Coh-Metrix によるテキスト理解に必要な語彙熟達度の数値化

―語彙知識の広さ・深さ・アクセス速度を中心に―

茨城県／筑波大学大学院在籍・日本学術振興会特別研究員　濱田 彰

概要　本研究は，テキストに含まれる単語の特性を数値化する Coh-Metrix を用いて，英検テキストの読解に必要な語彙熟達度を予測した。2つの調査を行い，語彙知識の広さ，深さ，およびアクセス速度を要求する語彙特性（e.g., 頻度・多義性・親密度）とテキスト理解度とのかかわりを検証した。

調査1では，英検1級から3級までのテキストを対象とし，各語彙特性が受験級によってどのように異なるのかを分析した。その結果，上位の受験級になるほど(a)低頻度語，(b) 下位概念に位置する名詞と動詞，(c) 意味を想起しにくい単語の割合が増加することがわかった。

学習者のテキスト理解度とテキストに含まれる単語の特性とのかかわりを検証した調査2では，単語の頻度・多様性・心像性が英検テキストの理解度に影響を与えることが示された。さらに，これらの指標を組み合わせた回帰モデルを利用することで，一定のテキスト理解度に到達するのに求められる語彙熟達度を予測できることが明らかとなった。

1 はじめに

英語で書かれた文章を理解するためには，何よりもまず英単語の知識が必要となる。語彙熟達度とテキスト理解度の相関関係を検証した研究では，語彙力のある読み手ほど，より良くテキストの内容を理解できることが示されている（e.g., Laufer, 1989, 1992; Qian, 2002; Schmitt, Jiang, & Grabe, 2011; Zhang, 2012）。しかしながら，学習者が英語を読むために学ぶべき単語は数多くあり，どんなテキストでも理解できるだけの語彙力を身につけることは非常に困難である。ゆえに，どれだけの語彙力があれば，どのような英文が読めるようになるのかという疑問が生まれるだろう。英検の場合，各受験級で出題されるテキストを理解するのに必要な語彙力を明示することは，学習者にとって単語を学習するための1つの指標になると思われる。

一般的なテキストに含まれる英単語は，その76%が頻度にして2,000語(注1)レベルに含まれており，86％が3,000語レベルになる（Nation, 2001）。したがって，英文を読むためには最高頻度語（e.g., the, a, I, have）から優先的に覚えるべきだと考えられる（Grabe, 2009）。しかしながら，単語の知識には，1つの単語についてどこまで深く知っているのかという側面も含まれる。また，覚えた英単語を実際の言語使用場面において流暢に使用できることも，言語パフォーマンスを左右することになるだろう。すなわち，テキストを理解するのに求められる語彙熟達度は，さまざまな観点から評価される必要がある。

このような背景から，近年では Coh-Metrix を応用して語彙熟達度を評価し（Crossley, Salsbury, McNamara, & Jarvis, 2011a, 2011b; Crossley, Salsbury, & McNamara, 2012），テキスト理解度とのかかわりを明らかにしようとする研究が注目されている（Crossley, Greenfield, & McNamara, 2008）。Coh-Metrix はコーパスに基づき，テキストに含まれている単語の特性（e.g., 頻度・多義性・親密度）を数値化することができる（e.g., Graesser, McNamara, Louwerse, & Cai, 2004）。例えば，高頻度語だけで書かれたテキストであっても，そのテキ

ストには多義語が多く含まれているといった分析が可能となる。このテキストを読ませる場合，学習者には多義語の知識がさらに求められることになるだろう。このように，Coh-Metrix を用いてテキストの特性を語彙の観点から数値化することで，学習者がめざすべき語彙熟達度を詳細に示すことができると考えられる。

本研究は Coh-Metrix によるテキスト分析を通して，英検各級で使用されている長文テキストを理解するのに求められる語彙知識の構成概念を予測し，既存の語彙テストとの対応関係を明らかにする。本稿では，まず語彙知識の構成概念について概観し，語彙知識と英文読解のかかわりを検証した先行研究をまとめる。続いて Coh-Metrix が範囲とするテキスト分析の方法について説明し，その発展的研究について触れる。以上の先行研究を基盤とし，本稿では 2 つの調査結果を報告する。調査 1 では英検テキストに含まれている単語の特性が各受験級でどのように異なるのかを検証し，調査 2 で使用するテキストを選定する。調査 2 では，英語学習者を対象とした実験を行い，Coh-Metrix により算出されたテキストの語彙特性とテキスト理解度，および語彙熟達度を測定するテストとのかかわりについて検証する。

2 先行研究

2.1 語彙熟達度の構成概念

英語学習者にとって，単語の知識は，英語運用能力を支える基礎の 1 つとなる。普段の授業で英語のテキストに触れる学習者は，単語の知識がどれくらいあれば英語で書かれた文章が読めるのかという疑問を持つだろう。ここでの「どれくらい」という言葉は，学習者がどれだけ多くの単語を知っているのかを指す。しかしながら，単語に含まれる情報はその意味概念だけでなく，1 つの単語について「どのような」側面の知識を持っていることが英文読解に求められるのかということも議論されている。このように，単語を知っているとは何を意味するのか，すなわち，語彙知識とはどのような概念で構成されているのかという問題に対し，これまで多くの研究者がさまざまな理論研究を行ってきた。

多くの研究者は，語彙知識をさまざまな観点に分割しており，それらは大きく (a) 広さ，(b) 深さ，および (c) 流暢さから成る（Daller, Milton, & Treffers-Daller, 2007）。本節ではこのような語彙知識の構成概念を総括し，それぞれの知識を測定する手法について述べる。

2.1.1 語彙知識の広さ

英文読解における語彙知識の広さ，いわゆる語彙サイズとは，「ある文字列を見てそれが英単語であるとわかり，その意味を思い出せる単語をどれだけ多く知っているか」と定義される（e.g., Grabe, 2009; Nation, 2001）。どれくらい多くの単語を知っていれば一般的なテキストが読めるようになるのかを検証した研究によると，文章に含まれる95%の単語を知っていることがテキスト理解には必要であり，そのためには3,000語 (注1) から5,000語の知識が求められる（Laufer, 1989, 1992）。また，テキストの理解度テストや単語の数え方を厳密に行った Hu and Nation（2000）では，テキストに含まれる98%から99%の単語を知っていることが必須であり，そのために必要となる語彙サイズは8,000語 (注1) から9,000語に上ると述べられている。

テキストに含まれている95％から99％の単語を知っていなければ，英語で書かれた文章を全く理解できないのかという点については議論の余地が残るものの（e.g., Nation, 2001; Schmitt et al., 2011），少なくとも語彙サイズが大きくなるほど，テキストの理解度は向上すると考えられている（e.g., Crossley et al., 2012）。ゆえに，学習者の語彙サイズを知ることは，その学習者の読解力を知ることにつながるため（e.g., 杉森, 2011），これまでさまざまな種類の語彙サイズテストが開発されてきた。日本人英語学習者を対象とした語彙サイズテストには「望月語彙サイズテスト」がある（望月, 1998）。このテストは7,000語 (注2) レベルまでの語彙サイズを測定でき，与えられた日本語に相当する英単語を選択肢の中から選ばせる出題形式となっている（図 1 参照）。問題として与えられる日本語は 2 問 1 組となっており，この 1 組について 6 個の英単語が選択肢として与えられる。図 1 の例で言うと，学習者は問 1 に対して(4) opinion を，問 2 に対して(3) noise を選択することが求められる。そして，推定語彙サイズは「正答率×実施したテストのレベル」で算出される。

1. 意見，考え			2. 音，物音，騒音		
(1) bank	(2) memory	(3) noise	(4) opinion	(5) skin	(6) wing

▶ 図 1 ：望月語彙サイズテスト第三版・2,000 語レベルの問題例（相澤・望月 , 2010）

sudden

beautiful	quick	surprising	thirsty	change	doctor	noise	school

▶ 図 2 ：Word Associates Format（Read, 1998）

2.1.2 語彙知識の深さ

杉森（2011）で指摘されているとおり，一般的な日本人英語学習者がよく用いる語彙学習方略は，英単語集を利用して語形（つづり）と意味の対応関係を暗記するというものである。日本のように外国語として英語を学ぶ環境では英語のインプット量が限られており，偶発的に語彙を学習する機会が少ないため，意図的に語彙サイズを増やす学習方略は必要不可欠である。しかしながら，学習者の語彙熟達度は，単語の語形と意味をどれだけ多く知っているかということのみでは説明されない。1 つの単語に含まれる情報は語形とその意味だけでなく，その語が持つ多義性・コロケーション・イディオムに関する知識や，語連想・反意語・同義語に関する知識も含まれる（e.g., Crossley et al., 2012; Grabe, 2009; Nation, 2001; Qian,1999, 2002; Read, 1998）。すなわち，1 つの単語についてどれだけ「深く」知っているかということも語彙熟達度を構成する概念となる。

語彙知識の深さは心内辞書の構造を反映していると考えられている（e.g., Daller et al., 2007; Nation, 2001; Read, 1998; Wolter, 2001）。単語の知識はそれぞれが意味的に関連づけられて記憶されており，語彙知識の深さが増すほど，単語同士の結びつきは母語話者に近い形で心内に記憶される（Wolter, 2001）。例えば，ある単語（e.g., dog）には上位概念（e.g., animal）と下位概念（e.g., pug）があり，それぞれが密接に結びついているほど，その学習者の語彙知識は深いということになる（e.g., Crossley et al., 2012）。同様に，多くの単語には多義性（e.g., lamb: animal vs. meat）があり，学習者は文脈の中でそのような語を解釈することを苦手とする（Elston-Guttler & Friederici, 2005）。文脈に応じて多義語などを適切に処理するためには，1 つの単語に対して，「子羊」と「羊の肉」といった複数の意味を関連づけて覚える必要がある（Verspoor & Lowie, 2003）。そして，複数の意味をネットワークのように関連づけて覚えるほど，学習者の語彙知識は精緻かつ頑健になる。

単語の意味的関連性に基づく語彙知識の深さを測定するテストとしては，Word Associates Format（WAF; Read, 1998）が用いられている。WAF は，語連想やコロケーションの知識を測定するテストとなっている。語連想の知識については，テスト項目となった単語の同義語や上位語，同位語，下位語を知っていなければ正解できない形式になっている。同様に，WAF ではコロケーションの知識についても問われ，単語が文脈の中でどのような形で使われていたかを学習することで培われた知識が試される（杉森, 2011）。出題形式は図 2 のとおりであり，刺激語（sudden）に対し左右のボックスから 4 語ずつ選択肢が与えられる。左側のボックスには刺激語と意味的に関係のある単語が入っており，右側のボックスには刺激語とコロケーションの関係にある単語が含まれている。そして，学習者はボックスにある 8 語から適切な 4 語を選択するよう指示される。

図 2 の例であれば，左側のボックスから quick と surprising を，右側のボックスからは change と noise を選択できると，適切に選択できた数に応じて 1 点が与えられる。

2.1.3 語彙知識の流暢さ

ここまで，単語はさまざまな知識（e.g., 語形・意味・語連想・コロケーション）から成ることを述べた。このような知識をコミュニケーションの中で効果的に利用するためには，語彙知識が保存されている心内辞書へのアクセス速度や自動性が重要になる（e.g., Crossley et al., 2011a, 2011b; Daller et al., 2007; Grabe, 2009; Perfetti, 2007）。すなわち，学習者は自身が持っている語彙知識を目的に応じて流暢に使用する必要があり，リーディングにおいては視認した単語の意味を素早く想起できることが求められる（i.e., 視認語彙；望月・相澤・投野, 2003）。

視認語彙の重要性は，Perfetti（2007）の Lexical Quality Hypothesis において特に強調されている。この仮説では，素早くかつ認知的な負荷を伴わずに語彙知識を検索できることが，読み手の読解力を強く予測するとしている。さらに，流暢な語彙処理を実現するためには，書記素・音素・意味・文法の知識が心内で強固に結合していることが求められるとしている。このような質の高い語彙知識を持っているほど，読み手はテキストに書かれた文字を認識すると同時にその意味を思い出すことができる。しかしながら語彙知識の質が低い読み手の場合，文字を認識してもその意味へのアクセスが起こりにくく，意味を思い出すのに余分な認知資源が必要になる。ゆえに語彙処理に続く読解プロセス（e.g., 文の構造を解析したり読み取った命題を理解表象へと統合すること）が妨げられ，十分なテキスト理解に至らないことが示唆されている。

このように，語彙知識の流暢さはリーディングにおいて重要な役割を担うものの，語彙知識へのアクセス速度を測定する妥当なテストは今のところ広く利用可能な状態にはなっていない（Iso, Aizawa, & Tagashira, 2012）。しかしながら，読解テストに時間制限を設けることで，間接的にではあるものの言語処理の流暢さを観察できることが報告されている（Zhang, 2012）。したがって本研究では，英文読解に必要とされる語彙知識の広さと深さに加え，流暢さ（i.e., 単語の意味へのアクセス速度）も扱うことにした。

2.2　英文読解と語彙知識

リーディングにはさまざまな認知プロセスがかかわる。テキストに明示的に書かれた情報をもとに英文を理解するプロセスとして，(a) 文字や単語を特定すること，(b) 特定した単語が持つ意味を心内辞書から検索すること，(c) 単語レベル・文レベルの統語情報を読み取ること，(d) 単語と統語情報から構築される命題を理解することなどが含まれる（Grabe, 2009）。これらの認知プロセスは，それぞれに対応した読み手の知識に支えられている。具体的には，文字や単語を見てその意味を想起するためには語彙知識が必要であり，文の構造を解析するためには文法知識が求められる。また，文章全体の理解を形成するためには背景知識も重要な役割を果たす。

読解プロセスにかかわる知識のうち，テキストの理解度に最も貢献するものとして，多くの研究者は語彙知識を挙げている。例えば Zhang（2012）は学習者の語彙知識の広さ・深さ，および文法知識を測定し，これらの知識が読解力とどのようにかかわるかを検証している。構造方程式モデリングによる分析の結果，読解力は語彙知識と文法知識で81％説明され，特に語彙知識が大きな役割を果たすことを明らかにした。同様に，Qian（2002）は語彙知識の広さと深さで読解力の67％を説明できるとしている。

実際に，語彙知識と英文読解との間にはどのような関係があるのだろうか。2.1節で述べたとおり，語彙知識は広さ，深さ，および流暢さという概念で構成されている。これらの知識とテキスト理解の関係性について検証した先行研究では，まず，語彙知識の広さとテキスト理解度は正の相関関係にあることが明らかにされている（e.g., Grabe, 2009）。すなわち，語彙サイズが大きくなるほどテキストの理解度は向上するという関係にある。

読解における語彙知識の深さの役割を検証した Qian（1999）は，学習者の語彙サイズが3,000語[注1]を超えると，今度は語彙サイズではなく，語彙知識の深さがテキスト理解の向上に必要になることを明らかにした。同様の結果は Qian（2002）でも得られており，語彙知識の深さは語彙サイズが十分にある学習者の読解力を構成する要素の 1 つになることが示されている。

一方，英語学習者の語彙知識の流暢さと読解力との関係を検証した研究は少なく，語彙知識へのアクセス速度が，読解力とどれくらい深くかかわるのかは今後の検証が待たれる。しかしながら，流暢な語彙処理がその他の読解プロセスを促進することを考慮すると（Perfetti, 2007），単語を見てその意味を素早く思い出せる学習者ほどテキストを効率よく理解できると考えられる。

2.3　Coh-Metrix

それでは，語彙知識の広さ・深さ・流暢さのそれぞれが必要になる英語の文章とはどのようなものだろうか。まず語彙知識の広さについて，語彙サイズが大きくなるほどテキストの理解度が向上するということは，使用頻度の低い単語を幅広く知っていることがテキストのより良い理解に求められることを意味する。なぜなら，低頻度語は学習者にとって未知語である可能性が高く，その割合が一定量を超え

るとテキスト理解は著しく阻害されるためである（e.g., Laufer, 1989, 1992）。

語彙知識の深さは WAF によって測定され，その得点が高い学習者ほどテキスト内容をよく理解できることが明らかにされている（e.g., Qian, 1999, 2002; Zhang, 2012）。WAF は語連想やコロケーションなど，単語同士の意味的関連性の知識を測定していることから（Read, 1998），読解テキストにそのような知識が求められる単語が多く含まれているほど，当該テキストを理解するのに語彙知識の深さが求められると考えられる。

語彙知識の流暢さ，すなわち単語の意味へのアクセス速度は，単語の親密度，具象性，および心像性の影響を受ける（e.g., Crossley et al., 2011a, 2011b）。単語親密度とは，ある単語がどの程度なじみあると感じられるかを反映した指標である。親密度の高い単語（e.g., dog, take, good）ほど容易に理解されることが明らかにされており（Grabe, 2009），逆に親密度の低い単語が多く含まれているテキストは理解しづらいと考えられる。同様に，抽象語は具象語よりも理解されにくいことから，テキストに含まれる単語の具象性もテキスト理解に影響を与えると予想される。また，単語が持つ心的イメージがどれくらい容易に喚起されやすいかを示す心像性（Ellis & Beaton, 1993）の高さも，単語の意味へのアクセス速度とかかわる。これら 3 つの指標は単語の頻度と正の相関関係にあるが（Schmitt & Meara, 1997），冠詞や代名詞といった高頻度語（e.g., a, the, this）よりも，内容語の方が親密度，具象性および心像性すべてにおいて高いことから（Crossley et al., 2011a），3 つの指標は単語のアクセス速度に独自の影響を与え，テキストの理解度を左右する要因になると思われる。

以上をまとめると，語彙知識の広さ・深さ・流暢さが求められるテキストには，(a) さまざまな低頻度語，(b) 多義語など語連想の知識が反映される単語，(c) 親密度・具象性・心像性の低い単語が多く含まれていることになる。このような言語特性をコーパスに基づき分析できるツールが，Memphis 大学の研究グループが開発した Coh-Metrix である。Coh-Metrix は，現在バージョン3.0が公開されており，テキストに含まれるさまざまな言語特性から，そのテキストの読み易さを評価するために開発されたウェブ上で利用できるツールである。特に，テキスト内容の意味的一貫性を，単語レベル・談話レベル・概念レベルで分析することができる（Graesser et al., 2004）。

単語レベルに限ってみると，まず Coh-Metrix 3.0 は，語彙知識の広さを反映する単語の頻度情報を，COBUILD コーパスから作成された CELEX Database（Baayen, Piepenbrock, & Gulikers, 1995）に基づいて数値化できる。また，語彙の多様性を示す指標として，TTR，VOCD，および MTLD を算出することができる。最も単純な指標である TTR は，テキストの総語数における異語数の割合を意味する。一般的にはこの値が高いほど，そのテキストにはさまざまな種類の単語が使用されており，相応の語彙サイズが読解に求められることになる。しかしながら，TTR の値はテキストの総語数が増えるほど低くなる傾向にあるため，総語数の異なるテキストを比較することはできない（杉森, 2011）。この問題を解決する指標として，現在最も妥当だとされているものが MTLD である（McCarthy & Jarvis, 2010）。

次に，語彙知識の深さを反映する語彙特性として，Coh-Metrix 3.0はテキストに含まれる単語の多義性の平均値と，それぞれの単語がどれだけ多くの上位語を持つかの平均値を算出できる。多義性の値は，テキストに含まれている内容語が持つ意味の平均値となっている。この値は WordNet（Miller, Beckwith, Fellbaum, Gross, & Miller, 1990）をもとに算出され，数値が高いほどテキストに多義性を持つ語が含まれていると解釈される[注3]（Graesser et al., 2004）。上位語数の算出方法は，同じく WordNet に基づき，名詞または動詞が持つ上位語の数を平均している。したがってこの値が低いということは，分析されたテキストには，あるカテゴリーの上位概念を表す単語（e.g., dog ⊂ animal）があまり使われていないことを意味する。一方，この値が高くなるほど，特定の対象を表す単語（e.g., pug）が使用されていることになる。

最後に，語彙知識へのアクセスしやすさを反映する語彙特性として，Coh-Metrix 3.0では MRC Psycholinguistic Database（Coltheart, 1981）に基づき，単語の親密度・具象性・心像性の平均値が算出される。まず親密度の値は，テキストに含まれている内容語の平均親密度となる。MRC にある親密度データは，ある単語を 1（一度も見たことがない）から 7（毎日見る）という 7 段階で評価したものに

基づく。したがって，この数値が高いほどテキストには親密度の高い語が含まれており，逆に数値が低くなるほど全体的に親密度の低い語が使用されていることになる。具象性の値も親密度と同様に計算され，数値が高いほどテキストの具象性が高く，数値が低くなるにしたがってテキスト内容の抽象度が増すことになる。心像性も同様の手法で数値化され，この値が高いほど心的イメージが浮かびやすい単語が多く含まれており，数値が低くなると心的イメージが浮かびにくい語がテキストに多く含まれているということになる。

これまで，Coh-Metrix を言語学習や言語テストに応用する試みが多くなされてきた。言語学習については，テキストの理解しやすさを Coh-Metrix により評価する研究が行われている。例えば，Crossley et al.（2008）は Coh-Metrix で検証したさまざまな言語特性（e.g., 単語の頻度・統語的複雑さ・文章の結束性）のうち，単語の頻度が最もテキストの理解しやすさに影響を与えることを示している。また Crossley, Louwerse, McCarthy, and McNamara（2007）では，学習者向けに簡易化されたテキストでは，一般向けのテキストよりも高頻度語や親密度の高い単語の使用割合が高いことを明らかにした。これらの結果は，テキストに含まれる単語の特性がテキスト理解度に大きな影響を与えることを示唆している。

言語テスティングの分野では，学習者の語彙熟達度を，Coh-Metrix によって数値化しようとする試みが行われている（Crossley et al., 2011a, 2011b）。一連の研究では，学習者のライティングやスピーキング・パフォーマンスを語彙知識の広さ・深さ・流暢さの観点から分析している。その結果，書き言葉の場合には単語の頻度・多様性・上位語数が語彙熟達度を予測し，話し言葉の場合には頻度・多様性・親密度・上位語数が語彙熟達度を予測することが明らかとなった。また，Crossley et al.（2012）では，TOEFL などで測定される英語熟達度と語彙熟達度の関係を検証しており，英語の熟達度が高くなるほど，学習者は頻度が低く多様な語彙を使用できるようになり，心像性と親密度の低い単語の知識を有することを明らかにした。すなわち，語彙熟達度の高い学習者は，頻度の低い単語や，下位概念を指す単語を幅広く知っており，かつ自身の語彙知識を流暢に運用できることが示されている。

2.4　本研究の枠組み

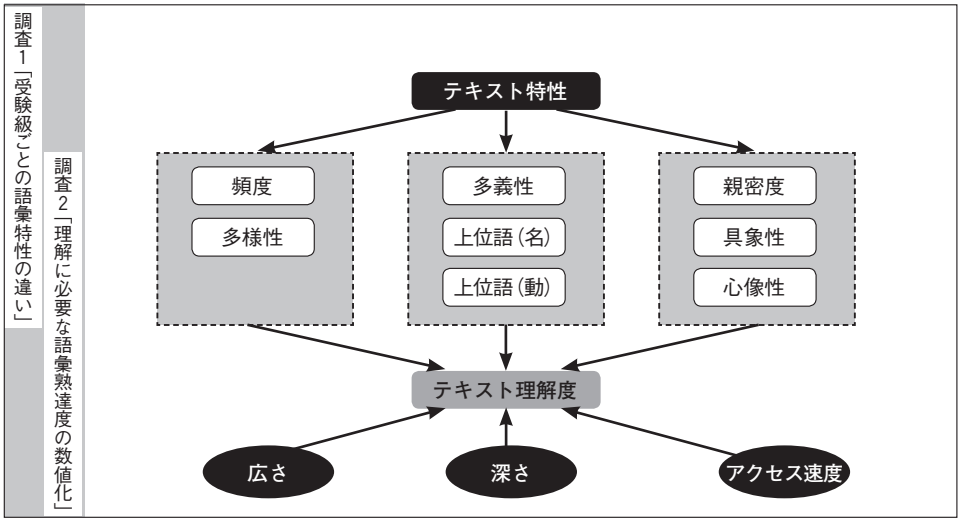
英文読解に必要な語彙知識とは何かを検証した先行研究では，多くの単語を深く知っており，かつ言語使用の場面で語彙知識を流暢に使いこなせることの重要性を指摘している（e.g., Crossley et al., 2007; Daller et al., 2007; Grabe, 2009; Nation, 2001; Perfetti, 2007; Qian, 1999, 2002）。語彙知識を構成する概念は単語の特性と密接にかかわっており，テキストに含まれている単語の特徴を分析できる Coh-Metrix は，それぞれのテキストを理解するのに求められる語彙熟達度の違いを可視化できると考えられる。したがって本研究では，英検テキストに含まれる単語の特徴が受験級ごとにどのように異なり，単語の特徴はテキストの理解度にどれだけの影響を与えるのかを検証する。また，語彙知識の広さと深さを測定するテストを用いて，英検テキストの理解に必要な語彙熟達度を数値化する（図 3 参照）。

3　調査 1

3.1　目的

調査 1 では，英検各級の長文読解問題のテキスト特性を，語彙知識の広さ・深さ・アクセス速度の観点から数値化し，テキスト理解に求められる語彙知識が各受験級でどのように異なるのかを明らかにする。そのため，1 級から 3 級までのテキストを扱い，受験級をテキスト難易度の指標とする。上位の受験級になるにつれて，テキスト理解に求められる語彙知識がどのように変化するのかを検証することを中心に，本調査では次に挙げるリサーチクエスチョン（RQ）に取り組む。

RQ1-1：語彙知識の広さにかかわる語彙特性（頻度・多様性）は受験級によってどのように異なるか。
RQ1-2：語彙知識の深さにかかわる語彙特性（多義性・上位語数）は受験級によってどのように異なるか。
RQ1-3：語彙知識へのアクセス速度にかかわる語彙特性（親密度・具象性・心像性）は受験級によってどのように異なるか。



▶ 図 3：調査1と調査2で検証する要因の関係図

3.2 方法

3.2.1 英検テキストの収集

Coh-Metrix 3.0によるテキスト分析を行うため，英検1級から3級までの内容一致選択問題で使用された長文読解テキストを収集した。分析対象としたテキストは，1998年第1回から2011年第3回までのもので，テキストのジャンルは物語文，説明文，および評論文であった。1級で使用されている超長文テキスト（600語以上）は，同じ1級で使用されている他のテキストと比べて総語数が明らかに違うため分析から除外された。また，電子メールや掲示といったジャンルのテキストは除外された。収集されたテキストの総語数，平均語数，およびリーダビリティを表1に示す。

3.2.2 分析対象とする語彙指標の選定

本調査では，語彙知識の広さ・深さ・アクセス速度の指標となる語彙特性を，先行研究（Crossley et al., 2011a, 2011b, 2012）に基づきそれぞれ選定した。

(1) 語彙知識の広さを反映する指標として，語彙の多様性を採用した。受験級ごとに分析対象となるテキストの総語数が大きく異なることを考慮し，本研究では MTLD の値を計算した（2.3節参照）。この値が高くなるほど，テキストに含まれている単語の種類は多くなる。学習者が多様な語彙を知っているのであれば，その学習者の語彙サイズは大きく，テキストに含まれる単語の多くを理解できると言える。しかし，さまざまな単語を知っているといっても簡単な単語のみしか使えないのであれば，語彙熟達度を正確に測定できるとは言えない（Daller et al., 2007）。

したがって，2つ目の指標として単語の頻度を加えた。Coh-Metrix 3.0ではテキストに含まれる内容語の平均頻度を数値化できるため，低頻度語がどの程度テキストに含まれているのかを調べることができる。頻度の値が低くなるほど，そのテキストに含まれる単語の平均頻度は低いと解釈される。

■ 表 1：調査1で分析したテキストの特徴

			平均語数		FKGL		FRE	
難易度	*n*	総語数	*M*	*SD*	*M*	*SD*	*M*	*SD*
1級	92	45,883	498.73	50.69	12.68	1.70	42.13	8.50
準1級	96	40,066	417.35	72.05	11.87	1.33	45.95	7.07
2級	74	26,264	354.92	22.11	9.32	1.09	59.98	6.19
準2級	54	15,668	290.15	18.28	7.94	1.00	67.25	5.53
3級	40	10,212	255.30	12.96	5.99	0.96	75.15	5.29

（注）*n* = テキストの数。 FKGL = Flesch-Kincaid Grade Level, FRE = Flesch Reading Ease

(2) 語彙知識の深さに関して，1つの単語がどれくらい多くの意味を持つのか（多義性）と，どれくらい多くの上位語を持つのか（名詞および動詞の上位語数）を算出した。

具体的には，多義語が持つ複数の意味に対して，それぞれの難易度や意味的距離が数値化された。上位語数に対しては，上位語ほど簡単に使用でき，下位語ほど難しくなるという意味的関係性および難易度を数値化した（Graesser et al., 2004を参照）。それぞれの値が高くなるほど，そのテキストには多義性のある単語，および上位語を多く持つ単語が含まれていることになる。

(3) 語彙知識へのアクセス速度を反映する指標として，単語の親密度，具象性および心像性を用いた。それぞれの値は100から700までの範囲にあり，700に近いほどテキストに含まれている単語の親密度・具象性・心像性は高いと見なされる。

3.2.3 手順

Coh-Metrix 3.0による語彙特性の数値化は，すべて2.3節で述べた Memphis 大学のウェブ上（http://cohmetrix.memphis.edu/cohmetrixpr/index.html）で行われた。ただし, Coh-Metrix 3.0が参照するコーパスデータである CELEX Database，WordNet，および MRC Psycholinguistic Database に登録されていない単語は自動的に除去されて分析された。

3.2.4 データの分析方法

受験級（難易度）ごとにテキストの語彙特性がどのように異なるのかを検証するため，それぞれの指標を従属変数，難易度を独立変数とした一元配置分散分析を8回繰り返した。調査1では統計的検定を8回繰り返したため，ボンフェローニによる有意水準の調整を行い（$\alpha = .006$），効果量[注4]と併せて結果を解釈することにした（Cohen, 1988）。さらに，一元配置分散分析で有意となった変数の下位検定として，Tukey HSD による多重比較を行った。

3.3 結果と考察

Coh-Metrix 3.0で算出された各語彙特性の記述統計を表2に示す。また，これらの結果を視覚的に表現したものが図4から図6に当たる。それぞれの図は語彙知識の広さ，深さ，およびアクセス速度を反映する各指標に対応している。

まず，図4が示すとおり，難易度の高い級になるほどテキストに含まれる内容語全体の頻度は低下し，使用されている単語の種類は豊富になっていることが読み取れる。したがって，3級から1級にかけてテキストの難易度が高くなるほど，求められる語彙知識の広さはより大きくなると予想できる。

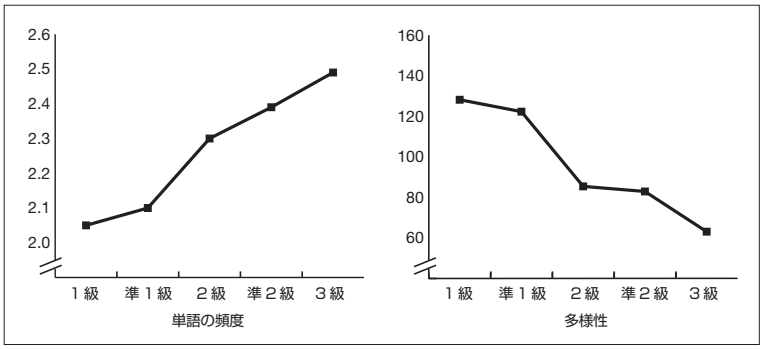
一方，語彙知識の深さについて図5を見ると，当初の予想とはいくぶん異なった傾向が現れているのがわかる。例えば，単語の多義性は難易度の低い級ほど高い傾向にあり，上位語数を多く持つ単語も受験級が高くなるほど直線的に増加するという結果にはなっていないようである。

単語の親密度・具象性・心像性といった語彙知識へのアクセス速度にかかわる語彙特性は，図6が示すとおり, 下位の受験級ほど高くなる傾向にあった。言い換えると，易しいテキストには意味を想起しやすい単語が多く含まれていることになる。

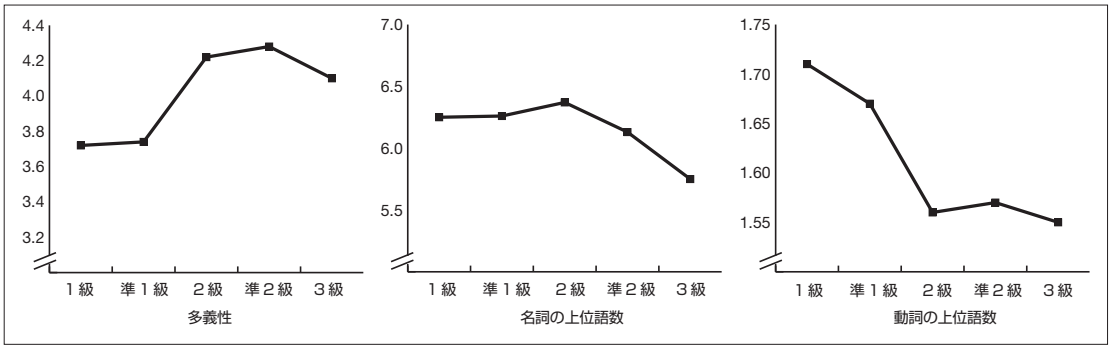
■表 2：各受験級の英検テキストにおける語彙特性の記述統計

	指標	1 級	準 1 級	2 級	準 2 級	3 級
広さ	頻度	2.05 (0.10)	2.10 (0.12)	2.30 (0.13)	2.39 (0.12)	2.49 (0.11)
	多様性	128.20 (24.84)	122.33 (26.40)	85.45 (16.53)	82.94 (12.30)	63.10 (10.67)
深さ	多義性	3.72 (0.32)	3.74 (0.27)	4.22 (0.33)	4.28 (0.42)	4.10 (0.39)
	上位語（名）	6.25 (0.50)	6.26 (0.51)	6.37 (0.59)	6.13 (0.61)	5.75 (0.54)
	上位語（動）	1.71 (0.15)	1.67 (0.12)	1.56 (0.15)	1.57 (0.14)	1.55 (0.18)
アクセス速度	親密度	561.55 (6.79)	565.23 (6.58)	573.19 (7.52)	578.08 (8.77)	586.68 (8.03)
	具象性	379.83 (19.89)	384.65 (18.42)	387.00 (23.71)	395.87 (19.13)	412.02 (17.67)
	心像性	409.50 (16.34)	414.13 (16.39)	414.04 (19.78)	427.40 (16.69)	445.42 (15.63)

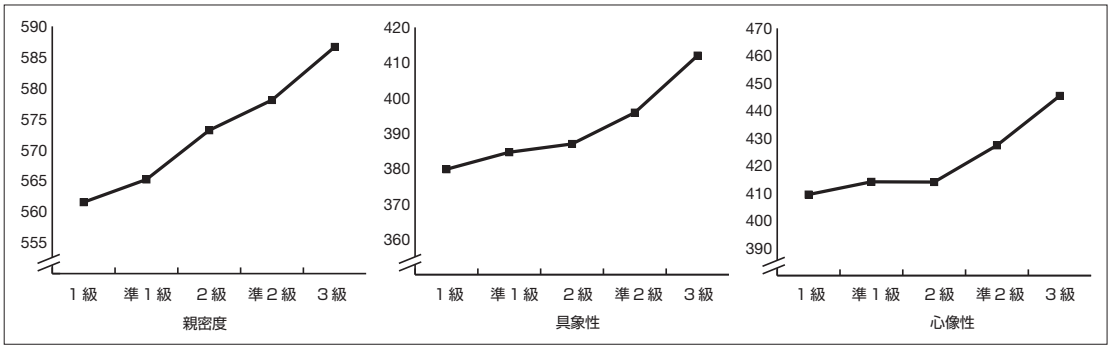
（注）カッコ内の数字は標準偏差を示している。「上位語（名）」は名詞の上位語数を，「上位語（動）」は動詞の上位語数をそれぞれ表している。



▶ 図 4 ：受験級ごとに求められる語彙知識の広さの違い



▶ 図 5 ：受験級ごとに求められる語彙知識の深さの違い



▶ 図 6 ：受験級ごとに求められる語彙知識の流暢さの違い

以上の解釈が統計的に支持されるのかを検証した結果を表 3 に示す。一元配置分散分析の結果，すべての語彙特性において難易度による有意な影響が見られ，受験級の違いによる効果量 (注5) も大きかった。続いて，各語彙特性がテキストの難易度によってどのように異なるのかを調べるために，Tukey HSD による多重比較を行った。その結果を表 4 に示す。

初めに，RQ1-1（英検各級のテキスト理解に求められる語彙知識の広さの違い）への回答として，英検テキストに用いられている単語の頻度や多様性は受験級によって異なることが示された。まず単語の頻度について，具体的には，上位の受験級ほど低頻度語の割合が段階的に増えることがわかった。一方，語彙の多様性については，難易度が高くなるほどさまざまな種類の単語が使用される傾向にあったものの，１級と準１級，および２級と準２級の間で有意な違いは見られなかった。

テキストに含まれる単語の頻度が低くなるほど，学習者は未知語に遭遇する確率が高くなることを考慮すると（e.g., Laufer, 1989, 1992），当初の予測どおり，低頻度語の使用割合が増える上級のテキストほど必要とされる語彙サイズは大きくなると考えら

■ 表 3 ：一元配置分散分析の結果

	指標	$F(4, 351)$	p	η^2
広さ	頻度	173.01**	< .001	.663
	多様性	114.72**	< .001	.567
深さ	多義性	46.83**	< .001	.348
	上位語（名）	9.29**	< .001	.096
	上位語（動）	17.96**	< .001	.170
アクセス速度	親密度	111.27**	< .001	.559
	具象性	20.89**	< .001	.192
	心像性	37.62**	< .001	.300

（注）*p< .05, **p< .01。有意水準はボンフェローニの方法で調整した。

■ 表 4 ：Tukey HSD による多重比較の結果

	指標	多重比較の結果	全体的な傾向
広さ	頻度	[1] < [準 1] < [2] < [準 2] < [3]	上位の受験級ほど使用される単語の頻度は低くなる
	多様性	[1・準 1] > [2・準 2] > [3]	上位の受験級ほど使用される単語の種類は多くなる
深さ	多義性	[1・準 1] < [2・準 2・3]	1 級と準 1 級では多義性のある単語が少なくなる
	上位語（名）	[1・準 1・2・準 2] > [3]	3 級のみ上位語を持つ名詞が少なくなる
	上位語（動）	[1・準 1] > [2・準 2・3]	1 級と準 1 級では上位語を持つ動詞が多くなる
アクセス速度	親密度	[1] < [準 1] < [2] < [準 2] < [3]	上位の受験級ほど親密度の低い単語が多くなる
	具象性	[1・準 1] < [準 2] < [3] | [2] < [3]	上位の受験級ほど具象性の低い単語が多くなる
	心像性	[1・準 1・2] < [準 2・3]	上位の受験級ほど心像性の低い単語が多くなる

れる。さらに表 2 を見ると，頻度の平均値の差は，2 級と準 2 級，および準 2 級と 3 級の間で同程度であるのに対し，準 1 級以降は頻度が急激に低下している。旺文社から刊行されている『英検全問題集』には，各級を受験する際の目安となる語彙サイズが表 5 のように提示されている。3 級から 2 級までは 1,500語ずつ語彙サイズが増加しているのに対し，2 級から 1 級までは約2,500語ずつ増加することからも，2 級までの試験と比べて，準 1 級と 1 級には相当の語彙サイズが求められることがわかる。これは Coh-Metrix による分析結果とも一致しており，3 級から 2 級へと難易度が高くなるにつれて求められる語彙サイズは直線的に増加するのに対し，準 1 級以降は必要な語彙サイズが飛躍的に増すと結論づけられる。また，表 5 が示す語彙サイズの変化とは異なり，準 1 級と 1 級の間で Coh-Metrix が示す頻度の値に大きな違いが見られないのは，調査 1 において英検 1 級の超長文テキストが除外されたことが影響していると思われる。

■ 表 5 ：英検各級を受験する目安となる語彙サイズ

難易度	必要な語彙サイズ	下位の受験級との差
1 級	10,000 語から 15,000 語	2,500 語から 7,500 語
準 1 級	7,500 語	2,400 語
2 級	5,100 語	1,500 語
準 2 級	3,600 語	1,500 語
3 級	2,100 語	

頻度の結果と同様に，英検テキストで使用されている語彙の多様性の指標は，上位の受験級ほどさまざまな単語がテキストに含まれるという傾向を示した。すなわち，1 級になるほど使用される単語の種類が豊富になることを意味する。先行研究で指摘されているとおり（e.g., Crossley et al., 2011a, 2011b），多様な語彙が使用されているほど多くの語彙サイズを必要とすることから，受験級の高さはテキスト理解の難易度を反映していることが確認された。逆に言えば，上級のテキストを理解するためには，より多くの語彙サイズが求められると予測される。

しかしながら，1 級と準 1 級，および 2 級と準 2 級の間には，MTLD の平均値に統計的な有意差が見

られなかった。この結果は，それぞれの受験級で使用される語彙の多様性に違いが見られないことを意味する。上で述べた語彙頻度の結果と合わせて考えると，1級と準1級，および2級と準2級で使用されているテキストの難易度を決定するのは単語の多様性ではなく，テキストに含まれる低頻度語の割合であるということがわかる。さらにMTLDの平均値を見てみると，2級と準1級の間で大きく多様性の値が上昇しており（表2および図4参照），この間でテキストの難易度が格段に上昇することが予測される。すなわち，頻度の結果と同様に，準1級からはテキストを理解するために求められる語彙サイズが非常に大きくなると言える。

次にRQ1-2（英検各級のテキスト理解に求められる語彙知識の深さの違い）について，英検テキストに用いられている単語の多義性や上位語数は受験級ごとに異なることが示された。具体的には，多重比較の結果から，2級から3級までのテキストと比べて，1級および準1級のテキストには多義性のある語が比較的少ないことがわかった。また，英検テキストに使用されている名詞が上位語をどれだけ多く持っているのかについては，3級とそれ以外のテキストとの間で有意差が見られ，3級のテキストには上位語数の少ない名詞が多く含まれているという結果になった。動詞の上位語数については，難しいテキスト（1級・準1級）と相対的に易しいテキスト（2級・準2級・3級）の間で有意差が見られ，1級と準1級では，その他のテキストと比べて上位語数の多い動詞の使用割合が高かった。

多義性を持つ単語の解釈は学習者にとって難しい読解プロセスであるため(Elston-Guttler & Friederici, 2005; Verspoor & Lowie, 2003)，多義性のある単語が多く使用されているほど，テキストの理解は難しくなると想定されたが，1級や準1級のテキストに比べて，2級から3級までの易しいテキストにおいて多義性のある単語の割合が多いという結果になった。これは，高頻度語（e.g., take, book, happy）ほど多義性を持つ単語が多く（Graesser et al., 2004），受験級が下位のテキストには，そのような高頻度語が多く含まれていたことが原因だと考えられる。一方，低頻度語の割合が高い1級や準1級のテキストでは，多義性を持つ語の割合が相対的に低くなったと言える。

ゆえにCoh-Metrix 3.0が示す多義性の数値は，テキストの難易度をうまく説明できないと予測される。例えば，Crossley et al.（2007, 2008）でも単語の多義性とテキストの難易度との関係性は示されていない。多義性を持つ単語の解釈が難しくなるのは，(a) よく知っている単語が別の意味（e.g., 第二義）で使用されており，(b) さらに多義性を解消するための文脈的サポートが低い場合である（e.g., Grabe, 2009）。Coh-Metrix 3.0ではこのような条件を分析することができないため，単語の多義性がテキスト理解の難易度に与える影響を調べる場合には，特定の単語が上記の条件を満たしているかを質的に見ていく必要があるだろう。

続いて，名詞や動詞が持つ上位語数の結果について考察する。Crossley et al.（2007）では，学習者向けに簡易化されたテキストに含まれている名詞や動詞は，上位概念の少ない一般的な単語となっていることが明らかにされている。また，“pug” という単語は「犬」や「動物」であることを含意するため，このような特定的な単語を解釈する場合，その語のカテゴリーや他の語とどのように関連するのかという知識を持っているほどテキスト理解は容易になる。3.2.2節にて言及したとおり，意味的に上位の概念を指す単語ほど簡単に使用され，下位の単語ほど使用するのが難しくなることを踏まえると（Crossley et al., 2011a, 2011b, 2012），下位概念を指す名詞や動詞が多く使用されているテキスト（1級や準1級）では，内容を理解するために深い語彙知識が求められると予測される。

最後にRQ1-3（英検各級のテキスト理解に求められる語彙知識の流暢さの違い）についても受験級によって有意差が見られ，英検テキストに用いられている単語の意味へのアクセスしやすさは受験級ごとに異なることが示された。多重比較の結果を見てみると，2級より簡単なテキストでは心像性の高い単語が多く含まれるようになり，また準2級より簡単なテキストでは単語の具象性も高くなることがわかった。さらに単語の親密度については，受験級の難易度が低くなるほど親密度の高い語が直線的に増えることが示された。

親密度，具象性，および心像性の高い単語ほど，その意味を素早く思い出しやすいことを考慮すると（e.g., Coltheart, 1981; Crossley et al., 2007），少なくとも2級と3級の間には3つの指標すべてで有意差があるため，それぞれのテキストの理解しやすさは

異なると予測される。特に，テキストを理解するための時間が制限され，単語や文法処理を素早く行わなければならない場合に（Zhang, 2012），単語の意味へのアクセス速度に影響を与えるこれらの語彙特性は非常に重要になると思われる。

3.4　調査 1 のまとめ

調査 1 では，英検 1 級から 3 級までのテキストを用いて，語彙知識の広さ・深さ・アクセス速度を反映する語彙特性が，各受験級でどのように異なるのかを検証した。特に，語彙知識の広さについては語彙頻度と多様性，語彙知識の深さについては単語の多義性と上位語数, アクセス速度については親密度, 具象性, および心像性に焦点を当てた比較を行った。

Coh-Metrix 3.0による分析の結果，テキストに使用されている単語の頻度は上位の受験級ほど低くなり，特に 2 級と準 1 級の間に大きな差があることがわかった。また難易度が高くなるほど多様な語彙が使用されている傾向にあったが，1 級と準 1 級，および 2 級と準 2 級の間にはその傾向が見られなかった。語彙知識の深さについては，1 級と準 1 級のテキストにおいて単語の多義性が低くなり，Coh-Metrix 3.0で算出される多義性の値がテキストの理解度を予測するとは限らないことが示唆された。また 3 級のテキストでは下位語として使用される名詞は少なく，それより上位の受験級になると特定の対象を指示する名詞が多くなることがわかった。動詞の性質については，難しいテキスト（1 級から準 1 級）と比較的易しいテキスト（2 級から 3 級）の境目で特定の意味のみを持つ動詞の使用頻度が変わっていた。最後に，語彙知識へのアクセスしやすさに影響を与える語彙特性として，親密度，具象性，および心像性の値は，上位の受験級になるにつれて低くなることがわかった。

以上の結果は，上位の受験級になるほど，英語学習者には (a) 低頻度語を幅広く知っていること（語彙知識の広さ），(b) それぞれの単語が他の単語とどのような意味的関係にあるのかを知っていること（語彙知識の深さ），および (c) 単語を見てその意味を素早く思い出せること（アクセス速度）が求められることを示している。しかしながら，これらの語彙特性のうち，どれが学習者のテキスト理解に大きくかかわるのかは明らかにされていない。したがって，続く調査 2 では，英語学習者に語彙特性の異なるテキストを読解させ，その理解度がテキストの語彙特性とどのようにかかわるのかを検証する。さらに，学習者の語彙サイズおよび語彙知識の深さを測定し，テキストの語彙特性・理解度・語彙熟達度との相関関係を分析することで，特定のテキストを読解するのに必要とされる語彙熟達度を予測する。

4　調査 2

4.1　目的

調査 2 では，Coh-Metrix 3.0で算出された語彙特性の指標が，テキスト理解度をどれくらい予測するのかを検証する。調査 1 では英検テキストで使用されている単語の特性が受験級ごとに大きく異なることが示されたことから，テキストを理解するのに必要となる語彙知識の側面も受験級に応じて異なる可能性がある。この可能性を検証するために，調査 2 では学習者の語彙知識の広さと深さを測定するテストを用いて，各受験級の語彙特性との対応関係を明らかにする。

調査 1 では 1 級から 3 級までのテキストを分析したが，調査 2 では 2 級と 3 級のテキストを扱う。1 級や準 1 級のような難易度の高いテキストを用いた場合，調査協力者に十分な語彙知識がなく，語彙熟達度の個人差にかかわらずテキスト内容を理解できない，いわゆる床面効果が出る恐れがある。また，準 2 級と 3 級の間では，単語の心像性 (注6) に大きな違いが見られなかったため，テキスト理解に求められる語彙知識の側面がうまく反映されないと予想される。語彙熟達度とテキスト理解の関係性を明確に示すため，テキストに含まれる単語の特性が大きく違う 2 級と 3 級のテキストを使用することにした。調査 2 で検証した RQ は以下の 2 つである。

RQ2-1：Coh-Metrix で分析した 2 級と 3 級テキストの語彙特性は，日本人英語学習者のテキスト理解度をどの程度予測するか。

RQ2-2：2 級と 3 級のテキストを読解するのに求められる語彙熟達度は，テキストに含まれる単語の特性とどのようにかかわるか。

4.2 方法

4.2.1 協力者

調査2に参加した協力者は，東京都内の私立大学，または茨城県にある国立大学に通う大学生51名であった。しかし，調査2で実施された複数のテストのうち1つでも受験しなかった協力者10名のデータを除外し，最終的に41名のデータを分析対象とした。協力者は少なくとも6年間英語を学んでおり，英語圏の教育機関に長期留学した経験はなかった。大学での専攻は哲学，経済学，英語教育学などさまざまであった。

4.2.2 マテリアル

4.2.2.1 読解テキスト

読解マテリアルとして用いた受験級は2級と3級であり，1998年第1回から2011年第3回までの過去問から選定された。英検テキストには手紙や電子メール形式の問題も含まれており，テキストジャンルを統一するため，2級のテキストはすべて第4問Bから選び，3級のテキストは第4問Cを対象とした。選定されたテキストの語数，文数，リーダビリティ，および語彙特性の平均値を表6に示す。

また，調査2で用いたテキストが調査1の結果に沿うものであることを確認するため，それぞれの語彙特性に対し*t*検定を行った。その結果，多義性と動詞の上位語数でのみ2級と3級の間で有意差が見られず，調査1と同じ結果になった。ゆえに，読解マテリアルとしてこれら2つの受験級を使用することとした。

4.2.2.2 語彙テスト

英検テキストの理解度と学習者の語彙熟達度との関係性を検証するため，語彙知識の広さと深さをそれぞれ測定できる語彙テストを利用した。語彙サイズの測定には望月語彙サイズテストの第三版（VLT；相澤・望月, 2010）を使用した。調査2の協力者が大学生で，既に一定量の語彙サイズを有していることを考慮し，1,000語レベルのテストは実施せず，2,000語レベルから6,000語レベルまでのテストを行った（最大値は130点）。また，学習者の語彙知識の深さを測定するためにRead（1998）のWAFを用いた（最大値は160点）。

4.2.3 手順

調査協力者は英語の授業中に，または個別にテストに取り組んだ。表7に示すとおり，調査2では3種類のテストが課され，すべて監督者の指示に従って順番どおりに遂行するよう求められた。

初めにVLTの冊子が配布され，問題形式と解答方法が説明された後（2.1.1節参照），協力者は2,000語レベルから6,000語レベルまでの5セクションを，それぞれ3分で解答するよう指示された。次にWAFの冊子が配布され，VLTと同様にテストの説明が行われた後（2.1.2節参照），すべての問題を25

■ 表6：調査2で使用したテキストの特徴

		2級（*n* = 40）	3級（*n* =40）	*t*（78）	*p*	*r*[注7]
語数		351.00（23.82）	255.30（12.96）	22.32**	<.001	.930
文数		19.68（1.87）	19.93（2.27）	0.54	.592	.062
FKGL		9.46（1.18）	5.99（0.96）	11.82**	<.001	.802
FRE		59.23（6.67）	75.15（5.29）	-14.46**	<.001	.854
広さ	頻度	2.31（0.13）	2.49（0.11）	6.64**	<.001	.601
	多様性	85.75（16.60）	63.09（10.67）	-4.46**	<.001	.451
深さ	多義性	4.18（0.25）	4.10（0.39）	-1.04	.301	.117
	上位語（名）	6.33（0.63）	5.75（0.54）	-4.44**	<.001	.450
	上位語（動）	1.55（0.18）	1.58（0.15）	-0.70	.488	.080
アクセス速度	親密度	573.10（7.63）	586.68（8.03）	-7.76**	<.001	.661
	具象性	382.40（23.97）	412.01（17.67）	-6.29**	<.001	.581
	心像性	409.10（19.60）	445.42（15.63）	-9.16**	<.001	.720

（注）*n* = テキストの数。*p < .05, **p < .01。ボンフェローニの修正による有意水準は α = .004。
FKGL = Flesch-Kincaid Grade Level, FRE = Flesch Reading Ease

■ 表 7：調査 2 の手順

テストの種類	目的	変数
(1) 望月語彙サイズテスト (VLT)	語彙サイズを測定	独立変数
(2) Word Associates Format (WAF)	語彙知識の深さを測定	独立変数
(3) 英検テキストの読解・筆記再生テスト	テキストの理解度を測定	従属変数

分で解答するよう指示された。

英検テキストの理解度は筆記再生テストによって測定された。テキストの読解および筆記再生テストのセクションでは，協力者はまず，40種類の冊子のうち 1 つをランダムに受け取った。読解後に内容理解問題を解かなければならないことが予告された後，協力者は 5 分間 (注8) で 1 つ目のテキストを読むよう指示された。以下に，テキスト読解の指示文を示す。

「次のページから英語の長文読解を始めます。読解後は本文に戻らず問題に取り組むので，内容をしっかり理解しながら読んでください。長文の読解時間は 5 分です。辞書を使用することはできません。試験終了後，配布した冊子はすべて回収しますので，必ず名前を書いてから読解を始めてください」

テキストの読解後に筆記再生テストの概要が説明され，協力者は15分で課題に取り組むよう促された。具体的な指示内容は以下のとおりである。

「いま読んでもらった英文の内容について，以下の 3 点に注意しながらその内容を日本語で忠実に再現して下さい。時間は15分です。(前のページに戻ることはできません)」

(1) 英文の内容に関して，覚えていることを残らず吐き出すように書いて下さい。
(2) 繰り返しや余分なことでも構いません。できるだけ多くのことを書いて下さい。
(3) 箇条書きにせず文章の形で書きましょう。

同様の手順で 2 つ目のテキスト読解および筆記再生テストが行われた。英検 2 級と 3 級のテキストを読解し再生する順番は，冊子間でカウンターバランスが取られた。すなわち，1回目の読解で半数の協力者が英検 2 級に取り組み，もう半数の協力者は 3 級のテキストを読解した。2回目の読解では，それぞれがもう一方の受験級に取り組むこととなった。

4.2.4　採点とデータの分析方法

語彙テストの採点はそれぞれ 2 値的に行われた。すなわち，1 つの項目に正答していれば 1 点が与えられた。VLT の信頼性係数 (Cronbach's α) は.953，WAF は.771と，ともに十分な値であった。

筆記再生テストの採点では，すべての読解マテリアルを Ikeno (1996) の基準に従い，調査者がアイデア・ユニット (IU) に分割した。この作業は調査者 1 名のみで行われたため，評価者内一貫性を高めるために，2 週間後にもう一度 IU の分割が行われた。一致率は95.44％であり，不一致点は Ikeno (1996) の基準を再度参照することで解決された。協力者のリコール・プロトコルは上記の IU の観点から，各 IU に対応する情報が再生されている場合に 1 点が与えられた。1 回目の採点時における一致率は92.48％であり，不一致点は再度採点し直すことで解決された。

RQ2-1に答えるため，筆記再生テストの成績を従属変数，Coh-Metrix で分析した語彙特性の値を独立変数とした強制投入法による重回帰分析を行った。その後，テキストの理解度をより良く説明できる回帰モデルを得るために，ステップワイズ法による重回帰分析を実施した。RQ2-2に対しては，筆記再生テストの成績を従属変数，VLT と WAF の成績を独立変数とした強制投入法による重回帰分析を行った。

本調査では分析対象となったテキスト数と協力者数が限られており，独立変数に対して十分なサンプル数 (注9) が得られなかったため，$p < .10$の場合にも結果を解釈することにした。

4.3　結果と考察

4.3.1　テキストに含まれる単語の特性とテキスト理解度の関係

筆記再生テストおよび VLT・WAF の成績を要約したものが表 8 である。まず RQ2-1を解明するため，テキストの語彙特性と筆記再生テストの関係を中心とした分析を行った。

■ 表 8 ：筆記再生テスト・VLT・WAF の記述統計

テストの種類	*n*	*M*	*SD*	*Min*	*Max*	95% CI
筆記再生（全体）	82	47.22	22.23	1.00	85.00	[41.84, 52.60]
2級テキスト	41	31.91	16.34	1.00	66.00	[26.21, 37.61]
3級テキスト	41	62.53	15.92	25.00	85.00	[56.97, 68.08]
VLT	41	100.88	17.98	66.00	129.00	[94.61, 107.16]
WAF	41	113.35	14.00	83.00	145.00	[108.47, 118.24]

（注）CI = confidence interval。VLT の平均点（100.88）は推定語彙サイズが 4,880 語であることを意味する。

表 9 は筆記再生テストの成績と各語彙特性との相関行列を示している。まず，筆記再生率との相関係数 (注10) に着目すると，頻度との間に中程度の相関関係があることがわかる（r = .513）。続いて親密度（r = .477），心像性（r = .463），多様性（r = -.461），具象性（r = .387），名詞の上位語数（r = -.315）という順に，有意な相関関係が見られた。一方，多義性（r = -.073）および動詞の上位語数（r = -.017）と筆記再生率の間に有意な相関係数は得られなかった。

以上の結果は，英検テキストの筆記再生率が下がるのは，(a) テキストに含まれる単語の平均頻度が低く多様な語彙が使用されている，(b) 上位語を多く持つ名詞の使用割合が高い，または (c) 使用されている単語の平均親密度・具象性・心像性が低い場合であることを表している。

調査 1 で予想されたとおり，テキストに含まれる単語の平均頻度が低く，使用される語彙が多様になるほどテキスト内容の再生率は減少することが確認された。2 つの指標は語彙知識の広さを反映しており，テキストに低頻度語が多く含まれていたり，使用されている単語が多様であったりするほど，学習者は自分にとって未知語である単語に遭遇する確率が高くなり，結果としてテキストの理解度が低下したと考えられる（e.g., Hu & Nation, 2000; Laufer, 1989, 1992; Nation, 2001; Schmitt et al., 2011）。

続いて，語彙知識の深さを反映する指標である，単語の多義性と，名詞および動詞の上位語数について，名詞の上位語数のみが筆記再生率と有意な相関関係にあることがわかった。調査 1 では，易しいテキストほど多義性のある単語が多く含まれているという結果が得られていたため，多義性の値はテキストの理解度を予測しない可能性が示唆されていた。調査 2 はこの予測を支持しており，単語の多義性はテキストに含まれる高頻度語の密度を反映していることが原因で（Graesser et al., 2004），テキストの理解度との相関関係が見られなかったと考えられる。

動詞の上位語数と筆記再生率の間に有意な相関関係が見られなかったのは，2級と 3 級のテキストに含まれる動詞の質に，初めから有意な差がなかったためである（表 6 参照）。これに対し，2 級と 3 級で違いのあった名詞の上位語数は，筆記再生率と負の相関関係にあり，1 つの単語（名詞）に含まれる

■ 表 9：筆記再生テストと語彙特性の相関関係（N = 82）

	M	*SD*	1	2	3	4	5	6	7	8	9
1. 筆記再生	47.22	22.23	—								
2. 多様性	74.78	18.06	-.461**	—							
3. 頻度	2.40	0.15	.513**	-.286**	—						
4. 多義性	4.14	0.34	-.073	.088	-.006	—					
5. 上位（名）	6.05	0.65	-.315**	.267*	-.536**	.189	—				
6. 上位（動）	1.56	0.16	-.017	.122	-.181	-.030	.088	—			
7. 親密度	580.41	9.73	.477**	-.353**	.707**	-.006	-.415**	.086	—		
8. 具象性	396.91	25.57	.387**	-.428**	.138	.001	-.183	-.003	.252*	—	
9. 心像性	427.43	25.23	.463**	-.477**	.303**	-.054	-.296**	.022	.419**	.927**	—

（注）*p < .05, **p < .01

上位語数が多いほど，そのテキストの理解度は低下することが示された。ただし名詞の上位語数は単語の頻度と負の相関関係にあり（r = -.536），上位語数の多い単語は低頻度語であるという傾向が見られた。したがって，名詞の上位語数がテキスト理解に与える影響は，単語の頻度の影響も同時に受けていることに留意する必要がある。

語彙知識へのアクセス速度にかかわる語彙特性として，本研究で取り上げた親密度，具象性，および心像性の値は，テキストの再生率とすべて有意な相関関係にあった。これらの語彙特性は単語の解釈しやすさを反映することから（e.g., Crossley et al., 2007），先行研究を支持するという結果になった。ただし，親密度の値は単語の頻度と強い相関関係にあり（r = .707），親密度の高い単語は高頻度語であるという傾向から，単語の親密度が独自にテキスト理解に影響を与えていたのかどうかをさらに分析する必要がある。

相関分析の結果を踏まえ，テキストを構成するさまざまな語彙特性のうち，どれがテキスト理解に独自の影響を与えていたのかを明らかにするため，筆記再生率を従属変数とし，テキストの語彙特性を独立変数とした強制投入法による重回帰分析を実施した。重回帰分析を行う前に，平井（2012）に従って，前提条件となる多重共線性の確認を行った（相関係数が r > .800の場合に多重共線性を疑った）。その結果，具象性と心像性の間に強い相関関係があり（r = .927），多重共線性があると見なした。具象性と筆記再生率の相関係数（r = .387）は，心像性との相関係数（r = .463）よりも低いため，具象性を以降の分析から除外した。

重回帰分析の結果，語彙知識の広さ（頻度・多様性），語彙知識の深さ（多義性・名詞の上位語数・動詞の上位語数），およびアクセス速度（親密度・心像性）にかかわる語彙特性７要因での回帰モデルが有意となり $F(7, 74) = 7.65, p < .001$，単語の多様性・頻度・心像性が有意な説明変数となった。表10にその結果を示す。

次に，データにより良くフィットした回帰式を得るため，ステップワイズ法による重回帰分析を同様の要因計画で行った。表11が示すとおり，単語の頻度・多様性・心像性以外の要因は除去され，回帰モデルは$F(3, 78) = 18.16, p < .001$で有意となり，最終的な決定係数（$R^2$）は41％となった。すなわち，単語の頻度・多様性・心像性の指標を組み合わせることで，英検２級と３級のテキスト理解度の41％が予測されることがわかった。

一連の結果は，英検テキストの読解に影響を与える語彙特性は単語の頻度，多様性，および心像性であったことを示している。まず，単語の頻度と多様

■ 表10：強制投入法による重回帰分析の結果

予測変数	*B*	95% CI	*SE B*	*β*	*t*	*p*
（定数）	-2.045	[-5.048, .958]	1.507		-1.36	.179
頻度	.554	[.140, .967]	.208	.382	2.67**	.009
多様性	-.003	[-.005, .000]	.001	-.249	-2.38*	.020
多義性	-.027	[-.141, .088]	.058	-.042	-0.46	.645
上位語数（名）	.013	[-.059, .085]	.036	.039	0.36	.721
上位語数（動）	.090	[-.160, .341]	.126	.069	0.72	.474
親密度	.001	[-.005, .007]	.003	.037	0.27	.790
心像性	.002	[.000, .004]	.001	.220	2.07*	.042

（注）R^2 = .42（N = 82, p < .001）, CI = confidence interval for B, *p < .05, **p < .01

■ 表11：ステップワイズ法による回帰分析の結果

予測変数	*B*	95% CI	*SE B*	*β*	*t*	*p*
（定数）	-1.479	[-2.453, -.506]	.489		-3.03**	.003
頻度	.540	[.273, .807]	.134	.373	4.03**	< .001
多様性	-.003	[-.005, -.001]	.001	-.243	-2.42*	.018
心像性	.002	[.000, .004]	.001	.234	2.32*	.023

（注）R^2 = .41（N = 82, p < .001）, CI = confidence interval for B, *p < .05, **p < .01

性は語彙知識の広さを反映する指標であり，テキストの中に低頻度語が多く含まれていたり，多様な語彙が使用されている場合にテキストの理解度は低下することが改めて確認された。相関分析の結果は，単語の頻度や多様性が他の語彙特性（e.g., 親密度）と相関関係にあることを示していたが，重回帰分析の結果，頻度や多様性はテキスト理解に独自の影響を与えていたことが明らかとなった。語彙知識の広さがテキスト理解に必須であることは多くの先行研究で支持されているが（e.g., Hu & Nation, 2000; Laufer, 1989, 1992; Nation, 2001; Schmitt et al., 2011），使用頻度の低い単語が使われているだけでなく，多様な語彙が使用されているテキストを理解するのにも語彙知識の広さが求められることがわかった。

一方，語彙知識の深さにかかわる単語の多義性と上位語数は，テキスト理解に独自の影響を与えないという結果になった。特に，相関分析では筆記再生率とのかかわりが示唆されていた名詞の上位語数が，重回帰分析により，テキストの理解度を予測する要因とはならないことが示された。名詞の上位語数は単語の頻度と有意な相関関係にあり（r = -.536），語彙頻度がテキスト理解度を説明する部分が大きいために，相対的に名詞の上位語数が説明する独自の部分が小さくなったと考えられる。単語の多義性については調査 1 で予測されたとおり，英検テキストの理解度とはかかわらないことが支持された。同様に，動詞の上位語数は 2 級と 3 級テキストの間で本質的な違いがなかったため，今回の調査ではテキスト理解とかかわらないという結果になった。

3 つ目の観点として，テキスト読解中に遭遇した単語をどれだけ効率よく処理できるかについては，単語の心像性のみがテキストの理解度を説明する独自の変数となることがわかった。重回帰分析には単語の親密度も要因計画に含められていたが，この結果は親密度の独自説明部分が心像性よりも小さかったことを意味する。表 9 にて示したとおり，単語の親密度は頻度と強い相関関係（r = .707）にあったため，単独での説明力が低くなったと考えられる。これに比べて，心像性と頻度の相関係数は低く（r = .303），心像性という語彙特性はテキスト理解に独自の影響を与えることが明らかになった。

以上の結果を踏まえ，今回の調査協力者と同等のレベルにある学習者が英検 2 級または 3 級を読解したときに予測される筆記再生率は，次の回帰式で求めることができる。

英検 2 級と 3 級の筆記再生率（%）
= 0.54×（頻度）− 0.003×（多様性）+ 0.002×（心像性）− 1.479［R^2 = .41］

例えば，Coh-Metrix 3.0で2011年度第 1 回英検 3 級の第 4 問C（*Roald Dahl*）を分析すると，単語の頻度は2.52，多様性は70.26，心像性は457.71となる。これらの値を上の数式に代入すると，予測される筆記再生率は58.64%となる。本調査でこの値に近い成績であった学習者のリコール・プロトコルを資料に示す。次節では，これだけのテキスト理解度に到達するのに，学習者にはどれくらいの語彙熟達度が求められるのかを分析する。

4.3.2 語彙熟達度とテキスト理解度の関係

RQ2-2を解明するため，英検 2 級と 3 級の筆記再生率と協力者が持つ語彙知識の広さおよび深さの関係に焦点を当てた分析を行った。表12は英検各級の筆記再生率および VLT・WAF の相関行列を示している。

全体的な傾向として，筆記再生率は VLT および WAF と中程度から強い相関関係にあった。語彙知識の広さがテキスト理解と密接な関係にあることは多くの先行研究で明らかにされており（e.g., Hu &

■ 表 12：筆記再生テスト・VLT・WAF の相関関係（N = 41）

	M	*SD*	1	2	3	4
1. VLT	100.88	17.98	—			
2. WAF	113.35	14.00	.765**	—		
3. 2級の筆記再生率	31.91	16.33	.630**	.734**	—	
4. 3級の筆記再生率	62.53	15.92	.604**	.604**	.690**	—

（注）*p < .05, **p < .01.

Nation, 2000; Laufer, 1989, 1992; Schmitt et al., 2011），中程度の相関が見られることも先行研究の指摘（e.g., Grabe, 2009）と一致していた。次に，語彙知識の深さを測定する WAF がテキスト理解度と正の相関関係にあったことは，Qian（1999, 2002）や Zhang（2012）と同様の結果であった。興味深いことに，WAF とテキストの再生率との相関係数は，VLT との相関係数よりも高い場合があり，語彙知識の広さが一定量を超えると，今度は語彙知識の深さがテキスト理解に必要になるという先行研究（e.g., Qian, 1999, 2002）を裏づけるものとなった。しかし，VLT と WAF は互いに強い相関関係（r = .765）にあるため，それぞれがテキスト理解度に対してどれだけの影響を独自に与えていたのかを調べる必要がある。

VLT と WAF がテキスト理解にどれだけ影響を与えていたのかを明らかにするため，英検 3 級と 2 級の筆記再生率をそれぞれ従属変数とした強制投入法による重回帰分析を実施した。VLT と WAF の間には多重共線性が存在する可能性もあるが，本調査では平井（2012）に従い，相関係数が.800を下回る場合には要因を除外せずに分析を行った。表13と表14にそれぞれの結果を示す。

英検 3 級のテキスト理解度については，有意傾向にとどまったものの，VLT と WAF がそれぞれ独自の影響を与えていた。これに対し，英検 2 級では WAF のみが筆記再生率を有意に予測する変数となった。すなわち，全体的な傾向として，語彙知識の深さが英検 2 級および 3 級のテキストを理解するのに必要であることが示唆された。

しかしながら，VLT と WAF の間には強い相関関係があることを考慮すると，WAF のスコアは語彙知識の広さもある程度反映しており，広さと深さを両方測定した総合的な語彙熟達度を示していると考えられる。また，VLT は語彙知識の深さを測定していない分，テキスト理解を独自に説明する部分が小さくなったのではないかと想定される。すなわち，英検 2 級のテキスト理解を VLT が予測しないという結果は，語彙知識の広さが 2 級のテキスト理解に必要ないということを意味するのではなく，英検のテキスト理解には語彙知識の広さと深さの両方が必要になることが示唆される。

さらに，語彙熟達度の高い学習者は自身が持っている単語の知識を流暢に利用できることから（Crossley et al., 2012），WAF で高得点を取れる学習者は語彙知識の流暢さも有していると想定される。本研究では語彙知識へのアクセス速度を測定するテストを実施できなかったものの，英検テキストの理解度には語彙知識へのアクセス速度に影響を与える心像性がかかわることが示されており（4.3.1節参照），心像性の低い単語が多く使用されているテキストでは語彙知識の流暢さも必要になると示唆される。

一方, 語彙知識の深さを反映する単語の特性（i.e., 多義性・上位語数）がテキストの理解度を予測しないという結果が出ているにもかかわらず，今回の調査では語彙知識の深さを測定している WAF の成績がテキスト理解度を予測するという結果になった。

■ 表13：強制投入法による重回帰分析の結果（3級テキスト）

予測変数	*B*	95% CI	*SE B*	*β*	*t*	*p*
（定数）	-.185	[-.555, .184]	.183		-1.02	.316
WAF	.004	[-.001, .009]	.002	.341	1.78	.083
VLT	.003	[.000, .006]	.002	.341	1.77	.085

（注）R^2 = .41（N = 41, p < .001）, CI = confidence interval for B, $^*p$ < .05, $^{**}p$ < .01
回帰モデルは F (2, 38) = 13.40, p < .001 で有意。

■ 表14：強制投入法による重回帰分析の結果（2級テキスト）

予測変数	*B*	95% CI	*SE B*	*β*	*t*	*p*
（定数）	-.637	[-.942, -.333]	.150		-4.24**	< .001
WAF	.007	[.003, .011]	.002	.608	3.60**	.001
VLT	.001	[-.001, .004]	.001	.165	0.98	.334

（注）R^2 = .55（N = 41, p < .001）, CI = confidence interval for B, $^*p$ < .05, $^{**}p$ < .01
回帰モデルは F (2, 38) = 23.29, p < .001 で有意。

このような矛盾は，Coh-Metrix が算出する指標とWAF が測定する知識が，厳密には一致していないことが原因だと考えられる。すなわち，英検テキストの理解には語彙知識の深さが求められると仮定した場合，具体的に単語の何を深く知っていればテキストの理解度が向上するのかは，Coh-Metrix で分析可能な多義性および名詞・動詞の上位語数という観点からはとらえられないと言える。

最後に，VLT と WAF のスコアから英検テキストの筆記再生率を予測するための回帰式を示す。表14で言及したとおり，VLT は有意な説明変数にならない場合があるものの，VLT は容易に利用可能なテストであることを考慮して数式に含めることとした。

英検 3 級の筆記再生率（%）
= 0.003×（VLT スコア）＋ 0.003×（WAF スコア）
［R^2 = .41］
英検 2 級の筆記再生率（%）
= 0.001×（VLT スコア）＋ 0.007×（WAF スコア）
− 0.637［R^2 = .55］

この回帰モデルを応用すると，英検 3 級および 2 級のテキスト理解に必要な語彙熟達度を推定することができる。例えば英検 3 級のテキストで60%（資料参照）という非常に高い筆記再生率を望む場合，学習者の語彙知識の深さ（WAT）が80点から100点の範囲にあれば，求められる語彙サイズは4,462語から5,231語（VLT 換算で90点から110点）という試算になる。

4.4 調査 2 のまとめ

調査 2 では，英検 2 級と 3 級のテキストを理解するのに求められる語彙知識の側面を明らかにするため，英語学習者を対象とした実験を行った。検証の観点として，調査 1 で分析したテキストの語彙特性，学習者の語彙熟達度，およびテキスト理解度との関係性を明らかにしようとした。実験の結果，英検 2 級と 3 級のテキスト理解には単語の頻度，多様性および心像性が影響を与えていることが示された。ゆえに，語彙知識の広さを測定する語彙テストが英検テキストの理解度を説明すると予想されたが，テキスト理解には語彙知識の広さと深さの両方がかかわることが示唆された。したがって，Coh-Metrix 3.0によるテキスト分析では，英文読解に求められる語彙知識の深さを反映する語彙特性をとらえきれないという限界点が示された。

5 結論と今後の課題

本研究において，Coh-Metrix 3.0によるテキスト分析および日本人英語学習者を対象に行った実験で得られた成果を，本研究の限界点と合わせて総括する。

本研究では，初めに，英検で用いられている長文読解テキストを対象として，Coh-Metrix 3.0で測定される各テキストの語彙特性が受験級によってどのように異なるのかを検証した。1 級から 3 級までのテキストを比較した結果，受験級によってテキストに含まれる単語の特性が変化し，テキストを理解するのに求められる語彙知識の側面も異なることが示唆された。調査 1 では特に，上位の受験級ほど (a) 使用されている単語の頻度が低くなり，単語の種類も豊富になる，(b) 単語の意味へアクセスしづらくなる，一方で (c) 単語の多義性や上位語数は受験級の難易度と一貫しないことを明らかにした。これらの結果から，難しい受験級で出題される長文読解テキストを理解するためには，さまざまな低頻度語を広く知っていること，そして，遭遇した単語を効率よく処理する語彙知識の流暢さが求められると考察した。さらに，英検テキストの理解には単語の多義性や語連想の知識が反映されにくいと予測した。

これらの予測が支持されるのか検証するため，調査 2 では英語学習者を対象に英検テキストの筆記再生テストおよび語彙テストを行った。Coh-Metrix 3.0で算出される語彙特性の指標がどの程度テキスト理解を説明できるか検証したところ，単語の頻度，多様性，および心像性が独自の影響を与えていることがわかった。この傾向は調査 1 の結果を支持しており，受験級が上位にあるテキストを読む際には，そのテキストに含まれる低頻度語やイメージしづらい単語がテキストの理解度を下げたことを意味する。さらに，英語学習者の語彙熟達度とのかかわりを分析した結果，主に語彙知識の深さがテキスト理解度を予測することが確認された。しかし，VLT と WAF の間には強い相関関係があり，WAF は学習者の語彙サイズをある程度反映していることを考慮す

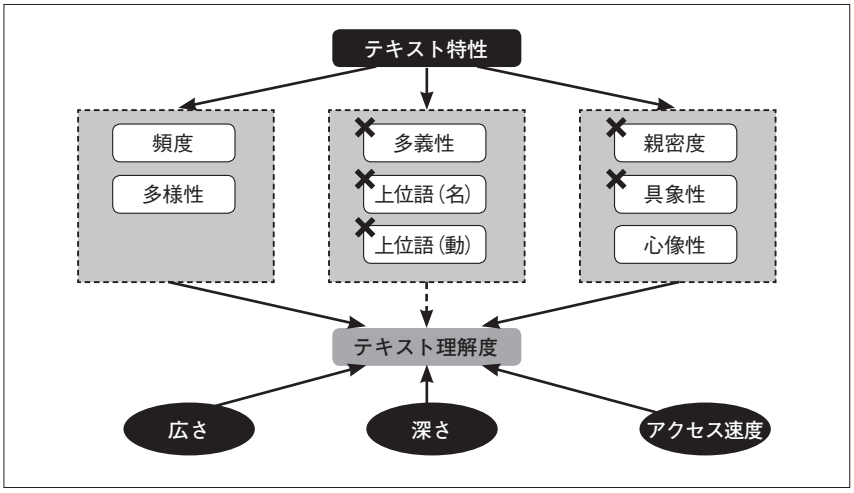
ると，難しい受験級のテキストを理解するためには語彙知識の広さと深さの両方が必要になると考察した（図7参照）。

しかしながら，本研究には留意すべき点がいくつか残されている。特に調査2では，分析対象となったテキスト数と協力者数が少なく，回帰モデルの信頼性や一般化可能性に課題が残る。さらに，同一の協力者が2種類のテキストを読解しており，データの独立性についても改善の余地がある。また，調査に参加した協力者は全員が大学生であり，本研究の結果は，英語力が限られた範囲にある学習者集団から得られたものであることに注意すべきである。これらの限界点を克服するためには，さまざまな熟達度にある学習者を対象に大規模な調査を行わなければならない。

もう1つの限界点として，本研究では，英検テキストの理解に必要な語彙熟達度を，語彙知識の広さ・深さ・アクセス速度という個別の観点から数値化することはできなかった。特に単語の意味へのアクセス速度を測定するテストはまだ実用段階になく，今後より妥当な語彙テストの開発が求められる。同様に，Coh-Metrix 3.0では語彙知識の深さを反映する語彙特性をとらえることができなかったため，学習者が持つ語彙知識の深さが反映されるような読解テストを作成するためには，方法論に関する検討が求められる。例えば3.3節で述べたとおり，多義語がテキスト理解に与える影響を検証するために，対象となる単語がどのような位置づけにあるのかを質的に観察する必要があるだろう。

最後に，Coh-Metrix を用いた英語教育研究および実践への示唆を述べる。本研究の最終的な目標である，英検テキストを読解するのに必要な語彙熟達度の数値化は，次のような手順で行うことができる。例えば Coh-Metrix で分析した英検2級の平均的なテキスト（表2参照。頻度 = 2.30，多様性 = 85.45，心像性 = 414.04）を本研究の協力者と同程度のレベルにある学習者に読ませると，予測される筆記再生率は（0.54 × 2.30）−（0.003 × 85.45）+（0.002 × 414.04）− 1.479 = 33.47％となる（4.3.1節参照）。それだけの理解度に達するためには，4.3.2節で示した回帰モデルを利用すると，WAF スコアが120点から125点の範囲にある場合，求められる語彙サイズは4,692語から6,000語になると推定される。

このような手順で，本研究で得られた回帰モデルを応用すると，新規に作成したテキストの読解に必要な語彙熟達度を予測することができる。例えば，英検3級の合格者と不合格者を区別して，本研究と同様の大規模な調査を行い，3級合格者（または不合格者）のテキスト理解度（長文読解問題の得点）を算出する。その際，学習者の語彙熟達度もさまざまな観点から測定しておく必要がある。そして，本研究と同様の手順でそれぞれのデータを分析し，より妥当な回帰モデルを得ることで，新しく作成した英検3級のテキストで合格点に到達するのに必要な語彙熟達度を提示することが可能になるだろう。



▶ 図 7：テキストに含まれる単語の特性および語彙知識とテキスト理解度のかかわり

謝　辞

本研究を発表する貴重な機会を与えてくださいました公益財団法人日本英語検定協会と関係者の皆様，ならびに選考委員の先生方に厚く御礼申し上げます。特に，助言者である池田央先生には，有益なご指導をいただき大変感謝しております。そして，筑波大学の卯城祐司先生，順天堂大学の小泉利恵先生，および福島大学の髙木修一先生には，本研究に対して実施方法から統計分析までさまざまなご助言をいただきました。また，研究室で共に学ぶ，名畑目真吾さん，長谷川佑介さん，木村雪乃さんにも多くの示唆と激励の言葉をいただきました。最後に，本調査を実施するにあたってご協力いただいた，東京経済大学の中川知佳子先生と学生の皆様に深く御礼申し上げます。

注

(1) ここでの語数はワードファミリー（word family）を指す。ワードファミリーの数え方では，基本形（e.g., happy）に加えて活用形（e.g., happier・happiest）および派生形（e.g., happily・happiness・unhappy）をまとめて 1 語と数える（Nation, 2001）。

(2) ここでの語数はレマ（lemma）を指す。レマの数え方では基本形とその活用形をまとめて 1 語と数える（Nation, 2001）。ワードファミリーで数えた語数をレマに換算する妥当な方法はないが，一般的にワードファミリーで数える方がレマ換算よりも語数は少なくなる。

(3) ただし，高頻度語ほど多くの意味を持つ傾向があるため，多義性の値の高さはテキストに含まれる高頻度語の多さを反映することがある（Graesser et al., 2004）。

(4) 効果量の値は（水本・竹内, 2008 を参照），ここでは受験級の違いがテキストに含まれる単語の各特性にどれだけ強い影響を及ぼしていたかを表している。

(5) 効果量 η^2 の値は「ある要因の平方和/全体平方和」で求められ，$.010 < \eta^2 < .060$ の場合は効果量小，$.060 < \eta^2 < .139$ の場合は効果量中，$.139 < \eta^2$ の場合は効果量大と解釈される（Cohen, 1988）。

(6) 他にも多義性や動詞の上位語数は 2 級と 3 級のテキスト間で違いが見られなかったものの，1 級や準 1 級のテキストを用いるよりは適切だと判断した。ゆえに本研究では多義性や動詞の上位語数がテキスト理解に与える影響を検証できるデザインではないことに留意されたい。

(7) 効果量 r の値は「$t^2/(t^2 + df)$ の平方根」で求められ，$.10 < r < .30$ の場合は効果量小，$.30 < r < .50$ の場合は効果量中，$.50 < r$ の場合は効果量大と見なされる（Cohen, 1988）。

(8) Zhang（2012）で指摘されているとおり，読解に制限時間を設定することで言語処理の流暢さを反映させられることから，本研究では比較的厳しい時間を設定するため，英語に熟達した日本人大学生 2 名を対象とした予備調査に基づき英検 2 級と 3 級の読解時間を決定した。具体的には，2 名の読解時間の平均を制限時間として使用した。

(9) 回帰分析では厳密な基準はないものの，多くのサンプル数を必要とする（平井, 2012）。具体的には，信頼性のある決定係数（R^2）を得るために $50 + k$（k = 独立変数の数）のサンプルが，各独立変数の有意性を検定するためには $104 + k$ 以上のサンプルが必要になる。したがって，本調査の結果をより精緻なものにするためには，少なくとも 120 以上のテキスト数と協力者数が求められる。

(10) 相関係数の解釈は平井（2012）に基づく：$r = .00$ 〜±.20（ほとんど相関がない），±.20 〜±.40（弱い相関がある），±.40 〜±.70（中程度の相関がある），±.70 〜±1.00（強い相関がある）。

参考文献（＊は引用文献）

＊相澤一美・望月正道（編著）.(2010).『英語語彙指導の実践アイデア集：活動例からテスト作成まで』. 東京：大修館書店 .

＊Baayen, R.H., Piepenbrock, R., & Gulikers, L.(1995). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium.

＊Cohen, J.(1988). *Statistical power analysis for the behavioral sciences*（2nd ed.）. Hillsdale, NJ: Lawrence Erlbaum Associates.

＊Coltheart, M.(1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 33*, 497-505.

＊Crossley, S.A., Greenfield, J., & McNamara, D. S.(2008). Assessing text readability using cognitively based indices. *TESOL Quarterly, 42*, 475-493.

＊Crossley, S.A., Louwerse, M.M., McCarthy, P.M., & McNamara, D.S.(2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal, 91*, 15-30.

＊Crossley, S.A., Salsbury, T., & McNamara, D.S.(2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing, 29*, 243-263.

＊Crossley, S.A., Salsbury, T., McNamara, D.S., & Jarvis, S.(2011a). Predicting lexical proficiency in language learners using computational indices. *Language*

Testing, 28, 561-580.
＊Crossley, S.A., Salsbury, T., McNamara, D.S., & Jarvis, S.(2011b). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly, 42*, 182-193.
＊Daller, H., Milton, J., & Treffers-Daller, J.(2007). *Modeling and assessing vocabulary knowledge*. Cambridge, England: Cambridge University Press.
＊Ellis, N.C., & Beaton, A.(1993). Psycholinguistic determinants of foreign language vocabulary acquisition. *Language Learning, 43*, 559-617.
＊Elston-Guttler, K.E., & Friederici, A.D.(2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language, 52*, 256-283.
＊Grabe, W.(2009). *Reading in a second language: Moving from theory to practice*. New York, NY: Cambridge University Press.
＊Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z.(2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, 36*, 193-202.
＊平井明代（編著）.(2012).『教育・心理系研究のためのデータ分析入門：理論と実践から学ぶ SPSS 活用法』. 東京：東京書籍.
＊Hu, M.H., & Nation, I.S.P.(2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*, 403-430.
＊Ikeno, O.(1996). The effects of text-structure-guiding questions on comprehension of texts with varying linguistic difficulties. *The Japan Association of College English Teachers Bulletin, 27*, 51-68.
＊Iso, T., Aizawa, K., & Tagashira, K.(2012). The development and validation of a test of lexical access. *Annual Review of English Language Education in Japan, 23*, 217-231.
＊Laufer, B.(1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman(Eds.), *Special language: From humans thinking to thinking machines*（pp.316-323). Clevedon, England: Multilingual Matters.
＊Laufer, B.(1992). How much lexis is necessary for reading comprehension? In P.J.L. Arnaud & H. Bejoint(Eds.), *Vocabulary and applied linguistics*(pp.126-132). London, England: Macmillan.
＊McCarthy, P.M., & Jarvis, S.(2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*, 381-392.
＊Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K.(1990). *Five papers on WordNet*. Princeton, NJ: Princeton University.
＊水本篤・竹内理.(2008).「研究論文における効果量の報告のために―基礎的概念と注意点―」.『英語教育研究』, 31 号 , 57-66.
＊望月正道.(1998).「日本人学習者のための英語語彙サイズテスト」.『語学教育研究所紀要』, 12 号 , 27-53.
＊望月正道・相澤一美・投野由紀夫.(2003).『英語語彙の指導マニュアル』. 東京：大修館書店.
＊Nation, I.S.P.(2001). *Learning vocabulary in another language*. New York, NY: Cambridge University Press.
＊Perfetti, C.A.(2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383.
＊Qian, D.D.(1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*, 282-308.
＊Qian, D.D.(2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*, 513-536.
＊Read, J.(1998). Validating a test to measure depth of vocabulary knowledge. In A.J. Kunnan(Ed.), *Validation in language assessment*(pp.41-60). Mahwah, NJ: Lawrence Erlbaum Associates.
＊Schmitt, N., Jiang, X., & Grabe, W.(2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*, 26-43.
＊Schmitt, N., & Meara, P.(1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition, 19*, 17-36.
＊杉森直樹.(2011).「語彙力の測定・評価」. 石川祥一・西田正・斉田智里（編著）.『テスティングと評価：4 技能の測定から大学入試まで』. 237-252. 東京：大修館書店.
＊Verspoor, M., & Lowie, W.(2003). Making sense of polysemous words. *Language Learning, 53*, 547-586.
＊Wolter, B.(2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*, 41-69.
＊Zhang, D.(2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. *The Modern Language Journal, 96*, 558-575.

資　料

英検 3 級テキストとリコール・プロトコルの例

Yuki's Wonderful Day in California（2000 年度第 1 回・問 4 C）

Last summer, Yuki and her family left Nagoya and went to Sacramento, California. It was really hot, but the air was dry, so Yuki liked the weather better than in Nagoya.

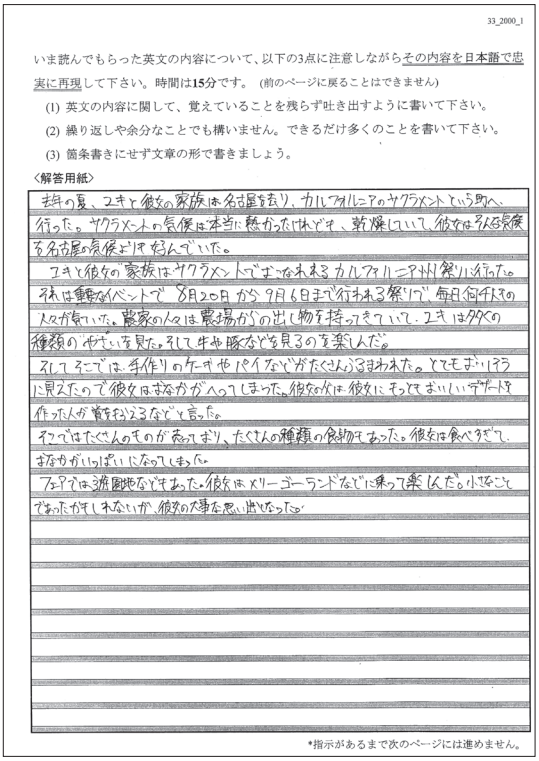
While Yuki and her family were there, they visited the California State Fair. It is an important event held every summer. That summer, it was held from August 20 to September 6. Many thousands of people visited the fair every day.

At the fair, Yuki saw lots of interesting things. Farmers were showing things from their farms. Yuki saw many kinds of vegetables. She was surprised to see how big some of them were. And it was fun to see cows, pigs, and other farm animals.

There were also lots of homemade cakes and pies at the fair. The desserts looked so delicious that Yuki became very hungry. Her father told her that the people who made the best desserts got prizes.

A lot of people were selling things at the fair. There were all kinds of foods, toys, and things that people made. Yuki bought a piece of pizza for lunch. It was very big, so she couldn't eat all of it.

There was even an amusement park at the fair. Yuki liked it the best because there were many rides and games. She rode the merry-go-round and played lots of games. She won a game and got a pretty doll. Yuki had a wonderful time at the fair.



33_2000_1

いま読んでもらった英文の内容について、以下の3点に注意しながらその内容を日本語で忠実に再現して下さい。時間は15分です。(前のページに戻ることはできません)

(1) 英文の内容に関して、覚えていることを残らず吐き出すように書いて下さい。
(2) 繰り返しや余分なことでも構いません。できるだけ多くのことを書いて下さい。
(3) 箇条書きにせず文章の形で書きましょう。

〈解答用紙〉

去年の夏、ユキと彼女の家族は名古屋を去り、カルフォルニアのサクラメントという町へ、行った。サクラメントの気候は本当に暑かったはずで、乾燥していて、彼女は名古屋の気候を名古屋の気候よりは好んでいた。

ユキと彼女の家族はサクラメントで行われるカルフォルニア州祭りへ行った。それは緑がベントで8月20日から9月6日まで行われる祭りで、毎日何千もの人々が集いた。農家の人々は農場からの出し物を持ってきていて、ユキは多くの種類のやさいを見た。そして牛や豚などを見るのを楽しんだ。

そしてそこでは、手作りのケーキやパイなどがたくさん売られた。とてもおいしそうに見えたので彼女はおなかがへってしまった。彼女の父は彼女にそれらのデザートを作った人が賞をおえるなどと言った。

そこではたくさんのものが売られており、たくさんの種類の食物であった。彼女は食べすぎて、おなかがいっぱいになってしまった。

フェアは遊園地などであった。彼女はメリーゴーランドなどに乗って楽しんだ。小さなことであったかもしれないが、彼女の大事な思い出となった。

*指示があるまで次のページには進めません。