

日本語を母語とする大学生英語学習者の
英語ライティングにみられる傾向の分析
— 語彙と共起表現に関する一考察

阿久津 純恵・青木 敦子

An Analysis of Japanese University Students' Written English:
Their Productive Vocabulary Knowledge and Its Influence on Preferred Phrases

AKUTSU Sumie, AOKI Atsuko

桜美林大学

桜美林論考『言語文化研究』第5号 2014年3月

The Journal of J. F. Oberlin University

Studies in Language and Culture, The Fifth Issue, March 2014

キーワード：英語ライティング、語彙、共起表現、コーパス、英語教育

要 約

本研究では、大学1・2年生の英語ライティングから学習者コーパスを構築し、日本語を母語とする英語学習者が頼りがちな英語表現の傾向を分析した。その結果、学生の産出語彙レベルの低さが、特定の共起表現の多用をもたらしていることが明らかとなった。本稿では、大学英語教育においてより効果的に英語ライティング力を高めるには、このような日本語話者の書く英語の特徴についての学習者の認識、さらに国際社会で通用する「英語らしい」英語とのギャップに対する意識を高める指導が必要であることを論じた。

Abstract

The purpose of this study is to investigate an approach to improve Japanese university students' writing ability in English based on the analysis of their writings. A small sized corpus was constructed using the writing data of Japanese university students, all of which was collected during the required English classes at universities in Tokyo. An analysis was made on the vocabulary levels and the phrases preferred or repeatedly used by the learners. The paper discusses that poor vocabulary productive knowledge is one of the major causes of students being apt to rely on particular phrases. It also suggests the necessity to raise students' language awareness of their own writings as well as more active usage of the receptive vocabulary knowledge.

1. はじめに

日本語を母語とする英語学習者の書く英語には、どのような誤りや傾向が見られるのか、またより効果的に英語ライティング力を高めていくためには、どのような教授法が適当であるのかについては、第2言語習得や英語教育において多くの研究が蓄積されている。今日の大学英語教育においては、グローバルに通用する英語力、より「英語らしい」英語を用いる能力の育成が、強く要請されているが、その一方で、多くの大学生の英語学習体験は、受験や資格試験勉強のために単語を暗記すること、つまり受容語彙の増強を中心とする傾向が強い。単語集にある単語を暗記することに多くの時間を費やしがちな日本人学習者の書く英語と「英語らしい」英語との間にあるギャップを学習者に意識させる指導を、英語ライティング指導における一つの効果的なアプローチとして提示したいというのが、本研究の目的である。

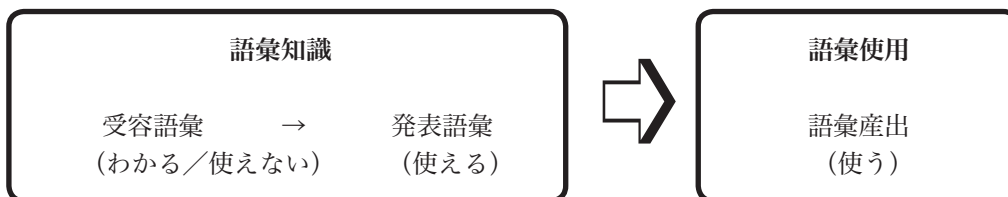
日本語話者の英語に関しては、コモンエラー分析、中間言語分析、L1の影響などの視点から、その特徴が論じられてきているが、主に、母語からの干渉としてさまざまな誤用が論じられることが多い。本研究では、日本語話者の英語ライティングにおける特徴を考察の対象とし、大学1・2年生の学習者コーパスを構築し分析を行った。誤用ではないが、日本語話者が頼りがちな英語表現の傾向を探り、何が「英語らしい」英語の獲得の妨げになっているのか、特にN-gram分析を中心に、学習者の発表語彙能力と英語表現の関係性を考察した。本稿では、学生の発表語彙のレベルの低さが、特定の表現の多用をもたらし、さらに、学生の書く英語と「英語らしい」英語を乖離させている一因となっていることを示したい。

2. 先行研究

語彙的な視点で言語教育を捉える重要性が強調された1980年代以降 (Schmitt, 1997)、語彙研究およびその指導法が多く提唱された。さらに、学習者の語彙力だけでなく、流暢さも育成するためには、連語知識が不可欠であることが指摘され、イディオムや句動詞などの指導が効果的であるとされ、コロケーションの知識養成が、大変重要であると論じられた (Schmitt, 2000)。

語彙学習は、学習者にとって最も重要な要素の1つであるが (Gass & Selinker, 1994)、この語彙知識は、受容語彙 (receptive vocabulary) と発表語彙 (productive vocabulary) に区別される (Nation, 2001)。受容語彙とは、前後関係がある中で、単語を認識し、意味を思い出せる語彙のことで、発表語彙とは、単語の意味を理解し、話したり書いたりする際に適切に使用することができる語彙のことである。さらに、石川 (2008) は、書き言葉の産出において、発表語彙と産出語彙を区別し、以下の図1のように位置づけ、この産出語彙の観察に学習者コーパスを用いた分析が有効であることを論じている。

図1 語彙知識と語彙産出の関係



あるグループに特徴的な表現の過剰使用や、過少使用される傾向にある表現などについては、母語話者コーパスと非母語話者コーパスの比較や、異なる母語を話すグループのコーパスを比較することで、L1の影響が考察されている (Granger & Rayson, 1998; 石川 2008)。母語話者と非母語話者、あるいは異なるL1グループの学習者コーパスを比較することによって、L1の影響を測定し、その要因を考察する分析の手法自体もまた、検証の対象となっている。さらに、テキストのジャンルの違いや、母語話者および学習者の属性の違いが、どのような影響となってテキストに現れるのか、それらの測定および分析方法について多くの研究が行われている。

3. 研究の方法

本研究においては、統一された条件の下で書かれたコーパスを作成することが目的ではなく、日本語話者の書く英語の傾向を測り、それを教授法に活かすことが目的であるため、3名の担当者によって、各大学で行われるそれぞれのライティング授業のシラバスに則って、データ収集を行った。

3.1 対象者

本研究では、東京の私立大学の1年生および2年生の学生が対象者である。全学生に、英語の統一カリキュラムを実施している大学の学生が含まれているため、学部や専攻は多岐にわたるが、対象者251名全員が、必修科目の英語クラス受講生である。

3.2 ライティング条件

ライティング課題は、各授業のシラバスに従い、授業の課題として書かれたものである。従って、ライティングの条件は統一されておらず、パラグラフやエッセイ、さらにパラグラフやエッセイの書き方を学習する前のライティング診断としてのフリーライティングなどが含まれている。また、学生が授業中に書いたものと、課題として授業外に書いたものがあり、さらに、学生が提出した手書きの原稿を第3者が電子データ化したものと、学生自身がデータファイルで提出した課題が混合している。自動翻訳を使用しないように指導した上で、辞書の使用は許可したという条件については、統一されている。

現在までのライティング課題データ収集で使用したライティングトピックのプロンプトは、以下の通りである。各大学、学部、クラスのレベルなどの条件により、プロンプトは、日本語で与えられた場合と、英語で与えられた場合がある。

- ・ 旅行
- ・ 携帯電話
- ・ 尊敬する人
- ・ Smoking
- ・ Someone you respect
- ・ Who is the person you respect the most?
- ・ These days an increasing number of children use cell phones. What is your opinion on this?
- ・ Some people are always in a hurry to go to places and get things done. Other people prefer to take their time and live life at a slower pace. Which do you prefer?
- ・ Do you agree or disagree with the following statement? “It is important for college students to have a part-time job.”
- ・ If you could change one important thing about your university, what would you change?

3.3 ライティングデータ収集

2011年度から2013年度にかけて収集された学生のライティング課題は、すべて電子データ化され、コーパスとしてまとめられた。前述したとおり、2種類のグループがあり、手書きの原稿を提出したグループと、パソコンに入力したライティングを提出したグループが混在している。結果に相違がでる可能性も想定されたため、それぞれの大学および提出の方法別にコーパスファイルを作成し、コンコダンス検索の下準備をした。手書きの原稿を提出したグループには、多くのスペルミスがあるが、電子化の段階でも修正せず、生のデータとして準備した。一方、学生自身がタイプしたものを収集したデータについては、提出する前に最低限、スペルチェックを行ってから提出するように指示した。以下が収集した英語ライティング情報である。

表1 学習者英語ライティング基本情報

学習者英語ライティング基本情報	
学生数	251
ライティング数	343
ライティング最大語数	289
ライティング最少語数	30
ライティング平均語数 (SD)	125 (47.1)
総語数	43,763

3. 4 学習者ライティングコーパス基本情報

コーパスは、Lancaster UniversityのUniversity Center for Computer Corpus Research on Language (UCREL) が提供しているCLAWS (the Constituent Likelihood Automatic Word-tagging System) を用いてPOSタグ付けし (CLAWS7 Tagset)、エクセルおよびコンコーダンスソフトウェアAntConcを用いて分析した。以下が、本稿で使用したコーパス情報である。

表2 学習者コーパス情報

学習者コーパス情報	
延べ語数 (Token)	43,763
異なり語数 (Type)	4,052
総センテンス数	3,908
1文あたりの平均語数	11.0
一語あたりの平均文字数	4.4

4. コーパス分析による使用語彙分析

本研究においては、ライティング条件が統一されていないことから、ネイティブコーパスとの比較は行わないこととし、日本人学生の英語ライティングの傾向を考察することを第一義的な目的として検証をすすめた。

4. 1 学習者コーパス語彙レベルチェック

まず、大学英語教育学会 (JACET) の基礎語彙改訂委員会により、British National Corpus コーパスと同委員会により構築された日本人学習者コーパスをもとに作成された、日本の英語学習者向けの基本語彙リスト語彙表JACET8000を用いて語彙レベルチェックを行った。今回の学習者コーパスにおける語彙のタイプ別頻度および各語彙の頻度は、それぞれ表3に示す通りである。

表3 JACET8000準拠語彙頻度表

語彙レベル	頻度 (type)	割合 (%)	頻度 (token)	割合 (%)	JACET 8000 レベル概要
JACET 1000	848	28.68	36,445	83.28	中学の英語教科書に頻出する基本単語
JACET 2000	496	16.77	2,030	4.64	高校初級レベルの単語
JACET 3000	270	9.13	876	2.00	高等学校英語教科書レベルの単語
JACET 4000	133	4.50	334	0.76	大学受験、大学一般教養の初級に相当
JACET 5000	103	3.48	287	0.66	難関大学受験、大学一般教養に相当
JACET 6000	89	3.01	213	0.49	英語を専門としない大学生やビジネスマンが目指すレベル

JACET 7000	64	2.16	146	0.33	英語専攻の大学生やビジネスマンが到達 目標とするレベル 日本人の英語学習者の一般的な単語学習 の最終到達目標
JACET 8000	52	1.76	100	0.23	
その他	902	30.50	3,332	7.61	—
合計	2957	100%	43,763	100%	—

この結果から指摘できることは、学生の使用語彙の少なさである。学生のライティングの語彙レベルを、語彙タイプ別の頻度および語彙使用頻度の点から判断すると、学生の使用語彙が中学英語教科書で学習する基本単語に偏っていることがわかる。使用語彙の80%以上がJACET1000レベルであり、大学一般教養レベルとされているJACET4000、JACET5000レベルの語彙使用は、それぞれ1%にも満たないことが明らかとなった。

4.2 POS分析

さらに、コンコーダンスを使用したPOS分析によると、品詞ごとの頻度について、以下のような結果がでた。

表4 学習者の品詞使用とその頻度

品詞	語数	比率 (%)
名詞	10,135	23
動詞	9,216	23
代名詞	4,731	11
前置詞	3,746	9
副詞	3,211	7
形容詞	3,045	7
冠詞	2,178	5
接続詞	2,503	6
その他	4,803	11

Granger & Rayson (1998) は、非母語話者の学習者コーパス International Corpus of Learner English (ICLE) と英語母語話者コーパス Louvain Corpus of Native English Essays (LOCNESS) を比較し、前置詞・形容詞・動詞の使用には大きな差はみられないが、英語学習者データには、冠詞・代名詞・副詞の過剰使用がみられ、接続詞・前置詞・名詞が過小使用される傾向があると報告している。この分析結果の比率と比較すると、本稿の分析対象である学習者コーパスでは、動詞と代名詞の使用比率において相違があり、それぞれ Granger & Rayson の研究結果よりも5%程度高い比率で動詞と代名詞が使用されていることがわかった (Granger & Rayson, 1998)。次に、この要因について語彙頻度表及びN-gram分析から考察したい。

4.3 語彙頻度表

日本語話者によるIの過剰使用については、すでにさまざまな研究によって指摘されており (Ishikawa, 2009; 小林 2010)、また、定冠詞the、前置詞to、of、inについても、高頻度語彙として現れる報告がなされている (阪上 2013)。本研究の分析対象とした学習者ライティングコーパスでも、表5の語彙頻度表から、同様の結果が指摘できる。ライティング課題の際に使用したプロンプトの影響により使用頻度が高いと思われる語は、ボールド体で示した。

表5 学習者コーパス語彙

順位	頻度	語
1	1607	i
2	1281	to
3	1103	is
4	1017	the
5	1001	a
6	943	and
7	756	in
8	672	of
9	664	for
10	518	it
11	509	my
12	463	can
13	409	she
14	406	he
15	403	are
16	391	so
17	391	we
18	386	was
19	384	have
20	372	children
21	366	that
22	364	time
23	353	cell
24	348	phone
25	348	they
26	322	very
27	288	with
28	265	when
29	259	do
30	259	not

最も使用頻度の高い代名詞 *I* については、自分の意見や経験を書くトピックが提示されたことが大きな一因となっていると考えられるが、これは *I* との 2-gram、3-gram の共起表現を調べると、より顕著な傾向となって現れる。

表6 *I* との共起表現頻度 (2-gram)

順位	頻度	2-gram
1	150	i think
2	147	i respect
3	122	i was
4	107	when i
5	90	i have
6	82	so i
7	81	i want
8	69	i am
9	58	i went
10	57	i can

表7 *I* との共起表現頻度 (3-gram)

順位	頻度	3-gram
1	74	i want to
2	47	i think that
3	47	i went to
4	43	when i was
5	34	i prefer to
6	29	i respect my
7	24	i do nt
8	22	i was a
9	17	i respect him
10	17	i respect is

誤用ではない *I* の過剰使用傾向は、より平易なセンテンス構造を導き、産出語彙レベルを低下させる一因となっていると推測できる。

4.4 N-gramによる共起表現分析

ある文字列の中で、N個の文字列または単語の組み合わせが、どの程度出現するのかを調査する言語モデルである N-gram 分析を使用して、2語から4語までの共起表現を調べると、ライティング課題の際に使用したプロンプトおよびパラグラフライティング指導による影響が、より顕著にあらわれた。ライティングプロンプトの影響を受けていると思われる表現はボールド体で、パラグラフライティングのスタイルが影響して使用頻度を上げていると思われる語は、さらにイタリック体で示してある。

表8 2-gram 頻度

順位	頻度	2-gram
1	203	cell phone
2	202	it is
3	150	i think
4	147	i respect
5	134	<i>there are</i>
6	129	<i>for example</i>
7	123	cell phones

表9 3-gram 頻度

順位	頻度	3-gram
1	74	i want to
2	70	a lot of
3	70	part time job
4	57	cell phone is
5	47	i think that
6	47	i went to
7	43	when i was

8	122	i was	8	34	i prefer to
9	121	is very	9	31	a slower pace
10	118	want to	10	30	a cell phone
11	117	part time	11	29	i respect my
12	110	he is	12	29	life at a
13	107	when i	13	29	part time jobs
14	105	is a	14	28	at a slower
15	94	so i	15	28	cell phones are
16	92	in the	16	27	he is a
17	90	i have	17	27	is good for
18	90	we can	18	27	there are three
19	88	she is	19	26	time job is
20	86	of the	20	25	in a hurry
21	84	in conclusion	21	24	a part time
22	84	my mother	22	24	are three reasons
23	82	a lot	23	24	high school student
24	81	do nt	24	24	i do nt
25	81	i want	25	24	she is a
26	80	high school	26	24	there are many
27	75	went to	27	22	and so on
28	73	in a	28	22	i was a
29	73	phone is	29	22	junior high school
30	71	have a	30	22	use cell phones

Biber & Barbieri (2007) は、高頻度に現れる語の連なりを単語連鎖 (lexical bundles) とよんでいるが、Cortes (2004) は、頻度の点においても多様性の点においても、4-gramの分析が有効であると論じている。特に、表10に示した4-gramの単語連鎖においては、現れる語彙が基本レベルに偏っていることに加え、ライティング課題のプロンプトおよびパラグラフライティングのフォーマットをふまえた表現がその大部分を占めてしまい、学習者自身が意味のある単語連鎖をほとんど使用できていないことがわかる。

表10 4-gram頻度

順位	頻度	4-gram
1	26	part time job is
2	24	there are three reasons
3	21	life at a slower
4	20	a part time job

5	20	at a slower pace
6	16	i prefer to take
7	16	when i was a
8	15	cell phone is very
9	15	live life at a
10	14	time job is good
11	13	i would like to
12	12	are a lot of
13	12	<i>i have three reasons</i>
14	12	i respect is my
15	12	the person whom i
16	12	there are a lot
17	11	a high school student
18	11	i think that it
19	11	person whom i respect
20	11	think that it is
21	10	always in a hurry
22	10	children use cell phones
23	10	i respect my mother
24	10	i want to be
25	10	job is good for
26	10	to take my time
27	9	and live life at
28	9	children should not have
29	9	i prefer to live
30	9	i think children should

たとえば、“It is important for college students to have a part-time job.”というプロンプトの場合、本文中で、“(have) a part time job”という表現をそのまま使用した場合が多く、それに対してpart-timeを副詞として用いる場合にみられる“work part-time”という表現の使用頻度は高くない。ほかにも、“Other people prefer to take their time and live life at a slower pace.”という表現が含まれているプロンプトの場合、“prefer to”を含む共起表現と“live life”を含む共起表現の頻出を引き起こしていることがうかがえる。“have a part-time job”や“live life”という表現に関しては、他の動詞や表現によるパラフレーズ、例えば、work part-time、lead one’s life等が可能であるが、それらの頻度が低いことから、学生の英語表現の選択が、プロンプトの表現に大きく引きずられる傾向にあるといえる。

英文プロンプトの影響の大きさと、センテンスを書き始めるための容易な手段として多

用されていると思われる I の使用から、パラフレイズに関する指導や、I の代わりに使用することができるセンテンスの主語について、指導を工夫する必要があることがわかった。

5. まとめと今後の課題

英語習熟度 British National Corpus および Bank of English という 2 つの大規模な英語コーパスが公開された 1990 年代以降、コーパス言語学は大きな発展を遂げてきている。コーパスの設計や構築には、さまざまな新しいアイデアや工夫が施され、コーパスを用いた第 2 言語習得研究や英語教育の現場での応用についても多くの提案がなされており、言語研究者にとって、まとまった量の電子化されたテキストは、大変重要な言語資料となっている。学習者コーパスから得られる様々な知見が、教員の経験や直感を超えて、学習者の実態を明らかにし (Granger, 1998; 投野 2005)、それぞれのグループの学習者に適した指導方法を工夫することができる。小規模のコーパス分析結果であっても、教員の経験や直感からくる学習者ライティングの傾向を、コーパス分析によって、より客観的に示すことによって、学習者の英語ライティングの際の視野を広げ、ライティング能力向上に寄与することがではないかと考える。

本稿では、日本語話者の大学生学習者ライティング英語の傾向を、コーパス化したデータを分析して、調査した。先行研究との比較、対照によって、学習者ライティング英語の傾向を考察したが、今後は、執筆条件やトピックが統一されていない学習者コーパスに対応する母語話者のライティングコーパスを構築し、非母語話者と母語話者の英語ライティングを比較分析する必要がある。継続して行っているライティングデータ収集については、英文プロンプトの工夫を行い、英文のプロンプトの影響をできるだけ少なくする方法を検討する必要がある。さらに、今回の学習者コーパスは、フリーライティング・パラグラフ・エッセイライティングが混在していたため、それぞれのジャンルに求められる定型表現の影響が大きく表れた。そのことを鑑み、今後は、ジャンル別のサブコーパス化についても検討したい。また、各学生の英語レベルや語彙レベルを測定し、学習者を英語習熟度別にグループ化し、その英語ライティングを比較することによって、それぞれの英語レベルにある学生が産出するライティングの傾向を調べることができると予測できる。さらに、アンケートやインタビューによって、個人の属性をコーパスデータに付与しながら、過剰使用語、過小使用語、コモンエラーなどの要因を探っていく質的な研究を組み合わせることが今後の課題である。学習者コーパスから、見慣れている比較的容易な単語に心理的に依存している日本語話者学習者の傾向が考察できた。今後、「英語らしい」英語を書くための教材作成に、この知見を活かしたいと思う。

参考文献

- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263–286.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397–423
- Gass, S. M. & Selinker, L. (1994). *Second language acquisition: An introductory course*. Amsterdam: John Benjamins.
- Granger, S. (Ed.). (1998). *Learner English on Computer*. London: Longman.
- Granger, S. & Rayson, P. (1998). Automatic lexical profiling of learner texts. In S. Granger (Ed.), *Learner English on Computer*, London: Longman, 119-131.
- Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. *Phraseology, Corpus Linguistics and Lexicography*, 87-100.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Schmitt, N. (1997). *Vocabulary: description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- 相澤 一美・石川 慎一郎・村田 年 (編) (2005). 『大学英語教育学会基本語リストに基づく JACET8000 英単語』東京: 桐原書店
- 石川 慎一郎 (2008). 『英語コーパスと言語教育』東京: 大修館書店
- 小林 雄一郎 (2010). 「日本人学習者の英作文における人称代名詞について」『言語処理学会第16回年次大会発表論文集』1074-1077.
- 阪上 辰也 (2013). 「日本人英語学習者のエッセイに見られる共起表現の分析」『広島外国語教育研究』159-169.
- 投野 由起夫 (2005). 「教材とコーパス」『立命館言語文化研究 16 巻4号』157-168.

付記: 本研究はJSPS 科研費 23652146 の助成を受けたものです。