

中高生の自由英作文における文法項目のレベル別過剰・過少使用の傾向

投野 由紀夫

東京外国語大学

y.tono@tufs.ac.jp

石井 康毅

成城大学

ishii@seiyo.ac.jp

1. はじめに

本研究の目的は、日本人中高生約1万人の自由英作文をもとに、添削前と添削後の使用文法項目の分布の変化を習熟度レベル別に比較することで、英語文法項目の過剰使用・過少使用の傾向を鳥瞰的に把握することである。近年、日本における英語教育の成果を客観的に評価するべく、文部科学省が中学・高校の全国学力調査などを大規模に実施しており、その中で4技能（聞く・話す・読む・書く）のうち、発信技能（話す・書く）において大きな課題があることが報告されている [1]。

日本人英語学習者の英語力は Education First (EF) の調査によると、2018年は88カ国中49位で、下から2番目の「低い」というカテゴリーに分類されている。これはシンガポール（3位）、香港（30位）、韓国（31位）に比して低く、一方、中国（47位）、台湾（48位）とはほぼ同位置にあった [2]。

2. 学習者の産出言語の分析

学習者言語の研究は第2言語習得研究で1960年代後半から特に中間言語 (interlanguage) の仕組みの解明という興味で実証的な研究が行われてきたが、大規模データによるコーパス分析の手法が確立してくる1990年代に学習者コーパス研究 (learner corpus research) として、独自の研究領域として発達した [3]。2000年代初頭までは学習者コーパスを用いた研究は特定の文法項目に焦点を置いたものが多かったが、ヨーロッパ言語共通参照枠 (CEFR) [4]の発表以後、CEFRレベルに基づく学習段階レベルの特徴を記述する第2言語学習者プロファイリング研究 (L2 learner profiling research) が注目を集め、ケンブリッジ大学の English Profile [5]、国内ではCEFR-Jプロジェクト [6]などがCEFRレベル別の言語材料の選定とレベ

ル判定に寄与する言語特徴の抽出を試みている。

3. CEFR 基準特性の研究

基準特性 (criterial feature) とは学習者のCEFRレベルを判断するのに有効なレベル特有の言語特徴を指す。CEFRはA1からC2までの言語能力レベルを技能ごとの能力記述子 (can-do descriptor) で規定しており、このレベル別能力記述子の順序構造が、項目応答理論 (厳密には Rasch モデル) を用いて客観的に決められている。CEFR-JもA1~B2をそれぞれ2~3段階に細分化した下位レベルの能力記述子をCEFRに倣って作成し、その妥当性検証を同様の手法で行った枠組である。基準特性となり得る具体的な言語材料をレベル別に選定する研究を総称して参照レベル記述 (reference level description: RLD) と呼ぶ。

この基準特性には主として以下の種類がある：

正の言語特徴: 語彙難度の高い単語の使用 [7]、法助動詞・関係代名詞などの特定の文法項目の発現、など

負の言語特徴: 冠詞の脱落、前置詞の誤りなど主としてエラー項目の増減がレベルを示す場合

正の言語使用分布: 法助動詞 can の能力と可能性の分布、などレベルが上がると母語話者の使用分布に近似する特徴、など [8]

本研究は、特にCEFR-Jプロジェクトの一環でCEFR-J Grammar Profile [9]の構築を行う過程において、中学・高校で指導される英語文法項目の網羅的抽出を試みた経験を活かし、英語学習者の自由英作文で出現する文法項目を網羅的に探索し、それらの添削前と添削後の使用比率の変化をもとに、CEFRレベル別に文法項目の過剰使用・過少使用の傾向分析を包括的に行うことを目的とする。

4. データ

本研究に使用するコーパスは Japanese EFL Learner Corpus (以下 JEFLC) [10]である。JEFLC は、中高生が書いた自由英作文 10,038 件が集められており、全体の総語数は 669,304 語である。全ての作文個々に添削データが作成されており、それをもとに編集距離を用いた自動エラータグ付与の研究 [11]などが行われ、現在は作文データに CEFR レベルが付与されたバージョンが存在し、CEFR-J Grammar Profile の元資料の一つとなっている。

5. 文法項目抽出

中学・高校の文法項目の抽出は、CEFR-J RLD プロジェクトで作成した抽出式に基づいて行われた [9][12]。これは、中学・高校で学習する主要文法項目をカバーした東京外国語大学佐野研究室作成の文法項目リスト [13]をベースに、English Grammar Profile [14]、British Council/EQUALS Core Inventory for General English [15]、および Threshold Level [16] に代表される機能・表現の文法項目を追加したもので、日本人が触れる文法項目をほぼ網羅したリストである。

表 1 に文法項目と、TreeTagger [17]で品詞タグを付与して一定の整形を行ったデータから各項目を抽出するのに使用した正規表現のサンプルを挙げる。

表 1: 文法項目と抽出式 (サンプル)

ID	文法項目	正規表現
26	none (不定代名詞)	¥bnone_NN_none¥b
49	比較級 and 比較級 (同語)	¥b(¥S+_ (JJR RBR)_ ¥S+) and_ CC_ and ¥1
66	過去進行 (肯定平叙文)	(was were)_ VBD_ be(?!(going_VVG_go to_TO_to gonna_VVG_gonna) ¥S+_ V_ ¥S+) ¥S+_ V.G_ ¥S+
145	助動詞+完了 (肯定平叙文)	(?!cannot¥b)¥S+_ MD_ ¥S+ have_VH_have ¥S+_ V.N ¥S+

最終的な文法項目の数は 263 項目、文種別 (肯定平叙文・否定平叙文・肯定疑問文・否定疑問文など) の異なる下位項目まで数え上げると 501 種である。

6. 使用状況の分析

過剰使用・過少使用の分析に関しては以下のように行った：

- (1) [CEFR の 4 レベル (A1~B2)] × [添削前・後]の計 8 種のサブコーパスを用意する。
- (2) 各サブコーパスにおいて、各文法項目 (5 節参照) の生起粗頻度とそれ以外の項目の粗頻度との比率を求める。
- (3) (2)の比率をもとに、(添削前) / (添削後)、(添削後) / (添削前) でオッズ比を求め、過剰使用、過少使用の傾向を見る。
- (4) 添削前の文法項目の生起粗頻度が 5 以上のものを対象にオッズ比のランキングを得る。

7. 各 CEFR レベルの過少使用の傾向

表 2 に A1 レベルの結果を示す。A1 レベルの作文の過少使用項目の多くが、A2 レベル以上で導入される文法項目である。それらの使用が少ないこと自体は教授順序の影響もあるので特に問題とは言えないが、A1 レベルの作文でも学習者が表現したい内容で必要となる複雑な文法項目が数多く存在するという事は示唆的である。

表 2: (添削後)/(添削前)のオッズ比上位項目 (A1)

ID	文法項目	オッズ比
82	助動詞+be+過去分詞	4.837
121	be able to ...	4.837
67	過去完了	4.569
170	過去分詞 (前置修飾)	4.567
62-1	現在完了 (否定平叙文)	3.219
12	These/Those + 名詞	3.018
79	will + be + 過去分詞	2.950
114	動詞 + 目 + V-ing	2.582
153	know/wonder wh-節 (whether 以外)	2.581
171	過去分詞 (後置修飾)	2.373

A2 レベルの過少使用の項目を表 3 に示す。A1 と共通する過少使用の項目 (be able to、否定の現在完了、前置修飾の過去分詞) が見られる他、特に否定や時制・相 (時制・完了相) に関する項目が A1, A2 に共通して目立つ。さらに A1, A2 とともに特徴的なのが、分詞の用法 (特に過去分詞を用いるもの) が多い点である。受動態もその一つで、

特に助動詞と用いたり **being** と用いたりする共起に問題がある。過去分詞の前置・後置修飾も添削後に補われるのが顕著な項目である。A2 独自のものとしては、関係代名詞の進んだ用法（非制限用法や **in which** のような前置詞が前に出る関係節）が拳がっており、全体に書く文章が複雑化していることが見て取れる。

表 3: (添削後)/(添削前)のオッズ比上位項目 (A2)

ID	文法項目	オッズ比
158	as soon as ...	4.049
108	being + 過去分詞	3.605
82-1	助動詞 + be + 過去分詞 (否定平叙文)	3.562
184	前置詞 + 関係代名詞	3.262
181	関係節 (非制限用法)	2.864
67-1	過去完了 (否定平叙文)	2.718
236	When ...?	2.446
121	be able to ...	2.408
62-1	現在完了 (否定平叙文)	2.320
170	過去分詞 (前置修飾)	2.264

表 4 は B1 レベルの項目である。A1, A2 と通して見られる項目としては、**be able to** と否定の現在完了がある。これらは B レベルになると克服して欲しい文法項目であるが、顕著な過少使用傾向を維持している。

表 4: (添削後)/(添削前)のオッズ比上位項目 (B1)

ID	文法項目	オッズ比
148	There + 助動詞 + be ...	4.537
184	前置詞 + 関係代名詞	4.125
216	仮定法過去完了	3.315
158	as soon as ...	3.252
181	関係節 (非制限用法)	3.081
108	being + 過去分詞	2.792
211	feel/hear/see + 目 + V-ing	2.368
87	get + 過去分詞	2.326
62-1	現在完了 (否定平叙文)	2.244
121	be able to ...	2.235

B1 独自の項目としては、仮定法過去完了、**feel/hear/see** + 目的語 + **V-ing** の知覚動詞構文、**get** + 過去分詞などが見られた。これらは A2 レベルで導入される文法項目であり、B1 レベル程度のテキストで表現する内容も複雑になってきて初めて使用機会が増えていることが伺えるが、実際は

過少使用の状況になっている。

8. 過剰使用の分析

紙数の関係で過剰使用の項目に関しては、(添削前) / (添削後) の比率を比較した A1 レベルの状況だけを表 5 に報告する。なお、過剰使用の文法項目は添削前の生起粗頻度が 5 以上では傾向がつかみにくかったため、頻度 10 以上の項目を対象としている。

表 5: (添削前)/(添削後)のオッズ比上位項目 (A1)

ID	文法項目	オッズ比
3	he/she is ...	3.932
4	we are ...	3.676
117-1	一般動詞命令文 (否定)	2.849
10	It is ...	2.188
58	be 動詞現在形	2.157
44	too ... to ...	1.959
10-1	It is not ...	1.733
146-1	there is/are/etc. not ...	1.628
59	一般動詞現在形 (三人称単数以外)	1.616
58-1	be 動詞現在形 (否定平叙文)	1.613

A1 レベルの過剰使用の傾向で顕著なのは、簡潔な定型表現が多いことである。学習者は導入期にこのような定型表現を活用して限られたストックの語彙・表現で対応しようとしていることがわかる。またこのリストを見ると、**be** 動詞が関連した構造が多い点が注目される。現在の中学英語教科書は、たいていの場合 **be** 動詞から導入をするものが主流であり、一般動詞は遅れて出現する。それによる **be** 動詞と主要な代名詞の主語 (**it, he, she, we**) などへの慣れが過剰使用を引き起こしているのではないか。

もう 1 つの大きな原因は、日本語と英語の文構造の違いである。日本語は「像は鼻が長い」でも知られるように主題話題 (**topic-comment**) 構造の主題優勢言語なのに対して、英語は主語優勢言語である。日本人は「A は B である」という文を作ろうとすると、主題話題として解釈しやすく、そのため「A is B」という発想から **be** 動詞使用が過剰傾向になると思われる。この点は添削前後の例文をパラレル・コーパスとして文対応を抜き出すことにより、添削前に **be** 動詞の用例が対応する添削後の英文では一般動詞で表現されているこ

とからもわかる。実例は口頭発表で紹介する。

9. 考察

本研究では中高生 1 万人の英作文を集めた JEFLL コーパスを用いて、添削前と添削後の作文に見られる中高の文法項目の網羅的な抽出を試み、その使用比率を比較してみた。

結果を見ると、そこにはいくつか興味深い傾向が観察された。第一に、時制と相、特に完了相（現在完了、過去完了）に関して、習得上の課題がある。ここで着目すべきは、過去分詞が完了相以外に助動詞+be+過去分詞、being+過去分詞、get+過去分詞のような態（voice）に関わる用法にも関連していることで、1つの形態が多様な機能を呈する文法項目において、習得が極めて困難で、結果的に回避（avoidance）の現象が顕著に見られると推察できる。

第二に、2つ以上の節を結びつける構造が特に過少使用される点である。A1 では間接疑問文（know/wonder+wh 節）、A2 では関係節（前置詞+関係代名詞、非制限用法）、B1 では関係節以外に仮定法過去完了、知覚構文など、従属節として埋め込み文になる複雑化が見られる。

これらの過少使用の実態が明らかになったことで、文法指導にも新しい知見が得られる。すなわち、単純に1つ1つの新出文法項目を提示・練習するだけでは不十分で、既習の文法項目との関連や組み合わせ（例：未来時制と受動態、進行相と受動態、など）を使いこなせるようになる指導が必要である。また、今回、過去分詞に関連する文法項目が多かったのは、過去分詞が時制、態、名詞の前置・後置修飾、主語の補部の位置にも来る、など極めて複雑多岐にわたる用法を持つからで、そのような形式と機能に関する習熟が特に B レベルになると必要になることが示唆された。

参考文献

- [1] 文部科学省 (2018). 「平成 29 年度英語力調査結果（中学 3 年生・高校 3 年生）」
- [2] Education First. (2018). 「EF 英語能力指数（Education First English Proficiency Index: EF EPI）」
- [3] Granger, S. (1998). *Learner English on Computer*. Harlow: Addyson Wesley Longman.
- [4] Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe.

[5] Hawkins, J. and Filipovic, L. (2012). *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. English Profile Studies 1. Cambridge: Cambridge University Press.

[6] 投野由紀夫（編）(2013). 『CAN-DO リスト作成・活用 英語到達度指標 CEFR-J ガイドブック』東京:大修館書店.

[7] Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 Vocabulary. *English Profile Journal*, 3, e1.

<https://doi.org/10.1017/S2041536212000013s>

[8] Harrison, J. and Barker, F. (2015). *English Profile in Practice*. English Profile Studies 5. Cambridge: Cambridge University Press.

[9] 投野由紀夫・石井康毅 (2015). 「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 回年次大会発表論文集』, pp. 884-887.

[10] 投野由紀夫（編著）(2007). 『日本人中高生 1 万人の英語コーパス JEFLL Corpus: 中高生が書く英文の実態とその分析』東京:小学館.

[11] 投野由紀夫・望月源 (2012). 「編集距離を用いた英文自動エラータグ付与ツールの開発と評価」峰岸真琴 他（編）『コーパスに基づく言語学教育研究報告』9, pp.71-92.

[12] 石井康毅・投野由紀夫 (2016). 「CEFR-J Grammar Profile のための文法項目頻度調査」『言語処理学会第 22 回年次大会発表論文集』, pp. 777-780.

[13] 佐野洋（編）(2005). 『文法項目別 BNC 用例集及び文法項目集（1.0 版）』. 東京外国語大学佐野研究室.

[14] *English Grammar Profile*. 2015. <http://www.englishprofile.org/english-grammar-profile>

[15] North, B., Ortega, A. and Sheehan, S. (2010). *A Core Inventory for General English*.

[16] van Ek, J. A. and Trim, J. L. M. (1991, 1998). *Threshold 1990*. Cambridge: Cambridge University Press.

[17] Schmid, H. (1994). Probabilistic Part-of-speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, pp. 45-49.