

学術論文読解を目的とした指導語彙の選定 An ESP Approach to Teaching Academic Vocabulary

キーワード：ESP、語彙、コーパス分析

石川有香・小山由紀江

ISHIKAWA Yuka & KOYAMA Yukie

1. はじめに

これまでのESP (English for Specific Purposes) 教育研究におけるニーズ分析の結果から、工学系大学においては、学術論文を読む力が広く求められていることが分かっている(中野, 2005; 小山, 2001 他)。論文読解力育成にはESP 語彙指導が不可欠であるが、本研究では、共通教育の段階で指導すべき語彙の選定方法を提示することとしたい。従来は、ジャンルに特化した語彙指導の内容や方法は教師の勘に頼っていたが、近年、当該分野のテキスト・コーパスに基づいて作成した語彙表を使用することが多くなっている。本稿でも、電気・機械・土木・生命物質など、様々な工学分野のテキストから構築した科学技術論文コーパスを分析し、語彙表を作成して、論文読解に必要な語彙を特定する。さらにその結果を既存の教育語彙表(JACET, 2003; Coxhead, 2000)と比較することによって、工学部の共通科目である英語授業において指導が必要な語彙を選定してゆく。

2. 論文テキストの語彙

Nation (2001)は、通常、テキストの約5%をいわゆる低頻度語が占めると言う。この語群には、どのテキストにも1度しか出てこないhapax legomenaを始め固有名詞や古語などが含まれるが、これらは非常に種類が多いため、習得の努力に比べてその成果はごく小さなものとなる。従って、低頻度語に関しては、むしろ、文脈から語義を推測するストラテジーを指導する方が効果的であるとされる(p. 19)。そのため、語彙指導の観点からは、低頻度語を除き、論文テキストに含まれる語を(1)基本語彙、(2)準専門語彙、(3)専門語彙、の3種に分類して論じることが多い。しかし、これら3種の定義については、先行研究の間でも一致を見ていない部分もある。本研究では、まず、上記3種の語彙定義を以下のように定める。

- (1) 基本語彙：論文テキストだけでなく、どのテキストにも繰り返し現れる高頻度語
- (2) 準専門語彙：専門分野に関わらず、さまざまな分野の論文テキストに共通に含まれている語彙で、基本語彙以外のもの。すなわち、論文テキストには高頻度で表れるが、他のテキストでは使用頻度が低く、論文テキストの特徴となる語
- (3) 専門語彙：特定の専門分野でのみ使用される語

ここで、準専門語彙と専門語彙について、その定義をさらに明確にしておきたい。この2種の語彙の分類については先行研究においても曖昧な場合が多く、境界線が明確ではないとされる(Mudraya, 2004)。先行研究の中には、準専門語彙をAcademic Word List (Coxhead, 2000) [以下AWL]と同義で使用するもの

も散見される。AWLは、ワード・ファミリー換算で570語からなる語彙群だが、学術テキストならば、通常、約10%をカバーするという(Coxhead, 2000)。一方で、新聞テキストは約4%(Chung & Nation, 2003)、また小説テキストは約2%(Coxhead, 2000)しかカバーしておらず、まさに、学術テキストの特徴語であるとされる。

ところが、論文テキストを科学分野に限定してみると、カバー率がかなり低くなることが指摘されている(Coxhead, 2000; Chung & Nation, 2003; 石川, 2006)。さらに、約180万語の科学技術論文テキストに2回以上出現する語を対象に、AWLの効率性を示す、いわゆるヒット率を調べた石川(2006)によれば、ETHIC, IMMIGRATE, GENDER, IDEOLOGY, LEGAL, LIBERAL(以下、大文字で表された語はレマ化されたものとする)など、AWLの570語のうち70語がまったくヒットしなかったと言う。これは、AWLが作成された際の基準コーパスが、芸術・法学・商学・科学の4分野からなるため、科学テキストに使用される語彙は4分の1しか反映されておらず、AWL自身が文系の分野に偏ってしまっているためであると推測される。石川(2006)では、AWLでヒットしなかった70語を科学技術論文で高い頻度で使用される語に入れ替えるだけで、カバー率が2.5%上昇することも指摘されている。

本稿では、工学系学術論文テキストのみからなる学術論文コーパスを独自に作成し、語彙表を作成するわけであるが、ここでは、3つ以上の工学系専門分野にまたがって、学術論文テキストに特徴的に表れる語彙群を、「準専門語彙」とすることとしたい。また、工学系学術論文の中でも、機械・生物・電気など、工学の特定分野でのみ使用される語は「専門語彙」とする。したがって、人文・法学などの分野で学術論文に使用される頻度が低い場合でも、工学系の多分野で高頻度で使用されている場合は、「準専門語彙」と定義している。各語彙の分類を図1に示す。

3. ESP教育と共通英語科目の語彙指導

我々は上で、論文テキストに表れる語彙の種類を確認した。では、これら3種のうち、共通科目としての英語授業において取り上げるべき語彙群とは、どういったものであろうか。ここでは、ESP教育の枠組みの中から考えてみたい。石川(2006)では、ESP教育のさまざまな定義を先行研究より概観し、大学院進学率の高い工学系大学で行われているESP教育の多くは、いわゆるEST(English for Science and Technology)教育ではなく、EAP(English for Academic Purposes)教育に分類されることを示した。本稿でも、こうした視点から工学系大学のESP教育を捉えることとする。

Dudley-Evans and St. John(1998)では、“ESP is defined to meet specific needs for the learner”が、ESP教育を決定づける主要な特徴としてあげられている。彼らに従うと、ESP教育では学習者や特定分野のディスコース・コミュニティなどのニーズを調査し、その結果に基づいて、コース・デザインを決定してゆくこととなる。上に述べたように、工学系大学においては、学術論文読解力の育成に対するニーズが高くなっているのだが、学術論文にはさまざまな種類の語彙が使用されており、論文読解力の育成を目的にして共通科目の英語授業で語彙指導を行う場合に、いったい、どのような語彙を選定してゆくべきであるかという問題に関して、これまで十分な議論がなされてきたとはいえない。

そもそも、工学系論文テキストと一口に言っても、そこには、機械工学・生物工学・情報工学などのさまざまな分野が存在している。さらに高度に細分化された専門学術分野では、特定のディスコース・コミ

ユニティーでのみ流通する語彙が用いられることが少なくない。しかしながら、我々はここで、こうした分野をすべて一纏めにし、工学系論文テキストというテキスト・ジャンルを設定し、その上で、どのような語彙を、共通教育で指導してゆくべきであるかを論じてゆくこととしたい。

先に、我々は、ニーズ分析が ESP 教育のコース・デザインに関与していることを確認した。このほかにも、Duddley-Evans & St. John (1998)は、コース・デザインに影響を与える要因として、現場の事情、教員の2つを挙げている。ここでは、これらの観点から ESP 語彙学習に適した語彙の選定方法を考えてゆく。

共通教育においては、学生の専攻別にクラス編成を行う場合もあるが、習熟度別クラス編成などでは、1つのクラスがさまざまな専攻の学生で構成されることになる。特定のジャンルでのみ使用される語彙や表現は、他分野を専攻する学生にとって、学習価値の低いものである。また、学習者の専攻が同じであっても、専門教育を十分に受けていない大学1年生・2年生の学生に対して、専門の知識のない英語教員が専門語彙・専門表現を指導するという授業形態は、学生にも教員にも、大きな負担を強いることになる。確かに、ある程度の数の専門語彙を導入することで、学生の学習意欲が高まるという一面もあるものの、こうした現場の事情・教員の事情がある場合には、労を上回る効果は期待できないことも、予測できよう。

現在、多くの工学系大学では、1・2年生に対する共通教育や、また、3・4年生や大学院生に対しても共通教育として設置されている英語科目は、主に英語教員が担当している。一方、論文テキストの読解指導は、ゼミや研究室単位で、主に専門教員が担当する専門科目として行われている。こうしたカリキュラム構成の下で共通英語科目に求められているニーズとは、論文テキスト読解のための基礎的英語力を育成するということになる。だとすれば、共通教育での語彙指導は、工学系のどの専攻分野でも論文読解時に必要となる「準専門語彙」と、「基本語彙」の中でも特に未習のものを中心に指導を行い、分野に特徴的な「専門語彙」は、専門教育の中で指導するべきであると言えよう。次節では、具体的な学術コーパスの構築と分析を踏まえ、将来どの専門分野に進むことになるようにも、工学部の学生ならば同様に、英語論文読解時に必要となるであろう語彙群を抽出し、共通英語で指導するべき語彙を選択してゆきたい。

4. 論文コーパス

本研究では、各学科の工学専門教員の協力を得て、工学系専門誌の中でも、より広範な分野に関係する論文誌を選択してもらい、論文コーパスを作成した〔以下 ASCJ コーパス〕。論文は、題名、著者名、所属、図表、数式、謝辞、参考文献などを手作業で削除し、本文のみをコーパス・データとしている。分析ツールには、Oxford WordSmith Tools 4.0 を使用し、各サブ・コーパスのサイズに差が出ないように、Wordlist に使用された Token 数が 520197～520860 となるように調整した。コーパスの概要を表1に示す。

4. 1 作業手順

ここでは、まず、各サブ・コーパスから頻度別ワードリストを作成した。染谷 (1998)のレマリストを用い、また、英米の綴字の差のある ANALYZE や BEHAVIOR などは、それぞれ ANALYSE や BEHAVIOUR にまとめた。

ところで、上で見たように、低頻度語とはテキストの5%を占めるとされる。そこで、頻度の高いものから順に並べ、累積頻度構成率が各サブ・コーパスの95%を超えたところまでを次の分析の対象とした。例

えば、生物コーパスでは、95%を初めて超えた語は、INCREASINGLY であり、これは生起回数が 11 であったため、10 以下の語を分析対象から除外している。同様に、電気コーパスは 11 以下、機械コーパスは 16 以下、物理コーパスは 15 以下、構造コーパスは 12 以下を除外した。

表1 ASCJ コーパス概要

雑誌	合計
構造 Engineering Structure 他 1 誌	520,656
電気 International Journal of Electric Power 他 5 誌	520,197
機械 Journal of Applied Mechanics	520,539
物理 Applied Physics	520,860
生物 Bio Engineering	520,707

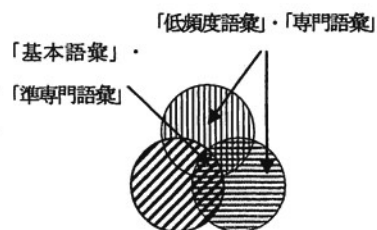


図1 工学系専門分野テキストの重なり語彙

次に、AWL の作成手順を踏まえ、出現したサブ・コーパスの数、すなわちレンジ数を分析の対象に加えることとした。各語のレンジを調査し、レンジ数が 2 以下の語を除外した。例えば、生物コーパスでは、200 以上の頻度の語語彙群中で、除外されたものは、以下のような語となる。

レンジ数 2	BIOMASS, SLUDGE, BIOREACTOR, DISSOLVE, ETHANOL, WASTEWATER, ANAEROBIC, AQUEOUS, MICROORGANISMS
レンジ数 1	CULTURE, PROTEIN, ENZYME, GLUCOSE, FERMENTATION, ADSORPTION, LYSOZYME, REFOLDING, AEROBIC, BIOFILM, IMMOBILIZE, LIPASE, PHENOL, FLASK, INHIBITION, HYDROLYSIS, YEAST, CHITOSAN, ACETATE, ENZYMATIC, BACTERIA, EFFLUENT, UPTAKE, NUTRIENT, CULTIVATION

一見して、明らかに生物工学の専門性が高い語が並ぶ。念のために、一般英語の教育語彙表である JACET8000 で確認してみても、大半の語は記載されていない。FERMENTATION, ADSORPTION, LYSOZYME, REFOLDING, AEROBIC, BIOFILM, IMMOBILIZE, LIPASE, PHENOL, FLASK, HYDROLYSIS, CHITOSAN, ACETATE, ENZYMATIC, EFFLUENT, UPTAKE, BIOMASS, SLUDGE, BIOREACTOR, WASTEWATER, ANAEROBIC, AQUEOUS, MICROORGANISMS は生物工学特有の語となる。しかし、上表のうち 11 語は、JACET8000 にも記載されていた。それらは、YEAST, ETHANOL, GLUCOSE, NUTRIENT, CULTIVATION, INHIBITION, ENZYME, BACTERIA, DISSOLVE, PROTEIN, CULTURE である。こうした 11 語は、確かに一般英語でも使用頻度の高いものであるが、工学系の論文で使用された場合には、意味が異なってくる場合が多い。例えば、CULTURE は、細胞などの「培養・培養物」を指し、DISSOLVE は「溶解」、INHIBITION も「化学反応の抑制・阻害」となる。我々は 2 章で「専門語彙」を「特定の専門ジャンルでのみ使用される語」と定義したわけだが、上記語群のように生物工学というジャンルに依拠した語彙は、専門語彙として専門授業で取り扱うべき語とし、工学系

大学生の論文読解力育成を目的とした共通英語の語彙教材では、取り上げないこととする。

4. 2 結果

以上の手順で作成した語彙表は、単に、生起頻度を合計したものではなく、専門語彙や低頻度語彙を除いたため、工学系の論文に共通して使用されている、基本語彙と準専門語彙から成り立っていると考えられる。語彙表の30位までを下の表2に示す。高頻度語は、どの専門分野にも使用されるため、これらの語のレンジ数はすべて、5となっている。

表2

順位	語	頻度	レンジ	順位	語	頻度	レンジ	順位	語	頻度	レンジ
1	THE	231095	5	11	AS	19438	5	21	FIG	9236	5
2	OF	102881	5	12	THIS	19096	5	22	WERE	8937	5
3	AND	67316	5	13	BY	18550	5	23	IT	8794	5
4	IN	59482	5	14	ARE	18541	5	24	SHOW	8756	5
5	A	56708	5	15	BE	17174	5	25	CAN	8545	5
6	TO	50110	5	16	AT	16062	5	26	WHICH	7999	5
7	IS	43519	5	17	WAS	15046	5	27	MODEL	7468	5
8	FOR	30428	5	18	ON	13981	5	28	VALUE	7293	5
9	WITH	20515	5	19	USE	13056	5	29	B	7183	5
10	THAT	20206	5	20	FROM	11844	5	30	RESULT	7105	5

この段階で得られた語彙表は、異なり語数で1963語となる。全リストは、EGSP Word List Ver.1[以下EGSPリスト1]として<http://www11.ocn.ne.jp/~yuka/egsp.html>に示す。

ところで、コーパス内のすべての語を頻度順に並べた語彙表と比較した場合、レンジを勘案することで、削除できた語群とは、どのようなものだったのだろうか。頻度別語彙表を作成し、累積頻度率が80%となる、頻度が543までの高頻度語の中で、削除された語を見てみると、以下のようなものとなっている(カッコ内は生起頻度数)。

BLADE (1233), ROTOR (1033), CULTURE (1001), PROTEIN (984), STIFFNESS (973), ENZYME (966), BC (825), CONCRETE (825), BIOMASS (757), WAKE (713), COMBUSTION (673), EARTHQUAKE (665), PLASMA (645), GLUCOSE (645), STATOR (635), EV (630), GROUND (608), FERMENTATION (599), GRAIN (599), SEISMIC (572)

上記の語群の中には、一般的英語にも、しばしば使用される語も散見される。しかしながら、上でも述べたように、論文テキストを見てみると、これらの多くが限定された用法となっていることが分かる。例えば、BLADEは、電気分野でも59回出現しているが、機械分野での出現が大部分を占め、1161回と高い

頻度を示している。コロケーションを見てみると、どちらの分野でも、“rotor blade”が多く、機械では、“impeller blade”，“turbine blade”などが続き、固定的な表現となっていることが分かる。こうしたものは、特定分野の1ジャンル、または、近い分野のみにまたがった2ジャンルに限定されているので、ここでは専門語彙とみなす。

4. 3 検証

以上のような手順で、工学系論文テキストに共通して現れる傾向の語彙を取り出した。これらは、工学系論文の基本語彙と準専門語彙と考えられる。ここでは、さまざまな種類のテキストに対する、EGSP リスト1のカバー率を見ておきたい。対象とするテキストは6種である。まず、科学論文テキストは、ASCJ コーパスとは別に、今回の分析で用いた論文誌数種を使用して、同手法で収集したものを使用した。第2に、科学専門誌テキストは、*Nature*より収集した。第3に、科学一般誌テキストとしては、科学全般を扱い、一般読者を対象として米国で発刊されている月刊誌 *Scientific American* から収集したテキストを用いる。科学論文テキストの場合と同様に、どちらも図・表等は除いている。第4に、高校教科書テキストは、*The CROWN English Series II New Edition*, *ACORN English Course II*, *English 21 II*, *APRICOT ENGLISH COURSE II*より収集した。第5に、新聞テキストは、*The New York Times*から収集し、最後に、児童文学は、Project Gutenberg より Frances Hodgson Burnett の *A Little Princess* の前半部分を使用している。すべて語数を3万語に統一した。語彙リスト1963語のカバー率の結果を表3と図2に示す。

表3

	カバー率
科学論文	83.32
科学専門誌	77.02
科学一般誌	73.48
高校教科書	67.65
新聞	64.93
児童文学	62.45

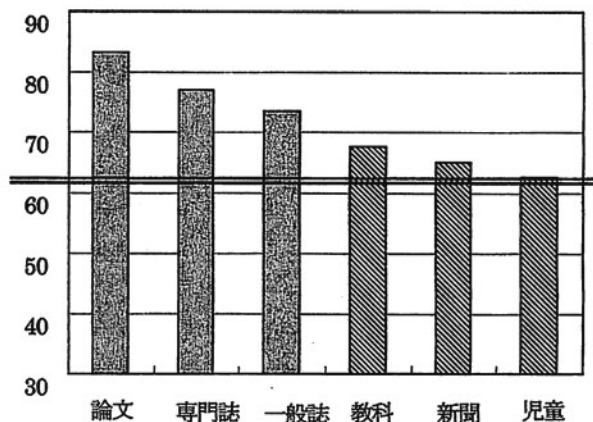


図2

図2に見るように、EGSP リスト1は、科学論文テキストを最も広くカバーしているが、同時に、右半分の一般的なテキストよりも、科学専門誌・一般科学誌を含め、左側の3つの科学テキストを広くカバーしていることが分かる。さらに、児童文学・新聞・高校教科書などのテキストにも共通に使用されている、基本語彙の部分をEGSP リスト1から取り除き、図2の二重線よりも上の部分に議論を限定すると、工学系に特化した準専門語彙の抽出も可能であるように思われる。次に、EGSP リスト1を加工して、工学系大学生に共通教育で指導すべき語彙の選定を行い、教育用語彙リストの作成を試みることにする。

5. 指導語彙リスト

語彙表を元に教材作成を行うにあたっては、特定コーパスの語彙分析だけに終始するのではなく、教育目的に則って効果的な教育語彙表に加工してゆくことが肝要となる。例えば、JACET8000 や SVL12000 などの既存語彙表は、BNC の語彙分析を元に行っているが、日本人学習者のために、それぞれ独自の視点から加工が行われている。ここでは、大学生を対象とし、共通教育で、論文読解力の育成を目的とした ESP 語彙学習に効果的な語彙表の作成をめざし、EGSP リスト1の加工を行う。

まず、本稿では、教材化を行う際の下準備として、以下の作業を行った。

- ・DISPLACEMENT と DISPLACEMENTS など、e-lemma に記載されていない語のレマ化、
- ・IS, ARE, WAS, WERE など、be 動詞と、HAVE, HAS, HAD のレマ化、
- ・アルファベットの1文字の削除、
- ・NON-, PRE-, MULTI-など、ハイフンでつながれた接頭辞の削除、

以上の作業で、リストの数は、1926 語になった。

ところで、教育現場で使用する語彙リストを作成する場合に、どのような語彙を既習語として省くかが問題となる。例えば、JACET8000 は、250 語を基本語として除き、別表として付記している。AWL は、GSL の高頻度 2000 語の後に学習する教育語彙表であるので、コーパス語彙分析によって作成されたリストから、GSL の高頻度 2000 語を除いている。特に科学技術英語テキストにおいては、語が、一般英語テキストとは異なる用法・意味で使用される傾向があることを踏まえ、ここでは、JACET8000 のレベル 1 にあたる、上位 1000 語のみを既習語をすることにした。既習語を取り除いた後の語彙表は EGSP リスト 2 として、上記のサイトで公開している。30 位までを下の表 4 に示す。

表 4

Ran	Word	Freq	Ran	Word	Freq	Ran	Word	Freq
1	FIG	9236	11	FUNCTION	3312	21	DETERMINE	2503
2	TEMPERATURE	5437	12	DUE	3257	22	RATIO	2499
3	FLOW	4747	13	ANALYSIS	3109	23	COEFFICIENT	2494
4	OBTAIN	4443	14	SOLUTION	2882	24	DISTRIBUTION	2453
5	LOAD	4288	15	EQ	2844	25	RESPECTIVELY	2411
6	STRUCTURE	3993	16	PHASE	2725	26	MODE	2342
7	DATA	3955	17	LAYER	2709	27	CONSTANT	2247
8	CONCENTRATION	3913	18	FREQUENCY	2707	28	RANGE	2237
9	PARAMETER	3856	19	OBSERVE	2618	29	FACTOR	2235
10	SAMPLE	3357	20	EQUATION	2570	30	MAXIMUM	2227

この EGSP リスト 2 に記載された語は、ワード・タイプ換算で 1341 語となる。上記で EGSP リスト 1 の検証に使用した 6 種のテキストを再び用いて、カバー率を調査した。結果を図 3 に示す。科学論文テキス

トでは23.4%、専門誌テキストは18.8%とカバー率は高くなっているが、一般英語である高校教科書は2.7%、新聞は4.5%、児童文学は2%と、かなり低い。また、科学一般雑誌は10.7%と、比較的高くなっている。したがって、EGSP リスト2は工学系論文に特化した準専門語彙であると同時に、一般的な科学テキストにも特徴的な語彙を包含した、共通教育にふさわしい教育語彙リストと思われる。今後は、ESP 教育の他のニーズにも即した語彙表との融合を図り、共通教育で指導すべき語彙教材を作成してゆきたい。

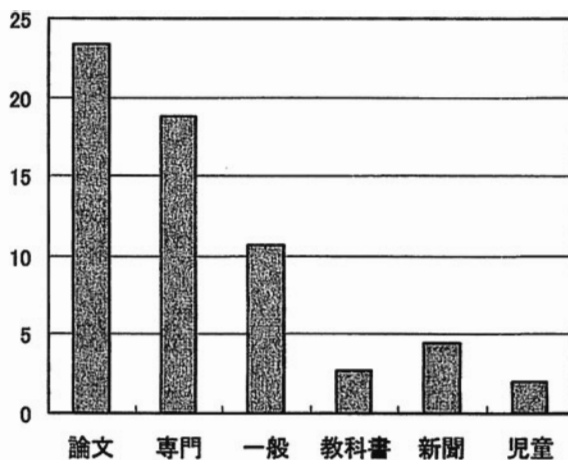


図3

(名古屋工業大学)

引用文献

- Chung, T., & Nation, I. S. P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-116.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.
- Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- 石川有香 (2006) 「工学系学術テキストコーパスにみる語彙の特徴」『広島女学院大学大学院言語文化論叢』第9号 127-143頁
- 小山由紀江 (2001) 『工学系大学における英語教育の現状解析と効率的システムの構築』(平成10年度～平成12年度 科学研究費補助金基盤研究(C) -1 研究成果報告書)
- Mudraya, O. V. (2004). Need for data-driven instruction of engineering English interface. *IEEE Transactions on Professional Communication, 47*(1) 65-70.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Massachusetts: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- 中野秀子 (2005) 「シラバスデザイン・教材例—九州共立大学工学部編—」横山彰三『ESP教授法に基づく大学専門英語教育のための効果的シラバスと教材開発の研究』(平成15年度～平成16年度科学研究費補助金(基盤研究(C))研究成果報告書) 27-33頁
- 染谷泰正 (1998) *English Lemma List (Version 1)* [Data file]. Available from Yasumasa Someya's Web site, http://www1.kamakuranet.ne.jp/someya/e_lemma.zip
- West, M. (1953). *A general service list of English words*. London: Longman.