

「論文」

エラータグ付き学習者コーパスを用いた 英語ライティングの分析： 中学・高等学校の学年進行に伴う特徴

阿部真理子

Abstract

This study investigates developmental patterns in the written English of Japanese EFL learners from the 7th to 12th grade by analyzing a corpus of approximately 30,000 words. The corpus was annotated with 37 types of error tags to examine error patterns and changes across different parts of speech and error type. The analysis revealed some findings. While overall error rates generally decreased with school year progression, there was a temporary increase in the ninth grade, third year of junior high school. This increase suggests that students at this stage attempt to use more complex linguistic features or newly learned grammatical structures. Regarding syntactic development, the analysis showed a clear development from simple to complex sentences, with senior high school students increasingly using more sophisticated structures involving conjunctions. These results provide insights into the characteristics of Japanese learners' linguistic development across different parts of speech and demonstrate qualitative changes in language use as students' progress through school years.

1. 研究背景

学習者言語のエラー分析は、1970年代の第二言語習得研究において、中心的な研究手法として位置づけられていた。学習者の誤用を分析することで、言語習得のプロセスや学習者特有の困難点を明らかにすることができると想定されていた。Corder (1967) は学習者のエラーが単なる間違いではなく、言語発達のプロセスを示す重要な情報を提供していると主張した。当時はデータの収集や分析手法に限界があり、大規模かつ多角的な検証は困難であったが、1990

年代には情報処理技術の急速な進歩により、大量の電子化テキストを効率的に処理することが可能になった。これを背景に、学習者コーパス研究 (Learner Corpus Research: LCR) という新たな研究分野が発展し始めた。LCR は、従来の研究では扱うことのできなかった規模の産出データ (学習者の書き言葉や話し言葉) を収集・電子化し、それらを計量的に分析することを可能にした。そして、このアプローチは、第二言語習得研究における仮説を実証的に検証するための新たな基盤となった。

国内における LCR には、日本語を母語とする英語学習者の書き言葉データを用いて形態素の習得順序研究を検証した Tono (2006) がある。研究の結果、日本語を母語とする英語学習者は冠詞の習得に困難を感じる一方、所有格の “s” の習得は初期段階から比較的容易であることが明らかになった。Izumi and Isahara (2004) も、日本語を母語とする英語学習者の話し言葉データを用いて形態素の習得順序を検証し、Tono (2006) と同じ傾向が得られることを確認した。また形態素の習得順序には母語が影響を与える点や、特に冠詞の習得に困難が見られるという知見は、Murakami and Alexopoulou (2016) において示されている。

このような形態素の習得順序研究に加えて、LCR は学習者言語の特徴をより包括的に理解するために様々な言語項目に焦点を当ててようになった。コーパス言語学的手法を用いて、学習者が言語発達の特定の段階において、どのような文法構造を使用しているのか、過剰使用 (overuse) や過少使用 (underuse) を含めて調査された (e.g., Gilquin et al., 2008; Götz, 2015; Granger, 1998; Granger et al., 2013; Ishikawa, 2013)。また、学習者言語の発達段階を特定するための分析も進められ、異なる習熟レベルの学習者を区別するための言語項目が特定される研究が行われた。例えば、会話の流れを促進する表現 (Hasselgren, 2001)、語彙指標 (Crossley et al., 2010a, 2010b, 2011)、品詞連鎖 (Tono, 2000)、品詞 (小林, 2007b; Marsden & David, 2008)、高頻度語 (小林, 2007a)、現在時制の動詞、話者の評価や態度を表す副詞、機能語 (小林, 2010) などが挙げられる。

これらの研究は、従来の小規模な分析では捉えられなかった学習者言語の特徴を明らかにしてきた。コーパスデータを活用する主な利点は、言語処理技術に基づいた客観的な分析を行うことで、研究結果の信頼性が向上する点にある (MacWhinney, 2000)。一般公開されたデータを利用することにより、研究の再現性が確保され、研究者間でデータを共有できることも強みの一つである。さらにはコーパスデータを用いた量的研究と質的研究を組み合わせることで、よ

り一般化可能性の高い知見を得ることができる (e.g., Myles & Mitchell, 2004)。一方、様々な言語項目の分析に加えて、学習者言語に対する理解を深めるためには、学習者が産出する言語に含まれるエラーの特徴を詳細に分析する必要があると言える。

先行研究では、日本語を母語とする学習者の習熟度情報付き英語ライティングコーパスにエラー情報を付与し、学習者言語をエラーの観点から多角的に分析する研究が実施されている (Abe, 2007a, 2007b)。これにより、習熟度に応じて学習者言語のエラーパターンがどのように変化するのが明らかになった。例えば、初級学習者は動詞関連のエラー (例：主語と動詞の一致) を、上級学習者は名詞関連のエラー (例：名詞の語彙選択) を多く産出する傾向があることが分かった。冠詞の正用率は、習熟度の発達とともに向上するものの他の言語項目と比較すると、日本語を母語とする英語学習者にとって、長期的に習得が困難な項目であることが示された。一方、動詞に関連する前置詞、主語と動詞の一致、アスペクト、名詞の屈折などのエラーは、学習が進むにつれて減少することが示された。

さらに阿部 (2007c) は、日本語を母語とする英語学習者の英作文における誤用を、品詞、三つのエラータイプ (誤形成、脱落、冗長)、習熟度の観点から分析した。第一に、誤形成エラーは正用率の高い品詞 (例：代名詞、形容詞、副詞、名詞、動詞) と関連し、習熟度の向上とともに頻度が減少することが示された。第二に、脱落エラーは学習者が徐々に正しく使用できるようになる品詞 (例：前置詞の補部や冠詞) と強い関連があり、変化のパターンは複雑であることが報告された。第三に、冗長エラーは主に接続詞と関連し、学習段階を通じて大きな変動は見られないが、習熟度が向上するにつれて徐々に減少する傾向が観察された。このようなカテゴリーごとの検証により、一部のエラーは習熟度の向上とともに減少する傾向があること、そして各エラータイプが習熟度において独特の変化パターンを示すことが明らかになった。特に、誤形成エラーの減少率、脱落エラーの複雑な変化パターン、および冗長エラーの緩やかな減少傾向は、学習者の言語発達段階を反映している。これらの特徴と習熟度との関係は、エラー分析が英語学習者の習熟度を客観的に区別し、発達段階を特定するための指標として機能することを示唆している。

近年では、学習者の言語使用におけるエラーを高度な統計処理技術を用いながら分析することで、どの言語項目がレベル判定に寄与するのかを明らかにする研究が進んでいる (石井・近藤, 2015; Kobayashi, 2014; Kobayashi & Abe,

2016)。エラーの頻度と学習者の習熟度の情報を利用したこれらの先行研究では、冠詞や動詞の語彙選択、前置詞の語彙選択などがレベル評価に有効な項目として挙げられている。また Mizumoto and Watari (2023) は、Cambridge Learner Corpus First Certificate in English (CLC FCE) データセットを用いて、日本語を母語とする英語学習者に特有なエラーを分析し、限定詞、前置詞、名詞の三つに注目すべきであると指摘している。このような研究は、教育者や教材開発者、テスト開発者に貴重な情報を提供することができるだけでなく、自動採点システムの構築にあたり、より精度の高い評価を行うための基礎的な資料を提供することができる。

本研究では、日本語を母語とする中高生の英作文における品詞別・誤用タイプ別のエラー分析(阿部, 2007c)を発展させ、学年進行に伴い英語ライティングがどのように変化するのか、そのプロセスを記述することを目的とする。阿部(2007c)ではエラー率の量的分析を中心としていたのに対し、本研究では文構造の複雑さや文法機能の変化に焦点を当て、質的側面から詳細に検証する。また、日本語使用の方略的な側面に注目し、これを単なるエラーではなく学習者言語の発達における指標の一つとして捉え直した。さらに、習得困難度を四つの枠組み(規則の内部構造の複雑さ、文法的機能の伝達の有無、母語との類似性、既習知識との関係)から解釈することで、中高生の英作文におけるエラーの特徴について、解釈を試みている。以上の背景を踏まえ、本研究では三つの観点から日本語を母語とする中高生の英語ライティングの実態解明を目指す。

1. 文構造の複雑さや文法機能の使用は、学年進行に伴いどのように変化するのか。
2. 日本語使用の方略的な側面はどのように変化するのか。
3. 日本語を母語とする英語学習者に特有の課題(冠詞、前置詞など)の発達過程には、どのような特徴があるのか。

2. エラータグ付き学習者コーパスの構築

2.1 分析データ

近年のAI技術の飛躍的な進展により、学習者エラータグの自動付与が現実のものとなりつつある。この技術は、タグ付与に要する人的・時間的コストを

大幅に削減することが期待されている。例えば、Cambridge University Press から、英語学習者の誤用とその誤用を英語母語話者が訂正した情報が対になったデータが公開されており (Nicholls et al., 2024), 学習者の原文と訂正文のデータから、誤用パターンを学習することが可能となっている。このような技術的進展を見据え、本研究では手動によって構築されたエラータグ付き学習者コーパスの分析を通じて、自動エラータグ付与の研究に資する知見の提供を目指す。

英語学習者のデータとして、Japanese EFL Learner (JEFL) コーパスを利用した。JEFL コーパスは、日本語を母語とする英語学習者の作文データを大規模に収集したものである (投野, 2007)。本研究では、このコーパスから中高一貫校の学習者による英作文データを抽出し、エラータグを付与したサブ・コーパスを構築した。英作文のトピックは、JEFL コーパスに収録されている5つのうち、「学園祭」に限定した。これはトピックによる言語使用の変動を最小限に抑えるためである。収集された英作文は、準備時間なし・参考資料なしという条件で20分間という制限時間内に書かれたものである。JEFL コーパスの作文タスクでは、必要に応じて日本語の使用が認められており、日本語で書かれた箇所は <jp></jp> のタグで囲まれている。これにより、英語力の制約によって特定の表現が避けられるのを防ぎ、より自然な学習者言語データを収集することが可能となっている。サブ・コーパスの構築には、中学1年から高校3年までの各学年から約5,000語 (誤差5語以内) を無作為に抽出し、合計30,000語規模のデータセットを作成した (表1参照)。

表 1. サブ・コーパスのサイズ

学年	人数	総語数	1人あたりの 平均語数	総語数 標準偏差	総文数	1文あたり の単語数	日本語の 使用頻度
中学1年	104	4,994	48.02	13.52	722	6.92	782
中学2年	77	5,004	64.99	24.62	643	7.78	354
中学3年	87	5,000	57.47	22.07	623	8.03	364
高校1年	46	4,997	108.63	32.48	566	8.83	232
高校2年	53	5,000	94.34	60.80	541	9.24	244
高校3年	55	5,005	91.00	38.43	484	10.34	143
合計	422	30,000	71.09	40.19	3,579	8.38	2,119

2.2 エラータグ付与の方法

エラータグ付き学習者コーパスを構築するにあたり、本研究では汎用性を重視した。その汎用性を確保するため、National Institute of Information and Communications Technology Japanese Learner English (NICT JLE) コーパスで採用されているエラータグガイドライン (和泉 他, 2004) を基盤とし、表 1 に示されたデータに対して表 2 のエラータグを付与した。ただし、中高生の英作文に特化したタグ体系とするため、以下の修正を加えた。第一に、前置詞の分類基準を変更した。NICT JLE コーパスでは「前置詞」(prp_lxc1) と「従属前置詞」(prp_lxc2) の区別があるが、本研究では「一般動詞以外の品詞と共起する前置詞」(prp_lxc1) と「一般動詞と共起する前置詞」(prp_lxc2) に分類した。第二に、スペリングエラーは頻度が高いため、その扱いを限定した。辞書に記載のない動詞の活用形 (例：“make” の過去形として “maked” と記している場合) は、動詞の活用エラー (v_inf) として扱った (例：<v_inf crr=“made”>maked</v_inf>)。また、文頭の大文字と小文字の誤り、複合語のスペースの有無、本来 1 語または 2 語であるべき単語の誤り (例：basket ball) はエラーとして扱わなかった。第三に、日本語の使用に関連して、特殊なケースの処理基準を決めた。中学 1 年に多く見られた現象であるが、名詞を日本語で記述しながら英語の冠詞を付与するケースについては、分析の対象としない。そして、誤って記された単語の品詞ではなく、正しいとされる品詞に基づいてエラータグを付与した。このタグ付与の方式は、NICT JLE のエラーガイドラインに準拠したものであり、Ellis (2005) においても言及されている方式である。例えば “afraided” という誤用は、正用が “afraid” であることから形容詞のエラー (aj_lxc) としてタグ付けを行った (例：<aj_lxc crr=“afraid”>afraided</aj_lxc>)。

次に信頼性の高いタグ付与を実現するために必要である明確な基準と手順を策定した。タグ付与基準として、各エラータイプの判定基準を具体的に決定し、複数のエラーが重複する場合の優先ルールを設けた。タグ付与の信頼性を確保するため、三つの原則を設定した。第一の原則として、文法項目に焦点を当て、スタイルやディスコースレベルのエラーは対象外とした。第二の原則では、最小限の修正で文法的に正しい形に修正可能な候補を挙げることにした。第三の原則として、明確な誤りのみを対象とし、判断が分かれるケースは除外する方針とした。タグ付与の手順としては、中学校・高等学校で教育経験のある教員 1 名が、四つのステップを踏みながら実施した。第一段階として、500 語のサ

表 2. エラータグ一覧表 (和泉 他, 2004 を改変)

品詞	エラーの種類	タグ	エラーの例
名詞 [n]	1. 活用	[n_inf]	*childerens / *housewives / *peoples
	2. 単複	[n_num]	many *kind of
	3. 格	[n_cs]	my *friend house
	4. 可算/不可算	[n_cnt]	*a music / *musics
	5. 語彙選択	[n_lxc]	a *type (a typewriter) / *catch ball
動詞 [v]	1. 活用	[v_inf]	*sleeped
	2. 主語動詞の人称・数の不一致	[v_agr]	And many of them *was crying.
	3. 形を選択	[v_fm]	I'm already *expect to the next festival so much.
	4. 時制	[v_tns]	I *eat breakfast this morning.
	5. 相	[v_asp]	The people *weren't knowing the reality.
	6. 態	[v_vo]	We *were played very well at the festival.
	7. 定形・不定形の選択	[v_fin]	<i>Prease</i> *visiting us school.
	8. 否定形	[v_ng]	*not well
	9. 疑問形	[v_qst]	think about *how should we use the room
	10. 補部	[v_cmp]	The advertisement makes people *to understand how dangerous drugs can be.
	11. 語彙選択	[v_lxc]	She *is black and short hair.
形容詞 [aj]	1. 活用	[aj_inf]	*more tall
	2. 原形・比較級・最上級の用法	[aj_us]	Jane is taller than Mary, but Mary is the *best basket ball player.
	3. 形容詞の数	[aj_num]	She worked hard to help the *poors.
	4. (修飾語として) 数量を表す形容詞	[aj_qnt]	There was *few traffic on the road.
	5. 補部	[aj_cmp]	It was kind of you *helping him.
	6. 語彙選択	[aj_lxc]	It is a *genius diamond.
副詞 [av]	1. 活用	[av_inf]	*more far
	2. 原形・比較級・最上級の用法	[av_us]	She came back *most quickly than me.
	3. 副詞の位置	[av_pst]	I have difficulty *often in understanding her.
	4. 語彙選択	[av_lxc]	He worked *hardly today.
前置詞 [prp]	1. 補部	[prp_cmp]	I look forward *to see you again.
	2. 語彙選択誤り 1	[prp_lxc1]	It was held *on June.
	3. 語彙選択誤り 2	[prp_lxc2]	Tom's teacher accused him *about cheating.

表 2. つづき

冠詞	1. 冠詞	[at]	It was * wonderful day.
[at]			
代名詞	1. 活用	[pn_inf]	*themselfes
	2. 数・性別の一致	[pn_agr]	It is a good book. I like *them.
	3. 格	[pn_cs]	*We school festival is very good.
	4. 語彙選択	[pn_lxc]	I was in Brass Band, *it played music in big hole.
接続詞	1. 語彙選択	[con_lxc]	Some of <jp>お客さん</jp> moved and began to cry, * they said “That was so nice!!”
[con]			
関係詞	1. 格	[rel_cs]	She is the girl *who smartphone was stolen.
	2. 語彙選択	[rel_lxc]	The cake * is on the table looks delicious.
[rel]			

ンプルデータを用いた試行を行い、タグ付与基準の確認と調整を行った。第二段階では、全データに対して1回目のエラータグ付与を実施した。第三段階では、1週間の時間的な間隔を置いた後、2回目のタグ付与を行った。最終段階では、1回目と2回目で不一致が見られた箇所を中心に確認作業を行い、エラータグを確定した。この一連の作業過程においては、タグ付与中に発見された基準の改善点を記録し、作業に反映した。なお、本研究で使用したデータは、前節で述べたように JEFLL コーパスから抽出されたもので、各学年の異なる学習者から収集された横断的 (cross-sectional) なものであり、同一の学習者を追跡した縦断的 (longitudinal) なデータではない。よって、本研究の結果は、集団としての傾向を示すものであり、個人内の発達過程を示すものではない。

これらの手順と確認プロセスにより、エラータグ付与の一貫性と再現性を向上させることを目指した。図 1 は、本研究におけるエラータグ付与の具体例を示している。次に、以下の例文に含まれる時制のエラー (Our class <v_tns crr=“made”>make</v_tns> a piramid’s inside.) を用いて、エラーの構造と意味を説明する。このエラーには、<v_tns></v_tns> (動詞の時制エラー) のタグを付与し、修正候補として crr=“made” が記されている。これは、過去に行われた学園祭について述べられているため、動詞の過去形 “made” を使用すべきところに現在形 “make” が誤用されていることを示す。なお、この例文中の “piramid” は学習者によるスペリングの誤りであるが、スペリングエラーは分析対象外としているため、エラータグは付与していない。

<s>We held a school festival <prp_lxc1 crr="on"></prp_lxc1> September 14 and 15.</s> <s>Our class <v_tns crr="made">make</v_tns> a pyramid's inside.</s> <s>It <v_tns crr="was">is</v_tns> very dark, and almost a ghost house.

注. <s>は文章の始まりを, </s>は文章の終わりを示す。

図 1. エラータグが付与されたデータの例

2.3 分析

エラー分析に加え、正用法の比率を算出するために Constituent Likelihood Automatic Word-tagging System (CLAWS) 7 の品詞タグを活用した。ただし、学習者の誤用により品詞タグが正しく付与されない場合や、国内で指導されている学校文法との違いが見られる場合には、手作業で修正を行った。分析では、まず品詞別エラー頻度を確認した。具体的には、名詞、動詞、形容詞、副詞、前置詞、冠詞、代名詞、接続詞について、エラーの総数を集計した。その上で、各品詞における使用総数（正用と誤用の合計）に対するエラー率を算出し、学年進行に伴うエラー率の変化を確認した。また、文の構造面では、単文から重文と複文への発達過程を確認するために、特に接続詞の使用パターンと文の結合方法を分析した。これらを通じて、品詞、文法といった多様な言語項目に注目し、学年ごとの学習者データを比較した。これにより、各学年に特徴的なエラーや、すべての学年に共通するエラーのパターンを確認することができた。また、各学年における日本語タグ (<jp>) の出現頻度を集計するとともに、日本語使用のパターンを分類した。

3. 結果

3.1 学年進行によるエラー率の変化

全体的なエラー率は、中学 1 年で 9.2% と最も高かった。中学 2 年では 6.9% まで減少したが、中学 3 年では 8.1% と再び上昇した。その後、高校 1 年で 7.7%、高校 2 年で 5.3% と減少傾向を示し、高校 3 年では多少上昇して 5.7% となった。この結果から、全体的な傾向としては学年が進行するにつれ、エラー率が低下するが、中学 3 年での一時的な上昇が特徴的な変化として観察された。図 2 に示されるように、この上昇はこの時期に最も高いピークを示す冠詞のエラー率 (42.1%) が影響していると考えられる。中学 3 年では、前置詞のエラー率 (21.3%) と接続詞のエラー率 (13.4%) も、それぞれ上昇している (図 2 参照)。この

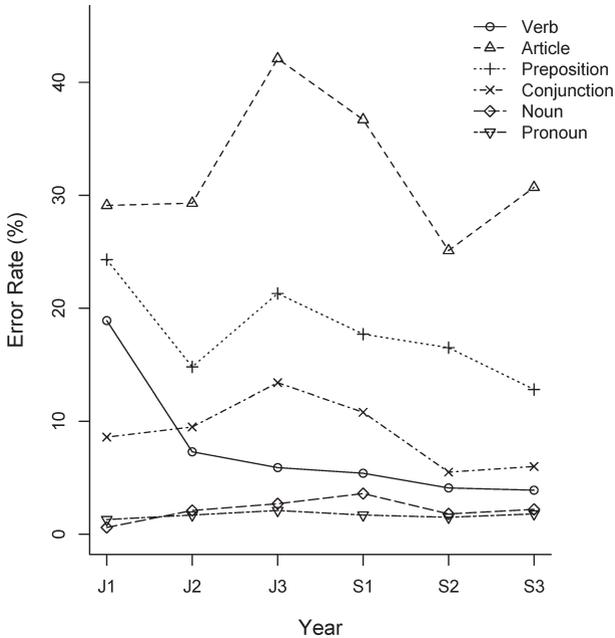


図 2. 学年進行による品詞別エラー率の変化

ような結果から、特に中学2年から中学3年への移行期において、品詞の特性に応じた指導上の配慮が必要であることが示唆される。

3.2 品詞別エラー率の特徴

品詞別のエラー率の高さによって、品詞を3つのグループに分類できる。一つ目は、高エラー率のグループ（冠詞 25.1～42.1%，前置詞 12.8～24.3%）である。二つ目は、中エラー率のグループ（接続詞 5.5～13.4%，動詞 3.9～18.9%）である。三つ目は、低エラー率のグループ（名詞 0.6～3.6%，代名詞 1.3～2.1%）である。品詞別の特徴としては、冠詞のエラー率は他の品詞と比較した場合、全学年を通じて最も高い。前置詞は、特に中学1年と中学3年において高い。接続詞は、中学3年でピークとなるが、それ以降は下降する傾向にある。一方、動詞、名詞、代名詞は低いエラー率を示しており、学年進行による大きな変化は見られない。この品詞別のエラー率の特徴は、CLC FCE データセットを利用した最近の学習者コーパス研究の結果（Mizumoto & Watari,

2023)とも一致する。

3.3 日本語使用の特徴と変化

本研究では、JEFLL コーパスにおける日本語の使用をエラーではなく、学習者言語の発達における方略的な言語使用として捉えた。母語の使用は学習者の言語使用を引き出し、より豊かな表現を可能にすると考えられる。サブ・コーパス全体の総文数 3,579 文のうち、2,119 件の日本語の使用が確認された。ただし、一文中に複数回、日本語が使用されているケースもある（例：<jp>当日は</jp> many <jp>人がきてました</jp>）。中学 1 年における日本語の使用は 782 件であったが、学年が上がるにつれて減少し、高校 3 年では 143 件となり、段階的な変化が確認された。分析の結果から、日本語の使用にも三つのパターンが特定された。第一に、文全体の日本語使用である（例：「おかしについての展示をしました」）。第二に、学園祭に特有の具体的な物の名称を示す日本語（例：「紙芝居」、「ダンボール」、「木材」、「忍者屋敷」）が、英語における語彙知識の不足を補う目的で使用されている。第三に、助詞や文末表現の日本語使用であった（例：「は」、「が」、「でした」）。そして、これらの日本語使用パターンは、学年に応じて異なる特徴を示した。例えば、中学 1 年では文全体の日本語使用が特徴的であった。中学生の時期は、基本的な名詞や動詞の日本語使用が中心であり、英文中に部分的に組み込まれていた。しかしながら特に高校 2 年では、他の学年と比べて特徴的な日本語使用のパターンが見られた。それは微妙な感情のニュアンスを示す日本語（例：「ぶっちゃけて」、「しらけてる」）の使用、および複雑な状況や動作を説明する日本語（例：「ほきょうしている」、「ピンタした」）の使用であった。高校 2 年における日本語の使用においては、感情表現が日本語使用全体の 10.7% を占め、複雑な状況説明に関する表現は 9.9% であった。さらには、高校 2 年と高校 3 年の間において質的な変化が見られた。高校 2 年では感情表現として基本的な語彙（happy, glad : 11 件）が使用されているのに対し、高校 3 年になるとより洗練された語彙（satisfied, delighted, delightful, pleased : 9 件）が使用されている。これは感情を英語で表現しようとする意欲の表れと解釈でき、それに伴い日本語使用が減少すると考えられる。また、高校 2 年と高校 3 年の間には、日本語使用の質的な違いだけでなく、量的な違いもあった。1 人あたりの日本語の使用回数は高校 2 年では 4.6 回であったが、高校 3 年では 2.6 回に減少している。高校 3 年では、日常的な感情は英語で表現し、日本語は文化祭に特有の名詞（例：“Anmitsu,”

“Yakisoba,” “Kanten”) に限定される傾向があった。このような日本語使用の変化は、学習者言語の発達を反映していると考えられる。初期段階では、学園祭の内容を伝達するための手段として日本語に依存する傾向があったが、学年の進行に伴い日本語の使用が減少し、より戦略的に日本語を使用する段階へと移行していたと言える。

4. 考察

4.1 高エラー率の品詞（冠詞、前置詞）

3.2 節で示された品詞間でのエラー率の違いは、各品詞の特性によって説明できる。特に、日本語に対応する概念がない品詞（冠詞、前置詞）は、エラー率が高い傾向があった。これらの特徴をより詳細に分析するため、この節では文法項目の習得困難度を説明する枠組みを使って解釈を試みる。まずは、3.1 節で示した中学3年での冠詞エラー率の上昇（42.1%）について触れる。この急激な上昇は、U字型発達曲線（e.g., Kellerman, 1985）の特徴と一致している。U字型発達曲線とは、学習者が新しい言語規則を取り入れる過程で一時的にエラー率が上昇し、その後熟達とともに再び減少するというパターンを指す。この現象は単に新しい文法項目への試行による結果とは解釈できない。むしろ、この時期に学習者が文章の複雑化を試みる中で冠詞の使用機会が増加し、その結果としてエラーも増加したと考えられる。

冠詞と前置詞の高いエラー率に関しては、白畑（2016）に示されている四つの観点（規則の内部構造の複雑さ、文法的機能の伝達の有無、母語との類似性、既習知識との関係）から説明できる。第一に、規則の内部構造の複雑さの観点からすると、冠詞は定冠詞・不定冠詞の選択、可算・不可算の区別など、複数の判断基準を同時に考慮する必要がある。前置詞に関しては、1つの前置詞が文脈に応じて多様な意味を持つために学習の混乱を招きやすい（例：空間 in the box, 時間 in winter, 抽象的状態 in anger, 慣用的表現 in fact）。また、語連鎖の構成語として多義性を帯びる場合（例：bring on, hold on）があるだけではなく、形容詞との組み合わせ（例：good at, interested in）も学習しなければならない。

冠詞と前置詞のエラー率の変動には、学習者の段階的な発達が影響しているとも考えられる。まず冠詞に関しては、中学1年と中学2年では基本的な使用（例：a book）に留まっている。しかし中学3年になると、既知と未知の概念（例：

Our class did a play whose name is <jp> 漂流教室 </jp>. The play is parody of TV.) や総称的なものを示す用法（例：Making <jp> 看板 </jp> is most difficult!）など、より複雑な判断を要する用法が使用されるため、エラー率の上昇が引き起こされると考えられる。前置詞に関しては、中学1年から中学2年にかけては、基本的な空間を示すための表現（例：in the classroom）が中心であるが、それ以降はより複雑な前置詞の用法も使用されるようになる。高校1年以降においてはこれらの用法に関する理解が徐々に定着するため、エラー率が低下すると考えられる。

第二に、文法的機能の伝達の観点からすると、両品詞とも具体的な意味内容よりも文法的機能としての役割を果たしている。冠詞の場合、新情報と既知情報の区別を示したり、普遍的な対象（例：the earth）であることを示したりする。前置詞に関しても、名詞句内の要素の関係性（例：the book on the desk）を示したり、動作の方向性や手段（例：I go to school by bus.）を示したりする。これらの役割は、具体的な物を表す名詞や動作を表す動詞とは異なり、文の要素の関係性を示す。そのため、個々の意味を覚えるだけではなく、文脈を通して他の語句とどのように関係づけられるかを理解する必要があり、結果的に冠詞と前置詞のエラー率が高くなると考えられる。

第三に、母語との類似性の観点からすると、冠詞は日本語に対応する文法上のカテゴリが存在しない。前置詞も日本語の助詞と一対一の対応をしていない（例：学校に行く go to school, 机の上に置く put on the desk, 3時に at 3）。第四に、既習知識との関係からすると、冠詞は規則を学習した後も文脈による判断が必要になるため、既習知識を単純に利用できない。例えば“I saw a dog yesterday. The dog was chasing a cat.”では新情報として不定冠詞を使用し、二文目では既に言及されている犬に対して定冠詞を使用する必要がある。また、“She goes to school every day.”という行為を表す場合は無冠詞でよいが、“She goes to the school to meet her teacher.”という例のように特定の学校を指す場合は定冠詞が必要である。さらには、“Water is essential for our life.”というように一般的な水を指す場合は無冠詞だが、“The water in this bottle is clean.”という例のように特定の水を指す場合は定冠詞を使う。このように文脈によって使うべき冠詞が変わるため、学習者は単純に規則を適用するだけでは不十分であり、文脈に応じた判断が求められる。また前置詞に関しては、同じ前置詞が異なる意味を持つことも多い。例えば、前置詞“on”に関しては多くの学習者にとって接触の意味を示す用法（例：put on）は馴染みが深い³が、継続の意味を示す用法（例：

move on) は新規性が高いと考えられる。このように冠詞と前置詞は全ての観点において、習得の難しさを示している。

4.2 中エラー率の品詞（接続詞）

接続詞の特徴について、文構造が複雑化する学年の進行段階に沿って説明する。第一段階（中学1年）は、単文に依存する時期である。この段階では、一文一文が独立しており、文と文のつながりが意識されていない。第二段階（中学2年）は並列構造が出現する時期である。名詞句の並列（例：We enjoyed games, food, and music.）が特徴的であり、等位接続詞を用いた単文の接続が試みられるようになる。第三段階（中学3年）はエラー率が最も高くなり、文の接続を試行錯誤する時期であると考えられる。接続詞を使用せずコマで文をつなげる傾向や、文の構造上適切ではない接続の方法（例：We make smile and my <jp>“Irassyaimase”</jp> or <jp>“Arigatougozaimashita”</jp>。）が観察された。このような例は、接続詞の使用と文構造の理解が深まっていないことを示している。

第四段階（高校1年）では依然としてエラー率が高いものの、接続詞を用いた重文と複文が使用されるようになってきている。1文あたりの接続詞の使用件数は、中学3年から高校1年にかけて多少の変化が確認された（中学3年：0.4件、高校1年：0.6件）。また“*But*,” “*Because*,” “*So*”などの接続詞を文頭に置く用法（例：Our class had a casino. But I couldn't join the casino. Because I had a tennis match that day.）が多用されており、文頭に接続詞を置く文の頻度は中学3年の19.5%から高校1年の25.1%へと増加している。接続詞を種類別に分析すると、等位接続詞の使用頻度は、中学3年の217件から高校1年の282件に増加している。特に“*and*”の使用は94件から143件へと増加し、複数の節を連結しようとする傾向が強まっていると考えられる。同様に、従属接続詞の使用も48件から63件へと増加している。さらに接続詞の使用が量的に増えるだけでなく、質的な変化が生じている。特に“*and*”を用いた単純な連結が増加する一方、“*because*”や“*if*”などの従属接続詞の使用が見られるようになり、単に文を長くするだけでなく、因果関係や条件関係といったより複雑な論理関係を表現しようとする傾向がある。

第五段階（高校2年）になると、エラー率が低下すると同時に、複文構造の発達が見られる。単文から接続詞や関係詞を用いた複文や重文へと発展する傾向がある（例：There are many games using our muscle and 競う who is the best.）。

より複雑な従属節の使用が増加し、統語的複雑さが発達しており、時間や原因を表す副詞節の使用も確認できる（例：I don't know what wasn't good, so I'm 悔しい）。また、文と文のつながりに関する意識が高まり、“However,” “But,” “So,” “Because”などの接続詞を用いて理由を説明し、論理的な文の展開を試みる例が見られる（例：We couldn't often? the *survece* to all of them. Because of the time limit.）。

第六段階（高校3年）は文構造が成熟する時期であると言える。複数の節を組み合わせた表現（例：On the first day, when only the students enjoy other classes' exhibition, I had a lot of fun with the matches.）や、名詞句を主語とする用法（例：What we had done was a food shop at one of the festival in this school.）、過去分詞による後置修飾を含む複雑な名詞句（例：The most exciting game was the one played by a girl named Momoko and a boy.）など、より高度な構文が使用されている。そして、高度な接続表現（例：as a result）の使用が可能となり、より論理的な文章構造が確立されている。このような発達は、文の長さや複雑さの向上だけではなく、洗練された思考を表現できるプロセスも示している。なお、関係詞節の使用においては、関係代名詞の完全な脱落（例：Some of my classmates sang pop songs and danced on the stage we made in the dark class room.）があったり、複雑な構文の使用に伴い、基本的なエラーが再出現したりしている。また、関係代名詞と代名詞の重複使用（例：My class did a show that we sang and danced in it.）、誤った非制限用法の関係詞節（例：Some boys including student in a lower year took part in the game, which intensity surprised me.）といったエラーが観察された。

4.3 中エラー率の品詞（動詞）

動詞と時制の使用に関しても、学年進行に合わせて発達段階を説明することができる。第一段階（中学1年）はbe動詞に依存する時期であり、二つのエラーパターンが観察された。一つは「be動詞＋動詞の原形」（例：Festival is go away.）であり、もう一つは一般動詞の代わりにbe動詞を使用するパターン（例：I am a 受付）であった。また、“It's”を過去時制の文脈でも使用するなど、時制の使い分けへの意識が低い。第二段階（中学2年）は、一般動詞の使用が拡大する時期であると同時に、動詞に続く前置詞の脱落が観察された。第三段階（中学3年）は動詞の構造が複雑化する時期であり、「一般動詞＋一般動詞」という誤った連鎖が観察された。また自動詞・他動詞の区別が曖昧であるため、前

置詞の不必要な挿入（例：discuss about the festival）も観察された。

第四段階（高校1年）は、受動態の使用を試行錯誤する時期であり、“were enjoy”から“were enjoyed”への変化が見られたが、能動態を使用すべき文脈での不適切な受動態の使用（例：That was a good festival because guests are enjoyed.）も観察された。受動態の形式的な理解は進んでいるが、適切な文脈での使用は発達途上にあることが示唆される。また、時制の使用が複雑化する様子が観察された。時制の操作に関しては発達途上にあり、完了形や過去進行形などの使用が十分に定着していない（例：I hadn't played when we practiced, but I could play on the stage.）。

第五段階（高校2年）は時制と相の使用が拡張される時期であり、完了形や仮定法（例：If I had come to this festival, I wouldn't have come a school.）など、より複雑な動詞形式の使用が試みられている。ただし、不必要な過去完了形を使用する例も観察されている（例：I bought about ten books and I had read it all day.）。また、過去の出来事を描写する際に、時制を一貫させることに苦心していることが示唆された（例：We practiced really hard for the school festival to play well. Many people came to our concert. I am so tired after it...）。特に客観的な事実と主観的な感想を同時に述べる場合に混乱しているようでもある（例：Our school festival is held on September 14th and 15th. The first day, only students can enjoy the festival. I think it was not necessary.）。これらの例は、時制と相の習得がまだ発達段階にあることを示す特徴と言える。

第六段階（高校3年）は、動詞が多様化する一方、受動態の使用が依然として課題となっており、基本的なエラーが再出現する時期であると言える。例えば、“A number of guests bought the sweets, our shop was flourish.”のような文では、“flourish”という洗練された動詞を使用しようとする過程で、本来は能動態で使うべき自動詞を誤って受動態にしたため、文法的に不適切な表現となっている。これは、より洗練された動詞を用いる際には、適切な態を選択することが難しいことを示唆している。高校3年のデータにおいて、態に関するエラーは4件検出されている。具体的には、“Our school festival hold on 14th, 15th September.”のようにbe動詞が欠落するケースや、“Our festival's symbol is changed every year.”のように能動態で表現すべきときに受動態を用いるケースなどが観察された。受動態に関するエラーは学年進行に伴い、件数が減少しているものの（高校1年：12件、高校2年：5件、高校3年：4件）、態の選択に関する課題が残っていることを示している。これらの現象は、より洗練された語彙を使

用する際に、文法的な正確性よりも意味の伝達を優先した結果、基本的な文法規則への注意力が低下したと考えられる。また語彙や表現の複雑化は、記述の内容にも影響を与えている。中学1年から2年では単純な事実描写が中心であり、時系列に沿った記述が特徴的であったが、中学3年から高校1年になるとより詳細な状況説明や因果関係の記述を試みるが多くなっている。

5. 結論

本研究は、日本語を母語とする中高生の英語ライティングにおけるエラー分析を行うことによって各学年における言語使用および発達上の特徴を明らかにした。また、各学年に特有のエラーパターンを報告した。これらの発見は、英語指導において発達段階に応じた指導方法の重要性を示唆するものでもある。結果として、具体的な意味内容を持つ言語項目（例：名詞、代名詞）は、中高6年間を通してエラー率が低かった一方、文法的な関係性を示す機能を持つ言語項目（例：冠詞、前置詞）は、規則の内部構造の複雑さ、文法的機能の伝達の有無、母語との類似性の低さ、既習知識との関係など、複数の要因に影響されることが示唆された。また学年進行に伴い、文構造の複雑化とともに使用できる表現の幅が広がっていくことも確認された。今後の課題として、作文のトピックやタスクの種類により分析結果に違いが出るのか、また異なる学校で得られたデータを用いた場合も同様の結果が得られるか、さらにどのような要因（例：母語、教科書、学習時間や方法）が学習者の英作文の発達過程に影響しているのかなどについても研究が必要となるだろう。

謝辞

本稿の執筆にあたり、貴重なご意見とご助言をいただいたすべての先生方に深く感謝申し上げます。また本研究に対し、多大なる支援をいただいたすべての方に感謝の意を表します。

参考文献

- Abe, M. (2007a). A corpus-based investigation of errors across proficiency levels in L2 spoken production. *JACET Journal*, 44, 1–14.
- Abe, M. (2007b). Grammatical errors across proficiency levels in L2 spoken and written English. *The Economic Journal of Takasaki City University of Economics*, 49, 117–129.

- 阿部真理子 (2007c). 「JEFLL コーパスに見る品詞別エラーの全体像」投野由紀夫 (編) 『日本人中高生一万人の英語コーパス “JEFLL Corpus” — 中高生が書く英文の実態とその分析—』 (pp. 146–158). 小学館.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5, 161–169.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2010a). The development of polysemy and frequency use in English second language speakers. *Language Learning*, 60(3), 573–605. <https://doi.org/10.1111/j.1467-9922.2010.00568.x>
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2010b). The development of semantic relations in second language speakers: A case for Latent Semantic Analysis. *Vigo International Journal of Applied Linguistics*, 7, 55–74.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243–263. <https://doi.org/10.1177/0265532211419331>
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.
- Gilquin, G., Papp, S., & Díez-Bedmar, M. (2008). Introduction. In G. Gilquin, S. Papp & M. Díez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. vii–xi). Rodopi.
- Götz, S. (2015). Tense and aspect errors in spoken learner English: Implications for language testing and assessment. In M. Callies & S. Götz. (Eds.), *Learner corpora in language testing and assessment* (pp. 191–216). Benjamins.
- Granger, S. (1998). The computerized learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3–18). Addison Wesley Longman.
- Granger, S., Gilquin, G., & Meunier, F. (Eds.). (2013). Twenty years of learner corpus research: Looking back, moving ahead. *Proceedings of the first learner corpus research conference (LCR 2011)*.
- Hasselgren, A. (2002). Learner corpora and language testing: Small words as markers of learner fluency. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 143–173). Benjamins.
- Hawkins, J. A., & Filipović, L. (2012). *Criterial features in L2 English: Specifying the reference levels of the Common European Framework*. Cambridge University Press.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world* (Vol. 1, pp. 91–118). School of Language and Communication, Kobe University.

- 石井雄隆・近藤悠介 (2015). 「文書分類の手法と一般化線形モデルを用いた英語ライティングにおける文法的誤りの影響」『人文科学とコンピュータシンポジウム』, 71-76.
- Izumi, E., & Isahara, H. (2004). Investigation into language learners' acquisition order based on an error analysis of a learner corpus. *Proceedings of the Interactive Workshop on Language e-Learning (IWLeL 2004)* (pp. 63-71).
- 和泉絵美・内元清貴・井佐原均 (編著) (2004). 『日本人1200人の英語スピーキングコーパス』アルク.
- Kellerman, E. (1985). If at first you do succeed. In S. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 345-353). Newbury House.
- 小林雄一郎 (2007a). 「The NICT JLE corpus と語彙研究」『英文学誌』, 49, 17-29.
- 小林雄一郎 (2007b). 「The NICT JLE Corpus における発達指標の研究—コレスポネンス分析によるタグ頻度解析」『言語処理学会第13回年次大会発表論文集』, 486-489.
- 小林雄一郎 (2010). 「コレスポネンス分析：データ間の構造を整理する」石川慎一郎・前田忠彦・山崎誠 (編) 『言語研究のための統計入門』 (pp. 245-264). くろしお出版.
- Kobayashi, Y. (2014). Computer-aided error analysis of L2 spoken English: A data mining approach. *Proceedings of the Conference on Language and Technology 2014*, 127-134.
- Kobayashi, Y., & Abe, M. (2016). Automated scoring of L2 spoken English with random forests. *Journal of Pan-Pacific Association of Applied Linguistics*, 20(1), 55-73.
- MacWhinney, B. (2000). *The CHILDS project: Tools for analyzing talk*. Erlbaum.
- Marsden, E., & David, A. (2008). Vocabulary use during conversation: A cross-sectional study of development from year 9 to year 13 among learners of Spanish and French. *Language Learning Journal*, 3(2), 181-198. <https://doi.org/10.1080/09571730802390031>
- Mizumoto, A., & Watari, Y. (2023). Identifying key grammatical errors of Japanese English as a foreign language learners in a learner corpus: Toward focused grammar instruction with data-driven learning. *Asia Pacific Journal of Corpus Research*, 4(1), 25-42. <https://doi.org/10.22925/apjcr.2023.4.1.25>
- Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition*, 38(3), 365-401. <https://doi.org/10.1017/S0272263115000352>
- Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4), 373-391. <https://doi.org/10.1191/0267658305sr252oa>
- Myles, F., & Mitchell, R. (2004). Using information technology to support empirical SLA research. *Journal of Applied Linguistics*, 1(2), 169-196.
- Nicholls, D., Caines, A., & Buttery, P. (2024). *The write & improve corpus 2024: Error-annotated and CEFR-labelled essays by learners of English*. Apollo - University of Cambridge Repository. <https://doi.org/10.17863/CAM.112997>

- Pendar, N., & Chapelle, C. (2008). Investigating the promise of learner corpora: Methodological issues. *CALICO Journal*, 25(2), 189–206.
- 白畑知彦 (2016). 「明示的文法指導, 明示的フィードバックが効果的な文法項目とそうでない文法項目一項目別に教え方を変えてみよう」 外国語教育メディア学会関西支部 2016 年度春季研究大会.
- Tono, Y. (2000). A corpus-based analysis of interlanguage development: Analyzing part-of-speech tag sequences of EFL learner corpora. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *PALC'99: Practical applications in language corpora* (pp. 323–340). Peter Lang.
- Tono, Y. (2006). L2 acquisition of grammatical morphemes. In T. McEnery, R. Xiao & Y. Tono, *Corpus-based language studies: An advanced resource book* (pp. 247–263). Routledge.
- 投野由紀夫 (編著) (2007). 『日本人中高生 1 万人の英語コーパス—中高生が書く英文の実態とその分析』 小学館.

(阿部真理子 岡山大学 Email: abema@okayama-u.ac.jp)