

第32回 研究助成

A 研究部門・報告 I・英語能力テストに関する研究

ライティングタスク(技能独立型 vs. 統合型)が 発表語彙とその測定に与える影響: TAALESによる語彙の洗練性分析を基に

研究者: 茨城県/筑波大学大学院 在籍 小室 竜也

《研究助言者: 寺内 一》

概要

本研究は英作文に含まれる発表語彙の洗練性を数値化するツール Tool for Automatic Analysis of Lexical Sophistication (TAALES) を用いて、技能独立型および技能統合型のライティングタスクと発表語彙のかかわりを検証した。

調査1では英検1級から3級で用いられているテキストを対象とし、受験級ごとにどのような洗練性指標が分類を予測するかを分析した。その結果、上位の受験級では心理言語学的な要因である長い反応時間や英語母語話者の習得年齢が高い語彙が使用されていることがわかった。

調査2では大学生を対象に独立型および統合型のライティングタスクを実施し、タスクの分類を予測する洗練性指標を分析した。その結果、発表語彙における語と語の結びつきの強度、音韻的に類似する語彙、内容語の多義性の3点がタスクを分類するのに寄与していることが示された。

1 研究の背景

4技能5領域の指導の中でも、リスニングとリーディングは受容技能、スピーキング(発表・やりとり)とライティングが産出・発表技能と分類される。さらに、リーディングが語彙認知や統語解析、推論生成などに下位区分できるように、語彙知識も連語や形態素などのいくつかの側面を持っていると考えられている(Nation, 2013)。

ライティングやスピーキングにおいて必要とされる語彙知識は発表に関する側面であり、発表語彙知識はスピーキングおよびライティングテストの得点を説明する重要な要素である(Kyle & Crossley, 2015, 2016)。ライティングにおける発表語彙を測定する方法の1つに、実際に生徒が書いた英文をいくつかの観点から分析するというものがある(Laufer & Nation, 1995)。この方法で多く用いられるのは特定のトピックが与えられ、それについて自分の意見を自由に書く技能独立型の形式である。現在、日本の英語教育の実践現場では4技能の統合的な指導が行われているが、複数技能を組み合わせたライティング活動における発表語彙の特徴は十分に明らかではない。そこで本研究では技能独立型と技能統合型のタスクの違いによって、作文中で用いられている発表語彙はどのような影響を受けているのかを検証することを目的とする。

本稿では2つの調査結果を報告する。調査1では、特定のトピックについて自分の意見を述べる形式(技能独立型)である英検のモデル英作文を語彙の洗練性分析ツール (Tool for Automatic Analysis of Lexical Sophistication; TAALES) によって調査することにより、各受験級における発表語彙の特徴を把握する。続く調査2では技能独立型および技能統合型(トピックに関する英文を読解後、自分の意見を書く)タスクを実施し、タスクの違いによる影響が発表語彙の特徴に与える影響をTAALESによって調査する。これらの2つの調査を通し、4技能を総合的・統合的に指

導する教育実践者に対して示唆を与えることを目的とする。

2 先行研究

2.1 発表語彙の測定

新学習指導要領では高校卒業時まで、小学校からの合計で約4,000から5,000語の語彙を指導することが示されている。語彙知識が言語運用能力の基礎の1つであることは多くの教育実践者に受け入れられており、学習者自身も語彙知識の重要性を自覚している。その語彙知識の中でも、低頻度の難しい語彙の知識を有していることは熟達した学習者であることを意味しており、洗練された語彙を知っている・使うことができることは言語熟達度の指標の1つとして機能する(Lu, 2010)。この考えに基づき、学習者がどのくらい語彙をたくさん知っているか測定する診断テストが多く開発された(e.g., Vocabulary Size Test; Nation & Beglar, 2007; New Vocabulary Level Test: Webb, Sasao, Ballance, 2017)。

日本人英語学習者の語彙サイズを診断するテストとして最も一般的なものは望月語彙サイズテスト(相澤・望月, 2010)である(図1)。これは頻度レベルごとに日本人英語学習者が知っておくべき8,000語の語彙をリスト化した大学英語教育学会基本語リストJACET8,000(石川 他, 2003)に基づき、1,000語レベルの頻度バンドごとにランダムに抽出した語彙の意味を選択肢の中から選ぶ。例えばcraftsmanは5,000語レベルに属しているため、craftsmanの意味を知っている学習者は5,000語よりも簡単な頻度の高い語彙は知っていると見なすことができる。しかし実際の測定は複雑であり、3,000語レベルの問題で誤りがあった学習者が7,000語レベルの問題をたまたま解答することができる可能性もある。このようにいくつかの問

題点は抱えているが、受容技能に関する語彙知識を測定する診断テストは様々なものが開発されている。しかし、学習者の発表語彙を測定する方法は受容語彙に比べてあまり発達していない。

発表語彙の測定には大きく2種類のアプローチがある。1つはこれまでに開発されてきた受容語彙サイズテストを援用し、発表語彙知識を直接測定するものである。もう1つは作文やスピーチといった学習者がアウトプット活動で実際に産出した成果物を間接的に分析する方法である。本節ではそれぞれの利点と欠点について総括する。

2.1.1 直接的な測定

Laufer and Nation(1999)のProductive Level Test (PLT)では、受容語彙サイズを測定するVocabulary Level Test(Nation, 1983)を援用した発表語彙サイズテストである。このテストでは頻度レベルごとに目標語がランダムに抽出され、適切な目標語だけが正しく産出されるように目標単語の一部の文字を提示した上で、受験者は目標語を母語から第二言語へ翻訳する。例えば、Laws are based upon the principle of jus _____.の中で目標単語justiceの数文字が提示される。この方法は頻度に基づいて、発表語彙知識を直接測定していると言える。数文字を与えることにより、テスト作成者は厳密に頻度レベルを統制することができるため、具体的な数値で発表語彙サイズが算出され直感的に理解しやすい。さらに直接的に発表語彙知識を測定することにより、受験者が自信のない語彙の使用を回避することをなくすことができる。

PLTは効率よく発表語彙知識を測定できるが、以下の2点の問題点がある。まず実際には外部から刺激が与えられた状態でのみ使用できる語彙知識を測定しているため、受容語彙知識の単なる翻訳テスト版であるとも考えられる。次に、その結果が必ずしも言語使用の場面で目標語が使えるかどうかとは関係がない可能性である。学習者が自らの意思で使いこなすことができる発表

■図1: 望月語彙サイズテストの例(相澤・望月, 2010)

1. 持っている			2. しなければならない		
(1) do	(2) get	(3) give	(4) have	(5) must	(6) raise

語彙知識を測定するためには、実際に英作文を書かせたりスピーチをさせたりしたアウトプットを分析する方法が必要である。

2.1.2 間接的な測定

発表語彙知識を測定する2つめの方法としては、学習者のアウトプットを様々な観点から分析する間接的な測定がある。具体的には、英作文で使われている語彙を4つの頻度レベル(1,000語, 2,000語, 大学語彙, それより頻度の低い語)に分け、それぞれのレベルに含まれる語彙の割合を表す語彙頻度プロフィール(Laufer & Nation, 1995)が多く用いられている。例えば、英作文100語のうち70語が1,000語レベル、20語が2,000語、5語が大学語彙、5語がその他の固有名詞で構成されるとすると、70%-20%-5%-5%と表される。

語彙頻度プロフィールを発表語彙能力の測定方法として考える前提には、熟達した学習者は難しい語彙を知っており使いこなせるというものがある。つまり、I think… I think…のように高頻度語を繰り返し使う学習者よりも、I suppose… I assume…のように低頻度語を使うことができる学習者の方が熟達した発表語彙知識を有していると言える。なお、語彙頻度プロフィールに必要なプログラムはPaul Nation氏のウェブページにて無料で使用可能である。他にもP_Lexというツールも無料で使うことができる(図2)。

語彙頻度プロフィールやP_Lexは言語使用の場面において産出された語彙を分析することか

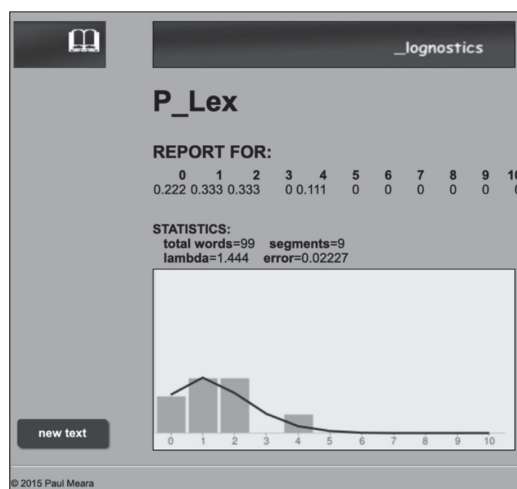
ら妥当性が高い一方、発表語彙サイズが大きいのかどうか直感的に分かりにくい。このデメリットを改善したのがKojima and Yamashita (2014)のS (http://kojima-vlab.org/lexical_richness/S_English.html)というプログラムである(図3)。Sは最終的なアウトプットで使用されている語彙を頻度レベルに基づいて分析し、書き手の発表語彙知識を算出する。例えば、ある生徒が書いた作文をSのプログラムで分析すると、2795.46という数値が出てきたとする。このことが意味するのは、書き手の発表語彙サイズは2795.46語であるということである。しかし、これらのプログラムは頻度レベルという観点のみに基づいており、語彙の豊かさの観点からは問題がある。

2.2 語彙の豊かさ (Lexical Richness)

語彙の豊かさという用語は1人の書き手が使用する語彙の数のことを指している。近年の第二言語でのライティングやスピーキングの研究の評価指標には複雑性、正確性、流暢性に語彙の豊かさという概念を加えた指標が用いられることが多い(Polio & Friedman, 2016)。

2.2.1 語彙の数え方

I like playing tennis. I am a member of the tennis club. I play tennis every day. という英作文ではすべての語を1語と数える方法で17語となる。しかし、人称代名詞のIと普通名詞のtennisは



■ 図2: P_Lexの分析結果の画面



■ 図3: Sの分析結果の画面

3回登場している。分かち書きされた語を1語とカウントする方法は総語数(token), 重複を1語とする数え方は異語数(type)とそれぞれ呼ばれている。

2.2.2 語彙の多様性

多様な種類の語彙を使うことができる学習者は、熟達した発表語彙知識があると考えられる(Laufer & Nation, 1995)。例えば、上述のI thinkだけでなくI supposeやI assumeなどの様々な表現方法を使いこなすことが熟達した書き手には可能である。このことから、語彙の多様性が高い英作文は高く評価されると言える。この多様性の指標には伝統的にTypeをTokenで割ったType Token Ratio (TTR)が用いられることが多い。

しかし、TTRには語数が長くなれば多様性は減少してしまうという問題点がある。例えば、熟達度の高い学習者は流暢性が高いため多くの語彙を産出することができるが、多くの語彙を産出すると冠詞や前置詞の割合が高まり、多様性が低くなってしまふ。つまり、本来であれば熟達度が低いと評価されるはずの、流暢性が低い作文やスピーチをした学習者の方が多様性のある語彙を用いることができると評価されてしまふ。文章の長さに影響を受けてしまふというTTRの欠点を改善するためにGuiraud IndexやMeasure of Textual Lexical Diversity (MTLD), vocD, HD-Dなどの様々な指標が提案され(Kyle, 2019)、ライティング研究ではその特徴に応じてそれぞれの指標が報告されている。

語彙の多様性は、作文中で使用されている語彙の難易度が低い場合の判断に根本的な問題がある。例えば、I love tennis very much. I want to be like Nishikori. という10 type 11 tokenの作文は $10/11 = 0.91$ TTRと計算され、The rate of juvenile delinquency is increasing at an alarming rate. という10 type 11 tokenの作文と同じ多様性レベルであるとみなされてしまふ。この問題点の解決策に語彙の洗練性指標が挙げられる。

2.2.3 語彙の洗練性

洗練された語彙は高度で難しい単語であると一般的に考えられ(Lafuer & Nation, 1995)、語

彙の多様性と類似した傾向を示すが、実際には語彙の洗練性は多様性とは異なるものである(Jarvis, 2013)。例えば、1人は3,000語レベルの易しい語彙を使って100語の作文をし、もう1人は5,000語レベルの難しい語彙で100語の作文をしたとする。前者は上限が3,000語レベルであるため5,000語レベルの語彙を使う後者よりも語彙の多様性は低くなり、高頻度の易しい語彙を使っている前者の方が洗練性も低いと言える。しかし上述のThe rate of juvenile delinquency is increasing at an alarming rate. の例のように、作文が同程度の多様性を示す場合には洗練性の指標を参照する必要がある。つまり、語彙の多様性が作文全体に占める特定の語彙の割合であるとするならば、語彙の洗練性は作文全体に占める難易度の高い洗練された語彙のことを指す(Kim, Crossley, & Kyle, 2018)。

語彙の洗練性は多面的であり、難易度の高低は簡単に決めることはできないが、歴史的には頻度に基づいて研究が発展してきた(e.g., Laufer & Nation, 1995)。2.1.2にて触れたように、作文中で5,000語レベルの語彙の割合が高くなれば洗練された語彙を使用しており、その書き手は熟達した発表語彙知識を有していると評価できる。

しかし、語彙の洗練性は頻度という観点のみでは定義することが難しい(Kyle & Crossley, 2015)。近年の研究では、頻度以外の観点として大きく次の4つが挙げられる(Kyle & Crossley, 2016): ① range (母語話者の参照コーパスにおいて、その語彙が収録されているファイルの数; McNamara, Crossley, & McCarthy, 2010), ② n-gram 頻度 (生起するn語の連続の頻度; Crossley et al., 2012), ③ 学術語彙 (Coxhead [2000] や Simpson-Vlach and Ellis [2010] の学術語彙リストに基づく頻度), ④ 心理言語学的要素 (具象性, 親密度, 心象性, 有意味性, Age of Acquisition; Crossley & Skalicky, 2019)。他にも、階層性 (上位語, 下位語) や多義性, 正書法・音韻的近傍語などの観点もある。

語彙の洗練性は多面的であるが、これらは類似した構成概念を測定している可能性がある。例えば、低頻度の語彙は親密度が低く、それゆえ母語話者であっても反応時間が長くなる。反応時間が長い語彙は綴りと音の対応関係が不規則で

あり、スペルが似たような単語である近傍語が多く存在する。さらに「ワイン」や「チーズ」のように日本語化された英単語も多く存在し、英単語としては頻度が低くとも日本人英語学習者はカタカナで知っている場合もある。それゆえ、語彙の洗練性に関する明確な定義は未だ議論が続いている。

2.3 技能統合型ライティングタスク

語彙知識は母語や第二言語のライティング能力において重要な役割を果たしている(Baba, 2009; Kim & Crossley, 2018)。ライティング能力は言語知識と方略的能力に大別され(Weigle, 2002)、発表語彙は前者の下位要素であり、特に洗練性は言語知識の重要な構成概念である(Bachman & Palmer, 1996)。なお、作文の目標設定やトピックに関する背景知識などは後者に分類されている。実際に、高等学校学習指導要領では文法的正確性や語句の選択だけでなく、書く前のブレイン・ストーミングや書いた作品を読み返して推敲するなどの養成も視野に入れている。つまり、作文をする学習者は使用する語彙の洗練性や文法の正確性だけでなく、書く題材に関する背景知識などを十分に活性化するなど、ミクロだけでなくマクロな視点を持つことが求められている。さらに第二言語での作文の場合、言語知識に認知資源を多く配分してしまうため母語で作文するときよりも内容に対してはあまり注意を向けることができない。英作文の指導者や評価者はこれらの点を経験的に理解しているが、明示的に認識することが重要である。

上述のライティング能力の構成概念は技能独立型の活動において当てはまるものであり、技能統合型とは異なる。技能独立型のライティングは例えば、「高校で制服を着用することに賛成か反対か」というトピックのみが与えられ、書き手自身でアイデアを生成しなければならないものである。しかし教室外の英語使用の場面に目を向けると、単に自分の考えや意見を書くだけでなく、Eメールの返信を書いたり、読んだ本の感想を書いたりする。このような技能統合型の活動は、独立型と比べて書き手に要求する能力が異なっている。例えば、独立型では提示された絵に

ついて感想を書く一方、統合型では提示された絵とその絵の作者が込めたエピソードなどの文章を読解(聴解)し、その内容に基づいて作文する。統合型ではリーディング・リスニングにおいて単に理解するためだけではなく、作文することを念頭に置いて情報を選択するため受容技能の重要性が高い(佐藤, 2017)。ただし、技能統合型の活動で必要となる能力は単なる受容技能と発表技能の組み合わせではなく、選択した情報に基づいた発表が求められている(Plakans, 2013)。

複数の技能を統合したとしても、評価の対象となるのは最終目標を設定した最後の技能であるライティングとなる(望月, 2015)。この理由は、生徒一人ひとりのリーディングの内容理解に加え、作文の評価をすることは現実的・実用的ではないためである。例えば、TED-Edの動画の内容をメモしながら最終的に要約をするという活動では、学習者が内容を理解したかどうかを確認する方法は学習者のとったメモとなる。しかし、5分の動画であっても30名のクラス全員のメモとアウトプットの両方を採点するには時間が非常にかかる。それゆえ、最終的なアウトプットを評価することが妥当性・実現可能性の観点から一般的である。

2.3.1 技能統合型の問題点

技能統合タスクを教室内で実施する際には、技能間の統合の強さに着目することが有用である(根岸, 2011)。例えば、Eメールを返信する場合は受信したメール内容を理解する必要があるため、技能が強く結びついていると考えられる。一方、運転免許取得の年齢を引き下げることに賛成意見を読み、それを参考に自分の意見を書くタスクの技能統合は弱いと言える。

技能統合型タスクは英語学習者が将来に出会う可能性の高い言語使用場面を想定しているため真正性が高く、教室内で多く取り入れられると考えられる。しかし、2つの欠点がある(Plakans, 2013)。1つはある程度高い熟達度が必要となることである。技能の統合が強い場合、提示されているソースを十分に理解できなければ何もアウトプットすることができない。つまり受容技能の熟達度が低く、例えばリスニングで提示されるキーワードを聴き逃してしまった場

合にはスピーキング能力がある程度あったとしても産出する内容がないため、熟達度を低く評価されてしまう。幅広い熟達度の受験者がいる英検で技能統合型テストが実施されていない理由の1つに、このデメリットが関係していると言える。もう1つのデメリットは発表技能そのものの診断目的には向かないというものである。例えば、生徒が技能統合型タスクで十分に作文することができなかった場合、その原因はライティングに関わる発表語彙知識が不十分だったためなのか、リーディングに関わる受容語彙知識が足りなかったためなのか曖昧となってしまふ。

一方で技能の統合が弱い場合、情報源となる読解テキストが存在することで学習者の文章構造や内容に関する処理を手助し(Plakans, 2008)、不慣れなトピックに関する情報資源を書き手に提供することから(Gebril, 2009)、独立型よりも多くのアイディアを生み出し自らの主張を裏付けることもできる。技能統合型ライティングに関する研究からは、語彙を始めとする言語的なサポートを得ることができるとするために作文中にソーステキストに読み戻ることが多いことや(Gebril & Plakans, 2016; Plakans & Gebril, 2012)、文法の複雑性(Biber et al., 2016)および語彙の洗練性が高く(Kyle & Crossley, 2016)、結束性の高さがライティングの得点を予測する(Guo, Crossley, & McNamara, 2013)といった特徴が明らかになっている。以下では発表語彙の洗練性に焦点を絞って統合が弱い技能統合型ライティング活動と語彙の洗練性の関係について明らかになっている特徴を述べる。

2.3.2 ライティングタスクと語彙の洗練性

第二言語学習者の英作文において、洗練性の高い難しい語彙を使ったものはライティング熟達度を判断する指標の1つとして機能する(Laufer & Nation, 1995)。この考えには、簡単な英語であっても作文の内容を重視するべきであるという批判があるかもしれないが、学術目的の英語や医療や司法といった特定の目的のための英語を用いる場合には洗練性の高い作文能力が求められる。このような背景から、近年では語彙の洗練性を分析するTAALESを用いた研究が注目されている(Crossley et al., 2019; Kim et al.,

2018)。TAALESはWeb上のサービスであるCoh-Metrixと異なり、使用者のPCにダウンロードする自然言語処理ツールの1つであり、400を超える大量な洗練性指標を高速で提示する。

2.2.3で述べたように語彙の洗練性は多面的であり明確に定義することは難しい。技能独立型ライティングで洗練性が重要な指標であることが述べられていても、それは頻度に基づくものなのか、それとも反応速度のような心理言語学的な要素に基づくものなのかは明らかにされてこなかった。そこでKyle and Crossley (2016)は複数の観点から語彙の洗練性を算出するためにTAALESを用いて、技能独立型と技能統合型ライティングタスクを比較した。その結果、技能独立型テストにおいては、得点の分散の36.8%は語彙の洗練性(BNC Written Rangeで16.7%, BNC Written Bigram Frequency Logarithmで17.8%)によって説明された。このことは、言語知識の下位区分である語彙の洗練性の重要性を支持するものである。説明率の多くを占めていたRangeは、低いほど特定の領域において使用される専門用語であることを意味しており、高い洗練性とみなされる。bigramは2語の連語のことを指し、適切なコロケーションを使用できているかどうかを意味している。連語の知識は語彙のネットワークだけでなく、文法知識とも結びついていることから(Sinclair, 1991)、bigram頻度が高い方が語彙的にも文法的にも豊富な言語知識を有しているとみなされる。

一方で、技能統合型ライティングテストは独立型よりも高い洗練性を示すが、得点の分散の8.3%しか語彙の洗練性によって説明されなかった(Kyle & Crossley, 2016)。この理由について、Kyle and CrossleyはBaba (2009)の結果について言及している。Baba (2009)では受容語彙知識の量と質が技能統合型ライティングの一種である要約課題において中程度の相関があること($r = .400$, $r = .340$)を報告している。Plakans (2013)で示されている通り、技能統合型の活動は単なる「受容 + 発表」というわけではないが、提示されているソーステキストを理解するために受容技能は一定の熟達度が必要となる。さらに、ソーステキストからの言語的な借用が可能であるため、文法的な誤りが少なく正確性が高いこ

とや、アイデアの生成を助けていたため流暢性が高いことも得点には関係していると考えられる。

2.4 Tool for Automatic Analysis of Lexical Sophistication

語彙の洗練性を加味して発表語彙を自動で算出するツールには語彙頻度プロフィールやP_Lexなどがあるが、これらはいずれも頻度の指標のみに基づいている。洗練性が多面的であることを踏まえると、rangeやbigram、心理言語学的な指標などを含めて詳細に分析する必要がある。TAALESはこれらの観点を含め、version 2.2では484の指標(2020年5月現在、version 2.8.1のベータ版で1100を超える指標を算出可能)を提供するツールである。このツールは、①誰でも無料で使用することができ、②Windows, macOS, Linuxといった複数のOS上で動作させることができる。TAALESのインストールはNLP for the Social Sciencesのウェブサイトから行うことができる。

TAALESで使用されている指標は次のとおりである。

(1) 頻度

語彙の洗練性指標として歴史的に用いられてきたのは、コーパスにおける頻度である(e.g., Laufer & Nation, 1995)。TAALESが参照するコーパスはBNC, SUBTREXus, COCA, MRC Psycholinguistics Databaseである。British National Corpus(BNC)は約1億語のイギリス英語を収集した均衡コーパスであり、SUBTREXusは映画字幕のコーパス(話し言葉)である。Corpus of Contemporary American English(COCA)は2020年5月現在、8つのジャンル(spoken, fiction, popular magazines, newspapers, academic texts, TV and Movies subtitle, blogs, web page)からサブコーパスが構成されている約10億語の現代アメリカ英語のコーパスである。なお、TAALES(version 2.2)ではacademic, fiction, magazine, news, spokenの5つから選択する。最後のMRC Psycholinguistics Databaseは心理言語学で用いられる指標だけでなく(e.g., familiarity, meaningfulness, imageability),

シラブル数やストレスパターンなどについても26のオプションから細かく条件が設定されている。

(2) Range

Rangeはコーパスに収録されているテキストの数のことを指す(Gries, 2008)。例えば、readやmakeといった語彙はどのようなテキストでも出現する可能性が高いため、rangeは高い。一方で、extremistやadherenceといった語彙は出現するテキストの範囲が限定的であると考えられる。それゆえ、rangeの値が低い語彙は洗練性が高いと考えられる。

(3) N-gram 頻度

N-gramとは一定の文字数で構成される語の塊のことであり、nは任意の数字を示す。例えば、bigramは2語の塊を意味している。kick the bucketのようなイディオム、on the other handやin my opinionといった連語表現を自由に使うことができる学習者は熟達度が高いと判断することができる(Sivanova-Chantria & Pellicér-Sanchez, 2018)。N-gramのような連語に関する知識は単語同士のつながりを表し、語彙知識の質を指す。N-gramを含む語と語の結びつきの強度は、MIスコア(mutual information score, 相互情報量)という統計値によって計算される。MIスコアは中心語と共起語が互いに相手の情報をどの程度持っているかを表す。つまり「Xと言えばY」「Yと言えばX」のように、片方の語それぞれに、あらかじめ互いの語の情報が取り込まれているような強い結びつきが成立していることが評価される。

(4) 学術語彙

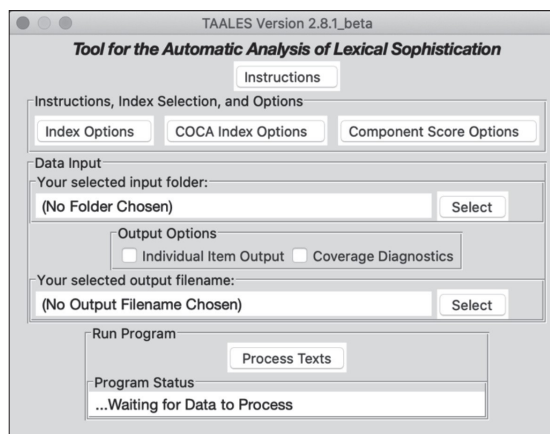
学術場面において使用される語彙は難易度が高く、洗練性が高い。学術語彙をまとめた語彙リストとしてはCoxhead(2000)のAcademic Word ListやSimpson-Vlach and Ellis(2010)のAcademic Formulas Listが挙げられ、TAALESの参照コーパスとして用いられている。

(5) 心理言語学的要素

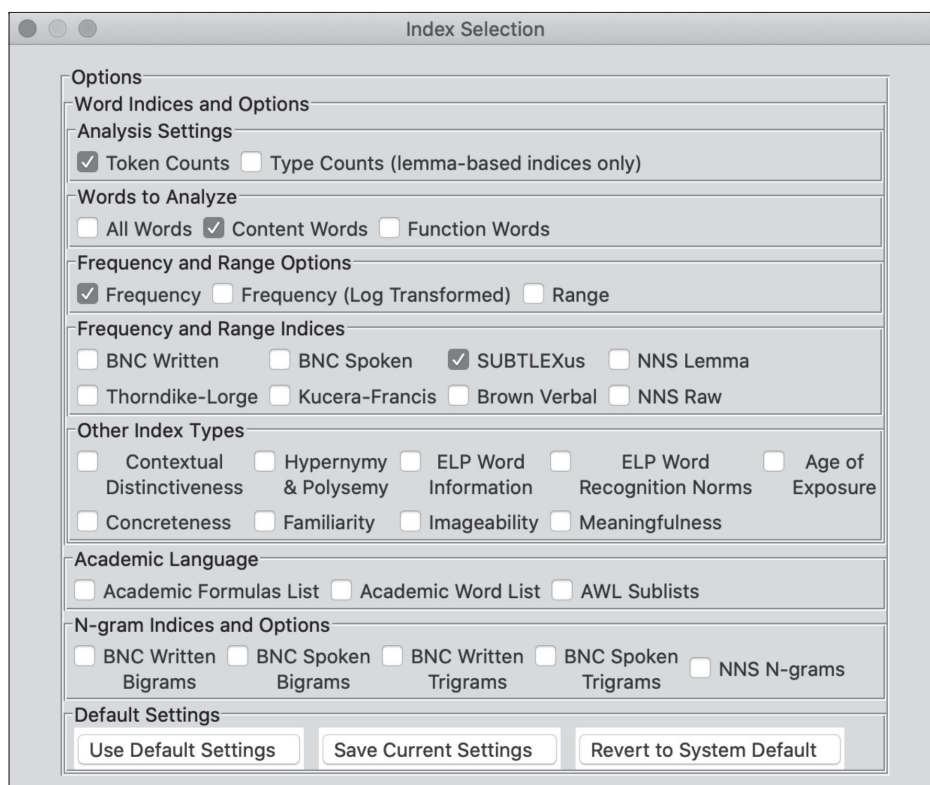
第二言語学習者の語彙知識は長期記憶に格納されており、この語彙に関する記憶体系はメン

タルレキシコンと呼ばれている。このメンタルレキシコンをモデル化する心理言語学の手法として、提示された目標語が非単語か実在語かを判断する反応時間を測定する方法が一般的に用いられている。反応時間に関する研究からは、目標語の特徴によって①頻度効果(反応時間は高頻度語よりも低頻度語の方が長い)、②正書法準拠効果(正書法に準拠した非単語はランダムな文字列の非単語よりも反応時間が長い)、③近傍語サイズ効果などの影響が得られることが明らかになっている。近傍語とは単語を構成する1文字を同一文字位置で別の文字に置き換えた際に得られる単語のことを指す(Andrews, 1989)。例えば、bankの近傍語はrank, tank, sankなどである。近傍語の数が多いほど反応時間は長くなる。TAALESではMRC Psycholinguistics Databaseに基づいてこれらの心理言語学的指標を算出する。Guo et al. (2013)は心理言語学的要素をCoh-Metrixで算出し、第二言語学習者の技能統合型英作文の得点の分散の8.5%を説明することができたことを述べている。

以上のように、TAALESを用いることにより様々な観点から語彙の洗練性を分析することができる。図4および図5にTAALESの使用画面を示す。.txtファイルを用意すればマウス操作で一括で簡単に分析することが可能である。



■図4: TAALES version 2.8.1 (ベータ版)のインターフェイス



■図5: TAALES version 2.8.1 (ベータ版)の指標選択画面

2.5 本研究の目的と構成

発表語彙知識の測定には、語彙サイズテストを用いる直接的な方法と英作文を用いる間接的な方法がある。前者は実際の言語使用を反映していない可能性がある点、後者は頻度の観点のみに基づいている点にそれぞれ問題がある。さらに、近年の技能統合型活動に注目が集まっている中で、技能統合型ライティングタスクと発表語彙との関係は不透明であった。そこで本研究では、発表語彙の間接的なアプローチの問題点をTAALESによって算出される豊富な洗練性指標を用いることによって解決し、技能独立型と統合型のタスクの違いが発表語彙に与える影響を明らかにする。この目的を達成するため、本研究では2つの調査を行う。調査1では技能独立型である英検のモデル英作文における発表語彙の特徴を明らかにする。調査2では、技能独立型および統合型タスクが発表語彙に与える影響を調査する。

3 調査1:英検のモデル英作文の分析

調査1では英検のモデル作文における発表語彙の特徴をTAALESによって数値化し、各受験級で求められる語彙の洗練性がどのように異なるのかを明らかにすることを目的とする。そのため1級から3級までのモデル英作文を用いて、発表語彙に関わる洗練性指標はどのように異なるかを検証する。検証課題(Research Questions: RQs)は次の2点である。

RQ1-1 TAALESによって算出される洗練性指標のうち、どの指標が受験級を予測するか。

RQ1-2 TAALESによって算出される洗練性指標は受験級によってどのように異なるか。

■表1: 調査1で分析したモデル英作文の特徴

受験級	n	平均語数		FKGL		S	
		M	SD	M	SD	M	M
1級	22	226.59	16.59	12.49	1.24	11684.18	2069.93
準1級	13	147.38	2.56	11.07	1.00	5319.24	2518.29
2級	13	92.77	4.02	7.17	0.88	2014.4	632.57
準2級	10	57.1	2.7	4.58	1.24	1870.09	602.06
3級	10	31.5	1.96	4.8	1.60	-	-

(注)S(Kojima & Yamashita, 2014)による分析で妥当な結果を得るためには最低50語が必要であるため、3級の結果は記載していない。

■表2: 調査1の記述統計量

受験級	1級	準1級	2級	準2級	3級
洗練性指標	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)
LD Mean RT	646.90 (8.56)	629.66 (3.92)	614.33 (4.85)	609.28 (6.09)	601.20 (7.81)
Kuperman AoA CW	7.39 (5.66)	6.69 (4.48)	5.60 (3.00)	5.03 (3.93)	4.72 (4.31)
MRC Familiarity CW	569.74 (3.85)	576.22 (5.76)	592.64 (5.43)	596.24 (8.09)	594.84 (10.23)

3.1 方法

3.1.1 英検テキストの収集

TAALESによるテキスト分析を行うため、英検1級から3級までのモデル英作文を収集した。分析対象としたテキストの詳細および、頻度の観点に基づくSの結果を表1にまとめる。この結果は受験級が高くなれば使用されている語彙も低頻度語彙であることを示している。

3.1.2 手順と分析

TAALESによるテキストの分析に先立ち、固有名詞や記号などは除去された。その後、洗練性のどの指標が英検の受験級を予測するのかを知るために語数を除いた指標で決定木分析が行われた。決定木分析は説明変数を用いて段階的にグループ分けをしていく分析方法である。今回の説明変数はTAALESによって算出される洗練性の指標である。決定木のノード(枝分かれ節点)に含まれなかった指標について詳細に検証するために多変量分散分析(multivariate analysis of variance: MANOVA)を行った。MANOVAは複数の独立変数と2つ以上の従属変数を分析することができる手法である。なお、TAALESによって算出される洗練性指標は484という多量の指標が示されるため、決定木のノードおよびMANOVAの従属変数に含まれた指標以外の記述統計については割愛する(表2)。

3.2 結果と考察

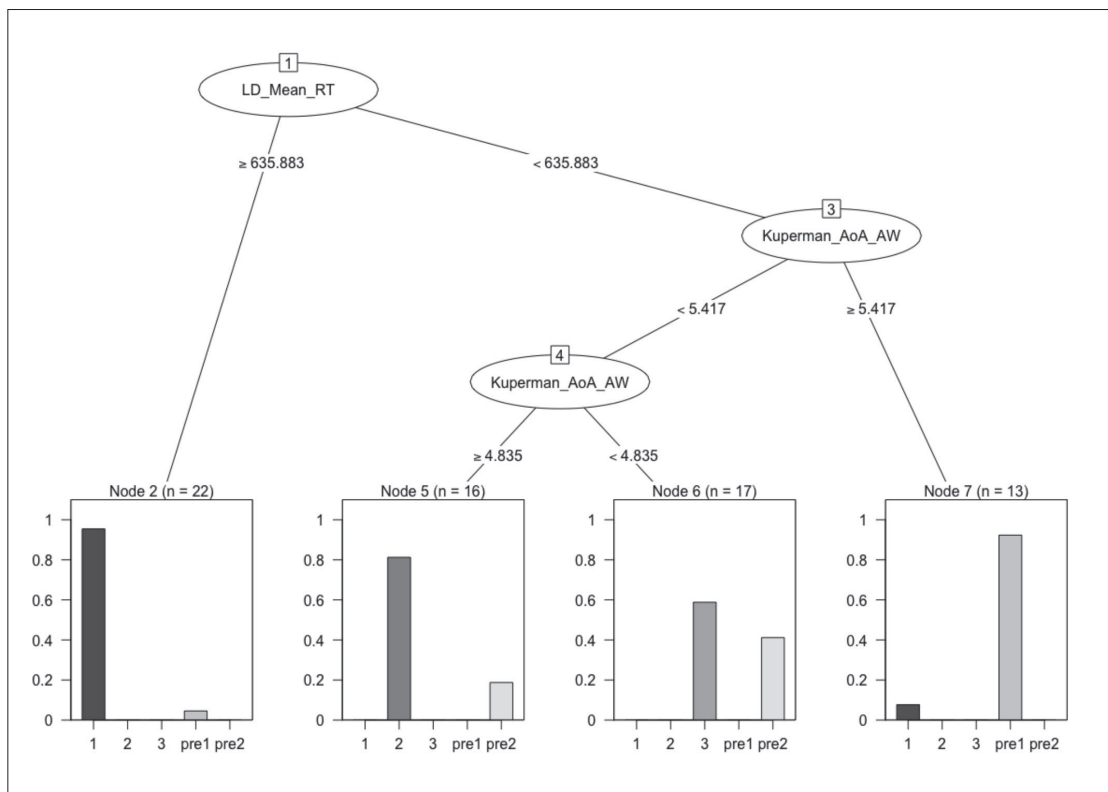
決定木分析の結果を図6に示す。英検の級の分類を最も予測したのはResponse Time(RT: その語が実在語か非単語か判断するために有する反応時間のことを指し、反応時間が長いものは洗練された難易度の高い語彙であると考えられる)であった。英検1級では635.883ミリ秒以上の反応時間を有するような洗練された語彙を使うことができることが求められる。第2および第3のノードはAge of Acquisition(AoA: その語を母語話者が習得する年齢のことを指し、高い年齢で習得する語彙ほど洗練された語彙であると考えられる)であった。つまり英検準1級では5.417歳以降、英検2級では4.835-5.417歳に習得する語彙を使える

ことが求められているという結果が得られた。

これらのノードはいずれも心理言語学の指標であり、母語話者を対象にしたデータベースを基にしている。RTは語彙認知やメンタルレキシコンへのアクセススピードにも関連する指標、AoAはインプットの頻度と関連する指標である。前者の結果から、母語話者であってもその単語が存在するかどうか6秒以上判断に迷う単語を英検1級のモデル作文で使用しているため、要求される発表語彙のレベルは非常に高いと言える。また、アクセススピードには頻度効果が影響する。第2ノードとなっているAoAの指標は、母語話者が生まれてから5歳ごろまでに触れるインプット量とほぼ同義であり、そのインプットには頻度の高いものと低いものが含まれている。それゆえ、第1ノードと第2ノードは頻度という共通点を持つ類似した構成概念であると考えられる。

認知言語学で提唱された用法基盤モデルに基づく、外部から提供されるインプットの頻度が高いものほど習得される傾向にある。この頻度効果に関して、Conklin(2019)は低頻度語彙ほど頻度効果の恩恵が大きいことを述べている。例えば、1回の授業の中で高頻度のreadやgiveを繰り返し登場させるよりも、低頻度語彙であるextremistやadherenceを複数回登場させる方が遭遇回数の効果が大きく見える。しかし、授業時間が限られている日本の教室現場の多くでは、頻度効果が見られるほど大量のインプットは提供されていないのが現実である。むしろ語彙は明示的な方法で指導されることが多い。さらに、明示的・意図的に指導される語彙は言語使用場面を想定し、高頻度で用いられるものが多い。

今回の決定木分析の結果は、洗練性指標が歴史的に頻度に基づくものであったこと(e.g., Laufer & Nation, 1995)とも一致する。この分析で明らかになった点は、頻度指標が単なるコーパスでの出現頻度ではなく、英語母語話者の反応時間や習得年齢と関係することである。つまり、モデル作文の作者が母語話者であった可能性や言語使用場面で高頻度な語彙は母語話者も使用することが関係している可能性も考えられる。RQ1に対して、受験級の予測には反応時間と習得年齢の2つが寄与しているという解答が得られた。



■ 図6: 調査1の決定木分析の結果

RQ2は受験級の間で洗練性指標はどのように異なるのかを扱っている。このRQに解答するため、対応のない1要因の計画で、決定木のノードを独立変数としてMANOVAを実施した。なお、英検1級のみサンプル数が多かったことから、BoxのM検定において異質と判断されないようにサンプルをランダムに削除し、他の受験級とサンプル数を等しくした。記述統計を表2に示す。Pillaiのトレースによる結果に基づくと受験級の洗練性指標は有意に異なっていた、 $F(12, 135) = 7.43, p < .001, \eta^2 = .398$ 。共分散行列の等質性は非有意であったが、Leneneの検定でMRC Familiarity CW(内用語の親密度)が5%水準で有意であり、等分散性が満たされていなかった。それゆえ、結果の解釈は決定木分析のノードに現れた2つに限定する。1変量分散分析による各指標の結果、LD Mean RTで $F(4, 45) = 78.36, p < .001, \eta^2 = .87$ 、Kuperman AoA CWで $F(4, 45) = 64.62, p < .001, \eta^2 = .85$ は0.1%水準で有意であった。それぞれの多重比較を見ると、反応時間は1級と準1級($p < .001$)、準1級と2級(p

$< .001$)で高い受験級で有意に反応時間の長い語彙を使用しており、2級と準2級($p = .558$)、準2級と3級($p = .120$)では有意な差が見られなかった。この傾向は習得年齢についても同様であり、1級と準1級($p = .024$)、準1級と2級($p < .001$)でそれぞれ高い受験級で習得年齢の高い語彙が使用されていたが、2級と準2級($p = .101$)、3級と準2級($p = .648$)では非有意であった。この結果と図6を組み合わせると、母語話者が5歳前後で受容語彙として習得する語彙は英検準1級の発表語彙として求められており、大人の母語話者であっても6秒以上判断に迷う語彙は英検1級の発表語彙の特徴であると言える。しかし、これらの洗練性の指標は英検2級から3級のモデル作文では有意な差が見られなかったことから、英検2級までは文法や語数が重視されていると考えられる。準1級と1級ではいずれも受験級が高いほど洗練された語彙が有意に多く使用されていた。これらの結果はいずれも、技能独立型の英作文における発表語彙は母語話者の心理言語学的な側面が重視されており、1級と準1級および準1級と

2級で洗練性が異なることを示している。つまり、発表語彙の洗練性は熟達度を反映する指標であると主張する先行研究(e.g., Guo et al., 2013; Kyle & Crossley, 2016)と一致する結果が得られた。

語彙の洗練性が有意に高くなるのは英検の準1級以上であった。このことから、英検3級や2級を受験する中高生を指導している教育実践者は語彙の洗練性(言語的複雑性)よりも語数(流暢性)と文法の正確性を重視することにより、限られた授業時間の中で効率的にライティングを指導することができると考えられる。教育現場でのライティング指導では、表現する内容がなければ十分な量の作文をすることができないため、Weigle(2002)の分類における方略的能力がより重視されていると言える。例えば、作文のアイデアを生成するためのプレライティング活動として、ペア・グループでのブレインストーミング(スピーキング)や英語で書かれたウェブを使った情報の収集(リーディング)が行われる。これらの活動はライティングだけでなく他の技能を組み合わせた技能統合型活動である。続く調査2では、このような技能統合型タスクに着目して、その特徴を語彙の洗練性の観点から明らかにする。

4 調査2

4.1 目的

調査2ではTAALESによって算出される洗練性指標に基づいて、技能独立型と統合型のタスクの違いと発表語彙の関係を明らかにする。RQsは以下の2点である。

RQ2-1 ライティングタスクの違いはどの洗練性指標によって予測されるか。

RQ2-2 ライティングタスクの違いによって発表語彙の特徴はどのように異なるのか。

4.2 方法

4.2.1 参加者

調査2に参加した協力者は、東京都内の私立大学、または茨城県の国立大学に通う大学生33名であった。協力者は3ヶ月程度の海外渡航経験はあるものの英語圏の教育機関に長期間留学してはいなかった。またアンケートによるプロフィールの調査の結果、英語熟達度はCEFRでB2からC1レベルであると推定された。

4.2.2 マテリアル

4.2.2.1 技能統合型タスク

統合型では技能間の結びつき(根岸, 2011)が重要な概念となる。調査2で対象とする技能統合型ライティングタスクは特定のトピックと英文が与えられ、書き手はそれをもとに自分の意見を述べるもので、技能間の結びつきは弱いものを対象とする。この理由は、受容技能における熟達度の影響をできるだけ少なくするためである。使用したソーステキストは「自動車の運転免許を取得する年齢を引き上げるべきか」というトピックで賛成意見と反対意見の両方を含む2つである。表3にテキストの特徴を示す。

4.2.2.2 技能独立型タスク

独立型タスクでは、統合型タスクと同様に「自動車の運転免許を取得する年齢を引き上げることを題材に、賛成か反対か意見を述べるタスクを実施した。

■表3: 調査2で使用するソーステキストの特徴

難易度	語数	推定 CEFR-J	ARI	VperSent	AvrDiff	BperA	FKGL	S
賛成	112	B1.2	7.68 (A2.2)	3.43 (C1)	1.57 (B1.2)	0.13 (A2.2)	8.8	1844.19
反対	96	A2.2	5.81 (A1.3)	2.50 (B1.2)	1.58 (B1.2)	0.12 (A2.1)	6.1	3318.00

(注) 読みやすさの指標等は CVLA ver1.1 (Uchida & Negishi, 2018) にて算出。

4.2.3 手順

調査協力者は個別に2つのタスクをそれぞれ約20分で取り組んだ。いずれのタスクもZoomを通してコンピュータ上で行われた。作文中の画面は収録されており、書き手がリアルタイムに行っている推敲作業なども記録された。統合型タスクではソーステキストを協力者が印刷し、手元で参照しながらタスクが行われたが、辞書の使用は禁止された。協力者は最初に独立型に取り組み、その後統合型に取り組んだ。

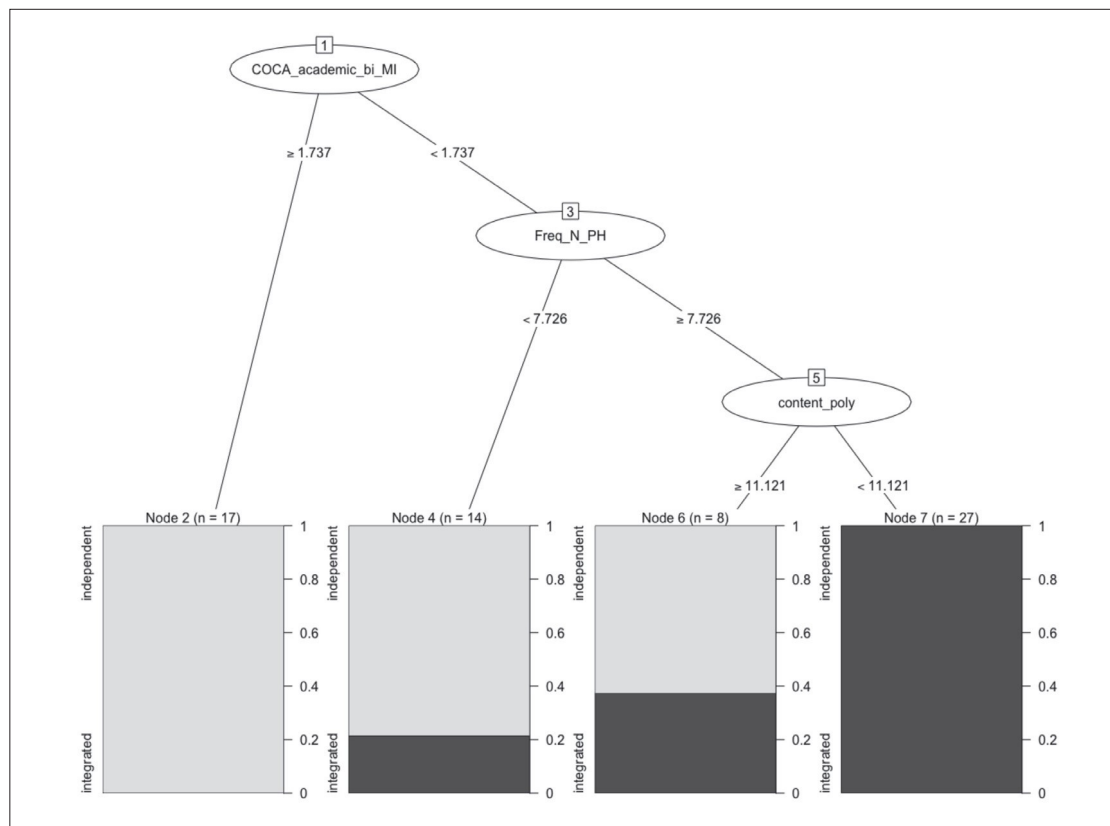
4.2.4. 採点と分析

調査1と同様にTAALESによるテキストの分析に先立ち、固有名詞や記号などは除去された。RQ1の焦点であるタスク分類を予測する洗練性指標を選出するために、説明変数をTAALESによって算出される洗練性の指標として決定木分析が行われた。その後、RQ2に解答するため、AntConcを用いてbigramおよびtrigramを抽出した。

4.3 結果と考察

4.3.1 RQ1:2種類の英作文に含まれる洗練性指標と分類の予測

TAALESによる洗練性指標算出の結果を表4、決定木分析の結果を図7にそれぞれ示す。2つのライティングタスクの分類を最も予測したのはCOCA Academic Bigram MI(アメリカ英語を収録したコーパスであるCOCAの学術サブコーパスにおける、2語の連続のn-gramの結びつきの強度)であった。技能独立型ではMIスコアの高いbigramを産出していると言える。第2のノードは音韻に関する近傍語頻度(近傍語頻度が大きいほど、形式的に類似した語彙が多く洗練性が高い)であった。そして第3のノードには内容語の多義性が含まれた。技能統合型の作文の多くは内容語の語義が11よりも小さい語彙を使用し、技能独立型の作文では多義性の高い語彙を使用していることが示された。



■ 図7: 調査2の決定木分析の結果

■表4: 調査2の記述統計量

タスク	技能統合型		技能独立型	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
洗練性指標				
語数	134.79	32.49	122.94	43.74
COCA Academic Bigram Association Strength (MI)	1.44	0.17	1.66	0.23
COCA Academic Trigram Bigram to Unigram Association Strength (MI ²)	8.08	0.30	8.35	0.33
Phonological Neighborhood Frequency (homophones excluded)	7.89	0.14	7.84	0.18
Polysemy CW	10.14	1.06	9.95	1.95

第1ノードを適切に解釈することは難しいが、技能統合型タスクの方が独立型よりもMIスコアの高い連語を使用していることが示された。この結果はTAALESを用いて発表語彙を分析したKyle and Crossley (2015, 2016)とは異なる。この理由として考えられるのは、認知資源の配分である。独立型タスクでは、未発達の連語よりもむしろ発表語彙として定着し自動化している単語や自由連結を使うことで、新たにアイディアを生成するための認知資源を節約していた。連語の発達は単語と同様に受容・発表の連続体であると捉えると(Meara & Bell, 2001)、未発達の項目を使用するには一定以上の認知資源を消費しなければならないと考えられる。統合型では新たにアイディアを生成することなくソーステキストのアイディアを利用することができるため、豊富な認知資源を形式面に配分することができた。それゆえ、発表語彙として未発達な連語を用いることができたと考えられる。

第2ノードは、音韻に関する近傍語頻度が7.726よりも大きければ統合型、小さければ独立型であることが多いことを示している。つまり、統合型タスクでは洗練性の高い語彙が用いられていたと言える。しかし近傍語頻度と反応時間の関係について検証した先行研究では、母語話者を対象とすることが多く、語彙認知の受容に関わる指標であることに注意が必要である。反応時間の研究では先行刺激が促進もしくは抑制の影響を与えることが示されているが(e.g., Andrews, 1989)、本研究でも統合型タスクのソーステキストで提示された語彙が発表語彙のメンタルレキシコンを活性化させた可能性がある。つまり形

式的に類似した語が刺激として提示されていれば、書記素や音韻に関するメンタルレキシコンを刺激し、ライティングでも使うことができるようになったと考えられる。ただし、作文中にソーステキストを参照することができたため、方略的な語彙使用があった可能性は十分に高く、読解時に提示されていた語彙が作文時にどのように使用されているかは検証の余地がある。

最後のノードである内容語の多義性は、英語語彙に関する大規模データベースであるWordNet (Fellbaum, 1998)に基づいている。独立型では11,121個以上の語義を有している語が多いが、統合型では語義がそれよりも少なく洗練性の高い語彙が多いことを意味している。この第3ノードは、語彙の洗練性本来の意味で用いられる指標である。Laufer and Nation (1995)では洗練性の概念を使用頻度と対応づけて語彙頻度プロフィールを作成しているが「洗練された語彙」が意味するのは、その難しさである。洗練性の概念に基づくと、作文で複雑な表現を使うことができれば優れているとみなすため、平易な言葉で内容を伝えることに主眼を置くコミュニケーション活動では批判されることが多い。古くはルネサンス時代において洗練された語彙は「インク壺言葉」として批判されていた。ルネサンスでは、より専門的で多義性が低いようなラテン語やギリシア語由来の語彙を知識人が用い、多義性の高いゲルマン語由来の語彙は庶民が用いるものであった。インク壺言葉は知識人の難解な言い回しをインクの匂いがすると批判したものであるが、発表語彙の洗練性が高いほど言語熟達度が高いことを示していると言える。

母語話者であっても意見の分かれる語彙の洗

練性であるが、全く不要であるわけではない。例えば、計量的に作文を分析する上では、作文全体の長さに影響を受けてしまったり同じ長さを比較する際にも問題が生じたりする多義性を補完する役割を担っている。さらに、洗練性には頻度の観点だけでなく、調査2で示されたようなMIスコアや近傍語サイズ、内容語の多義性などの様々な指標が存在するため、多角的に発表語彙を捉えることができると言える。

4.3.2 RQ2:n-gramに基づく発表語彙の特徴

RQ2ではタスクの違いによる発表語彙の特徴を扱っているが決定木分析の結果から、単語だけでなく連語について検討することの重要性が示された。チャンクや定型表現と呼ばれる連語はしばしば熟達度の指標の1つとして使用される(e.g., Siyanova-Chanturia, & Pellicer-Sanchez, 2018)。母語話者や熟達度の高い学習者であれば、by and large という表現は個々の単語に分割せずに全体的に捉えて意味を解釈することができるが、初学者は単語を単語に分解して意味を捉えようとし、不透明な表現は正しく理解することができないことが多い。それゆえ、連語を発表語彙として習得していることは熟達度の指標の1つであるとみなすことができる。ここではbigramに関する分析をAntConcによってbigramおよびtrigramを抽出し、それぞれのタスクにおける発表語彙の特性を明らかにする。特徴語分析の結果、上位5項目は独立型と統合型でほとんど同じn-gramが生成された(Appendix 1)。

TAALESで出力されたMIスコアが1.737以上のbigramとそれ以下の2つに分けると、独立型ではshould, I, think, Japan, disagree, beが特徴語として提示された。コンコーダンスに基づくと、shouldは主語にgovernmentを取り、(not) raiseと共起することが多かった。このコロケーションはトピックに由来するものである。2つめの特徴語であるIはthink, agree, disagreeと共起し、thinkはfor this reasonやso, thereforeに後続する一方、賛否について述べる場合には切り出し表現として機能していた。コンコーダンスプロットを見ると、I thinkは中盤から後半にかけて出現しており、まとめの表現として機能していると考えられる。一方、統合型の特徴語と

してはshould, disagree, Japan, we, who, too, can, thinkが確認された。コンコーダンスラインにて、shouldは独立型と同様に、governmentを主語に選択しraiseと共起していたが、weを主語とする例も多く見られた。当該のweはifやsoに導かれ、must, should, canなどの助動詞と共起していた。つまり、読み手を抱き込むようにした書き手の主張が強調されている。そして、tooのクラスターとしてはyoungやdanger、そして不定詞のto driveが見られた。このtoo A to doの構文は独立型でも数件見られたが、ソーステキストで多く使用されていた。

ここではtoo A to doの構文のようにスロットを埋めるプレハブ表現に着目する。スロットを提供するタイプの連語は全てを一から組み立てる必要がなく認知資源を大きく消費しないため、アイデア生成の必要がある独立型タスクで多く見られると予想される。しかし実際には、このタイプの連語は統合型の特徴語として確認された。この理由として、ソーステキストの存在が影響していると考えられる。具体的にはRQ1にて得られた近傍語サイズの結果と同様に、読解によってメンタルレキシコンが活性化されたため、産出に至ったと考えられる。RQ1では単語レベルの可能性であったが、連語においても活性化の可能性が示された。

両者に見られたI think (that)はMIスコアが3を超えており、書き手の意見を表す機能がある。統合型では作文の冒頭と中盤から後半で使用されることが多く、さらに賛成・反対の立場を述べた上でandやbutといった接続詞に後続してI thinkが用いられていた。この理由はソーステキストが存在する場合に、自分の意見とテキストで示されている事実や意見を区別するためである。本調査の協力者はアカデミックライティングなどの授業を受講している大学生であり、剽窃に関する十分な知識を有しており、立場を表す機能表現を発表語彙知識として有していた可能性がある。

4.4 調査2のまとめ

調査2では2種類のライティングタスク(技能独立型と技能統合型)を比較し、TAALESによって

算出される語彙の洗練性という観点から発表語彙の特徴を検証した。決定木分析の結果、タスクの分類を予測する洗練性の指標として、bigramにおけるMIスコア、音韻の近傍語サイズ、内容語の多義性の3つが挙げられた。また発表語彙の中でもbigramに焦点を当てた結果、その使用にはソーステキストの存在が影響している可能性が示された。

5 結論

本研究は2つの調査から構成されていた。調査1は英検のモデル英作文を分析し、2級以下では洗練性に有意な差は見られなかったことから、教育現場では英作文の正確性と流暢性に焦点を当てる必要性を示した。調査2では2種類のタスクが発表語彙に与える影響について検証した。調査1と異なる指標が分類を予測していたが、この理由はソーステキストの存在に起因すると考えられる。技能統合型タスクは真正性が高く、特に中学校での指導では統合型タスクを行うように教科書が設計されており、近年の4技能統合型の指導で広く教育実践現場で取り入れられる。しかし、タスクの遂行には一定以上の熟達度が求められる点(Plakans, 2013)、技能統合の強度(Cumming, 2014; 根岸, 2011)が高い場合には受容技能を活用する段階で問題を抱えてしまう学習者がいる可能性に留意しなければならない。それゆえ、受容の段階ではジグソーやディクトコンボといった協働学習を取り入れ、足場かけを提供しながら受容に関するタスクを乗り越え、発表タスクに問い組むという方法が有効であるだろう。

2つのライティングタスクの分類に最も寄与していた洗練性指標はn-gramの結びつきの強さであった。この結果には調査協力が者がこれまでに受けてきた語彙指導が関係している。本調査の協力は日本の中学校・高校で6年間英語を学習した大学生であった。連語表現の指導は主にディスコースマーカーを用いたリーディングで扱われるが、現在の中高でのライティング教育ではword boxにてスロットを埋める形式の定型表現を足場かけとして提示することが多い。言語使用を通じた付随的な学習法は大量に存在する連

語を学習する上で形式・意味・機能を結びつけるのに効果的であり(Nation, 2013)、受容技能を通じた付随的語彙学習よりも発表技能を通じた付随的学習の方が高い関与負荷を有するため効果的である(Laufer & Hulstijn, 2001)。しかし付随的語彙学習はそもそも効率がよいわけではない(Webb, 2008)。現在の中高での連語の指導が受容において明示的に、発表において暗示的に指導されることが多いため、連語表現が発表語彙として未発達であったと考えられる。そこで連語を指導する際には、高頻度語に焦点を当てて学習することができ、学習の効率の高い意図的な方法と組み合わせることが肝要となる(Nation, 2013)。例えば、大学でのアカデミックライティングにおいて、分野に高頻度な専門用語を含めた連語表現を中心とした明示的な方法で指導すると、学習効率を高めることができると考えられる。

本研究では言語の複雑性を表す語彙の洗練性に着目したが、この指標は英検1級から準1級レベルの高い熟達度を有する大学生や大学院生の学術的な文章を評価する際に特に有効である。学術目的の英語では分野における低頻度の専門用語が繰り返し使用される傾向があるため、頻度の観点だけでは発表語彙を妥当に評価することは難しい。そこで英語のレポートの採点では、TAALESによる洗練性指標を評価観点の1つに用いることができる。その際、心理言語学的な反応時間や習得年齢を参考にすることが効果的であることが本研究から示唆された。

本研究の限界点は大きく3つある。1つは作文のトピックに関する点である。間接的な発表語彙の測定では題材に影響を受けてしまうため複数のトピックを用いるべきであるが、本研究では1つのみであった。今後は複数のトピックを用いた調査が必要である。2つめは調査サンプルに関する点である。本調査の参加者は大学生のみを対象としており英語熟達度が限定的であるため、日本の英語学習者に結果を一般化することは難しい。最後にTAALESの問題点に関して、膨大な数の指標が算出されるため詳細な分析を行える点では有益であるが、使用されているコーパスの特徴や指標の意味が分かりにくいものも多い。ユーザーフレンドリーなツールとしては頻度に基づいたSが挙げられるが、洗練性の様々な側面

を考慮した直感的に分かりやすいツールの開発 | が期待される。

Appendix 1 AntConcによって抽出された調査2のn-gram

ランク	bigram項目(頻度)		trigram項目(頻度)	
	統合型	独立型	統合型	独立型
1	driving age (74)	the driving (64)	the driving age (64)	the driving age (58)
2	the driving (64)	driving age (62)	raise the driving (34)	raise the driving (32)
3	in Japan (50)	in Japan (58)	age in Japan (26)	age in Japan (26)
4	raise the (36)	car accidents (44)	driving age in (22)	driving age in (24)
5	I think (36)	young people (44)	should raise the (16)	cause car accidents (18)

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に厚く御礼申し上げます。特に助言者である寺内一先生には、非常に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、共に学ぶ神村幸蔵さん、小木曾智子さん、佐々木大和さん、前田啓貴さんには多くの示唆とサポートをいただきました。本研究にご協力いただいた皆様に深くお礼申し上げます。

参考文献(*は引用文献)

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 802-814. <https://doi.org/10.1037/0278-7393.15.5.802>.
- Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of second language writing*, 18(3), 191-208. <https://doi.org/10.1016/j.jslw.2009.05.003>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Beglar, D., & Nation, P. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13. https://www.lex tutor.ca/tests/nation_beglar_size_2007.pdf
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639-668. <https://doi.org/10.1093/applin/amu059>
- Fellbaum, C. (1998) *WordNet: An Electronic Lexical Database*. MIT Press.
- Conklin, K. (2019). Processing single-word and multiword items. In S. Webb, (Ed.), *The Routledge Handbook of Vocabulary Studies*, (pp. 174-188). Routledge.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Crossley, S. A., & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of Salsbury, Crossley & McNamara (2011). *Language Teaching*, 52(3), 385-405. <https://doi.org/10.1017/S0261444817000362>
- Cumming, A. (2014). Assessing integrated skills. In A. Kunnan (Ed.), *The companion to language assessment* (vol. I : Abilities, contexts, and learners; Part 2: Assessing abilities), (pp. 216-229). John Wiley & Sons.
- Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Journal of Language Testing*, 26(4), 507-531. <http://dx.doi.org/10.1177/0265532209340188>.
- Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for academic purposes*, 24, 78-88. <https://doi.org/10.1016/j.jeap.2016.10.001>
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2), 228-244. [https://doi.org/10.1016/0749-596X\(90\)90074-A](https://doi.org/10.1016/0749-596X(90)90074-A)
- Granger, S. (2019). Formulaic sequences in learner corpora: Collocations and lexical bundles. In Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.), *Understanding formulaic language a second language acquisition perspective*. (pp. 228-247). Routledge
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both

参考文献 (*は引用文献)

- integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218-238. <https://doi.org/10.1016/j.asw.2013.05.002>
- Hunston, S. (2002). *Corpora in applied linguistics*. Ernst Klett Sprachen.
- Ishikawa, S., Uemura, T., Kaneda, M., Shimizu, S., Sugimori, N., Tono, Y., ... Murata, M. (2003). *Japan Association of College English Teachers (JACET) 8000: JACET list of 8000 basic words*. JACET.
- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63, 87-106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, 102(1), 120-141. <https://doi.org/10.1111/modl.12447>
- Kojima, M., & Yamashita, J. (2014). Reliability of lexical richness measures based on word lists in short second language productions. *System*, 42, 23-33. <https://doi.org/10.1016/j.system.2013.10.019>
- Kyle, K. (2019). Measuring Lexical Richness. In S. Webb, (Ed.), *The Routledge Handbook of Vocabulary Studies*, 454-476. Routledge.
- Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12-24. <https://doi.org/10.1016/j.jslw.2016.10.003>
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarterly*, 49(4), 757-786. <https://doi.org/10.1002/tesq.194>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. <https://doi.org/10.1093/applin/16.3.307>
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51. <https://doi.org/10.1177/026553229901600103>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27, 57-86. doi:10.1177/0741088309351547
- Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5-19. <http://www.lognostics.co.uk/vlibrary/meara&bell2001.pdf>
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I.S.P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12-25. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/1983-Testing-and-teaching.pdf>
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13(2), 111-129. <https://doi.org/10.1016/j.asw.2008.07.001>
- Plakans, L. (2013). Assessment of integrated skills. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. (pp. 205-212). John Wiley & Sons.
- Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18-34. <https://doi.org/10.1016/j.asw.2011.09.002>
- Polio, C., & Friedman, D. A. (2016). *Understanding, evaluating, and conducting second language writing research*. Taylor & Francis.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. <https://doi.org/10.1093/applin/amp058>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (2018). Introduction. Formulaic language: Setting the scene. In Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (Eds.), *Understanding formulaic language a second language acquisition perspective*. (pp. 1-15). Routledge.
- Uchida, S. and M. Negishi (2018) Assigning CEFR-J levels to English texts based on textual features. In Y. Tono and H. Isahara (eds.) *Proceedings of the 4th Asia Pacific Corpus Linguistics Conference (APCLC 2018)*, (pp. 463-467). <http://hdl.handle.net/2324/2244112>
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20, 232-245. <http://nflrc.hawaii.edu/rfl/October2008/webb/webb.pdf>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics*, 168(1), 33-69. <https://doi.org/10.1075/itl.168.1.02web>
- Weigle, S. C. (2002). *Assessing writing*. Ernst Klett Sprachen.
- 相澤一美・望月正道. (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』. 大修館書店
- 根岸雅史 (2011). 「技能統合の評価をどうするか」. 『英語教育』第60巻2号 (5月号), 29-31. 大修館書店.
- 佐藤敬典. (2018). 「テスト作成に必要な理論を学ぼう」. 小泉利恵, 印南洋, 深澤真. (編著). 『事例でわかる 英語テスト作成ガイド』. (pp. 76-79). 大修館書店.
- 望月昭彦. (2015). 「ライティングと他技能の統合型の活動の評価」. 望月昭彦, 深澤真, 印南洋, 小泉利恵. (編著). 『英語4技能評価の理論と実践—Can-Do・観点別評価から技能統合的活動の評価まで』. (pp. 92-101). 大修館書店.