

## デンタル・トリビューンの英語語彙レベル

田嶋 倫雄

### Vocabulary levels of Dental Tribune International

Michio Tajima

#### Abstract

This study intends to find a desirable vocabulary level to comprehend professional journals in dentistry. The corpus was constructed by gathering articles from the e-newsletters of Dental Tribune International which were analyzed using Vocabulary Profilers (VP) and v8an. The VP result showed that the K3 words level has an 84.80% coverage of the corpus and v8an, 75.27%. This indicates that the basic 3000 words, the high school graduate level in Japan, are not enough to comprehend professional journals without checking words and learners need to acquire the whole word list of VP or v8an and specific dental words to achieve an acceptable level of comprehension, supposedly a 95% text coverage. The result reaffirms the importance of focus on EAP and/or ESP for effective vocabulary learning.

**Key words**: vocabulary level, corpus, dental words, ESP

#### 1. はじめに

第二言語または外国語で、専門分野の情報誌、雑誌、学術論文などが容易に通読できると考える者は少ないであろう。内容に伴い語彙が専門的になり難易度が上がるのも想像するのは難しくない。専門的内容の記事の難易度は、もちろん内容に関わることが第一に理由としてあげられるが、語彙そのものの知識が必要であることは疑う必要もない。

学術的内容のみでなく、専門的な分野の記事を英文で読み込んでいくとなれば、その分野の専門知識にともない専門用語の習得は必須である。一般分野の英文が読める程度の英語語彙力では、到底太刀打ちできないと考えるのが至極当然であろう。

外国語学習において語彙の習得は重要な要因の一つといえる。Morris and Cobb (2004) は、

語彙力はその言語の成績と相関している点を紹介しているし、Akbarian(2010)は語彙量(size)と語彙理解度(depth)は関連しているとし、言語学習上級者は語彙力が高ければ高いほど語の深い理解度があることを示している。語彙力はある程度は言語能力の指標といっても良さそうなほどである。

ではその問題となる英語の語彙の難易度はどの程度なのか。専門用語とされるのは全体の語彙の数に対し何割くらいあるのか。全体としての語彙知識はどのくらいあれば、特定の分野のテキストが読みこなせるのか。本論では、オンライン版 Dental Tribune International (DTI) を例にとり、内容理解を前提としてどの程度の語彙レベルで、どのくらいの英語語彙数が理想的習得であるのかを、探り出すことを目的としている。日本語を母国語とする者が、英語で書かれた歯科医向けの雑誌や専門誌を読むこと

が、日本において不可欠であるかは別問題として、そういった活字を読む際に必要であろうと想定される語彙数とレベルを知っておくのは意義あることといえよう。

## 2. 英語語彙レベルについて

英語全般において語彙レベルの指標は数多くある。最も著名なものは Michael West (1953) による A General Service List (GSL) と Coxhead (2000) の The Academic Word Level (AWL) を使用した Range and Frequency program であろう (現在は VocabProfilers (VP) としてオンラインで使用可能)。GSL は半世紀以上も前のものであるため、更新の末 2013 年 10 月に最新版が発表されている。さらに、現在の英語に合わせたものを作成し直した Charles Browne (2013) の A New General Service List (NGSL) もある。

日本人英語学習者向けに作成された JACET8000 や, SVL12000 (スペースアルク英語) なども代表例として挙げられよう。SVL12000 は 12 レベル分けされた語彙が、各レベルに 1,000 語あり、大学受験前に覚える英単語がレベル 5 とされている。アルクは大学受験には基本語からはじめ 5,000 語の語彙を必要と判断しているのである。また、Oxford University Press や Cambridge University Press も headwords という語を使用して、語彙レベルに合わせたレベル別読み物が揃えられているが、語彙の選定方法や語彙リストは、一般には公表されていないそうである。

GSL も NGSL も JACET8000 も、インターネット上のサービスで、英文テキストの語彙レベルを確認できる。例えば、GSL で調べる際は、VP classic で使用される GSL+AWL だけでなく、VP Compleat 版 (2013 年 10 月現在) がある。Compleat 版には classic も含まれていて、他に GSL+AWL を統合して使いにくさを

修正した BNL, や Brown の NGSL がある。また、British National Corpus (BNC) や Corpus of Contemporary American English (COCA) などが元になっているコーパスを統合した 25 段階レベルを基本的にそれぞれ 1,000 語ずつのリスト、合計 25,000 語程度で構成されている BNC-COCA-25 などからも選んで分析ができる。また、リストに含まれない語 (offlist) も表示することができる。

VP ではテキストを分析する際に、家族語 (word family) で表示する。同じ語幹を持つ語であるなら、語の基本形、三人称単数形、過去形、進行形、形容詞形、名詞形、接頭辞や接尾辞の含む語などは一つの家族とみなされる。例えば、accept, accepts, accepted, accepting, acceptable, acceptance, unacceptable である。しかし、家族語は、同じ語幹を持つ語は品詞の違いがあるにしても同じ家族扱いされてしまうので、必ずしも機械分析上一致しない結果が出現するのは避けられない。例えば、book という語を動詞であっても名詞であっても、一つとみなしてしまうという弱点がある。

JACET8000 分析プログラム (v8an) もそれぞれのレベルを 1,000 語ずつで分け、8 段階の語彙レベルとそれ以上の Over 8, カンマなどが含まれる省略形 (Cont. forms), 数字や数字と記号 (Non-words), 固有名詞 (Proper nouns) などの分類で、語彙のレベルを確認できる。家族語表示である VP に対し、v8an ではテキストをレマ化して語彙をランク分けしている。レマ化とは、辞書で確認できる見出し語で同じ品詞の活用形は一つにまとめられているということである。例えば、動詞の look と名詞としての look は別の語としてカウントされているはずなのである。JACET8000 は 1 億語の BNC と日本人英語学習者の環境を踏まえて作成されたサブコーパスを頼りに、コンピューター使用して作成された語彙リストであり、人的判断に

頼って作成された JACET4000 の改良版である (そのため、より正確と考えてもいいだろう)。よって VP と v8an の両方の語彙レベル確認はある程度の差があるというのが前提とするべきである。

英文を読む際に、未知の語彙を調べて意味確認をせず、文脈から意味を推測することで、ほぼ問題なく読み進めることができるとされる語彙カバー率は、95%といわれている (Laufer, 1992, Browne, 2013)。さらに Read(2000)は、英語を母国語としない者は最低でも95%の語彙を知っていることが必要としている。より認知負荷を下げるには、限りなく100%の語彙カバー率に近づけるのが重要であろう。実際、Van Zeeland and Schmitt (2013)は、98%が十分なカバー率と述べてもいる。一般的な英文でも95%の語彙を知っている状態になるには、3,000語程度の語彙力が必要とされる (Laufer, 1992)。語彙とは関係なく専門的な内容の文脈を理解するための知識も必要であろう。しかし、3,000語の語彙力を必要最低語彙力と見立て、それ以上の語彙とさらに専門用語の語彙力が、専門分野の記事や学術論文を理解するには必要だとするのは妥当のようである。

Browne (2013)はGSLとNGSLのCambridge English Corpus (CEC) 語彙カバー率を測っているが、GSLが84.24% (1964 word families)でNGSLが90.34% (2368 word families)と結果を出している。一般的な英文においてはGSLのカバー率は84%と判断していいだろう。もちろん固有名詞などは既知語とされているので、それらを認識できればカバー率はもっと高くなる。読解において語彙カバー率95%を必要とするなら、GSLだけでは一般英語は十分な理解ができるだけの語彙力でないことを意味する。必要最低限として3,000語というのも納得がいく。

そこでDTIを辞書を参照せずに読めるため

の95%の語彙力とは、どの程度の語彙レベルなのか疑問となる。何語レベルの語彙を習得しておくことが理想なのかがある程度判断できれば、今後の語彙目標設定値としての指針とすることもできるであろう。

### 3. コーパス構築と語彙レベル確認

独自のDTIコーパス構築には、手始めとしてDTIからのメーリングリストで紹介される代表的な記事で平成25年11月4日から平成26年4月14日までに配信されたものを収集することとした。

DTIのメーリングリストは、一回の配信で厳選された最新記事が四件、企業紹介欄が一件、DTIのオンライン版の紹介や広告などが混ざっているのが通常のものである。今回の語彙レベル確認では、各メーリングリストで配信される四つの記事のみを収集し取り扱うことにした。Top Story, News, Business, Dental Tribune Partner, E-paperなどのカテゴリがあり、出どころは、Dental TribuneのInternational版, America, Europe, Asia Pacific, Middle East & Africa版であったりする。各記事のタイトルと、記事の出典も含めた記事中に記載された地名・人名を全て分析に含めたが、その他の要因としての、記事執筆者名、企業紹介記事、広告、記事中の写真の説明、e-paper (print edition online)などはすべて省略し、テキストを調整した。収集した全ての記事の総語彙数が三万語を超えた時点で、本論のための収集は終了とした。ただし、三万語を超えた回のメーリングリストにおいては、その回の英文をコーパスに含めた。

語彙レベル確認には、VPとv8anの両方を採用し、テキストを分析にかけ、コーパスの語彙レベル分け、頻出数、累積率などを比較検討した。さらに目安として、テキストのカバー率が95%に近づく語彙レベルを確認した。

### 4. 結 果

総語彙数は、VP の結果では 31,281 語、v8an では 30,923 語と、数百語の差が生じた (Table 1, Table 2 参照)。これら数百語の差は特殊文字の受け入れ具合や、ハイフンによる連語や行送り語などが考えられるが、全体のレベル別出現

頻度の割合に、大きな問題を起こすとは考えにくいため、そのまま受け入れることとした。

語種 (types : 同じ語を一語と数える) と語彙の総出現数 (token : 同一語も出現するごとにカウントする) の割合である Type-token ratio が低いと、語彙の幅が少ないといえる。高いと、テキスト量に対し、多様に富んだ語彙を多く使

Table 1. VocabProfilers (VP) output with BNC-COCA-25

Freq. Level	Families(%)	Types (%)	Tokens (%)	Cumul. token%
K-1 Words :	631(27.88)	1122(27.83)	19481(62.28)	62.28
K-2 Words :	453(20.02)	776(19.25)	4097(13.10)	75.38
K-3 Words :	528(23.33)	786(19.50)	2946( 9.42)	84.80
K-4 Words :	190( 8.40)	220( 5.46)	560( 1.79)	86.59
K-5 Words :	103( 4.55)	119( 2.95)	482( 1.54)	88.13
K-6 Words :	68( 3.00)	77( 1.91)	627( 2.00)	90.13
K-7 Words :	58( 2.56)	65( 1.61)	191( 0.61)	90.74
K-8 Words :	52( 2.30)	55( 1.36)	107( 0.34)	91.08
K-9 Words :	23( 1.02)	23( 0.57)	32( 0.10)	91.18
K-10 Words :	32( 1.41)	35( 0.87)	92( 0.29)	91.47
K-11 Words :	22( 0.97)	28( 0.69)	47( 0.15)	91.62
K-12 Words :	13( 0.57)	13( 0.32)	19( 0.06)	91.68
K-13 Words :	15( 0.66)	17( 0.42)	105( 0.34)	92.02
K-14 Words :	11( 0.49)	11( 0.27)	22( 0.07)	92.09
K-15 Words :	4( 0.18)	4( 0.10)	5( 0.02)	92.11
K-16 Words :	9( 0.40)	10( 0.25)	19( 0.06)	92.17
K-17 Words :	5( 0.22)	5( 0.12)	24( 0.08)	92.25
K-18 Words :	6( 0.27)	6( 0.15)	6( 0.02)	92.27
K-19 Words :	17( 0.75)	17( 0.42)	36( 0.12)	92.39
K-20 Words :	4( 0.18)	4( 0.10)	7( 0.02)	92.41
K-21 Words :	5( 0.22)	5( 0.12)	8( 0.03)	92.44
K-22 Words :	2( 0.09)	2( 0.05)	7( 0.02)	92.46
K-23 Words :	4( 0.18)	4( 0.10)	10( 0.03)	92.49
K-24 Words :	1( 0.04)	1( 0.02)	3( 0.01)	92.50
K-25 Words :	7( 0.31)	8( 0.20)	23( 0.07)	92.57
Off-List :	??	809(20.07)	2325( 7.43)	100
Total (unrounded)	2263+?	4031( 100)	31281( 100)	100

Families は家族語, Types はそれぞれの同一語, Tokens は総出現数。Cumul. token は Tokens の累積。Type-token ratio : 0.13, Tokens per type : 7.76

Table 2. v8an-revised web edition output

	indexes	%	tokens	%	Cumul. token%
level 1	857	19.669	19488	63.021	63.021
level 2	501	11.499	2781	8.993	72.014
level 3	284	6.518	1008	3.260	75.274
level 4	405	9.295	1864	6.028	81.302
level 5	200	4.590	639	2.066	83.368
level 6	131	3.007	386	1.248	84.616
level 7	85	1.951	614	1.986	86.602
level 8	58	1.331	120	0.388	86.990
over 8	833	19.119	1555	5.029	92.019
cont. forms	18	0.413	183	0.592	92.611
non-words	245	5.623	587	1.898	94.509
proper nouns	740	16.984	1698	5.491	100
total	4357	100	30923	100	
i/t=14.09%					

index は見出し語, tokens は総出現数, Cumul. tokens は, Tokens の累積。

用していることが分かる。VPの結果の Table 1 にも表示してある 0.13 は決して高くないので、見出し語だけで判断すると、DTI の語彙の多様性は低いということが伺える。

また、Table 2 で表示してある「i/t」は、indexes の語数と tokens の語数の割合を表す指標で、「indexes の総語数 ÷ tokens の総語数 × 100」で算出される。indexes はレマ化された語の出現数を表し、tokens は通常の語数の数え方で表された語の出現数である。例えば、「look, looked, looking」の 3 語は、index 換算で 1, token 換算で 3 となる。「i/t」の値が小さいということは、同じ語が多用されていることを表しており、一般的には平易な文といえる。清水 (2003) によれば英字新聞が 10% をやっと超える程度としていて、v8an の結果の 14.09% は、DTI の方が語彙の繰り返しが多いことを示唆している。

VP の結果の Off-List の延べ語数 (Tokens) は 7.43% で、語種 (Types) は 20.07% にもなっ

ている。Off-List の語を確認してみると、「&」や「@」など理解が容易な記号、「africa」、「japan」、「kenya」など地名、「julie」、「michael」、「scott」などの人名、「overexpression」、「school-children」、「worldwide」らのように複合語を一つの語として語の間にスペースを置かずに表記したものや、単純な入力ミスにより二語が一語扱いされてしまっているものも多く含まれているので、一概に判断はできない。これらの比較的理解容易な語を調整すると実際の語彙レベル表は基礎的なレベルの語の割合が増えるであろう。その反面、歯学系の用語と判断できる語も多く Off-List の中に含まれている。「abutment」、「gingivalis」、「mandibular」、「occlusal」、「prosthodontics」などは歯科分野でもよく見かける用語に入るといえよう。これらの基礎的な語と歯科用語を単純に半分の割合と仮定し、基礎語を含めるならば 95% のカバー率に達するのは、K-10 words のレベルといえる。

数字や記号などの non-words として表して

いる v8an の tokens は全体の 1.898%あり、また人名、地名、省略形などの proper nouns は 5.491%である。これらを既知語と捉え、さらに省略形の cont. forms を含めると 100%になるわけだが、問題となるのは Over 8 の語彙であろう。つまり、Over 8 の語彙次第であるが、この部分を専門分野によって変動するであろう語彙とし、それらを既習語とするならば、DTI は理解できると言えそうである。単純に考えるならば、VP で見ても、v8an でみても特定分野の語彙を身に付けていると仮定して 10,000 語彙の習得で DTI は読みこなせそうである。

日本の高等学校の英語教育で学習すると思われる約 3,000 語と同じ数字の 3,000 語を超えるレベルを目安にすると、VP の K-3 words までで DTI の語彙カバー率が 84.8%、v8an の Level 3 までで 75.274%しか及んでいない。二つの語彙リストのうち、基本語からみた 3,000 語にこれほどの違いが出るのも問題があるが、前述した Off-List, non-words, proper nouns のうち理解できる語が多く含まれることを考慮しても、語彙カバー率 95%には残念ながら及ばないだろう。

VP の Off-List や v8an の proper nouns ではなく、難易度の高い語のリスト内に含まれている語彙という意味で、VP の K-4 から K-25 までのレベルに含まれるとされる語は 7.95%であり、v8an の Level 4 から Over 8 までの語は 16.745%も占める。

## 5. 考 察

英文を読む際に、辞書などの利用による語彙確認をせずとも、理解できると推定される 95%の語彙カバー率に達するにはどの程度の語彙力が必要であるか。VP の結果は、K-25 words のレベルでも 92.57%である (Table 1)。また、v8an の結果では、Level 8 では 86.99%である (Table 2)。つまり、v8an の結果を見ただけでも

語彙数が 8,000 あってもまだ十分ではない。語の意味を調べることなく、容易に理解するだけの英語力というものを目指すのは、至難の業であることが明白である。

VP の語の総出現数累積割合が 90%を K-6 words で超えているものの、K-25 words までいっても 93%に満たしていない。つまり 25,000 語まで習得しなくても、6,000 か 7,000 語レベルで、9 割は理解可能な語彙だと判断し、それ以降は常に辞書などで確認することを前提とすれば良いのかもしれない。同様に、DTI の 9 割の語彙を習得したといえるようにするには、JACET8000 の語彙リスト全てを網羅することが相当すると判断ができる。しかし、それを実行するのは気が遠くなる作業であるし、そもそも現実的ではないだろう。

別の問題として、レベルの高くない語も複合語または連語としては専門語彙になるという点がある。例えば、v8an では「chief」は Level 2 で、「complaint」は Level 4 の語であり、難易度は高くない。しかし、歯学分野では「chief complaint」は「主訴」と訳され一般的にはおそらく多くの者が「主な訴え」などと和訳するであろう。また「lower」も「jaw」も Level 3 の語であるが、「lower jaw」は「したあご」とする一般語ではなく、「下顎」など複合名詞の例を挙げたらきりが無い。よって語彙レベルだけで判断できない歯学用語の難しさも考慮しなくてはならない。

専門用語とはどう判断されるのか。大別すると機械的に判断する方法と、人的努力により判断する方法がある。特定の分野に精通する専門家や専門用語集、さらにコンピューターで出現頻度を確認するなど四つの方法を提案した Chung and Nation (2004) も、九つの方法で測定し英文内で使われる専門用語の抽出を比較した Chujo & Utiyama (2006) も、また Vocabulary classifications, Keyword analysis, term

extraction, systematic classifications の四つの方法のうち、二つを組み合わせた hybrid method を紹介している Kwary (2011) などもあるが、どれも結果的には人間の目で確認し、選定することを排除できていない。つまり専門用語であるという判断は、コンピューターの測定だけで機械的にはできず、専門とする者が判断する事が必須と考えて良いのであろう。

## 6. 最後

語彙レベル確認ソフトにより、DTI で読まれる英文の語彙の難易度はある程度確認できた。出来るだけ多くの数の語彙を、可能な限り手早い習得を目指すという学習効率を考慮するならば、つまり頻出度の高い語彙から理解し習得を目指すのが得策とするなら、本論で触れた語彙レベルと頻出数の確認作業は、間違いなく有効である。

本論で扱った DTI の全ての記事の語彙数は三万語を超える程度の小さいコーパスであるが、継続することでより正確な語彙レベルの確認ができるようになるであろう。また、DTI だけでなく、他の歯科系英文テキストも精査することで更にこの分野の語彙レベルが明確にできるであろう。頻出度による語彙の重要度も精査することと合わせ、歯科分野における英語語彙の抽出に貢献できることを切に願う。

### 参考文献

- Akbarian, I. (2010) The relationship between vocabulary size and depth for ESP/EAP learners. *System* 38, 391-401
- Browne, C. (2013) A new general service vocabulary list: Helping students help themselves. *The Language Teacher* 37(6), 43
- Coxhead, A. (2000) A New Academic Word List. *TESOL Quarterly* 34(2), 213-238
- Chung, T.M., & Nation, P. (2004) Identifying technical vocabulary. *System* 32, 251-263
- Chujo, K., & Utiyama, M. (2006) Selecting level-specific specialized vocabulary using statistical measures. *System* 34, 255-269
- Kwary, D.A. (2011) A hybrid method for determining technical vocabulary. *System* 39, 175-185
- Laufer, B. (1992) How much lexis is necessary for reading comprehension? In Arnaud, P., & Bejoint, H. (eds), *Vocabulary and applied linguistics*, London: Macmillan, 125-132
- Morris, L., & Cobb, T. (2004) Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System* 32, 75-87
- Read, J. (2000) *Assessing vocabulary*. Cambridge: Cambridge UP
- Van Zeeland, H., & Schmitt, N. (2013) Lexical coverage in L1 and L2 listening comprehension. The same or different from reading comprehension? *Applied Linguistics* 34(4), 457-479
- Vocabulary Profilers. The complete web VP! Retrieved July 17, 2014, from <http://www.lex-tutor.ca/vp/bnc/v8an>. Revised web edition. <http://www.tcp-ip.or.jp/~shim/j8web/j8web.cgi>
- West, M. (1953) *A General Service List of English Words*. London: Longman
- 清水伸一 (2003) JACET8000 付属 CD-ROM のプログラムデータ使用法. 第 42 回 JACET 全国大会. <http://www.tcp-ip.or.jp/~shim/J8.htm>
- スペースアルク英語. レベル別語彙リスト SVL 12000. Retrieved July 17, 2014, from <http://www.alc.co.jp/eng/vocab/svl/>