

京都大学学術論文コーパスを用いた学術語彙リストの作成

金丸 敏幸¹ 笹尾 洋介¹ 田地野 彰¹

京都大学¹

(kanamaru@hi.h.kyoto-u.ac.jp, ysasaojp@yahoo.co.jp, akira@tajino.mbox.media.kyoto-u.ac.jp)

供することにより，社会の発展に寄与するものとする。」

1 はじめに

現在，日本の大学は大きな変革の時期にあり，その中において，大学英語教育のあり方も大きく問われている．しかしその一方で，大学の英語教育の目標としては，「大学を卒業したら仕事で英語が使える」という抽象的なレベルに設定されており，具体的な目標は各大学に委ねられているという現状にある．

このような状況の下，われわれは，大学英語教育における目的として「学術研究に資する英語教育」を念頭において，そのために役立つ言語資源として「京都大学学術論文コーパス」を構築した．「京都大学学術論文コーパス」は，京都大学の各専門分野の研究者に，学生へ推薦する学術雑誌を選定してもらい，それらの雑誌に掲載された学術論文の中から無作為に抽出した論文から成り立っている．また，われわれは，このコーパスに出現する単語の屈折形を原型に統一し，一般語彙とされる約 2,000 語を除去した各種学術論文データベースも構築した．さらに，これらの学術論文データベースから，各語彙の分野ごとの出現頻度を統計的に処理することによって，学術研究に資する英語教育に役立つ 3 種類の学術語彙リストを作成した

以下，本稿では，大学英語教育の目的を概観した上で，われわれが構築した学術論文コーパスと学術論文データベースについて説明し，作成した学術語彙リストと既存の語彙リストとを比較した結果について述べる．

2 学術目的から見た大学英語教育

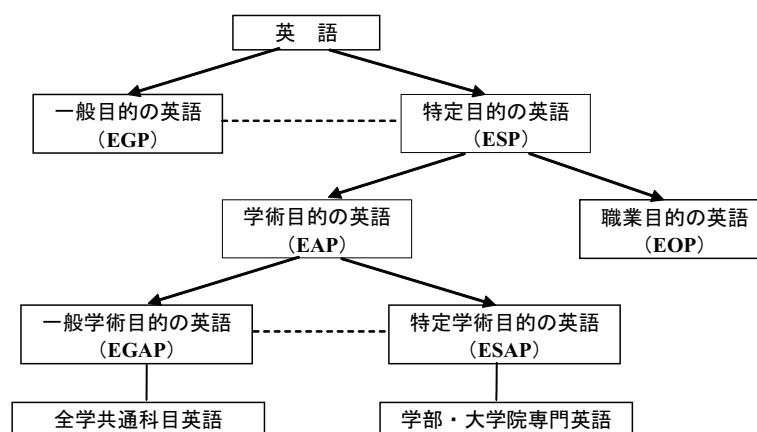
現在，日本の各大学では，独自の教育方針や教育理念に応じた英語教育が実践されているものと考えられる．大学における英語教育を考えるとときに重要となるのは，大学がどのような存在意義を持つかという点である．この時，2006 年に改正された「教育基本法」において新たに規定された大学に関する項目を考慮する必要がある．「教育基本法」の第七条では，次のように大学を規定している．

「大学は，学術の中心として，高い教養と専門的能力を培うとともに，深く真理を探究して新たな知見を創造し，これらの成果を広く社会に提

この観点に基づいて，大学を学術研究の場として考えると，大学英語教育の目的には「学術研究に資する英語教育」が含まれていると言える（田地野，水光 2005）．この意味において，現在，日本の大学英語教育の一つの大きな特徴となっている「特定目的の英語」(ESP: English for Specific Purposes) 研究の知見（大学英語教育学会実態調査委員会 2003; 田中 2007）が参考となる．ESP 研究の知見から大学英語教育のカリキュラムを検討すると，主に 1, 2 年生次を対象とする一般教育課程における英語教育においては，3 年生以降の専門教育課程における英語教育との連携を視野に入れる必要がある．両者の関係を整理するための理論的枠組みとしては，図 1 が参考となる（田地野，水光 2005）．

この理論的枠組みに従えば，「英語」はまず，「一般目的の英語」(EGP: English for General Purposes) と「特定目的の英語」(ESP) に大きく分けられる．そのうち，「特定目的の英語」については，「学術目的の英語」(EAP: English for Academic Purposes) と「職業目的の英語」(EOP: English for Occupational Purposes) とに細分化される．両者の区別がどのようなものであるかについては，例えば同じ医学のための英語 (ESP) を考えてみると理解しやすい．同じ医学英語でも，大学や研究所などでの学術研究に必要な「学術目的の英語」と，病院などでの診療時に，患者との会話で必要となる「職業目的の英語」とでは，使用する語彙や表現の点において，互いに異なった言語的特徴を示す．これは，「学術目的の英語」と「職業目的の英語」とで，専門家集団（プロフェッショナル・ディスコース・コミュニティ）とジャンルが，それぞれ異なっていることに起因する．

「学術目的の英語」は，さらに，「一般学術目的の英語」(EGAP: English for General Academic Purposes) と「特定学術目的の英語」(ESAP: English for Specific Academic Purposes) とに分類される．「一般学術目的の英語」は各専門分野に共通する，一般的な学術的言語技能を対象とする一方，「特定学術目的の英語」は理学や法学といった，ある特定の専門分野において必要となる学術的言語技能を対象とする．したがって，大学の英語教育カリキュラムの観点から考えると，前者の



*点線は連続体を示す。

図1 大学英語教育の目的と分類 (田地野, 水光 2005)

EGAP は、一般教育課程の英語が、後者の ESAP は、学部・大学院における専門課程の英語が、それぞれ対象となると考えられる。ここで注目すべき重要な点は、EGAP と ESAP が、はっきりと区別されるものではなく、図中で点線によって示されているように、連続体をなしており、有機的に関連づけられることが期待されていることである。

この理論的枠組みに基づいて、学術目的の英語教育に資する言語資源として、われわれは英語学術コーパス(「京都大学学術論文コーパス」)を構築した。次節では、このコーパスについて説明する。

3 京都大学学術論文コーパスの構築

英語学習における語彙に関する知識の重要性は、一般にも認められている通りである (Nation 2001; Read 1990)。例えば、英語文献の読解において未知語の割合が、全体の延べ語数の 2 パーセントを超えた場合、その文献の内容を十分に理解することが困難になるという指摘がある (Nation 2001)。また、専門的な話題を扱う英文作成の課題遂行に際して、書き手の持つ専門語彙の知識が、文法といった文構成の言語的側面や、心理的側面に影響を与えることが実験により示唆されている (田地野 2007)。

このように英語学習において、語彙知識は種々の言語技能の基盤となる。特に学問の場である総合研究大学においては、学術文献・論文において高頻度・広範囲に使用される学術語彙の学習がきわめて重要である。そこで、われわれは、京都大学の全 10 学部の教職員・学生の協力を得て、特に学術語彙の学習に資するための言語資源として、「京都大学学術論文コーパス」を構

表 1 各分野ごとの語彙数 (語)

学部	語彙数	学部	語彙数
総合人間学部	1,458,201	理学部	1,168,664
文学部	1,424,208	医学部	533,354
教育学部	1,006,735	薬学部	579,719
法学部	1,884,772	工学部	770,426
経済学部	1,180,444	農学部	640,942
		10 学部合計	10,647,465

築した。学術論文コーパスの構築については、図 2 に示されている手順に従った (田地野, 寺内, 笹尾, マスワナ 2007)。

まず、京都大学の総合人間学部、文学部、教育学部、法学部、経済学部、理学部、医学部、薬学部、工学部、農学部の各専門分野の研究者 (専任教員) の協力を得て、それぞれの分野の学部生・院生に推薦する英語の学術雑誌を 5~15 種類、選定してもらった。次に、選定してもらった雑誌から、主として過去 2 年間に掲載された学術論文を、1 雑誌あたり 7~16 本ずつ無作為に抽出した。抽出した論文を、人手によりテキスト化して、各分野ごとの論文コーパスを構築し、最後に、これらをまとめて全学共通学術論文コーパスを構築した。各分野のコーパスに含まれる語数を、表 1 に示す。

4 各種学術語彙リストの作成

われわれは、さらに、前節の図 2 の手順にそって、学術論文コーパスから種々の学術論文データベースを構築した。まず、各分野のコーパスに存在する単語を、

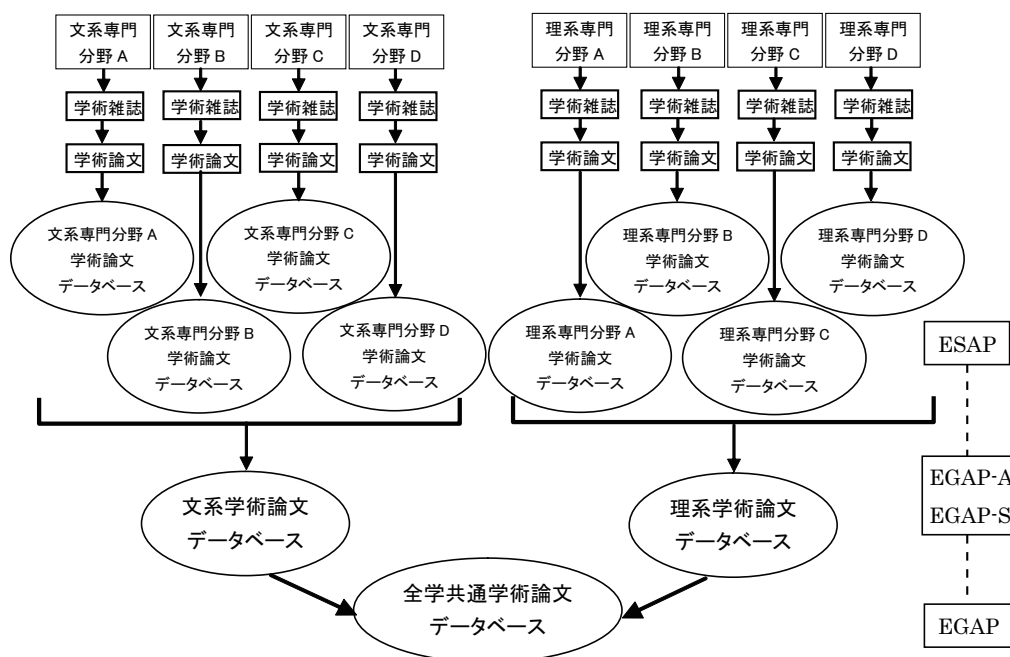


図2 学術論文コーパスの構築手順 (田地野ほか 2007)

人手により屈折型を原型に統一し、General Service List(West 1953)に掲載されている一般語彙(約2,000語)を取り除いて、各分野ごとの論文データベースを、次いで、これらを文系、理系ごとにまとめた論文データベースを、最後に両者をまとめた全学共通学術論文データベースをそれぞれ構築した。

ここで構築した論文データベースを用いて、学術語彙の種類に応じた種々の語彙リストを作成した。図1のEGAPとESAPとの関係性を参考にすれば、学術語彙は、少なくとも、以下に示す4種類に分類できる(田地野 2008)。

1. EGAP 語彙 ... 特定の専門分野に関係なく学術文献に共通して使用される一般学術目的英語語彙
2. ESAP 語彙 ... 特定の専門分野の文献に特徴的に使用される特定学術目的英語語彙
3. EGAP-A 語彙 ... EGAPとESAPとの中間に位置する文系(Arts)専門分野に共通する文系共通学術語彙
4. EGAP-S 語彙 ... 理系(Sciences)専門分野に共通する理系共通学術語彙

他にも、人文系、社会科学系、医薬系、理工農系、および生命科学系(ライフサイエンス辞書プロジェクト 2005)など、学際的な語彙分類による語彙リストの開発も考えられる(田地野 2008)が、今回は、この分類

に従い、各分野のESAP語彙を除いた、EGAP語彙リスト、EGAP-A語彙リスト、EGAP-S語彙リストの3種類の語彙リストを作成した。

今回の作業においては、文系分野を「総合人間学部文系、文学部、教育学部、法学部、経済学部」の5学部、理系分野を「総合人間学部理系、理学部、医学部、薬学部、工学部、農学部」の6学部として設定した。

まず、EGAP語彙リストの作成手順について説明する。EGAP語彙とは、文系・理系にかかわらず出現する語彙であると言えるため、文系分野と理系分野に偏りなく、高頻度に出現している語彙が相応しい。そこで、文系学部3学部以上かつ理系学部3学部以上に、1回以上出現するものをEGAP語彙の候補となる条件として設定し、これを満たす語彙を抽出した。抽出した語彙の文系分野での出現頻度数と理系分野での出現頻度数を集計し、両者の調和平均を算出して、値の大きい方から上位2,000語に対し、今回は教育的観点に基づく判断を手で行って、EGAP語彙を選別した。調和平均の値を基準としたのは、文系分野と理系分野の偏りが、なるべく少ない方が値が大きくなるためである。選別後のEGAP語彙は、ワードファミリー換算で477語となった。

次に、EGAP-A語彙リストとEGAP-S語彙リストの作成手順について説明する。学術語彙から先に出出したEGAP語彙を除外したうえで、文系・理系ごとに語彙の補完類似度を算出した。補完類似度は、特定分

表 2 EGAP 語彙と JACET8000 との重なり (%)

	JACET 8000		リスト外
	1～4000	4001～8000	
EGAP	43.0	39.2	17.8
EGAP-A	15.4	53.8	30.8
EGAP-S	9.9	30.3	59.8

野に偏って出現する単語の特徴度を示す尺度として、他の尺度と比較して有効性の高いことが知られている(内山, 中條, 山本, 井佐原 2004)。次に, 補完類似度の値の大きい方から上位 1,500 語ずつ, EGAP と同様に, 教育的観点から, 人手で EGAP-A, EGAP-S 語彙を選別した。選別後の EGAP-A 語彙, EGAP-S 語彙は, ワードファミリー換算でそれぞれ 312 語, 323 語であった。従って, EGAP 語彙, EGAP-A 語彙, EGAP-S 語彙の 3 種類を合わせた広義の EGAP 語彙は 1,112 語となった。

5 各種語彙リストの分析

われわれは, 前節で作成した, EGAP 語彙, EGAP-A 語彙, EGAP-S 語彙の各リストと, 信頼性の高い語彙リストとして定評のある「JACET8000」(大学英語教育学会基本語改訂委員会 2003) との重なりを調査した。この結果を, 表 2 に示す。

表 2 によると, EGAP 語彙の約 8 割, また EGAP-A 語彙の約 7 割が「JACET8000」の中に含まれており, 文系の学生には「JACET8000」が専門課程へと繋がる語彙の学習に対し, 有効となりうることが予想される。ただし, 「JACET8000」でカバーできない EGAP 語彙や EGAP-A 語彙も存在するため, 「JACET8000」が万能というわけではない。一方, EGAP-S 語彙は「JACET8000」の中に 4 割程度しか存在せず, リスト外の語が多く含まれていることが分かる。従って, 「JACET8000」を語彙学習の教材として指導したとしても, 200 語近い EGAP-S 語彙は未習のままとなってしまう。このような現状を考えると, 理系学生を対象とした, 理系分野の基礎語彙を指導するための整備・充実が急務であると言える。

6 おわりに

本稿では, 「特定目的の英語」研究の知見に基づき, われわれが構築した大学の英語教育に資する種々の言語資源を紹介した。今回作成した 3 種類の語彙リスト

は, 実際の学術論文を基にしているため, これまでの語彙リストに比べ, 学術目的の英語語彙教育に対してより有効である可能性が高い。特に, これまでの語彙リストは文系分野に偏っていると思われるため, 理系分野の学術語彙 (EGAP-S) までカバーしている本語彙リストは, 有用であると思われる。

現在, われわれは, 構築した学術論文コーパスを基にして, 英語教育に有効となる, さらなる言語資源の作成・研究を行っている。今後は, 人文系, 社会科学系, 医薬系, 理工農系, および生命科学系といった学際的な語彙分類による語彙リストなども視野に入れて, 開発を進めていく予定である。

参考文献

- 大学英語教育学会基本語改訂委員会 (編) (2003). 大学英語教育学会基本語リスト JACET List of 8000 Basic Words. 大学英語教育会.
- 大学英語教育学会実態調査委員会 (2003). わが国の外国語・英語教育に関する実態の総合的研究 - 大学の外国語・英語教員個人編 - . 丹精社.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press, Cambridge.
- ライフサイエンス辞書プロジェクト (編) (2005). ライフサイエンス必須英和辞典. 羊土社.
- Read, J. (1990). *Assessing vocabulary*. Cambridge University Press, Cambridge.
- 田地野彰 (2007). “専門分野のライティングにおける語彙知識の重要性 - 経済ニュースの記事の和文英訳を通して - .” 西堀わか子, 田地野彰 (編), 英語の授業実践研究, pp. 67-73. 奈良女子大学国際交流センター.
- 田地野彰 (2008). “理系学生に必要な語彙とは - 学術語彙データベースに基づいて - .” 英語教育, 57 (3), pp. 16-19.
- 田地野彰, 水光雅則 (2005). “大学英語教育への提言 - カリキュラム開発へのシステムアプローチ - .” 竹蓋幸生, 水光雅則 (編), これからの大学英語教育, pp. 1-46. 岩波書店.
- 田地野彰, 寺内一, 笹尾洋介, マスワナ紗矢子 (2007). “総合研究大学における英語学術語彙リスト開発の意義 - EAP カリキュラムデザインの観点から - .” 京都大学高等教育研究, 13 (1), pp. 121-131.
- 田中慎也 (2007). 国家戦略としての「大学英語」教育. 三修社.
- 内山将夫, 中條清美, 山本英子, 井佐原均 (2004). “英語教育のための分野特徴単語の選定尺度の比較.” 自然言語処理, 11 (3), pp. 165-197.
- West, M. (1953). *A general service list of English words*. Longman, Green & Co., London.