

# JACET8000 と電子版ニューヨーク・タイムズ紙掲載の新刊書第1章の語彙

上 村 俊 彦

## JACET8000 and Vocabulary of the First Chapters on the New York Times Website

Uemura, Toshihiko

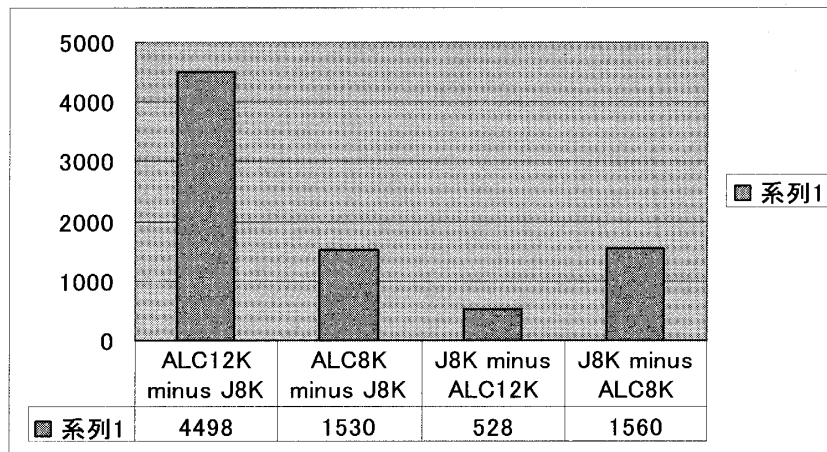
### 1. はじめに

英語圏の大人向け英語を読む時にどのような語彙が必要か。また、日本の大学レベルの語彙リストは、英語を母語とする人々が書いた英文を読む場合、どの程度有効であるか。本稿は、これらの疑問を考察するために、英語圏の大人向け英文テキスト資料として2005年3月13日から8月14日まで電子版 New York Times 紙上に掲載された新刊図書の新刊第1章103冊分の英文テキストの分析をおこなう。また、この英文テキストを読む際に既存の英語語彙リストはどの程度有効であるかを検証するために、大学英語教育学会の JACET 基礎語彙（通称 JACET8000）とアルク社の ALC Standard Vocabulary List（通称 ALC SVL）とを用いて検証を試みる。

### 2. JACET8000 vs. ALC SVL

JACET8000 に採用され ALC SVL には採用されていない語形、逆に ALC SVL に採用され JACET8000 には採用されていない語形を調べるために、ソフトウェア WordSmith Tools Version 4.（以下、WS4）の stop list に2つの語彙リストを登録して相互の出現頻度順語形リストを出力した。グラフ1は、JACET8000 に含まれず、ALC SVL（12000語）中あるいは ALC SVL 上位8000語中に存在する語形、逆に ALC SVL（12000語）あるいは ALC SVL（上位8000語）に含まれず、JACET8000 に存在する語形の数を示したものである。

グラフ1. JACET8000 vs. ALC SVL



N.B. ALC12K は ALC SVL 全体（12000語）、ALC8K は ALC SVL 上位8000語を、J8K は JACET8000 を示す。

グラフから明らかなように、ALC SVL (上位8000語) と JACET8000 の約8割強の語形は共通である。JACET8000 に存在せず、ALC SVL (12000語) のみにある語形は約45000、逆に JACET8000 のみにある語形は約530であった。BNC データと BNC レマ・リストとして使われた Kilgarriff のレマ・リストは、両語彙リストの語形を確定する際に利用されている。このため、JACET8000 と ALC SVL(12000語) に採用された多くの語形は共通である。

表1は、2つの語彙リストに採用された語形を比較するためにまとめたものである。表1から推測できるように、Kilgarriff のレマ・リストの語形の多くが JACET8000 と ALC SVL(12000語) に認められる。しかし、名詞+s (outskirts, thanks, works), 英語綴り (catalogue, enquire, whisky), 短縮形 (exam, math, photo) など、JACET8000 のほうが Kilgarriff のレマ・リスト語形をより忠実に採用している。なお、Kilgarriff レマ・リストと JACET8000 では、ともに “congratulation” (基本形) を採用しているのに対して、ALC SVL(12000語) では “congratulations” (複数形) を採用している。ALC SVL(12000語) 選定時に使用したコーパスデータの語形の出現頻度では、“outskirt, thank, work” については基本形、“congratulation” に関しては複数形が有意に高頻度であったため採用されたのであろうか。

表1. 語形比較

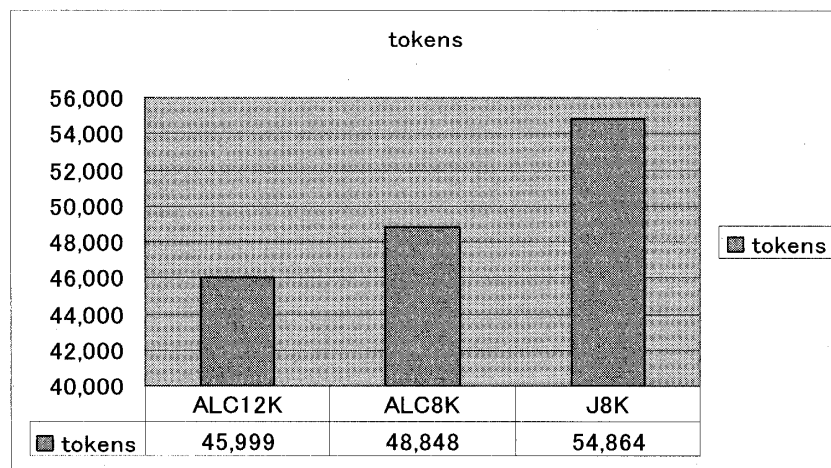
Kilgarriff's Lemma	JACET8000	ALC SVL
catalogue	catalogue	catalog
congratulation	congratulation	congratulations
disc	disc	
enquire	enquire	inquire
Euro	Euro	
exam	exam	(examination)
OK		OK
okay	okay	
outskirts	outskirts	outskirt
photo	photo	(photograph)
telecommunication	telecommunication	
thanks	thanks	(thank)
whisky	whisky	whiskey
works	works	(work)
	Apartheid	
	Internet	
	math	

JACET8000 では、語彙選定時に利用した英文コーパスデータの特徴が反映したためか、ALC SVL(12000語) にはない語形 “Euro, Apartheid, Internet” が採用されている。なお、JACET8000 の語形はすべて原形や基本形が採用されている。一方で、固有名詞 (地名・人名、関連の形容詞)、曜日名・月名、数詞、頭字語、省略語、アポストロフィ (') 付き語などは、出現頻度が高い場合であっても採録されていない。同音語については、ALC SVL(12000語) では May と may, March と march, Miss と miss は別の語形としてそれぞれカウントされている。JACET8000 では、このような語源・語義にもとづく区別はおこなわれておらずひとつの語形とみなされている。

### 3. テキスト・カバー率

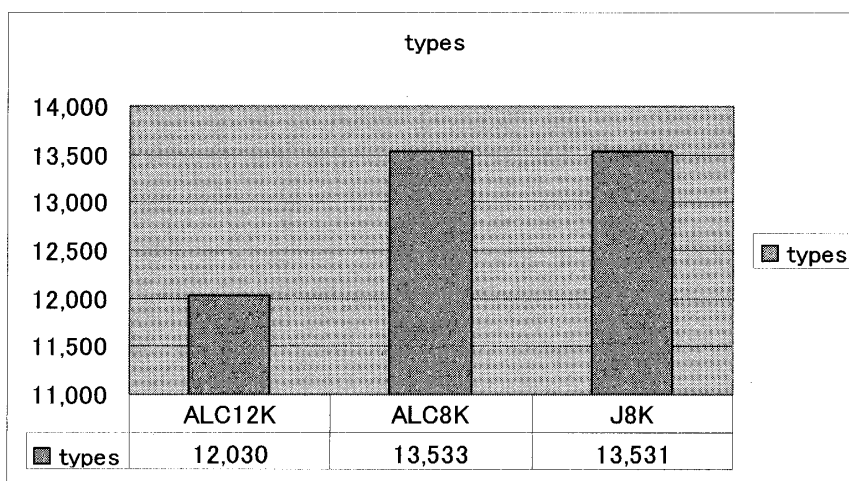
電子版 New York Times 紙掲載の新刊103冊の第1章からなる英文テキスト・データを WS4 にインプットし、wordlist コマンド実行後に得られる stat data によると、token 数は200,721語、type 数は20,102語であった。グラフ2は、ALC SVL(12000語)、ALC SVL (上位8000語)、JACET8000 を、それぞれWS4 の stop list にした場合にヒットしなかった未掲載の token 数である。

グラフ2. リスト未掲載の token 数



グラフ2によると、JACET8000 を WS4 の stop list とした場合に未掲載の token 数が最も多い。また、グラフ2から明らかなように、ALC SVL (上位8000語) と JACET8000 については、token レベルの出現頻度の総数に関して大きな違いが認められる。しかし、グラフ3が示すよ

グラフ3. リスト未掲載の type 数



うに、語彙リストALC SVL (上位8000語) と JACET8000 でカバーされなかった type (語形、以下語形) の数については、両者はほぼ同数で大きな違いが認められなかった。なお、JACET8000 と ALC SVL(12000語)では、グラフ3で該当する語形の数の開きは約1500語となっている。しかし、この数値は両語彙リスト収録の語形の数の違い4000語に遙かに及ばない。

グラフ4は、3つの語彙リストが新刊書103冊の第1章テキスト・データの語彙をどの程度カ

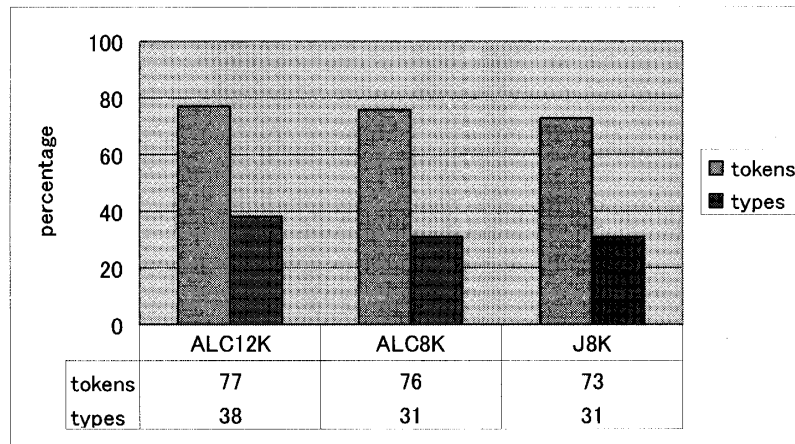
バーするか示している。カバー率 (%) は、WS4 の wordlist コマンドにより出力された stat data をもとに次の計算式で求めた。

$$\text{token カバー率} = \text{stoplist tokens removed} \div \text{tokens used for word list}$$

$$\text{type カバー率} = \text{stoplist types removed} \div \text{types (distinct words)}$$

グラフ4によると、各語彙リストは token レベルで新刊書103冊の第1章英文テキスト・データの語彙73%から77%をカバーしている。Laufer(1989)や Nation(2001)によると、外国語としての英文テキストを十分読みこなすためには、英文テキストの語彙の95%以上が既知であることが好ましい。ニューヨーク・タイムズ紙で紹介された新刊図書を読むためには、大学生レベルの語彙表である JACET8000 や ALC SVL(12000語)でカバーされる出現頻度の高い語形とともに、さらに豊富なボキャブラリーが必要とされる。ちなみに、type レベルのカバー率は31%から38%であった。

グラフ4. カバー率



#### 4. JACET8000 未掲載の NYT103 語形

WS4 によって JACET8000 未掲載語として出力された NYT103 頻度順リストを調べると、出現頻度が10回以上の語形は828語 (全頻度31,354回) であった。その中で新刊書103冊中、10冊以上に出現した語形 (レンジ10以上) は482語 (全頻度24,785回)、30冊以上に出現した語形 (レンジ30以上) は87語<sup>1)</sup> (全頻度15,866回) であった。換言すると、103冊中30冊以上の新刊書で10回以上使われている語形87語は、NYT103 テキスト全体で出現頻度10回以上の語形828語の11% (87/828) にすぎないが、全出現頻度では51% (15,866/31,354) を占める。これらの中には、多くの JACET8000 基本語の変化形や派生形が存在している。ちなみに、レンジ、出現頻度ともに10の語形482語について見ると、それぞれの比率は58% (482/828)、79% (24,785/31,354) となる。

NYT103 の JACET8000 未掲載語であっても、実際にはその原形や基本形は既知のものが多い。JACET8000 原形とその変化形、基本語とその派生形についての知識を高めることで、このような「未知の語形」の数を大幅に減少させることができる。JACET8000 でカバーされず、WS4 の出現頻度順リストに頻度10回以上の語形として現れた828語中、NYT103 の10冊以上 (レンジ10以上) に出現した482語を、語形の意味や形態論上の特徴をもとに大別したものが表2である。

表2. JACET8000 がカバーしなかった語形

	range over 30 texts		range over 10 texts	
	word form	total freq	word form	total freq
動詞の活用形	49	10922	254	15215
名詞の複数形	16	1547	108	3454
人名・地名等	2	299	36	1226
数詞・曜日・月名	13	2250	39	2887
所有格 's	0	0	5	94
縮約形	7	848	31	1737
接頭辞 -er, -est	0	0	9	172
計	87	15866	482	24785

すでに述べたように、JACET8000 で採択された語は、動詞の原形、名詞・形容詞等の基本形である。表2の「動詞の活用形」、「名詞の複数形」、「数詞・曜日・月名」、「所有格's」の付いた語形、縮約形（例'd, 't, 'veなど）、接頭辞（-er, -est, -ness）の付いた派生語は、その出現頻度の高さにもかかわらずJACET8000の「語の定義」によりリストから排除されている。同様に、高出現頻度の人名や地名、これらに由来する語形もJACET8000から排除されている。

未知の語彙が多い「大人の英文テキスト」を読んで概要を理解するためには、既知の語彙、縮約形、「所有格's」の付いた語形が文構造の中でどのような位置に置かれ、どのような統語上の役割を果たしているのかを認識できることが重要となる。同時に、人名または地名由来の語形が英文テキストの中でどのような文体上の役割を担っているか知ることも重要である。NYT103全体で出現頻度が30回以上の地名、人名または地名由来の語形としてWS4が出力した語形は、

(地名関連) YORK(38:122) AMERICA(27:45) ENGLISH(21:46) FRENCH(21:43) EUROPE(21:35) PARIS(20:37) BRITISH(15:39) ISRAEL(4:30)

(人名関連) JOHN(27:64) JAMES(16:32) ROBERT(15:61) WILLIAM(15:33) VIRGINIA(13:36) JACK(10:32) ROOSEVELT(8:36) HARRY(7:54) ROY(6:89) RUTH(5:57) DISNEY(3:71) EISNER(3:43) CARTER(3:32) TAFT(2:35) KENNEDY(1:35) MATHEWSON(1:30) MCGRAW(1:30)

(WS4 出力リストから直接引用した語形の綴りはすべて大文字)

であった。ただし、括弧の中の数字は、語形が出現したテキスト数（レンジ）と出現頻度を表す。

NYT103の多くは、フィクションや人文科学または社会科学系分野のノン・フィクション新刊書の第1章である。ISRAELを除く地名や関連語形は、15冊以上に見られることから比較的に多くの新刊書で言及されている。上記の（地名関連）から、どの国家や地域がよく言及されているか推定できる。また、（人名関連）から、どのような人名がフィクションまたはノン・フィクションの登場人名として多用されているかわかる。NYT103にはテキスト数（レンジ）が10以下の（人名関連）も多く含まれている。出現頻度の割に限られた英文テキストのみに出現する人名は、特定の人物に関する言及や記述の文中に用いられていることが推測される。このような地名や人名によって形成される「英文テキスト」のコンテキストを把握することは、実際の英文テキストの正しい理解の前提となる。

5. 家族語

Coaxhead の語彙リスト Academic Word List (以下, AWL)<sup>2)</sup> では, 570語の基本語と形態論上, または語源上関連する「家族語」の存在を重視している。たとえば, analyse (米語綴り analyze) を例にとると, 家族語 (analysed, analyser, analysers, analyses, analysing, analysis, analyst, analysts, analytic, analytical, analytically, analyze, analyzed, analyzes, analyzing) が存在する。英語学習者が体験的に学んだ語形成の規則を AWL の「家族語」に当てはめてみると, 一見未知と思われた語形が形態論上あるいは意味論上共通性を持った既知の原形や基本形と「家族関係」にあることが推測される場合が多々ある。

日本の学生の英語語彙力<sup>3)</sup> は必ずしも高くはない。日本の学生がその語彙力不足を補って英文テキストを読むためには, 「検定済み学習用英文テキスト」の学習を通じて学んだ基本的な英文法や英語語彙の語義, 動詞の変化形や名詞の複数形, 形容詞・副詞の派生形などの体系的な語形成の規則や統語論上の文法規則を手がかりとすべきであろう。

6. NYT103 の動詞コロケーション

WS4 の concordance コマンドには, 特定の語形のコロケーション・パターンを数値データで示す機能がある。wordlist によって出力された頻度順リストをレマ化し, 出現頻度が50回以上の動詞<sup>4)</sup> を選び出した。助動詞としても使われる “be, do, have” を除いた一般動詞の中で最も出現頻度が高かったのは “say” であった。以下のグラフ 5, グラフ 6 は, concordance コマンドで出力された collocates, patterns である。レマ化された “say” の出現頻度は671回であったが, 原形 “say” の頻度は149回であった。グラフ 5 から, “say” の前後でどのような語形が共起したか知ることができる。

グラフ 5. Collocates

N	Word	With	elation	Total	tal	Left	al	Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	SAY	say	0.000	151	1	0	0	1	0	0	0	0	149	0	0	0	1	0	1
2	TO	say	0.000	66	50	16	4	4	1	0	0	41	0	2	4	2	6	1	
3	THE	say	0.000	53	19	34	9	5	4	1	0	0	6	12	7	3	1		
4	I	say	0.000	52	30	22	3	6	4	3	14	0	6	6	5	3	1		
5	YOU	say	0.000	26	17	9	0	1	0	6	10	0	4	2	1	1	1		
6	HE	say	0.000	24	12	12	1	2	2	7	0	0	3	5	0	4	1		
7	A	say	0.000	23	7	16	5	1	1	0	0	0	2	2	5	4	1		
8	THAT	say	0.000	23	3	20	1	1	1	0	0	0	15	2	0	3	1		
9	AND	say	0.000	22	11	11	3	0	2	2	4	0	3	4	2	0	1		
10	IS	say	0.000	21	8	13	1	0	3	4	0	0	1	2	6	4	1		
11	OF	say	0.000	19	6	13	3	1	1	1	0	0	0	2	2	2	1		
12	NOT	say	0.000	18	9	9	1	1	2	0	5	0	0	1	6	1	1		
13	IN	say	0.000	16	4	12	0	3	1	0	0	0	6	1	0	0	1		
14	IT	say	0.000	16	9	7	1	5	2	1	0	0	4	0	1	2	1		
15	SHE	say	0.000	15	7	8	0	0	1	6	0	0	2	3	1	1	1		
16	WAS	say	0.000	15	9	6	4	1	2	2	0	0	0	1	3	2	1		
17	AS	say	0.000	13	10	3	1	1	4	4	0	0	0	1	0	1	1		
18	DIDNT	say	0.000	13	10	3	1	1	1	1	6	0	0	2	0	0	1		
19	MY	say	0.000	13	8	5	1	1	6	0	0	0	1	0	1	2	1		
20	WOULD	say	0.000	13	13	0	1	0	0	12	0	0	0	0	0	0	1		
21	DO	say	0.000	11	8	3	2	0	0	6	0	0	0	3	0	0	1		
22	IF	say	0.000	11	7	4	2	1	0	4	0	0	3	0	0	1	1		
23	WHAT	say	0.000	11	7	4	2	2	2	0	1	0	2	1	0	0	1		
24	ABOUT	say	0.000	10	0	10	0	0	0	0	0	0	3	3	1	1	1		
25	NO	say	0.000	10	4	6	3	1	0	0	0	0	2	1	2	1	1		
26	PEOPLE	say	0.000	10	7	3	2	1	1	2	1	0	0	2	1	1	1		
27	THEY	say	0.000	10	6	4	0	0	0	1	5	0	2	0	0	1	1		
28	THIS	say	0.000	10	2	8	0	0	1	1	0	0	4	4	0	0	1		
29	AT	say	0.000	9	3	6	2	0	1	0	0	0	4	0	0	1	1		
30	BUT	say	0.000	9	7	2	2	2	2	0	1	0	0	0	1	1	1		
31	HAVE	say	0.000	9	5	4	1	0	2	2	0	0	2	1	0	1	1		
32	ME	say	0.000	9	5	4	2	2	1	0	0	0	2	1	0	1	1		
33	WE	say	0.000	9	5	4	0	1	0	4	1	0	2	0	0	0	1		

また、グラフ6からNYT103における“say”の統語構造がうかがえる。ただし、NYT103の中で、5回以上の頻度で3つの語形連鎖（スリー・グラム）として認められたのは、“to say that”（7回）のみであった。その1例をあげる。

“... would be an exaggeration *to say that the* Genoese willed Caffa into existence, but not a large exaggeration. (Kelly, J. (2005) *The Great Mortality*)” (イタリアックス筆者)

グラフ6. Patterns

語形“say”（レマ化済み）の次に出現頻度が多かったのは“like”（レマ化済み）の475回である。ただし、今回の分析では、NYT103 データを POS タグ付けしていないため、“like”（前置詞）の用例を多く含んでいる。表3は、NYT103 データの中での出現頻度が50回を超える動詞コロケーションのうち、WS4によって頻度が5回以上のスリー・グラムをリスト化したものである。

表3. Clusters

word form	cluster 1	cluster 2	cluster 3
LIKE	WOULD LIKE TO 8	THEY LIKE TO 5	
SAID	I SAID I 7	THEN HE SAID 6	HE SAID AND 6
	AND HE SAID 5	AND SHE SAID 5	HE SAID I 5
MADE	HE MADE NO 7	MADE IT CLEAR 6	
KNOW	I DON'T KNOW 17	KNOW WHAT TO 7	TO KNOW WHAT 7
	DO YOU KNOW 7	KNOW HOW TO 6	WANT TO KNOW 5
	I KNOW THAT 5		
SEE	TO SEE THE 7	TO SEE WHAT 5	TO SEE A 5
THOUGHT	THOUGHT I WAS 5		
GET	HE COULD GET 5	TO GET UP 5	GOING TO GET 5
CAME	THE LORD CAME 7	CAME TO THE 6	I CAME TO 5
SAY	TO SAY THAT 7		
GO	TO GO TO 9	WANTED TO GO 8	TO GO AND 6
	GO TO THE 5		

英語学習者向けの多くの文法書や ESL/EFL 辞書にはさまざまな動詞句パターンが列記されている。しかし、実際の英文テキストで繰り返し出現する語形連鎖は、表3からも明らかなように限られたパターンである。学生に出現頻度の高い語形連鎖を優先的に学ばせることは、学習負荷の軽減と学習の効率化につながる。この視点は、大学英語教育の現場においても重要視すべきである。

## 7. 終わりに変えて

語彙リスト JACET8000 や ALC SVL を手がかりとして、「大人の英語テキスト」ニューヨーク・タイムズ紙で紹介された新刊書第1章の英文テキスト・コーパスの語彙についての考察をおこなった。本稿から明らかなように、コーパス言語学的手法を用いた語彙研究は英語の研究や教育面でさまざまな応用が可能である。ESL/EFL でカバーすべき出現頻度の高い語形についての研究については、多くの優れた研究がなされている。今後は、このような出現頻度の高い語形の語彙連鎖パターンやその統語構造上の特徴などについて検討すべきであり、さらに研究を進めたい。

### [注]

- 1) JACET8000 がカバーしなかった NYT103 テキスト・データの中で、103冊中30以上の第1章テキスト・データ中に現れた語形（レンジ30以上）は以下の87語であった。ただし、括弧の中の数字は、語形が出現したテキスト数（レンジ）と出現頻度を表す。

（動詞の活用形）

WAS(99:2403) IS(99:1090) HAD(97:1584) WERE(96:706) BEEN(93:461) DID(83:226) ARE(79:450) MADE(79:214) SAID(78:444) HAS(69:260) CALLED(63:127) WENT(59:102) CAME(58:151) TOLD(57:135) FELT(56:100) KNEW(55:108) SEEMED(51:108) TOOK(50:108) TURNED(49:81) LOOKED(48:98) WANTED(47:105) GOING(47:103) TAKEN(47:61) GOT(46:120) BEGAN(45:68) ASKED(44:103) GAVE(44:70) LOOKING(42:65) SEEN(41:67) HAVING(40:67) GONE(40:64) KEPT(39:63) STOOD(39:57) BECAME(38:71) TAKING(38:46) AM(37:92) HELD(37:50) LIVED(36:66) HEARD(36:57) PLAYED(34:49) DONE(34:43) DOES(33:68) SITTING(33:50) SENT(33:47) STARTED(31:54) TRIED(31:50) SPENT(30:40)

（名詞の複数形）

YEARS(78:262) DAYS(62:118) EYES(53:118) MEN(52:129) TIMES(52:84) THINGS(48:109) OTHERS(48:76) FRIENDS(45:79) CHILDREN(44:125) MONTHS(43:69) WORDS(40:77) FEET(34:52) HANDS(33:75) WOMEN(33:63) PARENTS(32:65) LIVES(30:46)

（数詞、曜日・月名）

ONE(99:702) FIRST(97:374) TWO(86:290) THREE(77:174) FOUR(59:115) FIVE(58:128) TWENTY(51:93) SECOND(47:77) SIX(45:78) TEN(35:51) EIGHT(33:50) HUNDRED(32:71)

（国名・地名・人名、関連の形容詞）

AMERICAN(44:170) YORK(38:122)

（接辞：-er, -est -ness）

EARLIER(41:25)



(縮約形)

DIDN'T(55:177) DON'T(49:170) IT'S(44:163) WASN'T(41:80) COULDN'T(36:68) THAT'S(33:64) I'M(31:126)

2) Academic Word List (AWL) の“570 word families”に West の General Service List (GSL)の2000語は含まれていない。Coaxheadによると、AWLは大学レベルで英語を外国語として学ぶ人々の使用を想定した語彙リストであり、GSLは習得済みの基本語彙と考えている。なお、GSL+AWLによる英文テキスト・カバー率については、上村(2003)を参照。

3) 日本の学生の多くは、文部科学省検定済み英語教科書で英語を学び、大学入試センター試験を体験して大学生となっている。大学入試センター試験に出題される問題は、文部科学省の『学習指導要領』のもとで検定された高校教科書に基づき出題されている。『高等学校学習指導要領』によると、文部科学省検定済み英語教科書で学ぶことのできる英語ボキャブラリーは2,700語程度である。

中学・高校の英語教科書として文部科学省の検定に合格するためには、中学英語教科書の語彙は900語程度、高校英語教科書の語彙は2700語程度(中学レベルの900語に加えて、「英語Ⅰ」400語、「英語Ⅱ」500語、「リーディング」900語)でなければいけない。平均的な日本人学生の高校終了時レベルの語彙力は、JACET8000、ALC SVL(12000語)はもとより、多様な語彙によって構成されているNYT103英文テキストに遙かに届かない。

4) WS4 の wordlist コマンドで得られた頻度順リストの語彙をレマ化し、その中で出現頻度が50回以上の動詞の用法がある語形を出現頻度順に並べると以下ようになる。

BE(101:5982) HAVE(102:2564) SAY(63:671) DO(81:648) LIKE(93:475) GO(64:435)  
MAKE(66:416) KNOW(66:416) SEE(67:355) COME(65:345) TAKE(57:333) GET(63:327)  
THINK(52:326) TELL(41:254) GIVE(37:194) FEEL(39:176) BECOME(48:169) FIND(47:160)  
LOVE(39:136) KEEP(36:128) WORK(56:126) FACE(48:123) NAME(54:116) PUT(55:114)  
READ(42:107) SET(56:103) END(50:94) WANT(43:94) LET(46:84) OPEN(45:77) STATE(35:74)  
TALK(39:64) PLAY(24:56) CALL(37:55) LIVE(36:52) TURN(31:50)

#### 参照文献

Davies, M. 2005 “Vocabulary Range and Text Coverage: Insights from the Forthcoming Routledge Frequency Dictionary of Spanish” in Eddinton, D. ed. *Selected Proceedings of the Seventh Hispanic Linguistics Symposium* <http://www.lingref.com/cpp/hls/7/paper1091.pdf>

大学英語教育学会基本語改定委員会編(2003)『大学英語教育学会基本語リスト』東京：大学英語教育学会  
文部科学省『中学校学習指導要領』(平成10年12月告示，15年12月一部改正)第2章各教科第9節

外国語 [http://www.mext.go.jp/b\\_menu/shuppan/sonota/990301/03122602/010.htm](http://www.mext.go.jp/b_menu/shuppan/sonota/990301/03122602/010.htm)

-----『高校学習指導要領』(平成11年3月告示，14年5月，15年4月，15年12月一部改正)2章  
普通教育に関する各教科 第8節外国語

[http://www.mext.go.jp/b\\_menu/shuppan/sonota/990301/03122603/009.htm](http://www.mext.go.jp/b_menu/shuppan/sonota/990301/03122603/009.htm)

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge, Cambridge University Press.

Nation, P. and Waring R. <http://www1.harenet.ne.jp/~waring/papers/cup.html>

上村俊彦(2003)「JACET8000のテキストカバー率：語彙表と英文テキストによる検証」県立長崎シーボルト大学『国際情報学部紀要』第4号 pp.57-66.

(英文テキスト・コーパス・データ)

The New York Times Books First Chapters

<http://www.nytimes.com/pages/books/chapters/index.html>

(語彙リスト関連 URL)

ALC Standard Vocabulary List

[http://www.alc.co.jp/goi/PW\\_top\\_all.htm](http://www.alc.co.jp/goi/PW_top_all.htm)

Bauman, J. The General Service List

<http://www.jbauman.com/aboutgsl.html>

Coxhead, A. The Academic Word List

<http://language.massey.ac.nz/staff/awl/awlinfo.shtml>

Kilgarriff, A. BNC database and word frequency lists

<http://www.itri.brighton.ac.uk/~Adam.Kilgarriff/bnc-readme.html>