

Latent Semantic Analysis (LSA) による 空所補充型読解テストの解明

一文レベルの意味的関連度を観点として一

茨城県／筑波大学大学院在籍 名畑目 真吾

概要

本研究は、概念間の意味的関連度を算出する LSA を用いて、空所補充型の英文読解問題の特性を解明しようとしたものである。2つの調査を行い、空所が含まれる文の意味的関連度と項目の難しさとのかかわりを検証した。

調査1では、英検の複数の受験級で用いられている問題を対象とした分析を行った。その結果、上位の級では下位の級と比べて、空所を含む文とその前後の文、および異なるパラグラフに含まれる文との意味的関連度が有意に低いことが示された。

調査2では、空所補充問題を学習者に解答させ、そのデータから算出される項目の難しさと空所が含まれる文の意味的関連度のかかわりを検証した。その結果、熟達度の低い学習者では、空所が含まれる文と正答となる語の意味的関連度が高い場合にその項目の正答を導きやすくなる可能性が指摘された。

本研究における2つの調査から、空所補充読解問題の難易度には、空所が含まれる文の意味的関連度がさまざまな形でかかわることが示唆された。

1 はじめに

テキスト内の情報のつながりを示す指標の1つに、意味的関連度がある。これは、テキストに含まれるある語や文によって伝達される意味概念が、テキストのその他の語や文が持つ意味概念とどの程度関連しているかを表すものである。近年の第1言語読解研究においては、テキスト情報間の意味的関連度は Latent Semantic Analysis (LSA) と呼ばれる手法により数値として算出され、その値が読み手

の読解プロセスやテキスト理解へ影響を与えることが明らかになっている。しかし、第2言語読解分野では、LSA を用いた研究は極めて少なく、いまだこの手法は広く認知されていない。

本研究では、この LSA を用いて算出されるテキスト情報間の意味的関連度の観点から、空所補充型の英文読解テストの特性を解明することを試みる。多肢選択式テストや筆記再生テストなどさまざまなテスト形式がある中で空所補充テストを扱う理由は、このテストではテキストの一部を削除して項目が作成されるため、削除される部分やそれが含まれる文が持つ意味的関連度が項目の難しさとかかわる可能性が考えられるためである。これまで、空所補充読解テストについては削除される単語の品詞などさまざまな観点で研究がなされてきたが(2.1.2節参照)、本研究ではテキスト情報間の意味的関連度という新たなアプローチで検証を行う。

本稿の構成は以下に述べるとおりである。まず、空所補充読解テスト、および LSA に関する先行研究を概観する。そしてそれらの知見に基づき、1つの予備調査と2つの調査を行う。予備調査では、複数の英語能力試験に含まれる空所補充読解問題の中から、LSA に適したものを選定することを目的とする。そして、予備調査で選定された問題について調査1で実際に LSA を行い、空所が含まれる文の意味的関連度について問題間の違いを検証する。調査2では、英語学習者を対象とした実験を通じて、LSA によって算出される空所が含まれる文の意味的関連度の値と項目の難しさとのかかわりについて検証する。

2 先行研究

2.1 空所補充読解テスト

2.1.1 空所補充読解テストの概要

あるテキスト内に複数の空欄を設け、その空欄に当てはまる語を埋める形式の問題は、大きくクローズテスト (cloze test) と呼ばれる形式に分類される。クローズテストは Wilson L. Taylor によって 1953年に開発されたテストであり (Taylor, 1953)、ゲシュタルト心理学と言語の余剰性にその理論的根拠を置く (川畑, 2001)。

クローズテストは外国語教育に導入された当初、学習者の総合的な言語運用能力を測定する有用なテストとして支持された (Oller & Conrad, 1971)。しかし、その後そのような主張に対する反論もあり (Alderson, 1979, 1980)、これまでにその妥当性の改善を目的としてクローズテストのさまざまな変種が提案されてきた。その代表的なものとして、「意図的クローズテスト」と呼ばれるものがある。当初のクローズテストがテキスト中の語を一定間隔 (5~10語ごと) で削除する形式であるのに対し (機械的クローズテストと呼ばれる)、意図的クローズテストでは作成者によってどの語が削除されるのかが決定される。実際の英語能力試験で見られる形式のほとんどは意図的クローズテストであり、本研究でもこの意図的クローズテストを対象とした調査を行う。Alderson (2000) では、クローズテストという名称は冒頭に述べた機械的削除のテストにのみ用いられるべきであり、意図的削除を用いたテストは機械的削除のテストとは異なる能力を測定しているため、クローズテストとして言及されるべきではない (p.208) と主張され、意図的削除を用いたテストを「空所補充テスト」(gap-filling test) として言及している。本稿ではこの主張に基づき、意図的削除を用いたクローズテストを「空所補充テスト」と呼ぶこととする。空所補充テストはさらに、自由記述式と多肢選択式に分類される。

2.1.2 空所補充読解テストの特徴

空所補充テストについては、これまでに多くの研究がなされてきた。空所補充テストをクローズテストと比較した Bachman (1985) は、これら2つのテストではその信頼性、他の読解テストとの相関、

受験者群の弁別において大きな差異はないが、後者の方が前者よりも難易度が高いことを示した。Bachman はこれら2つのテスト形式の難易度の違いについて、それぞれのテストで削除される語、つまりテスト項目の特徴の違いが影響していることを指摘している。

クローズテストと空所補充テスト (自由記述式・多肢選択式) に含まれる項目について、Abraham and Chapelle (1992) ではその難易度に影響し得る項目特徴が詳細に検証されている。この研究では、英語を第2言語とする大学生を対象とした実験を行い、クローズテストと空所補充テストではその難易度に影響する項目の特徴が異なることを示した。本研究の焦点である空所補充テストについては、空所と解答の手がかりとなる情報の距離 (解答に必要な文脈の量)、正答となる語の品詞や長さなどが項目難易度にかかわる要因として指摘されている (表1参照)。

■ 表1: 空所補充読解テストの項目難易度に影響を与える要因 (Abraham & Chapelle, 1992, p.470 より一部改変)

	易しい	難しい
空所と解答の手がかり情報の位置	近い ↔	遠い
正答となる語が機能語か内容語か	機能語 ↔	内容語
正答となる語の長さ	短い ↔	長い
正答として許容される語の数	2つ以上 ↔	1つ
正答となる語の語形変化が可能か (活用形の数)	1つ ↔	2つ以上

(注) 多肢選択式では機能語/内容語要因のみが影響していた。

また、Kobayashi (2002) では Abraham and Chapelle の枠組みをもとに、さらに「項目が測定する知識」などの要因を加えて項目難易度とのかかわりを検証している。この研究では、項目が測定する知識を成句の知識、統語的知識、意味的知識、談話的知識、一般的知識の5つに分類している。この研究ではクローズテストのみが検証対象となっているが、項目が測定する知識の枠組みは、本研究において空所補充テストの選定を行う際に有用となるため (詳細は2.3節で後述する)、ここで言及しておく。

さらに、空所補充テストについてはその解答プロセスに関する研究もこれまでいくつか行われてい

る。Yoshida (1997) では、日本人大学生を対象に、空所補充読解テストの解答プロセスを発話プロトコル法^(注1)により検証した。その結果、難易度の低い項目は空所の前後の情報や文法的な制約を利用した「局所的処理」によって解答されるのに対し、難易度の高い項目はテキストの主題など、より広い文脈からの情報を用いた「大局的処理」によって解答されることを明らかにした。このような「項目の難易度」と「解答の手がかりとなる情報の位置（文脈の量）」の関係は、上記の表1で示した Abraham and Chapelle (1992) の結果と整合するものである。さらに、同じく日本人大学生を対象とした Yamashita (2003) では、空所補充テストの解答プロセスについて、学習者の英語熟達度を観点とした検証が行われている。この研究では、熟達度の高い受験者は熟達度の低い受験者と比べて、より広い文脈からの情報を解答に利用する傾向があることが明らかにされた。

以上に述べたように、空所補充読解テストについてはこれまでその項目や解答プロセスの特徴に関してさまざまな研究が行われているが、これらの研究の結果に共通する知見として、空所補充読解テストでは「解答に必要な文脈の量」がその項目難易度にかかわる重要な要因の1つであるということが挙げられる。本研究では、以下で述べる LSA によって、空所が含まれる文の意味的関連度を算出する際にこの特徴を考慮する。

2.2 Latent Semantic Analysis (LSA)

2.2.1 LSA とは

LSA とは、大規模な言語コーパスに基づいた計算により、単語や文が表す概念間の意味的類似性を決定する数学的・統計的手法である。LSA は心理学領域で頻繁に用いられる手法であり、近年の国外におけるテキスト読解研究では、分析の中に LSA を含めている研究が数多く存在する (e.g., Graesser & McNamara, 2011; Todaro, Millis, & Dandotkar, 2010; Trabasso & Wiley, 2009; Wolfe, Magliano, & Larsen, 2005; Zwaan & Madden, 2004)。しかしながら、国内研究ではいまだ LSA を用いた研究は少なく、海外の関連分野ほどには LSA が浸透していないことが示唆されている (猪原・楠見, 2011)。したがって、本研究において日本の英語教育研究分野へ LSA を応用することは、意欲的・探索的な試みで

あると言える。

LSA においては、分析対象となるテキストセグメント（語や文）がコーパスによって形成される高次元の意味空間内で数値列（ベクトル）として表象され、それらの間のコサインが算出される^(注2)。このコサインの値が概念間の距離、つまり意味的関連度を表す。この意味的関連度は -1.0 から $+1.0$ までの値を取り、1 に近づくほど概念間の意味的な関連性が高いと解釈される。理論的な説明のみでは理解し難い内容であるため、以下の具体例をもとに LSA によってどのような分析が可能なのかを把握していただきたい。

Dennis (2007) では、LSA が実行可能なコロラド大学のウェブ (<http://lsa.colorado.edu/>) 上の複数のアプリケーション (Near Neighbors, Matrix Comparison, Sentence Comparison, One-To-Many Comparison, Pairwise Comparison) を紹介し、それらによってどのような分析が可能かをわかりやすく述べている。ここでは Dennis で説明されている事例のうち、LSA の基本的な説明に適したものとして、Near Neighbors, Sentence Comparison を用いた例を取り上げる。

まず、Near Neighbors を使った分析では、ターゲットとなる単語と意味的関連度の高い語のリストの提示が行われる。表2は dog をターゲットとしたときの例であり、ターゲットと意味的関連度が高い語として10の単語が提示されている^(注3)。LSA では、ターゲットの類義関係や上位語・下位語などの階層関係だけではなく、ターゲットと同じ文脈で共起しやすい語という観点から分析が行われる。例え

■ 表2: dog と意味的関連度の高い10語 (Dennis, 2007, p.61 より引用)

意味的関連度の値	語
.99	dog
.86	barked
.86	dogs
.81	wagging
.80	collie
.79	leash
.74	barking
.74	lassie
.72	kennel
.71	wag

ば、表2ではターゲットの下位語に当たる collie に加え、barked, leash, wag などターゲットと同じ文脈で共起しやすいと考えられる語が高い意味的関連度を持つと評価されている。

上記では単語対単語の意味的関連度について具体例を挙げたが、LSA では単語よりも大きい単位での意味的関連の分析が可能である。ここでは、Sentence Comparison の例を取り上げてみる。表3では、あるテキストに含まれる4文について、各文とその前後の文の意味的関連度が算出されている。例えば、文1と文2ではそれらの意味的関連度が.24と評価されていることがわかるであろう。

■表3: 文間の意味的関連度 (Dennis, 2007, p.64 より引用)

意味的関連度の値	文
.24	1: Incy wincy spider climbed up the water spout.
.91	2: Down came the rain and washed the spider out.
.20	3: Out came the sunshine and dried up all the rain.
	4: So incy wincy spider climbed up the spout again.

その他の分析ツールについて、Matrix comparison では複数の分析対象について最大ペア数の比較、One-To-Many Comparison ではある分析対象とその他複数の対象との比較、Pairwise Comparison では分析対象のペアでの順次比較 (e.g., 文1と文2, 文3と文4...) などの分析が可能である。すべての分析ツールで共通するのは、分析の際に意味空間を構築するコーパスや要因数の設定が必要となることである。ここで述べる LSA の説明はほんの概要程度であり、実際に用いる際はさらに重みづけの選択などを考慮する必要がある。本稿では紙面の都合上これらの説明は割愛するが、詳しくは Dennis (2007) が収録されている Handbook of Latent Semantic Analysis や中村 (2008) をご覧いただきたい。

本研究では、LSA を用いてテキスト情報間の意味的な類似性を推定し、それが空所補充読解問題の

項目の性質とどのようにかわるかを検証する^(注4)。以下では、本研究と関連する先行研究として、LSA と文章理解の関連を扱った研究をレビューする。

2.2.2 LSA と文章読解

LSA と文章読解のかかわりについては、これまで英語母語話者を対象とした研究が多く行われており、LSA によって算出される概念間の意味的関連度がテキストの特徴や読み手のテキスト処理にかかわることが明らかになっている。Foltz, Kintsch, and Landauer (1998) では、LSA は2つの文がどれだけ意味的に関連したトピックについて議論しているかを測定できるため、文章の一貫性を評価する指標となり得ると述べられている。実際にこの研究では、一貫性が高くなるように作成されたテキストほど隣接する文間の意味的関連度が高いという結果が示されている。

また、Wolfe et al. (2005) では因果的な関連度と LSA による意味的関連度の程度が異なる2文のペアを作成し^(注5)、この2つの関連度が文の読解時間と記憶に与える影響を検証した。その結果、2文間の高い意味的関連度は文情報の記憶に対して促進効果があること、因果的関連度より影響は弱いものの、文の読解時間に対してもいくらかの促進効果があることが示された。さらに、この研究ではより長い文章を用いて、局所的・大局的な意味的関連度を算出している。局所的な意味的関連度は「テキスト内のある文とその直前の文の意味的関連度」、大局的な意味的関連度は「ある文とその直前の文を除いたそれまでのすべての文の意味的関連度」として算出され、大局的な意味的関連度が文の適合度判断 (文の内容がそれまで読んだ内容とどの程度適合しているかの判断) に影響する可能性が示唆されている。

Today et al. (2010) では Wolfe et al. (2005) と同じ実験材料を用いて、意味的関連度と因果的関連度が2文間の一貫性の判断に与える影響について検証を行った。実験の結果、因果的関連と意味的関連の両方が読み手の一貫性判断に影響するが、LSA によって算出された意味的関連度は読解能力の低い読み手の一貫性判断に影響を与えやすいことが示された。

上記に述べた研究結果をまとめると、LSA によって測定される概念間の意味的な関連性が高いほど、読み手はそれらの間につながりがあると判断し、ま

た、それらの処理や理解が促されることがこれまでの研究によって明らかにされている。そのため、テキストの一部を削除して作成される空所補充問題の解答（難易度）は、空所を含む文と削除される語がどれだけ意味的に関連したものであるか、空所を含む文とその前後の文がどの程度意味的に関連したものであるかなどによって影響される可能性が考えられる。また、その影響は英語能力のような学習者の要因によって異なる可能性も考えられる。

2.2.3 LSA の利点と限界点

ここまで LSA の概要、および関連する先行研究について述べたが、LSA の利点としてどのようなことが挙げられるのだろうか。まず、その利点として、概念間の意味的関連度を大規模なコーパスに基づいて自動的に算出できる点がある。このことにより、手作業による算出の手間がなくなるだけでなく、意味的な関連度を算出できる単語ペアの数が飛躍的に増加する（猪原・楠見, 2011）。また、LSA は比較対象に含まれる語の重複や類義・階層関係を反映しているだけでなく、同じ文脈で、もしくはある単語と共起しやすい語という観点から意味的関連度を測定できるため、より精緻な関連度の表示が可能となる。そして、語や文などさまざまな単位を分析対象とできる点も利点である。このことにより、単なる単語対単語のような小さいレベルの意味的関連度だけでなく、文対文、パラグラフ対パラグラフのような相対的に大きな概念間での意味的関連度や、単語対文、文対パラグラフのような異なる単位間での意味的関連度を算出できる。

一方、LSA の限界点として知られているのは、語順のような統語情報、談話標識やテキストの構造・修辭的な要因は分析に考慮されていないことである（Foltz, 2007）。そのため、Therefore などで行なわれた 2 文もそれらの間の意味的関連度が低く評価されてしまうこともある。より正確な分析には上述の要因が考慮される必要があるが、これまでの先行研究の概観から示されるように、LSA によって算出される概念間の意味的関連度は読み手のテキスト理解プロセスとかがわりがあることは明白である。

2.3 本研究と先行研究のつながり

空所補充読解問題に関する先行研究では、解答に

必要な文脈量や正答の語の長さなどさまざまな項目の特徴を観点とした検証がなされてきた（2.1.2 節参照）。しかしながら、空所を含む文が正答となる語やテキスト内のその他の文とどの程度意味的に関連しているかという観点では、これまで検証が行われてこなかった。本研究では新たな項目特徴の観点として、LSA によって算出される空所を含む文の意味的関連度を扱う。その上で、空所を含む文のさまざまな意味的関連度と項目の難しさとのかかわりを検証する。

まず、予備調査として Kobayashi (2002) の「項目によって測定される知識」の枠組み（2.1.2 節参照）に基づき、複数の英語能力テストで扱われている空所補充読解問題の項目分類を行う。2.2.3 節で述べたように、LSA は概念間の意味的関連度を算出する一方で、その値には統語情報や談話標識などの要素は加味されていない。そのため、LSA による意味的関連度の値と項目難易度のかかわりを検証するには、項目が統語的知識や談話的知識などではなく、意味的知識を測定する問題を対象とするのが好ましい。よって、予備調査では、意味的知識を問う項目を多く含む空所補充読解問題を選定することを目的とする。

そして、予備調査で選定された空所補充問題について、調査 1 で LSA による空所を含む文の意味的関連度の算出を行う。空所補充問題を扱った先行研究では解答に必要な文脈量が項目の難易度にかかわることが指摘されていることから（2.1.2 節参照）、調査 1 では局所的なレベルから大局的なレベルまでの意味的関連度（2.2.2 節参照）を算出し、問題間での比較検証を行う。

最後に、調査 2 では実際に学習者に空所補充問題を解答させ、そのデータから算出される項目の難易度と空所が含まれる文の局所的・大局的な意味的関連度のかかわりについて、学習者の英語能力も要因に含めて検証する。

3 予備調査

3.1 目的

予備調査の目的は、実際の英語能力テストで用いられている空所補充読解問題について項目特徴の分類を行い、調査 1 で行う LSA に適したマテリアル

を選出することである。なお、本調査では調査対象を明確にするために、「説明文・物語文などを扱い、文章内に設けられた複数の空所に適切な語や句を補充する問題」を空所補充読解問題として定義した。そのため、(a) 語彙・文法テストに分類されるような単文での空所補充問題、(b) 会話文を用いた空所補充問題、(c) テキストの要約文を用いた空所補充問題、(d) ある文をテキスト内に挿入する位置として適切なものを選択する問題などは本研究の調査対象から除外した。

3.2 方法

3.2.1 マテリアル

まず、英検、TOEIC、センター試験、TOEFL、IELTS の5つのテスト^(注6)を調査し、そのうち上記で定義した空所補充読解問題が扱われていた英検、TOEIC、センター試験の3つを分析対象とした。具体的には、これら3つのテストから表4に挙げるテキスト・問題項目数を分析対象とした。なお、英検では3級から5級までに本調査で対象とした空所補充読解問題に分類される問題はなかったため、準2級から1級に含まれていた問題を分析対象とした^(注7)。分析対象とした問題に用いられていたテキストの語数とリーダビリティの平均を表5に示す。

■ 表4：予備調査で分析対象としたテスト

テスト	テキスト数	項目数	
英検	24	75	2011年度第1回から2010年度第2回までの3回分/準2級から1級まで
TOEIC	24	72	TOEIC 新公式問題集 vol.2-4 (ETS, 2007, 2008, 2009)
センター試験	9	23	2011年～2003年までの過去問題

■ 表5：各テストに用いられていたテキストの語数とリーダビリティ

	語数		FKGL		FRE	
	M	SD	M	SD	M	SD
英検	281.38	66.45	10.20	2.65	53.57	14.48
TOEIC	111.75	20.15	10.44	2.75	49.33	15.75
センター試験	210.67	119.32	9.30	1.62	60.19	8.79

(注) FKGL = Flesch-Kincaid Grade Level ; FRE = Flesch Reading Ease.

表4、5に示す空所補充読解問題はすべて多肢選択形式であった。1つのテキスト当たりの平均語数は英検で最も多く、TOEICではその半分以下だった。しかしながら、リーダビリティによって評定されるテキストの難易度については、3つのテスト間で大きな違いは見られなかった。テキストタイプについては、英検では物語文(準2級のみ)、説明文、評論文が使用されており、センター試験では主に説明文、TOEICでは手紙・メール形式の文章が多用されているものの、新聞記事や求人広告なども含めた実用的な文章が用いられていた。

3.2.2 分類手順

上記の表4に示した項目について、Kobayashi (2002)における「項目が測定する知識」の枠組みを援用した分類を行う。予備調査の目的はLSAに適したマテリアルを選出することであるため、5つの知識(成句の知識、統語的知識、意味的知識、談話的知識、一般的知識)のうち、「意味的知識」に分類される項目が多く含まれるテストを選定する。なお、実際の能力テストではテキスト外の知識である「一般的知識」で解答されるような問題は含まれていないと想定されるため、本調査では一般的知識は扱わないこととした。その他の4つの知識の分類基準は、本調査では表6のように設定された^(注8)。

■ 表6：項目が測定する知識の分類基準

成句の知識	イディオムやコロケーションに関する知識が問われる場合
統語的知識	機能語が削除項目になる場合や、動詞の活用などが問われる場合
意味的知識	内容語が削除項目になる場合など、文脈に沿った意味が問われる場合
談話的知識	接続詞・談話標識が削除項目になる場合など、一貫性・結束性の知識が問われる場合

各テストにつき調査者を含めた2名ずつで分類を行った。評価者間の一致度は、英検では98.67%、TOEICでは88.9%、センター試験では100%であった。Kobayashi (2002)では1つの項目によって複数の知識が測定される場合を考慮し、すべての項目について分析を行うのではなく、評価者間の一致度が高かったもの(66.7%以上/6人中4人)のみを分析の対象としている。本調査においてもこの点を

考慮し、2人の評価者の判断が一致した項目のみを分類対象とした。

3.3 結果と考察

各テストにおける項目が測定する知識の分類結果を表7に示す。

■表7: 各テストにおける項目が測定する知識の分類

	英検 (k = 74)	TOEIC (k = 64)	センター試験 (k = 23)
(1) 成句の知識	0%	1.56%	0%
(2) 統語的知識	0%	42.19%	0%
(3) 意味的知識	95.95%	43.75%	65.22%
(4) 談話的知識	4.05%	12.50%	34.78%

(注) 分類対象となった項目は、評価者間で判断が一致したもののみ。k = 項目の数。

英検の空所補充読解問題では、いくらか談話的知識を問う項目が含まれていたものの、ほとんどの項目が文脈に合致する意味として適切な内容語、もしくは複数の内容語から成る句を選択するものであり、それらは意味的知識を問う項目として分類された。センター試験においても半分以上は文脈に合致する意味を問うものであり、その他は談話標識などが選択肢となり談話的な知識を問うものであった^(注9)。TOEICでは文脈上適切な内容語を選択する項目が多かったことに加え、動詞の適切な活用を問う項目なども多く含まれていたことから、項目が測定する知識として統語的知識、意味的知識の2つが8割以上を占めていた。

以上の結果から、意味的知識を問う項目を多く含み、全体の項目数も多い英検の空所補充読解問題をLSAに適した問題として調査1で扱うこととする。調査1では、予備調査で用いた24のテキスト(75項目)を分析対象とする。

4 調査 1

4.1 目的

調査1では、英検の空所補充読解問題のテキストについて、空所が含まれる文の意味的関連度をLSAにより算出する。意味的関連度については、局所的なレベルから大局的なレベルまで複数の値を算出する。また、準2級から1級までの複数の受験級の問題を扱い、受験級を問題の難易度の指標とする。受験級(問題の難易度)が上がるにつれて、空所が含まれる文の局所的・大局的なレベルの意味的関連度がどのように変化するかを検証することを調査1の目的とする。本調査におけるリサーチクエスション(RQ)は以下のとおりである。

RQ1: 空所が含まれる文の意味的関連度の値は、問題の難易度(受験級)によって異なるか。

4.2 方法

分析対象となるのは、予備調査で選定された英検準2級~1級までの空所補充読解問題24題(75項目)である。これらの問題で用いられていたテキストの平均語数とリーダビリティを、受験級ごとに表8に示す。

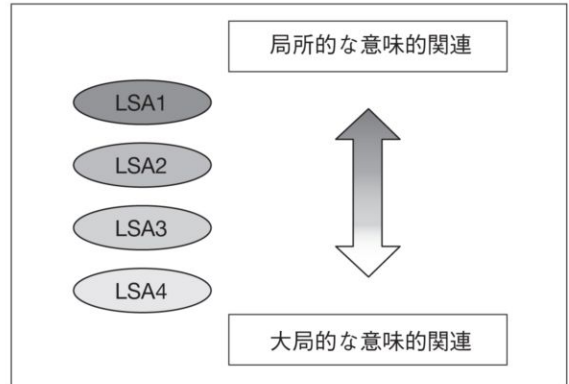
本調査では、Wolfe et al. (2005)の局所的・大局的な意味的関連度の算出方法を参考にしながら、各項目について「空所が含まれる文」を中心とした異なる4つの意味的関連度(LSA類似度)を算出した(以下、LSA1, LSA2, LSA3, LSA4とする)。最も局所的なレベルの意味的関連度として、「空所が含まれる文」と「正答となる語」の意味的関連度をLSA1とした。LSA2は「空所が含まれる文」とその「前後の文」の意味的関連度の平均とした。LSA3として、「空所が含まれる文」とそれが含ま

■表8: 調査1で用いたテキストの語数とリーダビリティ

	準2級 (n = 6)		2級 (n = 6)		準1級 (n = 6)		1級 (n = 6)	
	M	SD	M	SD	M	SD	M	SD
語数	204.33	63.56	318.50	5.68	249.83	7.28	352.83	9.54
FKGL	6.90	0.95	9.02	1.21	12.35	1.27	12.53	1.26
FRE	71.45	5.83	59.47	6.86	40.30	9.54	43.05	4.82

(注) n = テキストの数; FKGL = Flesch-Kincaid Grade Level; FRE = Flesch Reading Ease.

れる「パラグラフ内の2文以上離れた（＝前後の文を除いた）文」との意味的関連度の平均を算出した。そして、最も大局的なレベルの意味的関連度として、「空所が含まれる文」とそれが含まれるパラグラフ以外の「異なるパラグラフ内の文」の意味的関連度の平均を LSA4 として算出した。これら4つの LSA 類似度の関係を図示したものが図1、4つの LSA 類似度の分析対象を視覚的に示したものが図2である。なお、LSA1の「空所が含まれる文」は正答となる語や句を含まず、LSA2～LSA4の「空所が含まれる文」は正答となる語や句を含んだ状態で分析を行った。



▶ 図1：4つの LSA 類似度の関係

LSA1

Red Telephone Boxes-

At the beginning of the 20th century, few people in Britain had a telephone in their home. The only way most people could make calls was to use a public telephone. At first, most public telephones were in stores, and people had to pay the store clerk to use them. **As time passed, though, public telephone boxes began to be built (outside).** These allowed people to make calls on the street without being heard by other people. They also protected people from the rain. They were usually painted bright red. For many people, these red telephone boxes became a symbol of British life.

One problem with telephone boxes was that people sometimes damaged them by breaking their windows, writing on them, or trying to steal money from the telephones. In the 1980s and 1990s, it became more and more expensive to keep Britain's red telephone boxes in good condition. The telephone company began to replace them with ones that were easier to look after. However, many people did not like the (design) of these new telephone boxes. They preferred the look of the old red ones.

Some of the red telephone boxes have been sold to people who use them in (different) ways. Some have become decorations in gardens. Red telephone boxes have also been used in clothing stores as changing rooms. On one beach, there are even telephone-box showers for people who want to wash the sand off their feet. Communities that still have working red telephone boxes are often very proud of them and do their best to look after them.

LSA2

Red Telephone Boxes-

At the beginning of the 20th century, few people in Britain had a telephone in their home. The only way most people could make calls was to use a public telephone. At first, most public telephones were in stores, and people had to pay the store clerk to use them. **As time passed, though, public telephone boxes began to be built (outside).** These allowed people to make calls on the street without being heard by other people. They also protected people from the rain. They were usually painted bright red. For many people, these red telephone boxes became a symbol of British life.

One problem with telephone boxes was that people sometimes damaged them by breaking their windows, writing on them, or trying to steal money from the telephones. In the 1980s and 1990s, it became more and more expensive to keep Britain's red telephone boxes in good condition. The telephone company began to replace them with ones that were easier to look after. However, many people did not like the (design) of these new telephone boxes. They preferred the look of the old red ones.

Some of the red telephone boxes have been sold to people who use them in (different) ways. Some have become decorations in gardens. Red telephone boxes have also been used in clothing stores as changing rooms. On one beach, there are even telephone-box showers for people who want to wash the sand off their feet. Communities that still have working red telephone boxes are often very proud of them and do their best to look after them.

LSA3

Red Telephone Boxes-

At the beginning of the 20th century, few people in Britain had a telephone in their home. The only way most people could make calls was to use a public telephone. At first, most public telephones were in stores, and people had to pay the store clerk to use them. **As time passed, though, public telephone boxes began to be built (outside).** These allowed people to make calls on the street without being heard by other people. They also protected people from the rain. They were usually painted bright red. For many people, these red telephone boxes became a symbol of British life.

One problem with telephone boxes was that people sometimes damaged them by breaking their windows, writing on them, or trying to steal money from the telephones. In the 1980s and 1990s, it became more and more expensive to keep Britain's red telephone boxes in good condition. The telephone company began to replace them with ones that were easier to look after. However, many people did not like the (design) of these new telephone boxes. They preferred the look of the old red ones.

Some of the red telephone boxes have been sold to people who use them in (different) ways. Some have become decorations in gardens. Red telephone boxes have also been used in clothing stores as changing rooms. On one beach, there are even telephone-box showers for people who want to wash the sand off their feet. Communities that still have working red telephone boxes are often very proud of them and do their best to look after them.

LSA4

Red Telephone Boxes-

At the beginning of the 20th century, few people in Britain had a telephone in their home. The only way most people could make calls was to use a public telephone. At first, most public telephones were in stores, and people had to pay the store clerk to use them. **As time passed, though, public telephone boxes began to be built (outside).** These allowed people to make calls on the street without being heard by other people. They also protected people from the rain. They were usually painted bright red. For many people, these red telephone boxes became a symbol of British life.

One problem with telephone boxes was that people sometimes damaged them by breaking their windows, writing on them, or trying to steal money from the telephones. In the 1980s and 1990s, it became more and more expensive to keep Britain's red telephone boxes in good condition. The telephone company began to replace them with ones that were easier to look after. However, many people did not like the (design) of these new telephone boxes. They preferred the look of the old red ones.

Some of the red telephone boxes have been sold to people who use them in (different) ways. Some have become decorations in gardens. Red telephone boxes have also been used in clothing stores as changing rooms. On one beach, there are even telephone-box showers for people who want to wash the sand off their feet. Communities that still have working red telephone boxes are often very proud of them and do their best to look after them.

(注) 網掛けの部分が空所を含む文。四角の囲みが空所を含む文と比較対象となる部分。

▶ 図2：4つの LSA 類似度の分析対象

LSA 類似度の算出はすべて2.2.1節で述べたコロラド大学のウェブ上で行った。LSAの基本的な設定については先行研究の手法・結果に基づき、意味空間（＝LSAが基づくコーパス）を General reading up to first year college、要因数（縮約次元数）を300

要因とした (Foltz et al., 1998; Landauer & Dumais, 1997; Wolfe et al. 2005) (注10, 11)。以上の設定のもと、4つの異なる LSA 類似度を全75項目に対して算出した。

4.3 結果と考察

受験級別の4つのLSA類似度の平均値を表9、図3に示す。

RQへの回答のために、算出されたLSA類似度を従属変数として、4(受験級：準2, 2, 準1, 1) × 4 (LSA：1, 2, 3, 4)の混合計画の二元配置分散分析を行った。その結果、受験級の主効果、 $F(3, 71) = 4.91, p = .004, \eta_p^2 = .17$ 、LSAの主効果、 $F(2.61, 185.22) = 35.13, p < .001, \eta_p^2 = .33$ 、受験級 × LSAの交互作用、 $F(7.82, 185.22) = 2.01, p = .049, \eta_p^2 = .08$ 、が有意であった。

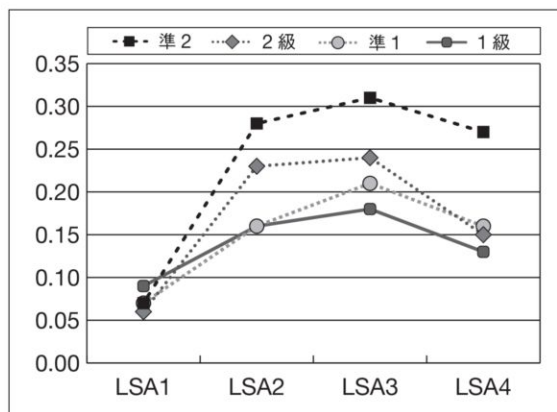
交互作用の下位検定として、まず各受験級におけるLSAタイプの比較を行った。その結果、準2級、2級、準1級においてLSAタイプの単純主効果が有意であり ($ps < .05$)、1級においては有意傾向であった ($p = .070$)。多重比較の結果を表10に示す。多重比較の結果をまとめて述べると、(a) 準2級、2級、準1級ではLSA1が他のLSA類似度よりも有意に低い値を示した、(b) 2級においてはLSA4がLSA2, LSA3よりも低い値を示したことが明らかに

なった。しかしながら、本調査では受験級(異なる難易度の問題)間でLSA1~LSA4の値に違いが見られるかがRQの焦点であるため、議論を明確にするために(a)(b)の結果について本文中で詳細な考察を述べることは避け、注釈に記載することとする^(注12)。

RQの回答に対してより重要となるのは、各LSAタイプにおける受験級間での比較である。下位検定の結果、受験級の単純主効果はLSA2, LSA4では有意であったが ($ps < .05$)、LSA1, LSA3では有意ではなかった ($ps > .10$)。多重比較の結果を表11に示す。有意差・有意傾向に着目すると、多重比較の結果は以下の2点にまとめられる：(c) LSA4において、準2級の値が他の3つの受験級よりも高かった、(d) LSA2において、準2級の値は準1級、1級よりも有意に高かった。つまり、準2級で用いられていたテキストでは「空所が含まれる文と異なるパラグラフに含まれる文」という最も大局的な意味的関連度が他の級と比べて高く、さらに「空所が含まれる文とその前後の文」というやや局所的な意味的関連

■ 表9：受験級別のLSA類似度

	k	LSA1		LSA2		LSA3		LSA4	
		M	SD	M	SD	M	SD	M	SD
準2	15	.07	.08	.28	.13	.31	.18	.27	.12
2	24	.06	.10	.23	.14	.24	.15	.15	.09
準1	18	.07	.06	.16	.13	.21	.15	.16	.07
1	18	.09	.10	.16	.09	.18	.13	.13	.07



▶ 図3：受験級別のLSA類似度

■ 表10：各受験級におけるLSA類似度の比較

LSA類似度	Sig. of F			
	準2	2	準1	1
1 - 2	.000**	.000**	.013*	.057
1 - 3	.000**	.000**	.001**	.037
1 - 4	.000**	.000**	.002**	.163
2 - 3	.531	.675	.165	.670
2 - 4	.769	.003**	.969	.294
3 - 4	.314	.002**	.109	.147

(注) ** $p < .008$, * $p < .016$. 多重比較の有意水準はBonferroniの修正を用いて設定した^(注13)。

度が準1級、1級よりも高いことが示された。

■表11: 各 LSA 類似度における受験級の比較

Grade	Sig. of F			
	LSA1	LSA2	LSA3	LSA4
準2 - 2	1.00	1.00	1.00	.001**
準2 - 準1	1.00	.048**	.460	.003**
準2 - 1	1.00	.055*	.100	.000**
2 - 準1	1.00	.513	1.00	1.00
2 - 1	1.00	.575	1.00	1.00
準1 - 1	1.00	1.00	1.00	1.00

(注) ** $p < .05$, * $p < .10$

上記で述べた各 LSA 類似度における受験級の比較から、空所が含まれる文の大局的な意味的関連度の違いが問題の難しさとかかわる可能性が示唆される。また、より難易度の高い問題（準1級、1級）では、大局的な意味的関連度に加えて、空所が含まれる文とその前後の文の意味的関連度の低さも問題の難しさの要因の1つとなっている可能性が指摘される。

これらの結果は準2級と他の級の間に、テキストの平均語数やリーダビリティの違い（表8参照）だけではなく、LSA4, LSA2 で測定されるような「空所が含まれる文の意味的関連度」の違いが存在していることを意味し、それらの意味的関連度が空所補充読解問題の難易度にかかわることを示唆するものである。

4.4 調査1のまとめ

調査1では、英検準2級から1級までで扱われていた空所補充読解問題のテキストについて LSA を行い、空所が含まれる文と「その前後の文」(LSA2)、および「異なるパラグラフに含まれる文」(LSA4)の意味的関連度が受験級によって異なることが示された。下位の受験級（準2級）と比較して、上位の受験級（準1級・1級）ではこれら2つの意味的関連度が低く、このような意味的関連度の違いが受験級間の難易度の差にかかわっている可能性がある。

続く調査2では、実際に英語学習者に空所補充読解問題の解答を行わせ、得られた解答データから算出される項目難易度と4つの LSA 類似度のかかわりについて検証を行う。

5 調査2

5.1 目的

調査2では、学習者の解答データに基づいて算出される項目の難しさと空所が含まれる文の LSA 類似度のかかわりを検証する。本調査では、学習者全体の解答データに基づく分析に加え、学習者を英語熟達度別に2つの群に分けた分析も行う。英語母語話者を対象とした Todaro et al. (2010) では、文間の LSA 類似度は読解能力の高い読み手よりも低い読み手における一貫性判断に影響を与えやすいことが示されている（2.2.2節参照）。この知見をもとに考えると、空所が含まれる文の LSA 類似度が項目の難しさに与える影響は、学習者の英語能力によって異なる可能性がある。そのため、本調査では、LSA 類似度と項目の難しさのかかわりを学習者の熟達度別に検証する分析を含めた。調査2の RQ は以下の2つである。

RQ2: 空所が含まれる文の意味的関連度の値は、学習者の解答データから算出される項目の難しさと関連があるか。

RQ3: 空所が含まれる文の意味的関連度の値と学習者の解答データから算出される項目の難しさのかかわりは、学習者の英語熟達度によって異なるか。

調査1では複数の受験級の問題を扱ったが、調査2では扱う問題の受験級を統一する。複数の受験級の問題を扱った場合、解答データから算出される項目の難しさが空所を含む文の意味的関連度によるものなのか、受験級のその他の性質（テキストの長さや語彙の難しさ）によるものなのかがあいまいになる可能性がある。それらの性質がほぼ同等である単一の受験級内においても意味的関連度が項目の難しさにかかわることが明らかにされれば、項目の難しさに対する意味的関連度の影響をより明確に示すことができる。

5.2 方法

5.2.1 協力者

協力者は、茨城県内の高校に通う1年生121名である。本調査は通常の英語の授業（50分）を利用し

て行われた。

5.2.2 マテリアル・手順

調査2においても、調査1と同様に英検の過去問題で用いられていた空所補充読解問題を使用した。扱う問題の受験級は、協力者の学年・調査校の学力水準を考慮して準2級のものとした。その他の問題の選定基準は以下のとおりである。

- テキストタイプおよびテキストの長さが協力者の解答姿勢や結果に影響を与えるのを避けるため、準2級で用いられている物語文と説明文のうち、説明文のみを対象とした^(注14)。
- LSAでは談話標識による文のつながりなどは反映されないため、談話標識や接続詞が解答に含まれる問題は避けた。

このような基準に加え、調査実施時間と問題項目数^(注15)を考慮して、計7つの空所補充読解問題(21項目)を選定した。使用した問題の例を資料に示す。テキストの提示順序のカウンターバランスを考慮し、7つのテキストの順番を入れ替えた冊子を7パターン作成した。

7つの冊子はランダムに配布され、協力者は制限時間の45分以内で7つの空所補充読解問題に解答するように指示された。なお、問題がすべて英検の準2級の過去問を用いていること、テキストが全部で7つあることなどは表紙の留意事項にて事前に伝えられた。

5.2.3 分析と採点

解答は正答項目を1、誤答項目を0とする2値データとして扱った。本研究では古典的項目分析に加えて、項目応答理論のモデルの1つであるラッシュモデルによる分析を用いて項目の難易度を算出したが、2つの分析で結果は大きく変わらなかったため、次節では古典的項目分析による結果のみを報告する^(注16)。

空所が含まれる文の意味的関連度は、調査1と同様に4種類のLSA類似度を算出した。本調査で用いた21項目の4つのLSA類似度の平均値は、表12に示すとおりである。

本調査では、項目の難しさと複数のLSA類似度(LSA1, LSA2, LSA3, LSA4)の関連を検証するために相関分析を行う。変数が3つ以上ある場合、相関係数には変数間の相関を考慮に入れない単相関

■ 表12: 調査2で用いた21項目のLSA類似度

LSA1		LSA2		LSA3		LSA4	
M	SD	M	SD	M	SD	M	SD
.07	.10	.20	.11	.25	.11	.23	.12

と、変数間の相関を考慮に入れた偏相関の2種類がある(中尾, 2010)。これら2つの相関係数のうちどちらを用いるかを判断するために、4つのLSA類似度間で単相関分析を行った。その結果、いくつかのLSA類似度間に中程度の有意な正の相関が見られたことから^(注17)、これらのLSA類似度には重複する部分があり、相互に独立するものではないと考えられる。そのため、LSA類似度間の相関を考慮し、項目難易度と各LSA類似度独自の関連性を検証するために、本調査では単相関分析ではなく偏相関分析を用いた。

5.3 結果

表13に、7つのテキスト、21項目に対する解答の記述統計を示す。平均点より、これらの問題項目がおおむね協力者のレベルに適した難易度であったことが推察される。

■ 表13: 空所補充問題7題の記述統計

N	M	SD	Max	Min
121	13.50	5.03	21	2

(注) 満点は21点である。

5.3.1 学習者全体の分析

テスト全体の信頼性については高い値が得られた($\alpha = .87$)。古典的項目分析では、ある項目の難しさを項目容易度(item facility; 項目の正答者数を全受験者数で割った値)として算出する。つまり、項目容易度はその値が高いほど項目が易しいということになる。全21項目の項目容易度の記述統計を表14に示す。

■ 表14: 21項目の項目容易度の記述統計

M	SD	Max	Min
.65	.14	.85	.30

各項目の容易度を従属変数とし、4種のLSA類似度を独立変数とした偏相関分析を行った。分析の際、制御変数は独立変数となるLSA類似度(e.g.,

LSA1) 以外の LSA 類似度 (e.g., LSA2, LSA3, LSA4) とした。分析の結果を表15に示す。LSA1 と LSA3 に項目容易度との弱い相関が見られた ($r = .36, p = -.38$) が、LSA2, LSA4 と項目容易度に相関関係は見られなかった ($r = -.07, p = -.03$)。LSA1, LSA3 で見られた項目容易度との弱い相関関係は、正負の関係が逆であることに留意されたい。しかしながら、これらの相関係数はすべて有意ではなかった ($p > .10$)。

■ 表15: LSA 類似度と項目容易度の偏相関係数

	項目容易度
LSA1	.36
LSA2	-.07
LSA3	-.38
LSA4	.03
M	-.02
SD	.30

上記の結果から、LSA1, LSA3 が空所補充読解問題に含まれる項目の容易度と弱い関連があることが示唆されるが、その相関係数は有意ではなかった。また、他の LSA 類似度も項目の難しさとの相関関係が確認されなかった。このことから、LSA によって測定される空所が含まれる文の意味的関連度と学習者全体の解答データから算出される項目容易度のかかわりは、ほとんどないと言えるだろう (RQ2)。

しかしながら、さらなる疑問として挙げられるのは、この結果が本調査の協力者間で異なるかどうかである。以下では、協力者を英語熟達度によって上位群と下位群に分類し、それぞれの群において算出される項目の難しさと空所が含まれる文の意味的関連度の相関関係が上記の結果と比べて変化するかについて検証を行う。

5.3.2 学習者の熟達度別の分析

空所を含む文の意味的関連度と項目の難しさの関係をさらに検証するために、学習者を英語熟達度によって上下群に分類し (上位群, $n = 60$; 下位群, $n = 60$) (注18)、各群の解答データから算出される項目容易度と LSA 類似度の偏相関分析を行った。受験者の英語熟達度の測定は、本調査実施時と同時期に受けた CASEC (Computerized Assessment System for English Communication) の総合得点を用いた (注19)。

5.3.1節の分析と同様に、各群の解答データから算出される項目容易度を従属変数とし、4種の LSA 類似度を独立変数とした偏相関分析を行った。分析の結果を表16に示す。熟達度上位群では、上述の結果と同様に LSA1 と LSA3 で弱い相関が見られたが、LSA2, LSA4 では相関関係が見られず、相関係数もすべて有意ではなかった。下位群でもほぼ同様の結果が得られたが、上位群と異なり、下位群では LSA1 と項目容易度の間に中程度の正の相関が見られ、相関係数は有意傾向であった ($r = .43, p = .076$)。下位群における LSA1 と項目容易度の相関係数は上位群におけるそれと比べて実質的に大きく変わるものではないものの、同じサンプル数でありながら相関係数の有意性に違いが見られていた。この結果は、LSA1 が本調査の協力者のうち、熟達度の低い群の解答データから算出される項目容易度に関連している可能性を示唆する (RQ3)。

■ 表16: LSA 類似度と項目容易度の偏相関係数 (熟達度別)

	項目容易度	
	上位群 ($n = 60$)	下位群 ($n = 60$)
LSA1	.37	.43*
LSA2	.02	-.09
LSA3	-.26	-.35
LSA4	-.04	-.12
M	.02	-.19
SD	.26	.14

* $p < .10$

5.4 考察

以上の分析結果では、学習者全体のデータから算出される項目の難しさと空所が含まれる文の LSA 類似度のかかわりはほとんどないことが示されたが、学習者を熟達度別に分類した結果では、下位群において LSA1 (空所が含まれる文と正答の語の意味的関連度) と項目容易度と中程度の正の相関と相関係数の有意傾向が確認された。つまり、熟達度下位群においては、LSA1 の値が高くなると項目への解答が易くなるという傾向があったことが示唆される。この傾向は直観的に納得できるものであり、正答となる語と空所を含む文の意味的関連度が高い、つまり、空所に入る語とその直近の文脈に意味的関連度の高い語が含まれていると、学習者は選択

肢の中から正答を選択しやすいという関係を意味するものである。以下の表17に、LSA1の値が高く、項目が易しかった問題の一部を挙げる。

■ 表17: LSA1の値が高く項目が易しかった問題の例

The method is quite easy. First, wrap what you want to cook in aluminum foil, then place it on a hot part of the engine of your car. Remember that the engine has a strong smell, so always wrap the food tightly in order to protect its (). As you drive, the engine will get hotter and begin to cook the food. After an hour or so, stop your car and check the food. If it is cooked, then you can eat it right away.
1. taste* 2. color 3. size 4. shape

(注) 空所が含まれる文を太字で示し、正答選択肢には*を付している。実際の問題では、この前後にいくつかのパラグラフがある。この項目のLSA1は0.32、項目容易度は0.75である。

この例では、正答となる taste と意味的な関連が高い smell (LSA 類似度 = .85), food (LSA 類似度 = .35) などが空所に含まれる文に存在していた。このような正答の語と直近の文脈の意味的関連度の高さが、項目の易しさにかかわっていたと考えられる。このような関係が熟達度下位群において見られたことは、英語母語話者を対象とした Todaro et al. (2010) の知見 (読解能力の低い読み手ほど文の一貫性判断に意味的関連度が影響しやすい) と部分的に重なるものであり、意味的関連度が項目の解答に与える影響は学習者の英語能力とかかわりがある可能性が指摘される。

本調査の結果を前述の調査1と比較すると、興味深い示唆が得られる。調査1では、準2級の問題の特徴として他の受験級よりもLSA2, LSA4の値が高いことが示されたが、LSA1については受験級間で統計的な有意差はなく、問題の難易度とのかかわりは示唆されていなかった。一方で、調査2ではLSA1が熟達度下位群における項目の難易度とかかわりがある可能性が示唆された。調査1では異なる受験級を難易度の指標としていたのに対し、調査2では同じ級(準2級)内の解答データをもとに項目の難易度を算出していたことから、「異なる受験級間」と「同じ受験級内」では項目の難しさと空所を含む文の意味的関連度とかかわり方が異なる可能性が考えられる。また、調査1で確認されたような

LSA 類似度の統計的な差と、テスト場面において学習者の解答に影響する、もしくは学習者が敏感に知覚する LSA 類似度の差が異なる可能性もある。つまり、調査2で問題に解答した学習者、特に熟達度の低い学習者は、調査1で難易度とのかかわりが示唆された LSA2 や LSA4 のような意味的関連度よりも LSA1 に依存して解答を行う傾向にあり、たとえその値が統計的に有意にはならないような違いであっても項目の解答に影響したのかもしれない。

本調査では統計的に明確な結果が得られていないため、あくまで可能性の示唆にとどめるが、以上に述べたように調査2の結果は、英語熟達度の低い学習者において「正答の語と空所が含まれる文の意味的関連度」が高くなるほど項目が易しくなる傾向を示すものである(注20)。

最後に、LSA3が項目容易度と弱い負の相関を見せたことについて触れたい。これは、空所を含む文とパラグラフ内の2文以上離れた文の意味的関連度が高くなると項目が難しくなるという関係である。この関係は、調査1で得られたLSA2, LSA4と受験級のかかわりや、上記のLSA1と項目容易度のかかわりと逆の傾向を示すものである。この理由についてはいくつかの可能性が考えられるものの、LSA3と項目の難しさの関連について統計的に有意な結果は得られていないため、この点は今後さらなる実験データに基づいて議論される必要があると考える。

5.5 調査2のまとめ

調査2では、学習者の解答データから推定される項目の難しさと空所が含まれる文の意味的関連度のかかわりを検証した。分析の結果、学習者全体の解答データでは項目の難しさと空所を含む文の意味的関連度のかかわりはほとんど見られなかったが、熟達度下位群の解答データに基づく分析では、空所を含む文と正答となる語の意味的関連度が高いほど、項目が易しくなる傾向が見られた。

6 結論と今後の課題

本研究では、英検で用いられている空所補充読解問題を対象として、LSAによって測定される空所を含む文の意味的関連度と項目の難しさのかかわり

を検証した。準2級から1級までの問題を対象とし、各級における複数のLSA類似度の違いを検証した調査1では、空所が含まれる文と「その前後の文」の意味的関連度(LSA2)、および「異なるパラグラフに含まれる文」との意味的関連度(LSA4)が、準2級において上位の級(準1級、1級)よりも有意に高いことが示され、それらのLSA類似度と問題の難しさとのかわりかかわりが示唆された。調査1では、異なる受験級を難易度の指標としていたため、この結果は厳密に言えば、項目の難しさとのかわりというよりは異なる受験級で用いられるテキストの特徴と解釈する方が適切かもしれない。つまり、調査1の結果は、より上位の受験級における空所補充読解問題では、「空所を含む文とその前後の文の意味的関連」や「空所を含む文の大局的な意味的関連」が低いテキストが用いられていたことを示すものであると解釈される。

一方、学習者の解答データから算出される項目容易度を扱った調査2では、熟達度下位群においてLSA1と項目容易度に中程度の正の相関と相関係数の有意傾向が確認された。つまり、熟達度下位群においては空所を含む文と正答となる語の意味的関連度が高い場合、その問題への正答率が高まる傾向があった。このことから、熟達度の低い学習者は「空所を含む文」と「正答となる語」の意味的関連度に依存した解答をしていた可能性が指摘される。

本研究におけるこれら2つの調査の結果は、LSAによって測定される空所を含む文の意味的関連度が、さまざまな形で空所補充読解問題の難易度にかかわることを示すものである。そのため、LSAによって測定される意味的関連度は、空所補充読解問題作成における客観的・科学的な指標の1つとなる可能性がある。例えば、空所補充読解問題に用いるテキストについては、ある文の前後の文との意味的関連度や異なるパラグラフに含まれる文との意味的関連度を観点として選定を行ったり、実際にその文に空所を設ける際には、その文と削除する語の意味的関連度や受験者の英語熟達度を考慮したりすることで、空所補充読解問題における項目の難易度が調整され得るであろう。

最後に、本研究の課題およびLSAを用いた今後の英語教育研究への示唆を述べる。本研究の調査2は、用いる問題の受験級や協力者の学年を統一して行われたが、それらの要因が異なる実験を行った場

合、本研究の結果と比較してどのような結果が得られるのかは興味深い研究課題である。例えば、より熟達度の高い学習者を対象に2級、準1級などの問題を扱った場合、LSA1ではなく他のLSA類似度が項目の難易度と何らかの相関関係を見せるかもしれない。このような可能性は、問題の項目数や受験者数をさらに多くした実験により今後検証される必要があるだろう。また、本研究はLSAによる空所が含まれる文の意味的関連度と項目難易度の関係に焦点を当てたが、実際の項目の解答には意味的関連度だけでなく、文の接続関係の理解などがかわることが考えられる。Foltz(2007)が示唆するように、LSAによる分析をさらに正確なものにするためには、そのような要因が分析とともに考慮されることが望まれる。今後、この点も考慮しながらLSAを用いることで、より精緻なテスト特性の解明へとつながることが期待される。本研究で用いたLSAは、英語教育におけるテスト研究、読解プロセス研究に応用される可能性が大いにあるものの、その分析はコーパスに基づいて行われることに留意されたい。本研究ではLSAのウェブサイトを利用して可能な英語母語話者の言語コーパスを使用した。今後英語教育分野でLSAが用いられる際は英語学習者のコーパスに基づいた分析も検討される必要がある。今後の英語教育研究におけるLSAの利用可能性向上のためにも、英語学習者の大規模コーパスに基づいてLSAが行える条件・環境の整備が求められるであろう。

謝 辞

本研究を行う貴重な機会を与えてくださいました(財)日本英語検定協会の皆様、ならびに選考委員の先生方に厚く御礼申し上げます。とりわけ、私の研究を担当していただきました池田央先生には、有益なご助言・ご指導をいただき深く感謝いたします。そして、筑波大学大学院の卯城祐司先生には、調査の実施から報告書執筆に至るまで親身なご指導をいただきました。心より感謝申し上げます。また、研究室の先輩である高木修一さん、清水遥さんにも多くのご助言と励ましをいただきました。この場を借りて御礼の言葉を送らせていただきます。最後に、本調査にご協力いただきました、茗溪学園英語科の松崎秀彰先生、ならびに学生の皆様に深く御礼申し上げます。

注

- (1) 実験協力者に頭の中で考えていることを発話させる手法。テスト研究に限って述べれば、項目解答の際に解答の理由や手がかりを報告させる。
- (2) 実際は単語×文脈行列の作成、特異値分解を経て意味空間の構築が行われるが、それらの詳細な説明は猪原・楠見 (2011)、中村 (2008) を参照されたい。
- (3) 提示される単語の数は調整可能である。
- (4) Landauer, Foltz, and Laham (1998) によると、LSA は 2 つの見方が可能である。1 つは、(a) 単語間や文章間の意味的関連性を推定する手法としての見方である。もう 1 つは、(b) 人間の知識獲得・活用プロセスのシミュレーションモデルとしての見方、つまり、LSA による単語や文章の意味表象は、人間のさまざまな認知活動をシミュレーションすることができるという見方である。本研究では、これら 2 つの見方のうち、(a) 概念間の意味的関連性を推定する手法として LSA を解釈する。
- (5) 例えば、Wolfe et al. (2005) では以下の 2 文は因果的関連・意味的関連がともに高いと評価されている：Malcolm collapsed while treating patients in his office. He was carried unconscious to a hospital. (Wolfe et al., 2005, p.172)
- (6) TOEFL については The Official Guide to the TOEFL Test (3rd Ed.) Authentic TOEFL iBT Practice Tests を、IELTS については Official IELTS Practice Materials 1, 2 を調査対象とした。その他のテストについては表 4 に示すとおりである。
- (7) 具体的には、準 2 級の大問 4、2 級の大問 3、準 1 級・1 級の大問 2 が対象となった。
- (8) Kobayashi (2002) では項目が測定する知識の明確な分類基準は記述されていないため、表 6 の分類基準は Kobayashi の結果を参考に本研究で独自に設定されたものである。
- (9) センター試験では、他のテストと比べて複数の知識が測定されると考えられる場合が多く見られた。しかし、本調査では分類の一貫性を保つため、より強く測定されると考えられる知識へ分類するという基準を用いていた。そのため、センター試験において分類の一致率が高いことは、必ずしもセンター試験の項目がある特定の知識を測定する傾向が強かったことを意味するものではない。
- (10) LSA はすべて One-To-Many Comparison により行い、分析単位を考慮して適宜重みづけを行った。
- (11) 固有名詞は、その繰り返しにより不当に高い意味的関連度を算出するおそれがあるため、削除して分析を行った。また、LSA に用いたコーパスに存在しない単語も削除して分析を行った。
- (12) (a) については、LSA1 では「空所が含まれる文」に対して「語」が比較対象になるのに対し、他の LSA では「文」が比較対象になるという違いが影響している可能性がある。つまり、「文」よりも「語」の持つ意味情報量が小さく、相対的に LSA1 で算出される意味的関連度が小さくなったのかもしれない。しかしながら、1 級ではそのような違いが見られていないため、ある程度意味的関連度の違いを反映している部分もあるだろう。(b) については、他の受験級と比べて、2 級においては空所を含む文の「前後の文との意味的関連度」および「パラグラフ内の文との意味的関連」と「異なるパラグラフ内の文との意味的関連」の差が大きいことを強調する結果であると解釈される。
- (13) 有意水準 5%、10% をそれぞれ 6 (比較を行う回数) で割り、0.8%、1.6% に設定した。
- (14) 物語文と比べて、説明文の方がやや長い。
- (15) 本調査では、テストの信頼性や分析の安定性を考慮して、20 項目以上を確保できるようにした。
- (16) ラッシュモデル分析は Winsteps (ver.3.72) を用いて行った。なお、熟達度別の分析では、協力者を 2 つの群に分ける操作により項目困難度 (logit 値) を算出するためのサンプルサイズが不足してしまうため、ラッシュモデル分析は行わなかった。
- (17) LSA2 と LSA4、LSA3 と LSA4 に中程度の正の相関が見られた ($r = .51, p < .05$; $r = .66, p < .05$)。
- (18) 熟達度別の分析では、本調査で使用した熟達度テストを受けていない 1 名のデータを削除した。
- (19) CASEC は語彙、表現、リスニングのセクションから構成されるため、ここで指す熟達度とは読解熟達度ではなくより一般的な英語熟達度である。
- (20) LSA1 のみが下位群における項目の難しさと関連があったという結果については、LSA1 の算出方法が他の LSA 類似度のそれと異なる (LSA1 のみが「正答の語」を比較対象としている) ことが理由であるとも考えられる。しかしながら、LSA1 と同様に正答の語を中心とした方法でそれぞれ LSA2~LSA4 を算出し (「正答の語」と「空所を含む文の前後の文」を LSA2、「正答の語」と「それが含まれるパラグラフ内の文」を LSA3、「正答の語」と「異なるパラグラフに含まれる文」を LSA4 とした)、下位群の項目容易度との偏相関を再分析した結果、LSA1 のみ中程度の有意な正の相関が見られ ($r = .47, p = .049$)、その他の LSA 類似度の相関の正負、相関係数の大きさ、有意性についても 5.3.2 節で述べた結果とほぼ同様の結果が得られた。

参考文献 (*は引用文献)

- * Abraham, R.G., & Chapelle, C.A. (1992). The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76, 468-479.
- * Alderson, J.C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13, 219-223.
- * Alderson, J.C. (1980). Native and nonnative speaker performance on cloze tests. *Language Learning*, 30, 59-76.
- * Alderson, J.C. (2000). *Assessing Reading*. Cambridge, UK: Cambridge University Press.
- Bachman, L.F. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16, 61-70.
- * Bachman, L.F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 19, 535-556.
- * Dennis, S. (2007). How to use the LSA web site. In T.K. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp.57-70). Mahwah, NJ: Lawrence Erlbaum Associates.
- * English Testing Service. (2007). 『TOEIC®テスト新公式問題集 Vol.2』. 東京: 財団法人国際ビジネスコミュニケーション協会.
- * English Testing Service. (2008). 『TOEIC®テスト新公式問題集 Vol.3』. 東京: 財団法人国際ビジネスコミュニケーション協会.
- * English Testing Service. (2009). 『TOEIC®テスト新公式問題集 Vol.4』. 東京: 財団法人国際ビジネスコミュニケーション協会.
- * Foltz, P.W. (2007). Discourse coherence and LSA. In T.K. Landauer, D.S. McNamara, S. Dennis, and W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp.167-184). Mahwah, NJ: Lawrence Erlbaum Associates.
- * Foltz, P.W., Kintsch, W., & Landauer, T.K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, 25, 285-307.
- * Graesser, A.C., & McNamara, D.S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371-398.
- 平井明代. (2010). 『テスト問題・教材再利用のすすめ: TEASY 理論編』. 東京: 丸善プラネット.
- * 猪原敬介・楠見孝. (2011). 「潜在意味解析に基づく概念間類似度の心理学的妥当性—言語統計解析アプローチの効用と限界—」. 『心理学評論』, 54号, 101-122.
- * 川畑松晴. (2001). 「クローズテスト」. 門田修平・野呂忠司 (編). 『英語リーディングの認知メカニズム』 (pp.286-297). 東京: 東京図書.
- * Kobayashi, M. (2002). Cloze test revisited: Exploring item characteristic with special attention to scoring methods. *The Modern Language Journal*, 86, 571-586.
- * Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- * Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- * 中村健太郎. (2008). 「潜在意味解析」. 豊田秀樹 (編). 『データマイニング入門—R で学ぶ最新データ解析—』 (pp.271-293). 東京: 東京図書.
- * 中尾桂子. (2010). 「相関分析: データの関連を見る」. 石川慎一郎・前田忠彦・山崎誠 (編). 『言語研究のための統計入門』 (pp.85-104). 東京: くろしお出版.
- * Oller, J.W., & Conrad, C.A. (1971). The cloze techniques and ESL proficiency. *Language Learning*, 21, 183-195.
- * Taylor, W.L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- * Todaro, S., Millis, K., & Dandotkar, S. (2010). The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes*, 47, 421-446.
- * Trabasso, T., & Wiley, J. (2009). What happens at reunions? Exploring causal connections and their role in reunion effects. *Discourse Processes*, 46, 269-308.
- * Wolfe, M.B.W., Magliano, J.P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, 39, 165-187.
- * Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20, 267-293.
- * Yoshida, S. (1997). Strategies in answering cloze items: An analysis of learners' think-aloud protocols. *JACET Bulletin*, 28, 207-222.
- * Zwaan, R.A., & Madden, C.J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 283-288.

調査2で使用した空所補充読解問題の例

A Little Good News

When we watch the news on TV, we usually hear about such things as wars, accidents, and crimes. While it is important to know about these things, many people would like to hear more good news. An American woman named Geri Weis-Corbley decided in 1997 that more happy news should be (1). Since that year, she has been running the *Good News Network* website, which offers good news from around the world.

Weis-Corbley used to work for a TV company, filming and writing news stories. Based on her experience, she thought that most people would like to hear more good news. Every day, she searches the Internet for happy stories. She then puts them on her website. In September 2008, there was a lot of bad news. Several large American banks got into trouble, and many companies had to go out of business. Because of these (2), visitors to the *Good News Network* increased. This shows that when most news is bad, there are many people who want to see some good news instead.

Weis-Corbley believes that the media should be more (3). For example, when a small number of people do something bad, it is always in the news. She argues that we need to have good news as well as bad news so we can think about the good things in life. Tal Ben-Shahar, a professor of psychology at Harvard University, has a similar opinion. He says that only hearing bad news makes people see everything in a negative way. Sometimes, hearing good news can help us to have a more positive point of view.

-
- (1) 1 reported* 2 invented 3 corrected 4 blocked
 - (2) 1 replies 2 secrets 3 promises 4 difficulties*
 - (3) 1 expensive 2 balanced* 3 serious 4 dramatic

(英検準 2 級2010 年第 1 回より引用。正答選択肢には*を付している)