

これからの英語語彙指導：単語の精選と効果的提示

－ コーパス処理に基づく高校生用英単語集の編纂をふまえて －

石川 慎一郎

要旨

近年急速な発展を見せているコーパス処理の手法を応用すれば、大量の言語データに基づいて信頼性の高い語彙リストを構築することができる。しかし、単に高頻度語を精選して並べるだけでは語彙学習の実はなかなかあがらない。従来の語彙指導は「何を覚えるべきか」を示して終わる場合が少なくなかったが、学習者の動機や能力の多様化をふまえ、これからは、覚えるべき語を「どう効果的に提示するか」という問題もまた真剣に考えてゆく必要があるだろう。

日本人学習者の英語語彙学習の実態を考えた場合、高等学校で使用する英単語集が非常に大きな影響を持っていることに気づく。本稿では、高校生用英単語集の編纂過程を紹介しながら、学習語彙の精選とその効果的な提示方法のありかたについて具体的に検証してゆきたい。

はじめに：日本人学習者と英語語彙表

英語力の基礎として語彙力が重要であることは言を待たないが、日本の英語教育においては、体系的な語彙指導がほとんど行われていない。たとえば中学校では、授業内で新出単語の発音や意味を指導するが、3年間で扱う語はわずか900語程度である。高等学校では、大学入学試験準備の必要もあって、何らかの英単語集を副教材として指定する機会が多いが、実際の語彙学習は生徒に一任しているのが現状である。大学においては、語彙指導自体がほとんど存在しない。

こうした日本人学習者の語彙学習の現状をふまえると、高等学校の段階で使用される英単語集が非常に大きな重要性を持っていることに気づく。大学入試を目指して単語集で集中的に学習した語彙が学習者の将来の英語語彙力の根幹となるのである。このことは、信頼性の高い高校生用英単語集の開発の必要性を示す。

単語集あるいは語彙表を開発する際には、信頼できる語彙頻度データを利用することがなによりも重要となる。実際、近年のコーパス処理技術の進歩に伴い、コーパス準拠型の語彙表も増えてきた。たとえば Leech, Rayson, &

Wilson (2001) は、1 億語の British National Corpus (BNC) に基づくきわめて信頼性の高い語彙リストである。また、JACET 基本語改訂委員会 (編) (2003) は、日本人学習者にとって重要性の高い言語資料 (中学校・高等学校・大学の教科書、各種資格試験など) を収集した独自のサブコーパスと BNC の頻度データを統計的に比較・調整して作られた教育語彙表である。しかし、前者は研究者用、後者は主として大学生用であって、高校生用英単語集へのコーパス活用の試みははまだ限定的なレベルにとどまっている。

以下では、単語選定・語義選定・用例選定の全面にわたってコーパスを活用した石川 (2004b) の編集過程をふりかえりながら、広く中高大におけるこれからの語彙指導の方向性について考えてゆきたい。

コーパスの構築と頻度データの処理

コーパスの価値は単にサイズが大きいことではなく、利用目的との合致性で決まる。今回は、大学入試を目指す高等学校中・上級生用の単語集を開発するという目的から、過去 4 年間の国内各大学の入試問題データを収集し、コーパス化した。データサイズは延べ語数 (tokens) で 5,411,927 語、異なり語数 (type) で 45,864 語である。

さらに個々の type ごとにその頻度を計算し、エラー語 (選択肢記号の a, b, c など) などを除外して、作業の基礎となる語彙頻度リストを作成した。同リストを概観してわかることは、構成語彙が一握りの高頻度語と無数の低頻度語に二極化するということである。これは入試英文の特徴というよりも、自然言語に共通する傾向性である。下記は、頻度上位 10 語のリストである。

なお、表中の Raw Freq は単純頻度、PMW Freq は百万語あたり (Per Million Words) の換算頻度、Share は全体に占める累計占有率を示す。頻度順位 3 位の“to”を例にすると、全 5,411,927 語のコーパスにおける単純頻度は 156,919 回であるが、分母を 100 万語とした場合の換算頻度は 31073 回となる。そして、頻度順位 1 位の“the”から頻度順位 3 位の“to”まで、上位 3 語の構成比の累計は全体の 11.25%となる。

表 1. 上位 10 語の頻度

Rank	Token	Raw Freq	PMW Freq	Share
1	the	283809	52441	5.24
2	a	168163	31073	8.35
3	to	156919	28995	11.25
4	of	143660	26545	13.91
5	and	111893	20675	15.97

6	in	101946	18837	17.86
7	is	67973	12560	19.11
8	that	65351	12075	20.32
9	I	58481	10806	21.4
10	it	54982	10159	22.42

表 1 を一瞥するとわかるように、ランク 1 位の定冠詞“the”の PMW 頻度は 5 万回を超えるが、順位が下がるにつれて頻度は急速に減少し、ランク 10 位となる“it”の頻度は“the”の 5 分の 1 に過ぎない。こうした急激な頻度下降は、上位 200 語の PMW 頻度変化を示した図 1 から明らかである。

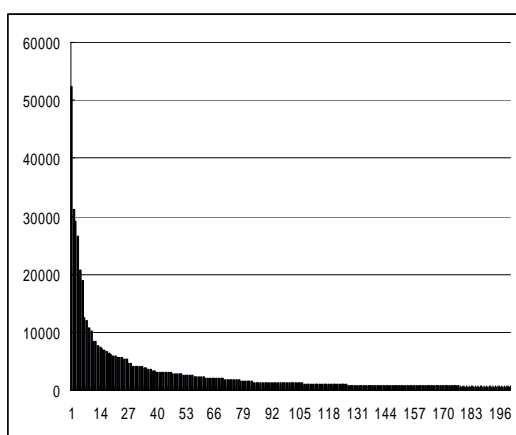


図 1 . 上位 200 語の頻度

単語集の有効性を示す指標として、英文に対する「カバー率」がよく問題にされる。カバー率をあげるためには低頻度の語まで幅広く収録すればよいわけであるが、無限に収録語数を増やすことは現実的でない。また、学習効率という点を考えると、出来る限り無駄な語を省き、一定の「ヒット率」を確保する必要もある（石川 (1998)）。

結局、基本となる頻度リストのなかで、どのレベルの語までを収録対象とするかは難しい判断にならざるを得ない。表 2 は 1000 語刻みの境界語とその語の頻度、およびランク 1 位の語から当該語までの累計占有率 (%) を示している。

表 2 . 各占有率の該当語リスト

Rank	Token	PMW freq	Share
1000	stopped	112	75.99
2000	anxiety	52	83.57
3000	tour	32	87.55
4000	loudly	21	90.11
5000	pointing	15	91.88
6000	openly	11	93.19
7000	hybrid	9	94.20
8000	boards	7	94.99

上表より、カバー率 90%を得るためには最低でも約 4000 語が必要であり、カバー率を 90%からさらに上げるためには、大幅に語を増加しないといけな

いことがわかる。今回は、カバー率とヒット率の妥当な接点を求め、頻度リストで上位 4500 語を収録範囲と決めた。しかし、その中には一般の高校生にとって既知と判断される基本語が散在している。そこで 4500 語からそれらを削除し、全体で 1161 語を見出し語とした。具体的には 1001~2000 位が 401 語、2001~3000 位が 390 語、3001~4000 位が 370 語、4001~45000 位が 45 語で、これらで全体の 91% をカバーしている。

コーパス・データに基づく用例選定

単語を立体的に理解させるためには、英単語と訳語の情報のみならず、その語が使われている談話的状况を伝えることが重要である。既存の単語集の多くは、いわゆる「例文」を添えることでこの問題に対処しようとしている。

しかしながら、コーパス準拠の観点から見ると、500 万語のデータの中でたまたま 1 度出現した例文を典型性の保証のないまま用例とすることには問題がある。また、教育的観点から見ても、長い例文の採用は記憶の焦点を拡散させる危惧がある。そこで今回は、例文の代わりとして、短い例句 (collocational phrase) を示すこととし、そこにコーパス解析をふまえた語と語の共起情報を盛り込むこととした。

たとえば名詞“crisis”をターゲットワード (NODE) と考えた場合、その直前位置 (N-1)、および直後位置 (N+1) に出現する語の頻度は次表のようになる (頻度 4 以上に限って示している)。

表 3 . “crisis”の共起語

Rank	N-1 Collocation		N+1 Collocation	
	Token	Raw Freq	Token	Raw freq
1	the	10	and	7
2	oil	10	in	7
3	a	8	is	5
4	of	7	of	5
5	energy	6	the	5
6	financial	6	a	4
7	environmental	5	we	4
8	identity	5	---	---
9	economic	4	---	---

このことから、大学入試英文コーパスでは、“crisis”の直前位置には内容語 (content words) として oil、energy、financial、environmental、identity、economic などが共起することがわかる。すべてを例句として採用

することはできないので、同一意味領域の語は上位ランク語のみを収録することとし、energy は oil で、economic は financial で代表させる方針を取った。またランク 4 位の of はコンコーダスラインを検証すると a sense of ~ の語形が多いことがわかった。

一方、直後位置には and、in、of の 3 種の機能語 (functional words) が共起しているが、コンコーダスラインを検証すると、and には節が、in と of には年号が続き、とくに“crisis”と結合性の高い典型的共起は確認されなかった。以上より、最終的に下記の 3 つの例句を採用した。

oil [financial、 environmental] crisis 石油[財政、環境の]危機
identity crisis アイデンティティの危機 (*自分の個性などが失われて
しまうこと)
a sense of crisis 危機感

こうした方針をすべての語に適用することで、従来のセンテンススペースの「例文」以上に、ターゲット語の共起環境を効果的に提示することが可能になった。

記憶の手がかりを与える

以上で概略したように、精選された語を精選された例句とともに示すことは単語集の基本である。しかし、学習者の能力や学習動機の高さを考慮すれば、これからの語彙指導は、学習者が語を記憶するプロセスにもっと積極的に関与すべきだと考えられる。石川 (2004b) では、記憶を促進するための工夫として、語源メモとカタカナ併記という 2 つの方策を採った。

たとえば“doubt”は「疑い」という意味の名詞であるが、これを機械的に覚えることは初学者には容易ではないだろう。記憶に手がかりを与える方法としては、語源解説を加えることが考えられる。

しかし、従来の辞書や単語集の語源解説は、読み手がすでに十分な語源知識を持っていることを前提にして書かれているため、初学者には有効に機能していない場合も多かった。“doubt”の語源に関して、それが「dubious と関連語」(研究社『新英和中辞典』6 版)であることや、「語源は仏語で元は douter。ラテン語の dubitare 由来」(*Longman Dictionary of Contemporary English*, 4th ed.) であることは学問的に正確ではあっても高校生には適切ではない。

そこで石川 (2004b) では、高校生の既知語とからめて語源の説明をし、それを断片的なコード表記ではなく、原則としてセンテンススペースの解説文として記載することとした。また、いわゆる「カタカナ日本語」の普及を念頭に、可

能な限りカタカナ語を併記することとした。“doubt”の場合、語源メモは下記のようなになる。

真実とうそ、2つの (doub→double) 間で揺れ動く気持ち。トランプの「ダウト」はここから。

学習者のカタカナ語の知識は予想以上に豊富であり、場合によってはカタカナ語をうまく選んで記載するだけでその語の意味と用法をごく自然に示すことができる。たとえば「変える・移す」という意味の“shift”には下記のようなメモを添えた。

車の速度を変化させる「シフトレバー」、パソコンで機能を切り替える「シフトボタン」はこの単語。

こうしたカタカナ語の提示により、“shift”の表面的な語義だけでなく、それが主として特定の選択幅の中での位置的・質的変移を含意することや、具体的事物・抽象的概念の両方を目的語にとりうるものが平易に示されるのである。

このほかにも、“warm”については運動前の「ウォームアップ」、 “receive”についてはパレーの「レシーブ」、名詞“bear”についてはぬいぐるみの「テディベア」、 “gather”については「ギャザースカート」、 “acid”については酸味の強い果実で知られる「アセロラ」、 “tail”については「ポニーテール」を挙げなど、多くのカタカナ語を解説に利用した。

語彙学習は単調で面白みのない作業と思われがちであるが、語と語の意外な結びつきを示してやれば、学習者の興味と好奇心を刺激することができる。また、学習者の既知の知識、すなわちスキーマと新出語を結びつけることで、記憶は促進され、その定着も高まるものと期待される。

語彙学習と音声教材

言語の本質は音声である。語彙学習においても、正しい発音を聞いて、語義と発音を連携させることが重要である。単語集が音声 CD を付属させるのはもはや珍しいことではないが、今回の英単語集では、より効果的な学習の展開を可能にすべく、音声の録音方法に工夫を凝らした。

一般に単語集の音声 CD は、「“crisis” (休止) “危険” (休止) “guide” (休止) “導く” ...」というように英語と日本語を交互に読み上げるものが多い。しかし、「英語→日本語」という 2 サイクル構成では肝心の英単語を聞く機会が限られるため、「英語→日本語→英語」という 3 サイクルでの録音を試みた。

学習の初步の段階では、英語・日本語を聞こえたとおりにそのまま口真似する。そして学習が進むと、英語と日本語を互いに転換していくような訓練が有効である。まず、英語音声聞いて訳語を思い浮かべ、それを日本語音声で確認する。ついで日本語音声から再度英語を思い浮かべ、最後にそれを英語音声で再確認するのである。音声 CD はポーズ時間を短くして録音しているので、英和置換と和英置換をすばやく連続的に行う必要がある。

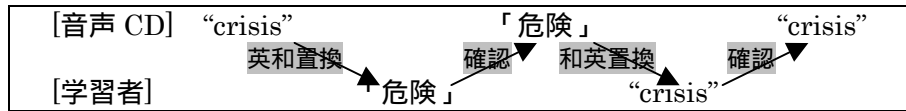


図 2. 学習者とモデル音声のインタラクションモデル

このように、音声 CD をうまく使えば、学習者とモデル音声の間に擬似的なインタラクションを生み出すことができ、転換シャドーイングの訓練にも有効である。繰り返し練習を行えば、英単語と日本語の訳語の間に反射的な反応メカニズムを作ることが可能になるであろう。

もちろん、自然言語における語のふるまいは実に多用であり、わずか 1 つや 2 つの訳語を反射的に記憶したからといって、その語を完全に理解したことにはならない。しかし、中核的な訳語を確実に修得しておくことは、今後、幅広い語義の広がりや語法を学習してゆく上で不可欠な基盤となるものである。

語彙学習とコンピュータ

学習者を取りまくメディア環境の急激な変化を考えれば、今後の語彙指導はコンピュータとの連携をいっそう深めていく必要がある。石川 (2004b) では、音声 CD に加えて CD-ROM を添付し、収録語の Excel®データ、CSV (カン

マ・タブ区切り) データ、および語彙学習ソフトウェアをバンドルした。電子媒体の語彙リストは、学習者と単語の多様な触れ合いの場を提供する。たとえば、書籍版とは別に自分の学習範囲だけを選んでプリントアウトしたり、単語あるいは訳語の列だけを非

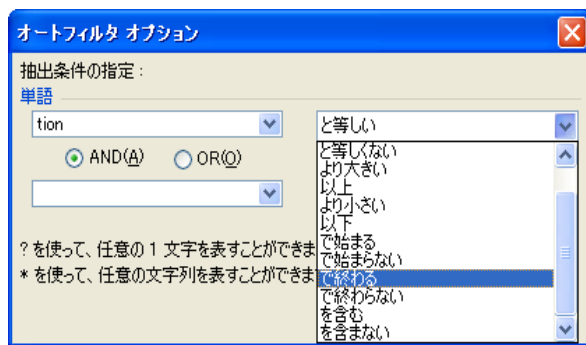


図 3. オートフィルタオプション指定画面

表示にしてチェックテストを作ったり、自分で列を追加して学習日や記憶チェック欄などを設けたりすることもできる。また、アルファベットや頻度順に自由な並べ替えを行うことも容易である。

さらに「オートフィルタ」機能を使用すれば、特定条件に合致する語を一気に抽出することも可能である。まず、単語列のオートフィルタボタンを押し、プルメニューから「オプション」を選ぶ。たとえば、単語ボックスに tion と入れ、条件ボックスから「～で終わる」を選ぶと、名詞接尾辞 tion で終わる語だけが自動抽出される(図4)。

1	A	B	C	D	E
48	11	generation	(名)	世代	
60	11	pollution	(名)	汚染	
68	11	revolution	(名)	革命	
83	12	condition	(名)	条件	
150	13	organization	(名)	組織	
187	14	function	(名)	機能	ファンクション
248	15	situation	(名)	状況	シチュエーション
201	16	competition	(名)	競争	
324	17	vacation	(名)	休暇	バケーション
348	18	mention	(動)	言ふ	
364	19	civilization	(名)	文明	

図4. オートフィルタ実行結果

このように「オートフィルタ」機能を用いれば、特定の形態的特徴を共有する一連の語を抜き出すことができ、語源学習に有効である。「立つ」という意味の語根 st を含む語を抽出したり、否定接辞 in を含む語を抽出したりすることで、各種接辞・語根の働きをより立体的に理解することが可能になるだろう。

ソフトウェアについては、既存の語彙学習ソフトを改良する形で、ソフト作者と共同開発を行った(元になったソフトウェアの概略については石川(2004a)を参照)。CD-ROM に収録されている「P-Study System VITAL 4500 エディション」は、書籍版に収録されているすべての語について、表示



図5. 語彙学習ソフトウェア画面例

される英単語を制限時間内に入力する「Speed Replyモード」、英語と日本語を交互に表示させて記憶を確認する「カードモード」、表示される日本語に対して適当な英語を選ぶ「和英4択モード」、英語に対して適当な日本語を選ぶ「英和4択モード」、日本語に対して適当な英語のスペルを入力する「和英入力モード」の

4つのモードで立体的な学習を行うことが可能である。

また、本ソフトウェアでは学習履歴が自動的に蓄積され、日々の学習時間やその都度の成績などがグラフで一覧できるようになっている。「インターネット



ランキング」機能もついているので、自分の学習時間・成績などをサイトに自動送信して、全国の学習者中でランキングを取得することもできる。こうし

た工夫は、単調な語彙学習に新鮮な刺激を添えてくれるものと言えよう。

図6. インターネットランキング画面例

おわりに

以上、本稿では、高校生用英単語集の編纂過程を紹介しながら、今後の語彙指導のありかたについて論考を行ってきた。ここで述べたことは、高等学校という学習段階に制限されるものではなく、中学校、大学、あるいは一般社会人を対象にした語彙指導においても同じく重要なことである。また、英語に限らず、他の外国語においても同様のアプローチは有効であろう。

本稿の最後に、今後の語彙指導の方向性を考える上で、とくに重要と思われる観点を列挙しておくことにしたい。

- (1) コーパスに基づく頻度情報の処理
- (2) コーパスに基づく共起情報・語用情報の処理
- (3) 語源情報の有効活用
- (4) カタカナ知識の有効活用
- (5) 学習者のスキーマとの密接な関連付け
- (6) 音声教材との有機的連携
- (7) 転換シャドーイング手法を応用した反射訓練の組み込み
- (8) 語彙リストの電子媒体化
- (9) CALL教材との有機的連携

冒頭でも触れたように、語彙指導は、これまでの正規の英語教育の枠組の中で必ずしも十分な対応がなされてこなかった分野である。上記の観点を生かし、語彙表作成から語彙指導、さらには診断・評価までも視野に入れた総合的語彙指導システムの構築が今後の課題となろう。

注

本稿は、同志社大学言語文化教育研究センター主催講演会「コーパスを利用した新しい英語語彙指導」（2004年12月22日、於：同志社大学京田辺校地香柏館第3共同研究室）における発表内容の一部である。

当日の司会の労をおとりいただいた同センター北尾謙治教授、および貴重なご助言をいただいた出席者各位に御礼を申し上げたい。

引用・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』. 東京：大学英語教育学会.
- 石川慎一郎. (1998). 「英語コミュニケーションと語彙——大学入試用単語集の有効性の検証——」. 『言語文化学会論集』, 11, 3-19.
- Ishikawa, S. (2003). Harmonization of a scientific viewpoint and an educational viewpoint in compiling a word list: A case study of 'JACET list of 8000 basic words.' *ASIALEX '03 Tokyo Proceedings*, 72-377.
- 石川慎一郎. (2004a). 「JACET8000 および各種教育ソフトウェアを用いた大学での語彙指導」. 大学英語教育学会基本語改訂委員会（編）. 『JACET8000 活用ハンドブック』 (pp.7-14). 東京：大学英語教育学会.
- 石川慎一郎. (2004b). 『VITAL4500 英単語・熟語』. 東京：文英堂.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London, UK: Macmillan.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. Harlow, UK: Pearson Education Limited.
- Mochizuki, M. (2003). JACET 8000 compared with other vocabulary lists. *ASIALEX '03 Tokyo Proceedings*, 378-383.
- Murata, M. (2003). Background of revision of JACET word list. *ASIALEX '03 Tokyo Proceedings*, 356-359.
- Murata, M., Tono, Y., & Yamada, S. (Eds.). (2003). *ASIALEX '03 Tokyo proceedings: Dictionaries and language learning - How can*

dictionaries help human and machine learning? Tokyo: Asian Association for Lexicography.

Shimizu, S. (2003). Data processing for building the JACET 8000 vocabulary list. *ASIALEX '03 Tokyo Proceedings*, 360-365.

Sugimori, N. (2003). Using frequency statistics to adjust the rank order of words in the JACET list of 8000 basic words. *ASIALEX '03 Tokyo Proceedings*, 366-371.