新JACET8000の頻度上位3000語*

上　村　俊　彦

The Top 3000 Words of New JACET 8000*

Toshihiko UEMURA

Abstract

Large English corpora such as the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the American National Corpus (ANC) are available for many linguists and English teaching professionals to make high frequency wordlists. Using the latest vocabulary data extracted from BNC and COCA, TOEFL and TOEIC test preparation textbooks, English newspaper articles and other English resources, the Japan Association of College English Teacher (JACET) plans to publish in 2016 the fifth version of its basic vocabulary list, which will be named as New JACET8000. The present study examines how efectively its top 3000 words cover high frequent words found in five large spoken English corpora and English graded readers.

Key Words

high frequency words, JACET8000, spoken corpus, graded readers

1．はじめに

British National Corpus（BNC）[1] の公開以後、英語の大規模コーパスをもとにした出現頻度の高い語彙や語連鎖に関する研究が進んでいる。Nation (2004: pp.3-13）は、BNCの頻度上位1～3000語（BNC3000）とGeneral Service List（GSL）+Academic Word List（AWL）の英文テキストカバー率の比較をおこなった。Adolphs and Schimitt (2004: pp.39-49）は、口語英語コーパスCambridge and Nottingham Corpus of Discourse in English（CANCODE）の頻度上位1～2000語はCANCODEの約94%をカバーすることを報告している。Adolphs and Carter (2013: pp.22-36)は、BNCの口語パート1000万語を分析して、語連鎖（２、３、４、５、

＊本研究は、科学研究費助成事業（基盤研究（C）課題番号25370644「英語発信技能評価システムの構築とその応用研究」）の助成金の一部を使っておこなわれた。

6 語）の出現頻度上位10位までのリストとその頻度とを明らかにした。Jones and Waller (2015: pp.72-73）は、Brigham Young University-British National Corpus（BYU-BNC）を分析して、英文テキストジャンル（口語英語、フィクション、雑誌、新聞、学術英語）の中で 現在完了形（have+past participle）が最も高頻度に出現するのは新聞英語であることを明らかにした。Timmis (2015: p.42)は、近年のコーパス研究の成果をもとに、頻度上位１～2000語と連鎖する語との共起関係の重要性を指摘した。

「大学英語教育学会基本語彙リスト第４版」（通称　JACET8000）は、2003年に大学英語教育学会（JACET）から刊行された。すでに10年が過ぎたため、同学会は大規模コーパスのBNCやCorpus of Contemporary American English （COCA)、日本人英語学習者に必見の試験英語データ（TOEFL、TOEIC、英検)、新聞記事、学術文献データなどからなるコーパスデータベースを新たに構築し、2016年刊行をめどに基本語彙リスト（新JACET8000）の策定作業をおこなった。(望月他編著（近刊)) 本稿では、新JACET8000の頻度上位1～3000語（以下、JACET3000）の英文テキストカバー率の信頼性についての検証を行う。

２．JACET8000の高頻度語

オンラインソフトウェアv8an 2015（正式公開前版）を使用することで、英文テキストの語彙の頻度別の分布状況を調べることができる。図1は、ペンギン多読用英語リーダー（Penguin Graded Readers、以下PGR）のLevel 3からLevel 6のタイトルからレベルごとに10タイトル[2) 3)]を選び、v8an 2015（正式公開前版）を用いて各レベルの英文テキストデータに対するJACET3000のカバー率を調べた結果である。ただし、図中のP３～P６のPはPGR、数字はタイトルのLevel ３～６を示す。なお、PGRの難易度はP６が最上位である。

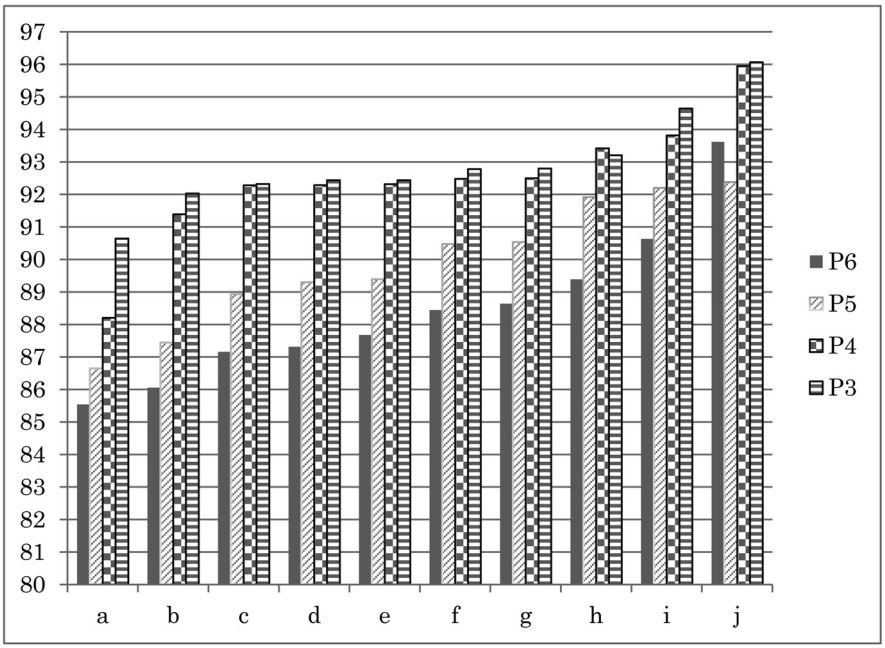

図１　v8an 2015によるペンギン多読用リーダーのカバー率

JACET3000によるPGRテキストのカバー率は、P３（91%～96%）、P４（88%～96%）、P５（87%～92%）、P６（86%～94%）となった。（小数点１位四捨五入）若干の例外（h、J）は認められるが、PGRテキストの難易度（P３＜P６）とテキストに占める頻度順位１～3000語との間には、易しいレベルのPGRタイトルの英文テキストの方がJACET8000によってカバーされる率は高いという傾向が認められる。言い換えると、PGRのレベルが上がり使用可能な語彙が増加すると、JACET3000による英文テキストの語彙に対するカバー率は下がる。ただし、図1からも明らかなように、P３とP４の各タイトルのカバー率の差異は、a（2.44%）以外は１%以内と僅差に留まっている。

高頻度語彙リストは、構築されたコーパスデータの特性に左右されるために常に同一の語学リストとはならないが、汎用性の高い高頻度語彙リストを策定する試みは以前からおこなわれてきた。West（1953）のGeneral Service List（GSL）は、英文テキストの頻出語彙リストの草分けとして知られている。Brezina and Gablasova（2013）は、大規模コーパスデータにもとづく現代版GSL[4)]を提案している。（以下、newGSL）newGSLで使用されたコーパスデータは構築時期の順に、1960年代のLancaster-Oslo-Bergen Corpus (LOB)、1990年代のBNC、2005-70年のBE06 Corpus of British English（BE06）と、2012年のEnTenTen12の４コーパスである。ちなみに、４つの中で最大のコーパスはEnTenTen12で、オンライン上で集められた120億語相当の英文テキストデータから構成されている。なお、newGSLでは、BNC口語パート1000万語を口語英語データとして利用している。

３．カバー率の検証

テキストカバー率の観点から、JACET3000とnewGSL（2251語）の検証をおこなった。なお、newGSLの語数は、WS6出力データ（レマ処理済み後の語数）にもとづいた。

３．１　Longman Vocabulary Checker

ウェッブサイトLongman Dictionary Online上では、Longman 9000 keywordsをもとにした英文テキストのボキャブラリーチェッカ（Longman Vocabulary Checker、以下LVC）の利用ができる。表２は、LVCの高頻度語（high frequency words）3000語と、Longman社の２つの英語辞典の語義記述語（dictionary defining vocabulary）リスト、newGSLとJACET3000との一致率を調べたものである。（以下、*Longman Dictionary of Contemporary English*の語義記述語をLDOCE、*Longman Advanced American Dictionary*の語義記述語をLAADと略記する。）

表１　Longman core vocabulary - high frequency カバー率

	LDOCE	LAAD	newGSL	JACET3000
word	2171	1585	1753	2069
%	68.92	69.76	75.43	68.51

４つの語彙リストの中で、LVCの高頻度語3000語（Longman core vocabulary - high frequency）と最も一致したのはnewGSLであった。一致率は、newGSLが75.43%、その他の３者は僅差で68.51～69.76%となった。newGSLには及ばなかったものの、JACET3000によるカバー率はLongman 社の語義記述語（LDOCE、LADD）とほぼ同一水準となった。

LVC出力の高頻度語一致語彙リストに、aboveやIは含まれていない。LVCでは、aboveは中頻度語（Longman core vocabulary - medium frequency）に分類されているが、Iは高頻度語（1～3000語)、中頻度語（3001～6000語）のいずれにも含まれていない。Leech, Rayson, and Wilson (2001)によると、BNCの総合頻度リストでは、aboveは739位（137回)、Iは11位（8875回）で、両者はともに高頻度語に該当する。(同掲書 List 1.2参照。ただし、頻度は100万語換算) ウェッブサイトLongman Dictionary Online上で、LCVにおける語頻度判定基準について言及した情報はないため、これら2語の頻度レベル判定の妥当性については検証できない。また、toは両レベルのいずれでもヒットする。[5] プログラム上のミスの可能性が高い。LVCの語彙頻度基準とカバー率算出プログラムは再検討の余地がある。

3．2　大規模口語英語コーパスによる検証

英文テキストを電子ファイル化して語彙の使用頻度やコロケーション情報を出力するソフトウェアとして、AntConc（Anthony, L.)、Sketch Engine（Kilgarriff, et al. 2014）やWordSmith Tools V.6（Lexical Analysis Software Ltd.、以下WS6) がよく利用されている。

本研究で使用した大規模口語英語コーパスは、American National Corpus（ANC）の公開口語コーパス（Face-to-faceコーパスと Switchboardコーパス)、Michigan Corpus of Academic Spoken English（MICASE)、British Academic Spoken English（BASE)、British Component of the International Corpus of English（ICE-GB（spoken)）である。表2は、5つの口語コーパスをWS6で出力した結果である。Switchboardは、総token数では最大であるが、ファイル平均のtype数では最小であった。

表2　5つの口語コーパス

	file size	N. of files	token		type		100万語換算
			total N.	average*	total N.	average*	
Face_to_face	1187676	93	199083	2141	8559	469	904
ICE-GB_Spoken	3410154	500	639445	2114	23186	637	1089
Switchboard	15719014	2256	2958436	1311	25931	341	865
MICASE	9741421	152	1795531	11813	34211	1443	1017
BASE	9049456	199	1650990	8296	34812	1257	1161

* average：1ファイルあたりの語数

各コーパス中に100万語換算で100回以上の割合で出現する語[6]（表2「100万語換算」参照）を抽出し、5つのリストを合成したものを口語英語コーパスの高頻度候補語（type換算で1804語）として検証の対象とした。(ただし、1804語はレマ処理済みの語数。) JACET3000とnew GSLとをWS6のストップリストとして、それぞれの語彙リストによるカバー率を、

（ストップリストと一致した異語数）÷（全異語 1804語）×　100

で求めた。(表3参照)

表３．JACET3000とnew GSLによる口語英語コーパスカバー率

	types removed	coverage (%)
JACET3000	1255	69.57
newGSL	1166	65.63

ストップリストによって排除されずに出力された語彙リストの特徴を見るために、WS6出力データから以下の語を除いた。

(1) 固有名詞由来語（例America, American, Britain, British, California, Carolina, Christmas, Dallas, England, Europe, European, France, French, German, Germany, Indian, Iraq, Iraqi, Japan, Kuwait, London, Michigan, Plano, Roman, Rome, Soviet, Spanish, Texas, York; Charlotte, David, Heseltine, Hurd, Hussein, John, Martin, Matthews, Michael, Michaela, Paul, Peter, Saddam, Speckman, Thatcher, Tony, William）
(2) アルファベット文字（Iとaを除く。alpha, delta含む）
(3) 頭辞語（例 TV）
(4) 省略形（例 'd, 'll, 'm, n't, 's, 've; Mr., Mrs., OK）
(5) 接辞（例 non-, re-）
(6) イギリス英語綴り（例 behaviour, centre, defence, honourable, labour）
(7) 数詞（例 eighteen, eighteenth, eighty, eleven, fifteen, fifty, forth, forty, fourteen, nineteen, nineteenth, ninety, seventeen, seventy, sixty, thirty, twelve, twentieth）
(8) 外国語（例 monsieur）

下記のリスト1はJACET3000、リスト2はnew GSLをストップリストとした結果である。

リスト1
（ストップリスト　JACET3000)

acute, ally, antigen, arsenal, assess, assessment, awful, bind, bunch, clinical, consent, conservative, consultation, context, creek, curve, defense, diabetes, diagram, dirt, downtown, electron, email, equation, essentially, faculty, fluid, gentleman, goat, grind, guerrilla, gulf, gun, handout, hip, hypothesis, infect, jury, kidney, lord, marginal, minus, molecule, naught, neat, nerve, node, ought, output, particle, penalty, phase, porch, prime, probability, protein, punishment, quote, rend, reunion, scary, scheme, semester, tissue, transcript, troop, urine, variable, versus, virus (70語)

（口語表現）alright, cos, cuz, em, gonna, goodness, gotta, grandma, hum, kinda, mama, uhm, wanna, yep (14語)

リスト2
（ストップリスト　newGSL）

accord, acute, aircraft, ally, antigen, anybody, anymore, arsenal, assumption, aunt, awful, baseball, basically, behavior, best, better, bunch, campus, cent, center, chase, color, commission, complicate, confuse, consent, conservative, cook, cousin, crack, crazy, creek, crop, curve, dear, defense, diabetes, diagram, dirt, downtown, due, electron,

engineering, equation, favorite, feedback, fiction, fluid, gene, gentleman, girlfriend, goat, grade, graduate, grandfather, grandmother, grandparent, grind, guerrilla, gulf, handout, highway, him, hip, historian, honey, hopefully, hypothesis, including, infect, insurance, junior, jury, kidney, kind, labor, lake, lately, least, lecture, lord, lot, mad, marginal, me, mine, minus, molecule, moving, naught, neat, neighbor, neighborhood, nerve, node, nurse, organization, ours, oxygen, particle, penalty, percent, pet, poem, porch, presentation, probability, program, punishment, queen, quote, realize, recognize, recycle, recycling, renal, rend, reunion, revolution, royal, scare, scary, scream, secretary, self, semester, seminar, sir, slide, smoke, snow, sort, speaker, tape, teenager, tennis, theater, them, tire, tissue, tonight, toward, transcript, troop, unite, upset, urine, us, used, vacation, variable, versus, virus, zero (154 語)

（口語表現）ah, alright, bye, cos, cuz, daddy, em, exam, gonna, goodness, gotta, grandma, granny, hey, hi, hum, kinda, lab, mama, math, mom, oh, okay, uhm, wanna, yeah, yep (27 語)

JACET3000とnewGSLとのカバー率を比較すると、前者が約5%上となった。また、JACET3000はnewGSLよりも口語表現に分類される未掲載語が少なかった。

リスト1とリスト2を見ると、ストップリストによって排除されずに両リストに残った語の中には教育・学習関連語（handout, semester, transcript）や学術専門語（antigen, assessment, diabetes, diagram, electron, equation, hypothesis, infect, molecule, probability, tissue, urine, virus)、軍・防衛関連語（ally, arsenal, defense, guerilla, troop）が多く含まれている。これらの語が1804語に入ったのは、2つの口語英語コーパス（BASEとMICASE）が大学講義やアカデミックな会話をコーパスデータ化したことに起因している。JACET3000にはprotein、newGSLにはgeneとoxygenとが採録されていない。また、教育・学習関連語（exam, lab, math）は、JACET3000で採録、newGSLで未採録となっている。なお、リスト2には、人称代名詞（him, me, mine, ours, them, us）や親族関係語（aunt, cousin, grandfather, grandmother, grandparent）が認められる。これらの語もnewGSLには採録されていないことになる。

3．3　PGRL6による検証

ペンギン多読用英語リーダーのレベル6（Penguin Graded Reader Level 6、以下PGRL6）のタイトルから任意に21冊[7)]を選び、英文テキストを電子ファイル化し、全ファイルをWS6で１つの語彙リストとした。このWS6出力リストから、出現頻度が100万語換算で100回以上の1495語を抽出した。（ただし、1495語はレマ処理済みの語数。）WS6で、JACET3000とnewGSLのそれぞれをストップリストとし、両リストによるカバー率をそれぞれ算出したところ、JACET3000のカバー率はnewGSLよりも約５%高い数値となった。（表４参照）

表4　JACET3000とnewGSLによるPGRL6カバー率

	types removed	coverage (%)
JACET3000	985	65.95
newGSL	899	60.13

以下のリスト3とリスト4は、JACET3000またはnewGSLによって除外されずに出力されたWS6の語彙リストから、２．３の除外対象となる語を削除したものである。（対象語の例については、2.3の(1)〜(8)参照。）

リスト3
（ストップリスト　JACET3000）

beer, beg, bullet, bumble, carriage, charm, cheek, chemist, churchyard, cigarette, cliff, consultant, convict, courthouse, courtroom, cruel, dealer, embryo, faint, forgive, gamekeeper, gentleman, ghetto, grave, grind, gun, handsome, juror, jury, kiss, lean, lighter, madam, molecule, murder, negro, pause, pleasant, priest, protein, rifle, sergeant, servant, softly, stern, tightly, tobacco, trench, tribe, twist, ugly, wed, workhouse（53語）
（口語表現）momma

リスト4
（ストップリスト　newGSL）

angry, asleep, aunt, beg, bell, bend, beside, best, better, bomb, born, bullet, bumble, calm, captain, carriage, center, charm, cheek, chemist, churchyard, cigarette, clever, cliff, coat, color, convict, courthouse, courtroom, crazy, cruel, dealer, dear, digital, dirty, embryo, excuse, faint, forgive, fortune, frighten, gamekeeper, gate, gene, gentleman, ghetto, glad, glance, goodbye, grass, grateful, grave, gray, grind, guilty, handsome, happiness, him, horror, hungry, hurry, joke, juror, jury, kind, kiss, knife, laboratory, lean, least, lighter, lip, lonely, lot, loud, madam, me, meter, mine, molecule, moving, murder, negro, neighbor, nervous, nod, nurse, pale, pause, pile, pleasant, priest, quietly, ray, realize, recognize, rifle, rough, scream, sergeant, servant, shelf, shine, shout, silence, silent, smell, smoke, softly, stamp, stare, stern, surprised, swim, them, tightly, tired, tobacco, tonight, toward, trench, tribe, twist, ugly, uncle, uncomfortable, unhappy, upset, upstairs, us, used, valley, whisper, workhouse（134語）
（口語表現）ah, momma, oh

JACET3000とnewGSLによるPGRL6（1495語）のカバー率は、口語英語コーパス（1804語）の場合よりも低くなった。これは、PGRL6のデータサイズに起因している可能性がある。PGRL6はデータサイズが小さいため、１タイトルの出現頻度が15回程度でも100万語換算で100回以上となる。限られたPGRL6タイトルで頻出した語の中には、科学関連語（embryo, molecule）や司法関連語（convict, courthouse, courtroom, juror, jury）などがある。embryo（14回）は*Brave new world*、*molecule*（15回）は*Double helix*、*convict*（17回）は*Great expectations*、juror（51回）は*Runaway jury*、workhouse（21回）は*Oliver Twist*のみに現れた。上記タイトル以外で、これらの語は100万語換算で100回以上の割合で現れていない。なお、サンプリングベースで網羅的ではないが、dealer（16回）とdigital（32回）は*Business @ the speed of thought*、chemist（14回）とgene（22回）とprotein（19回）とは*Double helix*、negro（14回）は*I know why the caged bird sings*、ghetto（20回）は*Schindler's list*に限定される頻出語であった。ちなみに、courtroom（17回）とjury（66回）は、２タイトル

（*Chamber, the*と*Runaway jury*）に偏在している。（斜体字はタイトル名、括弧内の数字は実際の出現頻度を示す。）

4．高頻度語彙リストとしてのJACET3000、newGSL

本研究では、大規模口語英語コーパス（話し言葉コーパス）と多読用英語リーダーの英文テキストコーパス（書き言葉コーパス）から、出現頻度が100万語換算で100回以上となる語彙リストを抽出し、JACET3000とnewGSLとをストップリストとした場合のテキストカバー率検証をおこなってきた。JACET3000は、話し言葉コーパスと書き言葉コーパスの両方でカバー率がnewGSLを5%程度凌ぐ結果となった。また、JACET3000は、newGSLよりも多くの口語表現を含んでいることが明らかとなった。ただし、JACET3000はnewGSLよりもtype数で約800語多いことを考慮すると、当然の結果かも知れない。

newGSLのリストには、一部の人称代名詞や親族関係語が含まれていない。newGSLでは、ベースとなったコーパスの1つであるEnTenTen12のデータ量の突出がその要因と考えられる。特定ジャンルの大規模データが語彙リスト策定時に影響したため、そのデータ特性がnewGSLにもそのまま反映したことが危惧される。これは、newGSLの今後の課題であろう。

JACET3000やnewGSLをストップリストにすることで、検証用コーパスの語彙リストからは多くの高頻度語が削除された。さらに、固有名詞由来語、頭辞語、省略語、異綴り、数詞などを除くと、検証用コーパスから作成された語彙リストに最終的に残ったものは中頻度または低頻度の候補語群となる。これらの語グループは、検証対象となったコーパスのデータ特徴を示していた。中頻度または低頻度グループの語彙項目は、ベースとなるコーパスによって変化するために、汎用性の高い中頻度や低頻度の語彙リストを作成することは高頻度語の場合以上に難しいことが推測できる。（これらの語の暫定的な意味論的な解釈については、3.2、3.3を参照。）

5．終わりに

本稿では、JACET3000にもとづく口語英語コーパスと多読用英語リーダーコーパスの分析を通じて、英文テキストにおける高頻度語とその他の語の出現傾向や分布状況について考察した。高頻度語の多義性や語連鎖の特性については、上村（2014）の考察を拡大する方向で今後さらなる検証をおこないたい。

注

1．BNCは、リリース順にBNC 1.0 (1995)、BNC World Edition (2000)、BNC XML Edition (2007) がある。

2．PGR Level 3 〜 6 の各10タイトルの第 1 章から本文40ページを目処にテキストデータ化したところ、各英文テキストのtype数は、1000語前後となった。以下は、タイトルリスト。

Penguin Level 3

Beddall, F.	2008	*A history of Britain*
Chaucer, G.	2008	*Canterbury tales*
Degnan-Veness, C.	2003	*Martin Luther King*
Gilchrist, C	1998	*Princess Diana*

Lerner, A. J.	2008	*My fair lady*
Mikes, G.	1999	*How to be an alien*
Poe, E.	2008	*Fall of the house of Usher and other stories, the*
Shipton, V.	2008	*Grey owl*
Stoker, B.	2008	*Dracula*
Twain, M.	2008	*Jim Smiley and his jumping frog and other stories*

Penguin Level 4

Beddall, F.	2008	*Alexander the great*
Cerasini, M.	2008	*Cinderella man*
Forsyth, F.	2008	*Days of Jackal*
Gram, D.	2008	*Gladiator*
Mansfield C.	2008	*Doll's house and other stories, the*
Marlove, C.	2008	*Doctor Faustus*
Maule, D.	2008	*Inventions that changed the world*
Mitchell, M.	2008	*Gone with the wind*
Puzo, M.	2008	*God father*
Shute, N.	2008	*On the beach*

Penguin Level 5

Conrad, J.	2007	*Heart of darkness*
Curtis, R.	2008	*Four weddings and a funeral*
Du Maurier, D.	2008	*Jamaica inn*
Grisham, J.	2008	*Brethren, the*
Highsmith, P.	2008	*Ripley's game*
Kerouac, J.	2008	*On the road*
King, S.	1999	*Body, the*
Lawrence, D.	2008	*British and American short stories*
Leroux, G.	2008	*Phantom of the opera, the*
Wells, H.	2008	*Invisible man, the*

Penguin Level 6

Bryant, S.	2008	*Business @ the speed of thought*
Cartledge, H. A.	2008	*Brave new world*
Doss, L.	2008	*Great expectations*
Durham, R.H.	2008	*Cry, the beloved country*
Gladwin, M.	2008	*East of Eden*
Harmes, S.	2008	*Chamber, the*
Kehl, J.	2008	*I know why the caged bird sings*
Maule, D.	2008	*Double helix, the*
Strange, J.	2008	*Beach, the*

Tolstoy, R.	2008	*Anna Karenina*

3．V8an 2015出力例（Anna Kareninaテキストの入力結果の一部）

Results of "Chapter 1 Affairs of..."

	level 1	level 2	level 3	level 4	level 5	level 6	level 7	level 8	over 8	cont. forms	non-words	proper nouns	total
indexes	545	193	82	52	29	12	16	10	41	14	3	28	1025
%	53.171	18.829	8	5.073	2.829	1.171	1.561	0.976	4	1.366	0.293	2.732	100
tokens	5677	360	133	118	72	18	22	34	55	175	3	438	7105
%	79.901	5.067	1.872	1.661	1.013	0.253	0.31	0.479	0.774	2.463	0.042	6.165	100

i/t = 14.426 %

(1) I 93
(1) Monday 1
(1) a 140
(1) about 23
(1) abroad 1

4．Brezina1 and Gablasova (2013)以外にも、Browne, C., Culligan, B. and Phillips, J.(2013)がCambridge English Corpus (CEC)をもとに策定したNew General Service Listがある。(http://www.newgeneralservicelist.org/参照)

5．LVCでは、Longman 9000 keywordsをLongman core vocabulary (high frequency、medium frequency、low frequency）の3レベルに分けている。調査時点では、toはhigh frequency、medium frequencyの両レベルでヒットしている。(ソフトウェアがレベル判定時に参照する基準リストの不整合ではないだろうか。)

6．出現頻度の100万語換算は、McEnery and Hardie（2012:pp.49-50）に準拠した。

nf=(number of examples of the word in the whole corpus)÷(size of corpus)×1,000,000

7．使用したPGR6は、注1掲載の10タイトルと下記の11タイトルで、第1章から本文40ページを目処にテキストデータ化した。WS6によると、typeの平均値は1345、最小値は1097　最大値は1539であった。

Collins, M.	2008	*Saving private Ryan*
Collins, W.	2008	*Woman in white, the*
Dahl, R.	2011	*Man from the south and other stories*
Dickens, C.	2008	*Oliver Twist*
Fielding, H.	2008	*Henry Fielding*
Flaubert, G.	2008	*Madam Bovary*
Gaskell, E.	2009	*North and South*
Grisham, J.	2011	*Runaway jury*
Grisham, J.	2011	*Testament, the*
Keneally, T.	2008	*Schindler's list*
Rice, C.	2008	*Remains of the day*

参照文献

（英　文）

Adam Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Pavel Rychlý, P., and Suchomel, V. (2014).
"The Sketch Engine: ten years on in *Lexicography*" July 2014, Volume 1, Issue 1, pp 7-36.

Adolphs, S. and Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal*. NY: Routledge

Adolphs, S. and Schimitt, N. (2004). "Vocabulary coverage according to spoken discourse context" in Bogaards, P. and Laufer, B. (2004). *Vocabulary in a second language*. Amsterdam: John Benjamins Publishing Company.

Brezina1, V. and Gablasova, D. (2013) . "Is there a core general vocabulary? Introducing the new general service list." *Applied Linguistics* (2013) pp.1-13.(doi: 10.1093/applin/amt018)

Leech, G., Rayson, P., and Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British national corpus*. Oxon: Routledge

Mayor, M. ed. (2009). Longman defining vocabulary in *Longman dictionary of contemporary English*,. Harlow: Pearson Education Ltd. pp.2060-2071.

McEnery, T. and Hardie, A. (2013). *Corpus linguistics*. Cambridge: Cambridge University Press.

Moon, R. (2012). "What can a corpus tell us about lexis?" in O'Keeffe, A. and McCarthy, M. (2012). *The Routledge handbook of corpus linguistics*. Oxon: Routledge. pp.197-211.

Nation, P. (2013). *What should every EFL teacher know?* Seoul: Compass Publishing.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London: Palgrave Macmillan.

Summers, D. ed. (2007). Longman American defining vocabulary in *Longman Advanced American dictionary*. Harlow: Pearson Education Ltd.pp.1853-1860.

Timmis, I. (2015). Corpus linguistics for ELT: Research and practice. London: Routledge. p.42

West, M. (1953). *A general service list of English words*. London: Longman Group Ltd.

（和　文）

上村俊彦（2014）「グレイディッド・リーダーの語彙と文法」長崎県立大学国際情報学部『研究紀要』第15号 pp.171-184.

URL

（2015年 9 月29日現在）

（コーパス）

BASE	http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/history/
BNC	http://www.natcorp.ox.ac.uk/docs/URG/intro.html
ICE-GB（spoken）	http://www.ice-corpora.net/ice/icegb.htm

MICASE	http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple
OANC　Face-to-face	http://www.anc.org/data/oanc/download/
OANC　Switchboard	http://www.anc.org/data/oanc/download/

（ソフトウェア）

Longman 9000 keywords	http://global.longmandictionaries.com/vocabulary_checker
Oxford Text Checker	http://www.oxfordlearnersdictionaries.com/oxford_3000_profiler
v8an 2015	http://mochvocab.sakura.ne.jp/cgi-bin/j8web/j8web.cgi
WordSmith Tools	http://lexically.net/wordsmith/