

Development of English Vocabulary Lists for Science and Technology Students

**Yosuke Sasao
Mihoko Kato
David Levin
Eugene Ryan**

Abstract

This research aims to create English vocabulary lists for students of science and technology for the purpose of improving English education at Toyohashi University of Technology (TUT). We developed a 4.5-million-word corpus from English academic journals that academic staff at TUT considered influential in their fields of research. By examining the frequency and range, we selected 5,349 words that will be useful in effectively fostering engineering students' English proficiency. The results show that knowledge of these words gives an individual the ability to understand more than 90% of the vocabulary used in science and technology academic papers. This figure indicates the importance to undergraduates of learning these words in order to start graduate-level research, where students are often required to understand and publish in English academic journals.

工学英語語彙リストの開発

笹尾 洋介

加藤 三保子

レヴィン デイヴィッド

ライアン ユージン

1. はじめに

外国語としての英語運用能力を高めるためには、英語語彙知識の育成が極めて重要な役割を果たす（例：Nation 2001）。学生の語彙力向上のために、豊橋技術科学大学（以下、本学）では、平成24年度から学部1～2年生を対象にe-learningによる語彙力強化指導を積極的に行っている。しかしながら、当該システムで学習する語彙は一般的な語彙を多く含み、工学を専門とする学生が必要とする語彙を必ずしも効率よく網羅するものではない。最新の研究成果は英語で発表されることが多いため、学生の研究を成功に導くためにも、英語学術論文を効果的に読む能力を身に着けさせる必要がある。また、全学の8割を占める高専からの編入生に対しても、英語の基礎力としての語彙指導を行う必要がある。本稿の目的は、こうしたニーズを満たすため、本学の学生に有益な英語語彙を明らかにし、どの段階でどのような語彙を指導すれば教育的効果を最大限に引き出せるかを考察することにある。

本研究は、ESP（English for Specific Purposes：特定目的の英語）の観点から、より効果的な語彙指導を目的とした工学系英語語彙リスト作成を行った。具体的には、本学学生が学ぶのにふさわしい英語語彙を実証データに基づいて選定するために、本学の専門教員が推薦する英語学術誌をもとに延べ1,003本の英語学術誌から総語数約450万語のコーパスを作成した。当該コーパスより、広範囲かつ高頻度で使用される語を抽出した。

2. 語彙指導とESP

ESPとは、あるディスコース・コミュニティー（専門家集団）において明確かつ具体的な目的達成のために使用される英語に関する言語教育研究を指す（寺内 2000）。ESPはEAP（English

for Academic Purposes：学術目的の英語）と EOP（English for Occupational Purposes：職業目的の英語）に大別される（Dudley-Evans & St John 1998; Robinson 1991）。本学の基本理念が「技術を科学で裏付け、新たな技術を開発する学問、技術科学の教育・研究を使命とする」と規定されていることを考慮して、本稿では特に EAP に焦点を当てる。EAP は EGAP（English for General Academic Purposes：一般学術目的の英語）と ESAP（English for Specific Academic Purposes：特定学術目的の英語）に下位区分され、両者の間に有機的な関連性を持たせることが望まれる（田地野・水光 2005）。具体的には、工学を専門としない英語教員は EGAP を、工学を専門とする教員は ESAP をそれぞれ担当することが可能であろう。

本研究では、工学を専門としない英語教員が工学系の学生のニーズに合った語彙指導を行えるように、工学系学術論文において幅広く使用され、かつ英語教員が指導可能な語彙を抽出することにより、さらに効果的な語彙指導を目指す。

3. 学術語彙コーパス構築手順

本学学生にふさわしい工学系語彙を抽出するためには、工学分野の研究者の協力が不可欠である。そこで、本学の各専門分野（機械工学（1系）、電気・電子情報工学（2系）、情報・知能工学（3系）、環境・生命工学（4系）、建築・都市システム学（5系））の教員に当該分野で権威がある、または評価の高い英語学術雑誌を推薦してもらった。次に、その推薦雑誌から最新の論文（各雑誌 10～15 本）を無作為に抽出した。その際、特定の話題に関する特集が組まれている可能性を考慮し、各号から 1 本ずつを無作為に選び、最新号から順次さかのぼって論文を入手してデータ化した。電子データが入手できない論文は、ABBYY FineReader11（OCR ソフトウェア）を用いてテキストデータに変換した。電子データ化するに際し、以下の五つの要素を手作業で除去した。

1. 著者、所属、連絡先（E-mail アドレスなど）
2. ページ番号、ヘッダー、フッター（雑誌名・タイトル・著者名など）
3. 数式、化学式
4. 図表（ただし、見出しは残す）
5. 引用文献

また、文字化け等の修正も手作業で行った。

専門分野ごとの調査対象雑誌数、論文数、および語数が表 1 に示されている（各専門分野から選出された雑誌に関する詳細は、笹尾・加藤・鈴木（2014）を参照）。

表 1 分析対象雑誌数、論文数、語数

専門分野	雑誌数	論文数	語数
1系	23	230	910,508
2系	14	198	734,783
3系	12	180	937,209
4系	18	210	940,506
5系	15	185	1,062,378
計	82	1,003	4,585,384

4. 語彙リスト作成手順

作成された学術語彙コーパスは、Lawrence Anthony 氏の AntWordProfiler 1.4.0 を用いて、語の使用頻度と範囲を調査した。さらに染谷泰正氏のレマリスト e-lemma を用い、レマ化を行った。

抽出された語を表 2 に掲げる基準に基づいてリスト 0 からリスト 8 の合計 9 種類に分類した。なお、一般語彙との関係を見るために、Paul Nation 氏の BNC/COCA word family lists (以下、BNC/COCA リスト) を下位区分の基準に取り入れた。これは小説、新聞、会話など多様なジャンルを含む BNC (British National Corpus) と COCA (Corpus of Contemporary American English) の合計 5.5 億語のコーパスでの頻度情報を基に作成された、1,000 語ごとに区分されたリストである。表 2 内の「範囲 = 5」とは五つの専門分野の雑誌のすべてにおいて 1 回以上出現する語を、「範囲 = 3 以上」とは五つの専門分野のうち 3 分野以上で使用されている語をそれぞれ指し示す。

表 2 工学系語彙分類基準

リスト	範囲	BNC/COCA リスト	頻度	その他
0	5	-	-	機能語 (冠詞、代名詞、数詞)
1	5	-	-	機能語 (接続詞、助動詞、前置詞など)
2	5	1k	-	
3	5	2k	-	
4	5	3k	-	
5	5	4k-5k	-	
6	5	6k-10k	-	
7	3 以上	1k-10k	50 以上	
8	3 以上	1k-10k 以外	50 以上	

リスト 0 と 1 は機能語を集めたものである。リスト 2, 3, 4 はそれぞれ BNC/COCA リストの 1,000 語、2,000 語、3,000 語レベルに含まれるものである。先行研究は、日本人大学生の語彙力をおよそ 3,000 - 5,000 語であると見積もっており (Mochizuki & Aizawa 2000; Sasao 2008; Schmitt & Meara 1997)、この結果を踏まえるとリスト 4 までの語彙は本学入学前に修得済みであると考え

することもできる。しかしながら、上位 3,000 語は学術分野以外でも様々なジャンルにおいて高頻度で使用される語 (high-frequency vocabulary) であるため、言語学習の初期段階で習得してしまうことが推奨されている (Schmitt & Schmitt 2014)。よって、英語を苦手とする学生のために、基礎から重要語を学習できるように、これらの語彙リストをレベル別に設定することに教育的価値があると考えられる。

リスト 5 と 6 は上位 3,001 語から 10,000 語までの語を含んでおり、リスト 7 は、上位 10,000 語のうち、すべての工学分野において使用されるわけではないが、一定の頻度以上 (50 以上) で使用される語を集めたものである。未知語の意味を推測しながら多様なジャンルの文章を読めるようになるためには、9,000 語と固有名詞の知識が必要であるとの先行研究の結果をもとに考えると、10,000 語の語彙力を身に着けることは学習目標の一つと捉えることができる。リスト 1 から 7 は上位 10,000 語のうち、様々な分野の工学系学術論文において使用される語をまとめたものであるため、一般目的の英語においても重要な役割を果たすと同時に、工学を専門とする学生にとっても重要な語が選出されている。なお、リスト 8 は、BNC/COCA 語彙リストの上位 10,000 語に入っていないものの、一定以上の頻度と範囲をもって使用される語を収録したものである。

5. 結果と考察

表 2 の基準に従って抽出された語の数と、それらの語が全テキストに占める割合 (カバー率) が表 3 に示されている。

表 3 抽出語彙とそのカバー率¹⁾

リスト	語数	総頻度数	累積頻度数	累積カバー率
0	41	655,867	655,867	14.3
1	130	1,268,316	1,924,183	42.0
2	972	746,138	2,670,321	58.2
3	878	565,179	3,235,500	70.6
4	979	486,861	3,722,361	81.2
5	734	192,312	3,914,673	85.4
6	512	89,817	4,004,490	87.3
7	598	84,794	4,089,284	89.2
8	505	68,442	4,157,726	90.7
計	5,349	4,157,726		

表 3 より、本研究で抽出した 5,349 語 (レマ換算²⁾) の知識があれば、工学系分野の英語学術論文で使用される語の約 9 割をカバーできることが明らかとなった。抽出された語には、固有名詞、略語 (例: fig)、頭文字語 (例: DNA) および化学記号 (例: NaCl) が含まれていないことを考えると、この語彙リストには工学系英語論文読解において必要とされる英語のほとんどが網羅

されていると言えよう。これらの語に加えて、各専門分野で使用する専門語彙を習得することにより、工学系英語学術論文の内容を理解するのに大きな問題は生じないと考えられる。

表4は各語彙リストに含まれている語の例を示している。なお、例に挙げた語は、当該リスト内で使用頻度が最も高い5語である。このような工学系英語語彙リスト作成により、BNC/COCA リストなどの既存の一般語彙リストを使用して英語学習を行うよりも、工学系の学生にとって、より効果的・効率的に彼らのニーズに合った語彙を習得することが可能となる。

表4 各リストに含まれる語 (例)

リスト	例
0	the, a, that, this, we
1	of, be, and, in, to
2	use, show, high, time, also
3	result, model, value, increase, surface
4	cell, data, temperature, structure, method
5	parameter, particle, spectrum, dynamic, domain
6	polymer, shear, accord, vector, deformation
7	neuron, urban, flame, policy, magnetic
8	pathway, preset, cathode, annealing, diffraction

リスト0は、機能語、とくに、冠詞、数詞、代名詞を集めたものであるため、すべての学生が知っていると考えられる。リスト1にはリスト0に含まれない機能語が含まれている。これはほとんどの学生にとってはなじみ深いものかもしれないが、たとえば、whether、whereas、despiteといった論文で頻出する語には改めて注意する必要があるかもしれない。リスト2から6は、すべての専門分野で出現する語をBNC/COCA リストに基づいて分類したものである。リスト2や3は、中学・高校で学習した語が多く含まれているため、大学入学前の復習が必要な学生にとって適切なレベルであろう。リスト2と3の学習を終えた学生は、さらに4、5、6、7へと進んでいくことになる。

リスト1から7はBNC/COCA 語彙リストの上位10,000語から選出されたものである。Nation (2013)によると、10,000語より頻度の低い語は、ある特定の分野においてのみ使用される場合が多く、母語話者であっても知らない語が多く含まれると報告している。よって、本学において、工学を専門としない英語教員がEGAPとして対応できる語は、リスト1から7までの語となる。リスト8に含まれる語は、EGAP語彙ではなく、ESAP(工学専門)語彙としてとらえるのが適切であろう。カリキュラム開発の観点からは、リスト1から7までの語彙を英語教員が主体となって学部生に指導し、それ以外の専門語彙を大学院に進学後、各専門分野の教員が主体となって指導するといった形で、総合教育院が提供する英語科目と、各専門分野との間に有機的な連携を図ることが可能となるであろう。

今後は、リスト記載の語に意味や用例を付加することにより、本学学生に特化した英単語集の

作成が可能である。さらに、本研究で作成したコーパスを分析することにより、連語表現（例：et al）を抽出し、学習者に提示することは教育的価値があるであろう。そうすることにより、限られた時間内で、より効果的に必要とされる語彙を習得し、大学院進学後に円滑に研究活動に打ち込むことができると期待される。

6. おわりに

本研究は、本学において総合教育院が提供する英語科目のより一層の充実を目的として、工学系英語語彙の抽出を試みた。各系の専任教員からの推薦を受けた英語学術雑誌から約 450 万語のコーパスを構築し、各専門分野において広範囲・高頻度で使用されている語を約 5,500 語抽出した。それらの語を、BNC/COCA 語彙リストを基に 8 種類に分類し、易しい（頻度の高い）語から難しい（頻度の低い）語へと段階的に学習を進められるように工夫した。これらの語を習得することにより、英語学術論文の約 90% の語を理解することができ、固有名詞や専門用語などを加えると、内容を十分に理解できるレベルになると考えられる。リスト 1 から 7 に含まれる語は、EGAP 語彙と捉えることも可能であり、工学を専門としない英語教員が指導可能である。これらの語を学部学生に指導することにより、これまで以上に効果的な英語教育が実践できる。学部・大学院の一貫教育を特色とする本学においては、本研究の成果が活かされることにより、学生は大学院進学後に円滑に研究活動を進めることができる。

注

- 1) 平成 26 年度全国高専教育フォーラムにける口頭発表（「豊橋技術科学大学における工学系英語語彙リストの作成」、発表者：笹尾洋介・加藤三保子）と数値が異なっているのは、教育的観点から、英綴りを米綴りに統一したり（例：optimise を optimize に統一）、数詞（例：one, fourth）をレベル 0 にまとめたためである。
- 2) 本研究ではレマ換算を採用している。レマ換算とは、屈折形（inflection）をまとめて 1 語として数える数え方である。たとえば、play, plays, playing, played は異なる 4 語ではなく、1 語と数えている。それに対して、BNC/COCA 語彙リストでは、ワード・ファミリー換算が使用されており、屈折形と特定の派生形（derivative）をまとめて 1 語とみなしている。たとえば、play, plays, playing, played, player, players は異なる 6 語ではなく、1 語と数えられている。そのため、レマ換算の方がワード・ファミリー換算に比して語数が多くなる。本研究でレマ換算が使用されているのは、日本人学生は派生形を作る基本的な接辞の知識が必ずしもあるわけではないためである（Mochizuki & Aizawa 2000）。

付記

本稿は、平成 26 年度全国高専教育フォーラムにて行われた口頭発表（「豊橋技術科学大学における工学系英語語彙リストの作成」、発表者：笹尾洋介・加藤三保子）を加筆修正したものである。本研究は、「豊橋技術科学大学における科学技術英語語彙データベースの構築 — 中学・高校英語と工学系専門英語との連携を目指して —」と題する平成 25 年度豊橋技術科学大学教育研究活性化経費（研究代表者：笹尾洋介）の助成を受けて行われた研究成果の一部である。

引用文献

- Dudley-Evans, T., & St John, M. (1998). *Developments in ESP: A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291-304.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (Second ed.). Cambridge: Cambridge University Press.
- Robinson, P. (1991). *ESP today: A practitioner's guide*. Hemel Hempstead: Prentice Hall International.
- Sasao, Y. (2008). Estimating vocabulary size: Does test format make a difference? *JACET Journal*, 46, 63-76.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19(1), 17-36.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503.
- 笹尾洋介, 加藤三保子, 鈴木新一 (2014) 「豊橋技術科学大学における工学系英語語彙データベースの構築—より効果的な語彙指導を目指して—」豊橋技術科学大学総合教育院紀要『雲雀野』第36号, pp.35-44.
- 田地野彰, 水光雅則 (2005) 「大学英語教育への提言」, 竹蓋幸生, 水光雅則 (編) 『これからの大学英語教育』 pp.1-46. 東京: 岩波書店.
- 寺内一 (2000) 「ESPを知る」, 深山晶子 (編) 『ESPの理論と実践: これで日本の英語教育が変わる』 pp.9-32. 東京: 三修社.