

## JACET8000 : JACET4000との比較

望月正道<sup>1</sup>

### 1. はじめに

大学英語教育学会 (JACET) は、初級者から大学生までの語彙学習の指針として、大学英語教育学会基本語リストを出版している。1981年の初版の後、1983年に第2版、1993年に第3版、そして2003年に第4版と改訂を重ねている。村田年先生は第2版から第4版まで、このJACET基本語リストの編集に従事されてきた。特に、第3版では担当理事として、第4版では担当理事と基本語改訂委員会委員長を兼任されて、強力なリーダーシップを発揮されると同時に、細やかな気配りで、第4版の出版に尽力された。本稿では、最新の大学英語教育学会基本語リスト第4版を第3版と比較し、その有効性を検証する。

### 2. 編集方法の相違

大学英語教育学会基本語リストの編集方法は、第1版から第3版 (JACET4000) までと第4版 (JACET8000) では大きく異なる。その違いは、主観的判断の有無の相違といえる。第1版から3版までは、基本的に、Kučera and Francis (1967), Carroll et al. (1971), Francis and Kučera (1982) の頻度情報と *Longman Dictionary of Contemporary English First Edition* の定義語彙を参考に、委員の議論により基本語を決定していた。議論の対象とすべき語の方針は、生活基本用語、外来語として日本語になり広く用いられている語、日本文化および英米文化を顕著に表わす語、英語教育上必要と考えられる語 (comma, subjectなどの文法用語など) の4つが挙げられる。この方針に当てはまる語は、頻度情報が低くても、協議により採用された。このような議論を経て、JACET4000第1版では大学英語教育学会基本語として4,064語、第2版、第3版では3,990語を選出した。

それに対して、第4版 (JACET8000) は、主観的判断を含めず、客観的データのみにより編集されている。すなわち、British National Corpus (BNC) の最頻度レマリスト5,516語 (Kilgarriff, 1997) を基準データとし、それにJACET8000サブコーパスの頻度情報を比較することにより、日本人学習者のための基本語を選出するものである。JACET8000サブコーパスとは、大学英語学会基本語改訂委員会が独自に作成したコーパスである。中学校・高校英語検定教科書、雑誌・新聞、映画、児童文学、BBC・CNNなどのニュース、センター試験・STEP・TOEFL・TOEICなどの検定試験の各種コーパスを統合したものである。BNCでの頻度とJACET8000サブ

<sup>1</sup> 麗澤大学外国語学部

コーパスでの頻度を、対数尤度 (log-likelihood) で調整し、上位8,000語を選出した。さらに、日本人学習者のための語彙表という教育的配慮から、高校英語教科書コーパスから上位4,326語を取り出し、8,000語内での順位と高校英語教科書コーパス内での順位の差を求め、その差の2分の1を8,000語に反映し、順位を調整した。また、第4版では、語の定義により、数詞、不規則動詞の過去形・過去分詞、大文字で始まる語 (Iは例外) などは8,000語には含まれない。そのため、教育的配慮から、不規則な活用形、月名、曜日、数詞、国名などの250語をPlus250という別表にしている。

### 3. テキストカバー率の比較

ここではJACET4000とJACET8000をレベル別多読用教材 (graded readers) と英字新聞のテキストカバー率という観点から比較する。テキストカバー率は、Nation and Ming-tzu (1999) とNation (2002) で研究されている。

Nation and Ming-tzu (1999) は、Oxford Bookwormsシリーズの6つのレベルのheadwordsと固有名詞で各レベルの教材がどれくらいカバーできるかを調査している。各レベルのheadword数は、レベル1 (400)、レベル2 (700)、レベル3 (1,000)、レベル4 (1,400)、レベル5 (1,800)、レベル6 (2,500) である。各レベルから7冊ずつをサンプルし、その平均カバー率を求めている。結果は、表1の通りである。

レベル1～4は、読解が可能になるとされる (Laufer, 1992) カバー率95%に達していない。しかも固有名詞が既知語として処理されているので、実際のカバー率はさらに低いことになる。

Nation (2002) は、General Service List (West, 1953) の1,986語とAcademic word list (Coxhead, 2000) の570語を合わせた2,556語を、BNCの3,000語とテキストのカバー率で比較している。カバー率が比較されたのは、350万語の学術テキストコーパス、30万語の*Microeconomics* (Parkin, 1990) のコーパス、50万語のLund話し言葉コーパス、Project Gutenbergからの350万語の小説コーパスの4種類のコーパスである。結果は、表2の通りである。

4種類いずれのコーパスでもBNCの方が1～2%カバー率が高い。しかし、BNCの方が444語多く含んでいることを考慮すると、GSL+AWLのカバー率とほとんど差があるとは言えない。さらに、BNC3,000語には、初級英語学習者が学ぶべきalphabet, ant, arithmetic, astronaut, auntのような語が含まれていない。これらから、Nationは、初級学習者にはGSL+AWLが、上級学習者はBNCが適切であると結論づけている。

これらの先行研究から、レベル別多読用教材のheadwords、固有名詞を既知語として扱うと、そのレベルの90%以上をカバーすること、また、BNCの3,000語は、GSLとAWLの2,556語と同じ程度のカバー率であることが明らかになった。本稿では、BNCを1つの基準として編集したJACET8000とBNC以前の語彙表に基づいて編集したJACET4000を、レベル別多読用教材と新聞のカバー率で比較する。

この比較では同綴異義語は区別できないので、JACET4000のうち同綴異義語は1語のみを残

表1 Oxford Bookwormsシリーズのheadwordsと固有名詞のテキストカバー率

	stage 1	stage 2	stage 3	stage 4	stage 5	stage 6
平均	93.3%	87.6%	91.8%	94.5%	95.7%	96.4%
範囲	88.3— 95.9%	86.1— 90.1%	90.6— 93.3%	93.8— 96.7%	94.8— 96.9%	95.0— 97.5%

表2 GSL + AWL2, 556語とBNC3, 000語のカバー率

コーパス	学術		Parkin		話し言葉		小説	
語彙表	GSL AWL	BNC	GSL AWL	BNC	GSL AWL	BNC	GSL AWL	BNC
カバー率	85.5%	86.5%	91.2%	93.2%	91.4%	92.6%	88.5%	89.6%

し削除した。この結果、JACET4000は3,953語となる。これとほぼ同数となるように、JACET8000ではstage 1-3にPlus250を加え、stage 4の頻度順で750位まで、計4,000語を比較対象にした。したがって、JACET4000の3,953語とJACET8000の4,000語を比較することとなる。

この2つの語彙表をPaul Nationが開発した*Range* ([http://www.vuw.ac.nz/lals/staff/paul\\_nation/index.html](http://www.vuw.ac.nz/lals/staff/paul_nation/index.html)から入手可能)というテキスト分析プログラムのbasewordsとして用いる。*Range*は、テキストの延べ語数、異なり語数、word family数を算出する。basewordsにするとは、語に屈折形を加えたものを1語として、*Range*に認識させることである。たとえば、テキストにwake, wakes, waked, wakingという4つの語があるとすると、これらはすべてwakeのword familyとして数えられる。

レベル別多読用教材として使用したテキストは、*Oxford Bookworm*シリーズのstage 1 (400 headwords) からstage 5 (1,800 headwords) までの5レベルから各10冊ずつ、計50冊である。総延べ語数60万超のコーパスになる。英字新聞テキストは、*The New York Times*と*The Times*から、国際、国内、経済、文化、スポーツなどの記事を2日間分(2003年2月21日と5月8日)に渡り選び、コーパスとした。各新聞1日につき約1万語を採集し、英字新聞コーパスは4万語を越える程度である。

*Range*による分析結果は、表3の通りである。

JACET8000とJACET4000のどちらも、延べ語数では、レベル別多読用教材の各レベルで90%以上をカバーしている。異なり語数では、どちらもレベルが上がるにつれて、75%から65%ほどへとカバー率が減少している。 $\chi^2$ 検定を実施したところ、延べ語数 ( $\chi^2=.000$ ,  $df=1$ , ns.)でも異なり語数 ( $\chi^2=.710$ ,  $df=1$ , ns.)でも両者に有意の差は見られなかった。新聞では、JACET8000の方が延べ語数 ( $\chi^2=12.836$ ,  $df=1$ ,  $p<.001$ )でも異なり語数 ( $\chi^2=12.274$ ,  $df=1$ ,  $p<.001$ )でもJACET4000より有意に多くカバーしていることがわかった。この結果から、JACET8000の最初の4,000語は、レベル別多読用教材ではJACET4000とカバー率には違いがな

表3 JACET8000とJACET4000のカバー率

	延べ語数	異なり語数
GR Stage 1	57,832	1,821
J4カバー率	53,184(92.0%)	1,370(75.2%)
J8カバー率	53,247(92.1%)	1,370(75.2%)
GR Stage 2	64,559	2,437
J4カバー率	60,166(93.2%)	1,800(73.9%)
J8カバー率	60,351(93.5%)	1,814(74.4%)
GR Stage 3	102,536	3,415
J4カバー率	95,884(93.5%)	2,475(72.5%)
J8カバー率	95,608(93.3%)	2,447(71.7%)
GR Stage 4	155,861	5,015
J4カバー率	145,018(93.0%)	3,265(65.1%)
J8カバー率	145,283(93.2%)	3,219(64.2%)
GR Stage 5	230,588	5,854
J4カバー率	213,980(92.8%)	3,810(65.1%)
J8カバー率	213,657(92.7%)	3,736(63.8%)
Newspaper	42,724	7,172
J4カバー率	35,344(82.7%)	3,637(50.7%)
J8カバー率	36,303(85.0%)	3,942(55.0%)

く、英字新聞ではより多くをカバーする。言い換えると、JACET8000は、学習者に必要な語彙を頻度にとらわれず協議で採用したJACET4000と同程度に初級学習者向きのテキストをカバーできる。さらに、authenticな教材である英字新聞では、JACET4000より多くをカバーしている。このことから、JACET8000は、初級学習者から、よりauthenticな題材を学習する上級者にも適した語彙表であると言える。

#### 4. まとめ

本稿は、大学英語教育学会基本語リスト第4版JACET8000と第3版JACET4000とを、編集方法とテキストカバー率の観点から比較した。編集方法では、従来の頻度情報を元に教育的配慮から協議により候補語を追加したり、削除したりするという編集方法を改め、BNCとJACETサブコーパス頻度情報、対数尤度、高校英語教科書コーパス頻度によって、科学的客観的に語彙を選定する方法を採用した。テキストカバー率では、JACET8000の高頻度語4,000は、レベル別多読用教材でJACET4000と同程度に、英字新聞ではJACET4000より多くをカバーすることが判明し

た。このことから、JACET8000は、JACET4000を凌ぐ優れた語彙表と言えるだろう。

### 参考文献

- Carroll, J., Davies P., and Richman, B. (1971) *The American Heritage Word Frequency Book*. American Heritage Pub. Co.
- Francis, N. and Kučera, H. (1982) *Frequency Analysis of English Usage*. Houghton Mifflin Co.
- Kilgarriff, A. (1997) BNC database and word frequency lists. Available electronically at <http://www.itri.bton.ac.uk/~Adam.Kilgarriff/bnc-readme.html> (May 5 2003).
- Kučera, H., and Francis, N. (1967) *Computational Analysis of Present-day American English*. Brown University Press.
- Laufer, B. (1992) How much lexis is necessary for reading comprehension? In Arnaud, P. and Bejoint, H. (eds), *Vocabulary and Applied Linguistics*, London: Macmillan, 126-132.
- Nation, P. (2002) A study of the most frequent word families in the British National Corpus. Keynote address at Second Language Vocabulary Acquisition Colloquium, University of Leiden, March 15, 2002.
- Parkin, M. (1990) *Macroeconomics*. Mass: Addison-Wesley.
- Procter, P. (1978) Defining Vocabulary 2000 Words, *Longman Dictionary of Contemporary English*. Longman.
- 大学英語教育学会基本語改訂委員会 (2003) 『大学英語教育学会基本語リスト』大学英語教育学会.