

技能別及び総合的英語能力を推定する語彙テストの開発：

サイズ、構成、認知速度の融合

Development of a Vocabulary Test Battery Estimating English  
Skills and Proficiency:  
Integrating Vocabulary Size, Organization, and Access Speed

平成19-21年度科学研究費補助金

基盤研究（B）研究成果報告書

研究課題番号 19320084

研究代表者 望月正道

麗澤大学外国語学部教授

平成22年（2010年）3月

## はしがき

本書は、平成 19-21 年度科学研究費補助金基盤研究 (B) (研究代表者 望月正道: 研究課題番号 19320084) 「技能別及び総合的英語能力を推定する語彙テストの開発: サイズ、構成、認知速度の融合」の研究成果報告書である。

本研究は、研究テーマが示す通り、語彙知識を 3 つの次元から測定し、英語習熟度や技能を推定しようという試みである。語彙知識の 3 つの次元を測定するテストを開発し、その妥当性を検証する。次に、それぞれの語彙テストで英語習熟度や技能をどの程度推定できるかを分析しようとするものである。

研究組織は当初研究代表者のほかに 8 名の研究分担者と 1 名の研究協力者で開始したが、諸事情により平成 20 年度からは研究分担者 2 名が抜けた。この研究メンバーで 3 年間に 10 回のプロジェクト会議を開催した。会議では、研究の方向性、研究方法、進行状況、データの解釈などについて議論を重ねた。数多くの大学生諸君にはテストの受験およびデータの提供という協力をいただいた。また、研究途中での成果は、大学英語教育学会全国大会のシンポジウム、大学英語教育学会英語語彙研究会研究大会、本研独自の語彙研究フォーラムという場で公表してきた。これらの発表により参加者からの有益なフィードバックを得ることができ、それを元に研究を改良することができた。さらに、海外から語彙研究の第一人者であるスウォンジー大学の Paul Meara 博士と Tess Fitzpatrick 博士をそれぞれ平成 19 年度と 20 年度に招聘講演してもらい、研究への助言をいただいた。これも本研究に大きな影響を与えるものであった。これらの多数の方々のおかげで研究を行うことができたことを御礼申し上げる。

本研究報告書は、目指してきた語彙テスト開発過程とその成果を記すものである。解明できたことと同時に新たな疑問が生じてきたことも多いが、語彙知識とその運用能力と英語習熟度および技能の関係の解明の一助になれば幸いである。

平成 22 年 3 月

研究代表者 麗澤大学 望月正道

## 研究組織

研究代表者	望月正道 Mochizuki, Masamichi mmotizu@reitaku-u.ac.jp 第1章, 第2章2.1, 第3章3.1, 3.2.1, 3.3, 第6章担当	麗澤大学教授 Professor Reitaku University
研究分担者	村田年 Murata, Minoru murata.minoru@nifty.com	千葉大学名誉教授 (平成19年度) Professor Emeritus University of Chiba
研究分担者	上村俊彦 Uemura, Toshihiko yandi@sun.ac.jp 第2章2.3.2担当	長崎県立大学教授 Professor University of Nagasaki
研究分担者	相澤一美 Aizawa, Kazumi aizawa@cck.dendai.ac.jp 第4章4.1担当	東京電機大学教授 Professor Tokyo Denki University
研究分担者	投野由紀夫 Tono, Yukio y.tono@tufs.ac.jp	東京外国語大学准教授 (平成19年度) Associate Professor Tokyo University of Foreign Studies
研究分担者	杉森直樹 Sugimori, Naoki sugimori@mbox.kyoto-inet.or.jp 第5章5.2担当	立命館大学教授 Professor Ritsumeikan University
研究分担者	石川慎一郎 Ishikawa, Shin'ichiro iskwshin@kobe-u.ac.jp 第2章2.2, 2.3.1担当	神戸大学准教授 Associate Professor Kobe University
研究分担者	磯達夫 Iso, Tatsuo tiso@reitaku-u.ac.jp 第4章4.2, 4.3, 4.4担当	麗澤大学准教授 Associate Professor Reitaku University
研究分担者	小泉利恵 Koizumi, Rie rie-koizumi@mwa.biglobe.ne.jp 第3章3.2.2, 第5章5.1, 5.3担当	常磐大学専任講師 Lecturer Tokiwa University
研究協力者	清水伸一 Shimizu, Shin'ichi shim@sk.aitai.ne.jp	安城学園高等学校教諭 Teacher Anjo Gakuen High School

## 目次

はしがき

研究組織

### 第1章 概観

- 1.1 研究の目的
- 1.2 研究計画
- 1.3 報告書の構成

### 第2章 語彙サイズテスト

- 2.1 先行研究の概観
- 2.2 日本人大学生のための新しい語彙サイズテストの開発
- 2.3 語彙サイズテストは総合的英語力を推定しうるか
  - 2.3.1 第1調査の概要と結果
  - 2.3.2 第2調査の概要と結果

### 第3章 語彙構成テスト

- 3.1 先行研究の概観
- 3.2 語彙構成テストLOTの妥当性・信頼性検証
  - 3.2.1 WATによるLOTの妥当性検証
  - 3.2.2 プロトコル分析によるLOTの妥当性検証
- 3.3 LOTによるTOEICリーディング・リスニング得点の予測

### 第4章 アクセス速度測定テストの開発

- 4.1 先行研究の概観
- 4.2 アクセス速度測定テストの開発(1)
- 4.3 アクセス速度測定テストの開発(2)
- 4.4 結論

### 第5章 3種類のテストと英語熟達度とスキルの関係

- 5.1 英語熟達度全体の予測
  - 5.1.1 TOEICの予測
  - 5.1.2 CASECの予測

5.2 TOEIC Bridgeを用いた熟達度と語彙力との関係に関する調査

5.3 Versantを用いたスピーキング・リスニング力の予測

第6章 結論

謝辞

参考文献

付録

## 第1章 概観

### 1.1 研究の目的

言語能力モデルの中で語彙は、言語知識の多数ある構成要素のひとつとして位置づけられているにすぎないが (Bachman, 1990; Bachman & Palmer, 1996), 語彙力は言語能力を最もよく反映するものと考えられている。例えば, Alderson (2000)は, “Measures of readers’ vocabulary knowledge routinely correlate highly with measures of reading comprehension, and are often, indeed, the single best predictor of text comprehension.” (p.35)と述べ, 語彙知識が読解を予測する最善の指標であるとしている。

語彙知識は, 言語能力の一部にすぎないにもかかわらず, 言語習熟度を強く反映する事実から, 言語学習コースへの配置テストとして用いられている。ひとつの例は, *The Common European Framework of Reference for Languages* (Council of Europe, 2001)に基づく 14ヶ国語の外国語能力診断テスト (DIALANG)である。これは, 語彙サイズテスト, 自己診断テスト, 言語テストという 3つのテストから構成されている。受験者は, まず語彙サイズテストを受験し, その言語のおおよその習熟度が測定される。その測定値に基づき, 学習者に最も適したレベルの言語テストを受験する仕組みになっている (Alderson, 2005)。すなわち, DIALANG は, 語彙サイズテストを用いて受験者のレベルを判断し, そのレベルにふさわしい言語テストを提供することにより, よりよく言語能力を測定しようとしているのである。別の例としては, *The Vocabulary Levels Test* (Nation, 1990, 2001)はもともと語彙能力の診断テストとして開発されたものであるが, その優れた弁別性のため言語コースへの配置テストとして広く利用されている。

科学研究費補助金助成研究「大規模コーパスを用いた日本人英語学習者用の語彙リスト構築と教材分析システムの開発」(基盤研究 B, 平成 16 年～18 年度, 研究代表者相澤一美)では, 大規模コーパスにより JACET8000 の拡充・精選を行った。さらに, 最も頻度の高い 5,000 語を和英記述・和英選択・英和記述・英和選択という 4つの方式で, 学習者の語彙サイズを推定するテスト, JACET8000 語彙テスト (J8VST) を開発した。

しかしながら, 今までの研究で吟味されてきたのは, 語彙サイズという語彙知識の一側面と言語能力や技能の関係にすぎない。すなわち, 語彙知識とは, サイズ, 構成または知識の深さ, 認知速度という 3つの次元で議論されるべきものであるが (Meara, 1996; Henriksen, 1999; Daller, Milton, & Treffers-Daller, 2007), これまでの語彙知識を測定するテストは語彙サイズのみに焦点を当ててきた。もちろん, 語彙の知識の深さの測定を試みる研究もあるが (Read, 1993, 1998; Paribahkt & Wesche, 1993; Wesche & Paribahkt, 1996; Mochizuki & Aizawa, 2000; Greidanus, Beks, & Wakely, 2005; Greidanus, & Nienhuis, 2001; Meara & Wolter, 2004; 望月, 2006), 言語能力を予測するものとして用いられたものは数少ない。そのひとつである Qian (1999, 2002)は, *The Word Associates Test* (Read, 1993, 1998)が語彙サイズテストよりも読解力をよりよく説明できるかどうかを研究課題としている。Qian の最初の研究では, 語連想テストの方が語彙サイズテストよりも, よりよく読解力を説明するという結果を得たが, 第 2 の研究では 2つのテストの説明力に違いは見られなかった。これは, 語連想テストが同意語と共起語を選択させるという方式を取っており, 語彙サイズテストと同じような能力を測定しているからなのかもしれない。

語彙構成と語知識の深さは、別の概念であるという主張があるが(Meara & Wolter, 2004)、本研究ではこの2つは同じ語彙知識の状態を表すものとして、相互に入れ替え可能な用語として用いる。これは、研究者が本来用いた用語を採用するためである。語彙の構成のテストとして提案されている Lexical Organisation Test (LOT; 望月, 2006)は、サイズとは違う側面の測定を意図している。使用される語彙は高頻度語のみであり、それらの語どうしのつながりを問うものである。これは、中核的な意味だけでなく、周辺的な意味、さらに、語の共起性、使用頻度に至るまでさまざまな語の知識の側面を考慮しないと正解できないものになっている。このような語彙構成のテストは、語彙サイズテストでは測れない言語能力を測定していると考えられる。

単語の認知速度の研究は、心理学では語彙性判断課題、音読課題のような形式で広く行われているが、応用言語学分野ではほとんど例がない。唯一、スウォンジー大学の Paul Meara 教授が開発した Q\_Lex というコンピュータプログラムは単語の認知速度の測定を意図したもので、Coulson (2005)がその検証を行っている。しかしながら、Q\_Lex は高頻度語がでたらめな文字列の中に隠れていて、それを探し出すまでの時間を計測するという方法で、認知速度を測定するものであるため、多分にクイズ的な要素があり、単語の認知速度を十分に計測しているのか疑わしいものがある。Mochizuki (2006)は、高頻度語の認知速度を語彙性決定課題として測定しようとしたものだが、認知速度のテストとしての妥当性に問題があった。本研究では、単語の認知速度を、語形を認知するだけでなく心内辞書に貯蔵されている単語の意味にアクセスするという意味でアクセス速度という用語を用いる。

このような学術的研究を背景に、本研究は、語彙知識をサイズ、構成(知識の深さ)、認知(アクセス)速度という3つの次元で測定し、それぞれが言語能力やその下位構成要素である技能をどの程度予測・説明できるかを解明しようというものである。

## 1.2 研究計画：プロジェクト会議の開催と科研関連学会での発表

本研究は、平成19年度から21年度の3カ年にわたって実施した。各年度の研究計画は次の通りである。

平成19年度は、多面的語彙知識テストを構成する語彙サイズテスト、語彙構成テスト、語の認知速度テストを、妥当性検証のために用いる他のテストとともに大規模に実施することにより、それらのテストの妥当性及び信頼性を検証し、3つの語彙テストの改良を目指した。以下のような日程で研究を進めた。

**平成19年度第1回プロジェクト東京会議(平成19年5月12日)** 麗澤大学東京研究センターにて、研究目的、計画、方法の確認し、3つの語彙テストの実施方法、結果のデータの解釈方法について打ち合わせた。さらに、第2回以降のプロジェクト会議日程、12月に海外研究協力者を招聘して語彙テストフォーラムの開催について話し合いを行った。

**平成19年度第2回プロジェクト広島会議(平成19年9月6日)** 安田女子大学にて、各研究グループが研究の進捗状況について報告を行った。J8VST グループ(石川, 上村, 村田)は J8VST と Nation の Vocabulary Size Test の実施状況, LOT グループ(杉森, 望月)は LOT と WAT の実施状況, Word Recognition Test(WRET)グループ(相澤, 磯)は新語彙アクセス速度テストの開発状況について報告した。また、研究協

力者清水伸一氏から Web 版 J8VST の使用方法についての説明があった。

平成 19 年度第 3 回プロジェクト京都会議(平成 19 年 11 月 4 日) キャンパスプラザ京都にて、引き続き 3 グループによる研究の進捗状況の報告を行った。さらに、12 月に予定している語彙テストフォーラムの詳しい打ち合わせを行った。

#### 語彙テストフォーラムの開催

日時：平成 19 年 12 月 1 日(土) 午後 2 時～5 時

場所：麗澤大学東京研究センター

講演：午後 2:00～3:20

司会：投野由起夫(東京外国語大学)

Paul Meara(Swansea University)

“What vocabulary tests are intended to measure and what they actually measure”

シンポジウム：午後 3:40～5:00

「異なる語彙テストは異なる語彙の側面を測定しているのか」

司会：上村俊彦(長崎県立大学)

コメンテーター：Paul Meara (Swansea University)

報告者：石川慎一郎(神戸大学), 磯達夫(麗澤大学), 望月正道(麗澤大学)

フォーラムでは、海外研究協力者の Paul Meara 氏の講演とシンポジウムが行われた。シンポジウムでは、石川、磯、望月が科研プロジェクトの目的と進捗状況についての英語での報告を行い、それについてコメンテーターの Meara 氏からのコメント、およびフロアとの意見交換が行われた。60 名以上の参加があり、有意義なフォーラムとなった。

平成 19 年度第 4 回プロジェクト東京会議(平成 19 年 12 月 2 日) 東京電機大学工学部外国語系列会議室にて、前日開催した語彙テストフォーラムの総括を行った。J8VST は弁別力指数で弱いものを排除し、短時間でできるように項目数を減らす。アクセス速度テストは文字列問題を排除し、精度を高める。また、19 年度に行った研究の報告書内容の検討を行った。さらにサイズ、構成、アクセス速度の 3 つの分野で平成 19 年度の研究のまとめる。

平成 20 年度は、多面的語彙知識テストを構成する語彙サイズテスト、語彙構成テスト、語のアクセス速度テストを 500 人に実施すると共に、英語標準テストである Test of English for International Communication® (以下、公式試験も団体試験もすべて TOEIC と記す)をそれぞれ実施する。得られたデータを分析することにより、3 つの語彙知識テストが英語標準テストの得点やその構成要素である技能の得点をどの程度説明するものであるかを解明する。この関係を回帰式で表すことにより、語彙知識テストから英語習熟度や技能の習得度を推測することを試みる。以下の日程で研究を進める。



平成 20 年度第 1 回プロジェクト東京会議(平成 20 年 5 月 3 日) 麗澤大学東京研究センターにて、平成 20 年度の研究方法について、会議日程、19 年度報告書の内容検討を行った。研究方法については、3 つのテストをサーバーに置き、ウェブ上でデータを収集する方式が稼働できるようになった。会議日程としては、JACET リーディング研究会と JACET 英語語彙研究会の合同研究大会を 12 月に開催し、そのときに海外研究協力者の Tess Fitzpatrick 氏を招聘し、科研プロジェクトと関連する講演をお願いする。さらに、講演の翌日に大阪または京都で第 2 回プロジェクト会議を開く。第 3 回は 2 月に長崎で開催することが決まった。

#### 第 47 回大学英語教育学会全国大会シンポジウムの開催

日時：平成 20 年 9 月 15 日

場所：早稲田大学

テーマ：「語彙測定 of 3 次元的アプローチとその問題点」

司会・提案者：望月正道（麗澤大学）

提案者：石川慎一郎(神戸大学) 磯達夫(麗澤大学)

石川、望月、磯の 3 名が語彙サイズ、語彙構成、認知速度の測定方法とその問題点を提案し、フローと意見交換した。

#### 第 1 回 JACET リーディング研究会・英語語彙研究会合同研究大会

日時：平成 20 年 12 月 6 日（土曜）午前 10 時 50 分～午後 5 時 45 分

場所：関西学院大学 西宮上ヶ原キャンパス

11:00～11:30 Mochizuki Kaken Project Report (1): L2 Vocabulary Size, Motivation, and Proficiency 石川 慎一郎（神戸大学）

11:40～12:10 Mochizuki Kaken Project Report (2): How much of reading and listening test performances could be predicted by performances on a lexical organization test?: An exploratory study 望月 正道（麗澤大学）

13:00～13:30 Mochizuki Kaken Project Report (3): Development of an online word recognition test: Examination of the internal validity  
磯 達夫（麗澤大学）、相澤 一美（東京電機大学）

13:40～14:10 Mochizuki Kaken Project Report (4): Relationships Among Three Types of Vocabulary Tests and Speaking Test 小泉 利恵（常磐大学）

講演 15:40～17:40 Dr. Tess Fitzpatrick (Swansea University) Using word association responses to explore the nature of vocabulary knowledge

合同研究大会の研究発表では、石川、望月、磯、相澤、小泉の 5 名が語彙サイズ、語彙構成、認知速度の測定結果と英語習熟度の関係を発表し、フローと意見交換した。科研の招待講演者である

Fitzpatrick 氏は、単語連想研究の常識とも言えるべき前提に疑問を投げかける意義深い講演をされ、プロジェクトの今後に大いに参考となった。

**平成 20 年度第 2 回プロジェクト京都会議(平成 20 年 12 月 7 日)** ホテル京阪京都藤の間にて、前日の第 1 回 JACET リーディング研究会・英語語彙研究会合同研究大会の総括、20 年度の研究の総括、報告書の締め切りについて議論した。

**平成 20 年度第 3 回プロジェクト長崎会議(平成 21 年 2 月 11 日, 12 日)** 長崎ブリックホールにて、21 年度の研究報告書の検討、用語の統一、平成 21 年度の研究の進め方について議論した。

平成 21 年度は、20 年度で算出した語彙テストによる標準テストごとの得点推定の検証を行った。かつ、言語習熟度や各技能を最もよく推定するには 3 つの語彙テストをどのような割合で組み合わせればよいかを試算した。さらに、3 年間にわたる科研プロジェクトの研究報告書を作成した。

**平成 21 年度第 1 回プロジェクト東京会議(平成 21 年 5 月 2 日)** 麗澤大学東京研究センターにて、21 年度の研究計画、学会発表、科研報告書の作成、科研報告会について議論した。語彙サイズテスト J8VST と語彙構成テスト LOT は改訂を行わないことが確認された。科研報告書は 11 月末日を締め切りとし、12 月の会議で読み合わせをし、年内に完成させる。

#### **第 48 回大学英語教育学会全国大会シンポジウムの開催**

日時：平成 21 年 9 月 4 日

場所：北海学園大学

テーマ：多次的語彙テストによる英語能力の推定

司会および提案者：相澤一美(東京電機大学)

提案者：清水伸一(安城学園) 小泉利恵(常磐大学) 杉森直樹(立命館大学)

相澤、清水、小泉、杉森の 4 氏が科研プロジェクトの中間報告を行い、フロアーと意見交換した。

**平成 21 年度第 2 回プロジェクト札幌会議(平成 21 年 9 月 4 日)** 北海学園大学にて、大学英語教育学会全国大会におけるシンポジウムの総括、21 年度の研究の進捗、報告書作成の確認、科研プロジェクト報告会について議論した。

**平成 21 年度第 3 回プロジェクト札幌会議(平成 21 年 12 月 6 日)** 麗澤大学東京研究センターにて、3 年間にわたる科研プロジェクトの報告書の原稿検討を行った。原稿の改訂版を年内に提出することを確認した。

### 1.3 報告書の構成

この報告書は、7つの章から成る。第1章である、この章で研究全体を概観したあと、第2章では語彙サイズテストの研究について報告する。先行研究を振り返ったあと、J8VSTの妥当性・信頼性を検証する。つぎに、J8VSTによりTOEICリーディング・リスニング得点がどれくらい予測できるかの研究報告を行う。最後に、Criterionのライティング得点の予測についての研究を報告する。第3章では、語彙構成テストの研究を扱う。先行研究のあと、LOTの妥当性・信頼性の検証を行う。これはWAT(Read, 1993;1998)による併存的妥当性とプロトコル分析による妥当性検証である。つぎに、LOTによってどれくらいTOEICリーディング・リスニング得点が予測できるかを論じる。第4章では、アクセス速度テストの研究を論じる。先行研究のあと、この科研プロジェクトで開発を試みたアクセス速度測定テストWRECTとLEXATTについて報告する。第5章では、3つの語彙テストを用いてTOEIC得点をどれくらい予測できるのかを構造方程式モデルによって予測する。また、電話による英語スピーキングテストであるVersantによるスピーキング力、TOEICスピーキング・ライティング得点を語彙テストでどれくらい予測できるのかも考察する。第6章は、本研究の結論で、3次元の語彙テストが技能別および総合的英語能力をどの程度予測できるかについてまとめる。また、本研究の限界、今後の研究の方向性についても言及する。巻末には、3つの語彙テストをウェブ上で公開しているURLおよびテストの元資料を付録として掲載する。

## 第2章 語彙サイズテスト

### 2.1 先行研究の概観

第二言語としての英語学習者の語彙サイズを測定するテストとしては、チェックリスト・テスト (Meara & Jones, 1990) や The Vocabulary Levels Test (Nation, 1990; 2001) がよく利用されている。前者は、10,000 語までの語彙表から抽出した語に疑似語を加えた目標語を Yes「知っている」、あるいは、No「知らない」と回答していく形式のテストである。10,000 語までの語彙サイズの推定が可能である。後者は、2,000 語、3,000 語、5,000 語、10,000 語レベル、学術語彙という 5 つのレベルから 30 語ずつを抽出し、語義と組み合わせる形式のテストである。Nation 自身は、どのレベルの語彙で指導が必要かを判断するための語彙診断テストとして開発したもののだが、開発者の意図とは裏腹に、手軽に英語学習者の語彙サイズを推定できるために、10,000 語までの語彙サイズ推定に利用されることが多い。Nation (2008) は、このような傾向に歯止めをかけるためか、The Vocabulary Size Test を開発した。これは、The British National Corpus (BNC) の頻度情報に基づき、最も頻度の高い 1,000 語ごとに、14,000 語を 14 のレベルに分け、それぞれ 10 語ずつを 4 肢択一で選択させる形式のテストである。これにより、14,000 語までの語彙サイズの推定が可能となる。

このような語彙サイズテストは、いずれも目標語を提示し、その意味を表す選択肢を選らばせる、あるいは、目標語を知っているかいないかを自己申告させる形式である。しかしながら、語彙知識は単語により受容レベルから産出レベルまで異なる習得段階にあり、しかも、表出方法によっても結果として得られる語彙知識は異なる。Laufer, Elder, Hill, and Congdon (2004) と Laufer and Goldstein (2004) は、語彙知識は、受容と発表、再認と再生という 2 つの次元から、1) 受容・再認、2) 発表・再認または受容・再生、3) 発表・再生という 4 つの方式のテストを作成し、この順に難易度が上昇するとしている。すなわち、発表・再生が可能なレベルの語彙が一番習得が難しく、受容・再認が可能なレベルの語彙が一番習得が容易であるとされる。この考えに基づき、望月 (2007) は、JACET8000 の 5,000 語までの語を、受容・再認(英和選択)、受容・再生(英和記述)、発表・再認(和英選択)、発表・再生(和英記述) という 4 つのモードでテストする語彙テスト(J8VST)を開発した。これは、学習者の語彙サイズを 4 つのモードでそれぞれ 5,000 語まで推定することができる。本研究は、この J8VST を元に英語語彙サイズが、英語習熟度をどの程度予測可能かを調べるものである。

### 2.2 日本人大学生のための新しい語彙サイズテストの開発

#### 2.2.1 研究の目的

語彙力にはさまざまな側面があるが、中でも語彙サイズは最も明示的で測定しやすいものであり、これまでにさまざまなテストが開発されている。本研究では、大学英語教育学会が選定した JACET8000 語彙表を基準リストとして、日本人英語学習のための使いやすい語彙サイズテストを開発することを目的とする。

#### 2.2.2 考慮すべき問題

語彙サイズテストの開発にあたって考慮すべき問題は多い。まず、重要になるのは何をどのように尋ねるか、すなわちテストのモードの問題である。語彙サイズテストのモードには、言語転換の方向性と、受験者に問う語彙力に関わる。

言語について言えば、日本人学習者を対象とした英語の語彙サイズテストの場合、(1)英語から英語へ(例: 英文定義を見て該当単語を選ぶ)、(2)英語から日本語へ(例: 英単語を見てそれに相当する日本語の意味を答える)、(3)日本語から英語へ(例: 日本語の単語を見てそれに相当する英語を答える)という3つの方向性が考えられる。このうち、本研究では、実施の利便性を考え、(2)および(3)の方向に限って検証することとした。また、語彙力について言えば、当該語のスペルを実際にかくことができる再生的・産出的な側面と、当該語のスペルを見て意味が理解できる再認的・受容的な側面があることが知られている。

以上をまとめると、語彙サイズテストの主要なモードは表 2.1 の 4 種類ということになる。

表 2.1 語彙テストのモード

語彙知識タイプ	受容(英語→日本語)	発表(日本語→英語)
再生的・産出的	モード 1(英和記述)	モード 3(和英記述)
再認的・受容的	モード 2(英和選択)	モード 4(和英選択)

### 2.2.3 試行版 4 モードテストの開発と実施

望ましい語彙サイズテストを開発するためには、上記の 4 つのタイプの中から最も信頼性と有効性の高いものを選ぶ必要がある。そこで、4 モードすべてについて試行版テストを開発し、同一受験者に受験させ、スコアの関係性を調査することとした。

各テストでは、JACET8000 のレベル 1~5 の各々から 25 語を選び、合計 125 問を出題した(標準回答時間は 1 モードにつき 15~30 分程度)。いずれのテストでも問題に使用した語は同一である。テストは研究協力者の清水伸一氏によって cgi を用いてオンライン化された。これにより、インターネット環境さえあれば、受験者は場所を選ばず受験が可能になり、実施利便性が確保された。さらに、受験者の回答データはウェブ上に蓄積され、csv の形式で取り出せるようになったため、分析上の利便性も大幅に高まった。図 2.1 と 2.2 は試行テスト画面の例である。

英語の意味を表す日本語を選んで下さい。

001 /125 mention

1 意識する

2 影響する

3 言及する

4 男性的にする

5 わからない

英語の意味を表す日本語をタイプして下さい。

001 /125 mention

解答

図 2.1 試行版語彙サイズテスト(英和選択)

図 2.2 試行版語彙サイズテスト(英和記述)

実際の調査は2007年の前期に3つの大学(東京圏私立中堅大学, 関西圏国立総合大学, 九州圏公立大学)で行い, 全体で171件の有効回答データを得た。調査に参加した学生は4モードテストと後述する比較用の2種類の既存語彙テスト, 併せて6種類を受験した。なお, 4モードテストの実施にあたっては, 学習効果が発生しないよう, テスト間に異なる作業を行わせるなどの配慮を行った。また, 系列位置効果によって結果が影響されないよう, 受験者群によって4モードの順序を変更してテストを実施した。

## 2.2.4 試行版テストの評価

4モードテストの信頼性を検証するため, (1)内部一貫性検証, (2)既存の基準語彙サイズテストとの比較を実施した。

### 2.2.4.1 内部一貫性検証

一般に, テスト・アイテムの信頼性を見る際にはクロンバック  $\alpha$  という統計量が広く用いられている。これは, 合計点(尺度得点)を構成する部分得点が内的整合性を持つかどうか, 言い換えれば, 個々の設問がそれら全体によって調べようとする目的に整合しているかどうかを評価するスコアと言える。

上記の4モードテストの個々の設問ごとのスコアをサンプルとして内部一貫性検証を行ったところ, 表2.2の結果を得た。

表 2.2 内部一貫性

語彙知識タイプ	英語→日本語	日本語→英語	係数平均
再生的・産出的	.91	.98	.95
再認的・受容的	.88	.93	.91
係数平均	.90	.96	—

一般に, クロンバック  $\alpha$  は.7以上が望ましいとされるので, 今回の検証ではすべてのモードが一定の内部一貫性を持っていることが確認された。

ただし, 各モード別に係数の大小を見ると, 日本語を見てそれに相当する英単語を自分で考えてその綴りを書かせるモード3(和英産出)が.98と最も高い値を得ていることが分かる。また, 英語→日本語の平均は.90, 日本語→英語の平均は.96であることから, 言語転換方向としては日本語→英語のほうが信頼性が高い。さらに, 産出的テストの平均は.95, 受容的テストの平均は.91であることから, 語彙力としては産出的テストのほうが信頼性が高いことになる。

### 2.2.4.2 既存語彙サイズテストとの比較

既に述べたように, 語彙サイズテストについてはすでにさまざまなものが開発されており, 一部は広く使用されて検証報告も出ている。ここでは各種の語彙サイズテストの中から, 基準的テストとして, Schmitt(2000)の Vocabulary Levels Test (Version 1)と Nation(2008)の Vocabulary Size Test を使用した。両テストの出題形式の例を図2.3と2.4に示す。

1. game [ ] 2. winning [ ] 3. being born [ ]  (1) birth, (2) dust, (3) operation, (4) row, (5)sport, (6) victory	See: They <u>saw</u> it.  (1) cut (2) waited for (3)looked at (4) started
---	--

図 2.3 Schmitt (2000) VLT

図 2.4. Nation (2008) VST

出題の視点は、Schmitt の場合は語(句)と語のマッチング、Nation はセンテンスレベルの文脈をふまえた語のパラフレーズとなっている。いずれも語彙転換の方向性で言えば英語→英語であり、本研究が比較対象とした英語→日本語、日本語→英語とは異なるが、仮に、英和型ないし和英型のテストで英英型との相関が出たとするならば、それは当該モードの信頼性を補強するデータと言える。

相関検定の結果は表 2.3 の通りである。

表 2.3 相関行列表

基準テスト	英和記述	英和受容	和英産出	和英受容	Nation	Schmitt
Nation	.54	.43	.52	.61	1.0	.67
Schmitt	.59	.47	.56	.80	.67	1.0
相関平均	.57	.45	.54	.71	—	—

上記より、和英受容型がいずれの基準的語彙サイズテストに対しても比較的高い相関をもっていることが分かった。もちろん、Nation と Schmitt のテスト形態はいずれも選択式(つまりは受容型)であることから、産出テストより受容テストのほうが相関が高いことは予想がつくが、英和型よりも和英型のほうが良い結果が出たことは重要な結果と言える。

次に、各テストの関係性を概観するために、コレスポンデンス分析を実施した。コレスポンデンス分析とは、相関を最大化するようにデータ表の行アイテムと列アイテムのカテゴリを並べ替え、アイテムの類似性に基づきアイテムを分類する統計手法である。

分析の結果、第 1 アイテムスコアおよび第 2 アイテムスコアを 2 軸として、図 2.5 のようなアイテム布置図を得た。和英受容型(JE\_R)が 2 種類の基準語彙テストと近接した関係にあることが確認できる。

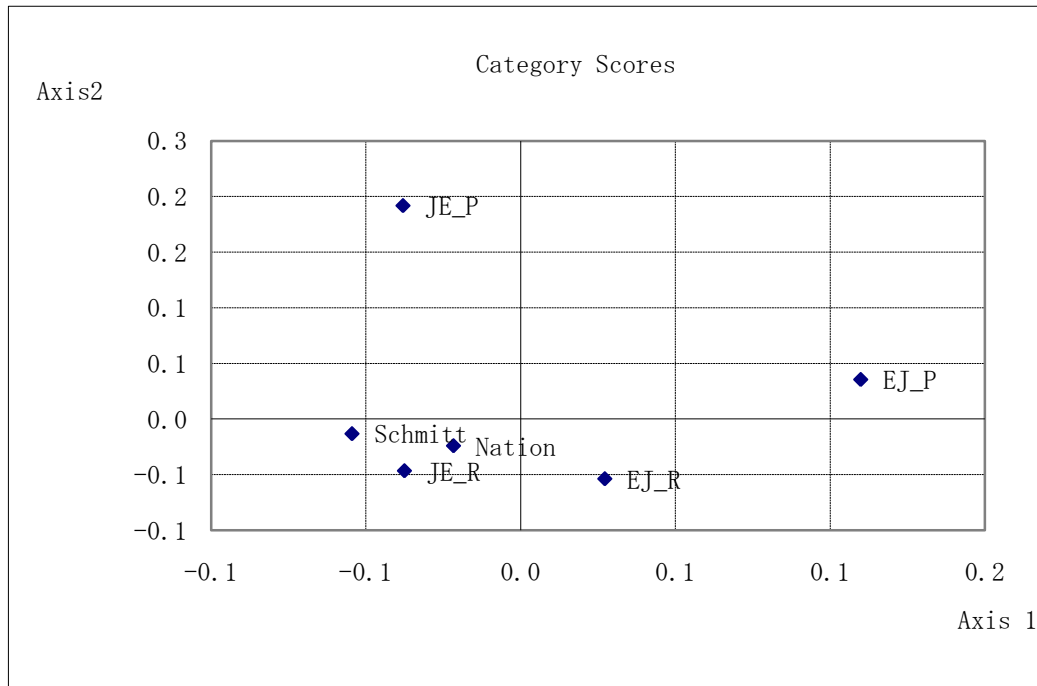


図 2.5. 布置図

### 2.2.5 小括

以上より、試行的に開発された 4 モードテストの中では、モード 4(和英受容型)が総合的に見て最も信頼性が高いと確認された。ゆえに、和英受容型を軸として語彙サイズテストのさらなる有効性検証を行うこととした。

## 2.3 語彙サイズテストは総合的英語力を推定しうるか

本節では、本プロジェクトで開発されたオンライン版語彙サイズテスト(以下 J8VST)を用いた 2 つの実証調査の結果について報告する。

第 1 調査は 2008 年に実施したもので、複数の大学に所属する 150 名の受験者を対象に J8VST と TOEIC テストの各セクションのスコアの関係を検討した(2.3.1 節)。

第 2 調査は 2009 年に実施したもので、単一大学の約 70 名の受験生を対象に J8VST、TOEIC テストと ETS の Criterion による課題英文ライティングのレベルスコアとの関係を検証した(2.3.2 節)。

後述するように、いずれの調査においても、J8VST のスコアは一定の信頼性を持って英語総合力を予測することが示唆された。

### 2.3.1 第 1 調査の概要と結果

#### 2.3.1.1 調査の実施

2.2 節において最も妥当な出題形式として選ばれた和英受容型の語彙サイズテストの実際的な有効性を検証するために、TOEIC テストとの相関検証を行った。調査は前述のオンライン版を使用した。和英受容型は図 2.6 に示すような選択式であるため、125 問の回答に要する時間は比較的短く、10~15 分で全問の回答が可能である。



日本語の意味を表す英語を選んで下さい。

001 / 125 話に出す,言及する,言及

- 1 manipulate
- 2 mention
- 3 minister
- 4 moderate
- 5 わからない

次へ

図 2.6 J8VST(オンライン版)

調査は 2008 年度前期に行われ, 受験者は 4 大学に所属する 1~2 年生 150 名であった。専門分野および英語力のバランスを取るため, データは私立大学人文系(38 名), 私立大学理科系(49 名), 国立大学人文・理科系(26 名), 公立大学人文系(37 名)の 4 種類から採取した。併せて実験の参加者には, 許諾書を取った上で, 直近の TOEIC テストのセクション別スコアの提供を求めた。

### 2.3.1.2 調査の結果と考察

まず, 各テストの記述統計量は表 2.4 の通りである。

表 2.4 記述統計量

テスト種別	N	平均	最大値	最小値	標準偏差
TOEIC	150	439.20	895	105	160.80
TOEIC リスニング	150	249.60	495	100	81.70
TOEIC リーディング	150	189.60	430	60	86.40
語彙サイズテスト	150	93.60	120	61	13.60

次にテスト間の相関性検定を行ったところ, 表 2.5 の結果を得た。

表 2.5 相関行列表

	TOEIC	TOEIC リスニング	TOEIC リーディング	語彙サイズテスト
TOEIC	1			
TOEIC リスニング	.95	1		
TOEIC リーディング	.96	.83	1	
語彙サイズテスト	.72	.64	.73	1

以上より、本研究で開発された語彙サイズテストは、TOEIC リスニングセクションスコアと.64、TOEIC リーディングセクションスコアと.73、TOEIC の総合スコアと.72 の相関を持つことが確認された。テストの性質上、語彙サイズテストとリーディングセクションの相関がある程度高いことは予想された結果であったが、リスニングセクションとも一定の相関が出ていることは注目される。

また、一般に相関係数の平方が説明力とされる。ここでは、TOEIC の総合点との相関係数.72 であるので、TOEIC 総合点の分散の約 51.8%が語彙サイズテストだけで説明されていることを示唆する。

本研究で開発された語彙サイズテストが、インターネット環境さえあればテスト用紙の準備や採点のコストが不要で、かつ、受験者にとってもわずか 10～15 分間で実施できることを考えれば、これは一応の満足のゆく結果であると結論できる。

### 2.3.1.3 総括

以上、本研究で明らかになったことをまとめておきたい。

- (1) 4 モードの語彙サイズテストの内部一貫性(クロンバック  $\alpha$ )は、和英産出(.98) > 和英受容(.93) > 英和産出(.91) > 英和受容(.88)の順となった。また、言語転換の方向としては和英が英和より、語彙力の問い方としては産出語彙力のほうが受容語彙力より、それぞれ信頼性が高いことが確認された。
- (2) 4 モードテストを 2 種類の英英型の基準語彙サイズテストと比較したところ、相関平均値( $r$ )は和英受容(.71) > 英和産出(.57) > 和英産出(.54) > 英和受容(.45)の順となった。
- (3) 以上より、内部一貫性・基準テストとの相関性・実施利便性などを総合的に判断すると、望ましい語彙サイズテストのモードは和英受容型(和英選択型)であることが確認された。
- (4) 和英受容型語彙サイズテストは、TOEIC のリーディングセクションスコアと.73、リスニングセクションスコアと.64 の相関を持ち、リーディングのみならず、リスニングとも一定の相関度を示す。また、TOEIC の総合スコアとの相関度は.72 となり、本テストは実用に耐えるものと結論できる。

もちろん、本研究には今後の課題もいくつか残されている。最も大きな問題は信頼性検証のサンプルの取り方である。一般に語彙サイズテストというのは万人向けの汎用テストとして開発されることが多いが、近年の大学生の英語力の多様化は著しく、今回の調査でも、大学種別によってスコアや相関値が大きく変化する傾向が確認された。

つまり、さまざまな大学をバランスよく組み合わせた理想的なサンプル群において TOEIC や基準的語彙サイズテストと高い相関度があっても、任意の 1 大学で実施した場合に同じ相関度が出ることは統計学上期待できない。本来の全体的分布が切断された場合、いわゆる切断効果が発生し、統計的な相関性が大きく損なわれるからである。汎用的に開発されたテストが個別的环境で使用しにくい場合があることは留意すべきであろう。

また、もう 1 つの問題は実施利便性のさらなる検討である。125 題という問題量は決して多いものではないが、多くの学習者に気軽に受験させるためには、現状の 125 題という問題数をさらに精選することが必要かもしれない。その際、推定の考え方を取り入れて、テストの枠組みに柔軟な可変性を持たせることが対策として考え

られる。

たとえば、テストの冒頭部で JACET8000 の 5 レベルから 5 語ずつを選び、合計 25 語程度の小規模なブローブ・テストを行わせる。この結果をオンラインで即時採点し、正答数が 100% のレベルと 0% のレベルについては以後の出題を取りやめる。そして、語彙サイズ推定の際には、当該レベルの正答率はそれぞれ 100% と 0% という値を推定値として代入する。こうした措置により、平易すぎる問題や難しすぎる問題を回答する無駄がなくなり、実施時間のさらなる圧縮が期待できる。ただし、問題数と信頼性がトレードオフの関係にあることは自明であるため、問題量の精選にはさらに広範で詳細な調査を行うことが不可欠であろう。

### 2.3.2 第 2 調査の概要と結果

前節(2.3.1)では、本研究で開発された語彙サイズテスト(J8VST)の 4 モードのうちで、語彙サイズ推定のツールとして最も有効なモードは日本語→英語の受容語彙力モードであることが明らかとなった。本節では、J8VST (和英受容型モード、以下、J8VST)を中心として、長崎県立大学シーボルト校国際情報学部国際交流学科(以下、本学科)に 2009 年度入学した学生の 1 年前期英語学習と英語語彙力の推移を中心に考察を試みる。

2009 年度本学科入学の学生は、TOEIC テストを 4 月の入学直後と 8 月上旬に、J8VST を 5 月上旬と 7 月下旬に、そして English Testing Service (ETS)の Criterion Online Writing Evaluation を用いた英文課題作成(以下、Criterion 課題)を 7 月下旬におこなった。まず 2 回の J8VST スコアについて概観し、その後 TOEIC テストスコアや Criterion 課題に対する判定レベルとの関係についての検証をおこなう。

#### 2.3.2.1 J8VST と TOEIC テスト

本学科 1 年次在籍者(81 名)のうち、J8VST を 5 月上旬と 7 月下旬の 2 回ともに受験した 74 名についてスコア比較を行った(基本統計量は、表 2.6 参照)。

両スコアは、平均点は 97.51 点から 99.96 点へ、最高点は 114 点から 114 点、最低点は 39 点から 64 点へと推移した。 $t$  検定の結果、受験者は、約 3 カ月弱の間に、2 ポイント程度であるが、有意に語彙サイズを増大させていることがわかった( $p < .001$ )。また、2 回の J8VST スコアには、.89 ( $p < .001$ )の強い相関が認められた。これは、J8VST が信頼できる語彙サイズテストであることを示している。

表 2.6 2 回の J8VST の基本統計量

	N	平均	標準偏差
May_J8	74	97.51	10.03
Jul_J8	74	99.96***	9.12

\*\*\* $p < .001$

2 回の TOEIC テストスコア(リスニングスコア、リーディングスコア、総合スコア)と 2 回の J8VST スコアとの相関をみた(表 2.7、表 2.8 参照。ただし、表 2.8 の表記はスペースの関係で一部略記)。相関は高い順に、リーディングスコア > 総合スコア > リスニングスコアとなり、前節(2.3.1.2)と同一の結果となった。

表 2.7 基本統計量

テスト種別	N	平均	最大値	最小値	標準偏差
Apr_Listening	63	256.90	495	170	59.64
Apr_Reading	63	209.21	440	105	64.83
Apr_Total	63	115.23	935	290	115.23
Aug_Listening	62	280.40	415	185	50.40
Aug_Reading	62	234.60	435	100	66.20
Aug_Total	62	515.00	815	310	109.55
May_J8	59	98.54	114	75	7.84
July_J8	57	100.61	114	74	7.19

表 2.8 相関行列表

	Apr_L	APR_R	Apr_T	Aug_L	Aug_R	Aug_T	May_J8	Jul_J8
Apr_L	1							
APR_R	.71**	1						
Apr_T	.92**	.93**	1					
Aug_L	.74**	.65**	.77**	1				
Aug_R	.61**	.75**	.76**	.76**	1			
Aug_T	.71**	.76**	.81**	.92**	.95**	1		
May_J8	.45**	.63**	.59**	.57**	.62**	.64**	1	
Jul_J8	.41**	.61**	.54**	.43**	.53**	.52**	.82**	1

\*\*  $p < .01$

### 2.3.2.2 J8VST と Criterion 課題

本学科 1 年生の CALL 演習クラス「CALL インテンシブ I」の時間中に、Criterion 課題 “Learn from Mistakes (Persuasive)” の作成と提出とを課した。課題の指示は、ETS 標準で以下の通りである。

“Do you agree or disagree with the following statement?

People always learn from their mistakes. Use specific reasons and details to support your answer.”

クラスの時間中に Criterion の学生用簡易マニュアルを配付し、操作の概要を説明後、受験者(63 名)は約 60 分の時間制限で Criterion 課題の英作文をおこない、オンラインで提出した。なお、英文作成時には辞書の使用を許可していない。

Criterion による英文ライティングレベル判定は 6 段階(最高 レベル 6)である。レベル判定は、受験者がオ

ンライン提出した英文ライティングそのものに対して行われたものであり、教員の助言による書き直し等の修正以前のものである。なお、本学科カリキュラムでは、「Basic Writing」(必修)は1年後期の開講科目であるため、課題提出時点で Criterion 課題を提出した受験者は英文ライティングに関する科目履修経験はない(レベル判定スコアの度数分布は図 2.7 参照)。

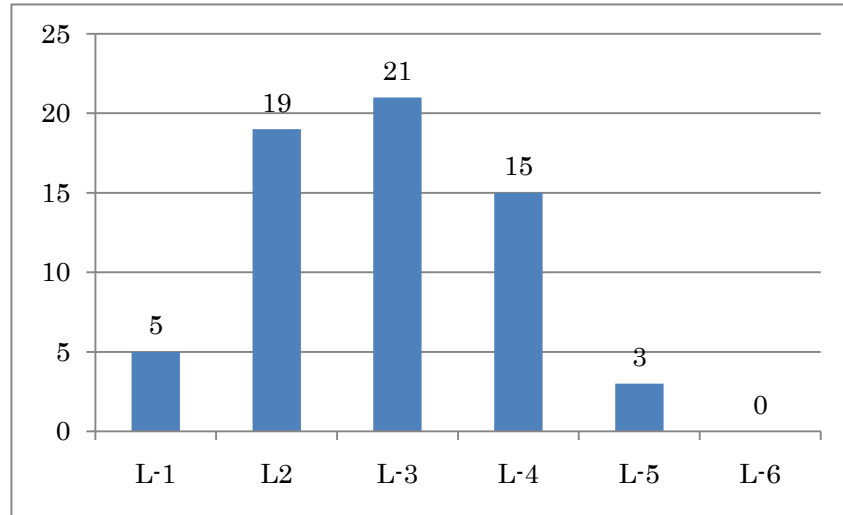


図 2.7 英文ライティングに対する判定レベルの分布

Criterion 課題の英文ライティングに対する判定レベルと J8VST(5 月, 7 月実施)のスコアについて相関を調べた(表 2.9, 表 2.10 参照)。Criterion 課題のレベルスコアと J8VST スコアとの間には, 5 月実施(May\_J8) .42, 7 月実施(Jul\_J8) .38 の相関が認められた( $p < .01$ )。

表 2.9 基本統計量

	N	平均値	最大値	最小値	標準偏差
C-Score	63	2.87	5	1	1.02
May_J8	61	98.54	114	75	7.84
Jul_J8	59	100.61	114	74	7.20

表 2.10 J8VST と Criterion 課題の相関

	C-Score	May_J8	Jul_J8
C-Score	1		
May_J8	.42**	1	
Jul_J8	.38**	.82**	1

\*\*  $p < .01$

ETS (2009)によると, 英文ライティングの評価は, 文法, 語法, メカニクス(mechanics), スタイル, 文章構

造と展開からなる診断的フィードバック分類 (Diagnostic Feedback Categories) に基づいておこなわれる。Criterion 課題のレベル判定には、このような語彙力以外のライティング要素に関する評価が加味されることを留意すべきであろう。また、Criterion による判定が同一であっても、J8VST や TOEIC テストのスコアは、必ずしも同レベルになるとはかぎらない。以下のサンプル A とサンプル B は Criterion 判定でレベル5であったが、J8VST (7 月実施) と TOEIC テスト(8 月実施)のスコアは、サンプル A (110, 615), サンプル B (114, 815)であった。両者は、J8VST (7 月実施)では4ポイント差で大差なかったが、TOEIC テストについては大きなスコア差が認められた。

[View Question](#)

Essay

I agree with an opinion that people always learn from their mistakes. There are following three reasons why I think so.  
 First, when I made mistakes, I think why did I make mistakes and what was wrong point in myself. I try to understand the reason of mistake using a lot of way, for example, I ask my friends, parents or teachers. When I find the reasons about it, I think that I have to change my thought or way in order not to make same mistakes again. I think the progress of thinking about mistakes gives you new chances. And next time you do the same things you can accomplish it with your best.  
 Second, many people think that there are many people who don't care about their mistakes. And the way of thinking is different from people to people, so it is impossible to think all person learn from their mistakes. However, I don't think so. When someone make mistakes, they feel a little sad and disappointment at least. Feeling such emotion is so important and I think it is also learning something. I think people who know such emotion can be more strong.  
 Finally, when they have mistake, even if they don't think that they can learn from their mistakes or they don't care about it at that time, they will notice the importance of their experience after the time passed. I think there are few times which you think about your wrong point if you make mistakes, but when you do the same things after the time you will remember your mistakes and try to go well next time.  
 That is why I agree with the opinion that people always learn from their mistakes. And I think it is important to learn from mistakes because it gives me courage and power to try the next chance.

図 2.8 サンプル A

[View Question](#)

Essay

I can't disagree with the idea that people always learn from mistakes. I think people sometimes learn from mistakes. There are following three reasons. First, human beings can't remember all the things. Everyone will forget something. mistake is a part of this. Even if we decided not to do such a fool thing again, we may forget about it tomorrow because of our ability to forget. This ability is very important to live, because we have a lot of things which we want to forget. Though it will certainly prevents us to remember our mistakes clearly. We do have the ability to forget. Second, sometimes we have to do the things which we know they are wrong. There are a lot of things that we can't solve only by reasons. Sometimes we will act following to our emotion, not reasons. I will suggest a example. There was a violent man. His girlfriend knew that keeping relationships with him was not good by her experience. However, her feeling said that she didn't want to live without him. She knew everything including he was not good for her but her emotion can't allow to do that. Thus, human beings do mistakes again and again even if he or she knew it would be a mistake. Third, we do have the times when we can't get away from mistakes. Even if we don't want to do that, maybe we have to do that again because of the lack of our ability to get away from mistakes. We will try to do well than before to get better result, but we can't always improve our fault so easily. As a result, we do learn something from mistakes. It's certain. However, we can't say that we always learn from mistakes because of our imperfection as a human.

図 2.9 サンプル B

上記のサンプル A とサンプル B に関する観察は、検証が可能となるデータ量を集積した後に改めて考察したい。代わりに、コーパス言語学の観点から Criterion 課題の語彙使用の分析を試みる。語彙分析ソフト

WordSmith Tools Version 5 (以下, WS5)で 63 名分の Criterion 課題の使用語彙分析をおこなった。WS5 による語彙リスト出力時には, 染谷レマリストによるレマ化と, JACET8000 を元にしたストップリストによるカバー率の検証をおこなった。

Criterion 課題で使用された語彙は, JACET8000 の出現頻度上位 3,000 語レベルで異なり語 53.36%, 総語数 79.40%のカバー率となった。ただし, 8,000 語すべてをストップリストとした場合

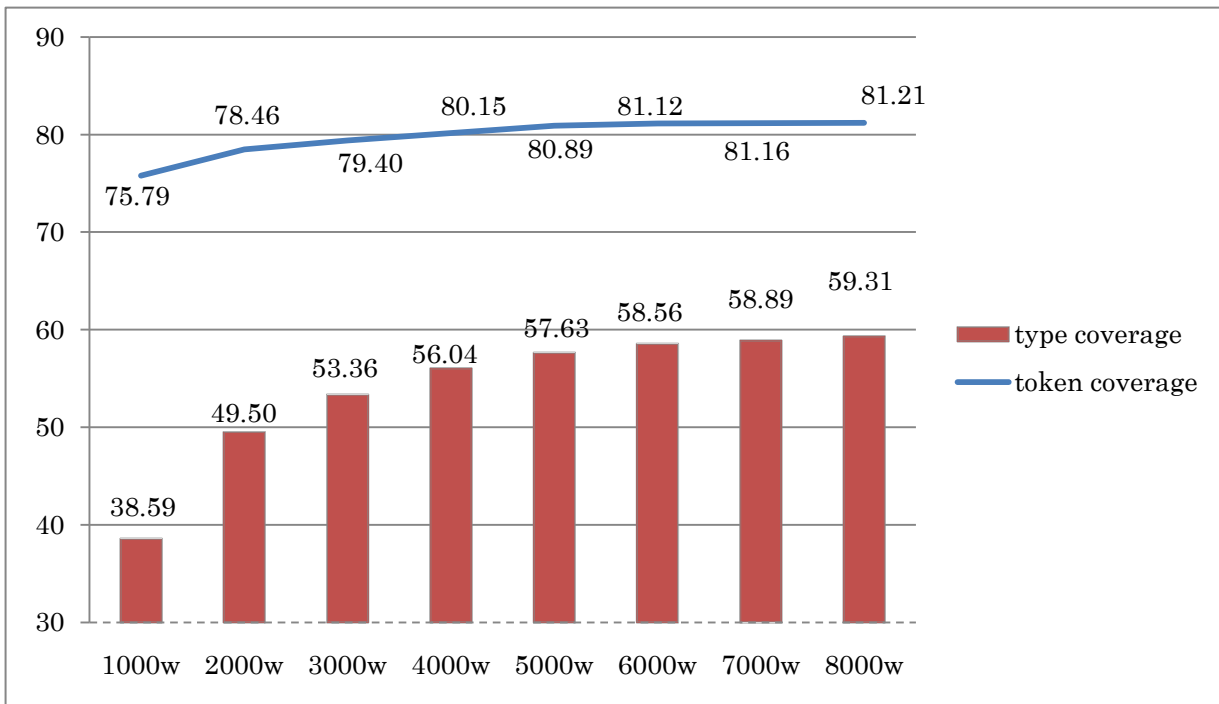


図 2.10 JACET8000 に基づく英文ライティング使用語彙の分析

であっても, 異なり語 59.31%, 総語数 81.21%のカバー率に留まった(図 2.10 参照)。なお, JACET8000 でカバーできなかった語彙の中には, *disappointment* (サンプル A) や *humanbeing*, *imperfection* (ともにサンプル B)などのスペルミス等が含まれている。

### 2.3.2.3 総括

以上, 本節で明らかになったことをまとめておきたい。

- (1)同一の受験者を対象として 5 月上旬と 7 月下旬に実施した J8VST のスコア間には高い相関 (.89) が認められた ( $p<.001$ )。これは再テスト法によって J8VST の信頼性を示すものである。
- (2)Criterion 課題のレベル判定スコアと J8VST スコアとの間には, 5 月実施(.42), 7 月実施(.38)の相関がみられた。 ( $p<.01$ )
- (3)Criterion 課題の使用語彙については, JACET8000 の出現頻度上位 3,000 語をストップリストとした場合のカバー率は, 異なり語 53.36%, 総語数 79.40%であった。しかし, 8,000 語すべてをストップリストで使用した場合であっても, 顕著なカバー率の向上とはならなかった。

第2調査では、4月から8月の短期間に同一の受験者がJ8VSTとTOEICテストとを各2回、前期終了時期にCriterion課題による英文ライティングを課した結果を分析した。J8VSTはTOEICテストのリーディングと最も高い相関となることは、すでに第1調査で明らかにされている。J8VSTのスコア推移を検討したところ、本学科における1年前期の英語学習は学生の語彙力向上に直結していないことが判明した。また、Criterion課題コーパスは、総語数レベルでは全テキストの80%がJACET8000の出現頻度上位4,000語でカバーされること、全JACET8000をストップリストとしてもテキストカバー率は82%以下に留まることが明かとなった。受験者が「英語を書く」視点から検討した語彙力は、JACET8000の出現頻度上位4,000語レベルであると推定される。J8VSTの125語は、JACET8000の出現頻度上位5,000語レベルまでの語彙から各レベル25語を抽出したものである(前節2.2.3参照)。よって、J8VSTは受験者の「産出語彙」評価のスケールとして有効に機能したと判断できるだろう。



## 第3章 語彙構成テスト

### 3.1 先行研究の概観

#### 3.1.1 語彙知識への2つのアプローチ

語彙知識の第2の次元、語彙構成、あるいは語彙知識の深さは、広さと深さのアプローチとして研究されることが多かった。Meara and Wolter (2004)は、広さと深さのアプローチは、2つの点で破綻していると主張している。第1点は、広さと深さという2つの構成要素が互いに関連し合っていない点である。図3.1では、棒グラフの棒ひとつひとつが単語を表わし、その長さが語の知識の深さを表わしている。図3.1が示すように、広さと深さは語彙の2つの異なる次元を表わし、関連し合っていない。すなわち、語彙の広さが広がったとしても深さに影響することはない。逆に、語彙の深さが深くなった語があったとしてもそれが広さに影響することはない。

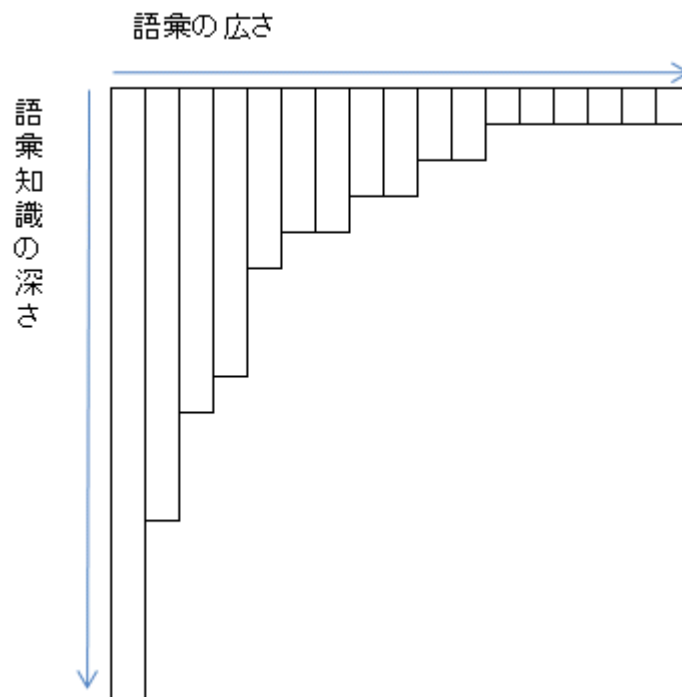


図 3.1 語彙の広さと深さのアプローチ(Meara & Wolter, 2004 を元に作成)

第2の問題点は、語彙の広さと深さを探求するにあたり、時間的制約から、より広くサンプルを採ろうとすると深さは浅くなり、逆により深くサンプルを採ろうとすると広さが狭くなるという制限である(Read, 2000; Meara & Wolter, 2004)。Nation(2001)の単語知識の枠組みによれば、1つの単語の知識は音声、綴り、語の構成、形と意味、概念、連想、文法、コロケーション、使用の制限という9つに分類され、さらにそれを受容的知識、産出的知識と2つに分けることにより、18の下位分類から成り立つ。この18の下位分類すべてについて、学習者の知識を探るのであれば、調査できる語数は制限される。逆に、より多くの語について調べるのであれば、調査できる語の下位区分は制限される。このように、語彙の広さと深さの調査はトレードオフの関係にあり、双方を満足する方法で調査することは困難である。

このように語彙知識の深さを探求するにあたり、広さと深さのアプローチの非効率性を考慮し、Meara and Wolter (2004)は、語彙サイズと構成のアプローチを提唱している。これは、本来 Meara(1996)に基づくもので、語彙の2つの構成要素であるサイズと構成が互いに関連しあっているというネットワークモデルである。図 3.2 では、四角が単語を表し、線が他の単語との結びつきを表す語彙ネットワークを示している。

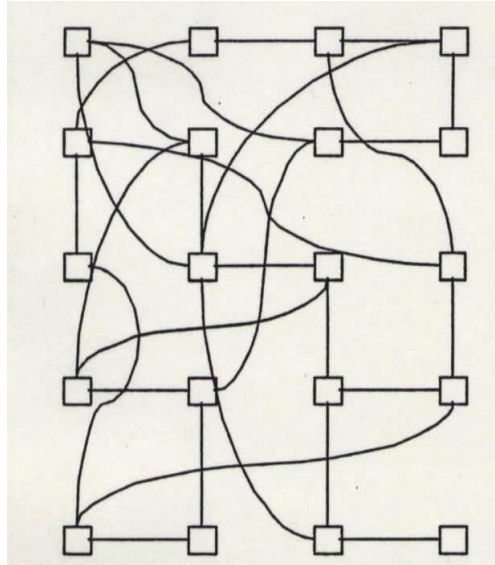


図 3.2 語彙サイズと構成のアプローチ(Meara & Wolter, 2004, p.89)

この語彙サイズと構成のアプローチでは、1つの構成要素の変化が他の構成要素と語彙全体に影響を与える。たとえば、語彙サイズが増加し、構成に変化がなければ、つまり単語と単語のつながりの数が増えなければ、語彙構成は疎になり、語彙全体も語数だけ多く、密度の低いネットワークとなる。逆に、語彙サイズは変化せず、構成が密になれば、つまり単語と単語のつながりの数が増加すれば、1つの単語が結びつく単語数が増えたことになり、語彙全体としては、小さいながらも密度の高いネットワークとなる。

Meara and Wolter (2004)は、語彙サイズと構成のアプローチの長所として、図 3.3 のように語彙構成のサンプルを綿密にかつ適切に採れば、語彙サイズの場合と同様に、ちいさなサンプルから語彙全体を推定することが可能であることを挙げている。

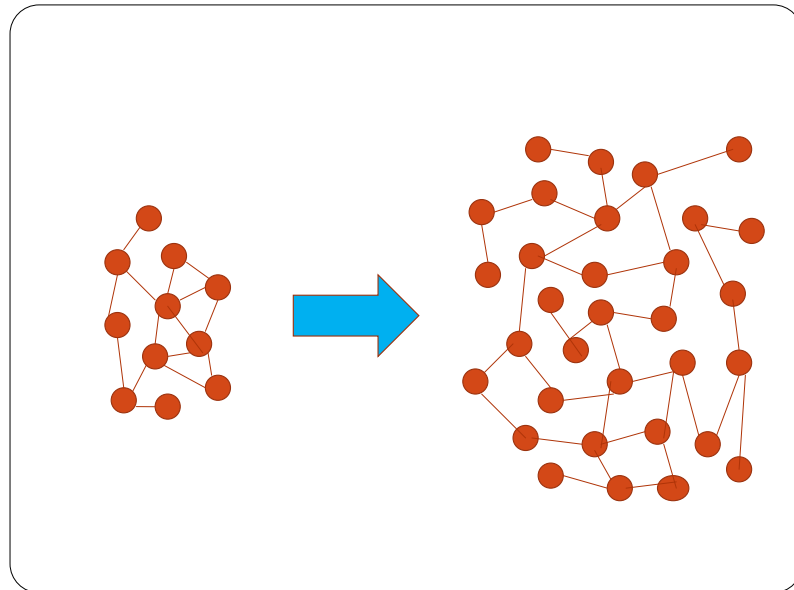


図 3.3 サンプルからの語彙構成の推定

### 3.1.2 語彙構成のテスト:V\_Links

Meara and Wolter (2004)は、語彙構成を測定するテストとして、V\_Links を開発している。このテストには、10語で1セットのテスト項目が20セットある。200語は最も頻度の高い1,000語から無作為に抽出されている。受験者はコンピュータ画面上に表示された10語の中からつながりがあると思う2語を1分以内にすべて見つけていく。2語のつながりの強さを1~4の4段階で評価する。Meara and Wolterは、このテストの特徴として、つぎの3つを挙げている。

- 1) 比較的短時間(30分)に多数の語(200語)をテストできる
- 2) 1セットで英語母語話者は平均して6つのつながりを見出す。20セットで120となり、見出したつながり数により英語母語話者と学習者を区別できる
- 3) 最も頻度の高い1,000語以内の語彙を用いているので、初級者から上級者まで幅広い学習者が受験することができる

逆に、問題点としてつぎの3つを挙げている。

- 1) 学習者は母語話者が選ばないつながりを選択する
- 2) つながりの強さの評価の問題。つながりは何でも強いと評価する受験者と弱いと評価する受験者が存在する
- 3) 時間制限の問題

Meara and Wolter (2004)は、V\_Links を147名の日本人英語学習者と英語母語話者に実施し、両者を区別できること、さらに、V\_Links の得点と語彙サイズテストの得点の相関は.30であることを報告している。この

結果は、語彙構成とサイズは独立していると解釈でき、V\_Links は語彙構成を測定するテストであるとしている。

しかしながら、Mochizuki (2005)は、習熟度の異なる2つの英語学習者グループにV\_Linksを実施したところ、その得点は両者を区別できなかったことを報告している。その原因として、Mochizuki は、2つのグループの語彙構成は最もよく使われる1,000語では違いがないためと推定した。それに基づき、彼は、最も頻度の高い2,000語から目標語を無作為に抽出することによりV\_Linksを改訂した。TOEIC平均得点350点のグループと500点のグループに改訂版V\_Linksを実施したところ、その得点は2つのグループを区別した。この結果は、2通りに解釈できる。第1の解釈は、得点の違いは2つのグループの語彙構成の違いを反映するものであり、改訂版V\_Linksは語彙構成を測定するテストであると考えられるというものである。第2の解釈は、得点の差は語彙構成の違いを反映したものではなく、2つのグループの語彙サイズの違いを反映しているというものである。この解釈では、改訂版V\_Linksは語彙構成のテストではなく、語彙サイズのテストと考えられる。

### 3.1.3 新しい語彙構成のテスト:LOT

望月(2006)は、V\_Linksの問題点として、つながりの定義があいまいなことと正解作成の妥当性の欠如の2点を挙げている。V\_Linksでは、つながりの定義を明確に示していない。その結果、何がつながりを構成するかについては受験者の判断に委ねられる。受験者のつながりの解釈はいちじるしく異なる。ただ単に2語が同じ品詞というだけで、つながりがあると考えた受験者もいる一方で、つながりであるとの判断がきわめて厳格でほとんどつながりはないと判断する受験者もいた。このように、つながりの定義が不明確であることは、受験者の回答行動に大きく影響し、受験者の語彙構成に関する知識を適切に抽出しているとは考えがたい。第2の正解作成の妥当性の欠如とは、V\_Linksでは英語母語話者20名にテストを実施し、その4割以上が共通して選択したつながりを正解としている。しかしながら、英語母語話者の大多数が選択した回答ならばともかく、4割以上という数値には、英語語彙知識の構成を測定しようとするテストの正解として妥当なものであるかについて疑問が残る。

Wolter (2005)は、V\_Linksの回答は1)英語母語話者と英語学習者のどちらもが選ぶつながり、2)英語母語話者は選ぶが英語学習者は少数しか選ばないつながり、3)英語学習者は選ぶが英語母語話者はほとんど選ばないつながりの3つに分類できるとしている。彼は1)は概念によるつながり、2)は英語特有のつながり、3)は学習者の母語に特有なつながりであり、英語母語話者と英語学習者を最もよく区別するのは、2)英語特有のつながりであるとしている。この考えに基づけば、V\_Linksで問題点とされた、つながりの定義のあいまいさを解消することができる。すなわち、英語特有なつながりはコロケーションと考えることができるので、それを選択する方式を語彙構成のテストに利用すればよいことになる。V\_Linksの第2の問題点、正解作成の妥当性の欠如は、英語母語話者の大多数が回答として選択するという基準を設ければ解決できる。

望月(2006)は、V\_Linksの問題点の解決策を踏まえ、新しい語彙構成のテスト、Lexical Organization Test (LOT)を提案している。このテストの作成方法は、まずJACET8000の最も頻度の高い1,000語から無作為に120語を選び、その120語に対してそれぞれCobuild Collocation Samplerからtスコアの高い共起語100語を集めた。120語それぞれに対して、100語の共起語リストから恣意的に2語を取り出し、1)sale

condition, 2) sale item のような 2 組のコロケーションを作成した。120 組のコロケーションを 20 名の英語母語話者に提示し、それぞれのペアに関して、どちらのコロケーションのほうがより強く結びついているかという質問をした。20 名中 17 名(85%)以上がより強く結びついていると一致した回答した 50 組を LOT に使用することとした。LOT は、図 3.4 に示すように、目標語と 2 つの共起語をコンピュータ画面上に三角形上に提示し、受験者は最も強く結びついていると考える 2 語を選ぶ形式のテストである。

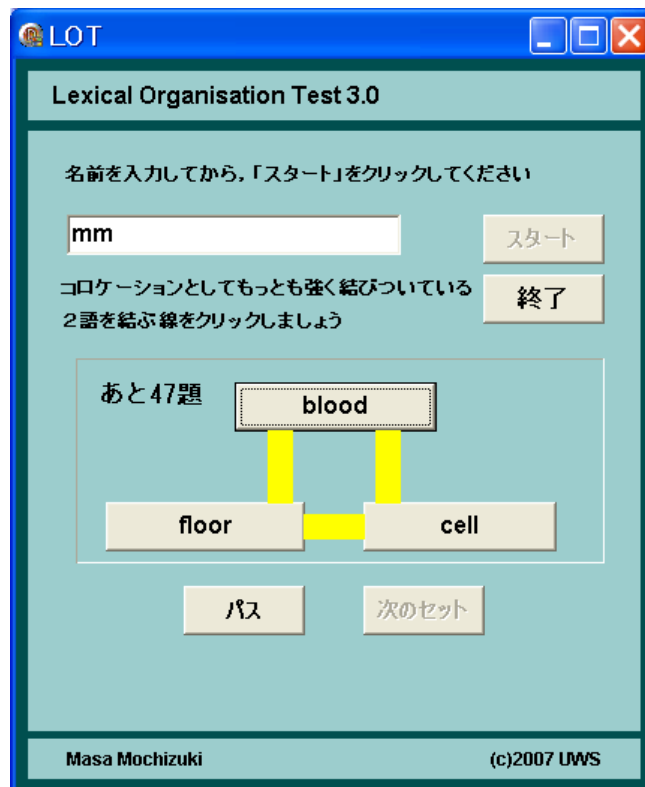


図 3.4 LOT

図 3.4 の例では、blood, floor, cell という 3 つの語が作り得る 3 つの語の組合せ、blood floor, blood cell, floor cell の中から最も強く結びついているコロケーションはどれかを選択しなければならない。この例では blood cell が最も強く結びついているコロケーションで、正解となる。LOT は 50 問から成り、各問 20 秒の制限時間内に回答しなければならない。

LOT が語彙構成を測定するテストという主張は、2 つの前提に基づいている。まず十分に発達した語彙構成をもつ英語話者は、そうでない英語話者よりも、単語間により多くの結びつきを見出すことができるという前提である。この前提に基づけば、blood, floor, cell という 3 語が提示されたときに、十分に発達した語彙構成をもつ英語話者は、この 3 語のもととなった、blood cell, cell floor という 2 組のコロケーションを見出すことができるはずである。第 2 の前提は、十分に発達した語彙構成をもつ英語話者は、そうでない英語話者よりも、つながりの強さについてより詳細な知識を持っているというものである。この前提に則れば、発達した語彙構成をもつ英語話者は、cell floor よりも blood cell のほうがより強く結びついていると判断できるはずである。LOT はこの 2 つの前提に基づき、語彙構成を測定していると考えられる。

望月(2006)は、LOT を英語母語話者と英語学習者 30 名ずつに実施し、LOT の得点は両者を区別するこ

とができることを明らかにしている。望月はこれは LOT が英語語彙構成を測定する証拠であると主張している。

### 3.2 語彙構成テスト LOT の妥当性・信頼性検証

#### 3.2.1 WAT による LOT の妥当性検証

望月(2006)は LOT を開発し、英語母語話者と英語学習者を区別できることから、LOT は英語語彙構成を測定するテストであると主張している。しかしながら、LOT の得点差は、英語母語話者と英語学習者の英語習熟度の違いを反映しているだけかもしれない。LOT が語彙構成を測定していることの妥当性を検証するのならば、より適切な方法を用いて検証する必要がある。

現在までのところ、語彙構成のテストと呼ばれるものは、上で紹介した V\_Links のみである。このテストが語彙構成を測定していないという立場から LOT が開発されたのであるから、V\_Links を LOT の妥当性検証の基準とすることはできない。語彙サイズと構成のアプローチから離れると、語彙の広さと深さのアプローチにおける深さのテストとして The Word Associates Test (Read, 1993, 1998: 以下 WAT)がある。WAT は 40 の形容詞を目標語として提示し、その語の意味を説明する語とその語と一緒に用いる語を 8 つの選択肢の中から合わせて 4 つ選ばせる形式のテストである。sudden の問題を例に挙げる。

**sudden**

beautiful	quick	surprising	thirsty
change	doctor	noise	school

目標語 sudden の意味を表すものは quick と surprising であり、sudden と一緒に用いる語は change と noise である。このテストは、ひとつの単語の持つさまざまな意味や共に使うコロケーションの知識を問う、語彙の深さを測定するものと考えられている。したがって、WAT を併存的妥当性検証の基準として用いることで、LOT が語彙構成を測定するテストとして妥当なものであるかどうかを判断することができると考えられる。

本研究は、LOT の結果を WAT の結果と比較することで、LOT の語彙構成テストとしての妥当性を検証することを目的とする。

#### 3.2.2 方法

関東地区 2 大学、関西地区 1 大学の 3 大学から 82 名の大学生が研究に参加した。彼らの英語習熟度は初中級レベルである。参加者は、各大学で筆記版 WAT を 40 分間で受験した。その後、コンピュータ教室で LOT を受験した。関東地区 1 大学の学生 18 名は、コンピュータ教室が使用できなかったため、LOT を研究者の研究室で受験した。LOT の受験時間は平均して 15 分程度である。

#### 3.2.3 結果

LOT と WAT の平均、標準偏差は、表 3.1 の通り、LOT は 50 点満点中 20.9 (標準偏差 5.0)、WAT は 160

点満点中 99.9 (標準偏差 13.4)である。LOTとWATには弱いが有意の相関が見られる( $r = .38, p < .01$ )。

図 3.5 は LOT と WAT の得点散布図である。

表 3.1 LOTとWATの基礎統計量

	LOT	WAT
平均	20.9	99.9
標準偏差	5	13.4
最大値	30	135
最小値	11	54

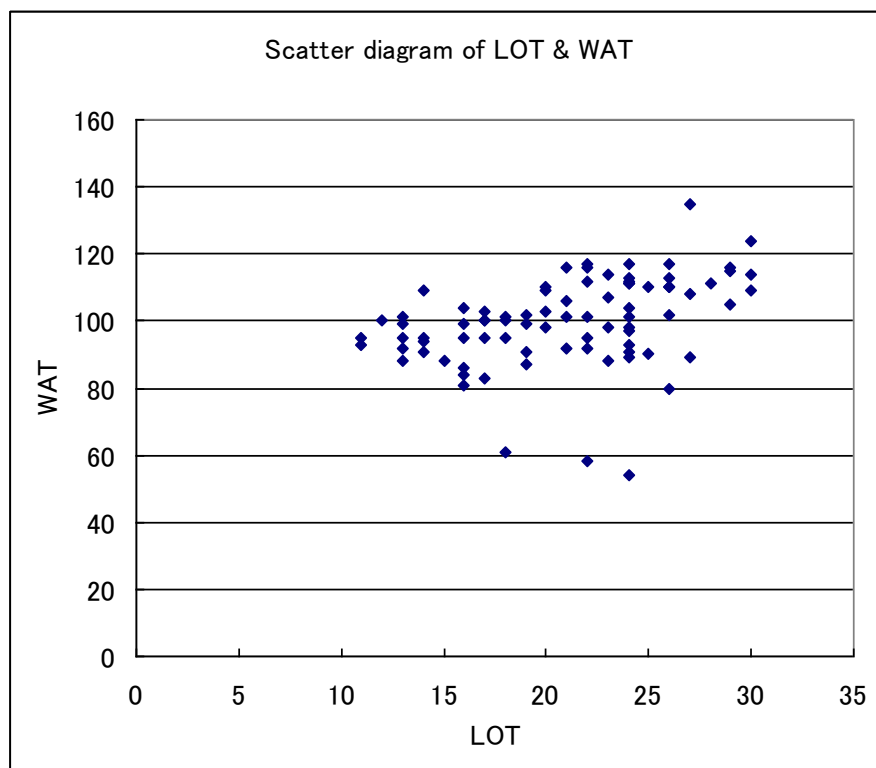


図 3.5 LOTとWATの得点散布図

### 3.2.4 考察

LOTは、併存的妥当性検証の基準としてのWATと $r = .38$ という有意の相関関係にある結果が得られた点について考察する。

まず、.38という弱い有意の正の相関関係が2つのテストにあることは、LOTもWATと同様に語彙知識の深さを測定しているものと考えられる。したがって、LOTは語彙構成を測定するテストとして妥当なものであると言える。しかしながら、ひとつのテストの得点がもうひとつのテストの分散に占める割合を示す決定係数は.14であり、きわめて小さい。これはひとつのテストの得点がもうひとつのテストの得点の14%程度しか説明できないことを意味している。このように考えると、LOTとWATは有意の相関があったとしても、かなり異なる能力を測定していると考えたほうがよい。

つぎに、2つのテスト間に強い相関関係が見られなかったこと、言い換えると、前段落で述べたように、2つのテストは、かなり異なる能力を測定していると考えられることについて考察する。

第1の原因として、WATはLOTほど異なる能力の受験者を弁別していないことが挙げられる。表3.1から求められるWATの変動係数は、.13であり、きわめて小さい。これは、WATが異なる能力の受験者をよく弁別していないことを示している。それに対して、LOTの変動係数は、.24とWATよりはるかに大きい。これは、WATのほうが異なる能力の受験者をよりよく区別していることを意味する。このことは、図3.6と3.7のLOTとWATの得点分布図から明らかに読み取れる。WATの得点は平均点付近に集中しているのに対して、LOTの得点は広く分布している。

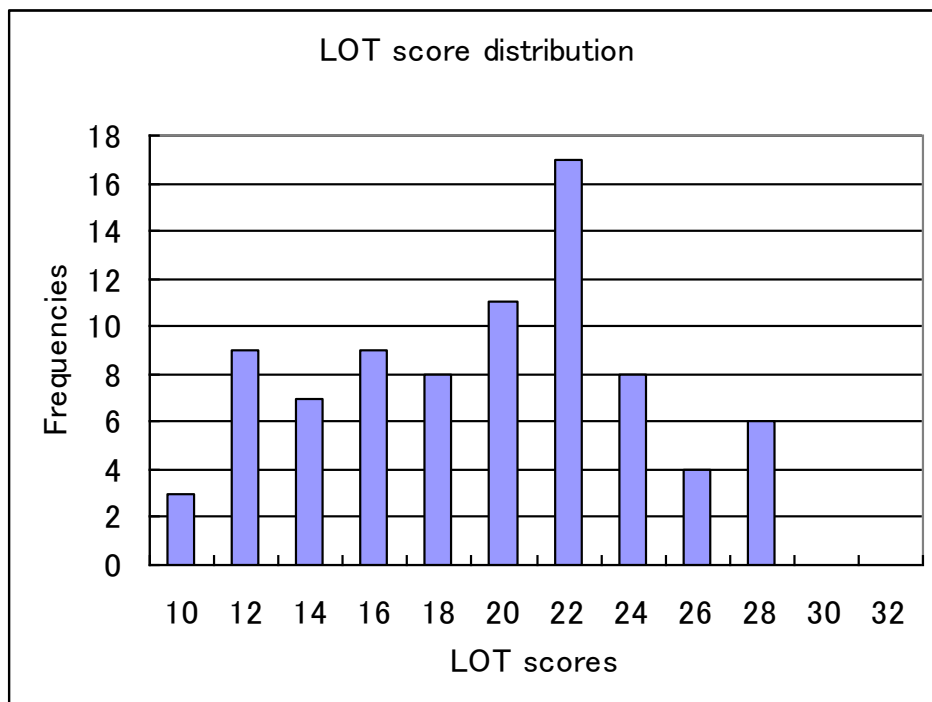


図 3.6 LOT の得点分布



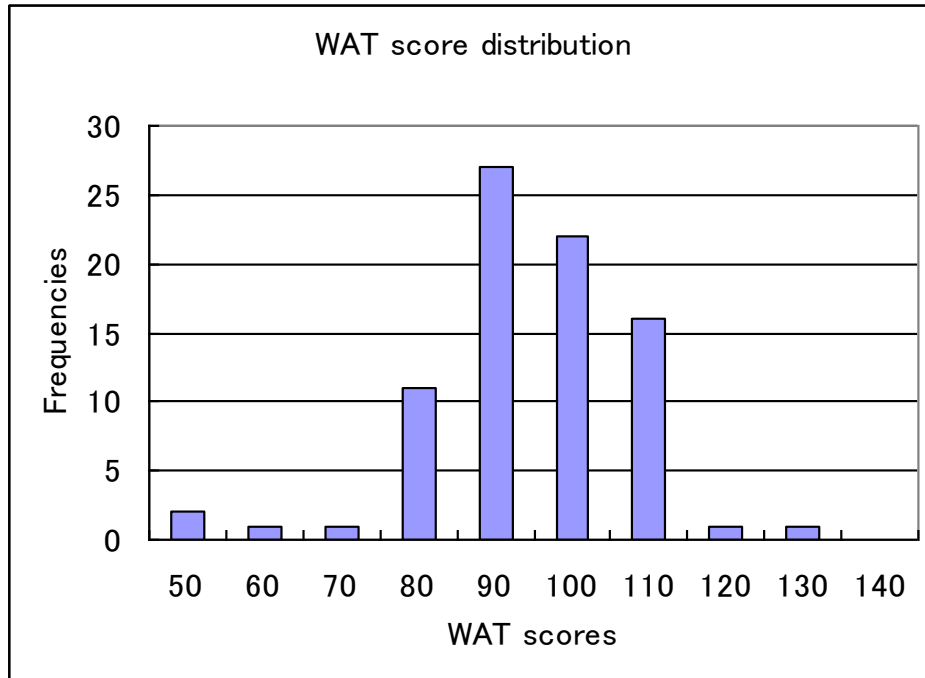


図 3.7 WAT の得点分布

この WAT の弁別力の低さは、あてずっぽうによる正解の可能性の高さが原因のひとつであると考えられる。WAT は、8 つの選択肢の中から 4 つの正解を選択する形式なので、まったくあてずっぽうに回答したとしても、50%の確率で正解できることになる。それに対して、LOT は、3 つの選択肢から 1 つの正解を選択する形式であるから、あてずっぽうによる正解確率は 33.3%である。この推量による正解確率の高さが、LOT よりも WAT のほうが能力の異なる受験者を弁別できない原因のひとつであると推察できる。

LOT と WAT の相関係数が 0.38 とそれほど高くない第 2 の原因は、WAT の目標語や選択肢に頻度の低い語彙が使われていることであると言える。言い換えると、WAT は語彙知識の深さだけではなく、低頻度語を知っているかいないかで得点が左右される語彙知識の広さ、サイズ、のテストでもあることが原因と考えられる。WAT では、40 の目標語のうち、5 分の 1 にあたる 8 語 (acute, chronic, crude, domestic, fertile, synthetic, coherent, ample) が低頻度語である。さらに、選択肢にも数多くの低頻度語が使われている (legal, smelly, rubbish, continuing, unplanned, solid, artificial, plentiful)。この主張を裏づけるために、LOT と WAT の目標語と選択肢の頻度レベルを JACET8000 のレベルで調べてみた。結果は、表 3.2 の通りである。

表 3.2 LOT and WAT choices の JACET8000 レベル

	1	2	3	4	5	6	7	8	NIL*
LOT	120	27	2	1	0	0	0	0	0
WAT	175	110	34	15	12	6	2	3	3
WTW*	6	10	8	7	6	2	1	0	0

\*WTW = WAT Target Words NIL = Not in the list

LOT の 150 語の選択肢のうち、120 語(80%)は最も頻度の高い 1,000 語である JACET8000 のレベル 1 に属し、27 語(18%)はレベル 2 に属している。すなわち、LOT でテストされる 150 語の 98%の語は JACET 順位 2,000 番以内の語である。それに対して、WAT の 360 語の目標語と選択肢のうち、175 語(49%)はレベル 1、110 語(31%)がレベル 2 に属している。すなわち、JACET 順位 2000 番以内にあるのは 80%で、残りの 20%は 2,000 語を越えた語彙ということになる。表 3.2 が示すように、レベル 3、4、5 だけでなく、レベル 6 以上の語や JACET8000 にない語も使われている。さらに、WAT の刺激語 40 語のうち 2,000 語以内の語は 16 語(40%)にすぎない。残りの 60%の語は 2,000 語レベル以上の語である。このような頻度だけでも WAT のほうが LOT よりも頻度の低い語を使用していることは明らかであるが、これを確証するために、LOT の語が使われているレベル 1 から 4 までに関して、LOT と WAT の語彙の頻度に差があるかどうかを調べるためにカイ 2 乗検定を実施した。結果は、1%水準で有意差が見られた( $\chi^2 = 36.57$ ,  $df = 3$ ,  $p < .001$ )で両者に有意の差があることが判明した。これにより、WAT は LOT より頻度の低い語を有意に多く使用していることが実証された。WAT の目標語の 60%は JACET 順位 2,000 番以降の語であり、2,000 語以上の語彙サイズがないと目標語の 60%はわからないことになる。さらに、目標語と選択肢の 20%は同様に 2,000 語以上の語彙レベルに属するもので、WAT で正しく解答するためには、語彙知識の深さと同様に、より大きい語彙サイズが求められていることになる。すなわち、WAT は語彙知識の深さのテストであると同様に、語彙サイズのテストであるとも言える。それに対して、LOT で使われている語の 98%は JACET 順位 2,000 番以内の語であり、LOT に正しく解答するためには、大きな語彙サイズは必要ではなく、より深い語の知識のみが求められる。これらから、LOT と WAT が低い相関関係にある第 2 の原因は、WAT は目標語や選択肢に頻度の低い語を用いることによって、語彙知識の深さを測定しているだけでなく、結果的に語彙サイズも測定していることと考えられる。

LOT と WAT に弱い相関しかないことの原因について 2 点考察した。1 つの原因は、WAT は推測による正解確率が 50%と LOT の 33%より高いために、能力の異なる受験者をよく弁別できていないことが考えられる。第 2 の原因は、WAT は最も頻度の高い 2,000 語以外の単語が目標語で 60%、目標語と選択肢の 20%で使われていて、これらの単語を知らないと WAT で正解するのがむずかしい。そのため、WAT は語彙知識の深さのテストであると同時に、語彙サイズのテストにもなっている。この 2 点から、2 つのテストの相関が低かったものと推定される。この 2 点について、LOT は語彙構成のテストとしてより望ましい特徴、すなわち推量による正解確率が低い、高頻度語のみを使用、を備えており、語彙構成のテストとして妥当なものであると考えられる。

### 3.2.5 結論

今回の研究は、語彙知識の深さのテスト WAT を併存的妥当性検証の基準として用いることで、新しい語彙構成のテスト LOT の妥当性の検証を試みた。2 つのテストの得点には、有意の弱い相関が見られ、LOT は語彙構成のテストとしてある程度妥当性があるものと考えられる。しかしながら、2 つのテストの決定係数は 0.144 ときわめて低く、両者は異なる能力を測定していると推定される。WAT は推量による正解確率が 50%と高いこと、最も頻度の高い 2,000 語以外の語を目標語で 60%、全体でも 20%と大量に使用していることから、語彙知識の深さ以外の能力を測定している可能性が大きい。これが LOT との相関を低めている原因と考えられる。この点を考慮すると、LOT のほうが語彙構成のテストとしてより妥当なものと主張できる。

### 3.2.2 プロトコル分析による LOT の妥当性検証

#### 3.2.2.1 語彙知識テストの妥当性検証

前節では併存的妥当性の観点から LOT の妥当性検証が行われた。本節では、受験者の内的プロセスと受験者の反応の観点から LOT の妥当性について検討する。

妥当性とは、「テストで測れるように意図した能力(構成概念)をどの程度測れているか」と一般には定義される。しかし、厳密には「テスト得点に基づいた解釈と使用の適切さが、実証的証拠と理論的根拠によって裏づけられる程度」または「その程度を統合的に評価した判断」(Messick, 1989)である。テストには高い妥当性が求められ、妥当性検証では、妥当性が高いという証拠を複数の観点から挙げ、妥当性の高さについて主張することになる(e.g., Bachman, 2005; Chapelle, Enright, & Jamieson, 2008; Kane, 2006)。妥当性検証にはさまざまな方法があるが、表 3.3 のように、6 つの要素にまとめられる。

表 3.3 Messick (1996)の妥当性の 6 要素と分析手法

要素	妥当性検証で吟味する側面	具体的な分析手法の例
内容的	テストの内容と測定領域の内容が一致しているか	タスク・カリキュラムの分析, 専門家の判断, 分析的枠組みの使用
実質的	受験者が項目・タスクにどう反応しているか	観察・アンケート, 面接・プロトコル分析による質的分析, 受験者が用いる方略・言語の報告, 談話分析
構造的	仮定された構成概念の構造と, データの構造が一致するか	項目応答理論, 信頼性, 因子分析, 構造方程式モデリング
一般化可能性	時間・グループ・受験状況・タスク・評価者の変化などにより, テストのプロセスと構造が変化するか	信頼性, 一般化可能性理論, 差別的項目機能(DIF), 分散分析
外的	テスト得点とテストの別のテスト・変数との関係はどうか	相関, 多特性・多方法の方法, 因子分析, 構造方程式モデリング, 分散分析
影響的	テスト得点を解釈・使用する際に社会的な影響はあるか	教室での観察・インタビュー・質問紙などの質的方法, 新聞などのメディアに現れる市民の意見の探索, 対象テストの記述, 介入前と介入後のテスト得点の比較, 指導用教材や記録の分析

注: 下線部が本節で用いる手法。小泉(2008b)より引用。

本報告書の他節においては、さまざまな受験者や教育文脈における「テスト間の相関」や「異なるグループ間での得点の比較」(ともに外的要素の検討)が報告されている。本節では、妥当性のより広い検討を目指し、(発話)プロトコル分析とアンケートを用いた実質的要素の分析を行う。

テスト受験時のプロセスを調べるために、今までに多くの研究がプロトコル分析を用いてきた。たとえば、語

彙テスト(Hoshino, 2009; Morimoto, 2007; Paul, Stallman, & O'Rourke, 1990), 文法問題(Dávid, 2007; 小泉, 2008a), C-test(Grotjahn, 1986), クローズテスト(Sasaki, 2000), リーディングテスト(Cohen & Upton, 2007; Nikolov, 2006; Rupp, Ferne, & Choi, 2006), リスニングテスト(In'nami, 2005; 島田, 2006), ライティングテスト(Nikolov, 2006)のプロトコル研究がある。また, 「受験者がテスト中に実際に用いているプロセスが, テスト作成時に意図したプロセスとどの程度一致するか」という観点は, テストの妥当性検証の中でも重要とされる。Kane (2006)は, テスト作成時に意図したようにテスト得点を解釈・使用するために必要なステップを述べている。Kane は, テストの信頼性が高かったときに次に必要になるステップとして, プロトコル分析などを用い, テスト得点の中に「得点の解釈を深刻にゆがめる, 構成概念とは無関係におこるばらつき(分散)がない」証拠を示すよう求めている。これは, たとえば, 語彙構成が測りたい能力(構成概念)であるテストにおいて, 全然別の知識や能力(例: 無関係に見えるパターンを読み取る能力)を持つか否かで得点が変わることがないという証拠を示すことである。そして, 構成概念とは関係ない分散がない(またはそのような分散が少ない)ことを示す証拠や, 構成概念と関係した分散が多いことを示す証拠を挙げられたときに, テスト得点(universe score)が意図したように解釈でき, 構成概念を示す得点(target score)として解釈できることになる(外挿: extrapolation: 小泉, 2008b 参照)と Kane は述べている。

### 3.2.2.2 語彙知識の構成と深さの測定

語彙知識の捉え方は研究者によってさまざまだが, その中に, (a)サイズと構成(Meara & Wolter, 2004), (b)広さと深さ(Anderson & Freebody, 1981)という2つがある。後者はより広く使われている(Read, 2004)が, 前者は後者の問題を指摘して提唱されたアプローチである(3.1.1 節参照)。

語彙構成の測定方法については, 2種類のテストがあり, V\_Links (Meara & Wolter, 2004)と Lexical Organisation Test (LOT: 語彙構成テスト)である。LOTはV\_Linksの改善を試みた形式であり, コンピュータ上で行う(3.1.3 節参照)。LOTでは, コロケーションになり得る3つの組み合わせの中から, もっとも強いコロケーションの組み合わせを選ぶことになる。テストされるコロケーションは英語特有のものである。LOTは50問から構成され, 各問20秒の制限時間がある。

語彙知識の深さの測定方法については, さまざまな方法が提唱されてきた(e.g., Read, 2000; Koizumi, 2005)。その中でよく使用されるのは The Word Associates Test (Read, 1993, 1998: 以下 WAT)である。WATは, 40の形容詞について, 複数の意味や同義語(synonym)と, その語と一緒に用いる(コロケーションとなる)語を, 8つの選択肢の中から4つ選ぶ形式である。同義語とコロケーションとなる語をそれぞれ何語選ぶかについては指定されていない。WATは, Readが作成したもとのテスト形式やその修正版が, Greidanus, Bogaards, van der Linden, Nienhuis, and de Wolf (2004), Qian (1999, 2002), Qian and Schedl (2004), Schoonen and Verhallen, 2008)などさまざまな研究に用いられている。

LOTとWATは2つの異なる語彙知識の捉え方に基づいたテストである。そのため測ろうと意図する構成概念についても異なり, WATは1語1語におけるコロケーションの知識が焦点であるのに対し, LOTではコロケーションを媒介にして語彙の全体的な構成を推論しようとする。しかし, 2つのテストは表面的には「コロケーションの知識」という構成概念を共有していると思われる。では, この2つのテストが引き出すプロセスは, どのような点が類似し, どのような点が異なるのであろうか。また, テスト受験者はどのような点が類似・相違してい

ると感じるのだろうか。LOT と WAT のテスト受験時のプロセスを調べた研究は知る限りまだない。そのため、本節でその点を吟味する。本節の目的は、LOT の妥当性検証の 1 観点として、LOT と WAT の受験時におけるプロセスを比較することにより、2 つのテストの構成概念、特に LOT の構成概念をより明らかにすることである。リサーチ・クエスチョン (RQs) は以下の 2 つである。

RQ1: LOT と WAT の受験時のプロセスはどのように類似し、異なるか。

RQ2: 受験者は、LOT と WAT のどのような点が類似し、異なると感じるか。

RQ1 は発話プロトコルを、RQ2 はアンケートでの受験者の回答を分析することで検討する。

### 3.2.2.3 研究方法

#### 受験者

関東地区の私立大学1校に通う 5 人の大学生と、大学の英語教員 3 名のあわせて 8 名が受験した。著者の呼びかけに対し受験を許可した 8 名であり、終了後謝礼を渡した。

表 3.4 のように、8 人の語彙サイズは 2240 語の幅(2653~4893 語)があった。語彙知識の広さ推定が 4,000 語台の参加者(P1~P4)を上位群、4,000 語未満の学習者(P5~P8)を下位群と考えた。

#### テストと手順

Cobb (n.d.)を利用して作成した筆記版 WAT (Read, 1998)と LOT (version 2.0)を用いた(3.2 節参照)。以下のような詳細な指示を与えた。

#### LOT の指示

これは単語が頭の中がどのようにつながっているかを見るテストです。3 つの単語がコンピュータ画面に現れます。コロケーションとしてもっとも強く結びつく 2 語を見つけてください。その 2 語を結んでいる線をクリックしてください。間違えたら、正しいと思う線をクリックしなおせば、訂正できます。どの 2 語がもっとも強く結びついているかわからない場合は、「パス」をクリックしましょう。全部で 50 セットあります。1 つのセットは 20 秒間表示されます。早く次のセットに進むときには、「次のセット」をクリックしましょう。それでは「次」をクリックしてください。

#### WAT での指示

問題の語と同じ意味を表す単語を①~④の中から、問題の語と一緒に使える単語を⑤~⑧の中から、合わせて 4 つ選びなさい。

例: sudden

- |            |         |              |          |
|------------|---------|--------------|----------|
| ①beautiful | √②quick | √③surprising | ④thirsty |
| √⑤change   | ⑥doctor | √⑦noise      | ⑧school  |

sudden は「突然の、思いがけない」という意味なので、quick, surprising がそれと近い意味を表します。また、sudden change, sudden noise という表現があります。したがって、②, ③, ⑤, ⑦の4つをマークします。①~④と⑤~⑧から選ぶ回答数の組み合わせは、1-3, 2-2, 3-1 のいずれかになります。

実施の際には、順序効果を相殺するために2つのテストの順番を半々に入れ替えた。プロトコルでは「頭に浮かんだことをそのまま口に出す」よう、特に、答えを選んだ理由は必ず述べるよう求めた。練習をした後に、参加者は2つのテストを受験しながら、同時にプロトコルを産出した。具体的な手順などは以下の通りである。

## 参加者への説明・指示

### (1) 調査の目的

・どのように問題を解いているか、その過程を調べることです。正答できたかどうかよりも、テストを解く過程に興味があります。この目的のために、テストを解いていただきます。

### (2) 本日、行っていただくこと

- ・テスト用紙への書き込み・メモは自由です。
- ・問題を解きながら、頭に浮かんだことをそのまま口に出しながら(頭の中の様子を実況中継しながら)、テストに解答してください。特に、答えを選んだ理由は必ず述べてください。分からない場合は「このところは分からないな。」で構いませんので、どんどん口に出してください。また、日本語・英語が混ざっていても構いません。不完全な表現になっても構いません。
- ・発話が意味のあるものか否かは、小泉が判断します。したがって、自分で意味のあるなしを判断せずに頭に浮かんだことすべてを報告してください。
- ・あなたの発話をもとに、あなた自身を評価することはありません。
- ・調査のためには自然な発話が必要です。頭に浮かんだことをどんどん口に出してください。
- ・発話の時間制限はありません。できるだけたくさん話してください。
- ・発話は全てテープに録音します。
- ・テスト後、アンケートを受けていただきます。
- ・問題番号を言ってください。
- ・ある程度考えて分からないときには、次へ行って構いません。
- ・小泉は同席しますが、あなたの発話についてあいづちをうったり、訂正したりすることはありません。実際はいいものと考え、自分のペースで発話を進めてください。
- ・テスト中に小泉が質問することがあります。
- ・問題番号を言ってください。

テスト終了後、受験者は、2つのテストの違いなどを尋ねるアンケートに答えた。

表 3.4 参加者の特徴: 語彙知識の広さ推定値が高かった順

No.	性別	専攻・職業	広さ	LOT 素点(%)	WAT 素点(%)
P1	男	英語教師	4893	22 (44.00)	102 (63.80)
P2	男	英語教師	4787	34 (68.00)	135 (84.40)
P3	男	英語教師	4787	32 (64.00)	138 (86.30)
P4	女	英語専攻	4613	24 (48.00)	140 (87.50)
P5	女	英語専攻	3827	20 (40.00)	75 (46.90)
P6	女	英語専攻	3653	23 (46.00)	81 (50.60)
P7	男	英語専攻	2733	18 (36.00)	44 (27.50)
P8	女	国際学専攻	2653	10 (20.00)	2 (38.80)
		平均	3993.25	22.875 (45.75)	97.13 (60.73)
		標準偏差	924.40	7.62 (15.25)	37.36 (23.36)
		満点	5000	50	160

注: 「広さ」は, J8VST(第2章参照)による推定で, レマ換算。P4がP1からP3までとあまり変わらないのは, 得点の天井効果と思われる。上位群と下位群の平均(標準偏差)は, 上位群で広さ 4770.00 (609.52), LOT 素点 28.00 (5.56), WAT 素点 128.75 (16.38)で, 下位群で広さ 3216.50 (115.98), LOT 素点 17.75 (5.89), WAT 素点 65.50 (17.95)で, 効果量の違いは非常に大きかった (広さ  $g = -3.08$ ; LOT  $g = -1.55$ ; WAT  $g = -3.20$ )。

## 分析

テープに録音したプロトコルは書き起こされ, Dávid (2007)に基づき表 3.5 のコード表を作成した。コード表は 6 つのカテゴリーからなり, そのうち, 2 から 5 までは, ある程度明確に何かを言及した「明示的な根拠の言及」に属している。

コーディングは, テスト項目についての言及のみを, 受験者が語った言葉に基づき行い, 語っていないことを推論してコーディングはしなかった。最初に言ったことを後で訂正しているときには訂正後の発言をコーディングした。テストの 1 項目ずつコーディングし, 2 つ以上のカテゴリーに当てはまる言及があった場合にはすべてをコーディングした。たとえば, 以下の例では, 2, 4, 3 とコーディングした。

(LOT) lay, stone, table (下線は正解)

「石でできたテーブル [(2)]。あるいは, 石を置くで, ええと lay stone といえるから [(4)]。

うーどうしようかな。どっちが強っていったら [(3)]」(P1)

評価者間信頼性については, 分析基準を話し合った後, テスト項目の 30.36% (510/[(50+160)\*8]; 510/[(LOT 項目数+WAT 項目数)\*8 人分]) を 2 人の評価者が独立に評価した。一致度で 78.43% (400/510), Cohen のカッパ係数で  $k = .73$  だった(算出手順は水本, n.d.を参照)。そのため, 評価者間信頼性は高いと考えた。不一致の部分は話し合い, 最終決定をした。残りは著者が判断した。

表 3.5 プロトコルのコード表

カテゴリー	説明・例
(1)言及なし	何も言わなかった場合、答えに関することを言わなかった場合(これにコーディングしたときには、他のカテゴリーにはコーディングしなかった)  例:ちょっとわからない、パス、(見出し語について)聞いたことがない、読めない
(2)意味の言及	全部の訳を言った場合、訳の一部を言った場合、意味について言及した場合(訳が間違っているがどの語を言っているかが捉えられる場合を含む)  例:~の意味がわからない、同じ意味のものを探そう。意味が違ったかな。何かの意味になった気がする
(3)複数のつながりの比較の言及	複数の組み合わせを比べて、つながりの強さを言及した場合  例:~よりはこの方が強い(気がする)、どっちが強いんだろう、一番言いそうだから、どっちとも結びつかないから、~の方が頻度的に強いから、全部つながりそう、全部可能性があるから
(4)単独のつながりの言及	単独の組み合わせについて、つながるかどうかを言及した場合  例:~は言うから、~は言わないから、これはある(ない)から、よく見る(見ない)から、聞いたことがあるから、熟語(コロケーション)だから。voteしたら popular になるから、関係がわからない、Country is independent (英語で言えるかを考えていると判断)
(5)日本語でのつながりの言及	例:~は日本語で言う(言わない)から、日本語でよく見る(見ない)から、日本語でこういう使い方をするから、日本語で関係がわからない、日本語で意味が通じるから、~は英語でも言うのかな
(6)その他	日本語か英語かはっきりしないつながり、何かをはっきりと言及しなかった場合(暗示的な根拠の言及)、消去法を使った場合(明示的に「消去法」と言っている時だけ)、音声不明瞭で聞こえない場合  例:なんとなく、直感で、これはよいから、自信があるから、勘で、ひびきからこっちかな

分析では、各受験者に対し、「カテゴリー内での使用数÷項目数」を計算した。たとえば、P5 の受験者は、LOT で(1)のカテゴリーを使った数が 5 個で、LOT の項目数が 50 個のため、その受験者は .10 (5÷50)となる。その後、全体での平均値を、全項目を入れた場合と、受験者が正答した項目だけで分析した場合で算出した。受験者が正答した項目だけで分析したときの方がより違いが見えやすかったため、本節ではそれを報告した。



プロトコル分析では参加者の人数が少ないのが普通であり(e.g., Dávid, 2007; Rupp et al., 2006 とともに  $N = 10$ ), 本節でも 8 名だった。そのため, 人数の違いで結果にあまり影響を受けないといわれている効果量 (Grissom & Kim, 2005) で解釈した。効果量とは, 平均値の差の程度など, 有意性検定の結果からではわからない「効果の程度・違いの程度」を示す値で, 最近使用されるようになってきている(e.g., Kline, 2004; 水本 & 竹内, 2008)。本節では, 効果量の中でも, 人数が少ないことでの偏りを修正する, バイアス修正済みの  $g$  (Grissom & Kim, 2005) を用い, Effect Size Calculator (CEM Centre, Durham University, 2008) で算出した。その解釈は, Cohen (1988) の基準に基づき, (絶対値で)  $g = .20$  以上  $.50$  未満ならば小程度,  $g = .50$  以上  $.80$  未満であれば中程度,  $g = .80$  以上であれば程度が大きいとし, 中程度以上のものを記述した。なお, もし今回の参加者人数 8 人で有意性検定 (paired  $t$ -test, 2-tailed) を 6 回行った場合は, 第 1 種の誤りを  $\alpha = .008$  ( $.05/6$ ), 2 つのテストの違いの程度が中程度とすると, 検定力 (power; 誤った帰無仮説をきちんと棄却できる割合) は 6.01% であり (G\*Power 3.0.10 を使用して計算; Faul, Erdfelder, Lang, & Buchner, 2007), 望ましいとされる 80% からは程遠く, 本来違いがある部分を見落とす可能性が高いことになる。しかし, 参考までにその結果も表 3.6 に挙げる。なお, 多変量分散分析については, 前提が満たされず使えなかった。記述統計等では SPSS (2003) を使用した。

### 3.2.2.4 結果と考察

#### プロトコル分析から (RQ1 の検討)

表 3.6 は, 答えが正解だった場合での LOT と WAT の各カテゴリーの使用割合を示している。効果量を見たところ LOT と WAT で中程度以上異なっていたのは, (3), (4), (6) だった。大きな効果量が見られたのは, (3) 複数のつながりの比較を言及したカテゴリーであった (LOT で 7%, WAT で 1%)。受験者が正解できた場合には, LOT では WAT よりも, 受験者が語のつながりを複数見つけ, どちらのつながりが強いかを多く考えていたことになる。よって, LOT は WAT よりも複数のリンクの比較を引き出せることが示された。複数のリンクの比較を引き出し, 語彙構成の知識を引き出すことは LOT の作成時に意図されていたことであり, そのような傾向が WAT より高かったことは, 作成意図がテストに反映されている証拠である。より高い割合で複数のつながりを比較させることができたのは, LOT では複数つながる可能性のある語をテストで提示し, 「コロケーションとしてもっとも強く結びつく 2 語」を見つけるよう指示したのに対し, WAT では複数つながる可能性のある語は LOT と同様に提示しているものの, コロケーションとしてつながるものはすべて選ぶため, どちらが強いというよりは, 語と語につながりがあるのはどれかを受験者は探すためだろう。ただし, LOT ではこのカテゴリーに当てはまったのは 7% のみで, WAT の 1% より高いものの, 一部に限られることがわかった。

表 3.6 正解のみの場合の割合での記述統計(N= 8)

カテゴリー	LOT		WAT		$g^a$	$t(7)$	$p$
	平均	$SD$	平均	$SD$			
(1)言及なし	.12	.07	.13	.16	-0.09	-0.22	.84
(2)意味の言及	.64	.18	.66	.26	-0.07	-0.33	.75
(3)複数のつながりの比較の言及	.07	.07	.01	.01	<u>1.18</u>	2.93	.02
(4)単独のつながりの言及	.24	.18	.36	.18	<u>-0.68</u>	-1.47	.19
(5)日本語でのつながりの言及	.01	.02	.01	.01	0.07	1.00	.35
(6)その他	.08	.12	.14	.10	<u>-0.59</u>	-2.20	.06

注:  $SD$  = 標準偏差. <sup>a</sup>Bias corrected  $g$ . 下線は,  $g = |0.5|$  以上. 1 テスト項目で複数のカテゴリーがコーディングされる可能性があるため, 平均の合計は 1.00 にならない。

次に, LOT と WAT の比較で効果量が中程度異なっていたのは(4), (6)だった。(4)単独のつながりについては, LOT よりも WAT で言及されていた(LOT で 24%, WAT で 36%)。単独のつながりを引き出すことは, WAT ではテストの目的に入っているため, WAT に肯定的な妥当性の証拠と考えられる。LOT では, 構成概念として一番意図されているのは「複数のつながりの比較」だが, 単独のつながりを 24%言及したとしても, LOT の妥当性を低める証拠にはならないと考える。それは, 受験者が複数のつながりを述べるときに, 単独のつながりを追加で言及することもあるだろうし, 複数のつながりの中の 1 つが圧倒的に強いと感じた場合には, (たとえ複数のつながりを頭の中で比較していても口頭では)単独のつながりのみを言及することもあるだろうからである。

(6)その他の言及については, LOT よりも WAT での割合が中程度高かった(LOT で 8%, WAT で 10%)。「その他」のカテゴリーに入った回答を分類してみると, 一番多かったものは, 「なんとなく」など何かをはっきりと言及しなかった場合で, LOT では 6%, WAT では 9%だった。

今まで効果量で中程度以上の違いが見られたカテゴリーについて述べてきた。次に, 使用された割合が高かったものについてまとめる。最も高かったものは(2)意味への言及であり, LOT で 64%, WAT で 66%と 6 割以上のテスト項目で意味が言及されていた。次に高い割合は, (4)単独のつながりの言及で, その次は(1)言及なしのカテゴリーだった。もっとも低かったカテゴリーは(5)日本語でのつながりの言及であった。(5)は, 2 つのテストともに 1%と少なく, テストでも測ることを意図していないため, この割合の低さは肯定的な妥当性の証拠である。

(1)言及なしが LOT で 12%, WAT で 13%だったことは, プロトコル分析では避けられないのかもしれない。なぜなら, 頭の中の実況中継をするよう受験者が頼まれたとしても, すべてを口にするとはいならず, また自動化されているプロセスは言及しにくい(できない)場合もあるだろうからである。

2 つのテストともに, 意味の言及が 6 割以上を占めていたことは, テストの構成概念である「複数のつながりの比較」と「単独のつながり」の言及の割合を単純に足して LOT で 31%, WAT で 37%であったことと比較すると, 高い割合と思われる。意味の言及が多いという結果は, 語彙の広さ・サイズではなく語彙構成・深さを測ることを意図する 2 つのテストの妥当性が低いことを示すのだろうか。

まず、WAT については、問題として提示する語の中に同義語が入っており意味が密接にかかわるため、問題ではない面もあるだろう。しかし、3.2.4 節で示唆されたように、WAT の目標語や選択肢に頻度の低い語彙が使われており、広さを測る度合いがより高いと考えるため、意味の言及の中には問題となるものもあろう。なお、WAT の項目を同義語とコロケーションを測る項目に分けて分析すれば、さらにコロケーションの項目を解く時のプロセスがわかるだろう。しかし、WAT は同義語から2語、コロケーションを2語などと数が決まっているわけではないため、受験者によっては同義語を選ぶところでコロケーションを考えていることもあるため、分離しての分析は難しいと判断して行わなかった。

LOT については、JACET8000 のもっとも頻度が高い1,000語から核になる語を選び、WAT とは違って頻度が高い語彙で構成されていたが、かなりの割合で意味の言及が見られた。この結果は、LOT では意味が構成概念には入っていないため、問題だと表面的には見えるかもしれない。しかし、必ずしもそうとは言えないかもしれない。その理由は2つある。第1に、受験者がテスト中に行う認知プロセスをすべてプロトコルとして産出するわけではないため、第2に、LOT には時間制限があり、話そうとするプロセスをすべて話し終えられないこともあるためである。第1点目は、LOT だけでなくWAT にもあてはまるだろう。

まず、受験者がテスト中に行う認知プロセスをすべてプロトコルとして産出するわけではないというプロトコル分析の限界点の関連で述べる。表3.7は、複数のつながりの比較をもっとも多く受験者(3名)から引き出した項目において、どのようなプロトコルが出ていたかを示している。その中の「意味の言及」にコーディングしたプロトコルを見ると、「科学的なお話」、「んー、お話って感じ」など、受験者が意味だけしか考えていなかったのか、それとも「単独のつながり」を考えたり、「複数のつながりの比較」をしたりした後に、訳だけを言及したのかはわからない。これは「単独の組み合わせ」のプロトコルでも同様で、「news story なんちゃらって言う。film などでよく言うね」と言った受験者は(a)(news story の頻度が高いなどために)他の組み合わせは見ずにそれを選んだのか、それとも、(b)他の組み合わせを比較し、他が news story よりもあまり言わないことを確認してから、news story を選んだのかかわらず、内的に後者の「複数の組み合わせの比較」をしている可能性もある。そのため、意味の言及が半分以上あったことが、テストの妥当性を直接低める証拠にはならないだろう。

表 3.7 LOT のある項目における 8 人のプロトコル

---

(LOT) news, science, story (下線は正解)

(1)言及なし(時間切れで、答えを選択せず。P8)

(2)意味の言及

科学のお話(science story を選択。P4)

んー, お話って感じ(news story を選択。P6)

科学のニュース。ニュース。テレビのニュース(news story を選択。P7)

(3)複数のつながりの比較

まあ, 同じカテゴリーに入るとすると, 話とニュースの方が結びつきが強いかな(news story を選択。P1)

news science かな。どっちもいけるってことかな(news story を選択。P2)

science news. storyよりはこっちの方が強い気がするので(news science を選択。P5)

(4)単独の組み合わせ

news story なんちゃらって言う。film などによく言うね(news story を選択。P3)

---

第 2 に, LOT の時間制限のためにつながりと言及できなかった可能性について考察する。表 3.8 は, プロトコルで(2)意味を言及した場合について, (2)だけを話しているか, その他のカテゴリーとともに話しているかを割合でまとめたものである。これを見ると, ほとんどが(2)が単独で産出されており, 意味を言っただけのプロトコルが多かったことを示している。ここで注目したいのは, WAT が LOT よりも(2)意味とともに(4)単独のつながりを言った割合が高かった点である。これは, WAT の特徴を示しているのかもしれないが, WAT には時間制限がなく, LOT にはあることを考えると, LOT では意味を言ったところで時間切れになり, つながりの言及までできなかったとも解釈できる。この点を厳密に検証するならば, 受験者が LOT の答えを選んだ後コンピュータ画面が次の問題に切り替わるのをストップし, 言えるプロセスを受験者が言い終わってから次の間に切り替える方法でプロトコルを全部引き出す方法はあるだろうが, 本節で使用した LOT は自動的に次の問題が提示されるために, 検証は難しかった。

表 3.8 正解のみの場合の割合での、意味の言及を行った割合の平均と標準偏差

	(2)	(2)+(3)	(2)+(4)	(2)+(5)	(2)+(6)	(2)+(4)+(6)
LOT	.50 (.21)	.04 (.06)	.07 (.08)	.00 (.00)	.02 (.02)	.00 (.00)
WAT	.37 (.18)	.00 (.01)	.21 (.20)	.00 (.00)	.06 (.05)	.01 (.01)

今まで 8 人の受験者の平均での結果で解釈してきた。次に、表 3.9 のように受験者を語彙知識の広さで分けてプロセスの違いを調べた。その結果、全員の受験者で分析したときと傾向が異なっていた点は 2 点あった。第 1 に、(3)複数のつながりの比較について、上位群と下位群ともに効果量で大きな違いがあった(下位群  $g = 0.82$ ; 上位群  $g = 1.45$ )が、上位群の方が違いの程度がより大きかった。つまり、上位群の方が下位群よりも、LOT で複数のつながりを比較した割合が高かった。これは、単一のつながりにおいては上位群・下位群間であまり違いが見られない(下位群  $g = -0.64$ ; 上位群  $g = -0.60$ )のと対照的だろう。上位群の方が下位群よりも LOT の得点が高く、点数の違いが大きかった( $g = -1.55$ ; 表 3.4.注参照)ことをあわせて考えると、上位群の方が LOT で複数のつながりの比較をより多く行うという結果は、3.1.3 節で述べた LOT の前提の 1 つの「十分に発達した語彙構成をもつ英語話者は、そうでない英語話者よりも、単語間により多くの結びつきを見出すことができる」ことの証拠となるだろう。

表 3.9 正解のみの場合の割合での記述統計: 語彙知識の広さで 2 群に分けた場合

	下位群( $n = 4$ )					上位群( $n = 4$ )				
	LOT		WAT		$g^a$	LOT		WAT		$g^a$
	平均	$SD$	平均	$SD$		平均	$SD$	平均	$SD$	
(1)言及なし	.12	.03	.06	.06	<u>1.09</u>	.12	.10	.20	.21	-0.44
(2)意味言及	.66	.21	.74	.23	-0.33	.62	.18	.57	.30	0.19
(3)複数つながりの比較	.08	.10	.01	.02	<u>0.82</u>	.07	.05	.01	.01	<u>1.45</u>
(4)単独組み合わせ	.21	.08	.31	.17	<u>-0.64</u>	.27	.25	.42	.19	<u>-0.60</u>
(5)日本語	.01	.03	.01	.02	0.10	.00	.00	.00	.00	NA
(6)その他	.10	.17	.20	.09	<u>-0.61</u>	.05	.04	.09	.10	<u>-0.51</u>

注:  $SD$  = 標準偏差. NA =  $SD$  が 0 のため、計算できなかった。

第 2 に、(1)言及なしのカテゴリにおいて、下位群では LOT で WAT よりも多く使用されており、違いの程度は大きかった(下位群  $g = 1.09$ )。一方、上位群では WAT で LOT よりも言及が多かったが、小さな程度の違いのみだった(上位群  $g = -0.44$ )。下位群で LOT での言及なしが多かったのは、時間制限がある中でプロセスを言語化できなかったためかもしれない。実際、下位群の受験者 1 名は「理由を述べる前に次問に進んでしまったことが度々あった(P5)」とアンケートで述べた。

最後に、本節と 3.2.3 節での結果の比較を試みたい。3.2.3 節では LOT と WAT の相関は  $r = .38$  ( $N = 82$ ) だった。本節では、人数が少ないがやってみたところ、 $r = .81$  ( $N = 8$ ,  $r_s = .86$ ) だった。また、本節では

J8VSTとLOT・WATの間では高い相関があり、LOTでは $r = .81$  ( $N = 8$ ,  $r_s = .71$ ), WATでは $r = .90$  ( $N = 8$ ,  $r_s = .73$ )だった。

LOTとWATの相関が3.2.3節と本節で異なるのはなぜだろうか。3つの可能性があるだろう。第1に考えられる理由は、受験者層の違いである。3.2.3節では大学生のみが対象で、LOTの得点の幅は10～28点だった。一方、本節では大学生と英語教師が参加し、幅は10～34点で、本節の方が人数は少ないものの、能力幅は広がった。

第2の理由としては、テスト受験時にプロトコルを行っていたか否かが挙げられる。プロトコルを行うことによって、(a)その中で学習が起こり(reactivity; Leow & Morgan-Short, 2004), (b)認知的負荷とそれに伴う疲労度が高くなり、(c)当て推量をする割合が低くなる可能性があると考えられる。本節の研究ではプロトコルを行うことで、学習者がテスト中に語のつながりに気づくことがあったかもしれない。また、プロトコルを産出することで通常テストを受けるよりも認知的負荷がかかり、疲労も普通よりは高くなったかもしれない。特にLOTについてのアンケートでは、「答えた後、なぜ選んだのかを答え、また問題を答えるというのが、時間制限がある中で頭の切りかえが難しかった(P8)」, 「パソコン [注:LOTのこと]は話すことを忘れてしまいそうになってしまうことがあった(P7)」と答えた受験者がおり、LOTでは特に認知的負荷がより高かったと思われる。最後に、当て推量をする割合については、確率的にはWATでは当て推量が半分の確率(4/8)で可能で、LOTでは当て推量が3分の1のみ可能である。しかし本節の研究ではプロトコルを人に聞かれることを意識したため、わからないときに適当に答える割合は、プロトコルを出さないときよりは少なくなっていた可能性もあるだろう。これらのテスト受験時にプロトコルを行うか否かの違いが、テスト間の相関の結果に何らかの影響をもたらした可能性もある。また上記で述べた本節での結果も、プロトコル生成により何らかの影響を受けている可能性があることは否めない。

LOTとWATの相関が3.2.3節と本節で異なる第3の理由は、テストの信頼性の違いがあるかもしれない。人数は少ないが本節のデータを使って算出したところ、LOTで $\alpha = .84$ , WATで $\alpha = .98$ と高い信頼性があったが、3.2.3節では報告されていない。

### アンケート分析から(RQ2の検討)

受験者が2つのテストの類似点・相違点をどのように捉えるかについて調べるために、アンケートで「2つのテストで同じ点、異なる点は何だと思えますか?」と「2つのテストを比較して何か感じることは何かありますか?」と尋ねた。その結果、テスト間の類似点としては、「コロケーションを問う所(P1)」, 調査形式の類似点として「頭の中で起こっていることをなるべく正確に話さなくてはいけない点(P4)」が挙げられた。

一方違いについては、LOTには時間制限があることを4名が挙げ、「制限時間があるのがつらかった(P7)」, 「WATは考える時間が自由だったので、パソコンのテストよりやりやすかった(P4)」などの感想があった。さらにWATの難易度はLOTよりも高いと4名が述べた。難易度が異なると感じる理由は、「WATの選択肢が[注:8個と]多い(P2)」ことのみ記述があったが、LOTは1問につき1つの答えを選び、WATは4つ選ぶという違いも受験者の捉え方に影響したかもしれない。また、コンピュータを使うか否か( $n = 2$ ), 「意味が問われるかどうか(P1)」も違いとして挙げられていた。なお、LOTの構成概念であり、RQ1での検討の中でWATともしっかりと大きく違っていた「複数のつながりの比較をするか否か」については、受験者からのアンケートでの言及

はなかった。

難易度については、受験者は WAT が LOT より難しいと感じていたが、実際は LOT の方が難しかった (LOT の正答率の平均は 46%, WAT は 61%)。このギャップは、LOT では、必ずしも正しいとは限らないが何らかの回答は選べるのに対し、WAT では、8 つの中から 4 つの回答を全部選ぶのが難しかったためだろう。また 3.2.4 節で述べたように、WAT の目標語や選択肢に頻度の低い語彙が使われているために難しく感じたのかもしれない。

### 3.2.2.5 結論

本節では、LOT の妥当性検証の 1 観点として、LOT と WAT 受験時におけるプロセスを比較することにより、2 つのテストの構成概念、特に LOT が測る構成概念を明らかにすることが目的だった。RQ1 の「LOT と WAT の受験時のプロセスはどのように類似し、異なるか」について、受験者が正答できた場合の受験時のプロトコルを調べたところ、主に相違点が 2 点、類似点が 2 点示された。第 1 の相違点は、LOT は WAT よりも、複数のつながりの比較を多く引き出せていたことである。これは特に語彙知識の広さが 4,000 語を越えた学習者において見られた。ただし、複数のつながりの比較を引き出した割合は、他のカテゴリーと比べると一部に限られていた。第 2 の相違点は、WAT は LOT よりも単独のつながりの言及が多かったことである。次に、第 1 の類似点は、意味への言及が 2 つのテストともに割合がもっとも高く、6 割以上を占めたことである。第 2 の類似点は、もっとも言及が少なかったのは、日本語でのつながりだったことである。

次に、RQ2 の「受験者は、LOT と WAT のどのような点が類似し、異なると感じるか」についてまとめる。2 つのテスト形式の類似点として挙げられたのは、コロケーションをテストする点であり、相違点は、LOT に時間制限があること、WAT が LOT よりも難易度が高いと感じられること、LOT でのコンピュータの使用、WAT で意味が問われる点だった。

これらの結果を妥当性検証と関連づけると、RQ1 での第 1 と第 2 の相違点は、LOT と WAT の構成概念から予測されることであり、妥当性の肯定的な証拠となるだろう。RQ1 での第 2 の類似点の「日本語でのつながり」の言及が少なかったことは、構成概念とは関係ない分散が少ないことを示す証拠(3.2.2.1 節参照)となり、これも妥当性の肯定的な証拠となるだろう。RQ1 の第 1 の類似点の「意味への言及」が多かった点については、プロトコルの解釈の限界と LOT の時間制限の存在のために、妥当性については寄与しないだろうと考える。そのため、全体的には 2 つのテストの妥当性の肯定的な証拠がいくつか示されたという結果であり、その証拠を基に、LOT と WAT の妥当性は高いという主張をより強く行うことができると考えられる。

今回のプロトコルとアンケートを用いた分析により、テスト間の得点の相関や、グループ間の得点の相違だけでは見えない、受験者の認知プロセスのある側面が吟味できたと考える。しかし、3.2.2.4 節の「意味の言及」の解釈で述べたように、プロトコル分析には、受験者が自分で言語化できたものしか分析できず、解釈できる範囲も限られることがあるという限界点もあり、さらに他の手法を用いて検討していく必要があるだろう。また、LOT と WAT のさらなる詳細な分析をするために、プロトコル分析の対象者を増やすこと、各テストの特徴をさらに見るために他の語彙テストと比較すること、妥当性検証の観点からは、先行研究に挙げた妥当性の観点をより包括に検討していくこと、テストの妥当性の証拠をより多くの観点から挙げていくことが求められる。また、本節での研究は、2 つのテストの妥当性検証に焦点を当て、その中で見られた違いをプロセスの違いとして述

べた。しかし、2つのテストはテスト形式が異なるだけでなく、そのテストに入っている語彙も異なる。そのため、テスト形式の違いによってテスト受験のプロセスがどのように変わるかを厳密に調べるためには、テストで使用する語を統制する必要がある。今後プロセスに焦点をより当てる場合には、その面を取り入れた検討が必要だろう。



### 3.3 語彙構成テストによる TOEIC リーディング得点・リスニング得点の予測

#### 3.3.1 研究の背景

語彙は一般的言語習熟度の信頼できる指標と考えられている。語彙で読解テスト得点のある程度を予測できることが報告されている(Koda, 1989; Laufer, 1992; Zhang & Annual, 2008)。前章で述べたように JACET8000 に基づく語彙サイズテスト J8VST は、TOEIC リーディング得点と  $r = .72$  の相関がある。語彙サイズは TOEIC リーディング得点の分散の約半分を説明できることになる。Henriksen, Albrechtsen and Haastrup (2004)によると、60名の ESL 学習者の語彙サイズと読解との間に  $r = .79 \sim .85$  の相関が見られたが、そのうち4名は例外的で語彙サイズが大きいにもかかわらず読解得点が高い、あるいは、語彙サイズが小さいにもかかわらず読解得点が高かった。後者は優れた推測能力と発達した語彙ネットワークのためではないかと推察している。同様に聴解もある程度予測できる。Bonk (2000)は語彙サイズテストと再生による聴解テスト得点の間に  $r = .45$  の相関があったことを報告している。しかしながら、これらの研究で問われている語彙知識は、サイズという側面のみである。語彙知識は、サイズ、深さ(構成)、認知速度という3つの側面から問いかけるべきであるが、サイズ以外の側面と言語技能の関係を調査した研究はごく少数である(Qian, 1999, 2002; Qian & Schedl, 2004)。

Qian (1999)は、語彙知識の深さもサイズと同様に読解に関与していると論じて、語彙サイズ、深さ、読解の3つの関係を調査した。語彙サイズが3000ワードファミリー以上の74名の中国人・韓国人 ESL 学習者が、4つのテストを受験した。テストは、TOEFL 読解テスト(4つの文章で設問20項目)、語彙サイズテストとして the Vocabulary Levels Test (Nation, 1990)、語彙知識の深さのテストとして the Word Associates Test (Read, 1998)、もうひとつの深さのテストとして自作の語彙形態素テストである。結果は、読解と語彙サイズの間で  $r = .78$ 、読解と WAT による語彙深さとの間で  $r = .82$ 、読解と形態素知識との間で  $r = .64$  のいずれも有意の相関があることがわかった。彼は、WAT による語彙深さのテストは、語彙サイズよりも読解をよりよく説明できるとしている。

Qian (2002)は、19の異なる母語を持ち、より広範な文化的背景をもつ217名の ESL 学習者を被験者として Qian (1999)を追試したものである。語彙形態素テストを TOEFL 語彙部門のテストに替えた以外は、同じテストを用いた。結果は、3つの語彙テストは TOEFL 読解といずれも同じような相関をもった。サイズ( $r = .74$ )、深さ( $r = .77$ )、TOEFL 語彙テスト( $r = .73$ )。彼は、1つの語彙テストだけよりも複数の語彙テストを用いることで、TOEFL 読解をよりよく説明できると主張している。

Qian and Schedl (2004)は、the Word Associates Test 形式の語彙知識の深さのテストを TOEFL 語彙部門の語彙を目標語として作成し、それと TOEFL 読解の相関を求めた。結果は  $r = .74$  で、Qian (1999, 2002)を裏付けるものである。

これらの研究は、語彙知識の深さのテストは、語彙サイズテストよりも読解をよりよく説明できる可能性があることを示唆している。懸念があるとすれば、The Word Associates Test を語彙知識の深さのテストとして用いることが適切であるかどうかということである。WAT と The Lexical Organization Test (望月, 2006)と  $r = .38$  の相関しかないことが報告されている(Mochizuki, 2007)。彼は2つの語彙知識の深さのテストの相関が低いことの理由のひとつとして、WAT では低頻度語が目標語や選択肢に多用されていて、WAT は語彙知識の深さだけでなく、語彙サイズテストにもなっていることを挙げている。LOT は、使用されている150語中147語

(98%)は最もよく頻度の高い 2,000 語内であるのに対して、WAT では 360 語中 285 語(79%)でしかない。WAT の残りの 75 語は、JACET8000 のレベル 3 から 8 にわたり、3 語はそのリスト外の語である。このことは、WAT で使用される語彙の 5 語に 1 語は 2,000 語レベル以上であり、語彙サイズが大きくないと高得点が望めないことを意味している。この意味で WAT は語彙知識の深さを測定しているだけでなく、語彙サイズも測定していると言える。

このように考えると、Qian (1999, 2002)は語彙知識の深さは読解に関与していると報告しているが、高い相関は WAT が語彙サイズを測定していることによるのかもしれない。したがって、本研究は語彙サイズに影響されない語彙知識の深さのそのものと読解の関係を明らかにしようとするものである。語彙知識の深さそのものとは、語彙サイズに影響されない語彙知識の深さと定義する。すなわち、2,000 語までの語が互いにどう結びつき合っているかの知識である。望月(2006)が開発した語彙構成テストと TOEIC のリーディング・リスニングの関係を探究することを目的とする。

リサーチクエスチョンは、次の 2 つである。

- (1) 語彙知識の深さそのものと TOEIC の読解の間には、Qian の研究で報告されていると同程度の相関関係が見られるか。
- (2) 語彙知識の深さそのものと TOEIC の聴解の間には、どのような関係が見られるか。

### 3.3.2 方法

#### 3.3.2.1 材料

語彙知識の深さのテストとして、提示された 3 語のうちで最も強いつながりをもつ 2 語を選ぶ The Lexical Organization Test (LOT) の Flash 版を用いる(図 3.8)。Flash 版 LOT は磯達夫氏がプログラムしたもので、テスト項目はオリジナル版と同じであるが、以下の 3 点で異なる。

- 1) 50 のテスト項目が無作為に提示される
- 2) テスト項目の 3 語の位置関係が固定されていない
- 3) 2 語を結ぶ線をクリックして回答するのではなく、中央のボールを 2 語を結ぶ線上へ移動させることにより回答する(図 3.9)。

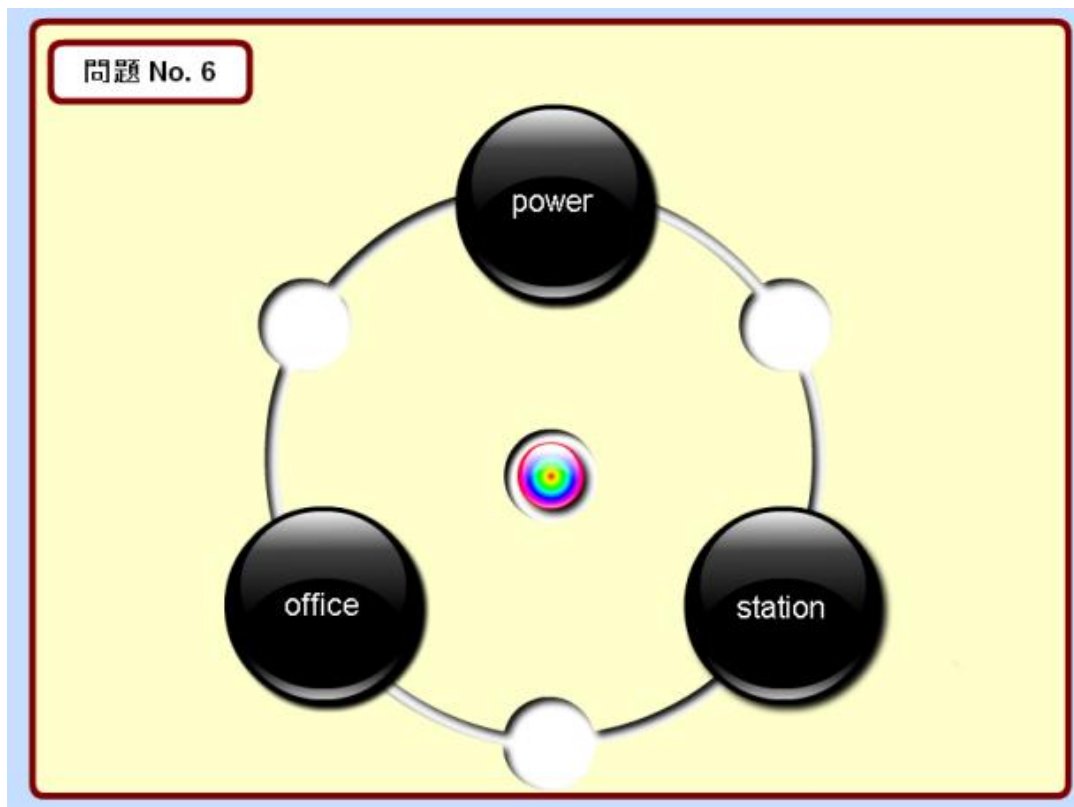


図 3.8 Flash 版 LOT

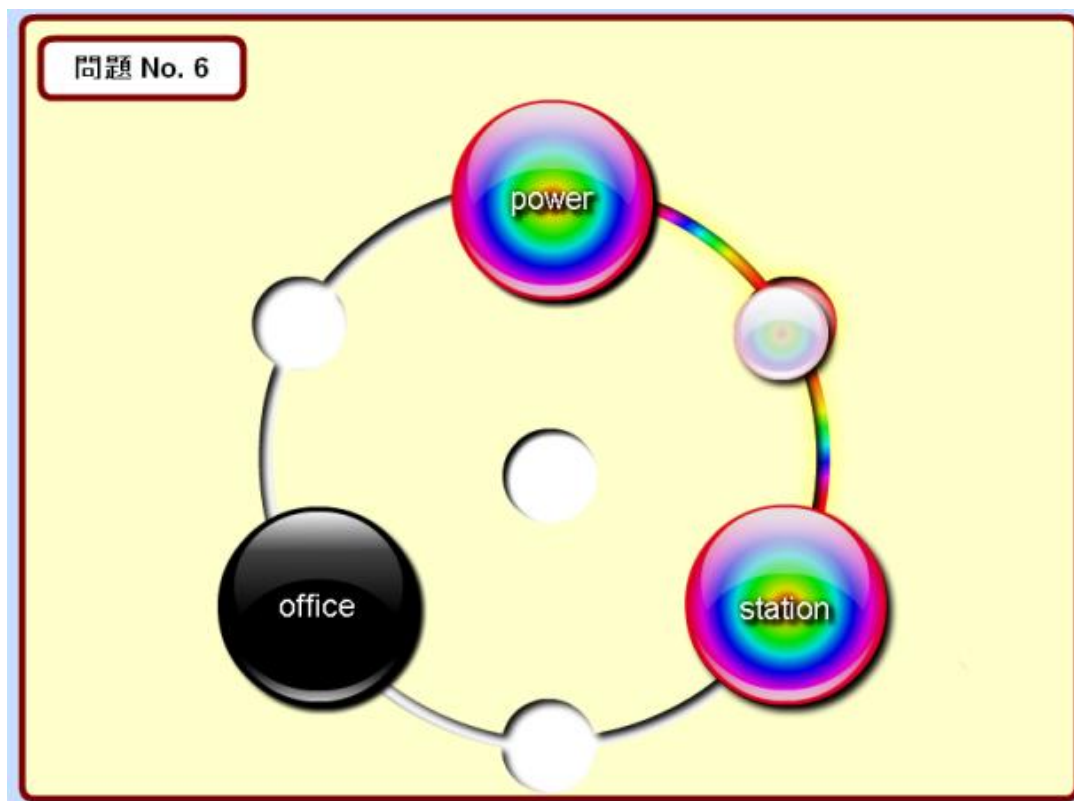


図 3.9 ボールを移動して 2 語を選択

### 3.3.2.2 参加者

日本の6つの大学から総計412名の大学生が研究に参加した。彼らは外国語, 工学, 国際情報など専攻が異なり, 年齢も18歳から22歳までの幅がある。彼らは, 語彙サイズテストとしてJ8VST, 語彙知識の深さ(構成)テストとしてLOT, 語彙認知速度テストとしてThe Lexical Access Time Test (Iso & Aizawa, 2008)をオンラインで受験した。その後, 自分のTOEIC得点をそれぞれの大学の研究分担者に報告した。しかしながら, 半数以上の参加者はいずれかのテストを受験しなかったり, TOEIC得点を申告しなかったりした。結果として, LOTを受験し, かつ, TOEIC得点を申告した187名を分析対象とする。

### 3.3.3 結果

TOEICとLOTの平均と標準偏差は, 表3.10に示す通りである。

表 3.10 TOEIC と LOT の平均と標準偏差

<i>n</i> = 187	TOEIC	Listening	Reading	LOT
<i>Mean</i>	422.2	240.7	181.6	21.0
<i>SD</i>	157.7	80.6	84.9	3.69
<i>Max</i>	895	495	430	29
<i>Min</i>	185	100	60	12

TOEICの平均は, リスニングが240.7, リーディングが181.6, 合計422.2である。最高得点は895, 最低は185である。一方, LOTの平均は50点満点中の21で, 最高得点が29, 最低が12である。

表3.11にTOEICとLOTの相関係数を示す。

表 3.11 TOEIC と LOT の相関係数

	TOEIC	Listening	Reading
correlation	0.55***	0.51***	0.54***

\*\*\* $p < .001$

TOEIC全体では $r = .55$ , リスニングとは $r = .51$ , リーディングとは $r = .54$ である。いずれも0.1%水準で有為の相関である。

図3.10~3.12は, TOEICとLOT得点の散布を示している。

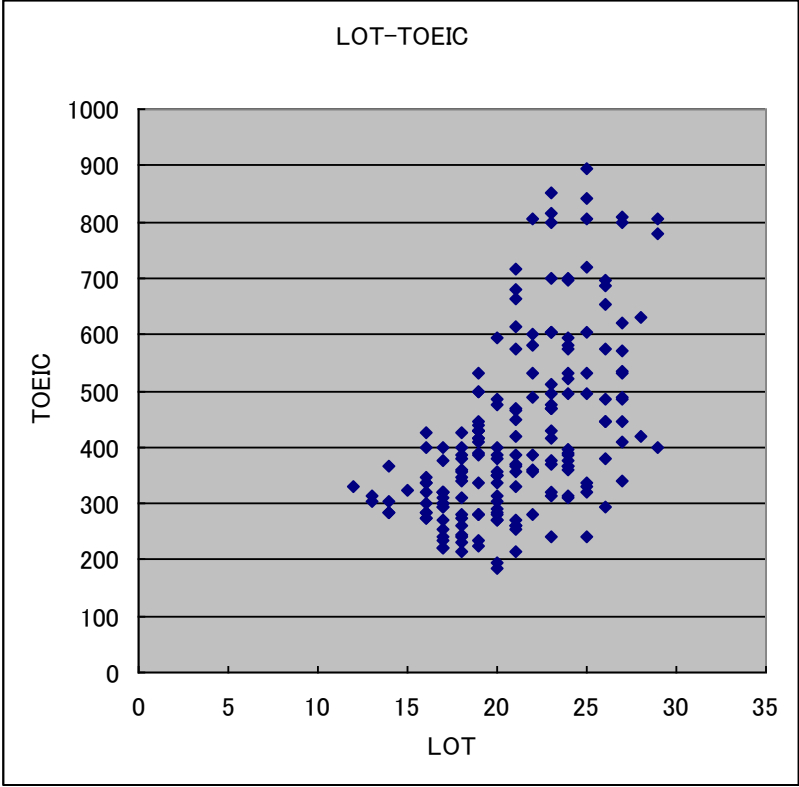


図 3.10 TOEIC 全体と LOT 得点の散布

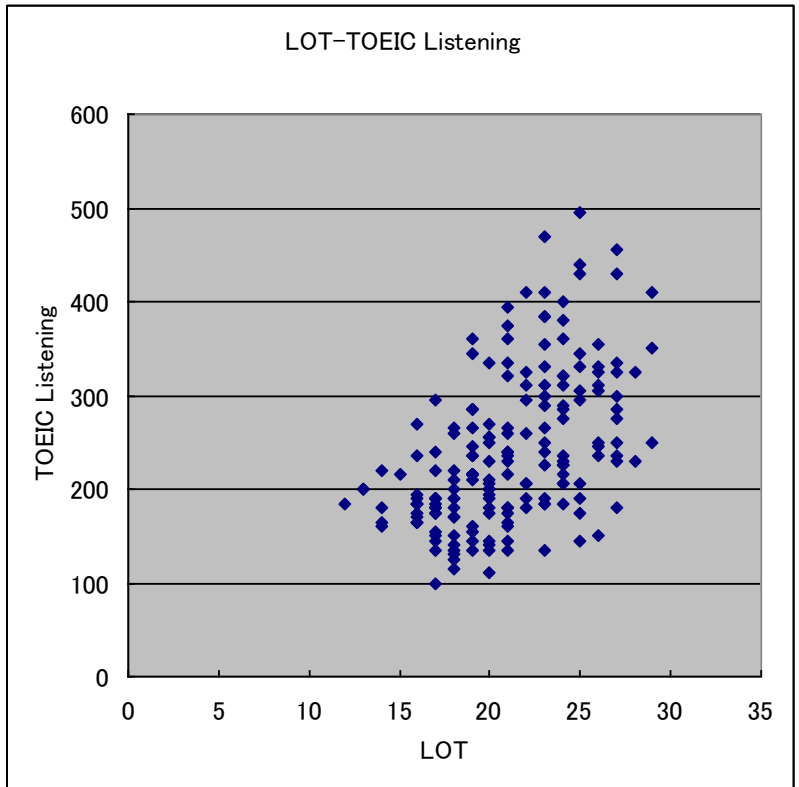


図 3.11 TOEIC リスニングと LOT 得点の散布

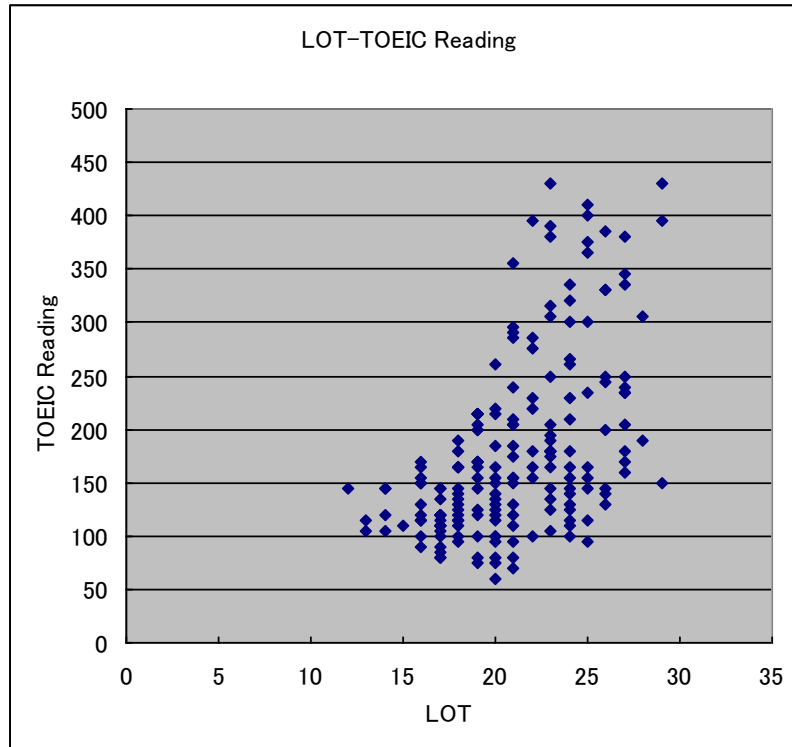


図 3.12 TOEIC リーディングと LOT 得点の散布

これらの結果は、LOT によって測定された語彙知識の深さそのものは、Qian(2002)で報告された TOEFL と WAT の間の相関( $r=.77$ )ほど高い相関を TOEIC 読解との間に持たないことを示している。LOT と TOEIC リスニングの相関は、リーディングとの相関より若干弱い相関をもつことが明らかになった。

### 3.3.4 考察

リサーチクエスチョンに対して結果が示唆することを考察する。まず、語彙知識の深さそのものと TOEIC の読解の間には、Qian の研究で報告されていると同程度の相関関係が見られるかというリサーチクエスチョン1に関して、否定の回答が得られたことについて考察する。これに関しては、LOT が語彙知識の深さのテストとしての構成のされ方と 2 点で関連していると指摘できる。第 1 に、LOT で使用されている語の頻度レベルは、TOEIC のリーディング部門のテキストを理解するのに必要な語彙のレベルを十分カバーしていなかったかもしれない。LOT は 50 の目標語とその共起語を使用している。目標語は最も頻度の高い 1,000 語から採用されていて、その共起語は 3 語を除いて最も頻度の高い 2,000 語から採用されている。最も頻度の高い 2,000 語は語彙知識の深さを測定するには十分ではあるが、読解を測定するには十分ではないのかもしれない。コーパス研究によると、最も頻度の高い 2,000 語は英語のテキストの 80% をカバーすることを示している。これは 2,000 語では 5 語に 1 語はカバーできないことを意味する。したがって、2,000 語で作られた語彙知識の深さ(構成)のテスト LOT は、TOEIC リーディングテストの理解を保証するには十分ではないと言えることができるかもしれない。

第 2 に、表 3.10 から明らかなように、LOT の標準偏差と得点範囲は、TOEIC リーディング得点に比べると、きわめて小さく狭い。LOT は満点が 50 であるのに対して、平均が 21 で標準偏差が 3.69、最高点が 29、最

低点が 12 である。これに対して、TOEIC リーディングは、満点が 495 であるのに対して、平均が 181.6 で、標準偏差が 84.9、最高点が 430、最低点が 60 である。LOT の場合、約 70% の被験者が 17 点から 25 点の範囲内にいる。これは、LOT は語彙知識の深さ(構成)という側面で被験者をうまく弁別していないことを示唆している。この事実が、LOT と TOEIC リーディング得点との間に中程度の相関しか見られない理由のひとつであるかもしれない。

それでは、LOT は被験者を十分に弁別できないのだろうか。習熟度の高い学習者にとってもいくつかの項目は難しすぎるようである。たとえば、TOEIC の最高点は 895 であるが、この得点をとった被験者の LOT の得点は 25 で、これは LOT の満点のちょうど半分である。TOEIC の 2 番目の最高点の学習者は 850 で、LOT の得点は 23 である。これは LOT の満点の半分以下である。このように LOT は、高い英語習熟度の学習者にも難しすぎるために、受験者を語彙知識の深さ(構成)の側面で十分に弁別できていないのかもしれない。

次に、LOT と TOEIC リスニングの得点の相関は、中程度で有意であったが、LOT と TOEIC リーディングのものよりも多少低かった。それでもそれは、Bonk が報告する  $r=.45$  よりも高いものである。しかし、これは単純には比較できない。Bonk は、語彙サイズとリスニング内容の再生の相関であるのに対して、本研究は語彙構成テストと多肢選択式リスニング内容理解問題との相関である。Bonk (2000) の用いた再生テストの方が多肢選択問題よりも難易度が高いと考えられるので、単純に相関は比較できない。

### 3.3.5 結論

本研究は語彙知識の深さのテストと TOEIC リーディングとリスニングテスト得点との関係を調査し、それぞれ中程度の有意の相関関係が見られることがわかった。相関が中程度だった原因としては、LOT が高頻度語しか使用していないことと LOT の弁別力の低さが考えられる。したがって、今後の研究課題としては、2,000 語より頻度の低い単語から成る語彙知識の深さのテストを開発し、そのテストとリーディングとリスニングテスト得点との関係を調査することが挙げられる。別の可能性としては、習熟度の異なる学習者をよりよく弁別できるように LOT を改訂し、それをを用いて、同様の調査を行なうことも考えられる。

## 第4章 アクセス速度測定テストの開発

本章は、このプロジェクトの一環として、語彙知識のうち、語彙アクセス(lexical access)の速度がコミュニケーション能力とどのような関連があるかを明らかにすることを目的とする。

### 4.1 先行研究の概観

#### 4.1.1 語彙知識の記述

語彙知識の記述的研究にはいくつかあるが、本節では代表的な2つの例を示す。まず、Ringbom (1987) は、語彙知識を図 4.1 の通り定義している。語彙知識は、6つの側面から構成され、語彙アクセス(accessibility)、形態音韻論(morphophonology)、統語(syntax)、意味論(semantics)、コロケーション(collocation)、語連想(association)の要素がある。この分類の特徴は、単語への素早いアクセス(accessibility)を取り上げている点である。語彙アクセスも、特定の限られた文脈でのみアクセスが可能なレベルから、文脈に関係なくアクセスすることができる段階までが断続的に存在していると説明している。また、意味(semantics)では、おおよその意味を知っている段階(knows approximate meaning only)から、すべての可能な意味を知っている段階(knows all possible meanings)に至るまでが、連続的に存在していることを示している。

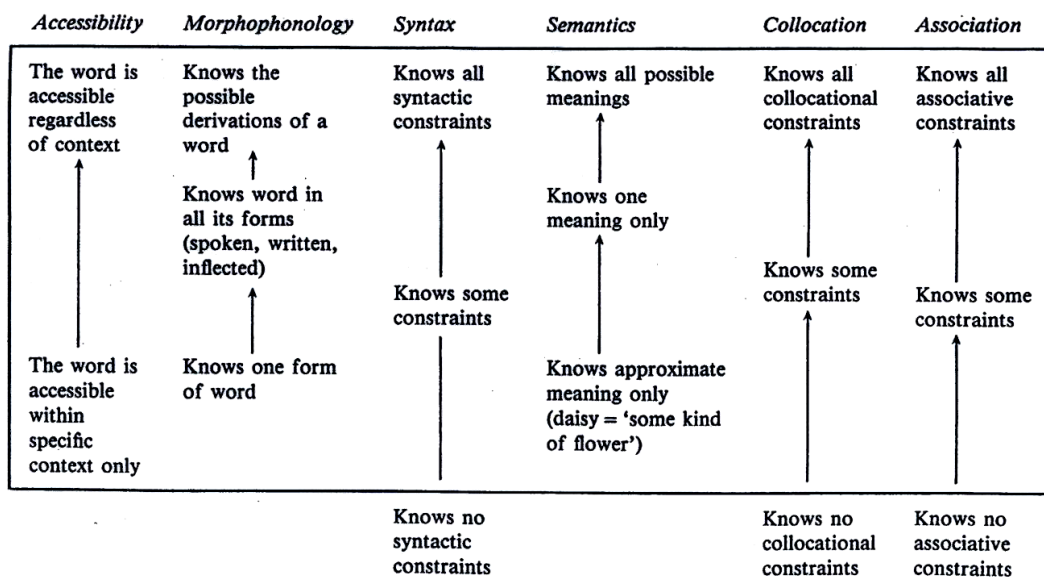


図 4.1 語彙知識の枠組み (Ringbom, 1987, p. 37)

語彙知識の記述としてよく取り上げられものには、Nation (2001)のものがある。Nation は、語彙知識を形(form)、意味(meaning)、使用(use)の3つに分け、それぞれを3つずつに下位分類している。さらに、これらの9つの分類には、産出的(productive)な面と受容的(receptive)な面の2つに分け、合計して18に分類している。

Nation (2001)とRingbom (1987)の語彙知識の枠組みを比較すると、Nationの18分類には語彙への素早いアクセスについて言及がない。しかし、L2のリーディング研究において、Grabe and Stoller (2002)や



Koda (2005)らは、語彙アクセス(accessibility)を重視している。この点を考慮すると、Nation の語彙知識のモデルは、語彙アクセスについての記述が抜け落ちてしまっている。

表 4.1 単語知識の記述 (Nation, 2001, p. 27 より)。

形 (Form)	発音 (spoken)	R	その語はどのように聞こえるか
		P	その語をどのように発音するか
	綴り (written)	R	その語はどんな語形をしているか
		P	その語はどのように綴られるか
	語構成 (word parts)	R	その語にどんな構成要素があるか
		P	その意味を表すためのどんな構成要素を使うか
意味 (Meaning)	語形と意味 (form and meaning)	R	その語はどんな意味を表すか
		P	その意味を表すためにどの語形を使用するか
	概念と指示物 (concept and referents)	R	その語の概念には何が含まれるか
		P	その語はどんな概念を示すか
	語連想 (associations)	R	その語はどんな語を連想させるか
		P	その語の代わりにどんな語を使えるか
使用 (Use)	文法的機能 (grammatical functions)	R	その語はどんな文型で使用されるか
		P	その語はどんな文型で使用できるか
	コロケーション (collocations)	R	その語はどんな語といつしよに使用されるか
		P	その語はどんな語といつしよに使用できるか
	使用時の制約 (constraints on use)	R	その語をどんな文脈でいつどれくらい頻繁に目にするか
		P	その語をどんな文脈でいつどれくらいの頻度で使用できるか

以上のような観点から、英語コミュニケーション能力と語彙アクセス(accessibility) , すなわち語彙への迅速な反応がどのように関係しているかは、今後の調査で明らかにされることが望まれる。

#### 4.1.2 単語の記憶

単語の記憶には、短期記憶(short-term memory), ワーキングメモリ(working memory), 長期記憶(long-term memory)が関与している。短期記憶は、20～30秒間、7単位程度の情報を保持する記憶現象である。例えば、電話番号のようなもので、繰り返さないと忘れてしまう。ワーキングメモリは、長期記憶から取り出した情報や、外部からの情報などを同時に処理する作業領域に対応した記憶領域である(苧坂, 2002)。この記憶は、発声をほとんど伴わない繰り返しである音韻ループ (phonological loop)の存在によって可能となる。この音韻ループに情報を貯蔵する能力は、言語を学習する能力の決定的な要因になっていると思われる(苧坂)。

長期記憶は、ワーキングメモリとは異なり、膨大な容量があり、一度蓄えられた情報は長期間貯蔵される。しかし、同じ長期記憶の中にも、すぐに忘れてしまうものから決して忘れないものまで、情報の保持には差があると言われている。

以上のように、語彙知識は、短期記憶で処理されて、ワーキングメモリを経由し、長期記憶で保持される。記憶に関する研究の結果、単語の学習に関連があるいくつかの原則(例えば、繰り返し、認知的な深さ、イメージ化、動機づけなど)が報告されており、語彙指導や教材作成に生かされている。

長期記憶の中で、特に語彙情報が蓄えられている部分を表すメタファーとして、心内辞書(mental lexicon)という用語が用いられてきた。心内辞書には、語彙の「綴り・発音・意味・統語などに関する情報」などが貯蔵されている。心内辞書に記載された内容は、語彙エントリー(lexical entry)と呼ばれる。これらの情報内容は、人間の情報処理システムで利用できる形式に符号化(coding)され、長期記憶に貯蔵される(御領, 1998)。さらに、単語が視覚的に提示されると、その情報は符号化の段階を経て心内辞書に到達し、既に蓄えられている情報と照合され、意味などの必要な情報が抽出される。このプロセスを、語彙アクセス(lexical access)と呼んでいる(御領)。

#### 4.1.3 語彙知識と語彙アクセス

ここで、関連用語について、定義を明確にしておきたい。まず、語彙アクセス(lexical access)は、単語が視覚的に提示されると、その情報が形態特徴符号、音韻的符号などいくつかの符号化の段階を経て心内辞書に到達し、すでに蓄えられている表象と照合され、意味などの必要な情報が抽出される過程を指す(御領, 1998)。本稿では、以上の観点から、「単語の意味や語の綴りを確認するまでの語彙アクセスの速度」を単語へのアクセス速度と定義することにする。

Potter, So, von Eckardt, and Feldman (1984)は、L1とL2の単語の結びつき方には2種類があると主張している。1つは、L1とL2の単語が直接結びついていて、翻訳はこの直接経路を用いて行われるという仮説(単語連結仮説)である。すなわち、日本人英語学習者の場合は、英単語(L2 単語)から概念表象にアクセスするには、日本語の単語(L1 単語)の語彙表象を経由して結びついているとする考え方である。

もう1つは、L2 単語も L1 と同様に共通の概念表象と結びついていて、翻訳は概念表象を媒介として行われるとする仮説(概念媒介仮説)である。

一般に、L2 の習熟度が低い場合は単語連結仮説が支持され、習熟度が高い場合は概念媒介仮説が支持される(Kroll & Stewart, 1994)。つまり、入門期の段階の学習者の場合は、単語連結仮説によって、L1とL2、またはL2とL1の間の翻訳プロセスを経るが、学習段階が発達するにつれて、L1でもL2でも単語は概念と結びつく度合いが高まり、L1とL2相互の翻訳プロセスへの依存度が低くなるので、語彙アクセスの速度も速くなると仮定できる。言い換えれば、英語力の高い学習者の方が低い学習者よりもアクセス速度が速く、同じ学習者でもよく知っている単語の方が知らない単語に比べてアクセス速度が速いという仮説が立てられる。

#### 4.1.4 アクセス速度に関する先行研究

アクセス速度を測定するためには、Mochizuki (2006)に見られるように語彙性判断課題(lexical decision

task)が用いられることが多い。表 4.2 には、日本語(L1)の語義と英語(L2)の綴りを使ってこのテストを行う場合に特に注意が必要な点がまとめてある。

表 4.2 語彙性判断テスト開発に関する変数 (Taft, 1991 より作表)

頻度効果 (frequency effect)	低頻度語より高頻度語の語彙判断に要する時間が短い。
語彙性効果 (lexical status effect)	文字列を単語と判断する方が非単語と判断するときより、反応時間は短い。
非単語の単語形式性効果 (non-word legality effect)	単語形式の綴り構造を持った Flink のような非単語の方が、LFMNI のようなランダムな文字列に対するよりも、語彙性判断に長い時間を要する。
単語類似性効果 (word similarity effect)	単語形式の綴り構造を持った非単語が実在の単語に類似している場合、その語を非単語と判断するのが困難になる。
反復プライミング効果 (repetition priming effect)	単語に対する反応潜時は、2 回目に提示されると早くなる。例えば、CAMEL が CAMEL に続いて提示されると、提示されないときよりも反応が早くなる。
意味的プライミング効果 (semantic priming effect)	意味的に関係のある単語が先行して提示されると、単語と判断されるのは容易になる。例えば、CAT を認知する潜時は SIX を先行提示したときよりも、DOG を先行提示したときの方が速くなる。
視覚的不鮮明化効果 (visual degradation effect)	文字列の視覚的鮮明度が低下すると、単語／非単語の判断が困難となる。例えば、単語をドットパターンで覆って提示する、刺激の輝度を低下させる、あるいは文字の大文字小文字を交互に並べるなど。

しかしながら、表 4.3 で Balota and Chumbley (1984)などが示すように、語彙アクセス速度の測定するための語彙性判断課題は、「課題に特殊な効果を測定している」という主張もある。

したがって、本研究においては、語彙アクセス速度の測定に語彙性判断課題は用いず、実際に英語として存在している語彙のみを用いることとした。しかしながら、語彙性判断課題を作成する際に留意すべき点のうちには、他の課題についても同様に留意すべき点がある。例えば、特に配慮が必要なのは、頻度効果・反復プライミング効果・意味的プライミング効果などが挙げられる。これらは、次節の語彙アクセス速度測定ツール開発における課題および問題作成時に可能な限り配慮することにした。

表 4.3 語彙性判断課題の適切さ.

Balota & Chumbley (1984, 1985)	語彙性判断課題に見られる基本的な効果の1つである頻度効果は、語彙へのアクセス過程自体ではなく、アクセス後の判断段階で生じると主張。語彙性判断課題では、リーディングにおける通常の語彙へのアクセスには関係のない、課題に特殊な効果を測定していることになる。
Balota & Chumbley (1984)	意味的分類課題と命名課題の結果から、語彙性判断課題で見られる頻度効果は、その課題に特殊な語彙判断後の段階で生じるに違いないと結論。
McCann & Besner (1987), McCann, Besner, & Davelaar (1988)	頻度効果はアクセス中に生じるのではなく、語彙へのアクセス後に作用する刺激の親近性や音韻の類似性から生じる。

## 4.2 アクセス速度測定テストの開発(1)

### 4.2.1 開発の目的

先行研究の概観からも明らかなように、英語学習者の語彙知識を測定するツールのうち、最も開発が進んでいないものがアクセス速度を測定するためのツールである。本研究において、アクセス速度は「語に対して何らかの作業を行う速度」(Laufer & Goldstein, 2004)として定義する。しかし、どのような作業を行う場合の速度が語彙に対するアクセス速度の測定に適しているかという問題に対する答えは明らかになっていない。つまり、アクセス速度は文字列の連続を語として認識するまでの時間を測定するべきか、それともその語の意味を認識するまでの時間を測定するべきか、という点が明らかになっていないのである。

しかし、語彙アクセス速度測定ツールは、その測定の結果が英語運用能力を反映しているべきである。つまり、英語運用能力が異なる受験者を識別できるツールであるべきだと考えられる。それは、語彙アクセスが語彙知識の一側面を担うと考えられるからであり、このことは、語彙知識のみならず、その量・質の影響を受ける一般的英語運用能力にも反映されるものであるという考えに基づいている。

さらに、語に対する何らかの作業は数ミリ秒から数秒というごく短時間に行われるものである。こういった作業の速度を測定する場合には、ツールの精度の高さが非常に重要である。しかしながら、精度を重視するあまり、広く一般に使用されないようなハードウェアやソフトウェアを利用することは、本研究におけるツール開発の意義に反すると考える。それは、アクセス速度が語彙知識の一面であるならば、それが実際の英語運用に影響を及ぼしているという観点から、アクセス速度を測定し、その結果を学習や教育にも応用してゆくべきであるという立場に基づいて、実用性を重視したツールを開発することを目的とした。そのため、本研究ではインターネット接続環境の整ったコンピュータ、インターネットブラウザ、入力デバイス(キーボードおよびマウス)を必要最小限の環境として、この範囲で可能な限り精度の高いツールを開発するものとした。

以上をふまえ、本研究は、以下の2つの目的を持って遂行された。

- (1) 語彙アクセス速度測定に最適な作業の特定
- (2) 一般的英語運用能力を識別できる精度の確立

#### 4.2.2 アクセス速度測定テスト WRECT の開発

上記の目的のために、Word Recognition Test (WRECT)を開発した。WRECT はマクロメディア社の Flash を用い、Action Script 言語によってプログラミングされた。Flash 技術を用いることにより、本研究には以下のような利点をもたらされた。

- (1) プログラム自体の容量が非常に小さいため、インターネット上での受験に最適である
- (2) 一般的なウェブ・ブラウザ（インターネット・エクスプローラーやサファリ）が導入済みであれば、WRECT 自体をインストールする作業が必要でない
- (3) テスト結果をインターネット経由で集中的に管理することができる
- (4) グラフィカルなインターフェースで、魅力的なテストを提供することができる。

テスト中に受験者が遂行するタスクは語形認識と語形および意味認識の2種類を用意した。さらに、語形認識タスクには、種類の異なる2種のタスクを作成し、テスト全体として、合計3つのタスクが用いられた。

タスク 1. 一つ目のタスクは、語形認識タスクであった。このタスクでは、語形を認識する速度を測定した。受験者には画面上に表示される 25 文字の文字列内に隠された英単語の数を申告することが課せられた。実際の指示文は、以下の通りであった。

このパートでは、25 文字のアルファベットの文字列内にいくつかの英単語が隠されているかを見つけてください。問題により、1 個から 4 個隠されています。

テストの前にルールを確認します。

- a) 3 文字以下の文字列は単語とはみなしません。
- b) 単語と他の単語の一部が重なることはありません。ぴったりとくっついていることはあるかもしれませんがありません。

回答は、マウスのクリックで行います。画面中央下よりに、青い丸があります。隠れている文字列を見つけるたびに、1回クリックしてください。カウンターの数値が1増えます。

もし、間違えてクリックしてしまった場合には、「-1」ボタンを押すと、数値が1減ります。全ての単語を見つけ終わったら、矢印のボタンを押してください。次の問題が始まります。



図 4.2 タスク1のサンプル画面

アクセス速度として測定される時間は、文字列の表示から一問毎の回答終了(矢印ボタン押下)までの時間とした。タスクは 5 問の練習問題および 30 問で構成された。受験者は、マウスを用い、単語を見つけるたびに、画面中央にある青いボタンを押すことが求められた。図 4.2 のサンプルでは、文字列中に **free** と **rule** の 2 語が隠されているので、この場合 2 個が正解である。

タスク 2. 二つ目のタスクは語形および意味へのアクセス速度を測定するタスクであった。このタスクでは、受験者は表示される 8 つの英単語の中から、他の 7 つの単語のトピックを表す上位語を選択することが求められた。このタスクもすべての操作をマウスで行うものとした。実際の指示文は以下の通りであった。

このパートでは、8 つの単語が表示され、あなたは仲間はずれの単語を探します。仲間はずれの単語は、他の 7 つの単語のトピックを表す単語になっています。つまり、ひとつの単語はカテゴリーを表し、他の 7 つの単語は、そのカテゴリーに属する単語となっています。

この仲間はずれの単語を見つけたら、それをクリックしてください。見えない部分に隠されているタイマーが止まります。次の問題に進むには、青い円をクリックしてください。この青い円をクリックすると同時に次の問題が始まります。

問題は全部で 35 問です。途中での休憩はありませんが、各問題に答え終わった後で青い円をクリックするタイミングは自由ですので、各自で休憩を取りながら進んでください。

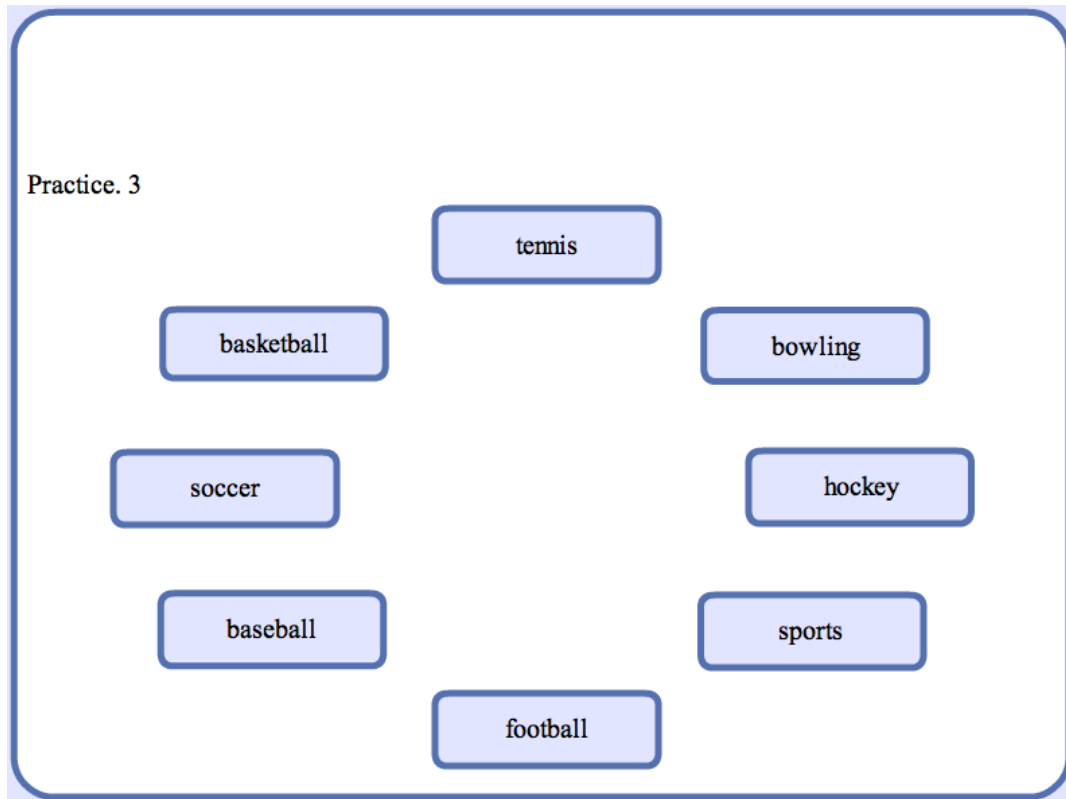


図 4.3 タスク2のサンプル画面

図 4.3 のサンプル画面では、tennis, bowling, hockey, sports, football, baseball, soccer, basketball の 8 つの単語のうち、上位語である sports が正解である。

このタスクでは、8 つの単語が表示されてから、正解であると思われる一つの単語をクリックするまでの時間を語彙アクセス時間として計測した。また、一つの問題が終了し、次の問題へと移行する際には、次の問題を表示するためのボタンのみが画面に表示され、それをクリックすることで受験者が問題表示のタイミングを自分で決定することができた。このボタンは、放射状の表示される 8 つの単語(図 4.3 を参照)の中央に位置するようにデザインした。これは、問題の開始から正解である単語までの距離をできる限り統一するための措置であり、30 問の問題でマウスの移動距離が大きく異ならないようにするための配慮であった。

タスク 3. 三つ目のタスクは、語形認識速度を測定するものであった。このタスクでは、受験者は、不完全な英単語を補完し実在する英単語とするために適切なアルファベットを 2 つの選択肢から選び、割り当てられたキーを入力することが求められた。実際に与えられた指示は以下の通りであった。

このパートでは、アルファベット 1 文字が欠けている単語が一語表示されます。単語の下に、アルファベットが 2 文字表示されますので、あなたは単語を完成させるのに適切なアルファベットがどちらであるかを出来るだけ速く答えます。

このパートでは、マウスをしません。キーボードの左キーと右キー、そして、スペースバーのみを利用します。選択肢であるアルファベットが左右に一語ずつ表示されますので、左側のアルファベットが答えだと思ったら、キーボードの左キーを、右側のアルファベットが答えだと思ったら、右キ

ーを出来るだけ速く押してください。キーを押した瞬間に画面が暗くなり、タイマーがストップします。次の問題に移るには、スペースバーを押してください。

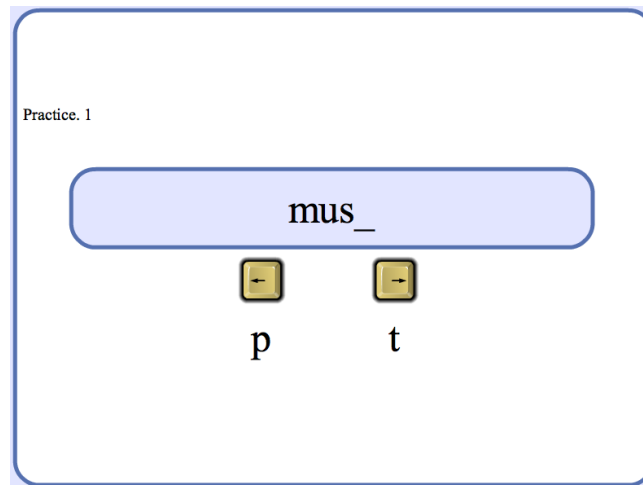


図 4.4 タスク3のサンプル画面

このタスクは、回答にキーボードのみを使用する。このタスクにおける語彙アクセス速度は、画面に文字列が表示されてから、キーボードの押下までの時間とした。

図 4.4 の例では、`mus_` の 4 文字目(アンダーバーの部分)を補完する適切なアルファベットは `p` であるか、`t` であるかを問う問題である。`musp` は実在しない英単語である一方、`must` は実在する英単語であるため、`t` (キーボードの右キー)が正解である。また、出題されるすべての不完全な語は、少なくとも 2 語以上のミニマルペアを持つ語であり(例:`must` – `musk`)、そのうちの一語のみが正解となるように工夫した。これは、習熟度の差による語彙アクセス速度の違いをより明確にするための工夫である。問題数は練習問題5問および測定対象の 30 問であった。

上記 3 つのタスクのうち、タスク 1 およびタスク 3 では、語彙アクセス速度の測定対象となるすべての目標語は、JACET8000 の 2,999 番以内の高頻度語に限定した。これは、語の頻度が語彙アクセスの速度に影響を及ぼすことが示唆されているためである(Wang & Koda, 2005)。しかしながら、タスク 2 においては、単語同士の意味概念の関係を重視し、頻度レベルの点で必ずしも統一がとれていない語も含まれていたが、すべての単語が可能な限り日常語であるようなトピック限定した。タスク 2 の練習問題5問を含む全 35 問の正解となる語は、以下の語である。

family, clothes, sports, instruments, time, city, movement, things, house, job, weather, color, speech, play, food, subjects, animal, number, kitchen, furniture, drink, travel, car, water, classroom, hospital, feeling, computer, country, body, music, landscape, insects, fruit, face.

全てのタスクの問題は、練習問題であるタスク開始直後の5問を除いては出題順がランダムに決定された。



さらにタスク 2 およびタスク 3 では選択肢の位置も問題ごとにランダムに配置された。

### 4.2.3 WRECT の妥当性の検証

#### 検証の目的

本研究で新規に開発された WRECT の妥当性は、英語習熟度の異なる受験者を識別できるかに焦点が当てられた。上にも述べた通り、語彙アクセス速度が語彙知識の一面を担っているのであれば、それは英語習熟度にも何らかの影響を与えていると考えられる。経験的にも英語母語話者がある単語を見て、その語の意味を想起するまでの時間は、非英語母語話者のその時間と比較した場合に短時間であることは正しいことのように思われる。そこで、語彙アクセス速度測定ツール WRECT の妥当性の検証は、英語母語話者の受験結果と日本人大学生の受験結果を比較することによって行われた。開発当初の目的と併せ、以下の研究課題を設定した。

- (1) WRECT は英語母語話者と日本人大学生の受験者とを識別することができるか。
- (2) WRECT の3つのタスクのうち、どのタスクが最も英語母語話者と日本母語話者とを識別するか。

#### 受験者

受験者は 55 名の日本人大学生と 11 名の英語母語話者であった。55 人の日本人大学生のうち、36 名は英語を専攻している学生であったのに対し、残りの 19 名は工学部に所属していた。また、11 名の英語母語話者は国内外の高等教育機関において英語教育に従事する教員であった。

#### 手順

日本人受験者は首都圏の 2 つの大学の各会場に設置されたコンピュータを用い、実施担当者からの口頭による調査実施の趣旨および受験の手順の説明を受けた後、インターネット上のテストサイトにアクセスし、WRECT を受験した。

母語話者の受験者は、電子メールによる受験依頼に記されたアドレスよりテストサイトにアクセスし、個人のコンピュータによって受験した。プログラムの性質上、コンピュータの性能による測定の誤差はほぼないものと仮定した。

#### 結果

表 4.4 は各タスクの基本統計量を示している。表中の総反応時間の算出にあたっては、各タスクのすべての問題で測定された語彙アクセス時間の総計から、回答が不正解であった場合の時間をのぞいて集計を行った。また、平均反応時間は総反応時間を正答数で割り、その数値を元に受験者グループ毎の平均値を算出した。なお、タスク 2 およびタスク 3 については、正解率が 80%未満の受験者は分析から除外した。

この表から明らかなように、タスク 1 は日本人受験者、英語母語話者ともに、正答率が 8 割を下回る結果となった(それぞれ 57.7%, 72.7%)。この結果は、この課題がアクセス速度を測定するには不適切であることを示している。このため、以降の分析からタスク 1 を除外することとした。

表 4.4 WRECT 各タスクの基本統計量

	グループ	人数	正答率		総反応時間		平均反応時間	
			%	SD	秒	SD	秒	SD
タスク	NNS	54	57.7	18.04	241.45	118.44	11.50	3.46
1	NS	11	72.7	15.20	254.41	155.75	9.58	4.99
タスク	NNS	37	89.6	5.82	169.94	38.45	6.34	1.48
2	NS	11	97.6	1.60	122.64	26.31	4.19	0.92
タスク	NNS	46	91.6	6.51	51.06	10.15	1.86	0.34
3	NS	11	98.8	2.25	50.25	13.32	1.70	0.45

タスク 2 およびタスク 3 の平均反応時間について、日本人受験者と英語母語話者の結果を比較した。 $t$  検定の結果、タスク 2 では日本人受験者と英語母語話者との間に有意差が認められた ( $t(26.93) = 5.85, p < .001$ )。しかしながら、タスク 3 では日本人受験者と英語母語話者との間に有意差は認められなかった ( $t(12.94) = 1.13, ns$ )。ここから、研究課題(1)について、WRECT は一部の課題(タスク 2)のみにおいて日本人受験者と英語母語話者とを識別できることが明らかになった。さらに、研究課題(2)に関しては、タスク 2 の意味判断課題が日本人受験者と英語母語話者を識別するのに最も適した課題であるという結果となった。しかしながら、後に考察で触れる通り、この結果には疑問が残る。

## 考察

本研究のために開発された語彙アクセス速度テストの結果にもとづいて、まず課題の難易度について考察をする。結果から明らかなどおり、タスク 1 については難易度の高さが目立った。25 文字の文字列内に 3 文字以上の英単語がいくつ隠されているかという課題は、開発者側の予想を遙かに上回る難しさであった。これは、開発の段階で、英語習熟度による語彙アクセス速度の違いをとらえるため、難易度を高く設定したことに起因する。つまり、タスク 1 は難易度を高くすることにより、習熟度の低い受験者は課題の遂行に長い時間を必要とするように意図して課題をデザインしていたが、本実験の結果、特に英語母語話者の正答率の低さが顕著に表す通り、語彙アクセス速度の測定をするために適切な難易度の課題でなかったことは明らかである。

また、タスク 3 については、結果を見る限り語彙アクセス速度には習熟度の違いが反映されなかった。この結果の解釈は二通りある。まず一つ目の解釈は、語彙アクセス速度は英語習熟度を反映しないという可能性である。これは語彙アクセス速度が語彙知識の一側面を担っており、語彙知識は英語習熟度とともに発達するという前提で調査を行ったが、この前提自体が誤りである可能性があることを示している。しかしながら、タスク 2 の結果を見る限り、日本人受験者と英語母語話者との間に有意差が認められたことから、この可能性は低いと思われる。

二つ目の解釈は、タスク 3 で与えられた課題の難易度が低かったということである。この課題では、ミニマルペアを持つ語を目標語とすることで、出題が `mus_` であれば、受験者が正解の可能性のある `must` と `musk` の両方にアクセスする必要性を仮定した。こうすることにより、語彙アクセス速度が速いと仮定される母語話者

はより速く複数の語にアクセスし、習熟度の低い受験者はより長いアクセス時間を必要とすると予想された。しかしながら、不完全な語を補完するアルファベットの選択肢が問題と同時に表示されるシステムであったため、開発者の意図とは裏腹に、一つの語に対するアクセス速度のみを測定していた可能性が否定できない。これにより課題の難易度が予想より低く、習熟度による語彙アクセス速度の差が検出できなかったものと思われる。

これらの点から、研究課題(2)である、語彙アクセス速度を測定するための最も適切である課題の種類についての結果も慎重に判断しなければならない。統計分析からは、タスク 2 で課された意味判断課題が最も適切な課題であるという結果が導かれた。しかしながら、上で述べたように、課題の難易度の設定のために、語形判断課題が機能していなかった可能性があり、この結果を持って意味判断課題が最も適切であるという結論は尚早である。

果たして、WRECT は語彙アクセス速度を測定するテストとして、妥当性があるとは言い難い結果であった。先行研究を元に、語彙性判断課題を廃止し、実在の英単語のみを対象として、また、特別なソフトやハードを用いないテスト開発を行ったが、結果的に多くの問題点が浮上した。第一に、課題そのものの難易度である。WRECT では 3 つの課題を設定したが、純粋に語彙アクセス速度を測定すること以上に、習熟度の差によるアクセス速度の違いに焦点が当たり、その結果が課題の難易度として反映されてしまった。今後は適切な課題の難易度を設定することが必要である。

第二の問題点は測定の精度に関する点である。WRECT のタスク 1 およびタスク 2 は、マウスを用いる課題であったが、アクセス速度の測定中にマウスを移動させることが必須であった。このため、純粋なアクセス速度に加え、マウスの移動の速度までもが測定されてしまっていた。この点に関しても、純粋な語彙アクセス速度の測定という本テストの根幹であるべき点に焦点が定まらなかった。また、タスク 3 については、キーボードを用いる課題であったが、回答が決まる時間(語彙アクセスの時間)に加え、左右どちらのキーを押すべきかの判断と実際にキーを押す時間が含まれていた。さらに、受験者からは、キーの押し間違い、つまり、意図していないキーを押してしまったという報告が多くなされた。

これらの問題点にからも、WRECT が語彙アクセス速度を測定するツールとして妥当なものでないことは明らかである。次節ではこれらの問題点や反省点を踏まえ、より妥当性のある語彙アクセス速度測定テストの開発の試みを紹介する。

## 4.3 アクセス速度測定テストの開発(2)

### 4.3.1 LEXATT 開発の目的

前節の WRECT の結果を踏まえ、その問題点・反省点を克服し、より妥当性のある語彙アクセス速度測定ツールである Lexical Access Time Test (LEXATT) の開発を試みた。LEXATT は WRECT の基本理念を踏襲し、さらに WRECT の問題点を克服するために、以下の点に焦点を当てて開発された。

- (1) 特殊なソフトやハードを用いずに、一般に普及しているレベルのコンピュータ上で測定できる
- (2) 測定の結果が英語習熟度の異なる受験者を識別できる
- (3) 複数レベルの語彙アクセス速度(語形のみおよび語形と意味へのアクセス)を測定できる

- (4) インターネット上に設置し、テスト自体のインストール作業をすることなく受験できる
- (5) テスト結果は自動で採点され、受験者にはテスト終了と同時に結果が表示される
- (6) 計測されるべき時間は、語彙アクセスに要する時間のみとし、実際に開発者側の意図する語を認識していたかどうかに対する回答時間は計測しない
- (7) したがって、回答のためにマウス等の移動が必須である場合にも、その移動時間は語彙アクセス速度には含まない。

WRECT の改訂版である LEXATT は以上の点を開発の目的とした。

### 4.3.2 LEXATT の開発

LEXATT には2つの課題が設定された。一つ目のタスクは、WRECT と同様に、語形を認識する速度を測定するタスクであった。二つ目のタスクは、語形及び意味の認識速度を測定するタスクであった。

#### タスク 1. 語形へのアクセスを測定するタスク

このタスクでは、10 文字の文字列の中に埋め込まれた英単語を発見し、その単語の 10 文字中の位置を答える課題である。目標語には JACET8000 のレベル1~3(3,000 語レベル以内)から無作為に抽出された 4 文字の英単語 40 個であった。この頻度による制限は、日本人大学生の英語学習者にとって未知語ではないことを念頭においた上で、頻度による語彙アクセス速度の変化(頻度効果)を統制するためのものである。

また、目標語はすべて 4 文字としたが、これは単語長が語形レベルでのアクセス時間に影響を与えると思われるため、単語長を統制し、かつ課題の難易度が高くなりすぎないようにするためのものである。

タスクの遂行にはマウスのみを使用するものとした。その際、マウスの移動時間を計測せず、さらにマウスボタンを押す際に必要とする指の力の影響を最小限にとどめるために、語彙アクセス速度の測定は、「マウスのボタンを完全に押してから離すまで」と設定された。マウスのボタンを完全に押すことを確実な条件とするために、各問題は、マウスのボタンが押されている間のみ表示され(図 4.5 参照)、目標語を見つけた時にマウスのボタンを放すと文字列は非表示となった。

このようにして計測された時間が実際に目標語を探すために費やされた時間であることを条件づけるために、10 文字の文字列は画面から見えなくなるが、その単語の一文字目がどこにあったかをクリックすることが求められた。この時間は記録として測定されず、アクセス速度の計測には含まれないものとした(図 4.6 参照)。

また、10 文字中の単語表示位置による影響を最小限にするため、各目標語は 1 文字目・3 文字目・5 文字目・7 文字目のいずれかに表示された。これら四つの表示位置は各 10 問ずつ用意された。

実際に受験者に示された指示文は以下の通りであった。

このパートでは、英単語を見つける速さを測定します。

10 文字のアルファベットの文字列の中に、4 文字の英単語が 1 つ隠れています。出来るだけ速くその単語を見つけてください。

解答手順は以下の通りです。

- (1) マウスの左ボタンで文字列表示ボタンを押すと、10 個のアルファベットが表示されます。ボタンを押している間はずっと表示されています。この間に 4 文字の英単語を 1 つつけてください。この間、文字列表示ボタンはピンク色になります。
- (2) 4 文字の英単語が見つかったらすぐにマウスの左ボタンを放してください。ボタンを押している間の時間が「語彙認知時間」として計測されます。
- (3) 次に、見つけた単語の一文字目がどこにあったかを、文字列表示枠(青い円)のうちの 1 つをクリックすることでさし示してください。

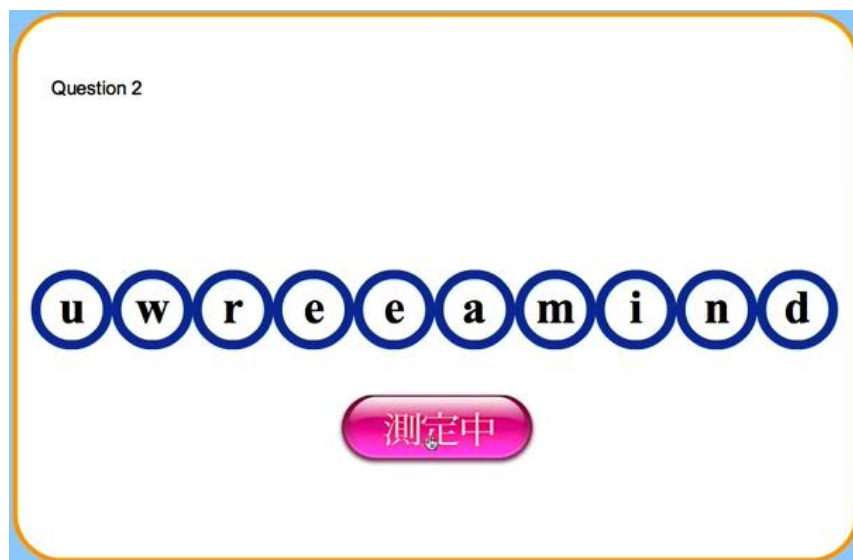


図 4.5 LEXATT タスク1(アクセス速度測定中)のサンプル画面



図 4.6 LEXATT タスク2(回答中)のサンプル画面

## タスク 2. 語形および意味へのアクセス

このタスクでは、受験者は表示される英単語を認識し、その語の意味を二つの選択肢から選ぶことが求められた。提示される目標語は、タスク 1 と同じく、JACET8000 のレベル1～3から抽出された。採用の条件は、4 文字の語であり、少なくとも一つ以上のミニマルペアがあることであった。これは、語の意味を選ぶ段階で、選択肢が正解である語の意味とミニマルペアである語の意味を提示し、正解である語を間違い無く認識していたことを確実にするためのものであった。

タスク 2 もタスク 1 と同様に、マウスのみ使って課題に取り組む仕組みとした。画面中央のボタン上で、マウスのボタンを押している間だけ問題が表示され(図 4.7 参照)、マウスのボタンを放した瞬間に問題が非表示となるが、この間の時間を語彙アクセスの時間とした。その後、画面には二つの日本語が表示され、受験者は直前に示された英単語の意味を選択した(図 4.8 参照)。この間の時間は測定されず、語彙アクセス時間には含まれなかった。

問題の表示時には、単語は左側から徐々に 1 秒かけて表示された。このことにより、単語の全容が明らかでない時点でも回答することができるようになるため、習熟度の高い受験者には有利であると思われた。

受験者に与えられた実際の指示は以下の通りであった。

このパートでは、単語とその意味を認知する速度を測定します。

解答手順は以下の通りです。

- (1) マウスの左ボタンで単語表示ボタンを押すと、オレンジの楕円の中に英単語がだんだんと見えてきます。その単語の意味がわかった時点で、マウスの左ボタンから手を離してください。ボタンを押してから、離すまでの時間を計測しています。
- (2) ボタンから手を離すと、意味表示枠に日本語が表示されます。表示された単語の意味である方の解答ボタンをクリックしてください。

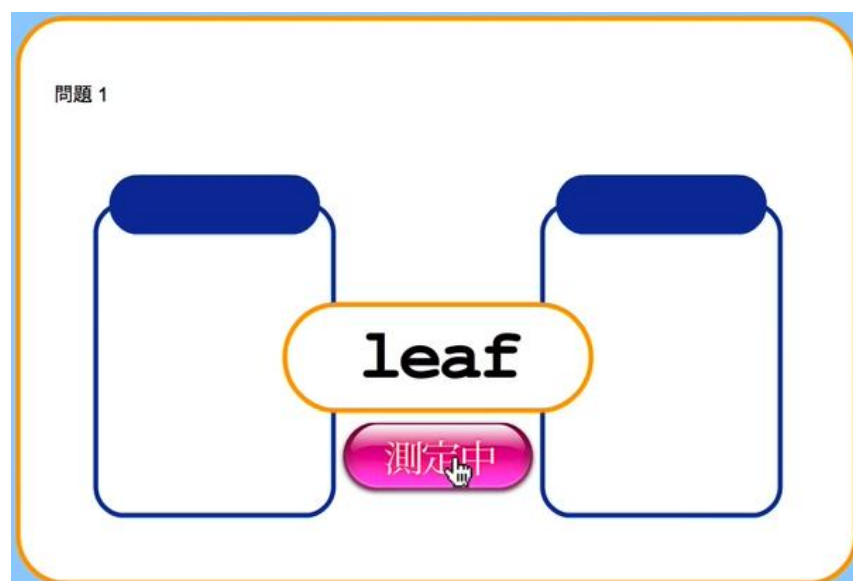


図 4.7 LEXATT タスク2(アクセス速度測定中)のサンプル画面

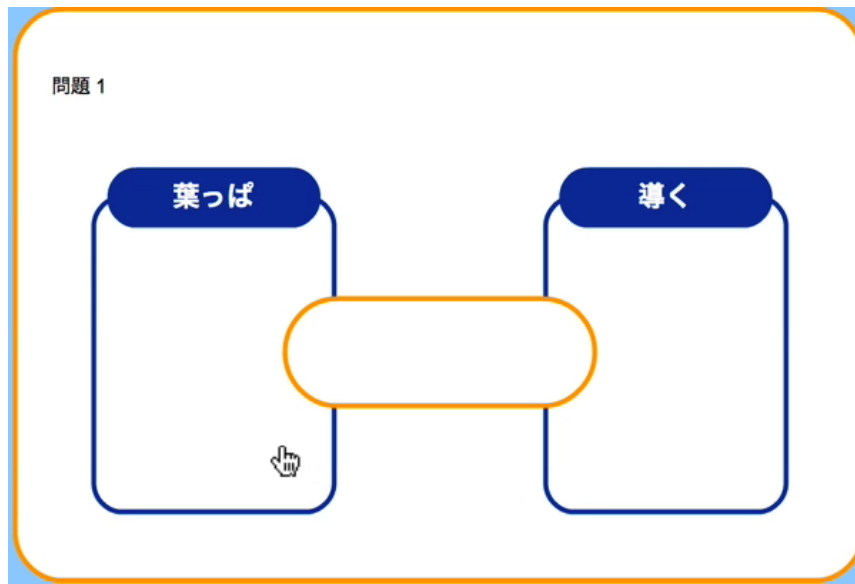


図 4.8 LEXATT タスク2(回答中)のサンプル画面

両タスクとも、実際の問題に回答する前段階として、5問の練習問題が準備された。受験者には練習問題を解いた直後に、5問中の正答率や平均語彙アクセス速度が表示された。

#### 4.3.3 LEXATT の妥当性の検証

WRECTと同様に、LEXATTの妥当性検証は、習熟度の異なる英語学習者を識別できるかという観点から行われた。しかしながら、タスク2で語彙の意味へのアクセスを計測するために、本テストの妥当性検証では、日本人英語教師とTOEICによる習熟度別の大学生グループ間との比較を行った。具体的な研究課題は以下の通りである。

- (1) LEXATTは習熟度の異なる英語学習者を識別できるか
- (2) 語形へのアクセスと意味へのアクセスでは、どちらがより適切な語彙アクセス速度測定の課題であるか

#### 受験者

受験者は255名の日本人英語学習者であった。うち15名は英語学または英語教育関連の修士号以上を持つ日本人英語教師であり、他の239名は日本人大学生であった。これら大学生は全員TOEICを受験しており、表4.5が示すように習熟度の差は広範囲であった。また、日本人英語教師は大学生の能力を超えるグループであると仮定した。そのため、修士号以上を持つ教員に限定した。

#### 分析方法

大学生の受験者はTOEICの成績に基づいて4つのグループに分けられた。それらのグループは、上位群、中上位群、中下位群、下位群であり、各グループのTOEIC平均点は上位群から順に718.45点、475.14点、

347.94 点, 257.45 点であった。また各グループの人数は 29 名, 69 名, 90 名, 51 名であった。これらのグループに教員グループをあわせ, 5 つのグループ間でタスク 1, タスク 2 それぞれの結果を比較した。

表 4.5 大学生受験者の習熟度 (TOEIC)

平均	標準偏差	最小	最大
410.29	147.05	185	895

## 結果

各グループのタスク 1 およびタスク 2 の結果 (平均値) は表 4.6 の通りである。タスク 1 では, 中下位群と下位群間を除き, 習熟度が上がるに従って, 平均語彙アクセス速度の値が小さくなっていることが分かる。しかしながら, タスク 2 においては, おおむね習熟度と語彙アクセス速度の平均値との間にタスク 1 と同様の関係がみられるものの, 教員と上位群および中下位群と下位群との間に逆転が起きている。

分散分析の結果, タスク 1 は  $F(4, 249) = 3.19$  ( $p < .05$ ), タスク 2 は  $F(4, 249) = 4.15$  ( $p < .05$ ) と, 両タスクともに有意であった。LSD 多重比較の結果, タスク 1 においては, 教員グループは中上位群, 中下位群, 下位群よりも有意に語彙アクセス速度が速く, また, 上位群グループと中下位群との間にも有意な差が認められた。

タスク 2 においては, 教員グループは中下位群との間に, また上位群は中下位群および下位群との間に有意差がみられた。さらに中上位群は中下位群よりも有意に速く語彙アクセスが行われていることが明らかになった。

これらの結果を研究課題と照らし合わせると, LEXATT の結果はおおむね習熟度の異なる英語学習者を識別できていた。しかしながら, 隣り合う習熟度グループ間には有意差が認められることは希であった。

また, タスク 1 の語形アクセス速度課題はタスク 2 の意味アクセス課題よりも習熟度の識別精度は高いという結果であった。

表 4.6 LEXATT タスクごとの平均値 (ミリ秒)

グループ	人数	タスク 1		タスク 2	
		平均	SD	平均	SD
教員	15	3473.33	807.03	523.53	135.19
上位群	29	3902.66	995.97	497.14	220.74
中上位群	69	4351.30	1449.51	598.38	292.99
中下位群	90	4645.70	1577.26	731.37	380.56
下位群	51	4438.10	1334.09	671.65	313.93



## 考察

LEXATT の結果は、おおむね習熟度の異なる受験者を識別できたが、すべてのグループ間に有意差が認められることは皆無であった。ここで、このような結果に至った原因について考察を加える。

まず、両タスクにおいて、最も習熟度が低いグループが次に習熟度が低いグループよりもよい結果を出していることについてであるが、これには2つの原因が考えられる。まず一つに、低習熟度グループの正答率の低さが上げられる。テストの性質上、不正解であった問題に費やした時間は語彙アクセス速度の集計に含まれなかったが、このため、他の問題よりも難しく、より時間がかかる問題があったとすれば、最も習熟度が低いグループはこれらの問題に不正解であったことが考えられる。これが事実であれば、低習熟度グループはより速く容易に答えることができる問題に対してのみその速度が計測されているということであり、平均的に語彙アクセス速度が速いという結果が出ても不思議ではない。

もう一つは、LEXATT が語彙アクセス速度を測定するための閾值的な習熟度が存在している可能性である。一つ目の可能性と関連する事柄でもあるが、目標語が受験者にとって未知語である可能性等が否定できないため、このような状況下にある習熟度の低い受験者の語彙アクセス速度を正確に測定できていない可能性は否定できない。

また、タスク 2 において、教員グループは上位群よりも語彙アクセス速度が遅いという結果になっているが、この点については、このタスクが年齢等による身体的な衰えが影響している可能性がある。また、目標語が徐々に表示される問題提示形式であったため、語の全体像が提示される前に答える受験者が少数であった場合には、習熟度による差が表面化しづらいことも否定できない。いずれの可能性においても、本調査の結果からは明らかにすることはできないが、この点は同様のテストを開発する際には考慮に入れるべき点である。

さらに、本調査ではグループ間の語彙アクセス速度の差に焦点を当てたが、グループ間の比較では測ることのできなかった個人差があったことも結果に大きな影響を与えていた可能性がある。具体的には、タスク 1 においては練習問題が 5 問あったにもかかわらず、最初の一問に約 26 秒かかった受験者が存在していた。他の問題には数秒で回答していることから、課題の趣旨が十分に理解できないままに課題に取り組んだ可能性があることを示唆している。

またタスク 2 においては、99 ミリ秒問題が提示された場合でも正解となる語を選択しているケースも見られた。これは明らかに、マウスの操作ミスであるにも関わらず、二者択一の問題形式であるがために正解してしまった例であろう。

## 4.4 結論

本研究は、語彙知識の一側面を担うと考えられているアクセス速度を測定するためのテストを開発することであった。一番目に開発された WRECT の問題点を克服し、改訂版として LEXATT の開発に至ったが、習熟度の異なる学習者を大まかに識別することはできたが、精度の面で問題が残っていると言わざるを得ない。また、個人差の中でも問題間で、アクセス速度に大きな差が見られた点も結果に影響を与えていたと考えられる。

これらの点を踏まえ、さらなる改訂版の開発が必要である。具体的に考慮すべき項目は以下の通りである。

- (1) 正答率が一定基準を下回った段階でテストを打ち切り、それらの受験者データを妥当性検証から除外する
- (2) 操作ミスや問題趣旨の不理解等を考慮し、最も速い速度 5 問と最も遅い速度 5 問をそれぞれ全員のデータから除外する
- (3) タスク 2 の問題提示方式を徐々に提示する方法から、全体を一度に提示する方法へ変更する
- (4) タスク 2 の選択肢の数を増やし、当て推量による正解を防ぐ
- (5) 母語の語彙アクセス速度を測定する課題を追加し、受験者個人のベースラインを調査することで、年齢による影響を排除する。

今後、上記5点に関する検討を重ね、測定の精度を高めることが必要である。

## 第5章 3種類のテストと英語習熟度とスキルの関係

### 5.1 英語習熟度全体の予測

#### 5.1.1 TOEIC の予測

##### 5.1.1.1 研究の目的

本節では、3種類の語彙知識テストと TOEIC リスニング・テストと TOEIC リーディング・テストの関係を吟味する。リサーチ・クエスチョン(RQs)は以下の2つであり、重回帰分析と共分散構造分析を用いて調べた。

RQ1: 3つの語彙テストは、TOEIC リスニング・リーディング得点をどの程度予測できるか。

RQ2: どの語彙テストが、TOEIC リスニング・リーディング得点をより予測できるか。

##### 5.1.1.2 方法

###### 受験者

日本の4つの大学から研究に参加した大学生総計416名のうち、3つの語彙テストをすべて受験し、かつ、TOEIC得点を申告した176名を分析対象とした。これは2.3節と3.3節での受験者とほぼ重なる。大学の内訳は、私立大学理系83名、私立大学文系35名、国立大学文系または理系27名、公立大学文系31名である。176名のうち1名が多変量正規性に関して極端な外れ値を示したため、175名で分析した。その1名は、他のテストは平均的である(TOEIC 総合得点400点、TOEIC リスニング得点295点、TOEIC リーディング得点105点、J8VST 72点、LOT 17点、LEXATT タスク2 457ミリ秒)が、LEXATT タスク1においては非常に小さく速い値(667ミリ秒)を示した私立大学文系学生だった。

###### 使用テストと手順

語彙サイズテストであるJ8VST、語彙構成テストであるLOT、語彙アクセス速度テストであるLEXATT タスク1とタスク2をウェブ上で受験した(詳しいテスト方法やその妥当性については第2, 3, 4章を参照)。語彙テストはLEXATT タスク1とタスク2、J8VST、LOTの順で実施した。TOEIC リーディング・リスニング得点は、受験者が自己申告した値を用いた。LEXATT タスク1とタスク2はともに反応速度を分析した。

###### 分析

語彙テスト得点とTOEIC得点(TOEIC 合計得点・リスニング・リーディング得点)の変数を用い、相関分析・重回帰分析・共分散構造分析(structural equation modeling: SEM)を行った。各得点との関係を他の要因を考慮せずに行う分析手法である相関分析では、ピアソンの積率相関係数を用いた。

各語彙テストによる予測率が相対的にどのように異なるかを調べるためには、重回帰分析とSEMを使用した。重回帰分析は、複数の独立変数(4種類の語彙テスト得点)から1つの従属変数(TOEIC 総合得点・リスニング・リーディング得点の中から1つ選択)を何パーセント説明できるか、複数ある独立変数の中でどれが一番従属変数を説明できるかを調べることができる。SEMが重回帰分析と異なる主な点は、(1)SEMが一般に重回帰分析と因子分析などを組み合わせた分析が可能で、観測された変数だけでなく、因子のような潜在的な

変数をモデルに入れ、誤差を取り除いた形で変数間の関係が吟味できること(狩野 & 三浦, 2002 など), (2)SEM では、適合度指標を見ることでデータがどの程度モデルと一致するかを吟味できる点, (3)SEM では、複数の独立変数(4 種類の語彙テスト得点)から複数の従属変数(TOEIC の合計及びセクション毎の得点)の予測を、1 回の分析で行えることなどである(豊田, 2007 などを参照)。

重回帰分析では変数増加法を用いた。独立変数によってどのくらいの割合の従属変数が説明できるかを調べる際には、 $R^2$  (決定係数)と修正済み  $R^2$  を使い、Cohen(1988, pp. 413-414)の基準で修正済み  $R^2$  を解釈した。具体的には .0196 未満ならばほとんど予測できず, .0196 以上 .13 未満ならば少し予測でき, .13 以上 .26 ならば予測が中程度でき, .26 以上ならば予測が多くできると解釈した。SEM では最尤法 (maximum likelihood method)を用いた。相関・重回帰分析は SPSS (2003)で、SEM は Amos™ 7.0 を用いた。

### 5.1.1.3 結果と考察

#### 記述統計

表 5.1 は、各変数の記述統計を示している。LEXATT タスク 1 得点以外は正規分布に近かったが、LEXATT タスク 1 得点は尖度が高かった。LEXATT タスク 1 得点と LEXATT タスク 2 得点について平方根変換・対数変換・逆数変換を試したが、最も正規分布に近くなる平方根変換の結果を表 5.1 に挙げた。

各変数の信頼性で計算可能なものは表 5.1 に挙げた。ほとんどの信頼性は高かったが、LOT では低かった ( $\alpha = .31$ )。これは、LOT 得点が今回の大学生ではあまりばらつきが見られなかったことが 1 つの理由であろう。5.1.2 節の大学生のみの場合 ( $n = 39$ )ではクロンバックのアルファ係数が  $\alpha = .56$ , 5.3 節の大学生・大学院生をまとめて扱った場合 ( $n = 100$ )では  $\alpha = .77$ , 3.2.2 節の大学生と英語教師をまとめて扱った場合 ( $n = 8$ )の分析では  $\alpha = .84$  と能力幅が広がるにつれて  $\alpha$  が高くなったことを考慮すると、大学生だけではばらつきが少ないと考えられる。これは、後の重回帰分析で、LOT 得点から TOEIC 得点の予測の度合いが J8VST より低い 1 つの理由と考えられる。LOT 得点の信頼性が低いため、今回の SEM では LOT を除き分析を行った。

なお、本節の研究における受験者の推定語彙サイズ(素点/125\*500 で推定)は 2440~4840 語(平均値 3776.23 語, 標準偏差 517.38 語)であり、この程度の語彙サイズの受験者の結果だということを念頭に置く必要がある。また、TOEIC 総合得点は 195~895 点(平均値 433.91 点, 標準偏差 155.92 点)でかなり幅広い受験者であったこともわかる。本節ではリスニング・リーディング得点の予測が焦点であるが、可能な範囲で TOEIC 総合得点の結果も参考までに挙げた。

表 5.1 変数の記述統計( $n = 175$ )

	平均	標準偏差	歪度	尖度	最小値	最大値	信頼性
TOEIC 総合得点	433.91	155.92	0.99	0.38	195	895	.90 <sup>a</sup>
TOEIC リスニング	247.51	79.14	0.72	0.20	100	495	--
TOEIC リーディング	186.40	84.09	1.13	0.59	60	430	--
J8VST ( $k = 125$ )	94.41	12.93	-0.45	-0.14	61	121	.90
J8VST (奇数項目 $k = 63$ )	46.93	6.53	-0.39	-0.04	30	61	.81
J8VST (偶数項目 $k = 62$ )	47.48	6.99	-0.55	-0.04	28	61	.84
LOT ( $k = 50$ )	21.12	3.70	-0.02	-0.61	12	29	.31
LEXATT タスク 1 ( $k = 40$ )	4477.15	1418.53	1.09	2.13	1891	10973	.74 <sup>b</sup>
LEXATT タスク 2 ( $k = 40$ )	621.33	338.25	1.20	0.79	194	1780	--
LEXATT タスク 1 変換後	66.13	10.21	0.58	0.58	43.49	104.75	--
LEXATT タスク 2 変換後	24.12	6.31	0.79	-0.24	13.93	42.19	--

注:  $k$  = 項目数. <sup>a</sup>TOEIC 総合得点の信頼性について、今回の受験者における信頼性は素データがないため分析できないが、総合得点はリスニング得点とリーディング得点を総合して総合得点を出していることを考え、リスニング得点とリーディング得点の一貫性という意味での信頼性をクロンバックの  $\alpha$  係数で算出した。

<sup>b</sup>LEXATT タスク 1 の問題は、文字表示枠の 1 枠目、3 枠目、5 枠目、7 枠目から始まる単語で正解が構成されている。各枠から始まる単語において正解した場合の認知処理時間が算出されているため、単語が 1 枠目、3 枠目、5 枠目、7 枠目で始まる認知処理時間が一貫しているかという観点での信頼性を、クロンバックの  $\alpha$  係数で算出した。

## 相関分析

語彙テスト得点と TOEIC リーディング・リスニング得点の相関で最も高かったのは、TOEIC リーディング得点と J8VST で .67 であった(表 5.2 参照)。最も相関が低かったのは TOEIC リーディング得点と LEXATT タスク 1 得点の -.19 だった。語彙テスト間の相関を見ると、サイズ・構成・アクセス速度間では高くても中程度 .56 (J8VST と LOT 間)であり、後の重回帰分析や SEM において多重共線性(multicollinearity)が問題になる程高いものはなかった。

表 5.2 変数間の相関

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)TOEIC	.95	.96	.68	.65	.65	.52	-.21	-.24	-.21	-.23
(2)TOEIC_L	--	.82	.63	.59	.61	.46	-.20	-.23	-.20	-.22
(3)TOEIC_R		--	.67	.65	.63	.54	<u>-.19</u>	-.22	-.20	-.21
(4)J8VST			--	.95	.96	.56	-.24	-.27	-.24	-.27
(5)J8VST 奇数				--	.83	.51	-.20	-.25	-.20	-.25
(6)J8VST 偶数					--	.57	-.26	-.27	-.26	-.27
(7)LOT						--	-.25	-.20	-.25	-.20
(8)LEXATT1							--	.43	.99	.41
(9)LEXATT2								--	.42	.99
(10)LEXATT1 変換									--	.41
(11)LEXATT2 変換										--

注: 下線のみ  $p < .05$ . それ以外は  $**p < .01$ .

### 重回帰分析

重回帰分析の結果を表 5.3 と表 5.4 に挙げる。サイズが語彙知識のより基本的な知識であることを考え、J8VST を最初に、次に LOT, LEXATT の順に重回帰式に入れて分析した。その結果  $R^2$  を見ると J8VST だけで TOEIC リスニング得点の 39%, リーディング得点の 45% を説明していた。LOT を入れると説明率は 2%, 4% 高くなった。LEXATT は予測にほとんど貢献せず、全体で 41%, 48% の分散が 3 つの語彙テストで説明された。Cohen (1988) の基準を用いると、J8VST による予測は多くできており、LOT による予測は小さな程度できていた。

表 5.4 を見ると、標準偏回帰係数(B)では、J8VST の係数が最も高いのは、リスニング得点とリーディング得点両方において共通していたが、LOT の寄与は、ややリーディングにおける値の方が高い傾向があった。

表 5.3 変数増加法で、独立変数を順に追加したときの決定係数

	TOEIC 全体			リスニング			リーディング		
	$R^2$	修正	$\Delta R^2$	$R^2$	修正	$\Delta R^2$	$R^2$	修正	$\Delta R^2$
	$R^2$			$R^2$			$R^2$		
J8VST	.46**	.46	--	.39**	.39	--	.45**	.44	--
J8VST+LOT	.49**	.48	.03	.41**	.40	.02	.48**	.48	.04
J8VST+LOT+LEXATT タスク 1+2	.49**	.48	.00	.41**	.40	.00	.48**	.47	.00

注: \*\* $p < .01$ .

表 5.4 4 つの変数を入れたときの偏回帰係数

TOEIC 総合得点	$B$	$SEB$	$\beta$	下限 <sup>a</sup>	上限 <sup>a</sup>
定数(Constant)	-358.46**	81.69	--	-519.72	-197.20
J8VST	6.67**	0.81	0.55	5.06	8.28
LOT	8.42**	2.81	0.20	2.86	13.98
LEXATT タスク 1	0.00	0.01	-0.01	-0.01	0.01
LEXATT タスク 2	-0.02	0.03	-0.04	-0.08	0.04

TOEIC リスニング	$B$	$SEB$	$\beta$	下限 <sup>a</sup>	上限 <sup>a</sup>
定数(Constant)	-111.68*	44.52	--	-199.57	-23.79
J8VST	3.22**	0.44	0.53	2.34	4.09
LOT	3.17*	1.53	0.15	0.14	6.20
LEXATT タスク 1	0.00	0.00	-0.02	-0.01	0.01
LEXATT タスク 2	-0.01	0.02	-0.05	-0.04	0.02

TOEIC リーディング	$B$	$SEB$	$\beta$	下限 <sup>a</sup>	上限 <sup>a</sup>
定数(Constant)	-246.78**	44.35	--	-334.32	-159.24
J8VST	3.45**	0.44	0.53	2.58	4.33
LOT	5.25**	1.53	0.23	2.23	8.27
LEXATT タスク 1	0.00	0.00	0.00	-0.01	0.01
LEXATT タスク 2	-0.01	0.02	-0.03	-0.04	0.02

注:<sup>a</sup> $B$ の95%信頼区間. \* $p < .05$ . \*\* $p < .01$ .

表 5.4 の結果を基に作成したのが表 5.5 の回帰式である。各テストの得点を式に代入すれば、TOEIC の予想得点が出る。しかし、重回帰分析での説明率がリスニング得点で41%、リーディング得点で48%と半分以下のため、信頼区間の幅は広く、あまり精密な点数の予測には役立たないことも示された。具体的には、

TOEIC の総合得点は 10～990 点の間の 5 点刻みで表示される(Educational Testing Service, 2009)が、平均的な受験者の場合の予測は 95%信頼区間で -93.58～953.99 点であり、受験者の語彙テストから予測される TOEIC 総合得点は、95%の確率で受験者の真の得点が 10 点から 954 点の間に入るとは言えるが、それ以上のことは言えないことが示された。同様に、TOEICリスニング・TOEICリーディングともに 5～495 点の 5 点刻みで表示されるが、本節での平均的な受験者の場合の予測はリスニングで-39.64～531.30 点、リーディングで-99.97～468.71 点と幅広いものであった。

表 5.5 語彙テスト得点から TOEIC 得点を予測する回帰式

<p>&lt;TOEIC 総合得点&gt;</p> <p>■点推定:</p> $-358.46 + 6.67 \times (\text{J8VST 得点}) + 8.42 \times (\text{LOT 得点}) + 0.00 \times (\text{LEXATT タスク 1 得点}) + (-0.02) \times (\text{LEXATT タスク 2 得点})$ <p>■95%信頼区間下限:</p> $-519.72 + 5.06 \times (\text{J8VST 得点}) + 2.86 \times (\text{LOT 得点}) + (-0.01) \times (\text{LEXATT タスク 1 得点}) + (-0.08) \times (\text{LEXATT タスク 2 得点})$ <p>■95%信頼区間上限</p> $-197.20 + 8.28 \times (\text{J8VST 得点}) + 13.98 \times (\text{LOT 得点}) + (0.01) \times (\text{LEXATT タスク 1 得点}) + (0.04) \times (\text{LEXATT タスク 2 得点})$ <p>■予測値(表 5.1.にある平均を四捨五入した値を重回帰式に代入したときに得られた予測値)</p> <p>430.20 (-93.58 to 953.99)</p>
<p>&lt;TOEIC リスニング得点&gt;</p> <p>■点推定:</p> $-111.68 + 3.22 \times (\text{J8VST 得点}) + 3.17 \times (\text{LOT 得点}) + 0.00 \times (\text{LEXATT タスク 1 得点}) + (-0.01) \times (\text{LEXATT タスク 2 得点})$ <p>■95%信頼区間下限:</p> $-199.57 + 2.34 \times (\text{J8VST 得点}) + 0.14 \times (\text{LOT 得点}) + (-0.01) \times (\text{LEXATT タスク 1 得点}) + (-0.04) \times (\text{LEXATT タスク 2 得点})$ <p>■95%信頼区間上限</p> $-23.79 + 4.09 \times (\text{J8VST 得点}) + 6.20 \times (\text{LOT 得点}) + (0.01) \times (\text{LEXATT タスク 1 得点}) + (0.02) \times (\text{LEXATT タスク 2 得点})$



■予測値 245.83 (-39.64 to 531.30)

<TOEIC リーディング得点>

■点推定:

$$-246.78 + 3.45 \times (\text{J8VST 得点}) + 5.25 \times (\text{LOT 得点}) \\ + 0.00 \times (\text{LEXATT タスク 1 得点}) + (-0.01) \times (\text{LEXATT タスク 2 得点})$$

■95%信頼区間下限:

$$-334.32 + 2.58 \times (\text{J8VST 得点}) + 2.23 \times (\text{LOT 得点}) \\ + (-0.01) \times (\text{LEXATT タスク 1 得点}) + (-0.04) \times (\text{LEXATT タスク 2 得点})$$

■95%信頼区間上限

$$-159.24 + 4.33 \times (\text{J8VST 得点}) + 8.27 \times (\text{LOT 得点}) + (0.01) \times (\text{LEXATT タスク 1 得点}) + \\ (0.02) \times (\text{LEXATT タスク 2 得点})$$

■予測値 184.37 (-99.97 to 468.71)

### 共分散構造分析(SEM)

ここでは、各テスト得点を観測変数(図 5.1 の四角)とした。潜在変数(図 5.1 の楕円)は 2 個あり、第 1 に、「J8VST の偶数番号での得点」と「J8VST の奇数番号での得点」の背後に「サイズ」を仮定し、第 2 に、「LEXATT タスク 1 得点」と「LEXATT タスク 2 得点」の背後に「アクセス速度」を仮定しモデル化した。上記で述べたように LOT の信頼性は低いためにモデルに入れなかった。

まず、SEM を行う前に、Kunnan (1998)に基づき 2 つの前提を確認した。1 つ目の「変数ごとの正規性」については、「LEXATT タスク 1 得点」以外は、歪度・尖度ともすべて  $|\pm 2|$  内だったため満たしていた。「LEXATT タスク 1 得点」については、平方根変換を行った値が正規性を満たしたため、その値を用いた。2 つ目の前提である「変数を組み合わせた時の正規性(多変量正規性)」については、多変量正規性の値が 1.66 で、1.96 を超えていず、5%水準で有意でなかったため、正規性は満たされたと考えた。モデル 1(図 5.2 参照)の分析後、「アクセス速度」から「TOEIC リスニング得点」と「TOEIC リーディング得点」へのパス係数が有意でなかったため、そのパスを除外したモデル 2 も分析した(図 5.3 参照)。次に、モデル 1 の適合度とモデル 2 の適合度は有意に異なるかを調べるために、表 5.6 の値を用いて  $\chi^2$  difference test を以下のように行った(Werner & Schermelleh-Engel, 2009 参照)。

$$\chi^2_{\text{diff}} = \chi^2_{\text{model}_2} - \chi^2_{\text{model}_1} = 5.87 - 5.04 = 0.83$$

$$df_{\text{diff}} = df_{\text{model}_2} - df_{\text{model}_1} = 7 - 5 = 2$$

$df_{\text{diff}} = 2$  のときには 5%水準で  $\chi^2_{\text{diff}} = 5.99$  以上ならば有意となるが、それを下回ったため、モデル 1 とモデ

モデル2の適合度は異なるとは言えないという結果になった。表5.6の他の適合度指標を見ると、モデル2はモデル1よりやや適合度がよく、「RMSEAの90%信頼区間の上限」以外の基準が満たされていた。さらにモデルがより単純でもあるため、モデル2を最終モデルとして採択し、今回のデータはモデル2に適合したと考えた。

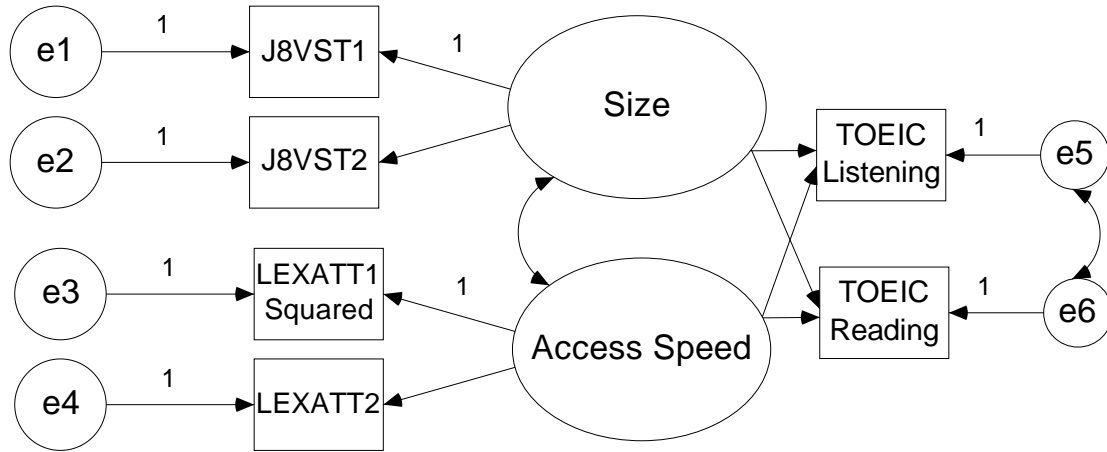


図5.1 モデル1の分析前のパス図

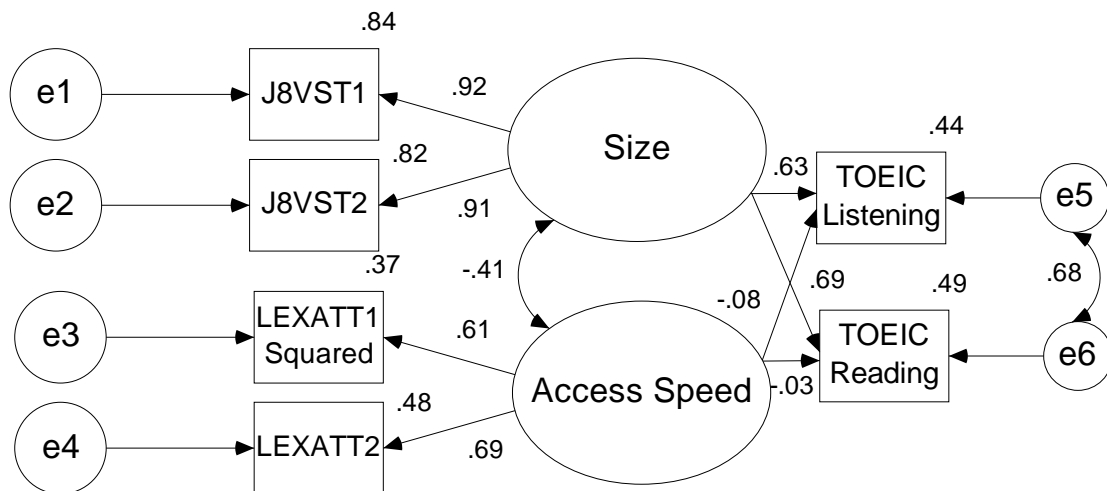


図5.2 モデル1の分析後のパス図

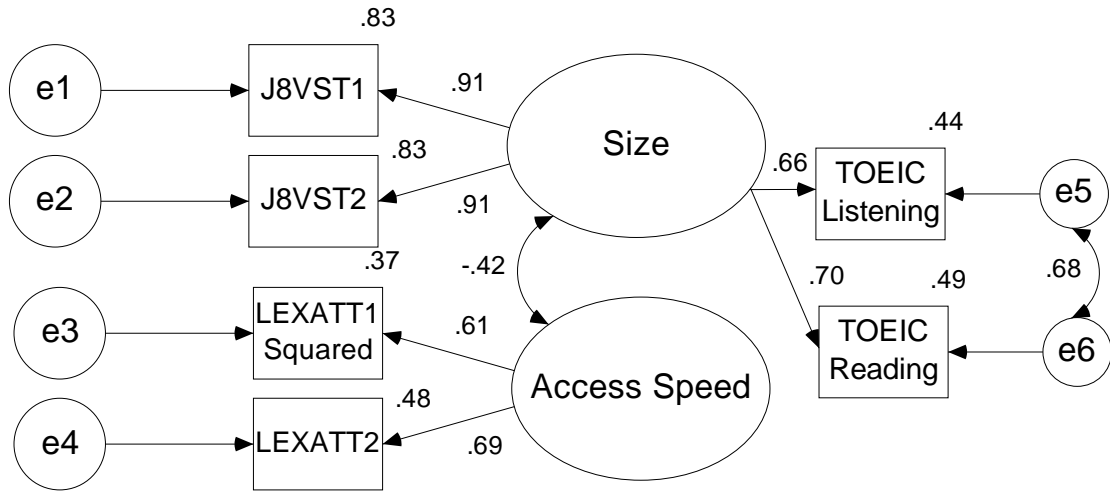


図5.3 モデル2:最終モデル

表 5.6 モデルの適合度の結果

	$\chi^2(df), p$	$\chi^2/df$	GFI	CFI	RMSEA (90% CI)	AIC	SRMR
基準	$p > .05$	$< 2.0$	$> .90$	$> .90$	$< .05$	より小さければ より良い	$< .08$
モデル 1	5.04 (5), .41	1.01	.99	1.00	.01 (.00 to .11)	37.04	.01
モデル 2	5.87 (7), .56	0.84	.99	1.00	.00 (.00 to .08)	33.87	.02

注:  $n = 175$ . GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = 信頼区間(Confidence Interval); AIC = Akaike Information Criterion; SRMR = Standardized Root Mean Square Residual. 基準は豊田 (2007), Arbuckle and Wothke (1995), Hu and Bentler (1999)を参照。SRMR は Plugins→Standardized RMR でダイアログボックスを開き、開いたままで分析をすると値が算出される。

モデル 2 によると、語彙サイズは TOEIC リスニング得点を 44%(長方形の右上の値 .44 より)、TOEIC リーディング得点を 49%予測できていた。予測率が有意に異なるかについては、「差に対する検定統計量」を Amos で出力し検討したところ、1.80 で 1.96 を超えていないため 5%水準で有意な違いがあるとは言えなかった。よって、語彙知識が英語技能に影響すると考えるならば、(アクセス速度の直接のパスはないため)語彙サイズがリスニング力、リーディング力に同程度の影響を与えていたと考えられる。なお、SEM で説明できた割合(TOEIC リスニング得点 44%、リーディング得点 49%)は、上記の重回帰分析で説明できた値のリスニング得点 39%、リーディング得点 45%よりも高かった。これは SEM では誤差を考慮し、潜在的な因子を仮定し

ているためと考えられる。

一方、アクセス速度は予測に貢献していなかった。しかし、アクセス速度をモデルから抜くと適合度指標の1つがかなり悪くなる(RMSEA = .10 [.00 to .25])ため、TOEIC得点を予測する際にサイズとアクセス速度の2要素で考えることは、よりよいモデル作成につながることを示している。

TOEICリスニング得点とTOEICリーディング得点の誤差変数の相関が .68 だったことから、語彙のサイズで説明されないリスニング得点とリーディング得点の関係は中程度あることも分かる。

さらに、サイズにおいては、J8VST1 得点と J8VST2 得点への負荷がともに .91 であり、2つの説明率が83%であることから、語彙サイズを安定して測定できていると考えられる。また、アクセス速度においては、LEXATT タスク 1 得点、タスク 2 得点への負荷が .61, .69 であり、2つの説明率が 37%, 48%であることから、アクセス速度を安定して測定できていると考えられる。サイズとアクセス速度のテストの構造的な妥当性の証拠の1つと考えてよいだろう(3.2.2.1.節参照)。

なお、各 TOEIC 得点の予測率は高いものの、半分以下であり、語彙知識以外の要素も関わることも確認しておきたい。今回は LOT の信頼性が低いためモデルに入れていないが、今後 LOT の信頼性を確保した上で再度検証が必要である。

#### 5.1.3.4 まとめ

本節では、3種類の語彙知識テストと TOEIC リーディング・リスニング得点がどのように関係しているかを吟味した。RQ1 の「3つの語彙テストは、TOEIC リーディング・リスニング得点をどの程度予測できるか」については、共分散構造分析(SEM)の結果で解釈すると、J8VST により、TOEIC リスニング得点の 44%、TOEIC リーディング得点の 49%を予測できており、この説明率は高いものだった。RQ2 の「どの語彙テストが、TOEIC リーディング・リスニング得点をより予測できるか」については、J8VST がより予測できていた。

## 5.1.2 CASEC 得点の予測

### 5.1.2.1 研究の目的

本節では、言語習熟度を測定するテストとして CASEC を用い、3 種類の語彙知識テストと CASEC の関係を吟味する。リサーチ・クエスチョン(RQs)は以下の 2 つであり、重回帰分析を用いて調べた。

RQ1:3 つの語彙テストは、CASEC 得点をどの程度予測できるか。

RQ2:どの語彙テストが、CASEC 得点をより予測できるか。

### 5.1.2.2 方法

#### 受験者

関東地区の私立大学の英語を専門とする学生 39 名であった。これは本報告書の 5.3 節での受験者と一部重なる。

#### 使用テストと手順

語彙サイズテストである J8VST, 語彙構成テストである LOT, 語彙アクセス速度テストである LEXATT タスク 1 とタスク 2 をウェブ上で 2008 年 7 月に実施した。語彙テストは LEXATT タスク 1 とタスク 2, J8VST, LOT の順で実施した。語彙テストの 1 週間後に CASEC を行った。

CASEC (Computerized Assessment System for English Communication)は、項目応答理論とコンピュータ適応型テストシステムを用いた、コミュニケーション能力判定テストである(野上, 2009)。セクション 1~4 から成り、それぞれ「語彙の知識」、「表現の知識・用法」、「聞いた内容の大意を理解する能力」、「具体的な情報を聞き取る能力」を測定するとされる。テスト方法は、セクション 1~3 は多肢選択式で、セクション 4 は聞いた英語を書き取りコンピュータに打ち込むディクテーション形式である。

#### 分析

語彙テスト得点と CASEC 得点を用いて、相関分析と重回帰分析を行った。共分散構造分析は人数が少ないため行わなかった。相関では、ピアソンの積率相関係数を算出した。

重回帰分析は、複数の独立変数(4 種類の語彙テスト得点)から 1 つの従属変数(CASEC 得点)を何パーセント説明できるか、複数ある独立変数の中でどれが一番従属変数を説明できるかを調べることができる。

重回帰分析では変数増加法を用いた。独立変数によってどのくらいの割合の従属変数が説明できるかを調べる際には、 $R^2$  (決定係数)と修正済み  $R^2$  を用い、Cohen (1988, pp. 413-414)の基準で修正済み  $R^2$  を解釈した。具体的には .0196 未満ならばほとんど予測できず、.0196 以上 .13 未満ならば少し予測でき、.13 以上 .26 ならば予測が中程度でき、.26 以上ならば予測が多くできると解釈した。分析には SPSS (2003)を使用した。

### 5.1.2.3 結果と考察

#### 記述統計

表 5.7 は、各変数の記述統計を示している。LEXATT タスク 2 得点以外は正規分布に近かったが、LEXATT タスク 2 得点は尖度が高かった。計算可能な各変数の信頼性のほとんどが高かったが、LOT では低めだった( $\alpha = .56$ )。これは、LOT 得点が今回の大学生ではあまりばらつきが見られなかったことが 1 つの理由であろう(5.1.1.3 節参照)。

なお、本節の研究における受験者の推定語彙サイズ(素点/125\*500 で推定)は 1680~4640 語(平均値 3454.36 語, 標準偏差 708.53 語)であり、この程度の語彙サイズの受験者だったことがわかる。これは、TOEIC リスニング・リーディング得点を予測した 5.1.1 節での受験者(2440~4840 語, 平均値 3,776.23 語, 標準偏差 517.38 語)と比較すると、本 CASEC の分析の方が、より語彙サイズの幅が広く、平均値が低く、標準偏差も大きかった。LOT の信頼性が、本 CASEC での受験者の方( $\alpha = .56$ )が 5.1.1.3 節の結果( $\alpha = .31$ )より高いのは、そのためかもしれない。なお、本節では CASEC 得点の予測が焦点であるが、可能な範囲で CASEC のセクションごとの得点での結果も参考までに挙げた。

表 5.7 変数の記述統計( $n = 39$ )

	平均	標準偏差	歪度	尖度	最小値	最大値	信頼性
CASEC 得点	499.51	120.18	0.13	0.62	242	836	.88 <sup>a</sup>
パート 1 得点	125.85	34.77	0.27	-0.02	60	212	--
パート 2 得点	123.00	34.98	-0.07	0.74	37	214	--
パート 3 得点	132.87	35.99	-0.07	0.03	41	210	--
パート 4 得点	117.79	34.84	-0.25	0.36	34	205	--
J8VST ( $k = 125$ )	86.36	17.71	-0.47	-0.34	42	116	.94
LOT ( $k = 50$ )	19.79	4.61	-0.03	0.92	9	33	.56
LEXATT タスク 1 ( $k = 40$ )	5189.31	1917.09	1.04	1.22	2057	10880	.81 <sup>b</sup>
LEXATT タスク 2 ( $k = 40$ )	971.46	374.44	0.99	2.31	336	2233	--

注: $k$  = 項目数. <sup>a</sup>CASEC の信頼性は、今回の受験者においてどうだったかは素データがないため分析できないが、総合得点は 4 つのセクションを総合して出すことを考え、4 セクション間の一貫性という意味での信頼性を、クロンバックの  $\alpha$  係数で算出した。<sup>b</sup>LEXATT タスク 1 の問題は、文字表示枠の 1 枠目、3 枠目、5 枠目、7 枠目から始まる単語で正解が構成されている。各枠から始まる単語において正解した場合の認知処理時間が算出されているため、単語が 1 枠目、3 枠目、5 枠目、7 枠目で始まる認知処理時間が一貫しているかという観点での信頼性を、クロンバックの  $\alpha$  係数で算出した。CASEC は 1000 点、CASEC パート 1, 2, 3, 4 は全て 250 点満点。

### 相関分析

語彙テスト得点と CASEC 得点の相関で最も高かったのは、J8VST とで .77 であり、最も低かったのは LEXATT タスク 1 得点の -.30 だった(表 5.8 参照)。CASEC パート 1 は語彙問題であるが、それと J8VST の相関が .82 と他のセクションとの相関より高かった。似た能力・知識を測っているもの同士の相関はそうでない相関よりも高くなるはずであり、そのような結果が出たことは、J8VST の妥当性の 1 つの証拠になると考えられる。語彙テスト間の相関を見ると、サイズ・構成・アクセス速さ間では相関は高く中程度(.47)であり、後の重回帰分析において多重共線性(multicollinearity)が問題になる程高いものはなかった。

表 5.8 変数間の相関

	CASE C1	CASE C2	CASE C3	CASE C4	J8VS T	LOT	LEXAT T タスク 1	LEXATT タスク 2
CASEC	.82**	.86**	.89**	.85**	.77**	.46**	-.30	-.36*
CASEC1	--	.67**	.60**	.54**	.82**	.54**	-.34*	-.29
CASEC2		--	.68**	.58**	.71**	.37*	-.09	-.36*
CASEC3			--	.77**	.60**	.37*	-.21	-.22
CASEC4				--	.51**	.30	-.37*	-.36*
J8VST					--	.47*	-.11	-.30
LOT						--	-.21	-.30
LEXATT タスク 1							--	.47**
LEXATT タスク 2								--

注: \* $p < .05$ . \*\* $p < .01$ .

### 重回帰分析

重回帰分析の結果が表 5.9 と表 5.10 である。サイズが語彙知識の基本的な知識であることを考え、J8VST を最初に、次に LOT, LEXATT の順に重回帰式に入れて分析した。その結果、J8VST だけで CASEC 得点の 59%を説明し、LOT が次の 1%を、最後の LEXATT は 4%を説明しており、全体で 64%の分散が 3 つの語彙テストで説明された。Cohen (1988)の基準を用いると、修正済み  $R^2$  では J8VST による予測が多くでき、LOT と LEXATT による予測が少しできていた。この割合は 5.1.1 節(TOEIC リスニング得点で 41%, TOEICリーディング得点で 48%)と 5.3 節(Versant 得点で 59%)の割合よりも高かった。その理由の 1 つは、CASEC が TOEIC よりも英語が苦手な層の能力の違いも弁別でき、受験者を幅広くばらつかせることができるために、語彙知識との関係が見えやすかったため、二つ目としては、語彙テストと CASEC はともにパソコン上で行う形式であり、テスト形式が共通していたためと考えられる。

表 5.10 を見ると、標準偏回帰係数では、J8VST の値のみが有意で、最も高かった。この結果は TOEIC リスニング得点とリーディング得点の予測の結果(5.1.1 節)と同じパターンだった。



表 5.9 変数増加法で、独立変数を順に追加したときの決定係数

	CASEC		
	$R^2$	修正済み $R^2$	$\Delta R^2$
J8VST	.59**	.58	--
J8VST+LOT	.61**	.58	.01
J8VST+LOT+LEXATT タスク 1+タスク 2	.64**	.60	.04

表 5.10 4 つの変数を入れたときの偏回帰係数

CASEC 得点	$B$	$SEB$	$\beta$	下限 <sup>a</sup>	上限 <sup>a</sup>
定数(Constant)	119.22	91.75	--	-67.23	305.67
J8VST	4.77**	0.80	.70	3.13	6.40
LOT	2.02	3.10	.08	-4.28	8.32
LEXATT タスク 1	-0.01	0.01	-.19	-0.03	0.00
LEXATT タスク 2	-0.01	0.04	-.03	-0.09	0.07

注: <sup>a</sup> $B$ の 95%信頼区間. \* $p < .05$ . \*\* $p < .01$ .

表 5.10 の結果を基に作成したのが表 5.11 の回帰式である。各テストの得点を式に代入すれば、CASEC 得点の予想得点が出る。しかし、重回帰分析での説明率が 64%と 3分の 2 以下のため、信頼区間の幅は広く、あまり精密な点数の予測には役立たないことも示された。具体的には、CASEC 得点は 0~1000 点の間で表示される(野上, 2009)が、平均的な受験者の場合の予測は 95%信頼区間で -106.85~1105.87 点であり、受験者の語彙テストから予測される CASEC 得点は、95%の確率で 0 点から 1000 点の間としか言えないことが示された。

表 5.11 語彙テスト得点から CASEC 得点を予測する回帰式

< CASEC 得点 >

■点推定:

$$119.22 + 4.77 \times (\text{J8VST 得点}) + 2.02 \times (\text{LOT 得点}) + (-0.01) \times (\text{LEXATT タスク 1 得点}) + (-0.01) \times (\text{LEXATT タスク 2 得点})$$

■95%信頼区間下限:

$$-67.23 + 3.13 \times (\text{J8VST 得点}) + (-4.28) \times (\text{LOT 得点}) + (-0.03) \times (\text{LEXATT タスク 1 得点}) + (-0.09) \times (\text{LEXATT タスク 2 得点})$$

**■95%信頼区間上限**

$$305.67 + 6.40 \times (\text{J8VST 得点}) + 8.32 \times (\text{LOT 得点}) + (0.00) \times (\text{LEXATT タスク 1 得点}) + (0.07) \times (\text{LEXATT タスク 2 得点})$$

**■予測値(表 5.7.にある平均を四捨五入した値を重回帰式に代入したときに得られた予測値)**

$$499.51 (-106.85 \text{ to } 1105.87)$$

**5.1.2.4 まとめ**

本節では、3種類の語彙知識テストとCASEC得点がどのように関係しているかを吟味した。RQ1の「3つの語彙テストは、CASEC得点をどの程度予測できるか」については、重回帰分析の結果で解釈すると、J8VSTとLOTとLEXATTにより、CASEC得点の64%を予測できており、修正済み $R^2$ で考えても58%と、説明率が高いものだった。RQ2の「どの語彙テストが、CASEC得点をより予測できるか」については、J8VSTがより予測できていた。今後、他の言語習熟度を測定するテストと語彙テストの関係を調べ、この説明率などがどの程度一般化できるかを調べる必要があるだろう。次の節では、その一貫として、言語習熟度を測定するテストとしてTOEIC Bridgeを用い、語彙テストの関係を調べる。

## 5.2 TOEIC Bridgeを用いた習熟度と語彙力との関係に関する調査

### 5.2.1 はじめに

本科研による研究では、日本人英語学習者を対象として、その語彙力を多面的に測定する語彙テストの開発を行っている。個々の語彙テストは、学習者の持つ語彙力を異なった観点から測定していると考えられるため、それぞれのテストによる語彙力の測定結果がどのような相関関係にあるかについての分析が必要である。また、それぞれの語彙力が、総合的な英語の習熟度とはどのように関連しているかについても調査を行う必要がある。そのため、本稿では、本科研プロジェクトにおいて開発された3種類の語彙テスト(LOT, LEXATT, J8VST)の相関関係を、日本人英語学習者を対象として調査するとともに、TOEIC Bridgeのスコアを用いて、これらのテストで測定される語彙力と英語習熟度との関連を分析する。

### 5.2.2 テストの実施方法

今回の調査は、2008年度前期に日本人大学生(理工系学部の1年生)を対象として実施したもので、調査対象の大学生は、4月の入学当初にTOEIC Bridgeによるプレースメントテストを受験した学生のうちの53名である。これらの受験者のTOEIC Bridgeの平均点は、128.1点で、この得点は概算でTOEIC 340点に相当する。TOEIC Bridgeは、TOEICに比べると問題のレベルも比較的易しいものであり、受験時間も短くなっている(60分間)。そのため、大学入学時の学生を対象とする場合や、長時間の受験時間が確保できない場合の実施に適している。TOEICを利用した場合、床面効果(floor effect)が生じて受験者の英語力の正確な測定が困難になるような場合にも、これらの問題点を回避できるという利点がある。テストは、リスニング・セクションとリーディング・セクションに分かれており、それぞれ50問出題される。リスニング・セクションは、パート 1: 写真問題、パート 2: 応答問題、パート 3: 短い話 で構成されている。リーディング・セクションは、パート 4: 不完全な文、パート 5: 読解 で構成されている。得点は180点満点で算出される。

上記の53名を対象として、コンピュータ教室での授業中に3種類の語彙テストを実施した。最初に語彙認識テストであるLEXATTを実施し、次に語彙サイズテストのJ8VSTを実施した。J8VSTについては和英選択モードのテストを実施した。最後に、語彙構成診断テストのLOTを実施した。実施時間は受験方法の解説時間を含め、3つのテストで約80分であった。それぞれの語彙テストの詳細については、各テストの解説の項を参照されたい。

### 5.2.3 結果と考察

#### 5.2.3.1 記述統計量

表1は、各テストの記述統計量を示している。今回の被験者は、TOEIC Bridgeのスコアによるクラス分けによって一定のレベルに分けられたグループであるため、当然のことながらTOEICの得点に関しては、ばらつきが少ない(TOEIC Bridge 平均点=128.1 標準偏差= 2.81)。それと比較すると、LOTやJ8VSTのスコア、LEXATTの反応時間は被験者間でのばらつきが大きくなる傾向が見られる(LOT平均点= 19 標準偏差= 3.43, LEXATT タスク1 平均反応時間= 11158 ms. 標準偏差 5062 ms, LEXATT タスク2 平均反応時間= 1838 ms ms. 標準偏差 614 ms)。

表5.12 各テストの記述統計量

	N	Minimum	Maximum	Mean	SD	Variance
LOT	53	12	28	19.02	3.43	11.78
LEXATT_タスク 1	53	24	39	33.77	3.76	14.17
LEXATT_タスク 2	53	30	38	35.34	1.79	3.22
LEXATT_タスク 1_RT (ms)	53	2757	11158	5062.34	1688.98	2852672.69
LEXATT_タスク 2_RT (ms)	53	236	1838	613.53	331.09	109623.60
VOC	53	60	110	85.81	12.41	154.19
TOEIC Bridge Total	53	124	132	128.11	2.81	7.91
TOEIC Bridge Listening	53	54	74	62.11	4.10	16.83
TOEIC Bridge Reading	53	58	74	66	3.94	15.53

### 5.2.3.2 テスト間の相関分析

次に、各テスト間の相関係数(Pearson)を求めた。それらを表5.13に示す。

LOTについては、他の語彙テストやTOEICの得点とは有意な相関が見られなかったが、これはLOTがコロケーション知識を問うテストであり、他の語彙テストとは異なるものを測定しているためと推測される。Meara and Wolter (2004)は、語彙構成を測定するテストとしてV\_Linksを開発し、147名の日本人英語学習者にこのテストを行ったところ、V\_linksの得点と語彙サイズの相関係数は、0.3以下であったと報告している。今回、有意ではないものの、LOTとJSVSTとの間の相関が $r = 0.259$ であったことは、同じ日本人英語学習者の語彙知識の構成度合いを表す数値として興味深いものである。

LEXATTのパート 1の得点(LEXATTタスク1)は、LEXATTのパート 2の得点(LEXATTタスク2)、JSVST、TOEIC Bridge及びTOEIC Bridge リーディング・セクションとの間にそれぞれ中程度の相関が見られた。LEXATT タスク 1は、ランダムな文字列の羅列から有意な英単語のスペリングを発見

表5.13. テスト間の相関係数

	LOT	LEXATT_ PT1	LEXATT_ PT2	LEXATT_PT1_ RT	LEXATT_PT2_ RT	J8VST	TOEIC Br_ Total	TOEIC Br_ Listening	TOEIC Br_ Reading
LOT	1.000								
LEXATT_PT1	.244	1.000							
LEXATT_PT2	.105	.267	1.000						
LEXATT_PT1_RT	-.103	.006	.125	1.000					
LEXATT_PT2_RT	-.030	.333*	.078	.458**	1.000				
J8VST	.259	.387**	.266	-.285*	-.080	1.000			
TOEIC Br_Total	.187	.402**	-.008	-.120	-.180	.330*	1.000		
TOEIC Br_Listening	-.088	-.140	-.313*	-.118	-.019	-.074	.399**	1.000	
TOEIC Br_Reading	.225	.433**	.320*	.037	-.109	.313*	.298*	-.756**	1.000

\* . Correlation is significant at the 0.05 level (2-tailed).

\*\* . Correlation is significant at the 0.01 level (2-tailed).

させるテストであるが、このような語形認識には、J8VSTで測定されるような語彙サイズが一定程度必要となることが示されている。また、このような能力がTOEIC Bridgeのリーディング・セクションの得点と相関があるということは、リーディングには、文字化された英単語に対する視認語彙(sight vocabulary)的な認識力が関係していることを示している。また、英単語の認識をさせてその意味を尋ねる形式のLEXATT タスク 2について

も、TOEIC Bridge リーディング・セクションとの間に弱い相関が見られ、リーディングと語彙力との関係がここでも示されている。

J8VSTについては、LEXATT タスク 1とTOEIC Bridge及びTOEIC Bridge リーディング・セクションとの間に弱～中程度の相関が見られ、これらのテストには語彙サイズがある程度必要であることを示している。

次に、LEXATTの反応時間(正答した問題のみ)と他のテストとの関係を見ると、LEXATT タスク 1の正反応時間(LEXATTタスク1\_RT)とJ8VSTのスコア間に弱い負の相関が見られた( $r = -.285$ )。反応時間が長い被験者ほど、単語の認識力が劣ると考えられることから、反応時間とJ8VSTとの間に負の相関があるということは、ランダムな文字列から有意な単語を素早く認識するアクセス能力は、語彙サイズと関係があるということを示している。なお、LEXATTのタスク 1とタスク 2との反応時間には中程度の相関が見られ( $r = .485$ )、2種類のテストが測定しているアクセス能力には共通性があることが認められた。しかしながら、これらの反応時間とTOEIC Bridgeの得点間には、有意な相関が見られなかった。

今回の被験者の特徴として、TOEIC Bridgeのリスニング・セクションとリーディング・セクションの得点間に強い負の相関が見られた。これは、今回の被験者グループがTOEIC Bridgeの得点でレベル分けされた集団であることに起因する切断効果がある原因であると考えられる。また、リスニング・セクションの得点は、LEXATTのタスク 2を除くと、他の語彙テストとは有意な相関が無く、今回の被験者は、自分の持っている語彙力をリスニングには十分に使用できていないと思われる。彼らがあまり英語が得意ではない学習者であったため、リーディング能力とリスニング能力がバランスよく発達していない学習者である可能性があり、TOEIC Bridge等のテストのスコアでクラス分けを行う際には、これらの特徴に注意する必要があると言えるであろう。

### 5.2.3.3 主成分分析

次に、本研究で実施した各種テストのスコアをもとに、LOTが測定するコロケーション知識、LEXATTが測定する語形及び意味へのアクセス能力とその速度、J8VSTが測定する語彙サイズ、TOEIC Bridgeで測定される総合的な英語力、を変数として、多変量解析の手法の一つである主成分分析を行った。

主成分分析に用いた変数については、1)LOTの得点、2)LEXATT タスク 1の値、3)LEXATT タスク 2の値、4)J8VSTの得点、5)TOEIC Bridgeの得点、の5つである。なお、LEXATTに関しては、テストの得点を採用した場合と、反応時間を採用した場合の2つに分けて、別々に分析を行った。分析に使用した統計ソフトはSPSS 13.0である。

### 5.2.3.4 主成分分析(1)

最初の主成分分析では、1)LOTの得点、2)LEXATT タスク 1の得点、3)LEXATT タスク 2の得点、4)VOCの得点、5)TOEIC Bridgeの得点、を変数として主成分分析を行った。

主成分分析の結果、表5.14に示すような成分が抽出された。固有値が1以上のものは第1主成分と第2主成分であり、累積寄与率は60.8%であった。表5.15は、第2主成分までの主成分負荷量を示している。第1主成分は、全てのテストの得点と正の相関があり、語彙力による総合的な英語力を示す能力と解釈できる。第2主成分は、LEXATT タスク 2の得点との相関が強く、語彙の意味知識を示すものと解釈できる。これらの値の主成分負荷プロット(成分プロット)を図1に示す。この図を見ると、語彙サイズを測定するJ8VSTと語形認

識力を測定するLEXATT タスク 1は、お互いに近接しており、これらのテストが測定している能力は類似している部分があることが示されている。また、TOEIC BridgeとLEXATT タスク 2はお互いに離れた場所に位置しており、関連性が低いことが示されている。

表5.14 説明された分散の合計

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.03	40.53	40.53	2.026	40.53	40.53
2	1.01	20.24	60.77	1.012	20.24	60.77
3	0.84	16.77	77.54			
4	0.62	12.34	89.88			
5	0.51	10.12	100.00			

Extraction Method: Principal Component Analysis.

表5.15 主成分負荷量

	Component	
	1	2
LOT	0.53	-0.11
LEXATT タスク 1	0.77	-0.02
LEXATT タスク 2	0.44	0.83
J8VST	0.74	0.08
TOEIC Br	0.63	-0.56

Extraction Method: Principal Component Analysis. 2 components extracted.

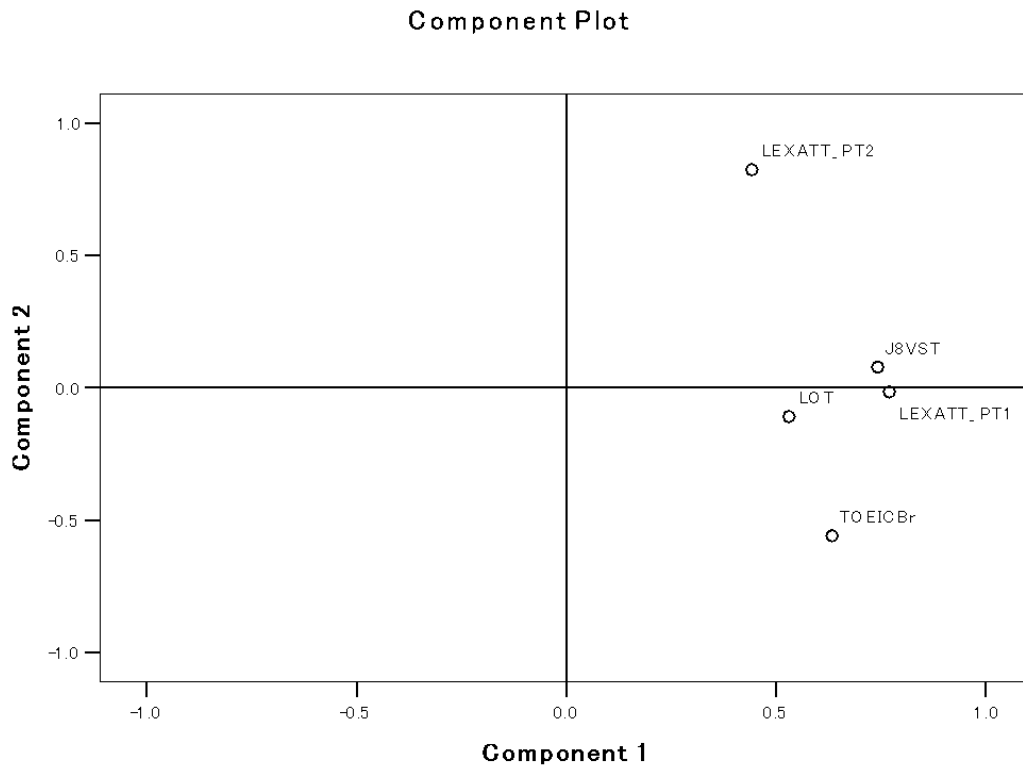


図5.4 主成分負荷プロット(LEXATT = スコア)

### 5.2.3.5 主成分分析(2)

2つ目の主成分分析では、(1)LOTの得点、2)LEXATT タスク 1の反応時間、3)LEXATT タスク 2の反応時間、4)J8VSTの得点、5)TOEIC Bridgeの得点、を変数として分析を行った。なお、2)及び3)に関しては、問題に正答した場合(正反応)のみの平均反応時間を求めた。また、反応時間に関しては、短いほど優れていると判断できるため、解釈を容易にするために反応時間に-1を乗じて負の数値に変換した。

主成分分析の結果、表5.16に示す成分が抽出された。固有値が1以上になったものは第1主成分と第2主成分であった。また、累積寄与率は第2主成分までで60.1%であった。表5.17は主成分負荷量を表している。これを見ると、第1主成分は主にLEXATT タスク 1の反応時間とJ8VSTの得点と強い関連があり、全てのテストの得点と正の相関があることから、語彙力をベースとする総合的な英語力を示す能力と解釈できる。第2主成分はLOTのスコアとの相関が最も強いことから、語彙のコロケーション能力を表す成分であると解釈される。これらの値の成分プロット(主成分負荷プロット)を図5.5に示す。この図を見ると、J8VSTとTOEIC Bridgeは近接しており、これらのテストが測る能力は近い関係にあることが示されている。また、LEXATTは、タスク 1とタスク 2が近い関係にあるものの、その他のテストとは離れた場所に位置しており、LEXATTが測定している語形や意味へのアクセス速度は、他のテストが測定している英語力とは異なるものであることが示されている。



表5.16 説明された分散の合計

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.83	36.66	36.66	1.83	36.66	36.66
2	1.17	23.46	60.12	1.17	23.46	60.12
3	0.82	16.32	76.44			
4	0.73	14.50	90.94			
5	0.45	9.06	100.00			

Extraction Method: Principal Component Analysis.

表5.17 主成分負荷量

	Component	
	1	2
LOT	0.45	0.54
LEXATT タスク 1_RT	0.70	-0.47
LEXATT タスク 2_RT	0.59	-0.63
J8VST	0.67	0.38
TOEIC Br	0.59	0.34

Extraction Method: Principal Component Analysis.

2 components extracted.

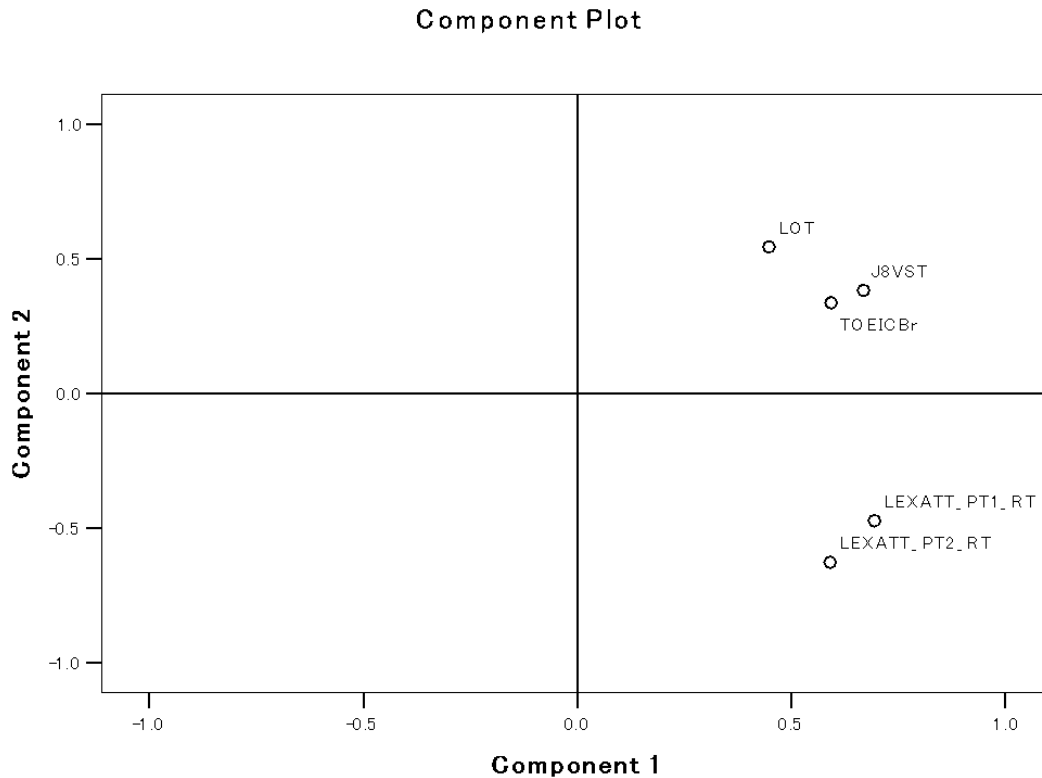


図5.5 主成分負荷プロット(LEXATT = 反応時間)

#### 5.2.4 まとめ

本研究では、3種類の語彙テストとTOEIC Bridgeの得点間の相関関係の分析を行った。その結果、今回のような英語力があまり高くない被験者では、TOEIC Bridgeで測定されるような総合的な英語習熟度と関係する語彙力は、語彙サイズ(J8VST)と語形認識力(LEXATT タスク 1得点)であり、語形や意味へのアクセス速度やコロケーションと能力はあまり関係が無い可能性が示された。また、今回、多くのテスト間に有意な相関関係が見られなかったり、相関係数が全体的に低くなり、主成分分析の累積寄与率も低くなったことについては、被験者数が少なかったことや、被験者の英語力が相対的に低かったことが影響している可能性がある。別章(3.3 語彙構成テストによるTOEICリーディング得点・リスニング得点の予測)の調査では、TOEICの平均点が422.2の大学生では、TOEICとLOTの相関は、 $r = .55$  ( $p < .001$ )であり、各セクションとLOTとの相関も、これとほぼ同じ程度であったことが報告されている。そのため、本研究の調査結果は、被験者の英語力が影響している可能性も捨てきれない。そのため、今後、より多くの学習者を対象とした調査を行って今回の分析結果を再検証することが必要である。

## 5.3 Versantを用いたスピーキング・リスニング力の予測

### 5.3.1 研究の目的

本節では、3種類の語彙知識テストと Versant 得点の関係を吟味する。Versant は英語コミュニケーション力(基本的な口頭言語能力, スピーキング・リスニング力)を測ることを意図したテストであり, 語彙テストで Versant 得点をどの程度予測できるかを調べる。リサーチ・クエスチョン(RQs)は以下であり, 重回帰分析と共分散構造分析を用いて調べた。

RQ1: 3つの語彙テストは, Versant 得点をどの程度予測できるか

RQ2: どの語彙テストが, Versant 得点をより予測できるか

### 5.3.2 方法

#### 受験者

日本全国の国立・私立大学, 海外の大学17校に通う, 大学生62名と大学院生38名の計100名であった。全員が, 3つの語彙テストと Versant の全てを受験した。大学生の内訳は, 私立大学文系58名, 国立大学文系3名, 公立大学文系1名で, 大学院生の内訳は, 国立大学文系23名, 私立大学文系7名, 国立大学理系4名, 海外の大学文系4名であった。これは5.1.2節での受験者と一部重なる。

#### 使用テストと手順

語彙サイズテストである J8VST, 語彙構成テストである LOT, 語彙アクセス速度テストである LEXATT タスク1とタスク2をウェブ上で受験した。語彙テストは LEXATT タスク1とタスク2, J8VST, LOT の順で実施した。Versant は, 語彙テストより前と後に受験した場合があったが, いずれの場合も1ヶ月以内にテストを受けた。2008年7月～2009年3月の間に実施した。

Versant は基本的な口頭言語能力(スピーキング・リスニング力)を測ることを意図したテスト(Pearson Education, 2009a)である。セクションは5つあり(表5.18参照), 電話を使って行われる(約13分)。Versant 受験後には, 総合得点, Structure, Vocabulary, Fluency, Pronunciation の下位得点が得られる。採点は人間の評価のパターンを組み込んだ機械が自動採点し, 約5分後には Web 上で結果が見られる。

受験者には, 受験前に Versant の内容について著者が説明するか, 受験前にテストの内容についてよく確認してから受験するよう求めるかのどちらかの方法をとった。Versant のテストの質については, Downey, Farhady, Present-Thomas, Suzuki, and Van Moere (2008)などを参照のこと。

表 5.18 Versant のセクション構成

Part	内容	項目数	自動採点
Part A	音読: 読まれた番号の文を音読する。	8 問	○
Part B	復唱: 聞こえた文を繰り返す。	16 問	○
Part C	質疑応答: 常識で答えられる質問に対して 1 語または 2, 3 語で答える。	16 問	○
Part D	文の構築: 3 つの語群を聞き、意味の通る文に並び替えて言う。	10 問	○
Part E	自由回答: 2 回質問を聞き、質問に対する意見(長めのモノローグ)を述べる。20 秒が準備時間で、40 秒で意見を述べる。	3 問	× <sup>a</sup>

注: <sup>a</sup> 自動採点ではなく点数も出ないが、テスト使用者が音声聞いて使うことができる。

## 分析

それぞれの変数と語彙テスト得点の変数を用いて、相関分析・重回帰分析・共分散構造分析(Structural equation modeling: SEM)を行った。相関は、ピアソンの積率相関係数を算出した。

Versant からは 5 種類の得点(Versant(総合)得点, 構造, 語彙, 流暢さ, 発音)が得られるが, 重回帰分析では Versant 得点を従属変数とし, 独立変数は 4 種類の語彙テスト得点とした。重回帰分析では変数増加法を用いた。独立変数によって何割の従属変数が説明されたかを調べる際には,  $R^2$ (決定係数)と修正済み  $R^2$  を用い, Cohen (1988, pp. 413-414)の基準で修正済み  $R^2$  を解釈した。具体的には .0196 未満ならばほとんど予測できず, .0196 以上 .13 未満ならば少し予測でき, .13 以上 .26 ならば予測が中程度でき, .26 以上ならば予測が多くできると解釈した。

SEM では最尤法(maximum likelihood method)を用いた。相関・重回帰分析は SPSS (2003)で, SEM は Amos<sup>TM</sup> 7.0 を用いて分析した。

なお, Koizumi (2008)では自由回答のパート E での発話を一部書き起こし, 談話分析的指標を用いて分析したが, 本節ではその結果は述べない。

### 5.3.3 結果と考察

表 5.19 は, 各変数の記述統計を示している。LEXATT タスク 1 得点以外は正規分布に近かったが, LEXATT タスク 1 得点は尖度が高かった。計算可能な各変数の信頼性はすべて高かった。

なお, 本節の研究における受験者の推定語彙サイズ(素点/125\*500 で推定)は 1680~4920 語(平均値 3919.60 語, 標準偏差 775.62 語)であり, この程度の語彙サイズの受験者であったことがわかる。これは, CASEC 得点を予測した 5.1.2 節での受験者層(1680~4640 語, 平均値 3454.36 語, 標準偏差 708.53 語)と比較すると, やや本 Versant 分析の方が能力の幅が広く, 平均値が高かった。LOT の信頼性が, 本 Versant での分析での受験者の方( $\alpha = .77$ )が 5.1.2 節での結果( $\alpha = .56$ )より高いのは, そのためかもしれない。なお, 本節の重回帰分析では Versant 得点が焦点であるが, SEM では Versant の下位得点をデータとして用いた。

表 5.19 変数の記述統計( $n = 100$ )

	平均	標準偏差	歪度	尖度	最小値	最大値	信頼性
Versant 得点	38.96	12.65	0.55	-0.64	20	69	.94 <sup>a</sup>
Versant 構造	38.41	12.63	0.23	-0.81	20	67	--
Versant 語彙	40.46	14.65	0.45	-0.71	20	79	--
Versant 流暢さ	37.76	14.28	0.75	-0.40	20	75	--
Versant 発音	40.97	12.44	1.05	0.30	26	80	--
J8VST ( $k = 125$ )	97.99	19.39	-0.80	-0.14	42	123	.96
J8VST (奇数項目 $k = 63$ )	48.81	10.24	-0.93	0.15	17	63	.93
J8VST (偶数項目 $k = 62$ )	49.18	9.49	-0.65	-0.42	25	62	.92
LOT ( $k = 50$ )	23.84	6.35	0.33	-0.01	9	41	.77
LEXATT タスク 1 ( $k = 40$ )	4518.91	1710.57	1.43	2.67	1884.00	10880.0	.81 <sup>b</sup>
LEXATT タスク 2 ( $k = 40$ )	803.56	361.92	1.02	1.66	274.00	2233.00	--
LEXATT タスク 1 変換後	66.17	11.92	0.86	1.04	43.41	104.31	--
LEXATT タスク 2 変換後	27.67	6.21	0.39	0.03	16.55	47.25	--

注:  $k =$  項目数. <sup>a</sup>Versant 得点は 4 つの下位得点を総合した得点であることを考え、4 つの下位得点の一貫性という意味での信頼性を、クロンバックの  $\alpha$  係数で算出した。<sup>b</sup>LEXATT タスク 1 の問題は、文字表示枠の 1 枠目、3 枠目、5 枠目、7 枠目から始まる単語で正解が構成されている。各枠から始まる単語において正解した場合の認知処理時間が算出されているため、単語が 1 枠目、3 枠目、5 枠目、7 枠目で始まる認知処理時間が一貫しているかという観点での信頼性を、クロンバックの  $\alpha$  係数で算出した。

## 相関分析

語彙テスト得点と Versant 得点の相関で最も高かったのは、J8VST とで .70 であった。最も相関が低かったのは LEXATT タスク 1 得点との -.27 だった(表 5.20 参照)。語彙テスト間の相関を見ると、サイズ・構成・アクセス速度間では相関は高く中程度(.73)であり、後の重回帰分析や SEM において多重共線性 (multicollinearity) が問題になる程高いものはなかった。一方 Versant の下位得点間では高い相関があり、後の SEM 分析において全てを変数と入れようとすると多重共線性が起こる可能性があった。

表 5.20 変数間の相関

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)Versant 得点	.91	.92	.94	.91	.70	.68	.69	.69	-.27	-.51	-.27	-.52
(2) Versant 構造	--	.86	.75	.69	.62	.60	.63	.66	<u>-.25</u>	-.47	<u>-.25</u>	-.48
(3) Versant 語彙		--	.80	.74	.70	.69	.68	.68	-.30	-.47	-.30	-.49
(4) Versant 流暢さ			--	.95	.67	.66	.66	.68	<u>-.25</u>	-.45	<u>-.26</u>	-.47
(5) Versant 発音				--	.61	.59	.60	.56	<u>-.20</u>	-.44	<u>-.21</u>	-.46
(6)J8VST					--	.98	.98	.73	<u>-.19</u>	-.46	<u>-.20</u>	-.47
(7)J8VST (奇数)						--	.93	.70	<u>-.22</u>	-.48	<u>-.22</u>	-.49
(8)J8VST (偶数)							--	.74	<u>-.15</u>	-.42	-.16	-.44
(9)LOT								--	<u>-.19</u>	-.43	-.19	-.43
(10)LEXATT タスク 1									--	.49	.99	.49
(11)LEXATT タスク 2										--	.50	.99
(12)LEXATT 1 変換											--	.51
(13)LEXATT 2 変換												--

注: 下線+イタリック = 有意でない. 下線のみ  $p < .05$ . それ以外は\*\* $p < .01$ .

### 重回帰分析

重回帰分析の結果が表 5.21 と表 5.22 である。サイズが語彙知識の基本的な知識であることを考え、J8VST を最初に、次に LOT, LEXATT の順に重回帰式に入れて分析した。その結果、J8VST だけで Versant 得点の 49%を説明し、LOT が次の 7%を、最後の LEXATT で 3%を予測でき、全体で 59%の分散が 3つの語彙テストで説明された。Cohen (1988)の基準を用いると、修正済み  $R^2$  では J8VST による予測が多くでき、LOT と LEXATT は少し予測ができた。

表 5.22 を見ると、標準偏回帰係数では、J8VST と LOT の係数が同じ値( $\beta = .35$ )で最も高かった。5.1.1 節と 5.1.2 節では LOT より J8VST の方が常に予測力が高かったという結果(表 5.4 と表 5.10 参照)と比較すると、この結果は、(a)従属変数にスピーキング力の要素が入り、語彙とスピーキング・リスニング力の関係は、リスニング力単独・リーディング力単独・習熟度全般と語彙の関係とは異なるため、(b)LOT の信頼性が高かったため、もしくは(c)受験者層が広いためかもしれない。その理由を探るために、さらなる検討が今後必要だろう。

表 5.21. 変数増加法で、独立変数を順に追加したときの決定係数

	全体		
	$R^2$	修正済み $R^2$	$\Delta R^2$
J8VST	.49**	.48	--
J8VST+LOT	.56**	.55	.07
J8VST+LOT+LEXATT タスク 1+タスク 2	.59**	.57	.03

表 5.22. 4 つの変数を入れたときの偏回帰係数

Versant 得点	$B$	$SEB$	$\beta$	下限 <sup>a</sup>	上限 <sup>a</sup>
定数(Constant)	5.90	6.19	--	-6.38	18.19
J8VST	0.23**	0.06	.35	0.10	0.36
LOT	0.71**	0.19	.35	0.32	1.09
LEXATT タスク 1	0.00	0.00	-.05	0.00	0.00
LEXATT タスク 2	-0.01	0.00	-.17	-0.01	0.00

注:<sup>a</sup> $B$ の95%信頼区間. \* $p < .05$ . \*\* $p < .01$ .

表 5.22 の結果を基に作成したのが表 5.23 の回帰式である。各テストの得点を式に代入すれば、Versant の予想得点が出る。しかし、重回帰分析での説明率が 59%と 6 割以下のため、信頼区間の幅は広く、あまり精密な点数の予測には役立たないことが示された。具体的には、Versant 得点は 20~90 点の間で表示される(Pearson Education, 2009b)が、平均的な受験者の場合の予測は 95%信頼区間で-4.75~82.67 点であり、受験者の語彙テストから予測される Versant 得点は、95%の確率で 20 点から 83 点の間とは言えるが、それ以上のことは言えないことが示された。

表 5.23 語彙テスト得点から Versant 得点を予測する回帰式

<Versant 得点>

■点推定:

$$5.90 + 0.23 \times (\text{J8VST 得点}) + 0.71 \times (\text{LOT 得点}) + 0.00 \times (\text{LEXATT タスク 1 得点}) + (-0.01) \times (\text{LEXATT タスク 2 得点})$$

■95%信頼区間下限:

$$-6.38 + 0.10 \times (\text{J8VST 得点}) + 0.32 \times (\text{LOT 得点}) + (0.00) \times (\text{LEXATT タスク 1 得点}) + (-0.01) \times (\text{LEXATT タスク 2 得点})$$

■95%信頼区間上限

$$18.19 + 0.36 \times (\text{J8VST 得点}) + 1.09 \times (\text{LOT 得点}) + (0.00) \times (\text{LEXATT タスク 1 得点}) + (0.00) \times (\text{LEXATT タスク 2 得点})$$

■予測値(表 5.19.にある平均を四捨五入した値を重回帰式に代入したときに得られた予測値)

$$38.96 (-4.75 \text{ to } 82.67)$$

さらに、今回の結果と似たデザインの先行研究(Koizumi, 2005)と比較すると、本節ではスピーキング・リスニング力が受容語彙サイズ(J8VST)で 49%説明できたのに対し、Koizumi ではスピーキング力が発表語彙サイズで 60%説明できていた。測る能力の違い(本節ではスピーキング・リスニング力, Koizumi ではスピーキング力), 受験者層の違い(本節では大学生・大学院生, Koizumi では中高生), テスト形式の違い(本節では日本語の意味から英語を選ばせる形式, Koizumi では日本語の意味から英語を書かせる形式)などにより、説明の程度が変わるのだろう。しかし、サイズのみでかなりの能力が予測できる面は一貫していた。ただ Koizumi の方がサイズだけで説明できた割合が高く(60%), 受験者が中高生で能力がより低めという結果は、Adams (1980)の述べた、受験者の能力がより低いと語彙のスピーキングへの寄与が大きいという傾向と一致していた。

ただ、本節の研究と Koizumi (2005)ともに、扱ったスピーキングはモノローグの基本的なものであり(Versant では、文の繰り返しや並び替え; Koizumi では、絵の描写), より高度なスピーキング能力(ディスカッション, ディベート)や対話能力を含めたスピーキングを扱って調べると、語彙から技能が予測できる割合は、本節や Koizumi の結果ほど高くない可能性はあるため、今後調べていく必要があるだろう。

### 共分散構造分析(SEM)

ここでは、各テスト得点を観測変数(図 5.6 の四角)とした。潜在変数(図 5.6 の楕円)は 3 個あり、第 1 に、「J8VST の偶数番号での得点」と「J8VST の奇数番号での得点」の背後に「サイズ」を仮定し、第 2 に、「LEXATT タスク 1 得点」と「LEXATT タスク 2 得点」の背後に「アクセス速度」を仮定し、第 3 に、「Versant 構造得点」と「Versant 発音得点」の背後に「スピーキング・リスニング能力」を仮定し、モデル化し



た。Versant の 2 つの下位得点全部を使わなかったのは、いくつかの変数間で高い相関が見られたため、多重共線性の問題を避けるためである。構造と発音の相関は(.69)であり、問題にならない値だった。

まず、SEM を行う前に、Kunnan (1998) に基づき 2 つの前提を確認した。1 つ目の「変数ごとの正規性」については、「LEXATT タスク 1 得点」以外は、歪度・尖度ともすべて  $|\pm 2|$  内だったため満たしていた。「LEXATT タスク 1 得点」については、平方根変換を行った値が正規性を満たしたため、その値を用いた。2 つ目の前提である「変数を組み合わせた時の正規性(多変量正規性)」については、多変量正規性の値が 8.19 で、1.96 を超え、5%水準で有意だった。そこで、Tabachnick and Fidell (2007) に基づきマハラノビスの距離の値から多変量外れ値を持つ受験者 1 人(24.215,  $p_1 = .001$ ,  $p_2 = .099$ )を除いて再分析したが、結果は除外する前とほとんど変わらなかった。そのため、その受験者も除外せず 100 名のデータの結果を以下報告した。

モデル1を分析したところLEXATT タスク 2得点の誤差変数が負の分散を示したため、そこを0に固定して再分析した(図5.7参照)。

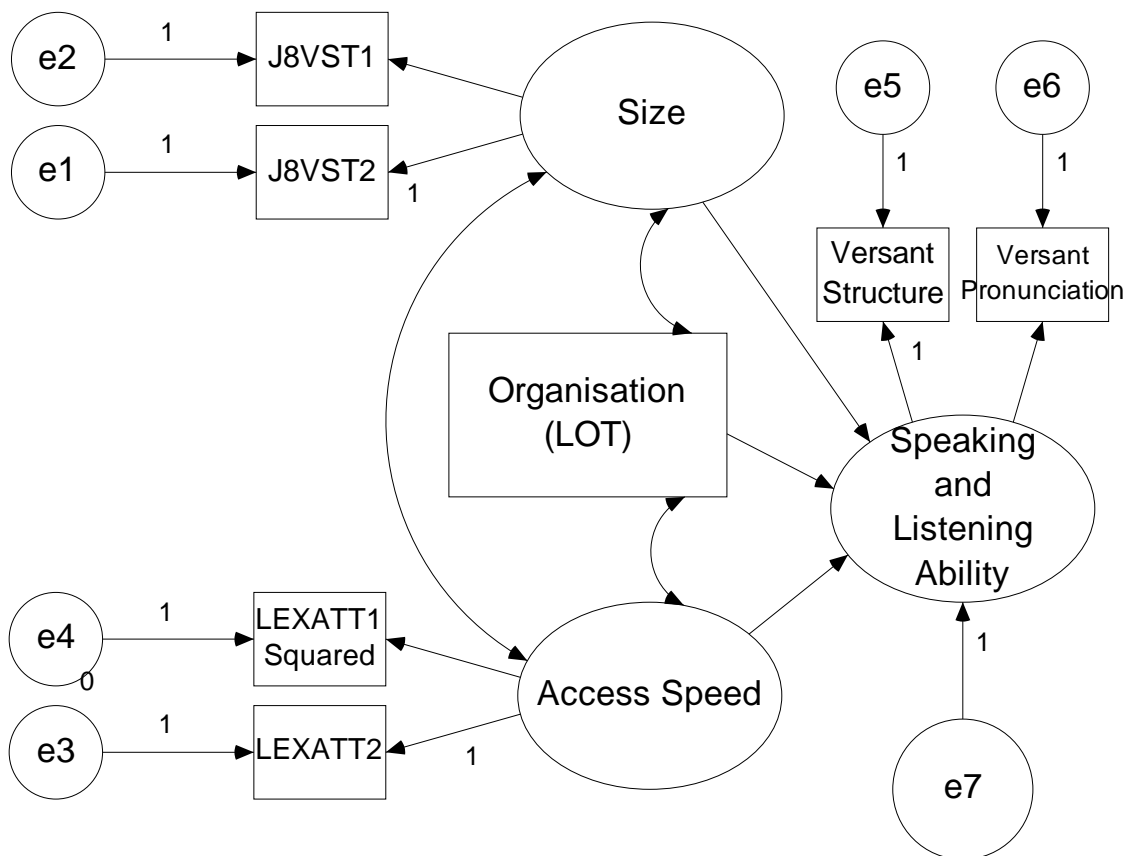


図 5.6. モデル 1 の分析前のパス図

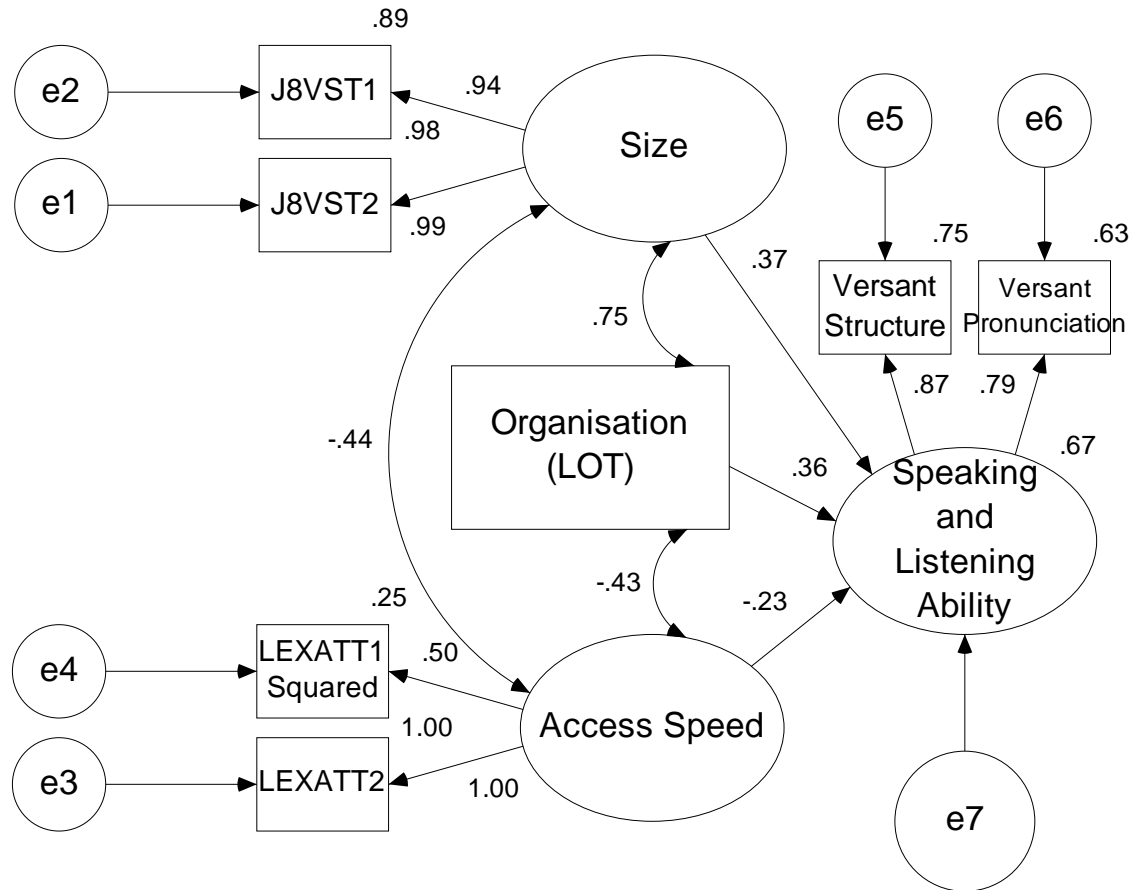


図5.7 モデル1の分析後のパス図:最終モデル

表 5.24 モデルの適合度の結果

	$\chi^2(df), p$	$\chi^2/df$	GFI	CFI	RMSEA (90%CI)	AIC	SRMR
基準	$p > .05$	$< 2.0$	$> .90$	$> .90$	$< .05$	小さいほど 良い	$< .08$
モデル 1	10.91 (10), .36	1.09	.97	1.00	.03 (.00 to .12)	46.91	.02

注:  $n = 100$ . GFI = Goodness of Fit Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; CI = 信頼区間(Confidence Interval); AIC = Akaike Information Criterion; SRMR = Standardized Root Mean Square Residual. 基準は豊田(2007), Arbuckle and Wothke (1995), Hu and Bentler (1999)を参照。SRMRはPlugins→Standardized RMRでダイアログボックスを開き、開いたままで分析をすると値が算出される。

表5.24を見ると、RMSEAの90%信頼区間の上限以外の基準が満たされており、今回のデータはモデルに適合したと考えた。モデル1におけるすべてのパス係数は5%水準で有意であり、モデル1を最終モデルとして採択した。

3種類の語彙知識からのスピーキング・リスニング能力へのパス係数が有意に異なるかを、「差に対する検定

統計量」を出力し検討した。差に対する検定統計量は絶対値の 1.96 を超えると 5%水準で有意な違いがあるとされる(豊田, 2007)。その結果, サイズからのパス係数(A), 構成からのパス係数(B), アクセス速度からのパス係数(C)での組み合わせでは, AB が 0.57, AC が-3.18, BC が-3.11 であり, A と B は異なるとは言えず, A と C, B と C が異なっていた。サイズと構成からの予測は同程度(順に.37, .36)で, サイズと構成はアクセス速度より(-.23)スピーキング・リスニング力を予測できることが分かった。語彙知識が英語技能に影響すると考えるならば, 語彙サイズと構成はアクセス速度よりもスピーキング・リスニング力に影響が強いという結果となった。語彙サイズが大きく語彙構成が優れている人は, 基本的スピーキング・リスニング力が高い傾向がかなり見られたということである。また構成の寄与がサイズの寄与と同程度だったことで, 口頭でのやりとりでは語彙を多く持っているのと同様に, 語彙のネットワークを知っていることが重要になり, 語の構成知識の役割が大きいたことが示唆された。

一方, TOEIC と CASEC の分析では予測にほとんど貢献していなかったアクセス速度も, サイズ・構成ほど強くないものの, 統計的に有意にある程度貢献していた(表 5.25 参照)。アクセス速度が遅い人(LEXATT の値が大きい人)は, スピーキング・リスニング力が低いという傾向が少し見られたということである。一方 TOEIC の分析においては, アクセス速度から TOEIC リスニング・リーディング得点へパス係数を引かないモデルが採択された(第 5.1.1 節)。スピーキング・リスニングでは即時性が高く, 速さ・流暢さが求められるため, アクセス速度の寄与がより大きく重要になるのだろう。

表 5.25 TOEIC (SEM), CASEC (重回帰分析), Versant (SEM)の分析結果

	パス係数			$R^2$	信頼性
	J8VST	LOT	LEXATT	全体	3種の語彙テスト
TOEIC リスニング	.66	--	.00	.44	.90, .31, .74
TOEIC リーディング	.70	--	.00	.49	.90, .31, .74
CASEC	.70	.08	-.19, -.03	.64	.94, .56, .81
Versant	.37	.36	-.23	.67	.96, .77, .81

SEM を用いた Versant の予測では, 3 種類の語彙からスピーキング・リスニング力が 67%(図 5.7 の「スピーキング・リスニング力」の円の右肩の値 .67 参照)を予測できていた。この割合は TOEIC の 40%台(44%, 49%)よりは高かった。結果が異なった理由の 1 つは対象技能が違うことだろうが, 参加者の受験に対する動機づけが異なることも違いの要因にかもしれない。本節の Versant の受験者は全員テストを受験することを自発的に申し出た者であるが, TOEIC の研究は授業の活動の中で行われた場合もあった。今後語彙テストを項目応答理論で分析し, 応答パターンがモデルに合わない受験者を特定して, 除外した後に再分析するなどの方法をとれば, 受験者要因の違いをよりなくす形で比較できるかもしれない。

さらに, 語彙サイズにおいては, J8VST の奇数問題の得点(J8VST1), 偶数問題の得点(J8VST2)への負荷が .94, .99 であり, 2 つの説明率が 89%, 98%であったことから, サイズを安定して測定できていると考えられる。語彙サイズテストの構造的な妥当性(3.2.2.1 節参照)の証拠の 1 つと考えてよいだろう。また, アクセス速度においては, LEXATT タスク 1 得点, タスク 2 得点への負荷が .50, 1.00 であり, 2 つの説明率が 25%,

100%であることから、アクセス速度にタスク 2 得点が貢献する割合が高いことが分かった。タスク 2 は単語の意味が分かるまでの速さを測る問題であり、スピーキング・リスニング力との関係でアクセス速度を考える際には、意味の要素を入れることが必要なのかもしれない。

なお、Versant 得点の予測率は高いものの、6 割強であり、語彙知識以外の要素も関わっていることも確認しておきたい。

#### 5.3.4 まとめ

本節では、3 種類の語彙知識テストと Versant 得点がどのように関係しているかを吟味した。RQ1 の「3 つの語彙テストは、Versant 得点をどの程度予測できるか」については、共分散構造分析(SEM)の結果で解釈すると、J8VST と LOT と LEXATT により、Versant 得点の 67%を予測できており、この説明率は高いものだった。RQ2 の「どの語彙テストが、Versant 得点をより予測できるか」については、J8VST と LOT が同程度にかなり予測できており、LEXATT はそれよりは寄与が少ないものの、有意にある程度の予測ができていた。

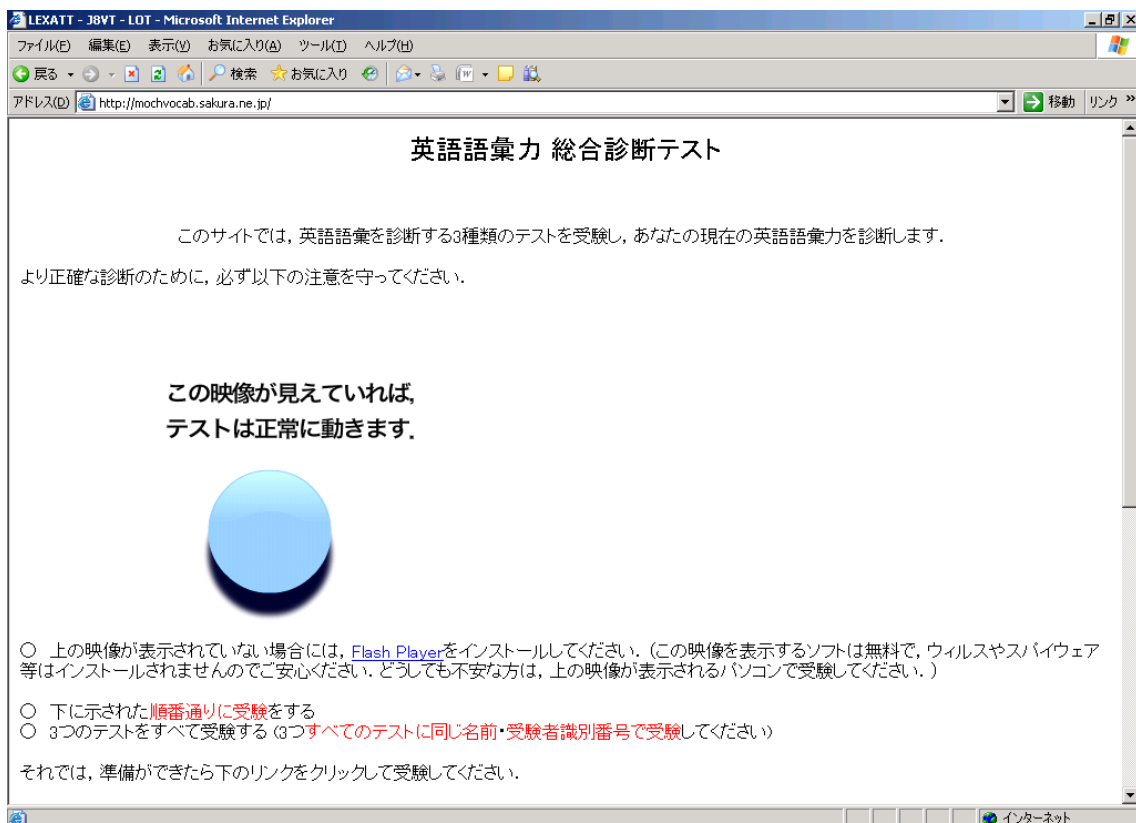
3 つの語彙知識はそれぞれスピーキング・リスニング力の予測に貢献することが示されたことで、語彙知識を捉えるときには、サイズだけでなく構成・アクセス速度の側面を入れるべきだと考えられる。3 側面で語彙知識を捉えることが予測力を高め、語彙研究の幅を広げることにもなる。別な受験者やテスト形式でさらに調べることで、語彙知識を増やすことが直接スピーキング能力を高めるという因果関係が想定できるのかについては、今後さらに調べる必要があるだろう。

## 第6章 結論

本研究は、平成19年度から3年間にわたり、語彙サイズ、語彙構成、アクセス速度という3つの次元から語彙知識を測定することにより、英語習熟度の推定を試みてきた。語彙サイズテストとして、『大学英語教育学会基本語リスト』(JACET8000)に基づく J8VST、語彙構成テストとしては、The Lexical Organisation Test(LOT, 望月, 2006)、アクセス速度のテストとしては、3年間の基盤研究で開発した The Lexical Access Time Test(LEXATT)を利用した。平成19年度は、3つのテストの妥当性を他の語彙テストとの相関を求める併存的妥当性によって検証した。平成20年度は、3つのテストと TOEIC, CASEC, VERSANT などの標準化された英語能力試験とを大学生に実施し、重回帰分析、共分散構造分析により、どの程度説明が可能かを吟味した。3つの開発した3つのテストは、ウェブ上で公開し、次のURLで利用可能である。

<http://mochvocab.sakura.ne.jp/>

このページにアクセスすると図 6.1 のような画面が現れる。



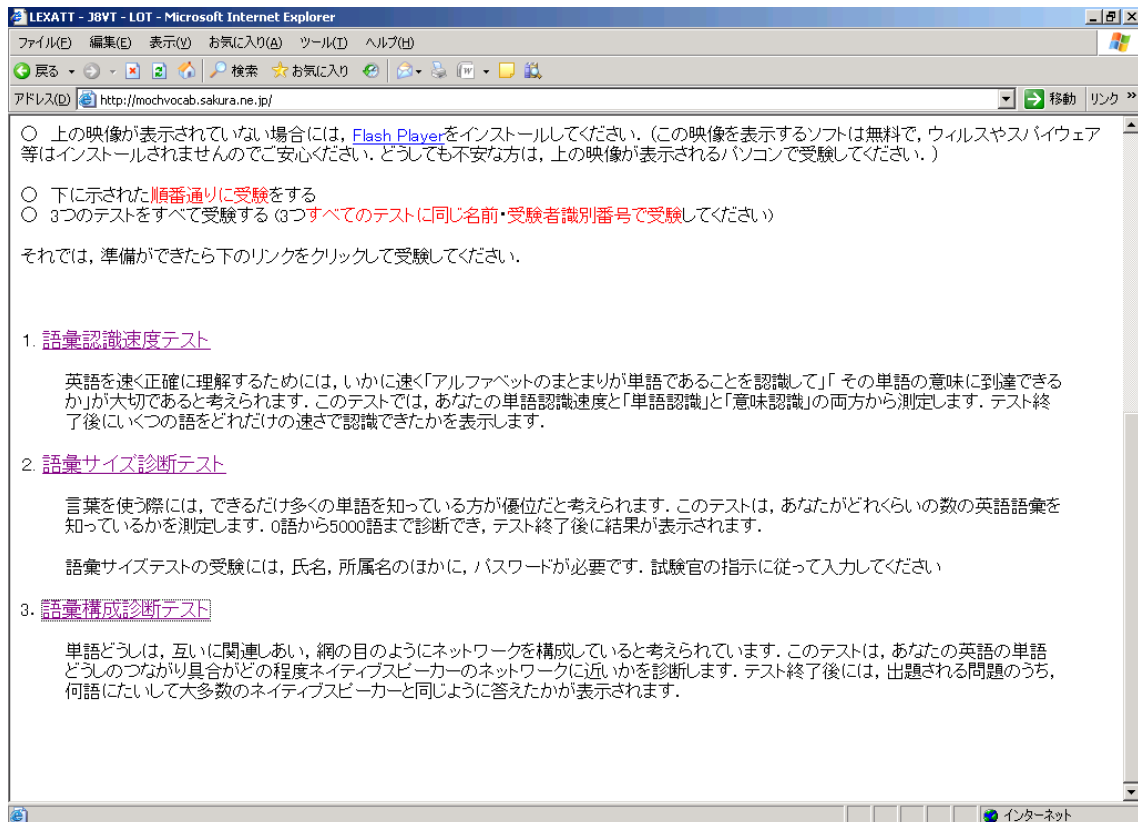


図 6.1 語彙テストウェブサイトページ

1 の語彙認識速度テストをクリックすると図 6.2 の画面が現れる。

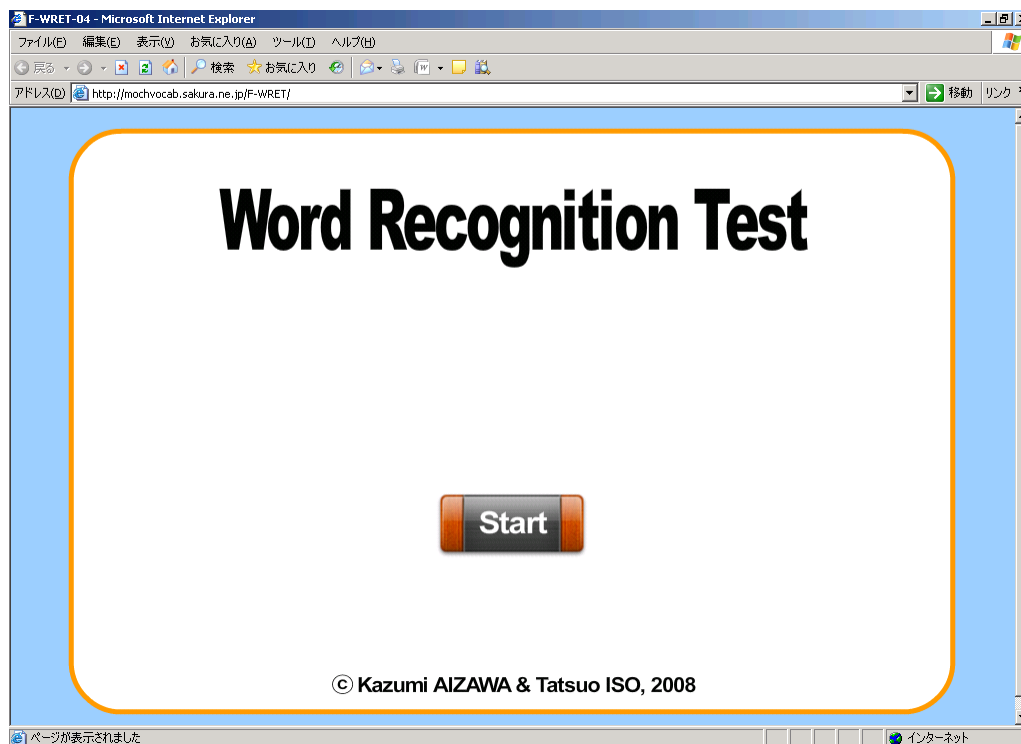
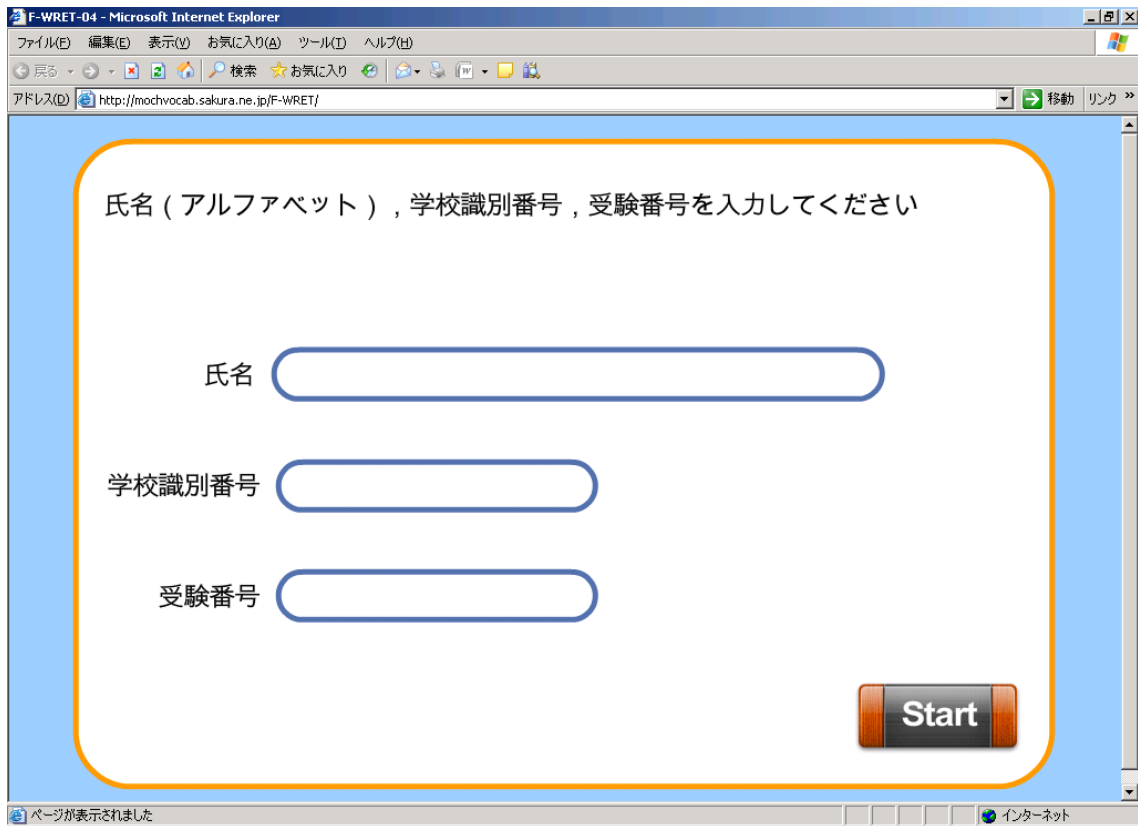


図 6.2 語彙認識速度テスト

「Start」 ボタンを押すと図 6.3 画面が現れる。



The screenshot shows a Microsoft Internet Explorer window titled "F-WRET-04 - Microsoft Internet Explorer". The address bar contains the URL "http://mochvocab.sakura.ne.jp/F-WRET/". The main content area has a light blue background with a white rounded rectangle containing the following text and form elements:

氏名 (アルファベット) , 学校識別番号 , 受験番号を入力してください

氏名

学校識別番号

受験番号

Start

At the bottom of the browser window, there is a status bar with the text "ページが表示されました" and "インターネット".

図 6.3 語彙認識速度テスト ID 入力画面

この画面では氏名をアルファベット, 学校識別番号には 3 ケタの数字, 受験番号は 10 桁の数字を入力する。そのあと, 「Start」 をクリックすると, テストが開始する。

図 6.1 の画面の「2. 語彙サイズ診断テスト」をクリックすると図 6.4 の画面が現れる。

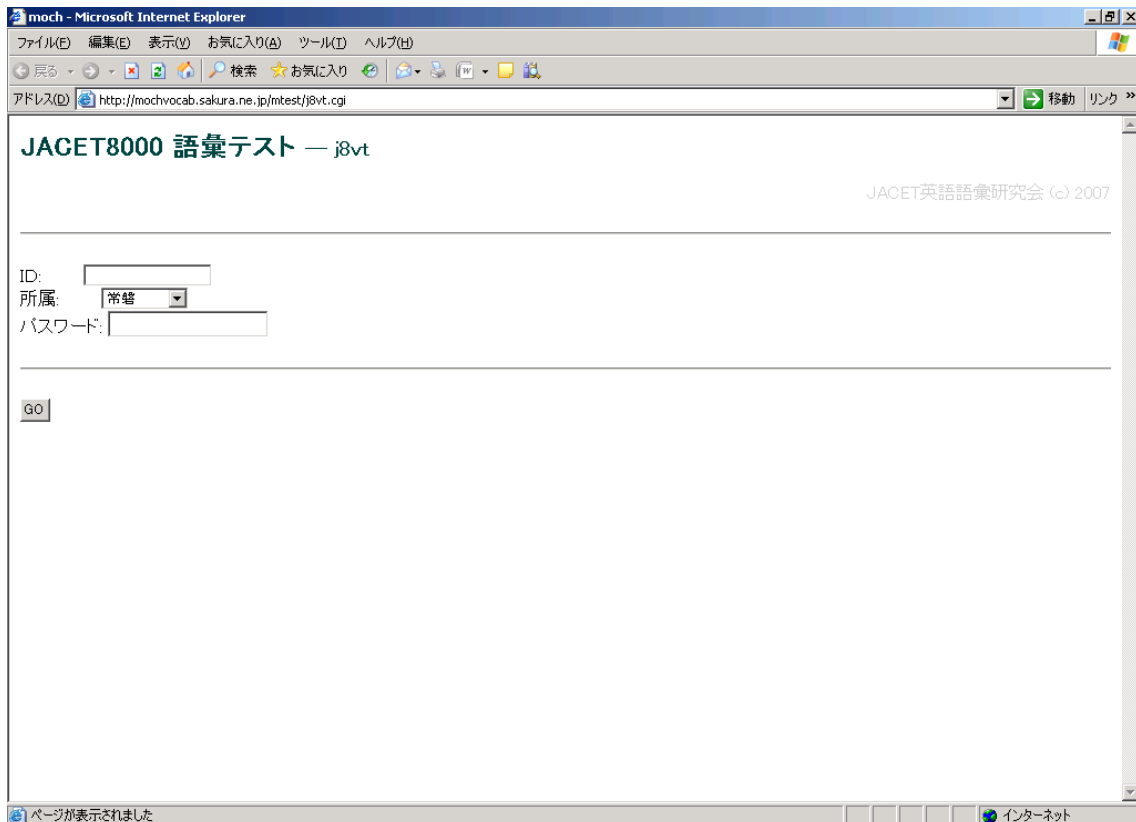


図 6.4 語彙サイズ測定テスト ID 入力画面

この画面では、ID に 10 桁の数字を入力，所属は下矢印をクリックし一番下の「安城」を選択，パスワードは「a」を入力する。そのあと、「GO」をクリックすると，テストが始まる。

図 6.1 の画面の「3. 語彙構成診断テスト」をクリックすると図 6.5 の画面が現れる。

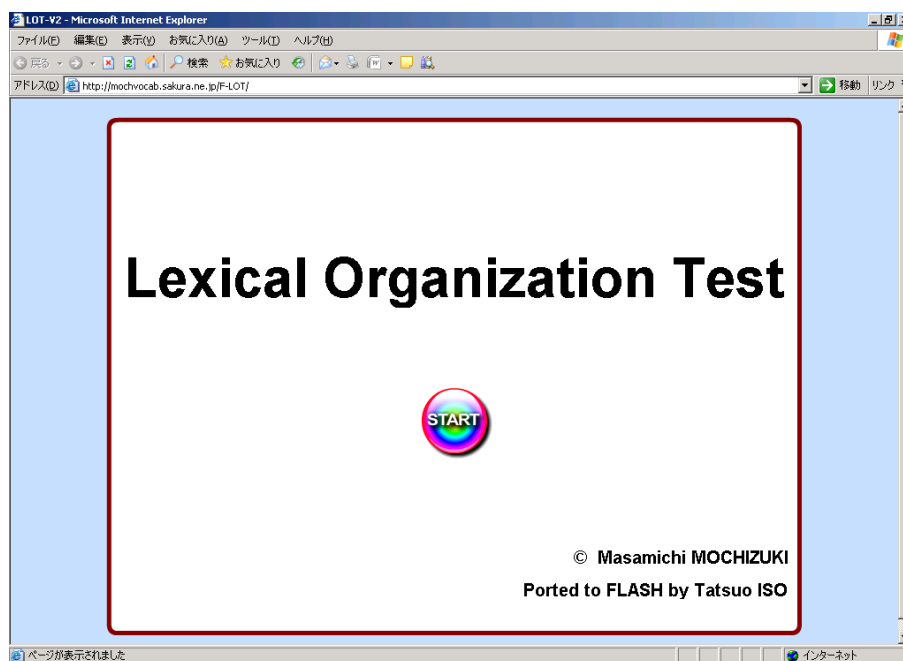


図 6.5 語彙構成診断テスト



「Start」 ボタンを押すと図 6.6 画面が現れる。

The screenshot shows a Microsoft Internet Explorer window with the address bar displaying 'http://mochvocab.sakura.ne.jp/F-LOT/'. The main content area has a light blue background and contains a white rectangular form with a red border. At the top of the form, the text reads '氏名 (アルファベット), 学校識別番号, 受験番号を入力してください'. Below this text are three input fields, each with a label to its left: '氏名', '学校識別番号', and '受験番号'. At the bottom right of the form is a 'Start' button with a gradient background and a shadow effect. The browser's status bar at the bottom shows 'ページが表示されました' and 'インターネット'.

図 6.6 語彙構成診断テスト ID 入力画面

この画面では氏名をアルファベット, 学校識別番号には 3 ケタの数字, 受験番号は 10 桁の数字を入力する。そのあと, 「Start」 をクリックすると, テストが開始する。

3 つの語彙テストの結果から, TOEIC の総合得点を予測する重回帰分析による回帰式は次のようになる。

$$\text{TOEIC 得点} = -358.46 + 6.67 \times (\text{J8VST 得点}) + 8.42 \times (\text{LOT 得点}) + 0.00 \times (\text{LEXATT Part 1 得点}) + (-0.02) \times (\text{LEXATT Part 2 得点})$$

その回帰式に J8VST, LOT, LEXATT の平均値をそれぞれ代入すると, TOEIC 得点 430.20 が予測できる。しかしながら, 95% の信頼区間は, -93.58 から 953.99 と, TOEIC の得点範囲をほぼすべて網羅するもので, 正確な予測がなされているとは言い難い。

共分散構造分析のひとつである構造方程式モデリングによる分析では, TOEIC 得点を説明する 2 つのモデルを検討した結果, 図 5.3 に示したモデルが採用された。このモデルでは, 語彙構成のテスト LOT は信頼性係数  $\alpha$  が .31 と低いために, 分析に加えていない。このモデルは, 構造方程式モデリングにより最適に適合したモデルを示したものである。図 5.3 が示すように, 語彙サイズとアクセス速度の間には, -.42 という中程度の負の関係が見られ, 語彙サイズが大きい学習者は, アクセス速度が速いという傾向が中程度あるという結果が得られた。しかしながら, アクセス速度は TOEIC リスニング・リーディングのいずれとも関係を持たず, 語彙サイズのみが TOEIC 得点を説明する語彙の側面となった。

このように TOEIC を予測する分析では、LOT の信頼性が低い点、また、アクセス速度は TOEIC 得点の予測にほとんど貢献しないという結果が得られた。これは CASEC を使用した研究でも同様の結果が得られた。

しかしながら、リスニングとスピーキング力を測定する VERSANT を用いた研究では、3つの語彙テストの信頼性係数  $\alpha$  は、J8VST(.96)、LOT(.77)、LEXATT(.81)と高い。構造方程式モデリングによる分析では、VERSANT の得点は、3つの語彙テストで合計 .59 が説明できる。パス係数は、それぞれ J8VST(.37)、LOT(.36)、LEXATT(-.23)である。これは3つの語彙テストが VERSANT の得点をバランスよく説明していることを示唆している。これもこの研究成果のひとつといえる。

本研究の目的は、語彙力を3つの次元から測定し、それを統合することによって、総合的な英語力及び技能別の英語力を予測することであった。TOEIC や CASEC の得点を総合的な英語力と考えた場合、それを予測するには3種類の語彙テストは必要なく、語彙サイズテストのみで十分に予測可能であるという結果を得た。これを説明する原因として3つの要因が挙げられる。

第1の要因として、語彙構成テスト LOT の信頼性の低さが挙げられる。LOT が信頼性の高いテストだったならば、構造方程式モデリングの分析に加えることができ、語彙サイズに加えて、TOEIC の得点を予測することができたかもしれない。

第2の要因として、被験者の語彙レベルの低さが挙げられるかもしれない。語彙サイズと構成の関係について、Meara (1996)は、語彙サイズが小さい場合、語彙構成は重要ではなく、語彙サイズのみが重要であると述べている。Meara によれば、英語の場合、小さい語彙サイズとは 5000 語から 6000 語である。これに従うと、この科研の被験者は語彙サイズが 5000 語以下と考えられるために、語彙サイズのみが重要である学習者であったかもしれない。そのため、語彙構成のテストの信頼性が高く分析に使用できたとしても、語彙サイズのみが英語総合力の予測に関わっている結果に違いはなかったかもしれない。

第3の要因として、アクセス速度のテストの妥当性の問題が挙げられる。前科研では語彙アクセス速度を測定するテストとして、ふつう認知速度の測定に用いるスーパーラボのような高価なソフトウェアを用いずに、実施できるプログラムの開発を目指した。しかしながら、無秩序な文字列に隠された高頻度語を見つけ出すという課題と提示された単語を認知し、その意味を2択で選ばせるという課題が真にアクセス速度を測定しているのかという妥当性は十分に検証されてはいない。

このように本研究は、語彙テストから英語能力の推定する方法を明らかにしたが、同時に3つの問題点も明らかになった。すなわち、語彙構成テスト LOT の信頼性が低いという問題点、アクセス速度テストの妥当性検証が不十分という問題点、語彙テストから英語能力を推定する場合信頼区間が広く、厳密な推定ができないという問題点である。今後、このような問題点を解決し、より正確な英語習熟度の推定ができる語彙テストを開発する研究が望まれる。

## 謝辞

本研究は、数多くの人たちの協力によって可能になりました。開発したテストを受験してくれた学生のみなさん、英語母語話者の方々、心より御礼申し上げます。特に、第3章に関しては、プロトコル分析では、コーディング等にお手伝いいただいた星野由子氏(東京富士大学)に感謝申し上げます。第5章に関しては、Versantを無料でご提供くださった Pearson Education 社と、分析方法などの助言をくださった印南洋氏(豊橋技術科学大学)に感謝申し上げます。

## 参考文献

- Adams, M. L. (1980). Five cooccurring factors in speaking proficiency. In J. R. Firth (Ed.), *Measuring spoken language proficiency* (pp.1-6). Washington, DC: Georgetown University Press.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp.77-117). Newark, DE: International Reading Association.
- Arbuckle, J. L., & Wothke, W. (1995). *Amos 4 user's guide*. Chicago: SmallWaters Corporation.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1-34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 26(3), 1178-1191.
- Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effect in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24, 89-106.
- Bonk, W.J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14, 14-31.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language™*. New York: Routledge.
- Cobb, T. (n.d.). Word Associates Test. Retrieved November 13, 2008 from <http://www.lexutor.ca/tests/associates/>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, A. D., & Upton, T. A. (2007). 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24, 209-250.
- Coulson, D. (2005, December). Recognition speed for basic L2 vocabulary. A paper read at JACET English Vocabulary Group Second Conference on Researching, Learning and Teaching Second Language Vocabulary, Chuo University Korakuen Campus.

- Council of Europe (2001). *The common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Curriculum, Evaluation and Management (CEM) Centre, Durham University. (2008). Effect size calculator. Retrieved November 15, 2008 from <http://www.cemcentre.org/renderpage.asp?linkID=30325017>
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Dávid, G. (2007). Investigating the performance of alternative types of grammar items. *Language Testing*, 24, 65-97.
- Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M., & Van Moere, A. (2008). Evaluation of the usefulness of the Versant for English Test: A Response. *Language Assessment Quarterly*, 5, 160-167.
- Educational Testing Service. (2009). 「TOEIC テスト結果について」. Retrieved May 10, 2009, from <http://www.toeic.or.jp/toeic/about/score/>
- ETS -2009 User manual Criterion online writing evaluation V. 9.1
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3:00 A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. Retrieved October 7, 2008, from <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>
- Grabe, W., & Stoller, F. (2002). *Teaching and researching reading: Applied linguistics in action*. New York: Longman.
- Greidanus, T., Bogaards, P., van der Linden, E., Nienhuis, L., & de Wolf, T. (2004). The construction and validation of a deep word knowledge test for advanced learners of French. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp.191-208). Amsterdam, the Netherlands: John Benjamins.
- Greidanus, T., Beks, B., & Wakely, R. (2005). Testing the development of French word knowledge by advanced Dutch- and English-speaking learners and native speakers. *The Modern Language Journal*, 89, 221-233.
- Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal*, 85, 567-577.
- Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing*, 3, 158-185.
- Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education*. Cambridge, UK:

Cambridge University Press.

- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21, 303-317.
- Henriksen, B., Albrechtsen, D., & Haastrup, K. (2004). The relationship between vocabulary size and reading comprehension in the L2. *Angles on the English Speaking World*, 4, 129-140.  
<http://criterion28.ets.org/cwe/News/CriterionUserManual.pdf>  
<http://www.lexically.net/downloads/version5/HTML/index.html>  
<http://www.lexically.net/wordsmith/>
- Hoshino, Y. (2009). Reactivity of thinking aloud in EFL learners' performance in vocabulary tests. *JLTA (Japan Language Testing Association) Journal*, 12, 16-25.
- Hu, L. -T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- In'nami, Y. (2005). The effects of task types on listening test performance: A retrospective study. *JLTA (Japan Language Testing Association) Journal*, 8, 1-20.
- Iso, T., & Aizawa, K. (2008, December). Mochizuki Kaken Project Report (3): Development of an online word recognition test: Examination of the internal validity. Paper presented at the 1st JACET Joint Research Forum on English Vocabulary and Reading, Kwansei Gakuin University, Hyogo, Japan.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement (4th ed., pp.17-64)*. Westport, CT: American Council on Education and Praeger.
- Katagiri, K. (2001). Developing the ten-minute vocabulary tests for quick and approximate estimates of general English ability of Japanese EFL learners. Unpublished PhD dissertation, Tokyo Gakugei University, Japan.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- Koda, K. (1989). The Effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign Language Annals*, 22, 529-40.
- Koda, K. (2005). *Insights into second language reading*. Cambridge: Cambridge University Press.
- Koizumi, R. (2005). *Relationships between productive vocabulary knowledge and speaking performance of Japanese learners of English at the novice level*. Unpublished Ph.D. dissertation, University of Tsukuba, Japan. Retrieved November 21, 2008, from <http://www.tulips.tsukuba.ac.jp/limedia/dlam/B25/B2599596/1.pdf>
- Koizumi, R. (2005). Predicting speaking ability from vocabulary knowledge. *JLTA (Japan Language Testing Association) Journal*, 7, 1-20.
- Koizumi, R. (2008, December). Mochizuki Kaken Project Report (4): Relationships among three types of vocabulary tests and speaking test. Paper presented at the 1st JACET Joint Research

Forum on English Vocabulary and Reading, Kwansei Gakuin University, Hyogo, Japan.

- Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174.
- Kunnan, A. J. (1998). An introduction to structural equation modelling for language assessment research. *Language Testing*, *15*, 295-332.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaud, J. L., & H. Béjoin, H. (Eds.), *Vocabulary and applied linguistics* (pp.126-132). London: Macmillan.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, *21*, 202-226.
- Laufer, B. & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*, 33-51.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, *54*, 399-436.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, *26*, 35-57.
- McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 14-24.
- McCann, R. S., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 693-706.
- Meara, P. (1992). *EFL vocabulary tests*. Centre for Applied Language Studies. University College Swansea.
- Meara, P. (1996). The dimension of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp.35-53). Cambridge: Cambridge University Press.
- Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, *4*, 142-154.
- Meara, P., & Jones, G. (1990). *Eurocentre vocabulary size test*. Zurich: Eurocentres Learning Service.
- Meara, P., & Wolter, B. (2004). V\_Links: Beyond vocabulary depth. *Angles on the English-speaking world*, *4*, 85-96.
- Messick, S. A. (1989). Validity. In R. L. Linn (Ed.). *Educational measurement* (3rd ed., pp.13-103). New York: National Council on Measurement in Education/American Council on Education.

- Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*, 241-256.
- Mochizuki, M. (2006). An attempt to create a word recognition test: WRET. Paper presented at the JACET English Vocabulary Research Group 3rd Conference on Researching, Learning & Teaching Second Language Vocabulary, Chuo University, Tokyo.
- Mochizuki, M. (2007). Validation of a test of lexical organization. A paper read at the Special Forum on Vocabulary Testing, December 1, 2007, Reitaku University Tokyo Research Center.
- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: an exploratory study. *System, 28*, 291-304.
- Morimoto, Y. (2007). Test-taking processes of vocabulary tests in context from the perspective of think-aloud analysis. *JLTA (Japan Language Testing Association) Journal, 10*, 68-87.
- Nation, P. (1990). *Teaching and learning vocabulary*. Wadsworth, OH: Heinle and Heinle.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston, MA: Heinle, Cengage Learning.
- Nikolov, M. (2006). Test-taking strategies of -12 and 13-year-old Hungarian learners of EFL: Why whales have migraines. *Language Learning, 56*, 1-51.
- Paribakht, T.S., & Marjorie, B.W. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*, 9-27.
- Paul, P. V., Stallman, A. C., & O'Rourke, J. P. (1990). Using three test formats to assess good and poor readers' word knowledge. Technical Report No. 509 Center for the Study of Reading, University of Illinois at Urbana-Champaign. (ERIC Document Reproduction Service No. ED322493)
- Pearson Education. (2009a). Versant™. Retrieved November 29, 2009, from <http://www.versant.jp/>
- Pearson Education. (2009b). Versant Results. Retrieved November 29, 2009, from [http://www.versant.jp/test\\_result.html](http://www.versant.jp/test_result.html)
- Potter, M. C., So, K., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23-38.
- Qian, D.D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review, 56*, 282-307.
- Qian, D.D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning, 52*, 513-536.
- Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge for assessing reading performance. *Language Testing, 21*, 28-52.



- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*, 355-371.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp.41-60). Mahwah, NJ: Lawrence Erlbaum Associates.
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209-227). Amsterdam, the Netherlands: John Benjamins.
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon, England: Multilingual Matters.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23*, 441-474.
- Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing, 17*, 85-114.
- Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing, 25*, 211-236.
- Scott, M. (2007). *WordSmith tools version 5*
- Statistical Package for the Social Sciences (SPSS). (2003). SPSS (Version 12.0.1) [Computer software]. Chicago: SPSS Inc.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. Needham Heights, MA: Allyn & Bacon.
- Taft, M. (1991). *Reading and the mental lexicon*. Hove, U.K.: Lawrence Erlbaum Associates Ltd.
- Tanaka, M., & Laufer, J. (2003). *Vocabulary levels test for Japanese learners of English*. Hiroshima University Information Media Center
- Wang, M., & Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning, 55*, 71-98.
- Werner, C., & Schermelleh-Engel, K. (2009). Deciding between competing models: Chi-square difference tests. Retrieved November 28, 2009, from [http://user.uni-frankfurt.de/~cswerner/sem/chisquare\\_diff\\_en.pdf](http://user.uni-frankfurt.de/~cswerner/sem/chisquare_diff_en.pdf)
- Wesche, M., & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review, 53*, 13-40.
- Zhang, L.J., & Annual, S.B. (2008). The Role of vocabulary in reading comprehension: The case of secondary school students learning English in Singapore. *RELC Journal, 39*, 51-76.
- 池田央. (1992). 『テストの科学：試験にかかわるすべての人に』. 東京：日本文化科学社.

- 石川慎一郎. (2003). 「妥当性の高い語彙力テストの形式について：JACET List of 8000 Basic Words 準拠語彙力テスト開発に向けた基礎的研究」. 『言語文化論叢』(千葉大学外国語センター), 12, 67-80.
- 石川慎一郎. (2005). 「大学生英語学習者の受容語彙力と発表語彙力の関係：語彙サイズテストおよびエッセイ・コーパス分析に基づくアプローチ」. 『中部地区英語教育学会紀要』 34, 337-344.
- 狩野裕 & 三浦麻子. (2002). 『Amos, EQS, CALIS によるグラフィカル多変量解析—目で見る共分散構造分析—(増補版)』. 京都：現代数学社.
- 芋坂満里子. (2002). 『ワーキングメモリー脳のメモ帳』. 東京：新曜社.
- 小泉利恵. (2008a). 「EDiT 文法診断テストの妥当性」. 金谷憲 (監修), 『新しい英語文法診断テストの提案』 (pp. 43-53). 東京：NPO 法人英語運用能力評価協会(ELPA).
- 小泉利恵. (2008b). 「より適切なテスト得点の解釈と使用を目指して：妥当性と妥当性検証法」. 『大学英語教育学会関東支部 2007 年度研究年報』, 4, 38-41. 東京：大学英語教育学会関東支部 (研究年報委員会).
- 御領謙. (1998). 「心内辞書観念から見た単語認知の問題」. 芋阪直行 (編), 『読み-脳と心の情報処理』 (pp. 133-150). 東京：朝倉書店.
- 静哲人. (2002). 『英語テスト作成の達人マニュアル』. 東京：大修館書店.
- 島田めぐみ. (2006). 「日本語聴解テストにおいて選択肢提示形式が解答過程に与える影響—発話思考法を用いて—」. *JLTA (Japan Language Testing Association) Journal*, 9, 101-116.
- 大学英語教育学会基本語改訂委員会(編). (2003). 『大学英語教育学会基本語リスト：JACET List of 8000 Basic Words』. 東京：大学英語教育学会.
- 投野由紀夫(編著). (1997). 『英語語彙習得論：ボキャブラリー学習を科学する』. 東京：河源社.
- 豊田秀樹 (編). (2007). 『共分散構造分析 [Amos 編]』. 東京書店.
- 野上康子. (2009). 「コンピュータ適応型テスト CASEC における項目の長期使用の影響について」. 『日本テスト学会誌』, 5, 145-164.
- 星野由子. (2008, December). 「発話プロトコルの有無が語彙テストのパフォーマンスに及ぼす影響」. 第 1 回 JACET リーディング研究会・英語語彙研究会合同大会. 関西学院大学.
- 水本篤. (n.d.). 「カ ッ パ 係 数」. Retrieved November 21, 2008 from <http://www.mizumot.com/stats/kappa.htm>
- 水本篤 & 竹内理. (2008). 「研究論文における効果量の報告のために—基礎的概念と注意点—」. 『関西英語教育学会紀要「英語教育研究」』, 31, 57-66.
- 望月正道. (1998). 「日本人英語学習者のための語彙サイズテスト」. 『財団法人語学教育研究所紀要』 12, 27-53.
- 望月正道. (2006, August). 「新しい語彙構成テスト開発の試み」. 第 32 回全国英語教育学会高知研究大会自由研究発表. 高知大学.
- 望月正道. (2007). JACET8000 語彙テストの開発(事例研究 1) 『大規模コーパスを用いた日本人英語

学習者用の語彙リスト構築と教材分析システムの開発:平成16~18年度科学研究費補助金基盤研究(B)研究成果報告書(研究課題番号16320076, 研究代表者 相澤一美)』(pp.134-145).

望月正道. (2008, September). 「語彙構成のテストの開発と改良:項目分析による検証」第47回大学英語教育学会全国大会シンポジウム「語彙知識の3次元的アプローチとその問題点」. 早稲田大学.

## 付録 A: J8VST の問題, 選択肢, 正解

例:001;2;mention;話に出す, 言及する, 言及, ;manipulate, mention, minister, moderate  
例の数字や単語はつぎのことを表す。

001 は, 問題番号

2 は, 正解 (この場合, manipulate, mention, minister, moderate の 2 番目である mention が正解であることを示す)

mention は, この項目の目標語

話に出す, 言及する, 言及は, 和英選択方式の問題

manipulate, mention, minister, moderate は, 和英選択方式の選択肢

001;2;mention;話に出す, 言及する, 言及, ;manipulate, mention, minister, moderate

002;1;war;戦争, ;war, week, wife, word

003;4;reply;返事をする, 返事, ;receive, relay, repeat, reply

004;3;office;事務所, 会社, ;object, occasion, office, owner

005;2;tell;言う, 教える, ;take, tell, think, try

006;2;white;白い, 白, ;wake, white, wink, write

007;2;nation;国家, 国民, ;name, nation, notice, notion

008;4;cross;十字架, 横切る, 交差させる, ;class, close, course, cross

009;4;color;色, 色づく, ;caller, career, collar, color

010;1;wide;広い, 広範な, ;wide, wild, wise, wit

011;3;wish;願う, 願い, ;wash, walk, wish, wonder

012;4;dry;乾いた, 乾かす, ;daunt, dead, dink, dry

013;4;such;そのような, ととても... な, ;same, several, so, such

014;2;boat;ボート, 小舟, 船, ;board, boat, boot, bought

015;2;forest;森林, ;farm, forest, form, fort

016;3;being;存在, 生き物, ;bean, beast, being, brain

017;4;sit;座る, ;set, shot, shit, sit

018;1;least;最も少ない, ;least, less, little, lit

019;1;glass;ガラス, グラス, ;glass, gloss, grass, gross

020;1;want;欲する, したい, 欠乏, ;want, weep, whip, win

021;2;very;とても, 大変, ;vary, very, vet, volley

022;3;week;週, ;waste, weak, week, west

023;3;south;南, ;sauce, self, south, space

024;2;century;世紀, ;cent, century, cloth, coast

025;2;so;そんなに,とても,それで,;second, so, some, such  
 026;3;pale;青白い,薄い色の,;pail, plain, pale, poor  
 027;2;phrase;語句,;passage, phrase, praise, pronoun  
 028;3;context;文脈,;composition, content, context, count  
 029;2;closely;接近して,近くで,;clearly, closely, counter, crushingly  
 030;3;pole;棒,柱,極,;pick, pile, pole, poll  
 031;3;giant;巨大な,巨人,;gang, genius, giant, god  
 032;4;fuel;燃料,;fill, flu, fossil, fuel  
 033;2;lecture;講義,;leather, lecture, legend, literature,  
 034;2;philosophy;哲学,;philology, philosophy, physiology, psychology  
 035;1;partner;共同出資の経営者,伴侶,;partner, patron, person, professor  
 036;2;declare;宣言する,申告する,;decade, declare, demand, deny  
 037;4;formal;正式な,公式な,;fair, female, foreign, formal  
 038;3;media;マスメディア,マスコミ,;mass, mathematic, media, multiple  
 039;1;aim;狙う,ねらい,;aim, assign, attend, avoid  
 040;3;code;記号,規範,暗号化する,;category, clone, code, cord  
 041;3;fault;落ち度,故障,欠点,;fail, false, fault, fool  
 042;3;plate;皿,板,;pan, plane, plate, pound  
 043;1;fairly;公平に,公正に,;fairly, flaringly, fearfully, freely  
 044;2;forever;永久に,永遠に,;flight, forever, forgive, fulfillment  
 045;4;truck;トラック,貨物自動車,;tractor, trailer, trouser, truck  
 046;2;desert;砂漠,見捨てる,;draw, desert, desire, dessert  
 047;4;surround;取り囲む,;signal, solve, spread, surround  
 048;4;traveler;旅行者,;target, trapper, task, traveler  
 049;3;gray;灰色の,灰色,;gay, glue, gray, grim  
 050;2;relation;関係,関連,;regulation, relation, relative, ratio  
 051;1;presidential;大統領の,社長の,;presidential, prior, professed, prudent  
 052;4;sexual;性の,;sensual, sensible, sensitive, sexual  
 053;2;officially;公式に,;objectively, officially, operationally, outrageously  
 054;1;subsidy;補助金,助成金,;subsidy, substitute, supply, support  
 055;3;controversy;論争,;chase, coalition, controversy, conversation  
 056;2;increasing;ますます増加する,;including, increasing, invisible, innovative  
 057;2;abolish;廃止する,;abandon, abolish, adapt, assert  
 058;2;deposit;預金,預金する,;depart, deposit, dispend, distrust  
 059;3;sleeve;そで,;scarf, shirt, sleeve, sore  
 060;3;attendance;出席,付き添い,;absence, assignment, attendance, attention

061;3;headquarters;本部, 本社, ;hatred, heaven, headquarters, honor  
 062;2;observer;観察者, 傍聴者, ;objector, observer, opportunity, opposite  
 063;2;chapter;章, ;caption, chapter, circle, counter  
 064;3;overlook;見渡す, 大目に見る, ;observe, organize, overlook, oversee  
 065;4;isolation;孤立, ;incentive, indent, infant, isolation  
 066;1;empirical;経験上の, 経験主義の, ;empirical, evident, experimental, explicit  
 067;1;decorate;装飾する, 飾る, ;decorate, defend, design, descend  
 068;1;therapy;治療法, ;therapy, treat, trauma, tryout  
 069;2;recommendation;推薦, ;reckon, recommendation, replace, reserve  
 070;2;personnel;社員, 人事課, ;peer, personnel, pirate, pupil  
 071;4;loyalty;忠誠, 忠誠心, ;lateral, lawful, liberal, loyalty  
 072;2;allowance;手当, 引当金, ;account, allowance, assistance, attempt  
 073;4;thinking;思考, ;temperament, testament, theory, thinking  
 074;4;moving;動いている, 引っ越しの, ;metric, mortal, motive, moving  
 075;1;sector;部門, 分野, ;sector, shape, statement, substance  
 076;1;landowner;地主, ;landowner, lender, liar, lord  
 077;2;plow;すき, 耕す, ;pave, plow, pour, prop  
 078;1;ache;痛み, 痛む, ;ache, acorn, arch, art  
 079;1;hammer;金づち, ;hammer, hangar, hummer, hunger  
 080;4;supporting;助演の, ;secondary, senior, supervise, supporting  
 081;3;lordship;君主, 貴族の地位, ;launch, librarian, lordship, loyalty  
 082;2;immense;巨大な, 多大な, ;illegal, immense, infamous, irregular  
 083;3;intake;摂取, 取り入れ, ;include, influx, intake, invoke  
 084;4;fog;霧, ;fear, feat, fill, fog  
 085;3;biology;生物学, ;bibliography, bioethics, biology, biography  
 086;3;pudding;プリン, ;pillar, protein, pudding, purity  
 087;2;betray;裏切る, ;beat, betray, board, brave  
 088;2;viewer;見物人, ;victor, viewer, virtue, visitor  
 089;1;lesbian;女性同性愛者, ;lesbian, lethal, lexicon, lunatic  
 090;4;outward;外へ, 外側へ, ;oral, outdrive, outlive, outward  
 091;1;cleaner;掃除機, 洗浄剤, ;cleaner, container, contributor, cylinder  
 092;1;collaboration;共同, 協力, ;collaboration, command, concession, corporation  
 093;2;receptor;受容体, 受信機, ;reception, receptor, response, retrieval  
 094;3;toxic;有毒の, ;tax, threat, toxic, treat  
 095;4;dwelling;住まい, ;deploy, destine, diva, dwelling  
 096;3;yarn;紡ぎ糸, ;yak, yard, yarn, yield

097;1;damn;呪う, ;damn, destroy, disclose, disport  
 098;3;bitterly;苦々しく, ;beadily, belatedly, bitterly, busily  
 099;2;planner;立案者, ;pioneer, planner, plowman, presenter  
 100;3;exploitation;開発, 搾取, ;emergency, evolution, exploitation, exposure  
 101;2;vault;金庫室, 地下納骨室, 丸天井, ;vale, vault, vent, vintage  
 102;3;coil;巻いたもの, 1 巻き, ;cent, coal, coil, circuit  
 103;4;conspicuous;人目をひく, 顕著な, ;circumstance, component, condiment, conspicuous  
 104;4;bust;胸像, 壊す, 逮捕する, ;beast, berth, breath, bust  
 105;2;allied;同盟している, ;achieved, allied, assigned, associated  
 106;1;simplify;単純にする, ;simplify, specify, sprinkle, suspect  
 107;3;melody;旋律, ;meadow, median, melody, menace  
 108;3;livestock;家畜, ;lack, launch, livestock, lottery  
 109;2;supposedly;たぶん, おそらく, ;spiritedly, supposedly, supply, suspiciously  
 110;2;wary;用心深い, 慎重な, ;wander, wary, winder, wise  
 111;4;triple;3 倍の, ;third, thrift, treble, triple  
 112;4;purse;財布, ハンドバック, ;pulse, pullet, purchase, purse  
 113;3;crouch;かがむ, ;crash, crime, crouch, cue  
 114;2;arena;競技場, 舞台, ;academy, arena, athlet, audience  
 115;1;facade;見せかけ, 正面, ;facade, face, fake, false  
 116;2;militant;好戦的な, ;method, militant, missionary, motivated  
 117;4;groom;花婿, ;glaze, grate, grave, groom  
 118;3;disgusting;気分の悪くなる, 実にはいやな, ;depressed, disguise, disgusting, dispel  
 119;2;hen;めんどり, ;hail, hen, honor, horde  
 120;1;beneficiary;受益者, 受取人, ;beneficiary, berserker, bouncer, browser  
 121;1;manipulation;操作, ;manipulation, manual, message, movement  
 122;1;vacuum;真空, 空白状態, ;vacuum, vale, vent, volcano  
 123;4;attempted;未遂の, ;accused, animated, assisted, attempted  
 124;2;demolish;破壊する, 取り壊す, ;depress, demolish, depose, dread  
 125;2;rot;腐る, 腐敗, くだらない, ;rift, rot, rubbish, ruin

## 付録 B: LOT の問題・正解

右端の数字は正解を表す。

- 12 は, 1 番目と 2 番目の組み合わせが正解
- 23 は, 2 番目と 3 番目の組み合わせが正解
- 13 は, 1 番目と 3 番目の組み合わせが正解。

00 は練習問題を表す。

例:00 football, manager, match, :13 では,  
この項目は練習問題であり, football match が正解

- 00 football, manager, match, :13
- 00 dark, horse, mouth, :12
- 00 jump, ship, space, :23
- 01 blue, smoke, thin, :12
- 02 believe, marriage, save, :23
- 03 advice, order, seek, :13
- 04 blood, cell, floor, :12
- 05 evidence, hard, present, :12
- 06 nuclear, plant, winter, :12
- 07 offer, place, take, :23
- 08 beat, egg, heart, :13
- 09 feature, film, history, :12
- 10 express, fly, train, :13
- 11 power, office, station, :13
- 12 case, king, test, :13
- 13 remain, silent, steady, :12
- 14 lose, sense, sight, :13
- 15 bed, look, river, :13
- 16 follow, suit, wish, :12
- 17 board, game, shop, :12
- 18 kiss, life, ring, :12
- 19 poor, quality, record, :12
- 20 expert, opinion, team, :12



- 21 hit, run, wave, :12
- 22 lay, stone, table, :13
- 23 news, science, story, :13
- 24 point, rate, view, :13
- 25 food, time, waste, :23
- 26 family, structure, value, :13
- 27 call, nurse, student, :23
- 28 head, nod, step, :12
- 29 disease, prevent, water, :12
- 30 back, chair, empty, :23
- 31 future, personal, relationship, :23
- 32 expect, popular, vote, :23
- 33 common, factor, practice, :13
- 34 condition, item, sale, :23
- 35 deal, new, tax, :12
- 36 hall, town, wood, :12
- 37 dream, holiday, special, :12
- 38 beach, private, school, :23
- 39 lack, process, understanding, :13
- 40 centre, trade, union, :23
- 41 break, clean, street, :12
- 42 high, promotion, season, :13
- 43 chance, company, enjoy, :23
- 44 career, international, path, :13
- 45 free, part, speech, :13
- 46 air, base, wall, :12
- 47 cut, price, supply, :12
- 48 car, heavy, rain, :23
- 49 attack, launch, lead, :12
- 50 approach, fresh, particular, :12

## 付録C: WRECT

### Phase 1

1	huusegaizegdpavdwcpmmqrdt	19	xqxzourruataalkvbxmeasyhbn
2	ucsbvchurchzihuwrongztgqh	20	shortksomeholpgteuraisefs
3	wbutwftcylschooltqsizeiov	21	laopgiaeffectekjiubgwdaud
4	iherncsayoworldhqhfamily	22	gfreeuwzovgsszjurulelgjyt
5	wlseeyxikhgiuijgebvcsomog	23	resultnlcpywynendblqobage
6	lrgvazbbjxdkfewxycwlightz	24	mswidexworryzylastpchance
7	pointuoetypezutxhaffectnw	25	omxxhhptfarfbuvavueofamuv
8	wamkoagreeeolotlsurewhile	26	yforvlefbwbpdifnshouldxb
9	cohgnfnubigdlvwbemgkqqhw	27	kmljfmajormbiwgirluadulta
10	nqviroadizqwfjwddguessmhp	28	nsuchzcowluhjszdtbkzufce
11	mhnsqqpdanceahalfnfsprit	29	jkduthrowkcaimageuzjukczs
12	jfutureysocialcyetxpeople	30	bedepxgmibmuchxucouldgfmu
13	uabzrdjumpoysqyofqsozikuo	31	rnbmdnjyiestrongiuhnpyzsl
14	twqbcguidexeuscommonqux	32	fthmissraeamlnatureolrglw
15	rnpwaitdijppyanswercofly	33	jobqvaddnuexpectwyaizasmz
16	whilezlquiethcounterjudge	34	vsahowodagecmeaniznaturez
17	sqqzkunderuvzsjuhkeqmqcin	35	qehwouldkorderhnavbovejls
18	hmbteazhaecizcyjrhjgnqown		

### Phase 2

1	family	mother	sister	uncle	grandmother	aunt	father	grandfather
2	clothes	shirt	skirt	suit	sweater	dress	pants	blouse
3	sports	football	basketball	baseball	hockey	bowling	soccer	tennis
4	instruments	piano	violin	horn	guitar	trumpet	trombone	clarinet
5	time	day	hour	minute	year	month	second	week
6	city	building	street	mayor	museum	school	factory	store
7	movement	run	walk	march	stroll	hike	step	stalk
8	things	book	watch	bottle	desk	machine	ruler	shoes
9	house	floor	wall	ceiling	window	corner	room	chimney
10	job	cleaner	writer	teacher	driver	builder	clark	teller
11	weather	rain	sunlight	cloud	hail	heat	mist	fog

12 color black blue brown red orange yellow purple  
 13 speech lisp exclaim murmur shout whine stutter whisper  
 14 play actor stage curtain lights script actress scene  
 15 food bread lettuce hotdog steak beans lemon peach  
 16 subjects science math history social studies arts biology geography  
 17 animal horse dog lion pig elephant spider cat  
 18 number nine sixteen four seventy-seven fifteen eighty thirty-two  
 19 kitchentable chairs refrigerator oven microwave stove sink  
 20 furniture table chair chest bed bookcase desk sofa  
 21 drink milk cola coffee cocoa orange juice water tea  
 22 travel subway metro bus airplane train boat ship  
 23 car wheels engine windows seats doors breaks tires  
 24 water river lake ocean sea well pond swimming pool  
 25 classroom teacher students books pens notebooks pencils desks  
 26 hospital doctor nurse beds patients medicine shots  
 x-rays  
 27 feeling happy sad angry nervous proud excited bored  
 28 computer CD-ROM keyboard hard drive monitor  
 mouse software floppy disk  
 29 Country China Japan Canada Germany France Kenya  
 Jordan  
 30 body chest arm neck face head leg finger  
 31 music jazz classical folk rock blues opera hip-hop  
 32 landscape beach hill mountain island desert valley plain  
 33 insects beetle ladybug butterfly dragonfly caterpillar  
 grasshopper firefly  
 34 fruit peach apple orange grape pear bananamelon  
 35 face eyes nose ears lips eyebrows cheeks chin

## Phase 3

1 mus\_ t p 6 wi\_e d t  
 2 re\_l a u 7 fa\_e c p  
 3 l\_ve o e 8 pa\_\_t r k  
 4 f\_nd i y 9 wi\_\_h t p  
 5 \_old h p 10 w\_\_at h s

11	__ore	m	z	24	sta__e	t	b
12	__ake	m	v	25	brea__	k	t
13	t__en	h	r	26	__eason	r	l
14	wa__t	n	p	27	f__llow	o	u
15	__ood	g	l	28	re__ordc	t	
16	__ight	r	j	29	re__ainm	p	
17	gre__t	a	o	30	repo__tr	n	
18	pla__e	c	k	31	poli__e	c	p
19	th__se	o	u	32	p__ayer	l	h
20	__eave	l	t	33	lea__er	d	t
21	__ight	m	k	34	__uffer	s	j
22	s__eak	p	d	35	li__ely	k	d
23	mo__thn		s				

## 付録 D: LEXATT

### Task 1 の文字列

1	tripiwasco	24	aiwiseqale
2	oimealrike	25	iumailbost
3	celpliftei	26	pakedietue
4	azaveefilm	27	fopyedgeeo
5	ihugehoce	28	kisceastoe
6	softujidee	29	bireviewoi
7	busyenusho	30	steckingu
8	fairopakea	31	tydejoinae
9	taskaqooti	32	suipwineea
10	riskobanqu	33	mongvoteiu
11	cookugacke	34	tormgameai
12	itemecacko	35	nindideaio
13	yardosaibu	36	ueathaonce
14	poemozecki	37	ewomeitime
15	calmishupo	38	oonryujust
16	uelockcoay	39	oawsouvery
17	eosnowvoon	40	imeamoneed
18	oaherobalt	41	uwreeamind
19	aidishnuck	42	ezastisign
20	iukicktwim	43	otulluwait
21	ueroofixon	44	acrueeplan
22	eopurebeed	45	ivackoloss
23	uabathtaik		

### Task 2

1	trip	旅行	わな
2	tail	しっぽ	青白い
3	kick	蹴る	拾う
4	snow	雪	知っている
5	tape	テープ	取る
6	play	遊ぶ	粘土
7	pool	プール	涼しい
8	case	入れ物	速度
9	past	過去	速い
10	meat	肉	会う
11	list	一覧表	最後の
12	date	日付	資料
13	bear	クマ	親愛なる…
14	tale	物語	しっぽ
15	ship	船	店
16	keep	保つ	深い
17	foot	足	食べ物
18	sell	売る	細胞
19	lift	左	持ち上げる
20	best	最高の	試験
21	rate	割合	遅い
22	rock	岩	鍵
23	heat	熱	座席
24	farm	農場	用紙
25	hope	希望	家
26	game	試合	門
27	moon	月	正午
28	wide	広い	果実酒
29	nice	米	良い
30	coat	上着	暗号
31	roof	屋根	植物の根
32	fish	魚	皿
33	leaf	葉っぱ	導く
34	poor	貧しい	注ぐ
35	fall	落ちる	失敗する
36	time	時間	種類

37	than	～より	その時
38	well	良い	意志
39	tall	背が高い	言う
40	live	生きる	好む
41	same	同じ	いくつかの
42	part	一部分	港
43	wife	妻	生活
44	king	王様	親切な
45	rule	規則	役割

技能別及び総合的英語能力を推定する語彙テストの開発：  
サイズ、構成、認知速度の融合  
Development of a Vocabulary Test Battery Estimating English Skills and  
Proficiency:  
Integrating Vocabulary Size, Organization, and Access Speed

平成19-21年度科学研究費補助金基盤研究（B）研究成果報告書  
研究課題番号 19320084

=====  
平成 22 年 3 月 20 日発行

発行者 科学研究費補助金基盤研究（B）

研究代表者 望月正道(麗澤大学)

〒277-8686 千葉県柏市光ヶ丘 2-1-1  
TEL 04-7173-3601

印刷所 昌美印刷株式会社

〒120-0005 東京都足立区綾瀬 2-26-7  
TEL 03-3690-3196

=====  
本書の一部あるいは全部を引用または複写製版する場合には、本書より引用したことをお断りください。