研究部門 ■ 報告 I

統合タスクにおける類似が スピーキングパフォーマンスに与える影響

一文章と質問内容の類似レベルに着目して一

茨城県立水戸第一高等学校 教諭 矢野 賢

申請時:茨城県立日立第二高等学校 教諭



本研究では, 英検二次面接試験問題を 用い, 統合タスクにおいて読んだこと の関係がパフォーマンスにどのような影

と話すことの関係がパフォーマンスにどのような影響を与えるのかについて,類似の枠組みを用いて比較検証を行った。

まず統合タスクに関する研究および類似に関する研究などについて概観し、これらの知見に基づき、調査を2つ行った。調査1では、インプットおよびアウトプットに用いるトピックの親密度および類似度の違いについて質問紙を用いて調査し分析を行った。調査2ではインプットに含まれる情報がどの程度アウトプットに使用されパフォーマンスに影響を与えているのかについて、実際のインタビューにおける発話をもとに調査および分析を行った。

この結果、読んだ内容と質問内容とが高次レベルで類似している組み合わせではパフォーマンスを促進した例があった一方、表面的にのみ類似していた場合には不適切な転移を行い誤った説明につながった例が見られた。



はじめに

新学習指導要領(高等学校編)(文部科学省, 2009a) においては、急速に進むグローバル化に対応するため、情報の受信のみならず発信する力が求められている。この背景には、OECD による PISA 調査などの結果から、知識・技能を活用する力が求められている(文部科学省, p.2)ことがある。これを受け、外国語科・英語科における必履修科目である「コミュニケーション英語 I」においては、「聞

いたり読んだりしたこと、学んだことや経験したことに基づき、情報や考えなどについて、話し合ったり意見の交換をしたりすることや、簡潔に書くことなどの統合的な言語活動」(文部科学省,2009b, p.5)を行うことが求められている。言い換えれば、4技能を有機的に関連づけた指導(読んだもの・聞いたものに基づき話す・書く活動)が求められていると言える。

このような統合タスクを用いた指導が求められる一方、これらに関する研究も進んできている。これまでの研究から、統合タスクにおいては、選択し統合するスキルが必要である(Brown, Iwashita, & McNamara, 2005)こと、トピックによってパフォーマンスが異なる(Brown et al., 2005; Frost, Elder, & Wigglesworth, 2011; Lee, 2006)ことなどが指摘されている。しかしながら、このような統合タスクを行う際にどのような情報を選択し統合がなされるかについては、筆者の知る限り十分な説明がなされていない。

本研究では、読んだことと話すことがどのような関係にあるのかをGentner(1983)やGentner、Ratterman and Forbus(1993)らの類似の枠組みを用いて比較する。具体的には、日本のEFL環境における学習者を対象とした複数の技能を含んだテストとして英検の二次面接試験を用いる。「読んだこと」および「聞いたこと」を踏まえた「話す」ことのパフォーマンスを評価するこのテストを使用し、統合タスクにおけるインプットとアウトプットの類似関係がスピーキングパフォーマンスにどのような影響を与えるのかについて検証する。

本稿の構成としては、まず統合タスクに関する研

究および類似に関する研究などについて概観する。次にこれらの知見に基づき、調査を2つ行う。調査1では、インプットおよびアウトプットに用いるトピックの親密度および類似度の違いについて調査する。この調査1の結果を踏まえ、調査2ではインプットに含まれる情報がどの程度アウトプットに使用されパフォーマンスに影響を与えているのかについて、3つの異なる分析を行いその関係について検証する。最後に、これらの結果をもとに教育的示唆を述べる。

2

先行研究

2.1 パフォーマンスにおける独立タスクと統合タスク

現在の ESL / EFL 環境における英語学習者を対象 としたテストにおいては、統合タスクが用いられる ようになってきている。例えば「英語圏への留学に おけるアカデミック場面での英語能力」を目的とし た TOEFL iBT のスピーキングテスト (English Testing Service, 2013), 「主に国際ビジネスの場面 における英語力の測定」を行う TOEIC のスピーキ ングテスト (ETS, 2006, 2007), 「日常生活および社 会生活における英語の理解および使用能力の測定」 を目的とした英検における二次面接試験(日本英語 検定協会,2013)、「与えられたパッセージの内容の 再話活動 | を用いた SRST (Hirai & Koizumi, 2009) など、さまざまな対象者に向け多様な目的で統合テ ストは用いられている。これらに共通しているの は、単に個別のスキルを測定する、あるいは言語知 識を測定することにとどまらず、留学先やビジネス など実際の状況におけるパフォーマンスを測定する ことに主眼を置いているという点である。

パフォーマンステストにおけるスピーキングタスクを他の技能との関連で考えた場合、これらを3つに分類することが可能である。例えば Lee(2006)は、スピーキングタスクを(1) スピーキングを単独で、または視覚的な情報をもとに話す独立タスク(Independent task); (2) 聞いたものに基づいて話すタスク(Listening-Speaking); (3) 読んだものに基づいて話すタスク(Reading-Speaking)の3つに分類している。Lee によれば、これらのタスクのうち独立タスクは、「話者の経験や一般知識などに依存

して答える必要がある」(pp.131-132)ものである。その一方,(2)や(3)のような統合タスクにおいては「話すためにまず講義やテキストなどを聞いたり読んだりして理解し、それから理解したことを示すために話をする準備を行う必要がある」(p.132)ものであると述べている。つまり統合タスクにおいては、理解したことを自分の持っている知識などと比較し、必要に応じてこれらの情報を統合することにより、話すための準備を行う必要があるのである。これらのタスクのうち、今回の研究においては、特に(3)の Reading-Speaking 型の統合タスクに焦点を当てる。

2.2 独立タスクにおける問題点

それぞれのタイプのタスクにおいては問題点が指 摘されている。独立タスクの問題点としては主に、 背景知識に依存するため扱うトピックが限定される こと、およびタスクの真正性の2点である。Brown et al. (2005) は TOEFL iBT のスピーキングテスト を用い、統合タスク (Reading-Speaking および Listening-Speaking) と独立タスクにおける評価ス ケールの妥当性について検討した。この研究の中 で Brown et al. は、アカデミックな状況においては、 テキストを読んだり講義を聞いたりしたものをス ピーキングパフォーマンスに統合するという認知的 に複雑なプロセスが必要とされるが、インプットを 与えない独立タスクを用いた場合には扱うトピック が限定され、かなり平凡なものになってしまう (p.1), と主張している。また, Frost et al. (2011) も同様に統合タスクによる評価の必要性を主張して いる。Frost et al. によれば、実際のコミュニカティ ブな行動においては複数のスキルや非言語の認知能 力が必要なのである。同様に Luoma (2004) は、 統合タスクが必要とされているのは、「真正性のあ るテストにおける言語使用を望んでいる」(p.43) ためであると述べている。つまり統合タスクを用い たテストは、個々のスキルを個別に測定することよ りも、アカデミックな状況や実際のコミュニケー ション場面などでのパフォーマンスを測定すること に重きを置いており、その実際的な必要性の高まり から用いられるようになってきているとも言える。

2.3 統合タスクにおける問題点

このように、統合タスクを用いたさまざまなテス

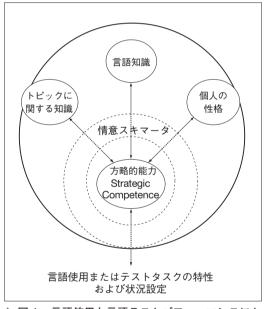
トは、言語的な側面だけでなく実際のコミュニケー ション場面における英語力の測定を行うという目的 から使用されるようになってきている。しかしなが ら、統合タスクにおいてもその問題点が指摘されて いる。その1つに、インプットのトピックによりそ のパフォーマンスの結果が異なることがある (Brown et al., 2005; Lee, 2006, Shohamy, 1983)。 1 ンプットのトピックによる違いは、個人における背 景知識の違いが影響を与えていると考えられる。統 合タスクにおいて与えられたインプットの理解にお いては個人の持っている背景知識がその助けとなり うるし、話したり書いたりする際においてもまた。 背景知識は影響を与える可能性のあることが考えら れる。この他、Brown et al. や Shohamy の研究で は、いずれもタスクやそのコミュニケーションのタ イプによってパフォーマンスが異なる点を指摘して いる。また、Frost et al. (2011) は、特に言語熟達 度の低い話者において、インプットのトピックによ りパフォーマンスにおける内容の誤り(ゆがみ)が 多く見られることを指摘している。つまり、個人の 言語熟達度の違いによって、インプットの情報にゆ がみが生じるのかどうかが変わってくる可能性があ るのである。

2.4 パフォーマンステストにおけるト ピックの役割

トピックはパフォーマンステストにおける構成要 素の1つであることは、Bachman and Palmer (1996. 2010) によってモデル化されている。Bachman and Palmer (1996) はパフォーマンステストにおい て考慮すべき構成要素として、トピックに関する知 識、言語に関する知識、そして個人の性格の3つを 挙げている (pp.61-76, 図1参照)。このモデルで は、メタ認知要素である方略的能力を用いることに より、(1) ゴールの設定;(2) 必要な知識の調査;(3) プランニングの順に行われることで言語使用が可能 になると想定されている。また、同じタスクを行っ た場合であっても、話者によって異なった方略がと られると考えられている。この過程における(2)必 要な知識の調査においては、トピックに関する知識 があるのかどうか、その知識はタスクの遂行に必要 なのか、などが検討される。Bachman and Palmer (1996) はこの考え方に基づき、「個別のスキルをそ れぞれ抽象的なものとして位置づけるよりも、むし

ろ言語能力とタスクの特定の組み合わせにおけるパフォーマンスであると考える方が,より有用である」 (p.75) と主張している。

タスクにおいて使用されるトピックの違いによってパフォーマンスの結果が異なることは、個人の持っているトピックに関する知識が異なっていることから、ある程度避けられないことであると考えられる。しかしながら、異なったトピックを使用することによってテストにおけるパフォーマンス結果が異なることは、テスト結果を一般化して解釈する場合には好ましくない。そのため、Brown et al. (2005)、Lee (2006)、Shohamy(1983)いずれの研究においても、統合タスクを用いたパフォーマンステストを行う際には、複数のタスクを併用して使用する必要があると結論づけている。



▶ 図 1:言語使用と言語テストパフォーマンスにおける構成要素 (Bachman & Palmer, 1996)

2.5 類似に関する研究

統合タスクにおいては、聞いたり読んだりしたことをもとにして話す要約タスク(e.g. Hirai & Koizumi, 2009)もあるが、本研究で取り上げる英検のスピーキングテストのように、読んだ内容に関連した事柄について、自分の意見などを述べるタスクもある。このようなタスクにおいては、読んだ内容と話す内容の関連がパフォーマンスに影響を与える可能性があると考えられる。それはタスク遂行にお

ける選択一統合の過程の中で、類似した複数の領域 間に知識の転移が起こる可能性があるためである。

では、類似しているということはどういうことであろうか。類似に関する研究においては、その類似した2つの領域において転移の元になる領域を「ベース領域」、選択した情報が転移される対象領域を「ターゲット領域」と呼んでいる(Gentner,1983; 鈴木,1996; 山﨑,2001; Zook,1991)。「電子」の仕組みを説明する際に、「衛星の仕組み」をその類似した例として示す場合、元になるその「衛星」に関する知識を「ベース領域」、「電子の仕組み」に関する知識を「ダーゲット領域」と呼んでいる。

また、類似とはベース領域とターゲット領域において共通した情報を含んでいるということであると考えることができる。Gentner(1983)は、この任意の領域に含まれる個々の対象物の単位を「オブジェクト」と呼んでいる。オブジェクトは、例えば「ウサギ」のような明確な物体(clear entities)であったり、「ウサギの耳」のようにある物体の一部であったり、また「ウサギの群れ」のように、組み合わせによる小さな単位としてとらえることもできる。

これらのオブジェクトは、属性 (attribute)を持つことがある。属性とは、論点 (argument)を取る述部 (predicate)が1つだけある場合のことを指している。例えば「そのウサギは大きい」といった場合、このウサギの属性は「大きい」である。これに対して、複数のオブジェクトが「関係を持っている」といった場合、論点を取る述部が2つ以上ある場合のことを指す。この関係は、そのレベルによって一次的関係 (First Order Relation: FOR)と高次の関係 (Higher Order Relation: HOR)に分けることが可能である。例えば、「xとyが衝突した」といった場合 FOR であるが、「xとyが衝突した」といった場合 FOR であるが、「xとyが衝突した結果炎上した」というように複数のオブジェクトが共通している場合、HOR であると言える。HOR に含まれる情報としては、空間的情報、命題的情報、動的な因果情報などがあると言われている。

さらに Gentner et al. (1993) は、オブジェクトの

属性レベルの類似の有無と、構造的類似のレベル(一次の関係か高次の関係か)によって、類似を4つに分類している。「牛乳と水は似ている」といった場合、水と牛乳はともに「液体」という属性を持っており、どちらも「飲む」といった行為で他のオブジェクトと関係づけることができる。このような場合、2つのオブジェクトは文字どおりの類似(Literal Similarity:LS)であるとされる。

これに対して、「衛星は電子に似ている」といった場合、「衛星」と「電子」は共通の属性は持たない。しかしながら、「衛星」は「惑星」の周りを「回転」し、「電子」もまた「原子核」の周りを「回転」すること、「電子」は「原子核」よりも「小さい」し、「衛星」も「惑星」よりも「小さい」、といったように、複数の関係構造が類似している。このように、異なった2つの領域間において他のオブジェクトとの関係構造が複数にわたって類似していたり、さらに「回転することによってエネルギーを生み出す」など、高次の関係が類似している場合、アナロジと呼ばれる。

また、属性レベルで類似していても低次の関係しか共有していない場合は、FORであるとされる。例えば、「野球」と「サッカー」は、ともに「スポーツ」という属性において「選手」が「ボールを使う」といったように、同じ属性を持っているが、より上位の関係構造は異なる場合のものを指す。

また、属性、関係、いずれも類似していない場合、例えば「太陽はオレンジに似ている」という関係は、単なる「丸い」という表面的類似(Surface Similarity:SS)である。

これらの関係レベルは「連続したものではあるが、類似について検討していく際には有用なものである」と Gentner et al. (1993) は述べている (p.527)。 Gentner et al. はこの類似の転移における特性として、SS はアナロジに比べて目立つために情報として取り出しやすいが、問題解決においては高次の関係を含んだアナロジの方がより参考になる重要な情報を含んでいるとした。

■ 表 1:類似に関する 4 つの分類 (Gentner et al., 1993)

オブジェクト レベルの類似	高次での構造的類似あり	高次での構造的類似なし
あり	文字どおりの類似 (Literal Similarity:LS) 例) 「牛乳」 と 「水」	低次の関係のみ類似 (First Order Relation:FOR) 例) 「野球」と「サッカー」
なし	アナロジ (Analogy:AN) 例)「衛星」と「電子」	表層的類似 (Surface Similarity:SS) 例) 「太陽」 と「オレンジ」

山﨑(2001)も同様に、「属性レベルでターゲット領域と類似している領域では類似はベース領域として検索されやすいが、ターゲット領域の問題解決にうまく適用できるためには関係レベルでの類似が必要である」(p.21)と述べている。問題解決に適用するには、ベースとターゲット間において、高次の構造的関係がある必要があるのである。これは、たとえ見た目が似ていたとしても構造が異なっていれば導かれる答えが異なるからであると考えられる。

この他、類似性の分類については2つの領域間の 距離という考え方を用いている研究もある。例え ば Gentner (1983)や Halpen, Hansen, and Riefer (1990)においては、一次か高次かという2つの分 類ではなく、領域間の距離をベース領域とターゲッ ト領域間で共有しているオブジェクトの数によって 決定するといった考え方を用いている。

このように、類似の枠組みは研究によってさまざまな考え方が存在しているが、今回の研究で取り上げる Reading-Speaking 型の統合タスクにおいても、これら類似の枠組みを参考にして考えることが可能である。例えばベースとして与える文章と話すためのタスクで問われる内容の間に「太陽」と「オレンジ」のような属性の類似があった場合には、ベース領域の情報が検索されやすいことが予想される。一方で、タスクの問題解決にベース知識を適用するためには、これらの間に「衛星」と「電子」のような高次の構造的関係が必要なのではないかと考えることができる。

2.6 EFL / ESL リーディング分野に おけるアナロジに関する研究

EFL/ESLにおけるリーディング研究においては、ベース領域として与えたテキストのターゲット領域への理解や記憶、知識の応用などへの影響を扱ったものが多く、これらの研究においてはアナロジが文章理解を促進しないという結果が多い(Brantmeier, 2005; Hammadou, 1990, 2000)。この理由としては認知的な負荷が高いことや類似性が気づかれにくい点が指摘されている。

また、「カメラの構造」と「目の構造」という 2 領域間における親密度を考慮した Yano(2011)では、外国語で書かれた 2 つの文章を続けて読んだ際に、最初に読んだ文章における領域(ベース領域)

が、続いて読んだ文章の領域(ターゲット領域)よりも親しみのあるテキストであった場合、推測課題の解決を促進した。しかし、山﨑(2001)が、「説明者にとってベース領域とターゲット領域のどちらがなじみ深いかは決まっていない」と述べているように、読み手の持っている背景知識によって文章の親しみやすさは異なると考えられる。このため、親密度が類似を用いた問題解決に与える影響は、トピックと読み手との関係によって異なることが考えられるが、統合タスクにおいても親密度がどのように類似した情報の選択一統合に影響を与えるかについては検討する必要がある。

2.7 先行研究と本研究のつながり

これまで述べたように、実際のコミュニケーション場面やアカデミックな場面への対応能力を測定することを目的として、統合タスクを含んだテストが多く用いられるようになってきている。しかしながら、これまでの統合タスクを用いたスピーキングパフォーマンスに関する研究においては、類似の枠組みを用いて検討した研究は筆者の知る限りでは十分でない。そのため、本研究において類似と統合タスクの関係を明らかにしたい。

3

研究課題

今回取り扱う英語検定試験の二次試験にて行われ ているReading-Speaking 型の統合タスクにおいて は、読んだ文章の内容から選択された情報がスピー キングにおいてどのように統合されるのかについて 調査する。特にこのタイプのタスクにおいて、イン プットとして読む文章内容の領域(ベース領域)を. 質問とその回答内容としてのスピーキング領域であ るターゲット領域へ適用する際に、親密度および類 似度が影響するのかについて明らかにする必要があ る。そのため調査1では、まず各領域の親密度およ びベース領域とターゲット領域間の類似性について 調査する。続く調査2では、統合タスクにおいて、 類似のレベルの違いがパフォーマンスにどのように 影響を与えるのかについて調べる。類似レベルの枠 組みとして、オブジェクトの数による一次、高次の 類似を用いる。

また、このような統合タスクにおいては、情報の

選択-統合が必要であることが先行研究(Brown et al., 2005)において指摘されているが、この際に、ベース領域に関して持つ親しみやすさや問われている内容の類似のレベルが、この選択一適用の過程において影響を与えるかどうかについても検討する。特に親しみやすさに関しては、より親しみやすい領域からなじみのない領域に転移が行われるという考え(Yano, 2011; Zook, 1991)がある一方、親しみやすさの度合いは学習者により異なる(山崎, 2001)。これらのことから、調査1、2の結果を踏まえ、「読む一話す」型のタスクにおけるパフォーマンス結果と親密度・類似度との関係についての考察を行う。

4

調査 1

4.1 調査 1 の目的

調査1の目的は、今回の参加者におけるベース領 域とターゲット領域間の親密度の違い、およびト ピック間の類似の違いを調べることである。統合タ スクにおいては、テキストから読み取った内容をも とに自分の考えを話すタスクを行う際に、それぞれ のトピックの違いがタスクにおけるパフォーマンス に影響すると考えられている(Bachman & Palmer, 1996; Brown, et al, 2005; Lee, 2006; Shohamy, 1983). また、2つの領域の関係においてベース領域から ターゲット領域へ情報の転移がなされる際には、そ の類似度によりその影響が異なるという考え (Gentner et al., 1993; Halpen et al., 1990; 山崎, 2001) がある。また、ベース領域とターゲット領域におけ る親密度の影響が考えられるが(山崎, 2001; Yano, 2011, Zook, 1991). 統合タスクにおける類似や親密 度との関係は十分には明らかではない。そのため調 査1では、今回の統合タスクにおいて最初に提示さ れるテキスト内容に関する領域をベース領域、質問 および話す内容に関する領域をターゲット領域と し,5件法による質問紙により,(1)各カード(ペア) における領域間の親密度評定. および(2)各ペアに おける類似度評定を行い、それらの結果をそれぞれ 比較した。

4.2 方法

4.2.1 協力者

まず、日本人高校2、3年生239名を対象に熟達

度テストを実施した。これらの生徒はすべて EFL 環境で英語を学んでいる。テストの内容は、英検準 2級および 2級筆記問題のうち、語彙、文法などに 関する多肢選択問題と文章理解問題を用い、計40問を出題した(M = 19.05; SD = 5.62)。この中から英 検準 2級以上の資格保持者(28名)の平均点を算出した(M = 24.65, SD = 5.27)。英検準 2級一次試験 合格者以上の熟達度の者を選ぶため、全受験者のうち資格保持者の平均点以上の得点であった者の中から、計57名の協力者を依頼した(M = 25.61, SD = 4.69)。これらの協力者のうち、調査 1、2 のいずれかの調査における実施上問題のあった 4 名のデータを除外し、最終的に53名を分析対象とした。

4.2.2 マテリアル

日本人高校生英語学習者の傾向を見るため、比較的平易であると考えられる英検準2級の面接問題を使用することとした。このテストで用いられているスピーキングテストには、ベース領域として読んだテキストの内容について答えるタスクと、読んだ内容に関連ないし類似した領域について自分の意見を述べるタスクの2つがある。前者のタスクはベース領域(テキスト)をもとに話す内容を準備できる一方、後者のタスクは話者の経験や一般知識のみに依存するのか、ベース領域から情報を選択して自分の知識と統合して応答をするのかについては、話者の方略によって異なることが予想される。

問題については、英検準2級過去6回全問題集(旺文社,2011a) および英検準2級二次試験・面接完全予想問題(旺文社,2011b) から23の問題を候補として検討した。まずこれらの問題において使用されているテキストについて、文の数、総語数、リーダビリティ(FRE/FKGL)について比較した(表2)。それぞれのテキストに含まれる文の数は2つを除きすべて4文であったため、3文のものと5文

■表 2:検討した全テキストにおける文の数,総語数、およびリーダビリティ

	М	SD
文の数	4.00	0.30
総語数	264.17	16.23
FRE	56.20	8.22
FKGL	8.60	1.21

(注) n = 23; M = 平均; SD = 標準偏差; FRE = Flesch Reading Ease; FKGL = Flesch Kincaid Grade Level.

■表 3:使用した各カードにおけるテキストの総語数、リーダビリティ、および質問の領域

カード	テキストの領域	質問文の領域	文の数	総語数	FRE	FKGL
Α	Recycling	Plastic bottles	4	259	51.2	9.2
В	Electric cars	Car navigation system	4	271	49.8	9.7
С	Sea travel	Traveling	4	257	52.0	9.2
D	SNS	On-line shopping	4	305	43.6	10.5

(注) FRE = Flesch Reading Ease; FKGL = Flesch Kincaid Grade Level; SNS = Social Networking Service.

■表4:各領域に関する領域と親密度評定の記述統計およびt検定の結果(n = 53)

	ベース		ターゲット			ベース―ターゲット			
カード	領域	М	SD	領域	М	SD	t	р	Δ
Α	Recycle	3.81	0.90	Plastic bottle	4.45	0.80	- 4.08	< .001	- 0.80
В	Car navigation	3.00	1.09	Electric car	2.30	1.08	3.51	.001	0.65
С	Sea travel	2.04	1.13	Travel	3.96	1.09	- 11.32	< .001	<u>- 1.76</u>
D	SNS	3.70	1.25	On-line shopping	3.68	1.27	0.09	.925	0.02

(注) M = 平均;SD = 標準偏差; $\Delta = M$ (ベース)-M(ターゲット)/ SD(ターゲット);下線の数字は効果量が|.50| を超えているもの。

のものは除外した。総語数およびリーダビリティを 比較し、同程度のテキストを4つ選定した(表3)。

4.2.3 調査方法

A~Dの各カードにおけるテキストと質問内容において扱われている各領域に関する親密度(計 8 領域)と、それぞれのカードにおける領域間の類似度の評定(計 4 ペア)に関して、質問紙を用い 5 件法による調査を行った。実際の調査におけるトピックおよび質問項目については資料 1 を参照のこと。親密度についてはベース領域とターゲット領域の各領域において、「身近にあるもの、詳しく知っているもの」であるかどうかを尋ねた。また類似度に関しては、ベース・ターゲットの 2 領域間がどの程度類似または関係していると思うかを尋ねた。なお、調査 1 については、調査 2 におけるスピーキングパフォーマンスへの影響を避けるため、スピーキングテストの実施後である、2012年12月初旬に行われた。説明の都合上調査 2 より先に報告する。

4.2.4 分析方法

親密度評定については、それぞれのカードにおけるベース領域(テキスト内容における領域)とターゲット領域(口頭質問の領域)をt検定(繰り返しあり)により比較した。また、類似度評定については、一要因分散分析(被験者間)を用い、カード間の類似度の差を比較した。

4.3 結果と考察

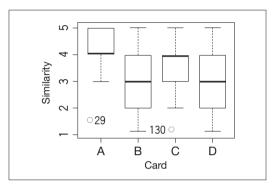
表 4 は、親密度評定の記述統計および、ベースとターゲット間の親密度の記述統計および t 検定の結果である。検定の結果、カード A、B、C において、ベースとターゲットの間に有意差が見られた。カード D において有意差は見られなかった。有意差のあった 3つのカードのうち、ベース領域の親密度が有意に高いのはカード B だけであった。カード A およびカード C は、ともにターゲット領域の親密度がベース領域よりも有意に高く、特にカード C においてはその差が最も大きかった。

次に類似度評定の一要因分散分析(被験者間要 因)の結果である。記述統計および多重比較の結果 については表5に示す。被験者間1要因分散分析の 結果, $F(3, 156) = 17.78, p < .000, \eta^2 = .07$ であり, ペア間に類似度の差が見られた。効果量は中で あった。ボンフェローニの修正による多重比較の結 果. カードAにおいてベース領域とターゲット領 域の類似が他の3つのカードよりも高く評定されて いた。次に評定が高かったのはカードCである。 カードBとカードDにおける類似度評定の間に差 は見られなかった (カードA>カードC>カード B = カード D)。この結果をグラフに表したのが図 2である。グラフにおいては外れ値が2つ見られた ため、これら2つのケースを除いて再分析を行った が、分散分析および多重比較の結果などにおいて結 果には大きな違いは見られなかった。

■表 5:類似度評定の記述統計および一要因分散が	ł
析の多重比較結果	

カード	平均	Α	В	С
Α	4.30 (0.72)	-		
В	3.15 (1.15)	.000 **		
С	3.62 (0.90)	.000 **	.033 *	
D	3.30 (1.22)	.000 **	1.000	.466

(注) n = 53; 括弧内の数値は標準偏差; ボンフェロー 二の修正による計算結果; ** < .01; * < .05



(注) n = 53; Similarty = 類似度評定; Card = カード 図 2:各カードにおける類似度評定の箱ひげ図

4.4 調査 1 のまとめ

調査1の分析において見られたカードごとの傾向 について表6にまとめた。この結果からわかるの は、まず、カードAおよびカードCはベース領域 よりもターゲット領域の親密度が高く, いずれも2 つの領域が類似していると評定されている点であ る。ただしその類似の度合いに差がある。この他, カード B およびカード C においてはそれぞれべー ス領域である Sea travel や SNS の親密度が他の領 域に比べ低く評定されていた F(3, 156) = 35.29, p < $.05, n^2 = .35$ (D - FA = D - FD > D - FB = ...カードC; すべてp < .01)。効果量は大であった。 特にカード C においては、ベース領域の親密度が 低いことがターゲット領域との差をもたらした可能 性も考えられる。これに対し、カードAにおける ベース領域は他のカードよりも高く評定されてい た。つまり、ベースそのものに対する親密度が高く 評定されていたのだが、さらにターゲット領域はそ れよりも高い評定を得ている。つまりベース、ター ゲットともに親密度の高い領域であったと言える。 この他、カードBのみに見られる特徴として、ベー ス領域の親密度がターゲット領域の親密度よりも高 いことが挙げられる。また、カード D については、 親密度の差は見られず、ベース領域とターゲット領域は類似していないと評定された。これらの特徴が どのように実際のパフォーマンスにおいて影響する のかについて、次の調査 2 においてさらに検証する。

■表6:調査1のまとめ

カード	親密度	類似度
А	B < T	0
В	B > T	—
С	B < T	0
D	B = T	

(注) B = ベース領域; T = ターゲット領域; ◎ = 最も高い; ○ = 高い

5 調査 2

5.1 調査2の目的

調査2の目的は、オブジェクト使用の有無 (OBJ) によるパフォーマンスへの影響に関する検証を行うことである。調査1の親密度評定の結果から、タスクにおいて使用されている2領域間においても、実際には親密度、類似度の評定の結果が異なることが示唆された。調査2では、これらの違いが実際のスピーキングパフォーマンスにどのような影響を与えるのかについて調べることを目的とする。

5.2 方法

5.2.1 協力者

調査2においては、調査1と同一の、53名の日本 人高校2、3年生が参加した。

5.2.2 マテリアル

調査1に用いたAからDの4枚のカードおよび、それらに対応する質問項目を用いた。質問項目については、テキストに関連した領域について意見を述べる質問(関連質問)および関連しない領域について意見を述べる質問(関連なし質問)の2つを用いた。各カードの質問におけるベース領域からターゲット領域への情報の転移を見るため、Gentner(1983)を参考にして2名の評価者による各カードにおけるオブジェクトの抽出を行った。相違点は抽出後の話し合いにより解消した。各質問項目、抽出

された実際のオブジェクトおよび用いた文章については資料2 および資料3 を参照のこと。

5.2.3 インタビュー調査の手順

インタビューテストは2回に分けて実施された。 第1回(カードA、B) に関しては2012年7月に、 第2回(カードC, D) に関しては9月中旬から10 月中旬にかけてそれぞれ行った。面接者による差を 排除するため、面接は筆者1名により実施した。実 験手順は英検の面接試験に準じた(図3)。まずテ ストの概要について説明を行い、その後口頭による 英語のインタビューを開始した。英語でのインタ ビューは、最初に名前の確認と挨拶や平易な会話を 1分程度行った。次に、ベース文となるテキストの 提示および黙読(20秒). 音読. テキスト内容に関 する質問、自分の意見を述べる質問(テキスト関連 あり), 自分の意見を述べる質問 (テキスト関連な し), の順に行った。1度のテストにつき2枚のカー ドを順に使用した。提示順についてはカードA-B 間. C-D 間でそれぞれカウンターバランスを取っ た。すべての質問が終了した後、それぞれのテキス トや質問に関する理解などについての感想を、日本 語で自由に述べてもらった。これらのインタ ビューの様子を、IC レコーダに録音した。協力者 には、実施後に謝礼として1人あたり350円相当の 菓子および飲料が贈られた。

5.2.4 採点および分析方法

録音データのうち、「自分の意見を述べる質問」における発話を書き起こした後、データベースソフト(Filemaker Pro Advanced ver. 12)に入力した。この発話について、評価者 2 名(筆者および英文学科卒業の英語指導者 1 名)によりデータ読み取りおよび採点を行った。不一致点は話し合いにより解消した。

分析については、3つに分けて行った。まず2つの質問タイプ(関連質問および関連なし質問)におけるオブジェクト使用の有無の違いについて、ウィルコクソンの順位和検定を用いて分析を行った。

次に、関連質問におけるオブジェクト使用とパフォーマンスとの関係について調べるため、設問ごとに指標を算出した。指標についてはスピーキングパフォーマンスに関する先行研究(Koizumi, 2005; Koizumi & Katagiri, 2007)を参考にして設定した。

具体的には、① 熟達度テスト点、② 総語数 (token), ③ AS ユニット数, ④ 節の数, ⑤ 節当た りの使用語数。⑥ 発話におけるオブジェクトの使 用有無 (OBJ), ⑦ 理由の適切さ, の7つである。 (1)の熟達度テスト点については、調査1で用いた熟 達度テストの得点をもとに、Linacre and Wright (2006) および平井(2010) を参考にし、ソフトウェ ア BIGSTEPS (ver.2.82) による項目応答理論 (IRT) を用いた分析を行い、その結果算出されたロジット 値を用いた。IRTとは、「受験者がテスト項目に応 答(正誤)していくパターンに基づいて、その受験 者の能力レベル (認識能力,知識,技能など)を確 率モデルに基づいて推定しようとする理論」(平井, p.78) である。② 総語数 (token) および ③ AS ユ ニット数. ④ 節の数. の3つについては. 発話量 の指標として用いた。AS ユニットの分割について は, Foster, Tonkyn, and Wigglesworth (2000) を参 考に行った。発話の複雑さの指標として、⑤節当 たりの使用語数を用いた。また、パッセージと回答 の類似の指標として、⑥ OBJ を用いた。パフォー マンスの内容に関する指標としては、⑦ 理由の適 切さを設定した。理由の適切さについては、2つ目 の質問である「自分の意見を述べる質問」において 理由を適切に答えているかどうかを、2件法により 2名の評価者が判定した。これらの結果については 評価者間信頼性を確認した (κ = .796)。「理由の適 切さ」については解釈の異なった点が見られたた め、話し合いにより解消した。各指標を算出した後 に、カードごとにスピアマンの相関分析を行った。 これらの結果については、調査1の結果を踏まえて 考察を加えた。最後に、2つのレベルの類似(一次 の類似関係;高次の類似関係)と関連質問における 理由の説明との関係について、フィッシャーの正確 確率検定を用いて各カードごとに分析し、実際の発



▶ 図 3:インタビューテストの概要

話と対比させながら解釈を行った。

5.3 結果と考察

5.3.1 関連質問におけるオブジェクトの使用(分析 1)

問題タイプによる OBJ を調べるため,関連質問 (4問) と関連なし質問 (4問) において,回答における 1 つ以上のオブジェクト使用の有無をカウントし比較した。記述統計については表 7 に記す。関連なし質問についてはオブジェクトを使用した回答の数が非常に少なく (M = 0.06) 分散の正規性が確保できなかったため,ウィルコクソンの符号順位和検定を用いた。検定の結果,z = -5.91,漸近有意確立p < .001 で有意差が見られた。この結果,関連質問において,関連なし質問に比べて有意にオブジェクトを多く使用していることがわかった(表 8)。

■表7:関連質問と関連なし質問におけるウィルコクソンの順位和検定の記述統計

質問	М	SD	Min.	Мах.
関連なし	0.06	0.23	0	1
関連あり	1.42	0.86	0	4

(注) n = 53; M = 平均; SD = 標準偏差; Min. = 最小值; Max. = 最大值

5.3.2 オブジェクト使用のスピーキングパフォーマンスへの影響(分析 2)

分析1の内容から、関連質問においてオブジェクトが有意に多く使用されていることが明らかになった。続く分析2では、これら関連質問におけるOBJがパフォーマンスにどのような影響を与えているかについて調べるため、4.2.4で述べた各指標に関してスピアマンの順位相関分析を行った。分析はカードごとに分けて行われた。各カードにおける指標間の相関係数および有意差のあったペアについては、表9のとおりである。

すべての指標について、OBJ と各指標との関係を中心に解釈を行った。その結果、次のようなことが明らかになった。

全カードに共通して、OBJ と熟達度の間には大 きな相関が見られなかった。ただし、カードBに おいては最も低く負の値であった。また、OBJと 総語数との間については、カードBが最も低く、 それ以外のすべてのカードにおいては有意な正の相 関が見られた。OBJ と AS ユニット数の間につい ては、カードA および C においてのみ正の相関が 見られた。オブジェクト使用と節の数についても カードBが最も低く負の値であり、それ以外のす べてのカードにおいて有意な正の相関が見られた。 このように、これらの結果は類似度評定の結果と同 様の傾向を示している。類似度評定の高かった カードにおいては OBJ と各指標との相関が有意に 高かった一方で、類似評定の最も低かったカードB においては OBJ と各指標間の相関が低く、熟達度 や理由の説明の適切さなど負の値のものが見られ た。

その他、OBJと節当たりの語数(複雑さの指標)との間においてはカード C および D においてのみ正の相関が見られた一方、カード B においては最も低く負の値(-0.04)であった。また、OBJと理由の適切さの間についてはカード A において正の相関が見られた一方、カード B では負の相関(-0.38)が見られた。カード C および D については有意な相関は見られなかった。OBJと複雑さ、理由の適切さとの相関においても、やはりカード B が共通して低かった。この結果もまた類似度評定の結果と同様の傾向であると言える。

これらについては調査1の親密度および類似度評定の結果と合わせ、さらに考察を加えて後述する。また、OBJと理由の適切さの間の関係については、次の分析3においてさらに詳しく調べる。

■表8:関連質問と関連なし質問におけるウィルコクソンの順位和検定の結果

		N	平均ランク	順位和
	負の順位	0	.00	0.00
関連あり質問 – 関連なし質問	正の順位	44	22.50	990.00
	同順位	9		
	合計	53		

■表9:各カードにおける指標間のスピアマン相関係数一覧(自分の意見を述べる質問)

■ X 3 · □ //	1 10 43	11 O 15 12 15		, > 1HIX	기사 있 듯	()) •//	37L C 21		
指標	1	2	3	4	5	6	7	М	SD
カードA									
1.熟達度								0.76	0.86
2. 総語数	.46**							8.55	5.85
3. AS	.24	.72**						2.09	0.99
4.節の数	.25	.79**	.83**					2.38	1.13
5.語数/節	.42**	.84 * *	.38**	.34*				3.42	1.42
6. OBJ	.01	.41 * *	.58**	.44**	.27			0.66	0.48
7.理由	.25	.48**	.52**	.41 * *	.34*	.48**		0.60	0.49
カードB									
1.熟達度								0.76	0.86
2.総語数	.26							8.92	6.49
3. AS	.14	.88**						2.21	1.03
4.節の数	.14	.84 * *	.86**					2.57	1.42
5.語数/節	.35 * *	.83**	.56**	.39 * *				3.28	1.40
6. OBJ	.13	.20	.29*	.24	04			0.08	0.27
7.理由	05	.50 * *	.47**	.46**	.44 * *	38 * *		0.64	0.48
カードC									
1.熟達度								0.76	0.86
2. 総語数	.20							9.85	6.37
3. AS	.15	.90 * *						2.43	1.10
4.節の数	.21	.92 * *	.89**					2.83	1.34
5.語数/節	.12	.76**	.57**	.47**				3.27	1.04
6. OBJ	.04	.38 * *	.28*	.31 *	.37 * *			0.17	0.38
7.理由	.16	.49**	.58**	.52**	.24	.16		0.74	0.45
カードD									
1.熟達度								0.76	0.86
2.総語数	.33*							9.96	6.24
3. AS	.18	.60**						2.62	1.00
4.節の数	.21	.62 * *	.91 * *					3.00	1.41
5.語数/節	.30 *	.35 * *	.20	.18				3.91	1.23
6. OBJ	.18	.37 * *	.22	.31 *	.05			0.51	0.50
7.理由	.38**	.12	.23	.27*	0.29*	.01		0.89	0.32

(注) n = 53. *p < .05, ** p < .01; M = 平均; SD = 標準偏差; AS = AS ユニット数; 語数/節 = 節当たり の使用語数; OBJ = オブジェクト使用の有無; 理由 = 適切な理由を述べているか否か

5.3.3 オブジェクト使用のレベルと理由の 適切さに関する分析(分析 3)

分析2の結果から、OBJは、発話量にかかわる 指標や適切な理由の説明などと有意な相関があると 考えられる。また、これらの相関分析の結果はカー ド間の類似度評定の結果と一致し、特に類似度評定 の低いカードBにおいては異なった傾向が見られ た。 今回のマテリアルにおけるベース領域はターゲット領域と関連したことを質問しているため、いずれも何らかの形でオブジェクトの類似を想定していると考えられる。しかしながら、関連質問における発話が必ずしもベースに類似しているものになるわけではなく、またその類似のレベルも異なる。Gentner et al. (1993) や Halpen et al. (1990) などの先行研究におけるオブジェクト使用においては、

その使用した数によって異なった類似のレベルが想 定されていたが、このレベルの違いも理由の説明に 影響を与えるのではないかと考えられる。

分析3においては、それぞれの発話におけるオブジェクトの使用数に基づき、(1)オブジェクト使用なし;(2)一次オブジェクト使用(オブジェクト1つ);(3)高次オブジェクト使用(複数のオブジェクトを使用)の3つのレベルに分類した。さらに、この3つのレベルと説明における適切な理由の有無についての関係について、フィッシャーの正確確率検定を用いて比較した。この分析結果については、表10にまとめた。

この結果、カードAの関連質問においては、オブジェクトを使用しなかった場合に適切な理由を説明した者の数が有意に少なかった一方で、高次のオブジェクトを使用した場合適切な理由を説明した者が有意に多かった(いずれもp<.01)。一次のオブジェクト使用については、有意差は見られなかった。カードAにおける適切な理由の説明を見ると、環境(environment)、ペットボトル(plastic bottle)、リサイクル(recycle)の3つのオブジェクトを用いている例が多く、インタビュー後の聞き取りにおいても、「プラスチックの、あれ、ペットボトル。environmentって本文に出てきて、これかって」など、ベース文の情報を参考にして答えを導き出したと考えられるコメントが見られた。このことから、構造的な類似を適用して発話を導き出している例が

あったと考えられる。実際の発話の例を図4に示す。このうち、plastic bottle は、質問文とベース文 双方に共通して含まれているオブジェクトであったため、内容の関連づけを促した結果理由の説明を行いやすくしたのではないかと考えられる。

続いて、類似度評定の最も低かったカード Bの 関連質問である。このカードにおいては、オブジェ クトを使用しなかった場合に適切な理由を説明した 者の数が有意に多かった (p < .01)。その一方で、 一次、高次にかかわらず、オブジェクトを使用した 場合には適切な理由の説明を行うことができな かった者が有意に多かった(いずれもp<.05)。こ れらオブジェクトを利用した発話の例を図5に示し た。これらの者は、いずれもベース文の中で述べて いるカーナビの話を適用して説明を試みている。し かしながら、質問は電気自動車についての意見を求 めているため、答えはいずれも的外れなものになっ てしまっているため、 適切な理由の説明にならな かったと考えられる。事後のコメントにおいても、 「電気自動車ですか? あれ、あ、そっか。ナビだ と思ってました。あそっか、あ、そうだ。ナビだと 思って答えてました。ナビの話を読んだので」とい うコメントがあったことなどを考えると、一次の表 面的な類似から問われている内容を推測して答えた 結果、的外れな答えになってしまったものと考えら

類似度評定の結果が中間に位置した2つのカード

■表10:関連質問におけるオブジェクトの使用と適切な理由の説明

	オブジェクト使用	一次オブジェクト	高次オブジェクト
	なし	使用	使用
カードA			
適切な理由なし	9 (-2.79) **	5 (-0.18)	0 (-2.74) **
適切な理由あり	9 (-2.79) **	15 (-0.18)	15 (- 2.74) **
カードB			
適切な理由なし	14 (-2.90) **	2 (-2.01) *	2 (-2.01) *
適切な理由あり	35 (-2.90) **	0 (-2.01) *	0 (-2.01) *
カードC			
適切な理由なし	10 (-1.59)	0 (-1.48)	0 (-0.49)
適切な理由あり	34 (-1.59)	8 (-1.48)	1 (- 0.49)
カードD			
適切な理由なし	2 (-0.63)	1 (-0.49)	0 (- 0.35)
適切な理由あり	24 (-0.63)	24 (-0.49)	2 (- 0.04)

(注) *<.05; ** < .01

{Could you say that again, please? (Rep. Q)} | Yes. | {Because,} because people :: thinking about the *environment* .. :: will be increase in the future. | And my *families* try :: to *recycle* the *pet bottle* and *can*. | (ST-1)

| Yes. | Because, { mmm... } today's *environment* is most remarkable ... people. | I think :: {pet bottle is ... } *recycling pet bottle* is important things for me. | (ST-2)

- (注)太字斜字体はオブジェクトと判断した箇所。ST-1 = 生徒の例1。ST-2 = 生徒の例2。
- ▶ 図 4:Card-Aの関連質問 (Do you think drinks in plastic bottles will be more popular in the future? Why? / Why not?) におけるオブジェクト使用の例

Yes. ... | Navigation systems is useful. | (ST-3)

{ Yes. Electric car is mmm ... } (Rep. Q) | Yes. | {...} Electric cars have *navigation*. | {...} navigation is {... mmm ...} don't need *map*. | (ST-4)

| Yes. | Because now we couldn't drive. | But if I drive, :: {err ...} I think :: I need a *navigation*, :: because { I ...} I couldn't remember {road ...} many roads, | so I think :: that's I need. | (ST-5)

- (注)太字斜字体はオブジェクトと判断した箇所。ST-3 = 生徒の例3。ST-4 = 生徒の例4。ST-5 = 生徒の例5。
- ▶ 図 5:Card-Bの関連質問 (Do you think more people will drive electric cars in the future? Why? / Why not?) におけるオブジェクト使用の例

(カ-FC およびカ-FD) の関連質問においては、 分析 2 における相関分析の結果同様、いずれもオブジェクトと理由の適切な説明に有意な関係は見られなかった。

5.4 調査 2 のまとめ

調査2においては、実際のパフォーマンスにおける OBJ がパフォーマンスにどのように影響を与えるかについて、適切な理由の説明との関係を中心に分析を行った。その結果、次のようなことがわかった。

分析 1 においては、それぞれの文章における 2 つの質問(関連質問と関連なし質問)において、OBJ について調べた。この結果、関連質問において文章中のオブジェクトを有意に多く使用していることがわかった。この結果は、関連質問を使用したねらいどおりの結果であり、特に驚くべき結果ではない。

続く分析2では、このOBJが、各カードにおける関連質問におけるパフォーマンスに与える影響について調べた。OBJによって、Bのカードを除いた多くのカードにおいて総語数やASユニット数、節の数など発話量に関する指標との正の相関が見られた。これらのカードだけを見ると、文章はパフォーマンスを量的に促進するベース知識となっていると考えることができる。しかしながら、複雑さの指標となる節当たりの語数について法則性は見られず、またカードBでは量的な有意な指標においても相関が見られなかった。理由の説明との相関に

おいてカードA(有意な正の相関)とカードB(有意な負の相関)という全く反対の結果が見られたことは、注目すべき結果であると言える。

この結果について、さらに分析3においてオブジェクトの使用を一次および高次に分類して分析を行った結果、特にカードAにおいては高次のオブジェクトを使用した場合に、適切な理由の説明ができた者が多かった一方で、カードBでは、一次、高次にかかわらず、オブジェクトを使用した場合には説明理由が適切ではないことがわかった。カードCおよびDにおいては、相関分析同様、オブシェクトの使用の理由の説明への影響は見られなかった。これはあくまでも推測であるが、カードCおよびDにおいてはいずれもカードA、Bよりも理由の説明の平均値が高く、OBJと理由の説明との相関が低いことを考えると、オブジェクトを使用しなくとも答えることが容易な設問であったと考えられる。

次に、これらそれぞれのカードにおける違いについて、調査1における親密度および類似度評定の結果と合わせて考察を加える。

6

総合考察

本研究においては、「読むー話す」型の統合タスクにおいて、文章内容と質問内容の類似の観点から、そのパフォーマンスに与える影響について、2つの

調査により検証を行った。これらの調査の結果、カードのトピックにより、オブジェクトがパフォーマンスに与える影響は異なっていた。先行研究においてもトピックの違いによってパフォーマンス結果が異なることが指摘されている。まずこの点について、調査1における親密度および類似度評定結果と比較して述べたい。

最も特徴の見られたのはオブジェクト使用と適切 な理由の説明との関係である。類似度の高い組み合 わせ (カードA) の場合、オブジェクトを使用する ことによってこの理由の適切な説明が促進されたと 考えることができる。その一方で、類似度の低い組 み合わせである (カードB) にもオブジェクトを使 用した例が見られた。ただしこれらの情報は誤った 理解につながる理由の説明に用いられている例が見 られた。情報の転移が適切に行われるには、ベース とターゲットは高次の類似をしている必要があるこ とが調査2(分析3)の結果からわかる。カードB は、親密度評定において唯一ベースがターゲットよ りも低く評定されていた。このような組み合わせの 場合. なじみのない内容について問われたときに. よりなじみのある内容から情報を得ようとしたと考 えられる。外国語で理由の説明を求められた際は十 分に質問内容を理解できないことも考えられること から、既に知っている内容から推測して説明を試み たとしても不思議ではない。そのため、電気自動車 とカーナビのように高次では類似していないが表面 的に類似している領域を参考にしてしまった場合 に、不適切な転移を行い誤った説明につながったと 考えられる。このことがオブジェクトの摘要が適切 な理由の説明を阻害するという結果につながったも のであろう。

本研究では、ベース領域を表面的である一次的な類似としてベースをとらえた場合と構造的な類似としてとらえた場合では、ターゲット領域として仮定したタスクパフォーマンス結果が異なる可能性が示唆された。統合タスクは、インプットとアウトプット間でそれぞれ異なった領域を扱う可能性があるが、これらの領域の類似関係を、そのレベルを含め十分に検討する必要がある。特にインプット(ベース)領域の情報を適切にアウトプット(ターゲット)領域に摘要されるためには、表面的な一次の類似ではなく、高次のレベルで類似している2領域を扱うことが必要である。

本研究で扱った英検の二次試験に見られるように、統合タスクを用いた複数の問題を併用する場合には、文章領域や質問領域など、扱う領域(トピック)が話者にとってどの程度なじみのあるものなのかを考慮するとともに、可能な限りで複数の問題を課すなどの工夫をすることも重要である。また、実際の授業やテストにおいて統合タスクを用いる場合においても、表面的にだけ似ているものをベースとして提示した場合には、続くターゲット領域を間違って理解してしまう可能性があるため、質問している内容領域との関係レベルでの類似があるのかどうかについて、十分に検討する必要がある。

最後に本研究の限界について述べる。本研究で は、類似レベルを想定した統合タスクの検討を行っ たが、過去に同様の先行研究がなかったことから、 探索的手法として、オブジェクト数を類似の指標と して用いることで分析を進めた。類似した領域のオ ブジェクトを新しい領域に用いてパフォーマンスを 行った者の数は非常に少なかった。そのため、今回 はノンパラメトリックな分析手法や相関分析などを 用いて分析を行った。今後、統合タスクにおいてい かに類似の利用を促すとともに、より多くのサンプ ル数を確保することでさらに詳細な分析が可能にな り、さらにその因果関係を探ることが可能になると 考えられる。また、第1回のインタビューテストで のカードA. B間と異なり、第2回のインタビュー テストで行ったカード C. および D においては. 類似の影響によるパフォーマンスへの影響の違いが 見られなかった。この原因としては、あくまでも推 測ではあるが、それぞれのカードにおけるトピック の違いの他, 1回目にカードA, Bに関する感想な どの聞き取りを行った際にベースとの類似について 気づいた可能性があったこと、カード C. D は 2 回 目のインタビューであったためテスト形式に慣れた ため、比較的落ち着いて考えることができたことな どが考えられる。

以上の問題点を解消し今後さらに同様の研究を行うことで、統合タスクにおける類似とパフォーマンスの関係が明らかになるものと考えられる。この他、「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」(名畑目、2012)として、Latent Semantic Analysis(Dennis、2007)などの手法も注目されており、類似の数値化を検討する手法として

の利用可能性が期待できる。今後統合タスクを類似 という視点でとらえた研究や実践が現れることを期 待したい。

謝辞

本研究を行う機会を与えてくださった公益財団法 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。特に本研究 を担当していただいた吉田研作先生には、有益なご 助言ならびにご指導をいただきました。心より感謝申し上げます。また、筑波大学大学院の卯城祐司先生ならびに平井明代先生には、大学院在籍時より継続的にご指導をいただきました。深く感謝いたします。また、スピーキングパフォーマンスのリサーチデザインおよび分析に関して順天堂大学の小泉利恵先生に多くのご助言と御教示をいただきました。この場を借りて、御礼の言葉を贈ります。

- *Bachman, L., & Palmer, A. (1996). Language testing in practice. Oxford: UK. Oxford University Press.
- *Bachman, L., & Palmer, A. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford: UK. Oxford University Press.
- *Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal, 1*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.0026-7902.2005.00264.x/abstract
- * Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tests. TOEFL Report, 29.
- *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.55-70). Mahwah, NJ: Laurence Erbium Associates.
- *English Testing Service. (2006). 『TOEIC® スピーキングテスト/ライティングテスト公式ガイド』. 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. (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. (2013). For test takers.
 Retrieved January 5, 2013 from
 http://www.ets.org/toefl/ibt/about/content/
- *Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. Applied Linguistics. 21, 354-375. Retrieved from http://applij.oxfordjournals.org/content/21/3/354. short
- *Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing*, 29(3), 345-369. doi:10.1177/0265532211424479
- *Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155-170. doi:10.1016/S0364-0213(83)80009-3

*Gentner, D., Ratterman, M.J., & Forbus, K.J. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive* psychology, 25, 524-575. Retrieved from http://groups.psych.northwestern.edu/gentner/ papers/GentnerRattForbus93.pdf

- *Halpen, D.F., Hansen, C., & Riefer, D. (1990).
 Analogies as an aid to understanding and memory.
 Journal of Educational Psychology, 82, 298-305.
- *Hammadou, J. (1990). The effects of analogy on French reading comprehension. *The French Review, 64*(2), 239-252. Retrieved from http://www. jstor.org/stable/10.2307/395869
- *Hammadou, J. (2000). The impact of analogy and content knowledge on reading comprehension: What helps, what hurts. *The Modern Language Journal*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/0026-7902.00051/abstract
- *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.
- *Hirai, A., & Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly*, 6, 151-167. doi: 10.1080/15434300902801925
- *Koizumi, R. (2005). Predicting speaking ability from vocabulary knowledge. *JLTA Journal*. Retrieved from http://www7b.biglobe.ne.jp/~koizumi/JLTA2005_Koizumi_Predicting_speaking_from_voc.pdf
- *Koizumi, R., & Katagiri, K. (2007). Changes in Speaking Performance of Japanese High School Students: The Case of an English Course at a SELHi. ARELE: Annual Review of English Language Education in Japan, 18, 81-90. Retrieved from http://ci.nii.ac.jp/naid/110008512322/
- *公益財団法人日本英語検定協会.(2013). Retrieved January 7, 2013 from http://www.eiken.or.jp/about/basis.html
- *Lee, Y.-W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing*, 23, 131-

166. doi:10.1191/0265532206lt325oa

- *Linacre, J.M., & Wright, B.D. (2006). A user's guide to bigsteps. Retrieved August 22, 2012, from http://www.winsteps.com/a/bigsteps.pdf.
- * Luoma, S. (2004). Assessing speaking. Cambridge: UK. Cambridge University Press.
- *文部科学省.(2009a).『高等学校学習指導要領』.
- *文部科学省.(2009b).『高等学校学習指導要領解説外国語編英語編』.
- *名畑目真吾.(2012). [Latent semantic analysis(LSA) による空所補充型読解テストの解明―文レベルの意味的関連度を観点として―]. STEP BULLETIN, vol.24, 42-58.
- *旺文社.(2011a).『英検準 2 級過去 6 回全問題集』. 東京: 旺文社.
- *旺文社.(2011b). 『英検準 2 級二次試験・面接完全予想問題』. 東京: 旺文社.

Radden, G., & Dirven, R. (2007). Cognitive English

- grammar. Amsterdam, Netherlands: John Benjamins Publishing Company.
- *Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning*, 33, 527-540.
- *鈴木宏昭.(1996).『類似と思考』東京:共立出版.
- *山﨑晃男.(2001). 「アナロジー」. 森敏昭(編著) 『おもしろ思考のラボラトリー』.(pp.15-33). 東京: 北大路書房.
- *Yano. K. (2011). The effects of visual aids and analogies on Japanese EFL reading comprehension. (Unpublished master's thesis). University of Tsukuba, Ibaraki: Japan.
- *Zook, K. (1991). Effects of analogical processes on learning and misrepresentation. *Educational Psychology Review, 3*. Retrieved from http://www. springerlink.com/index/M79825R577812532.pdf

資料 1:英検面接試験のトピックにかかわるアンケート

英検面接試験のトピックにかかわるアンケート

調査対象者:()年()組()番 氏名()さん

【このアンケートについての説明】

- このアンケートは、以前に英検面接試験調査に協力してくれた生徒を対象としたものです。
- 面接に使用されたトピックについて、以下の質問に答えてください。
- 答えに際しては、他の人と相談はせず、自分の考えに基づいて答えてください。
- 内容や氏名等の個人情報等は外部に公開されることはなく、あくまでも研究目的でのみ使用されます。

【調査1:トピックの親密度に関する質問】

次のそれぞれのことがらについて、自分にどれだけ親しみがあるのかを以下の基準から答えてください。

基準: 1.全く親しみがない:全く知らない、全然身近にないもの(こと)

2. どちらかというとあまり親しみがない: あまりよく知らないもの(こと)

3. どちらとも言えない

4. どちらかというと親しみがある:よく知っているもの(こと)

5. **とても親しみがある**: 詳しく知っている, とても身近なもの(こと)

(問 1) recycling (リサイクルすること)

1 2 3 4 5 全く どちらかというと どちらとも どちらかというと とても

親しみがない あまり親しみがない 言えない 親しみがある 親しみがある

(以下は回答欄および質問項目を省略しトピックのみ示す)

- (問 2) car navigation system (カーナビ)
- (問3) sea travel (海(船)での旅)
- (問 4) Social Networking Services (SNS: twitter, facebook, mixi などのコミュニケーションツール)
- (問 5) drinks in plastic bottles (ペットボトル飲料)
- (問 6) electric cars (電気自動車)
- (問 7) traveling (旅行)
- (問 8) on-line shopping (オンラインショッピング)

回答

【調査2:トピック間の類似】

次の 2 つの組み合わせは、似ている(または関係がある)と思いますか? 1から 5 の数字で答えてください。

 基準: 1.全く関係がない。
 全く似ていない

 2.あまり関係がない。
 あまり似ていない

3. どちらとも言えない

4. <u>どちらかというと関係がある</u>,すこし似ている5. とても関係がある,とてもよく似ている

(問9) recycling と drinks in plastic bottles (リサイクルとペットボトル飲料)

 1
 2
 3
 4
 5
 回答

 全く
 あまり
 どちらとも
 どちらかというと
 とても

 関係がない
 買係がある
 関係がある

(以下は回答欄および質問項目を省略しトピックの組み合わせのみ示す)

- (問10) electric car と car navigation system (電気自動車とカーナビ)
- (問11) sea travelと traveling (「船での旅行」と「旅行全般」)
- (問 12) on-line shopping と Social Network System (オンラインショッピングと SNS)

【その他】

面接試験の内容や質問項目に関して感じたことを自由にお書きください。

ありがとうございました。

資料 2:各カードにおいて使用したテキストと抽出したオブジェクトの一覧

カード	テキストの領域に関連した質問	テキストの領域に関連のない質問
А	"Do you think drinks in plastic bottles will be more popular in the future?" Why? (Why not?)	"There are many different kinds of housework to do. Do you do any housework?" Please tell me more. (Why not?)
В	"Do you think more people will drive electric cars in the future?" Why? (Why not?)	"There are a lot of foreign visitors to Japan. Would you like to show your town to them?" Please tell me more. (Why not?)
С	"Do you think traveling is a good way to relax?" Why? (Why not?)	"Today, many people use portable music players. Do you often listen to music on a portable music player?" Please tell me more. (Why not?)
D	"Do you think more people will do on-line shopping in the future?" Why? (Why not?)	There are many recycling stores in Japan now. Do you like buying used things? Why? (Why not?)

資料 3:各カードにおける質問項目

カード	テキスト	オブジェクト
А	Recycling These days, recycling is becoming common in people's daily lives. Most towns and cities collect cans and plastic bottles for recycling. Many families use products made from recycled paper, and in this way, they try to make the environment better. It is becoming more important to take care of the environment.	recycling, daily lives, towns, cities, cans, plastic bottles, families, products, recycled paper, environment
В	Car Navigation Systems People usually take a map with them when they visit unfamiliar places. These days, however, many cars have navigation systems. These systems show drivers where they are and how to reach their destinations. Drivers can even hear directions from the systems, so they can drive safely and comfortably without looking at a map.	map, unfamiliar places, car navigation systems, drivers, destinations, directions
С	Sea Travel Most people travel to foreign countries by air because it's the fastest way. These days, however, sea travel is becoming more popular. These are many luxurious cruise ships that are of almost the same quality as high-class hotels. Travelers enjoy many activities on the ships, so they can relax and have fun.	travel to foreign countries by air, sea travel, luxurious cruise ships, high-class hotels, travelers, many activities, ships
D	Social Networking Services Recently, new services for communication have started on the internet. They are called social networking services. With these services, people can exchange opinions and information with each other, and this makes making new friends easier. Social networking services are drawing more attention as a new tool of communication.	services for communication, internet, social networking services, people, opinions, information, new friends, attention, new tool of communication