

# 東北大学全学教育における英語読解用語彙の選定について

岡田 毅<sup>1)</sup>\*

1) 東北大学国際文化研究科

## 1. 序論

1980年代半ばから、応用言語学の観点に基づいて学術的な言語使用 (academic discourse) に関する研究が盛んになった。これは、コーパス言語学が大きな成果を挙げはじめ、言語使用域 (register) ごとに言語特徴は大きく異なるという事実を客観的に記述することが可能になったからである。言語使用域の典型的な区分は、「話し言葉」と「書き言葉」という伝統的な二分法を想起させるが、精密な設計に裏付けされた大規模なコーパスが、英語を中心に次々と開発・公開されるに至って、単純な二分法というよりもむしろ、具体的な言語使用の場という意味での使用域を正確に定義し分類することによって、直観的にしか捉えられてこなかった言語使用の特徴と差異が、科学的・帰納的に実態として記述できるようになってきたわけである。このような進展は、言語教育、とりわけESL (English as a second language) やEFL (English as a foreign language) 教育の場に大きな貢献をなすものと考えられるようになり、学習者が必要とするであろう種類の英語を教えるべきであるという実践的な英語教育の立場から注目を浴びることになった。

均衡コーパス (balanced corpus) に求められる代表性 (representativeness) は大きな問題であり、それ自体が重要な研究課題でもあり得る。均衡コーパスとは性格を異にする特殊コーパス (specialized corpus) は、「特定の研究目的のために、特定の言語使用域のデータを収集し整備したコーパス」と定義することができ、コーパス研究に関わるさまざまな研究成果と技法とを存分に応用し当てはめることのできる研究対象である。<sup>注1)</sup>

大学等の高等教育機関での勉学を目的とする英語学習者にとって、大学で用いられる話し言葉や書き言葉という使用域が最も重要なことは自明であり、彼らにとってのよりよい学習教材を提供するための準備として、これらの使用域に対する十分な記述と分析が求められてきており、多くの研究成果が発表され専門の学術雑誌が出版されてきている。<sup>注2)</sup> Biber (2006a, 2006b) などでは、これまで主に書き言葉を中心として進められてきた、学術目的の英語 (English for academic purposes, 以下EAP) の研究の射程を話し言葉にまで大きく拡張し、多角的な分析を総合的に行っている。

本研究は、EAPの中で、敢えて書き言葉に使用域を限定し、学習者が身に付けるべき英語語彙を新しい視点から改めて見出そうとするものである。それには以下のような理由がある。本研究の目的は、「学術目的を持った英語学習者」(learners with academic goals) という概念を「研究型大学である東北大学の1、2年次に在籍中の日本人大学生」に特定化し、英語母語話者や第2言語として英語を用いる大学生との間に境界を設け、国内外の研究成果を慎重に踏まえることによって、主に教養教育課程等で実施される大学英語教育におけるリーディング指導にとって最適な語彙群を認定し選出しようとするものであるために、書き言葉を第一義的な分析および教材作成の素材とするのである。また、ここでの研究は、これまで北海道大学、東京大学、京都大学などから公開されてきたような、学術語彙リスト<sup>注3)</sup>の形を目指すのではなく、東北大学の全学教育の中での実際の英語授業で採用されるべき(主・副を問わず)リーディング教材の開発支援へ

\* ) 連絡先：〒980-8576 宮城県仙台市青葉区川内41 東北大学大学院国際文化研究科 t-okada@intcul.tohoku.ac.jp

と繋がることを企図している。そのためには、多様な専門的背景を持つ英語教員に応じた教材の作成や選定に関わる一定の方向性を柔軟な姿勢で模索する必要がある。ここでは筆者が共同研究で推進中の統合的なコーパス解析システム (Okada and Sakamoto, 2010) との関連性が大きな意義を持つ。

本論の目的は (1) 大学英語教育における最近の議論を主に語彙の観点から整理し直し, (2) 実際の語彙リストの性格を踏まえながら, それらの出現傾向をコーパス分析を通して明らかにし, (3) 東北大学全学教育におけるリーディング教材開発のための指針を提案することにある。

## 2. 学術英語語彙

### 2-1. EAP 論争

従来のEAP学習のための語彙の選出に関わる研究では、英語母語話者と非英語母語話者という大きな区分を設けるのみで、日本人英語学習者とりわけ大学の学部学生という特定の学習者層を設定していない。本研究の課題は、学習者を「学術目的を持った英語学習者の中でも、非英語母語話者である日本人大学生」と明確に設定したうえで、Coxhead (2000) で提唱されたAWL (academic word list) に代表される英単語リストが、どのように有効であるのかを、大学英語教育学会 (JACET) の公表している語彙集や、京都大学や北海道大学が独自に提案している英語語彙リストとどのような相関関係を持つのかを精査することにある。それによって、東北大学の1, 2年次学生にとっての最良の英語リーディング教材の選定および作成の基礎を構築するというのが大きな課題である。

そのためには、(1) 従来のGSLやAWLは日本人大学生にとってどの程度有益なのか、(2) JACETや日本の大学から提案されている語彙は、学術目的の英語にとってどの程度有効なのか、を入念に調査する必要がある。また理念追求に終始するのではなく、実際のリーディング教材作成を射程に入れながら、最近のEAPに関する議論を、本学の学部1, 2年次学生に対する全学教育というコンテキストの中で精密化する必要がある。

竹蓋・水光 (2005) では、大学での英語教育の目標

をEAPと規定し、EGPは大学入学以前に習得を済ませ、入学後のEAP教育をEGAPとESAPとに大別するという明確な分類に基づいた提案がなされている。

…専門性が高ければ高いほど、各分野のESAPの差が大きくなることが予想される。逆に専門性がさほど高くない段階においては、分野間の差はほとんどないと見てよいであろう。たとえば論文の執筆や学会での口頭発表などに求められる言語技能は、どの専門分野にもある程度共通したものである。つまり、EAPの中に、各分野に共通するEGAPと、各分野に固有のESAPとを想定することが可能である。(竹蓋・水光, 2005: 12)

これはEAPをEGAP (一般学術目的の英語) とESAP (特定学術目的の英語) に下位区分するJordan (1997) の考え方を踏まえたものである。

### 2-2. EGAP と ESAP

竹蓋・水光は、大学での英語教育の目標をEAPと定めながらも、全学共通科目での英語をEGAPとし、3,4年次の学部やその後の大学院での専門英語をESAPと明確に設定している。この考え方は、日本の特に研究型大学における英語教育の目標と理念を明確に示すための分かりやすい指針と言える。しかし、「EGAPとESAPとは連続体を構成する」(同書:12)と、慎重な姿勢は示しているものの、Jordanの比較的単純な区分を全面的に安易に受け入れすぎているのではないかという疑問が残る。本当に、論文執筆や口頭発表の能力は専門分野で差の見られない共通のもののだろうか。また、この考え方は学部1,2年次の学生にとってのEGAP習得の目標を、どちらかと言えばESAPに近い側に置いているとは言えないだろうか。これに関しては、さらに以下のような論述が見られる。

ただしその場合、コース設計においては、授業で扱う「内容」と「学術的言語技能」とを区別して論じることが必要である。内容には、各専門分野での研究に即した題材やその分野の言語的な特徴などが含まれる。他方、学術的言語技能は、具体的には、た

例えば学術論文を執筆するためのアカデミックライティングや論文を読むためのアカデミックリーディング、講義を理解するためのアカデミックリスニング、学会での口頭発表で求められるオーラルプレゼンテーションなどの技能を含んでいる。

内容と学術的言語技能を区別して考えた場合、1～2年次生対照の全学教育科目の英語において、主として英文学、英語学、英語教育学を専門とする英語担当教員が各自の専門性のある程度活かしつつ、法学、工学、医学などの他分野専攻の学生を指導する状況が想定できる。この場合、1～2年次生については、授業で扱う教材や内容面で学生の専攻する専門分野との関連性は必ずしも高くないが、他方、学術的言語技能で学生の専門分野に通じる技能が射程に含まれることになる。そして3～4年次や大学院に進むにつれて、次第に内容と学術的言語技能の両面において専門分野との距離を縮めていくことになる。(同書: 12-13)

ここで言う、分野に特定の言語的な特徴というのは「内容」ではなく表現方法、好まれる構文や表現といった学術的言語「技能」なのではないのだろうか。また、「学術的言語技能」という能力は相当に専門的で、受容したり発信したりする学術的な内容が確立されていない1、2年次生にこれに基づいた英語学習への動機付けを期待することはできないというのも、全学教育の英語を担当する教員ならだれでもが頷首することであろう。

そして、上の引用部分の後段は、全学教育における英語教育の現状を余りにも安易に前提として受け入れすぎているのではなかろうか。「担当教員が各自の専門性のある程度活かしつつ」という表現は、現実的な諸問題からによるものということは理解できるが、まさにこの点こそが克服しがたい障害なのではなかろうか。

また、「内容面で学生の専攻する専門分野との関連性は必ずしも高くないが」というが、1、2年次の学生が自らの分野について、とりわけ英語で発信すべき専門的な知識やコンテンツを習得しているとは考えにくい。ここでも、中等教育における英語教育が、ESL

ではなくEFLとして不可避免的に内包しているEGP教育でのシミュレーションという問題が発生してしまう。要するに、「将来諸君は何某の専門分野でこのような英語を扱わねばならないから、そのために基礎となる学術的言語技能を身に付けなければならない」という一見もっともそうでも、非常に漠とした前提やシミュレーションに基づいた教育を展開しなければならないのである。また、竹蓋・水光は「ESAPの内容を視野に入れつつ、連続体であるEGAPとESAPの線上に」大学の英語教育の到達目標を設定している、と言うが、その線とは、後段で言うところの1、2年次から3、4年次、大学院という年次経過に影響を受けざるを得ない傾斜線ではなかろうか。全学教育における英語教育の目標を、彼らのスタンスで実現しようとするならば、学部や大学院の年次進行に合わせる必要はなく、むしろ専門的な知識が増加し、受容・発信すべきコンテンツに対する姿勢やニーズが明確に高まった3、4年次や大学院の段階で、EGAPをさまざまな専門性を持った英語教員が教えた方が効果が上がるということになる。

さらに「大学入学までにEGPを終え、…」(同書: 13)という前提は余りにも大胆すぎると感じられる。それでは、大学という場で用いられる英語は、全てEAPということになる。果たしてそうなのだろうか。「学術研究に資する英語教育を目指すことは、理念として学術研究を掲げている大学にとっては当然の論理的帰結である。」(同書: 13)は正しい主張であるが、大学で用いられる英語に幅を認め、「大学の世界で学術的な生活を送り、学術的な成果を上げることを最終的な目標に掲げるのが当然な大学生」にとって必要な英語能力とは何かという問題に取り組む必要がある。本論はそれに答えようとする試みである。

田地野(2004: 13)では、大学英語教育におけるEGPという概念中のG(General)の部分に不明瞭が残る、と指摘した上で、EGAPを「すべての専門分野に共通する学術的スキル」、ESAPを「ある特定の専門分野に必要な学術的スキル」と別個に設定するJordan(1997)の考え方を批判している(同書: 16-17)。そしてDudley-Evans and St John(1998)のように、GeneralとSpecificの関係を連続体として捉えようと

している。しかし、果たしてこの「EGAPとESAPとを連続体として捉える」という構図で、Gの部分の曖昧さが解消されることになるのだろうか。加えて、「すべての専門分野に共通する学術的技能」とは一体何を指すのが明瞭なのだろうか。「ある特定の専門分野」と言っても、専門分野間にも特徴的な差異を示す部分があっても、逆に重複する部分が存在することは容易に想像できるし、そもそも、特定の専門分野という単位を簡単に設定できるとは考えにくい。反対に、構文や表現や語彙の用法と言った言語的な諸特徴を綿密に記述することによって初めて、直観的にしか捉えられてこなかった「専門分野」という使用域別の分類が可能となるのであり、この点がEAPも含めたESPに対するコーパス研究の大きな課題となっていることを看過することはできない。例えば、伝統的な「生物学」と「医学」の中にも容易な分類を許さない複雑な下位分野が存在し、それら個々の「専門分野に必要な学術的技能」とは、むしろ高度に精密な定義を伴った専門用語群の習得なのではないかという考え方である。

従って、連続体という概念を持ち出しても、EGAPとESAPの区別は依然として不明瞭なままであり、「特定の専門分野」という概念が学術的英語使用の区分に有効でない以上、「すべての専門分野に共通する」Generalという言葉技能がどのようなもので、その習得を促すための教材等を蒐集すべきソースをどこに求めればいいのかという疑問は未解決のまま残ることになる。

そこで、EGAPをBiber (2006a) などの言う「大学での英語」 university Englishのうちで、大学生が入学前後に、それを通して情報を収集しなければならない大学からの各種の文書や、履修講義やコースの選択、履修の登録に必要な大学ならではの事務的文書や、専門課程での勉学を開始する前に背景としての知識を求めなければならない入門書文献を読むために必要な英語と、それらに関連する口頭説明や解説を理解するために必要な英語と定義する。本稿ではその中で、議論の対象を前者に限定し、「大学で学術目標を達成するために必要な、大学生活を設計するために必要な英語」読解力を養成するための指針を模索しようとするものである。<sup>注4)</sup>

### 3. コーパス分析

従来のEAP研究においては、その分析対象が学術的な書き言葉、とりわけ研究論文に限定されることが多かった。その結果、大学生が遭遇する多様な英語の諸特徴に関する詳細な研究はこれまでなされてきていない。Biber (2006a: 23-31) では、大学という使用域の中で、語彙使用のパターンを考察しているものであるが、そこでは使用域は授業でのティーチング (classroom teaching) と教科書 (textbook) という話し言葉と書き言葉の2極に大別されている。同書の後続章では、次に示すように、話し言葉の使用域を6個に、書き言葉のそれを10個に下位区分しているものの、それらの使用域を横断して調査されるのは文法的な多様性 (grammatical variation) であり、語彙使用の傾向ではない。

表1. Biber (2006a) によるEAP使用域の分類  
(Biber 2006a: 48などから作成)

Spoken Registers	classroom teaching
	class management
	labs
	office hours
	study groups
	service encounters
Written Registers	textbooks
	course packs
	syllabi, etc.
	institutional writing

前節で言及したような、大学生生活を設計し、学術的な目的を完遂するために必要な英文読解に求められる能力は、総合的に培われるべきであるのは当然のことであるが、本節では、それを構文や連語等ではなく最小の単位と考えられる語彙に分析の射程を限定し、いわゆる「一般使用目的のための英単語リスト」(General Service List, 以下GSL) と「学術目的のための英単語リスト」(以下AWL) の用いられ方の傾向を、TOEFL (Test of English as a Foreign Language) のReadingセクション中、および古典的な英文学作品中で対照的に分析することにより明らかにする。そしてこれらの語彙リストと、JACET語彙研究グループが2003年に公開した語彙リストや京都大学が2009年に公開した『京大学術語彙データベース 基本英単語1110』

(以下「京大語彙リスト」)との相関関係を分析する。

### 3-1. GSL, AWLの「語彙レベル」

#### 3-1-1 General Service List

GSLはWest (1953)において提案された、英語学習者にとって大いに役立つと考えられる一般使用 (general service) のための2000語を選出したリストである。このリストは頻度やその他の点において厳密性は欠くものの、長年にわたって英語教材作成にあたっての重要な指針であり続けた。特に2000個の項目を成すワードファミリー (word family) の定義が厳密ではないことに加えて、頻度情報の処理に一貫性がなく、word familyの見出し語的なheadwordの頻度が必ずしもその下位メンバーである派生形や屈折形の個々の出現回数を積算したものなのかには疑問の余地がある。従って、常にGSLには何語が含まれているのかという基本的な疑問が解消されないままであり続けてきたことも事実である。

各種の不整合を除去したGSLの改良版は1995年にJohn BaumanとBrent Culliganによって公開された (<http://jbauman.com/gsl.html>)。詳細は当該Webページに譲るが、そのリストでは、Brown Corpusに出現頻度を求め、Bauer and Nation (1993) のheadwordとその派生形 (derived form) に関わる基準を精密に適応し合計で2284個のheadwordを掲げている。<sup>注5)</sup>

#### 3-1-2 Academic Word List

Coxhead (2000) によって公開された570種のword familyから成るAWLは学術的なテキストに一般的に出現する単語のリストであり、350万語前後の専門コーパスの分析を通して作成されたものである。4つの主要分野 (芸術・経済・法律・科学) を7つに下位区分した合計28種類のサブコーパスから、出現の範囲と頻度を考慮して抽出されたこの570種のword familyは学術的文書のおおよそ10%前後の単語をカバーすると言われている (Coxhead, 2000: 213)。

本論での議論にとって大切なのは、このAWLがGSLとは重複しない語彙群から成っているという点である。CoxheadはGSLの源をWest (1953) に求めているが、前述のように我々はBauman and Culligan

(1995) 版のSGLを採用するので、後述のようにそれらをword familyとして同列に扱う場合、重複の有無についての慎重な検査が要求される。

AWLは一定以上の頻度で幅広い学問分野の文献に出現する語彙をリストするというのがその理念であり、例えば小説作品というような他の分野ではそれほど頻出しにくいことが予測される。つまり、AWLは大学での学問を進めるにあたって学習者に求められる基礎的で重要な語彙群と規定することもできる。なぜならば、AWLには特定の学問分野で用いられる専門用語は含まれておらず、大学生としての英語学習者が読解と文書作成のために広く身に付けていなければならない性格の語彙が選出されているからである。これが、全学教育における英語教育に求められる語彙を模索しようとする本論でのような議論がAWLにその立脚点を置く所以でもある。

ここで、word familyの概念を示すとすれば、それは「親単語」 (parent word) のもとに「家族構成員単語」 (family members) がまとめられた語群ということができる。つまり元となる単語 (stem) と、その屈折形 (inflection) と派生形 (derivative) とがword familyを構成することになる (Schmitt and McCarthy, 1997: 331)。そしてこのようなword familyについての知識は脳内辞書の項目間の連結の仕方と関連があるとも考えられ、学習者にとって有効な語彙獲得の手段として大きな注目を集め続けている。<sup>注6)</sup>

#### 3-1-3 JACET 82500

BNC (British National Corpus) および独自のサブコーパスを基にして作成され2003年に公開された「大学英語教育学会基本語リスト」(以下JACET 8000)は、1000語ずつのレベルに分けられた語彙リストであり、屈折形や派生形や短縮形は全て基本形に収斂されている。また、国名や月名や曜日名や数詞などは別立ての250語のリスト (JACET8000 plus250) に収録されている。本研究ではこの合計8250語のリストを総称してJACET8250と呼ぶ。<sup>注7)</sup> およそ、レベル1から3ないし4が大学受験までに必要な語彙レベルとされ、本学のような難関大学受験生はレベル5、つまりTOEICでは500点以上までの語彙を習得してきていると想定

してもよいと考えられる。これを裏付けるデータとして、東北大学学務審議会外国語委員会英語教科部会編『平成21年度全学教育科目「実践英語II」授業実践経過報告書』（p.12）では、東北大学文系学部生の7年間に亘るTOEICの平均点数が559.3点であるという報告がなされている。<sup>注8)</sup>

### 3-1-4 GSL, AWLのレベル検証

ここでは、JACET8250のレベル分けを利用して、GSLとAWLが果たしてどのような水準の単語をリストしているのかを検証してみよう。

図1. は2284個のGSLの中から数詞などを除いた2277語をJACET 8000 Level Marker (<http://www.tcp-ip.or.jp/~shim/J8LevelMarker/j8lm.cgi>) を用いてレベル分けした結果である。(横軸の9は8000語レベル外を示している。)

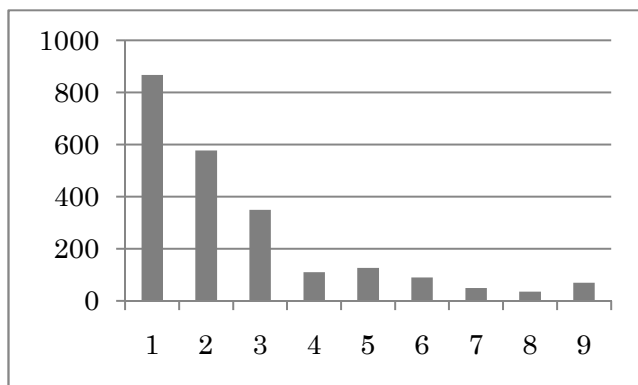


図1. GSLのJACET8000レベル分け

また、図2. はCoxhead (2000) で示された570のheadwordを構成する2872のfamily memberを同様にレベル分けした結果である。

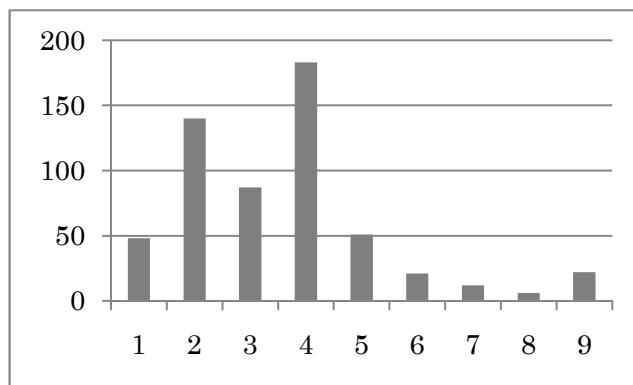


図2. AWLのJACET8000レベル分け

この2つのグラフから、(1) GSL中の単語の大半(78.8%)が高等学校終了まで、つまり大学入試センター試験受験レベルに相当するレベル3に含まれているのに対して、(2) AWLでのレベル3までは合計でも48.24%にしかならず、反対にそれより上のレベル4, 5, 6に分類される単語が44.73%にも達するということが読み取れる。

ここで深刻な問題が発生する。前述のようにAWLは570個headwordから成るリストであり、そのword family全てを展開すると合計で2872語に達する語彙群である。これらの語を単純にJACET Level Markerにかけても、図3. のような結果しかえられない。

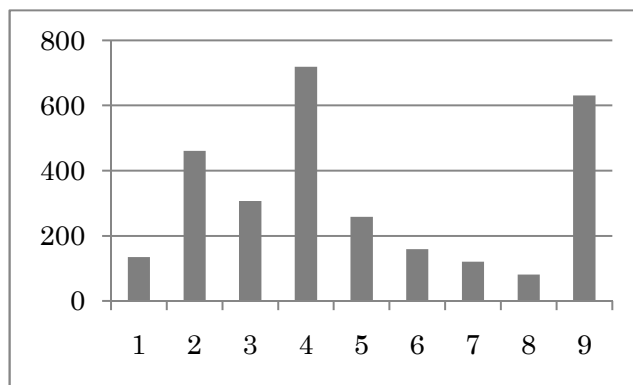


図3. AWLのfamily memberのレベル分け

ここからも、レベル4以上の語彙がAWLに多く含まれることは読み取れるが、問題はレベル判定を受けられない語群が21.97%にもなるという点である。これはheadwordに収斂されている各family memberに対するJACET 8000 Level Markerの処理に起因している。例えば、*coherent*はレベル5に分類されるのに対して、その派生形である*coherently*はレベル外と分類されてしまう。同様に動詞*derive*はレベル4になるが*derivation*も*derivations*もレベル判定を受けない。

このように、headwordとそのfamily memberに対する扱い方に多様性があるために、その問題を解決すべく次節以降では、一律のword family化をそれぞれのリストに施した上で、それらに属する語群のカバー率を検証することとする。そのデータ整備作業には、WebページCompleat Lexical Tutor v.6.2 (<http://www.lextutor.ca/>) から提供されているfamilizing機能 (Horst, et al. 2005) を用いる。

### 3-2. AWL, GSL, JACET8250のword family

Compleat Lexical Tutor v6.2のFamilizerは、BNC中の頻出上位20000語を元に、入力語をword familyリストに展開（または、複数の入力語を単一のheadwordに収斂）する機能を持っている。(1)は、*appropriate, injury, region*を与え、それぞれのword familyを表示した結果であり、反対に(2)は、family memberの*appropriately, uninjured, regional*を入力し、そのheadwordのみを求めた結果である。

- (1) appropriate  
 appropriacy  
 appropriately  
 appropriateness  
 inappropriacy  
 inappropriate  
 inappropriately
- injure  
 injured  
 injures  
 injuries  
 injuring  
 injury  
 uninjured  
 injurious
- region  
 regional  
 regionalist  
 regionalists  
 regionalism  
 regionally  
 regions

- (2) appropriate  
 injure  
 region

前述のようにGSLやAWL、ならびにJACET8250

と後出の「京大語彙リスト」には、項目としてのheadwordとfamily memberに関しての不均衡が存在するために、ここでは一律にFamilizerを利用したデータ統一を図る。

その際に、特にGSLとAWL間で慎重な処理が求められる。このFamilizerでword familyに展開されたAWLに属する総単語は3705個、GSLのそれは11544個となるが、family化のプロセスにおいて15個の重複が発生する。AWLは「GSLに含まれない学術的な語彙」とされているわけだから、この重複語を減じた3690語をAWLの展開リストとして採用する。なお、展開プロセス中の不具合を除去した後に得られるJACET8250のfamilyへの展開語総数は27168となる。

また、1110項目から成る「京大語彙リスト」のfamily member総数は5374であり、重複等を削除整備した後の結果としては、477項目の「文系・理系共通語彙リスト」が2757語に、311項目の「文系語彙リスト」が1430語に、322項目から成る「理系語彙リスト」が1227語にそれぞれ展開される。

表2. 「京大語彙リスト」の展開内訳

総項目数:	1110	文系・理系共通:	477項目
family member 総数:	5374語		2757語
		文系:	311項目 1430語
		理系:	322項目 1227語

### 3-3. 語彙カバー率

ここでは、実際のコーパス分析の例としてBiber et als. (2004) で公開されているコーパス (T2K-SWAL Corpus) 中の書き言葉領域部分と、Project Gutenberg (<http://www.gutenberg.org>) から無償公開されている電子データ中からEmily Brontëの作品*Wuthering Heights*を入手し、慎重にコーパス化したものを取り上げて、GSLとAWLのfamily wordによるテキストのカバー率を調査する。

ESLないしEFLの英語運用能力を測定するためのテストとしてのTOEFLのリスニング、リーディング部門で用いられるテキストが、実際に学術的な使用域（とりわけ大学での言語使用域）で用いられる英語の

性格を正確に反映しているのかを明らかにするために立ち上げられたTOEFL 2000 Spoken and Written Academic Language Projectで開発されたT2K-SWAL Corpusは述べ語総数2737200語から成るサイズで、その内訳は表3に示すとおりである。留意すべきはこのコーパスでは大学における学部・大学院の各レベルも一括して扱っているという点である (Biber et als., 2004: 6)。

T2K-SWAL Corpusの単語リスト中の「書き言葉寄り」(written-skewed)の単語で100万語あたりの出現回数が20回以上のものの合計1423項目(品詞の異なる同形単語は1語としてカウント)をfamilizeして、固有名詞やfamilizeのプロセス中での重複等を削除して得られる単語(種類)は6864個である。

表3. T2K-SWAL Corpusの構成

使用域	テキスト数	語数
<b>Spoken:</b>		
Class sessions	176	1248800
Classroom management	40	39300
Labs/in-class groups	17	88200
Office hours	11	50400
Study groups	25	141100
Service encounters	22	97700
<b>Total speech:</b>	<b>291</b>	<b>1665500</b>
<b>Written:</b>		
Textbooks	87	76600
Course packs	27	107200
Course management	21	52400
Institutional writing	37	151500
<b>Total writing:</b>	<b>172</b>	<b>1071700</b>
<b>TOTAL CORPUS:</b>	<b>423</b>	<b>2737200</b>

11544語のGSLのfamily memberのうち3721個がT2K-SWAL Corpusの6826語をカバーし、3690語に展開されたAWLは2097語をカバーする。また、GSLでもAWLでもカバーできなかった1008語のうちの911語はJACET8250のfamily memberでカバーでき、最終的に全くカバーできなかったものは97語にしか過ぎない。

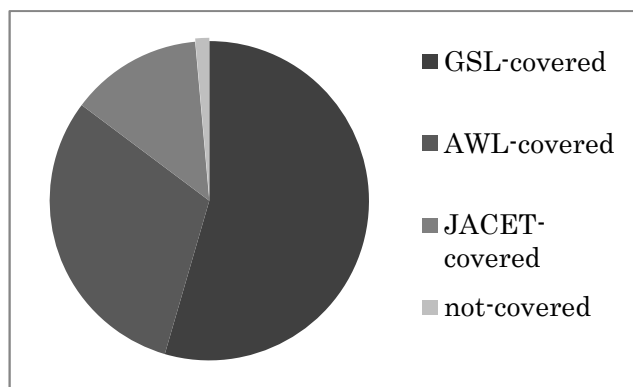


図4. T2K-SWAL Corpusのカバー比率

電子版の*Wuthering Heights*は述べ語数119329語、総異なり語数9640語であり、出現回数が1回のものを減じた単語種類総数は5007語となる。図5は上と同様にGSLとAWLのカバー率を調査した結果である。

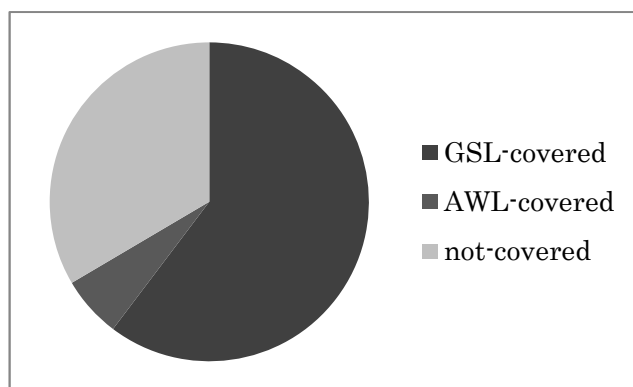


図5. *Wuthering Heights*のカバー比率

学術的な書き言葉コーパスと、英文学作品に対するそれぞれの語彙リストによるカバー率の差は統計的にも明らかな有意差を示す。特にAWLのカバー率は、T2K-SWAL Corpusでの30.56%に対して、*Wuthering Heights*でのそれはわずか6.23% (312語)にしか満たない。<sup>注9)</sup> また、注目すべきは我々が採用しているGSLは1953年版ではなく1995年版であるということである。すなわち、West (1953) が提唱したGSLは1930年代から40年代にかけての半世紀以上前の単語を含んでいるのに対して1995年版では新しい単語が含まれるようになっている。新しいGSLを19世紀英文学作品で検証しても、学術英語コーパスでのそれよりも高いカバー率 (60.23% : 3019語)を示すという結果は、GSLの特徴を明示するものであるし、旧版のGSLに



はheadwordの定義や処理の方法に問題が多く、それが新版では改良されているということの証拠ともいえる。

それでは、GSLとAWLの語彙リストでカバーできなかった単語群に注目してみよう。T2K-SWAL Corpusでの合計1008語のheadwordは合計259個で、これをJACET 8000 Level Markerで査定すると、以下のような結果が得られる。縦軸は語数を、横軸はJACETのレベル（9はレベル外）をそれぞれ示している。

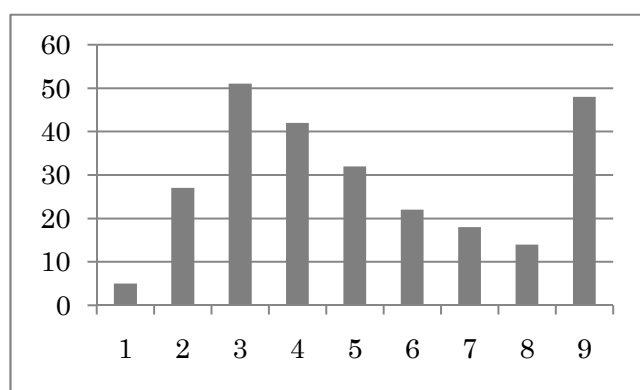


図6. T2K-SWAL CorpusのGSL, AWLでカバーできなかった単語のJACETレベル

また同様に *Wuthering Heights* でのカバー不能単語1276個のheadwordは1223個であり、そのレベルは図7のようになる。<sup>注10)</sup>

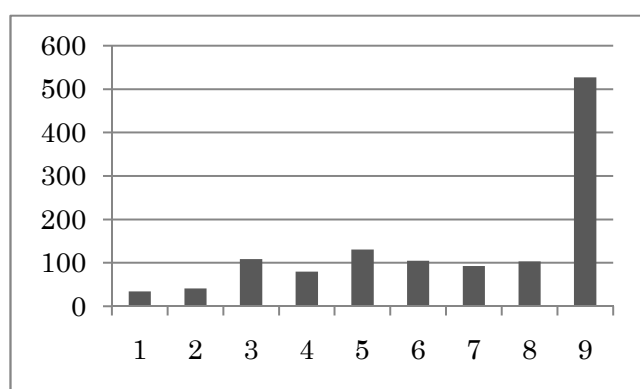


図7. *Wuthering Heights*のGSL, AWLでカバーできなかった単語のJACETレベル

特に注目されるのは、GSLでもAWLでもカバーできなかったT2K-SWAL Corpusの語彙の中で、レベル5以下のものが61.72%を占めるという点である。

対照的に *Wuthering Heights* でのそれでは、レベル外(9)が単独で圧倒的な割合(43.09%(527個))を占め、レベル5以下はわずかに32.29%である。

つまり、英文学作品にはJACETのレベルでは測定できない(3)で一例を示すようなheadwordが豊富であるのに対して、T2K-SWAL Corpusに対しては、(4)のような東北大学生なら知っていると思定できるレベル4ないし5以下の単語の多くがGSLやAWLのリストではカバーできないということである。

- (3)
- |            |             |
|------------|-------------|
| abide      | anon        |
| abject     | antipathy   |
| abode      | apathy      |
| abominable | apparition  |
| abstain    | appellation |
| abyss      | ardent      |
| acquiesce  | ascend      |
| affirm     | ascribe     |
| afore      | asunder     |
| afresh     | atrocious   |
| aggravate  | attire      |
| aghost     | aught       |
| ail        | averse      |
| ajar       | awhile      |
| amaze      | ballad      |
| amiable    | banister    |
| amiss      | beck        |
| anew       | benefit     |
| animate    | beguile     |
| annihilate | behold      |
| ...        | ...         |

- (4)
- |              |            |
|--------------|------------|
| absorb       | diet       |
| acid         | differ     |
| ancestor     | disabled   |
| apartment    | economy    |
| architecture | electronic |
| atom         | emergency  |
| award        | emotion    |
| barrier      | engage     |
| billion      | era        |
| brand        | evidence   |
| burden       | executive  |
| campaign     | facility   |
| career       | faculty    |
| congress     | fossil     |
| consequence  | fulfil     |
| constitution | gear       |
| continent    | gene       |
| convey       | geography  |
| crisis       | graduate   |
| criticism    | household  |
| ...          | ...        |

反対に、T2K-SWAL Corpus においてGSLとAWLでカバーできず、JACETのレベル外のheadwordは(5)に示す48語である。

- |     |               |               |
|-----|---------------|---------------|
| (5) | allele        | geologic      |
|     | amplitude     | handbook      |
|     | angular       | hypermedia    |
|     | artefact      | internship    |
|     | assimilate    | isotope       |
|     | assistantship | kibbutz       |
|     | basidiocarp   | kinetic       |
|     | benzene       | lifeworld     |
|     | canscapable   | magma         |
|     | carbonate     | magnet        |
|     | condense      | mister        |
|     | convection    | multicultural |
|     | correlate     | propositional |
|     | cortex        | proton        |
|     | coursework    | radial        |
|     | deconstruct   | requisite     |
|     | demography    | retain        |
|     | depose        | rotate        |
|     | dissertation  | schizophrenic |
|     | dynamisme     | sculpt        |
|     | evaporate     | seawater      |
|     | external      | silica        |
|     | fallacy       | specie        |
|     | forme         | transact      |
|     | ...           | ...           |

(5) にはコーパス中でのtypoが疑われるものが若干混入しているが、その他のheadwordは「京大語彙リスト」などで十分にカバーできる。<sup>注11)</sup>ところが、(3)のような文学作品中のJACETレベル外のheadwordに対して「京大語彙リスト」のカバー率は図8でのように驚くべき低さを示す。commonは1110語から成る「京大語彙リスト」中の「文系・理系共通」(477語)を示し、humは「文系」(311語)、sciは「理系」(322語)を表している。カバーできないheadwordは95.01%に当たる1162個であり、「文系・理系共通」と「文系」がそれぞれ2.04% (25個)、2.13% (26個)しかカバーせず、「京大語彙リスト」の「文系」で必要と指定している語彙であっても、実際の英文学作品中で、GSLでもAWLでもカバーできなかった単語のほんのわずかな割合にしか有効ではないことを示している。

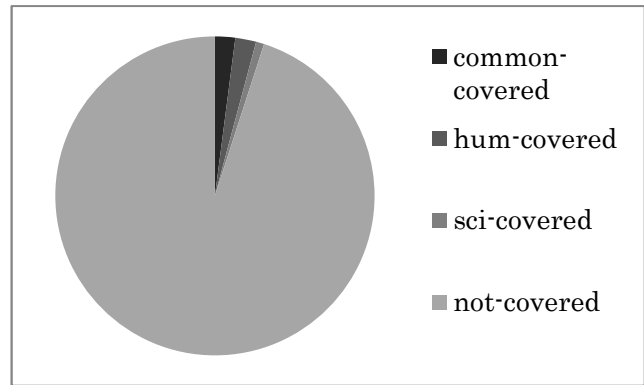


図8. *Wuthering Heights*中のGSL, AWLでカバーできなかったheadwordの「京大語彙リスト」カバー率

GSLやAWL等の英語圏の研究者たちが公表している語彙リストでも、JACET8250や「京大語彙リスト」のような日本国内の学会や大学が日本人学習者を想定して作成した語彙リストでもカバーできない図8中の最大領域こそを英文学作品に特有の語彙、と考えることができるかもしれないし、それは本論で言うところの全学教育における英語の守備範囲を逸脱している語彙の集団かもしれない。

しかし、個々の英語教員が独自の工夫を凝らして教材を選定し全学教育での英語の授業を展開している現状を鑑みるときに、この種の逸脱語彙に対する扱いこそが大きな意味を持つのではないだろうか。

#### 4. 教材作成への提案

ここでは、これまでの議論を踏まえ、東北大学全学教育の英語担当教員が共有できる資産としての英語教育用データベースおよび教材に向けての可能性について簡単に触れる。

##### 4-1. 全学教育用教材としての難易度

授業で採用すべき教材の難易度に関してはさまざまな要因が複雑に交錯しており、簡単にその基準を設定することは困難である。しかし、これまで述べてきたような各種の語彙リストを元にした語彙のレベルや、平成15年度から7年間に亘る文系4学部生全員が受験したTOEIC、平成21年度から開始したTOEFL-ITPの成績を、関連報告書を参照しながら学部毎に丁寧に分析していくことによって、多様な学問背景を持った

個々の英語教員が採用しようとする（特にリーディング訓練用の）教材に対する柔軟な注釈付けが実現できるはずである。全体的に高度ではない英文教材を取って選択せずとも、内容的に魅力があると教員が判断した教材に対して、本学（担当学部）の学生のレベルに沿った語彙注釈なり、一歩進んでの構文・表現に関わる注釈を与えることが保証されれば、選択される教材の幅には何の制約もなく、なおかつ注釈のレベルを変更できるようなデータベースシステムが準備されれば、担当する学部の異なる英語教員間での教材資源のシェアリングも可能となるはずである。

#### 4-2. 京大語彙リストのレベル検証

「京大語彙リスト」を word family に展開し、そのレベルを JACET 8000 Level Marker で分析した結果が図9（文系・理系共通: 2758個）、図10（文系: 1431個）、図11（理系: 1227個）である。

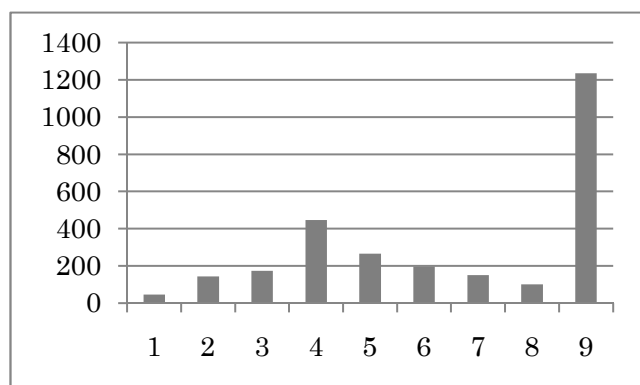


図9. 「京大語彙リスト」文系・理系共通レベル

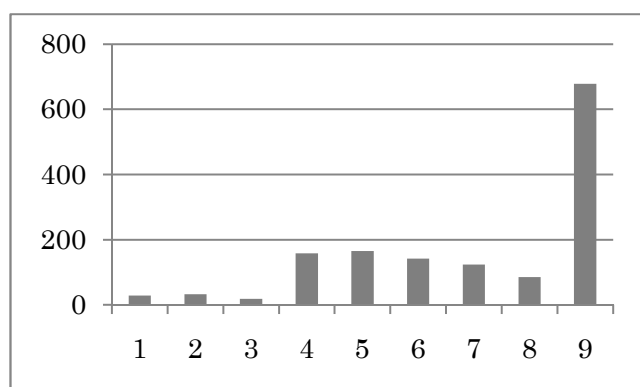


図10. 「京大語彙リスト」文系のJACETレベル

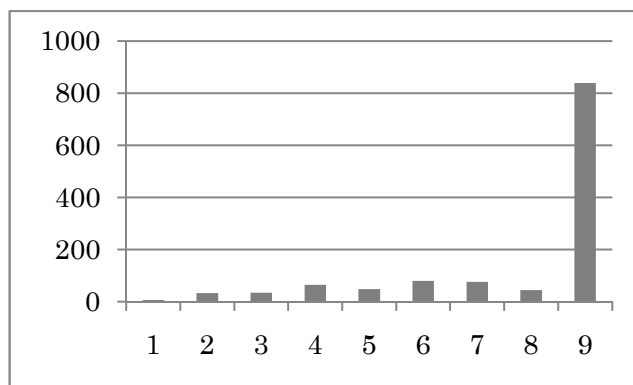


図11. 「京大語彙リスト」理系のJACETレベル

この結果から、「京大語彙リスト」全体のレベルは相当に高く、文系でかろうじてTOEICでは600点前後のレベル6までの語彙が38.2%（544個）を占めるが、理系などではレベル外の語彙が68.38%（839個）に達している。これらのレベル外単語は、専門のジャーナル等から収集された専門用語と考えることができる。しかし、「京大語彙リスト」では、それぞれの専門分野の京都大学教員が学部生・院生に必読と思われる英語学術雑誌（約170種）を選定してもらい（同書: 7）、得られたデータベースを元に1110語のリストを選出したに過ぎず、実際に英語の授業を担当する教員が「この語彙は必須」と指定したものではないという点は重要である。

#### 4-3. 教員による語彙注釈: コーパスシステムにおけるユーザーとしての教員

ここで、本論でこれまでに触れてこなかった、各々の語彙の特定コーパス中における出現頻度に関して簡単に述べる。対象文書中の単語の出現回数をどのように考慮に入れるかという問題に関して興味深い結果が得られる例を、*Stanford Introductory Seminar 2009-2010*を対象にしたのGSLとAWLのカバー率調査で示してみよう。総数121ページから成る学部新入生向けのセミナー（その理念や構想は本学の「基礎ゼミ」と共通している）のコースカタログの述べ語数は65576語で異なり語数は10599である。この単語リスト中の各単語の出現回数に着目し、コーパス研究で一般的に行われる「100万語あたりの200回以上の出現回数」という換算を当てはめると、実際の出現回数が14回以

上の単語536種類が選ばれる。この536個に対するGSLとAWL中の単語のカバー率を求めたのが図12のグラフである。

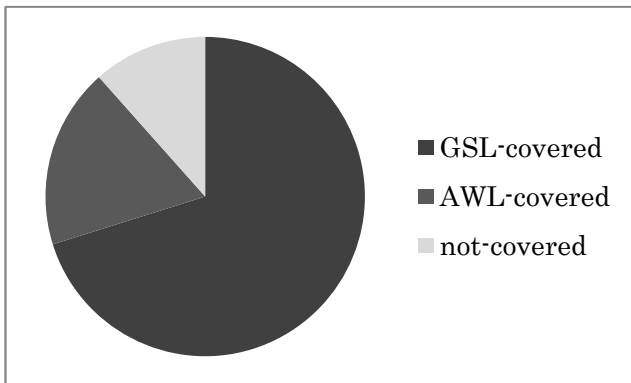


図12. 14回以上出現の単語に対するカバー率

SGLは70%以上をカバーし、AWLでもCoxhead(2000)の言う10%を大きく上回る18.28%のカバー率が得られる。

大きな問題は、word familyに展開されている570個のAWLのheadwordには、各々のfamily memberの出現回数が適切に反映されていないという点である。family memberの総数をカウントすることはあっても、個々の屈折形や派生形のコーパス中での出現頻度は考慮されておらず、それは究極的には「どのような単語を知っていれば、どの程度読めるのか」という問いに答えきれないことになる。例えばこの対象文書中での内容語としての最頻出語である*research* (276回)を知っているからよく読める、とは容易に判断できないし、*students* (239)、*course* (223)、*professor* (210)を知らない本学学生は存在しないわけであるから、頻出語は、例えば当該文書が何についてかかれたものかというような内容推測作業等には有益であっても、EGAP教育における読解力養成のための指針とはなりにくいのではないだろうか。反対に出現回数の少ない*psychiatry* (14)や、閾値以下でカバー率調査の対象にすらならなかった*ubiquitous* (1)や*stakeholder* (1)などを知らないために、重要な文書情報を受容できないという事態も想定できるのである。そのような低頻度語は前後の文脈情報を頼りに解釈可能、とする考え方もあり得るが、どの程度までの未知語の割合が、特定使用域の文書読解困難さに影響を与えるのかといっ

た重要な問題は今後の大きな研究課題でもある。

特定専門分野に固有の英語語彙は、全学教育の直接的な射程範囲には含まれないかもしれないが、第2節で述べたように、EGAPとEAPとの境界について、それを学部の学年進行のみならず、英文の専門性という観点からも捉え直すことが現実的ではないと我々が考える以上、多様な専攻分野の背景を持った英語教員が「学部1,2年次にこの程度までの語彙は習得してほしい」と考える範囲が重要である。そのためには、ただ漠然と語彙範囲を定めるのではなく、GSLやAWL等の語彙リストではカバーできない語彙の中でも、なおかつ東北大学の全学教育の英語担当教員が重要な学習項目と位置付ける語彙群を、読解指導の教材の中での注釈や、独自の解説的語彙リストとして教材化することができれば、それがまさに具体的な形で、本研究が目指すような「個々の教員の専門性を活かしながらの全学英語教育」にとっての語彙選択に繋がるのではないだろうか。

筆者が研究を推進中のRDBMS (Relational Data Base Management System)を中心に据えた統合的なコーパス処理システム (Okada and Sakamoto, 2010)では、システムユーザー層のひとつとして英語教員を想定している。既存の大規模コーパスを利用することも可能であるし、教員独自が収集した英文データをコーパス化して、利用ニーズに呼応した属性を付与するための柔軟な機能を持つ下位システムを開発中である。3階層モデル (three-tier model) を採用した当該システムでは、利用者が単にシステムのサービス享受するだけでなく、他の利用者と協働してコーパスの内容とシステムの機能とを拡張させていくというのが大きな特徴のひとつである。このような性格を持つコーパスシステムに、教材作成機能の一環としての語彙レベル判定や註釈付け等の機能を持たせることによって、多くの英語教員が共通の資源を共有し、東北大学全学教育の英語教育が目指すべき方向性そのものも模索しながらの、活発で背極的な協働体制が確立できると期待される。

## 5. 結論

本論の議論の目的は、限られた数の共通英語読解用教材を作成するためでもないし、また個別の「東北大語彙リスト」を作成するためのものもなかった。むしろ緩やかな教材候補群の集積としてのデータを、多くの英語教員が協働して持ち寄り、それをデータベース化した上で、語彙のレベル解析や、専門分野の背景を反映させての教員個人による柔軟な註釈付与や実際の教材出力に至るまでの包括的なシステムの開発可能性を示唆し、「東北大学の英語教員集団の考える」英語教育に相応しい教材の構想を模索することが、現在の文脈の中で最も肝要なのではないだろうか。

## 注

1. 特殊コーパスの研究から得られた当該使用域に関するさまざまな知見が、一般コーパス (general corpus) として代表性を保持しなければならない均衡コーパスの構築に対して大きな貢献をなす、という相乗効果にも意義を見出す必要がある。
2. *English for Specific Purposes*  
([www.elsevier.com/locate/esp](http://www.elsevier.com/locate/esp))  
*The Journal of English for Academic Purposes*  
([www.elsevier.com/locate/jeap](http://www.elsevier.com/locate/jeap))
3. 参考文献目録を参照
4. 米国のトップ10大学がWeb等に公開している外国人留学生向けの文書をコーパス化し、それに対する分析を現在行っている。
5. Brown Corpusでは同一の品詞標識を付与される意味的に無関係の同形異義語 (homograph) は1語としてカウントされる。この新しいGSLにおいては、このような同一品詞を持つ同形異義語に加えて、Brown Corpus中で異なった品詞標識を持つ同形で意味的に関連のある語同士も併せて1語として頻度を算出している。
6. 学習者が、屈折形はまだしも、品詞の異なる語にもなり得る派生形を容易に習得するわけではないという指摘もなされている (Schmitt and Zimmerman, 2002)。そこでは、単に word family に属する member の単語に触れさせるだけでは学習者の語彙力の増強は望めず、特に派生形についての指導と学習者の能

動的な産出訓練の必要性が強調されている。

7. JACET8250にはBNCからの語彙データが豊富に反映されているために、同コーパスの約10%を占める話し言葉データが含まれていることになる。この書き言葉対話し言葉の問題は本稿での議論の射程を大きく超えるものであるために、稿を改める必要がある。
8. しかし、深刻な問題は、東北大学の1, 2, 年次学生の語彙力をJACET8000におけるレベル5と安易に想定してよいのかという点である。ひとつ下のレベル4には、大学入試センター試験、実用英語検定試験なら2級程度の語彙力であるレベル3よりも上の3001位から4000位の単語が含まれていることになり、実際にはこのレベルの語彙力の学生も少なからず在籍しているのではないかと思われるからである。
9. T2K-SWAL Corpusに対するAWLのfamily memberのカバー率は顕著に高く、Coxhead (2000) の想定する学術的文書からの語彙群よりも、T2K-SWAL CorpusではEAPの性格が際立っているのかもしれない。
10. ここで詳述の余裕はないが、*Wuthering Heights* のカバー不能単語のheadwordへの収斂結果語数がT2K-SWAL Corpusでのそれ25.86%に比べて72.97%と極めて多いのには、文学作品ならではの語彙 (headword 種類) の多様性が現れているのかもしれない。
11. *kibbutz* のように、特定の分野に出現しがちな語が含まれているのはTOEFLのリーディング問題の出展に影響を受けていると考えられる。このような、語彙に対する典拠先によるバイアスの調査もBiber et. als. (2004) の大きな研究目的のひとつでもあった。  
\*本研究は科研費 (基盤 (C) 22500891) の助成を受けて行われている研究の一部を扱ったものである。

## 参考文献

- Bauer, L. and P. Nation. (1993) "Word Families", *International Journal of Lexicography: Vol.6*, pp.1-27.
- Bauman, J. and C. Brent. (1995) "General Service List", <http://jbauman.com/gsl.html> (accessed on Dec 30th, 2010) also available from <http://www.auburn.edu/~nunnath/engl6240/wlistgen.html> (accessed on Dec 30, 2010)

- Biber, D. (2006a) *University Language: A Corpus-based Study of Spoken and Written Registers*, John Benjamins.
- (2006b) "Stance in Spoken and Written University Registers", *Journal of English for Academic Purposes: Vol.5, No. 1*, pp.97-116.
- (2010) "Challenging Stereotypes about Academic Writing: Complexity, Elaboration, Explicitness", *Journal of English for Academic Purposes: Vol. 9, No.1*, pp.2-20.
- , S. M. Conrad., R. Reppen., P. Byrd., M. Helt., V. Clark., V. Cortes., E. Csomay, and A. Urzua. (2004) "Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus", Educational Testing Service.
- Coxhead, A. (2000) "A New Academic Word List", *TESOL Quarterly: Vol.34, No.2*, pp.213-238.
- Dudley-Evans, T, and M. St John. (1998) *Developments in English for Specific Purposes*, Cambridge University Press.
- Horst, M., T. Cobb. and I. Nicolae. (2005) "Expanding Academic Vocabulary with an Interactive On-line Database", *Language Learning & Technology: Vol. 9, No. 2*, pp.90-110.
- Hyland, K. (2006) *English for Academic Purposes: An Advanced Resource Book*, Routledge.
- and T. Polly. (2007) "Is There an 'Academic Vocabulary'?", *TESOL Quarterly: Vol. 41, No. 2*, pp. 235-253.
- Jordan, R. (1997) *English for Academic Purposes: A Guide and Resource Book for Teachers*, Cambridge University Press.
- (2002) "The Growth of EAP in Britain", *Journal of English for Academic Purposes: Vol. 1, No.1*, pp.69-78.
- Nation, P. (2001) *Learning Vocabulary in Another Language*, Cambridge University Press.
- Okada, T. and Y. Sakamoto. (2010) "A New RDBMS and Flexible POS Tagging for EFL Learners and Researchers: Designing a Corpus Analysis System Based on the Three-tier Model", *CAHE Journal of Higher Education Tohoku University: Vol.5*, pp.43-52.
- Schmitt, N. and M. McCarthy. (1997) *Vocabulary: Description, Acquisition, and Pedagogy*, Cambridge University Press.
- Schmitt, N. and C. B. Zimmerman. (2002) "Derivative Word Forms: What Do Learners Know?", *TESOL Quarterly: Vol. 36, No. 2*, pp.145-171.
- Vongpumivitch, V., J. Huang. and Y. Chang. (2009) "Frequency Analysis of the Words in the Academic Word List (AWL) and non-AWL Content Words in Applied Linguistics Research Papers", *English for Specific Purposes: Vol. 28, No. 1*, pp.33-41.
- West, M. (1953) *A General Service List of English Words*, Longman.
- 
- 北海道大学言語文化部英語教育系カリキュラム委員会 (1997) 「北海道大学英語語彙表」(<http://icarus.imc.hokudai.ac.jp/jugyo/huvl/>) (accessed Jan 06, 2011)
- Introductory Seminars: Course Catalog 2009-2010*, Stanford Introductory Studies (Office of the Vice Provost for Undergraduate Education, Stanford University, <http://introsems.stanford.edu>). (accessed Jan 05, 2011)
- 京都大学英語学術語彙研究グループ, 研究社 (2009) 『京大 学術語彙データベース: 基本英単語1110』 研究社
- 田地野彰 (2004) 「日本における大学英語教育の目的と目標について: ESP研究からの示唆」, 『京都大学大学院 人間・環境学研究科マルチメディア教育運営委員会 MM News』 7号, pp.11-21.
- 竹蓋幸生, 水光雅則 (2005) 『これからの大学英語教育』 岩波書店
- 東北大学学務審議会外国語委員会英語教科部会(編) (2010) 『平成21年度全学教育科目「実践英語II」授業実践経過報告書』
- 東京大学教養学部英語部会 (編) (2009) 『東大英単』 東京大学出版会