

コーパスベース単語リストを使った 大学英語ライティング教科書の語彙分析

An Analysis of Vocabulary in University English Writing Textbooks Using Corpus-based Word Lists

飯島優雅*1・堀江郁美*1
Yuka Iijima、 Ikumi Horie

Email: yijima@dokkyo.ac.jp, horie@dokkyo.ac.jp

キーワード：英語アカデミックライティング、コーパスベース単語リスト、教科書分析
Keywords: English academic writing, Corpus-based word list, Textbook analysis

英語ライティング科目担当教員は、科目目標や学生のニーズと教科書内容を照らし合わせ、様々な市販教科書から最適と思われる一冊を採択することが多いが、教科書内で使われている語彙が、どの程度使用頻度の高いものかの判断は印象や勘に頼ることが多い。本稿では、著者らが開発中であるオンラインのテキスト分析システムを用いて、大学生向けの英語アカデミックライティング教科書でどのような語彙が使用されているかを調べた。英語教育研究領域で信頼度の高いものとして広く活用されている3つのコーパスベースの単語リスト（General Service List, Academic Word List, JACET 8000）をもとにサンプル教科書4冊の分析を行った結果、対象英語レベルに関係なく、教科書内のモデルライティングの約7割は高頻度の約2000語で書かれていることがわかった。また、アカデミック単語の使用は、中上級向け洋書教科書で12-15%、初級中級以下の和書教科書では8%程度であった。さらに、本研究で活用したテキスト分析システムが、教科書の選定や教材開発などの過程において、語彙に関するデータに基づく意思決定をサポートする有用なツールとなりうることを示唆された。

English writing teachers often choose a textbook from various candidates on the market that appears to best suit the course objectives and students' needs. In such a decision making process, however, due to practical and technical reasons, many teachers would perhaps rely on their intuition when identifying whether the selected textbook uses high frequency vocabulary. It would be certainly helpful if teachers could know the level of vocabulary in candidate textbooks prior to their decisions. This study explored the vocabulary used in four sample university English writing textbooks, using the online text analysis system that the authors are developing. We used three corpus-based word lists for analysis (General Service List, Academic Word List, JACET 8000), which are widely used in the research body of English language teaching and learning. The study found that approximately 70% of the examined model texts were written in the most frequent 2000 words, regardless of the textbooks' target proficiency levels. Also, the use of academic vocabulary accounted for 12-15% of the model texts in the non-domestic textbooks for intermediate and advanced learners, and about 8% in the textbooks published domestically. This study also suggests that the online text analysis system can be potentially a useful support tool for making confident evidence-based decisions on material selection and development.

*1 獨協大学 経済学部

1. はじめに

学部教育の英語カリキュラムにおいて、アカデミックライティングを導入する日本の大学が増えつつあり、パラグラフやエッセイライティング、リサーチペーパーライティングを中心に、様々なライティング教科書が市販されるようになってきた。ライティング科目担当教員は、学生のニーズや科目目標と教科書内容を照らし合わせ、最適な一冊と思われる教科書を採択するが、その選定過程において、教科書内で使われている語彙が、どの程度使用頻度の高いものか、または、学生の語彙知識に合っているかの判断は、各教員の経験に基づく印象や勘に頼ることが多いのではないだろうか。

第一著者がコーディネーターを務める、本学の全学共通カリキュラム英語部門が提供するライティング科目¹においても、推薦教科書を選定する際は、基本的な学術的英語表現の使用法が学べることと、英語部門が定める Dokkyo EGAP Can-do List (一般学術目的の英語能力記述文) のライティング 42 項目ができるだけ多く網羅され⁽¹⁾、学生も教員も使いやすい内容と構成 (テーマ、モデルライティングやスキル練習問題の質と量、評価など) になっていることが主な基準となる⁽²⁾⁽³⁾。ただし選定過程で教科書の使用語彙も精査することは、時間的にも技術的にも難しく、教員らの経験知に基づく判断に頼らざるを得ないのが現状である。また、実際に教科書を授業で使ってみると、モデルライティングや練習問題に頻度の低い語彙が使われていることが少なくなく、学生の注意が、文章構成や論展開の作法などの本来の学習項目より語彙へ向いてしまうこともある。

このような課題は、教科書開発者が想定する学習者・授業像と実際の指導現場の違いによるものであり、市販教科書を使う限り避けることが難しい。しかし、教科書の使用語彙の水準については、あらかじめ教材選定の段階で、印象や勘だけでなく、エビデンスに基づく判断をする手立てがあれば、教員がより自信を持って適切な教材選定、授業設計、補助教材作成を行うことができる。

著者らは、教員の教材研究と学生の自律学習を支援するツールとして、オンラインでテキスト内の英単語分析を行うシステムを開発中である。このシステムはデジタル化されたテキストに含まれる単語を、あらかじめ設定した辞書や単語リストに照らし合わせ、テキスト中の単語分布を瞬時に視覚的かつ数値化して示す

機能を持つ⁽⁴⁾⁽⁵⁾⁽⁶⁾。インターネット上のテキストを教材化するための支援ツールとして開発を始めたものだが、教科書のような紙媒体のテキストの分析にも応用できると考えられる。

本稿では、このシステムを用いて、大学英語ライティング教科書でどのような英単語が使用されているかを考察する。本研究の結果は、教員個人だけでなく、大学英語プログラムにおける共通教材の選定や開発の担当者にとっても、有益な資料を提供できると考える。さらに、継続的に改善を続けているこのシステムの開発の方向性と有用性についても示唆を得ることができると考えられる。

2. 研究目的

本研究の目的は、アカデミックライティングスキルの学習を目的とした大学英語ライティング教科書の、モデルライティングで使用されている単語の水準を明らかにすることである。一般的かつ学術的言語コミュニケーションの文脈で使用頻度が高い英単語、つまり、英語学習者として日本人大学生が優先的に学ぶべき英単語が、学生がまずはお手本とすべきとされている教科書のモデルライティングにどの程度使われているか、その出現率を検証する。

3. 研究方法

3.1 ライティング教科書のサンプル抽出

全学共通カリキュラム部門ライティング科目は、1年生対象科目「Academic Writing I」の推薦教科書が全11冊、2年生対象の「Academic Writing II」の推薦教科書が全15冊ある。本研究では、まずこの推薦教科書の中から複数の教員に採択されている教科書を調べ、そこから4冊(教科書A-D)をサンプルとして抽出した。詳細は表1の通りである。

表1 分析対象のライティング教科書

教科書	対象学年・科目	教科書または出版社カタログに記載の対象英語習熟度	教科書または出版社カタログに記載の使用英単語の情報	対象クラス
A (2002) 和書	1年生 Academic Writing I lab: Paragraph	Pre-intermediate to intermediate (中級前かから中級)	識字際で辞書使用は最小限で済むよう、すでに学習者が知っていると思われる語彙と文構造で書かれている。	下位クラス TOEIC 400 点未満
B (2013) 洋書	2年生 Academic Writing II lab: Essay	中級, Common European Framework (CEF): B1 ⁽⁷⁾	なし	下位クラス
C (2008) 和書	2年生 Academic Writing II lab: Essay	記載なし	なし	下位クラス
D (2011) 洋書	2年生 Academic Writing II lab: Essay	CEF: B2 Upper Intermediate - C1 Advanced (中級上から上級)	なし	上位クラス TOEIC 400 点以上

¹ 本学の全学共通カリキュラム英語部門が提供するライティング科目は、1年生を対象とするパラグラフライティングの力を伸ばす科目「Academic Writing I」と、2年生を対象とするエッセイライティングの力を伸ばす科目「Academic Writing II」から成る。2015年度には、前者が19クラス、後者は15クラス開講され、20人を超える教員が授業を担当している。このように1つの科目が複数のクラスと教員で運営されるため、この科目では質保証の手立てとして、コースコーディネーター教員があらかじめ科目ごとに推薦教科書を複数冊指定し、授業担当教員はその中から1冊を採択することになっている。科目目標と学生の力に合う、質の高い推薦教科書の選定が、科目運営の重要なひとつの要素となっている。

3.2 コーパスベースの単語リスト

本研究では、大学英語ライティング教科書の中で、一般的かつ学術的言語コミュニケーションの文脈で使用頻度が高い英単語、つまり、英語学習者として日本人大学生が優先的に学ぶべき英単語や、すでに習得済みであろう単語が、どの程度使用されているかを検証することを目的とする。そのため、表2の3つのコーパスベースの単語リスト、*General Service List (GSL)*、*Academic Word List (AWL)*、*JACET 8000* を分析の基準とした⁽⁸⁾⁽⁹⁾⁽¹⁰⁾⁽¹¹⁾。これらの単語リストは、複数の大きなコーパスから使用頻度が高い英単語を集約したものと、応用言語学や英語教育学の研究領域で広く認められ活用されている信頼性の高いものである²。単語リストによっては、基としたコーパスにおいて頻出度の高い単語から順に、または特定の英語検定試験と関連付けて、単語を複数のサブリスト (Level) に分割しているリストもある(表2)。サブリスト化されていないGSLについては、本研究の便宜上、著者らが分割した。

最近の研究では、現代の言語使用を反映したコーパスを基に作成された、*New General Service List*⁽¹²⁾、*New Academic Word List*⁽¹³⁾が発表され、*新 JACET 8000*⁽¹⁴⁾も2016年の発表に向けて準備中とのことである。本研究では、抽出した教科書の出版年がこれらの新しい単語リストの発表前であることから、元の単語リストを使って分析する。

表2 テキスト分析に使用したコーパスベース単語リスト

単語リスト名	単語数	サブリスト分割方法
General Service List (West 1953; Baumann & Culligan 1995)	2284 語 ワードファミリー ³	Level 1-11 GSLはサブリスト化されていない。本研究の便宜上、頻度順に200単語ずつ分割した。Level 11 のみ 284 単語存在する。
Academic Word List (Coxhead 2000)	2872 語 (570 ワードファミリー)	Level 1-10 高頻度順に60ファミリーずつ分割されたものを、それぞれ展開している。そのため、各 Level に存在する単語数はそれぞれ異なる。
JACET 8000 (JACET 2003)	基本単語 8000語 + 固有名詞等 250語	固有名詞等を除き、基本単語が頻度順に並べられ、1000語ごとにLevel 1-8に分割されている。Level 9に重要固有名詞が250語存在する。Level 1-3は高校までに学習しておくべき単語。Level 4は大学一般教養の英語初級レベル。

² 単語リストの基になっているコーパスは、GSLはBrown Corpus (1961)、AWLはCoxheadが作成した約350万語を含むコーパスで、単語は人文、法律、経済、科学の4分野28科目400以上のテキストから集められたものである。JACET 8000は、あらゆる分野のイギリス英語の書き言葉と話し言葉1億語を集めたBritish National Corpusに加え、アメリカ英語、日本の英語教育に関連する資料のサブコーパスに基づいている。

³ ワードファミリーは、基本となる単語(例 watch)とその派生語(例 watches, watched, watching)を含めて1つと数える⁽¹⁵⁾。

3.3 テキスト分析システム

本研究では、著者らが開発中であるウェブベースの英文テキストの単語分析システム「ステップ de タンゴ!」⁽⁶⁾を使用した。このシステムは、本来、学生がウェブページや雑誌、新聞などの英文記事を自律的に読む際のサポートと、教員がウェブページを教材として利用する際にウェブページで使用されている語彙の難易度診断のサポートを目的として開発されたシステムである。このシステムは大きく分けて以下3点の機能を有する。

1. **語彙集作成**: 入力された英文記事の中から、ユーザが指定した単語リストの中に存在する/しない単語を抜き出し、英英辞書または英和辞書から単語の意味を抽出しユーザの語彙集を作成する。
2. **記事・単語の保存・閲覧**: (1)で抜き出された単語を記事中でハイライトし、(1)で作成した語彙集とともに保存・閲覧できる。
3. **英文記事内の単語レベル測定**: 入力された英文記事内で使用されている単語を、ユーザが指定した単語リストを用いて分析し、使用頻度に基づく単語レベルから英文記事の学習適切度を測定する⁽⁶⁾。

(1)と(2)の機能は英文記事を読むためのサポート機能であり、特に英語習熟度の低い学習者の自律的学習の大きな助けとなる。(3)は学習者だけでなく、教員にとっても有用な機能である。3.2節で説明した権威あるコーパスベースの単語リストを用いることにより、英文記事の単語レベルが学習者に適しているかを測定し、推定する。本研究では、(3)の英文記事内の単語レベル測定機能を用い、大学英語ライティング教科書のモデルテキストで使用されている、単語リストの英単語の出現率を調査、分析した。

次にシステムを用いた分析方法を説明する。このシステムは登録制のため、まずログインを行い「解析 de 単語」画面に移動する。この「解析 de 単語」画面でデータを入力し、テキスト分析、単語集作成、レベル判定などを行う⁴。「解析 de 単語」画面では、左部分に現れる「タイトル」「記事」の2つのテキスト入力画面と、右部分に現れる「単語リスト」「辞書」の選択画面の左右2つの画面からなる。まず、システム内では入力された英文テキストを空白で区切り単語を抽出する。次に、プリンストン大学開発の英語語彙データベースWordNet⁽¹⁶⁾を用い、この抽出された単語を基本形に戻し、重複を取り除く。そして、その基本形が各単語リストのどのレベルに存在するかを調べ、レベルごとに出現個数を算出した。最終的に、結果として単語リストのレベル順に出現個数が表示される(図1)。

⁴ この画面では、他に「学ん de 単語」「読ん de 単語」の3つのタグが表示され、クリックすることにより他の画面に移動することができる。

図1 オンラインテキスト分析システム「解析 de 単語」のテキスト分析結果表示の画面イメージ

3.4 テキスト分析の手順

まず、表1が示す教科書の各章をスキャナーで読み取り、モデルライティング（パラグラフまたはエッセイ）の本文テキストデータをデジタル化した。次に、教科書ごとに、すべてのモデルライティングのテキストを1つのファイルにまとめた。1冊につき約10~15のモデルライティングが集まった。この時点では、データ内にはスキャナーの読み取りミスが存在するため、人力で修正作業を行った。この後、データを前節で紹介したテキスト分析システムに入力し、GSW、AWL、JACET 8000と順番に照らし合わせ、テキストの中に、各単語リストの単語項目がいくつ出現したかの出現回数と、その回数のテキスト全体の単語数に占める割合を出現率として算出した。単語の出現回数は、テキスト分析システムでは「ヒット数」として示される（図1）。出現率は、単語の出現回数を、表2で紹介した各単語リストのレベル別に分け、各テキスト内の重複を取り除いた使用単語数で割って%を算出した。

4. 結果と考察

まず、単語リストのフィルターをかけない状態で、教科書のモデルテキストの「述べ使用単語数（重複あり）」と、その中で重複している単語を1つとして数えた「使用単語数（重複なし）」を整理した（表3）。述べ使用単語数の多さに比べ、使用単語数がかなり少なくなっており、同じ単語が複数のモデルテキストの中で繰り返し使われている傾向を示す結果となった。例えば、教科書Bはモデルライティングの数が他の教科書よりもかなり多いため、述べ使用単語数が11701語と多くなっているが、実際に使われている単語数を重複なしでみると、1900語と全体の約16%に留まっている。同じ単語を教科書内の様々なモデルライティングで使うことで、語彙インプットの量と機会を増やし、学習者の語彙習得を促そうという意図がうかがわれる。推測の域を出ないが、もしそうであれば、語彙習得促進の観点からは好ましい教科書であると言える。

表3 モデルライティングテキスト中の使用単語数と割合

教科書	述べ使用単語数（重複あり）	使用単語数（重複なし）	使用単語数のテキスト全体の割合
A	3781語	696語	18.41%
B	11701語	1900語	16.24%
C	2311語	640語	27.70%
D	3043語	746語	24.52%

なお、使用したテキスト分析システムの設計には、各単語リストの「述べ出現単語リスト単語数（重複あり）」を解析する機能が含まれていないため、本研究では、「述べ出現単語リスト単語数」を「述べ使用単語数」を割って算出する「各単語リストの単語がテキスト全体を占める割合」を明らかにすることはできなかった。これはこの機能がシステム開発時の利用目的にはなかったからで、もし教材の語彙の精査に今後も利用するのであれば、この点は今後の機能拡張へ向けての課題である。

次に各教科書のテキスト分析の結果を、単語リストごとに考察する。モデルテキストに単語リストの単語がどの程度使われているかを詳しく見るために、各レベルの単語の出現回数（重複なし）とテキスト全体の使用単語数（重複なし）から、出現率を計算したのが表4である。ここで言う「出現回数」とは、テキスト中で使用されている単語リストの英単語の総数のうち、重複使用されている項目を1つとして数えた、異なる単語の回数である。「出現率」は、この「出現回数」を表3で紹介した「使用単語数（重複なし）」で割って算出した。

4.1 GSL 単語の分布

すべてのサンプル教科書で全体的なGSL単語の出現率の合計は、教科書A-Dの順に、68.5%、63.1%、76%、67.4%と、教科書B以外は7割前後と、比較的高い結果となった（表4）。なお、GSLのLevel 1-11は分析の便宜上、著者らが単語を頻出度の高い順番に200単語ずつ分割したものである（表2）。特にLevel 1-5は最も使用頻度の高い基本1000語と言えるが、これらの単語の出現率は、教科書AからDの順に、61.5%、48.8%、66.9%、59.6%と、3冊で60%前後になっており、教科書B以外は平易な英語で書かれた内容理解が容易なモデルライティングと言ってよいだろう。これに比べ、Level 6-10の単語の出現率は4%以下と全体的に低かった（図2）。

GSLに含まれる英単語は基本の約2000語で、あらゆる文書の約80%はカバーできると言われているリストであることから⁽⁷⁾、どのレベルの英単語もまんべんなくテキスト中に使用されている可能性が高いと予想できた。しかし、今回調査した4冊のサンプル教科書については、教科書Cの76%以外は63.1-68.5%と70%未満の出現率に留まった。

表4 GSL 単語の出現個数(上段)と出現率(下段)

Level	1	2	3	4	5	6	7	8	9	10	11	合計
教科書A	166	123	66	48	24	17	11	9	6	2	5	477
	239%	177%	95%	69%	35%	24%	16%	13%	09%	03%	07%	685%
教科書B	251	235	189	143	108	75	65	50	42	25	15	1198
	132%	124%	10%	75%	57%	40%	34%	26%	22%	13%	08%	631%
教科書C	196	102	62	45	21	19	14	8	11	3	1	484
	306%	16%	10%	70%	33%	30%	22%	13%	17%	05%	02%	76%
教科書D	183	119	68	48	27	23	15	11	7	6	2	509
	245%	16%	91%	64%	36%	31%	20%	15%	09%	08%	03%	674%

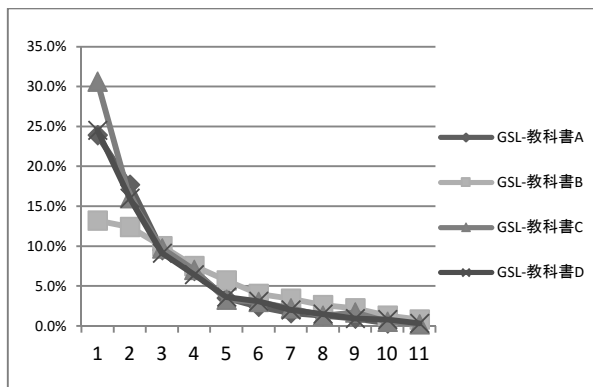


図2 GSL 単語 Level 1-11 の教科書モデルライティングでの分布

4.2 AWL 単語の分布

サンプル教科書4冊のモデルテキスト内のAWL単語Level 1-10の分布と出現率は、表5と図3が示す結果となった。テキスト全体で見ると、AWL単語の使用は、下位クラス用の和書教科書A、Cで約8%、下位クラス用の洋書教科書Bは12%、上位クラス用の洋書教科書Dは15%程度であることが分かった。Nation⁽¹⁷⁾はGSLの基本2000語があらゆるテキストの80%近くをカバーし、AWL単語は次の8-10%をカバーすると述べているが、本調査の4冊のサンプルにおけるAWL単語の出現率はこの指摘とほぼ一致する結果となった。しかしながら、これらが大学生向けライティング教科書であることと、一般的に学部生にとってアカデミック語彙の習得は難しいことを考えると⁽¹⁸⁾、本研究のサンプル教科書AとCのAWL単語使用率は低めであり、アカデミック語彙の使い方も学ぶライティング科目で使う場合には補助教材が必要であろう。

各教科書のAWL単語のレベル別分布を図3のグラフで見ると、4冊とも同様の傾向が見られたGSL単語(図2)とJACET 8000単語(図4)の分布とは、若干異なった分布を示す結果となった。全体としては右肩下がり、4冊とも使用頻度の高いLevel 1-2のAWL単語から順に使用している傾向が見られるが、Level 4-7の単語使用にはバラつきが見られるなど、これらの教科書のモデルライティングが必ずしも高頻度のアカデミック単語を優先的に使用して作成されているわけではな

いことがわかる。単語より題材が優先されている可能性がある。

表5 AWL 単語の出現個数(上段)と出現率(下段)

Level	1	2	3	4	5	6	7	8	9	10	合計
教科書A	10	11	6	10	5	7	5	2	1	1	48
	14%	16%	09%	14%	07%	10%	07%	03%	01%	01%	83%
教科書B	44	51	18	22	26	24	21	8	11	5	230
	23%	27%	10%	12%	14%	13%	11%	04%	06%	03%	121%
教科書C	15	10	6	6	5	0	4	2	1	3	52
	23%	16%	09%	09%	08%	00%	06%	03%	02%	05%	81%
教科書D	28	22	17	12	5	10	11	4	2	1	112
	38%	30%	23%	16%	07%	13%	15%	05%	03%	01%	150%

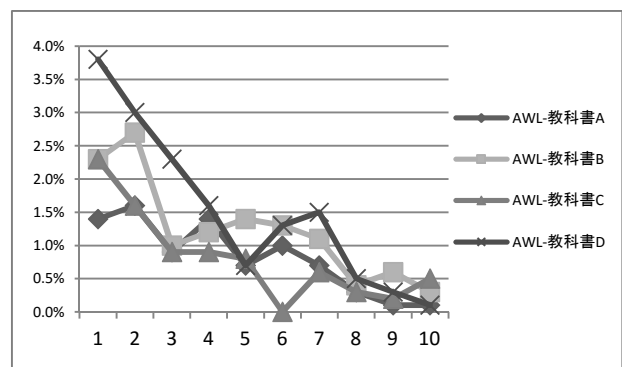


図3 AWL 単語 Level 1-10 の教科書モデルライティングでの分布

4.3 JACET 8000 単語の分布

JACET 8000単語リストの単語の分布と出現率は、教科書A-Dの順に、約89%、88%、94%、92%だった(表6)。この単語リストの単語出現率がGSL、AWLと比べて高くなっていたのは、単語リストに存在する総単語数の違いであると考えられる(表2)。JACET 8000のひとつのレベルには1000語含まれ、他の単語リストより大分多い。

GSL単語出現率と同様に、最も使用頻度の高い初めの基本1000単語(Level 1 中学校英語教科書レベル)の出現率を見ると、教科書A-Dの順に56.9%、41.5%、61.1%、54.8%となり、GSL単語よりも低い結果となった。JACETによれば、一般英文の約70%がこのLevel 1の単語で書かれているという。さらに高校初級レベルのLevel 2、高校英語教科書レベルのLevel 3も含めてLevel 1-3の単語の出現率を総合してみると、教科書A-Dの順で77.2%、72.1%、80.5%、77.9%となり、モデルライティングの7-8割は、学生たちが大学入学前に習得したであろう英単語で書かれていることになる。つまり、語彙の点では、大学生にとって読みやすいモデルテキストとなっている可能性が高いことが示唆された。

表6 JACET 8000 単語の出現個数(上段)と出現率(下段)

Level	1	2	3	4	5	6	7	8	9	
教科書 A	396	95	46	25	9	19	6	4	20	620
	56.9%	13.7%	6.6%	3.6%	1.3%	2.7%	0.9%	0.6%	2.9%	89.1%
教科書 B	789	378	203	91	64	38	37	28	46	1674
	41.5%	19.9%	10.7%	4.8%	3.4%	2.0%	2.0%	1.5%	2.4%	88.1%
教科書 C	391	83	41	34	10	7	6	2	25	599
	61.1%	13.0%	6.4%	5.3%	1.6%	1.1%	0.9%	0.3%	3.9%	93.6%
教科書 D	409	123	49	38	20	14	12	10	14	689
	54.8%	16.5%	6.6%	5.1%	2.7%	1.9%	1.6%	1.3%	1.9%	92.4%

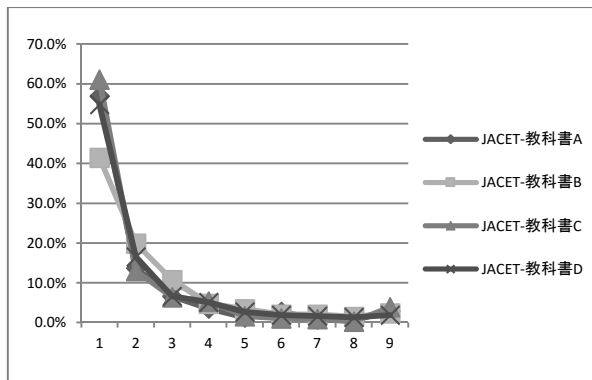


図4 JACET 8000 単語 Level1-9 の教科書モデルライティングでの分布

さらに 4.1-4.3 節で報告した結果を、4 冊のサンプル教科書の比較のために整理したものが表 7 である。大学で学ぶべき語彙がこれらの教科書でどの程度使用されているかを把握しやすくするため、JACET 8000 は、Level 1-3 (中学・高校英語教科書レベル 3000 語) と、Level 4-6 (大学一般教養初級から英語を専攻としない大学生が目指すべきレベル 3000 語)⁽¹⁾に分けて出現率を示した。大学で学ぶべき語彙を含むと考えられる AWL と JACET 8000 Level 4-6 の単語出現率に絞ると(表 7 網掛け列)、教科書 D (中級上～上級エッセイ) では前者が 15%、後者が 9.7%とその差が 5%強となった。脚注 2 の通り、AWL は学術文献に特化したコーパスから、JACET 8000 は学術書、新聞、小説、日常会話も含む広範囲のテキストを対象とした British National Corpus から頻出語を選択しており、この 2 つの単語リストは、基準となるコーパスの特性が大きく異なる。このことから、学術的な論証文ライティングの指導書である教科書 D は、より AWL 単語の使用が多くなっていると推測できる。アカデミック語彙の使い方も学ぶ大学 2 年生向けのライティング科目に適切な教科書であると言えるだろう。一方、教科書 B は AWL 単語 12.1%、JACET4-6 単語 10.2%と、アカデミック語彙と大学生が学ぶべき一般語彙がバランス良く使用されており、大学 1 年生向け中級レベルのパラグラフライティングの教科書として、語彙の観点からは適切であることが数値から示唆された。

表7 サンプル教科書内モデルライティングにおける各単語リストの単語の出現率

教科書	GSL	AWL	JACET 8000 Level 1-3	JACET 8000 Level 4-6
A	68.5%	8.3%	77.2%	7.6%
B	63.1%	12.1%	72.1%	10.2%
C	76%	8.1%	80.5%	8.0%
D	67.4%	15.0%	77.9%	9.7%

4.4 その他

テキスト分析の結果、いずれの単語リストにも入らなかった単語も多く存在した。例えば、Daisuke Matsuzaka などの個人名や、baseball などのスポーツ名、Asian, Japan などの地域名、国名、barbecue などの料理名などの固有名詞がほとんどであった。また、mobile、blog、CGI など現代では比較的使用頻度の高い単語がいずれのリストにも含まれなかった。これは、時代によって単語が新しく作られ、その反対に使用頻度が高かったものが使用されなくなることを表していると言えるだろう。これは今回の実験に使用した単語リストが古くなってきていることを示しており、より現代の言語使用を反映した新しい単語リストの発表が待たれるところである。

また、テキスト分析の際、いずれの単語リストにも入らなかった単語を分類することにより、その教科書のモデルライティングが提供している内容のある程度推測することができることがわかった。例えば、教科書 A には野球に関する名詞、baseball、pitcher、tournament が存在し、野球を題材としたモデルライティングがあることがわかる。教科書 B では、Amazon、Facebook、iPhone などコンピュータに関する単語が多く、教科書 C では、cram、classroom、graduation など、教育に関する題材を扱っていることが示された。

5. 結論と課題

本稿では、オンラインのテキスト分析システムを用い、大学英語ライティング教科書のモデルライティングにどのような単語が使用されているか、3つのコーパスベース単語リストに照らし合わせて検証した。分析の結果、サンプル教科書のモデルライティングの約 7 割は、GSL 単語、JACET 8000 Level 1-2 単語に表される一般的に使用頻度の高い基本 2000 語で書かれていることが分かった。これは上級学生を対象としている教科書についても見られた傾向である。このことは、モデルライティングを学生が読んで分析する際、内容を理解しやすい傾向を示しており、論展開や文章構成、引用のしかたなど、ライティングスキルの学習項目に注意を向けやすくなるため、好ましい傾向であると言える。

さらに、基本語よりも難易度の高い AWL 単語の分布については、和書 2 冊の教科書で 8%、洋書 2 冊の教科書で 12-15%程度の使用が認められた。教科書によって AWL のレベル別単語の出現率は異なり、必ずしも使用頻度の高い単語から順にモデルライティングで使われ

ているとは限らないことを示す結果となった。また、このように AWL 単語の出現率が低い教科書を使用する際は、学生がアカデミック語彙を発信語彙として使う練習は補助教材で行うなど指導の工夫が必要であろう。

本研究ではデータ化したテキストと元の教科書テキストの突き合わせと修正に時間がかかったため、限られた冊数を対象とする分析になった。今後はさらに分析対象とする教科書を増やし、ライティング科目で推薦する教科書の語彙水準の全体的な傾向を把握したい。また、テキスト分析システムの機能の精度を、*AntConc*⁽¹⁹⁾や *Range*⁽²⁰⁾など、他の同様の語彙分析ツールと比較して検証する必要がある。さらには、いずれの単語リストにも入らない単語(数を表す単語など)も多くあるため、これらの単語を分析時にどのように扱い、表示するかの工夫も必要である。

本研究のライティング教科書の分析は、このテキスト分析システムを、他の英語科目の教科書や教材分析にも同じ手法で容易に活用できる可能性を示した。教科書全ページの分析は時間的に現実的ではない場合でも、サンプル数ページを分析すれば、ある程度その教科書の使用語彙の傾向は把握できる。特に、推薦教科書や統一教科書、統一教材を導入している語学プログラムにとっては、教材選択の意思決定過程において有用な情報を提供するツールとなりうる。

本研究で利用したテキスト分析システムは、現代の英語使用を反映させた、New GSL⁽¹²⁾、New AWL⁽¹³⁾、新 JACET 8000⁽¹⁴⁾などの単語リストや、信頼できる他の単語リストと目的に応じて組み合わせることで、今後の利用方法が広がる。例えば、学生が書いたエッセイを AWL と JACET 8000 Level 4-6 を使って分析すれば、アカデミックで大学生にふさわしい語彙がどの程度使用されているかを、教員や学生自身が把握できる。学生が、自分が使う語彙の難易度を自己診断し、より洗練された語彙使用に挑戦して改善することができれば、有用なライティング力育成の支援につながる。また、ビジネス英語や経済英語など専門分野に特化した単語リストと AWL を組み合わせ、指定した単語が多く含まれるウェブサイトやインターネット上で自動的に抽出する仕組みにも応用できれば、効率よく自分の学習目的に適した記事を探し読む活動に生かすことができる。

応用言語学研究分野において、テキスト中の使用語彙分析研究は多いが、その手法は語彙研究の専門家でない教員や学生にとっては、日々の指導や学習活動に気軽に取り入れやすいとは言いがたい。教員と学生ユーザが使いやすい、研究と教育実践をつなぐツールとして、本テキスト分析システム開発を継続させたい。

謝辞

本研究は、獨協大学情報学研究所の助成と、第二著者の獨協大学研究奨励費(2015年度)の助成によるものである。

引用文献

- (1) 岡田圭子・飯島優雅(編) 継続的な英語教育改革の過程と成果—学士力育成に資する英語教育の充実. 獨協大学全学共通カリキュラム英語部門 2003-2012 年度実践報告書. pp. 106-107, 獨協大学全学共通カリキュラム部門(2013)
- (2) Nation, I. S. P., Macalister, J., *Language Curriculum Design*. NY: Routledge (2010)
- (3) Nitta, R., Gardner, S. "Consciousness-raising and practice in ELT coursebooks," *ELT Journal*, 59, 1, pp. 3-13 (2005)
- (4) Iijima, Y., Horie, I. "A word list generator program for using authentic texts in an academic English reading class," *Proceedings of the 9th International Conference on Information Technology Based Higher Education and Training*, pp. 407-412 (2010)
- (5) Horie, I. Kashiwabara, K., Yamaguchi, K., and Iijima, Y., "Personalized Teaching Material Generator Based on Word Set," *Information Technology Based Higher Education and Training*, pp. 343-348 (2010)
- (6) 堀江郁美, "適応型単語リストを用いた自律学習支援システム—英語学習に用いられる複数の語彙リストの比較—," 獨協大学情報学研究, pp. 64-71 (2014)
- (7) Council of Europe. "Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)" http://www.coe.int/t/dg4/linguistic/cadre1_en.asp
- (8) West, M., "A General Service List of English Words," Longman (1953)
- (9) Bauman, J., Culligan, B. "About the General Service List," <http://jbauman.com/gsl.html> (1995)
- (10) Coxhead, A., "A New Academic Word List," *TESOL Quarterly*, 34, pp. 213-238 (2000)
- (11) 大学英語教育学会基本語改訂委員会(編), "大学英語教育学会基本語リスト JACET List of 8000 Basic Words" (通称 JACET 8000), 大学英語教育学会 (2003)
- (12) Brezina, V., Gablasova, D. "Is There a Core General Vocabulary? Introducing the New General Service List," *Applied Linguistics*, 36, 1, pp. 1-22 (2015)
- (13) Gardner, D., Davies, M. "A New Academic Vocabulary List," *Applied Linguistics*, 35, 3, pp. 305-327 (2014)
- (14) 望月正道, 笠原究, 上村俊彦, "基本語改定特別委員会報告: JACET 新 JACET 8000 の提案<JACET 基本語改訂特別委員会>," *The JACET 54th International Convention Book*, p. 63 (2015.8)
- (15) Bauer, L., Nation, P. "Word families," *International journal of Lexicography*, 6, 4, pp. 253-279 (1993)
- (16) WordNet at Princeton University, <http://www.wordnet.princeton.edu>
- (17) Nation, I. S. P. *Learning vocabulary in another language*. New York: Cambridge University Press. (2001)
- (18) Hyland, K., Tse, P., "Is there an "academic vocabulary"?", *TESOL quarterly*, 41, 2, pp. 235-253 (2007)
- (19) Anthony, L. "AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom." In *Professional Communication Conference, 2005. IPCC 2005. Proceedings. International*, pp. 729-737. IEEE (2005)
- (20) Nation, I.S.P., Heatley, A., "Range: A program for the analysis of vocabulary in texts [software]" (2002)

(2015年9月29日受付)

(2015年12月2日採録)