係代名詞使用比率

問 33 は that 節の使用比率を分析した。中級者グループは 81.8%が、上級者グループは 95.7%と両グループとも高い使用比率になっていたことが観察された。that は指示語、代名詞または関係代名詞など、他の項目としても出現するので、特に文法項目を限定して抽出して、分析してある。

問 40 は形式主語の it という文法構造を、英文作成時に反映できているかを検討した。中級者グループ内での使用比率は 44.1%であり、上級者グループでは 63.3%であった。中級者グループにとって、与えられた会話文の日本語にない単語を英語にする際に困難が生じると思われる。日本語から英語に 1 対 1 対応になっていない英語特有の構文を即座に用いて作成することは、容易ではないと推測される。

### 3.3 分析結果

各理解度別グループ間の基本分析情報については、平均使用単語数、各問題の使用単語数、品詞別出現回数において、差は見られなかった。

中級者グループは、使用すべき主要な単語は上級グループとほぼ同じものを用いていた。上級者グループは、中級者グループと比べ主要な頻出単語は同じものを使用しているが、全体的に多様な種類の単語を使って文章を作成する傾向にある。また、文法項目に着目した特徴分析では、上級者グループはなるべく文法構造を明確に表現できる単語を使用することが観測された。理解度グループ間を分けているいくつかの要因において、使用単語の

種類の多様性と文法構造を考慮に入れた単語選択が大きく影響していることが示唆された。

#### 4 まとめ

本研究では、第二言語学習者の習得過程を理解度別に分類し、テキストマイニングを用いて、英文にどのような特徴があるかを分析した。今回、分析対象として作成された英文は比較的短い文章構成であったにもかかわらず、被験者にとっては、文法知識を反映させて作成すべきものとなった。中級者グループは、語彙力や文法知識が乏しいというよりも、難しい言い回しや文法構造を「回避」[6]する傾向にあることが示唆された。

今後は、単語中心の分析のみならず、質的データの定性分析を進めていき、第二言語学習者の理解度別特徴を特定化することで、誤答分析以外の言語習得分析を提示していきたい。今回は分析対象が会話文の単文を記述したものに限定してきたが、長い文章の分析を試みることで学習者の個人習熟度を一般化して習得過程のさらなる理解度別が分析できると思われる。

#### Appendix

下線部が英文法テストで回答を問われた部分に相当する。文末に示してあるのは英文法テストでの正答率である。

- 5. Napoleon was born when? 23%
- 9. The young man will succeed in the future. 50%
- 11. The poor hate to be pitied by the rich. 49%
- 13. The number of traffic accidents has been increasing for a couple of months. 60%
- 24. Unbelievable! Such a dirty man cannot have killed himself. 64%
- 28. Many Americans have not seen the sea. 65%
- 33. Experiments have shown that the substance is cancer-causing. 67%
- 39. Mary has been boyless for eight months. 67%
- 40. It is a great convenience to have a car. 58%
- 44. I am looking for a tie to go with this jacket. 28%

#### 謝辞

本研究は順天堂大学医学部一般教育共同研究費資金による、「テキストマイニング分析

における日本人英語学習者の英語習得過程研究」の成果の一部である。データ分析に際して、文法的解釈にコメントをくださった本多由紀さんに感謝の意を表する。

#### 参考文献

- [1] ロッド・エリス, 『第二言語習得序説』, 研究社, 2003.
- [2] Chamot, A, ‘Strategies in the acquisition of English structures by a child bilingual in Spanish and French’ in Andersen(ed.) 1979.
- [3] Taylor, D, ‘Errors and explanations’, *Applied Linguistics* 7:144-166, 1986.
- [4] 大場昌也, 『学習英文法 2005 T』よこはま TLO 出版局, 2005.
- [5] 野村総合研究所, “TRUE TELLER 英語版 Version 6.0.0 Build 3 A” 2007.
- [6] Hulstijn, J. and E. Marchena. ‘Avoidance: Grammatical or semantic causes’, *Studies in Second Language Acquisition* 11:242-55, 1989.