

大学入試英作文の語彙分析

—異なるコーパス間の比較から—

新潟県／長岡工業高等専門学校 教諭 占部 昌蔵

概要

本研究は、入試英作文の中で出題されることの多い和文英訳問題に的を絞って英文を収集し、それをコーパス化したものを分析し、その特徴を明らかにすること、および、高校生の書いた自由英作文コーパスと比較することによって、それぞれどのような相違点があるのかを探ることであった。主な分析に使用したソフトはAntConcとRangeである。主な分析結果は、次の4点である。

- (1) 入試英作文の方が、多くの種類の語を使用しているということ。
- (2) 頻度順で上位30位までの分析では、自由英作文の方が、わずかに内容語の使用が上位にきているということ。
- (3) 語彙難度分析では、大きな違いは示されなかったが、入試英作文の方が、難度の高い語をより多く使用しているということ。
- (4) 特徴語分析では、自由英作文に比べて入試英作文では what が顕著に使用されており、その品詞が関係代名詞であったということ。

最後に、データ収集の方法やタグ付けなど今後の課題を提示した。

1 はじめに

5月の連休や夏季休暇に大学生が母校の高校を訪問することはしばしばあり、教師との会話の中で、大学での英語の授業が大学入試に比べ簡単であるとこぼすのを聞くことがある。いわゆる大学入試とは、大学へ入学を希望する者への大学が行う選抜試験であり、何のために行われているのかというと、

大学入学後、その大学の教育課程に沿っていけるだけの学力を備えているかを判断するためで、その中の「英語」試験問題の目的は、受験者の英語能力がある一定の熟達度に達しているかどうかを調べるためであろう。であるなら、大学が求める学生の英語能力は大学によって違って来るであろうし、学部や学科によって違ってもおかしくはない。その結果、問題の難易度に差が出てくるのも当然である。そもそも、高等学校でどの程度の英語が教えられているのか、高等学校卒業時の生徒の英語能力がどの程度であるとかを大学入試問題作成において考慮する必要はあまりないのかもしれない。しかし、大学全入時代を迎え一部の大学では定員割れを起こしており、本来の大学入試の目的が機能しなくなっていることもあると聞く。

平成20年3月に文部科学省より告示された新小学校学習指導要領に外国語活動（英語教育）が新設され、このことに伴い、小学校と中学校の「英語教育連携」とか、小・中・高・大の「英語教育連携」という言葉がよく聞かれるようになってきた。このような現状があるのに、大学入試問題はこのままでよいのだろうか。また、高校と大学の英語教育連携とは、どのようなものを指すのであろうか。現在までに、そのような「連携」は行われてきたのであろうか。大学入試問題について、大学の教員と高等学校の教員は話す機会を持ってきたのだろうか。高等学校で教えられている英語と大学入試英語問題の間に、難易度においてどの程度の開きがあるのだろうか。これらの疑問の中で、特に、大学入試問題の「難易度」へ関心が強くなった。

そこで、本調査では、小規模ではあるが、大学入

試問題の英作文（和文英訳）問題の解答をコーパス化し、それと同時に高校生の書いた自由英作文のコーパスも作成し、これらを比較することによって、それぞれの特徴や相違点を明らかにすることにした。

2 先行研究

2.1 大学入試問題

大学入試問題については、今までにもよく議論されてきている。例えば、西本（2004）では、高校でしばしば行われている大学入試問題に沿った授業の問題点を指摘している。また、Watkins・河上・小林（1997）では、指導要領以上に高校の英語教育を支配しているとか、解答を公表していないこと、和文英訳問題においては日本語が抽象的で難しいことなどを問題に挙げている。静（2002）では、大学入試問題に関して、英文和訳問題の撤廃、英文のレベル適正化などが提言されている。静・井ノ森（2006）では、和訳問題の悪影響や入試問題全般の難しさについて議論されている。金谷（2009）では、国公立と私立の大学から18大学を選び、その英作文問題を1年度のみ調べているが、調査内容は出題形式、日本語の語数、自由英作文の制限語数などについてであって、問題の難易度についての具体的なデータは伴っていない。このように、さまざまな観点から問題提議はされてきているが、データを伴う実証的な研究というものはあまり見当たらない。

2.2 コーパスの定義

石川（2008）によると、「コーパスとは、電子化された大量の言語データベースのこと」であり、「コーパスを使った言語研究のことをコーパス言語学と呼ぶ」と定義されている。コンピュータの急速な進歩や普及により、コーパス言語学の研究は欧米を中心に、日本でも盛んに行われるようになってきている。Brown Corpus以降、英語コーパスは大規模化と多様化の一途をたどっていて、British National Corpus など数億語に達しているものや地域別コーパスなどが生まれている。そして、これらの大規模なコーパスを利用した研究は数多く行われてきている。

2.3 学習者コーパス研究

朝尾（1998）によると、学習者コーパスは「〔ある〕言語を外国語として学んでいる学習者が書いたり、話したりしたものを、コンピュータ可読な形式で大量に記録したもの」と定義されている。外国語として英語を話す人口の大きさを考えると、母語話者ではない学習者英語コーパスもあって不思議ではなく、International Corpus of Learner English（以下ICLE）などの大規模プロジェクトによるものが生まれてきている。言語習得が完了した母語話者による一般のコーパスに対して、学習者コーパスは、記録されたものの中にエラーを含んでおり、発達段階にある学習者の言語習得過程で形成される中間言語を分析することが可能となる。このことは、第二言語の学習プロセスを支配する原理を明らかにすることをその目標の1つにしているSLA研究に学習者コーパスが貢献できることを意味している。そして、ロングマン現代英英辞典などのELT辞書の開発にも貢献してきている。一方で、学習者コーパスの開発やそれを英語教育に応用した研究はまだ歴史が浅く、あまり多いとは言えないが、近年、増加傾向にある。例えば、国際的規模のものでは前述のICLE、日本国内のものでは、自由英作文をベースにしたJapanese EFL Learner Corpus（以下JEFLLCorpus）やスピーキングテストをベースにしたThe National Institute of Information and Communication Technology Japanese Learner English Corpus（以下NICT JLE Corpus）などがある。JEFLLCorpusを研究した投野（2007）においては、日本人（中高生）英語学習者に関する詳細な研究が報告されている。ただ、JEFLLCorpusにも短所があり、自由英作文のトピックが6種類しかないことや母集団に偏りがあり、レベルにおいて「高」が大半を占め「中」と「低」は全体の約18%でしかないことなどが挙げられる。

3 調査目的

以上の先行研究から、入試英作文に絞った研究は少なく、日本人高校生を対象としたコーパスを使用した研究も少ない。そこで、本調査では、入試英作文（和文英訳）から小規模英作文コーパスを構築し、語彙の頻度と傾向を調査することと高校2・3年生

の自由英作文による小規模学習者コーパスを構築し、入試英作文コーパスと比較、調査することを主たる目的とする。具体的な調査項目は次のとおりである。

- (1) 入試英作文において、どのくらいの種類の単語が使用されているのか。
- (2) 入試英作文において、その語彙難度はどの程度なのか。
- (3) 入試英作文と高校生の書いた自由英作文とを比較し、両者にどの程度相違があるのか。

4 調査方法

4.1 構築された英作文コーパス

今回の調査で構築されたコーパスは、次の2種類である。それぞれのコーパスのインデックスは次のとおりである。

(1) 入試英作文コーパス

出題元：国公立大学、私立大学

ジャンル：英作文（和文英訳）

データ数：148（校）

内容：国公立大学の2005～2008年、私立大学の2005年と2009年の入試問題から和文英訳問題を抽出。なお、国公立大学の2008年と私立大学の2005年についてはネイティブスピーカーによる修正文を含む。

(2) 学習者自由英作文コーパス

対象者：高校2・3年生（高等専門学校生も含む）

ジャンル：自由英作文

データ数：97（人）

内容：「My future」、「将来つきたい職業」などの課題を与えられた自由英作文。課題は過去の入試問題からいくつかを提示し選択して書いてもらった場合と、協力校の教師が独自で設定した場合と、検定教科書準拠問題集の課題の場合がある。英作文を書くときに、条件あり（制限時間20分、辞書使用不可）で書いてもらったのが5校分、条件なしの場合が5校分である。

本調査における学習者コーパス構築にあたっては、先行研究よりも英作文のトピックを増やすことと学校レベルのバランスを取ることを心がけた。その結果、トピックは18種類（資料1）に、レベルは

「高」3校、「中」4校、「低」3校となった。ただし、ここでいうレベルはおおまかなくくりでしかなく、英語能力のレベルを明確に示すデータはない。

4.2 分析ソフト

コンコダンサー（コーパス分析ソフトウェア）にはさまざまな種類がある。今回の分析に使用したソフトは、早稲田大学の Laurence Anthony の開発した AntConc^(注1) で、シンプルで使いやすい特徴がある。本調査では、単語の使用されている文脈、意味の追跡ができるコンコーダンス分析、どのような単語がどのくらいの頻度で出現するかをリストで表示できる単語頻度分析、他のコーパスに比べてそのコーパス中に特に頻度が高い語を抽出できる特徴語分析を行った。

語彙難度を調べるにあたっては、語彙レベル分析ソフトウェアとして有名な Range^(注2) を使用した。このソフトは、Paul Nation が開発したもので、英文を分析してその文中の語彙を West（1953）の General Service List の、基本語最上位1000語（以下 GSL1）と次位1000語（以下 GSL2）、Academic Word List（以下 AWL）とその他の4段階に区分する。その割合により書き手のレベルを推測することができる。

5 結果と考察

5.1 コーパスの総語数と異なり語数

2種類のコーパスの総語数（Tokens）と異なり語数（Types）を分析した。さらに総語数に占める異なり語数の比（Type/Token Ratio）も合わせて示したのが表1である。

■ 表1：両コーパスの総語数と異なり語数

コーパス	入試英作文	自由英作文
Tokens	7045	7090
Types	1512	1001
Type/Token Ratio	21.5	14.1

この結果から、入試英作文において、使用されている単語の種類は約1500種類ということがわかる。ただ、これは入試英作文4年分という規模での結果であって10年分であればさらに種類が増えるであろう。

う。今回の分析では、タイプ・トークン比が総語数の大きさに影響されることから、総語数の少ない入試英作文コーパスに合わせて総語数をほぼ同数にして両コーパスを比較している。この結果からは、テキストの語彙多様性の観点で見れば入試英作文の方が多様であることがわかる。ただ、自由英作文の方は課題の制限があるので語彙多様性にも制限がかり、このような差となった可能性はある。さらに、この結果を客観的に見るために、石川（2008）から引用したデータと比較してみる。なお、ここでの比較にはコーパスの規模が大きく異なるので、タイプ・トークン比ではなく、Guiraud 値を用いて比較したのが表2である。

■ 表 2：語彙多様性比較（両コーパスと日韓教科書）

コーパス	入試英作文	自由英作文	日本	韓国
Tokens	7045	7090	93260	185547
Types	1512	1001	5256	11111
Guiraud	18.0	11.9	17.2	25.8

この表からわかることは、入試英作文は語彙多様性で見れば、日本の教科書とほぼ同じレベルであるが、自由英作文の場合は、韓国や日本の教科書にはるかに及ばないということである。

5.2 コーパス別語彙頻度

両コーパスにおいて、どのような種類の単語をどのくらい使用しているかを見るために、それぞれのコーパスごとの頻度順単語リスト（上位30語）を表3に示す。

この表からわかることは、入試英作文のコーパスは上位30語の多くは機能語であるが、自由英作文の方は、内容語がわずかに多いことに気づく。自由英作文の方で english, like, think が多く使用されているが、これは英作文の課題から影響を受けたと思われる。また、because が多いのも通常意見を述べる英作文においてはどうしても理由が必要になるので一番使いやすい because が多用されたのではないかと考えられる。

■ 表 3：語彙頻度表

順位	入試英作文		自由英作文	
1	the	305	i	348
2	to	254	to	269
3	i	182	is	213
4	a	181	we	202
5	of	160	the	185
6	and	142	and	153
7	in	140	it	153
8	you	114	a	125
9	that	109	english	114
10	is	108	in	114
11	it	93	of	85
12	was	83	for	80
13	we	75	can	77
14	are	61	my	77
15	for	59	like	74
16	my	59	with	74
17	they	59	so	72
18	with	58	very	72
19	when	53	that	70
20	at	51	think	66
21	be	51	have	64
22	have	51	study	64
23	if	49	but	63
24	what	48	friends	60
25	people	45	if	60
26	not	44	because	51
27	as	43	are	50
28	on	41	be	48
29	can	38	school	47
30	or	38	they	44

5.3 語彙難度

両コーパスにおいて使用されている語彙は、どのレベルにどのくらい使用されているかを見るために、異なり語数の各区分に入る語数とその割合を表4に示す。

この表からわかることは、総語数で見ると入試英作文のコーパスも自由英作文の方もその割合が大きく異なるわけではないので、Range の分析からは入試で求められている英作文の方が自由英作文よりも語彙難度において特に高いというわけではない。しかし、異なり語数で見ると、入試英作文の方

■ 表 4：語彙難度表

	GSL1	GSL2	AWL	その他
入試英作文 (総語数)	6127 (86.9%)	344 (4.9%)	201 (2.9%)	373 (5.3%)
自由英作文 (総語数)	6328 (89.3%)	333 (4.7%)	75 (1.1%)	354 (4.9%)
入試英作文 (異語数)	933 (61.7%)	213 (14.1%)	126 (8.3%)	240 (15.9%)
自由英作文 (異語数)	664 (66.4%)	136 (13.6%)	41 (4.1%)	160 (15.9%)

(注) 表内の下段の括弧内の数字は各コーパスの総語数および異なり語数に対する割合(%)である。

GSL1：最も基本的な 1000 語

GSL2：次に基本的な 1000 語

AWL：Academic Word List

では GSL 1 のうち、933語使われているのに、自由英作文の方は、664語であり、GSL 2 でも、AWL でもその差は大きい。ということは、入試で求められる語彙と比較すると、高校生が自由英作文を書くときに使える語彙は、Range の分析で示す割合の差は小さいが、実はかなり少ないと言える。この差は詳しく調べてみると大きいことがわかる。例えば、入試英作文中の GSL2に属する213語の内、自由英作文の136語と重複しない語の中には consciously, preserve, photographer などがあり、これらの語はおそらく高校 2・3 年生なら理解可能な語だと思われるのだが、総語数約67万語の JEFLL Corpus 中にもこれらの語は存在しない。以上のことから日本の高校生が英作文を書くときに使える語彙というのは予想以上に少ないということである。

5.4 特徴語分析

入試英作文のコーパスでは頻繁に使用されながら、自由英作文の方ではあまり使用されず、その差が大きい語を上位15位まで示したのが表 5 である。

この結果、15位まですべてが中学校で学習する基本的な語であることがわかる。それゆえ、これらの語は、一般的な高校生レベルの学習者では理解可能だが自由英作文にはほとんど使えていないか、入試の英作文において特に頻繁に出題されているかのどちらかであることが推測される。そこで、1 位の what の使用例をコンコーダンス・ラインで確かめてみることにした。その結果から、品詞別にその頻度と例文を表 6 に示す。

■ 表 5：特徴語リスト

順位	Word	Keyness	頻度
1	what	34.61	48
2	his	34.44	26
3	you	31.36	114
4	just	29.14	22
5	had	27.34	26
6	as	26.61	43
7	was	25.16	83
8	the	22.66	305
9	from	20.33	32
10	of	18.88	160
11	while	18.55	14
12	would	18.40	25
13	into	15.90	12
14	an	14.84	33
15	japanese	13.96	32

■ 表 6：品詞別頻度と例文 (what を含む)

型	頻度	例文 (what を含む)
I	38	to believe what they can actually see
II	7	of time. What kind of book would you
III	2	their lives? What is real happiness?

この表からわかることは、what は型 I のように関係代名詞として使用されていることが圧倒的に多く、型 II や III の疑問形容詞や疑問代名詞として使用されることは少ないということである。このことは、入試問題の解答を公表していない現状があるので強くは言えないが、入試英作文出題者の中には関係代名詞 what を誘発する和文を入れている可能性があるということである。したがって和文英訳問題の中には、語彙で見ると比較的簡単だが高校生にとっては十分に使いこなせていない文法面の習熟度を問う問題が多く存在すると言える。

6 まとめ

本調査では、「3 調査の目的」で記したように、大学入試英作文（和文英訳問題）から小規模英作文コーパスを構築し、語彙の種類、頻度、難度やその傾向を調査すること、および高校 2・3 年生の自由英作文による小規模学習者コーパスを構築し、入試

英作文コーパスと比較, 調査することが目的であった。調査目的(1)~(3)については次の結果を見た。(1)今回構築した入試英作文コーパス中に出現する語彙の種類は約1500種類であり, 自由英作文の方は約1000種類であった。(2)入試英作文の語彙難度においては日本の教科書よりも易しいレベルにあると言える。全体として見ると, (3)高校生が自由英作文に使用する語彙の種類は予想以上に少なく, 入試英作文との差は予想よりも大きかった。

今回の調査結果から, 入試英作文で問われている語の種類は予想よりも少なく, 日本の教科書と同程度とも言える。高校生が理解できる語彙(受容語彙)は入試英作文で求められているよりも高いと思われる。しかし, その受容語彙と比較すると, 自分が運用できる語彙(産出語彙)力はまだまだ乏しく, この力をさらに向上させていく必要があると言える。

今後, より質の高い英作文を書けるようになっていくために, 少しずつでも産出語彙が豊かになるような英作文指導の工夫が必要と思われる。

今回の調査で構築した学習者コーパスは小規模であるため今回の調査結果を一般化することはできないが, 日本人高校生の書く自由英作文の特徴の一部を表していることには間違いなく, 有意義な分析結果を報告することができた。

7 今後の課題と展望

今回の調査研究を通じて, 今後の課題として下記の点を提示したい。

まず, データの収集に関して, 日本人高校生を代表する協力校数を倍増させる必要がある。先行研究においてレベルの偏りがあったので, 本調査ではバランスを取ることに成功したが, 協力校の確保には苦勞し, 結果的にはデータ数を伸ばすことはできなかった。また, 学習者のデータを詳しく調査できるならば, なお理想的である。個々の学習者の英語

力がどのくらいであることを客観的に示す指標があれば, タグを付与して, より厚みのある分析が可能になってくるであろう。

次に, エラーの取り扱いである。学習者の書いた英作文には, エラーが含まれていることが多い。今回の調査では, 語彙の種類, 頻度, 難度に焦点を当てたので, 文法的なエラーについて調べることはなかったが, 文法的なエラーにタグを付与することで, その傾向や特徴について調べることができる。タグを付与する作業は, 多大な労力が必要で, 今回の調査では見送ったが, 次回のコーパス構築においては検討したい。

また, 研究手法についてであるが, 同じ受験者が自由英作文と入試英作文の両方を行った場合でのデータを比較することも次回の調査では検討したい。

学習者コーパスの開発や研究は, 国内国外を問わず今後ますます盛んになると思われる。すでに, 辞書や教材開発ではその研究の成果が応用され, 近年多くの書物が出版されてきている。そして, 国内の書き言葉コーパスとして JEFLL Corpus は最大であり, 研究成果の一部は出版され, 公開もされていることから, 多くの英語教師の研究や指導に大いに貢献しているものと思われる。

本調査は小規模な調査ではあったが, この結果や手法が, 日本の高校生に英作文を指導する際や新たな研究を進める際において, 多くの英語教師に活用していただければ幸いである。

謝 辞

本研究の機会を与えてくださった(財)日本英語検定協会と選考委員の先生方, 特に貴重なご助言をくださった大友賢二先生に厚く御礼申し上げます。そしてデータ収集においては, 大分県立大分上野丘高等学校の麻生雄治先生, 新潟県立松代高等学校の松井市子先生をはじめ多くの先生方にご協力いただきました。心より感謝申し上げます。

注

(1) AntConc : http://www.antlab.sci.waseda.ac.jp/antconc_index.html, Version.3.2.1

(2) Range : <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>

参考文献 (*は引用文献)

- * 朝尾幸次郎.(1998).「英語教育とコーパス利用」.『英語コーパス言語学—基礎と実践—』. 233-248. 東京：研究社.
- Coxhead, A. (1998). *An academic word list. Occasional Publication, 18*. Wellington, New Zealand: School of Linguistics and Applied Language Studies, Victoria University of Wellington.
- * 石川慎一郎.(2008).『英語コーパスと言語教育学—データとしてのテキスト』. 東京：大修館書店.
- * 金谷憲.(2009).『教科書だけで大学入試は突破できる』. 東京：大修館書店.
- * 西本有逸.(2004).「英語嫌いを増やす入試対策の英語授業が増えている!」.『STEP 英語情報』11・12月号, 12-13.
旺文社.(2005).『2006年受験用全国大学入試問題正解 英語(国公立大編)』. 東京：旺文社.
旺文社.(2005).『2006年受験用全国大学入試問題正解 英語(私立大編)』. 東京：旺文社.
旺文社.(2006).『2007年受験用全国大学入試問題正解 英語(国公立大編)』. 東京：旺文社.
旺文社.(2007).『2008年受験用全国大学入試問題正解 英語(国公立大編)』. 東京：旺文社.
- 旺文社.(2008).『2009年受験用全国大学入試問題正解 英語(国公立大編)』. 東京：旺文社.
- 旺文社.(2009).『2010年受験用全国大学入試問題正解 英語(私立大編)』. 東京：旺文社.
- 齊藤俊雄・中村純作・赤野一郎.(1998).『英語コーパス言語学—基礎と実践—』. 東京：研究社.
- * 静哲人.(2002).『英語テスト作成の達人マニュアル』. 東京：大修館書店.
- * 静哲人・井ノ森高詩.(2006).「Special 対談 大学入試英語改革」.『STEP 英語情報』1・2月号, 14-19.
- * 投野由紀夫.(2007).『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—』. 東京：小学館.
- * Watkins, G.・河上道生・小林功.(1997).『これでいいのか大学入試英語(下)』. 東京：大修館書店.
- * West, M. (1953). *A general service list of English words with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London: Longman, Green and Company.

資料

資料1：自由英作文課題

- 1 「修学旅行で訪れた場所について」
- 2 「将来つきたい職業」
- 3 「スポーツと私」
- 4 「なぜ英語を勉強しなければならないか」
- 5 「最も印象に残る学校行事」
- 6 「携帯電話についての自分の意見」
- 7 「家庭や学校のルールについて」
- 8 「ボランティア活動について」
- 9 「学校で英語の学習にゲーム機を使うことの是非」
- 10 「事故で無人島に漂着して、優先的に行うことを3つ以上あげその理由を説明せよ」
- 11 「今までにあなたが経験した忘れがたい失敗について」
- 12 「My favorite sport」
- 13 「My favorite food」
- 14 「My dream」
- 15 「My favorite book or anime」
- 16 「My future」
- 17 「Alone or with friends」
- 18 「My high school memories」