

KANSAI GAIDAI UNIVERSITY

大規模言語テストにおける項目解答行動の分析の試み：韓国語母語話者を対象として

メタデータ	言語: Japanese 出版者: 関西外国語大学・関西外国語大学短期大学部 公開日: 2016-09-05 キーワード (Ja): 大規模言語テスト, 項目分析, 項目反応理論, 名義反応モデル, 韓国語母語話者 キーワード (En): 作成者: 伊藤, 理英, 大澤, 公一 メールアドレス: 所属: 関西外国語大学, 山口大学アドミッションセンター
URL	https://doi.org/10.18956/00006043

大規模言語テストにおける項目解答行動の分析の試み

— 韓国語母語話者を対象として —

伊 藤 理 英

大 澤 公 一

要 旨

韓国語母語話者を対象として、テスト項目の解答行動に影響を与えている諸要因を探索するための分析を行った。大学修学能力試験 (CSAT)、日本留学試験 (EJU)、日本語能力試験 (JLPT) に出題された非音声領域の既出757項目によるモニター試験データ (n=4,647) の一部に対して項目反応理論 (IRT) の拡張モデルである名義反応モデルを適用し、テスト項目における各選択肢の期待選択確率が受験者の日本語能力のレベルによってどのように変化するかを数理統計学的に分析した。各選択肢の期待選択確率の推移が韓国語母語話者にとって特徴的、あるいは日本語教育学的に分析することが有効と思われたテスト項目を15題取り上げ、韓国語母語話者によるテスト解答行動であるという点に留意しながら、各テスト項目の特性を言語学あるいは日本語教育学の観点から定性的な考察を行った。

キーワード：大規模言語テスト、項目分析、項目反応理論、名義反応モデル、韓国語母語話者

1. 目 的

国際交流基金 (2012) によると2012年度日本語教育機関調査結果における韓国語母語話者の日本語学習者数は840,187人であり、前回2009年度調査時の964,014人より12.8%減少している。これは海外の日本語学習者数全体 (3,985,669人、前回2009年度調査時3,651,232人より9.2%増加) の21.1%を占めており、国別では中国 (26.3%)、インドネシア (21.9%) に次いで第3位の規模となっている。この3年間で学習者数は減少しているものの、大韓民国 (以下韓国) では大学修学能力試験 (College Scholastic Ability Test; 以下CSAT)、日本留学試験 (Examination for Japanese University Admission for International Students; 以下EJU)、日本語能力試験 (Japanese Language Proficiency Test; 以下JLPT) といった大規模公的試験に対する社会的需要が存在する状況は変わっていない。

大規模公的試験、特に言語テストにおいては測定対象となる言語能力以外の受験者の特性

(例：母語や文化的背景、教育課程の特徴など)によってテスト項目の正答率に不公平が発生しないよう作題時に入念な配慮がなされている。しかしながら、日本語能力の測定においては受験者の母語による有利不利の存在も指摘されており、例えば漢字圏の受験者は非漢字圏の受験者と比較して文字や語彙のテスト項目に対する正答率が高くなったり、特定の母語や日本語教育課程の特徴に起因する要因から特異な解答行動を取ることもあり得る。本研究の目的は、日本語学習者数の上位に位置している韓国語母語話者における、日本語能力を測定するテスト項目の解答行動に影響を与える諸要因について、大規模公的試験の既出項目を用いたモニター試験のデータ分析から探索的に考察を行うことである。

2. 方法

2.1 モニター試験

CSAT、改定前EJU、改定前JLPTに出題された非音声領域の既出757項目を用いたモニター試験を韓国国内で実施した。得られたデータ ($n=4,647$) の一部に対して項目反応理論 (Item Response Theory, IRT; Lord, 1952; Lord & Novick, 1968; Lord, 1980) の拡張モデルである名義反応モデル (Nominal Response Model, NRM; Bock, 1972) を適用し、テスト項目の各選択肢の期待選択確率が受験者の日本語能力レベルの高低によってどのように変化するのかを定量的に分析した。調査対象となったテスト項目は項目反応理論によって尺度化されること、また受験者にとって可能な限り露出の低い状態のものが望ましいことから、入手可能な既出項目より可能な限り過去の試験で出題されたもの (受験者の目に触れている可能性が低いと考えられるもの) で、かつ一問一答の小問形式のものを優先的に抽出した。具体的なモニター試験の実施デザインについては大澤 (2011) を参照のこと。

2.2 名義反応モデル

名義反応モデルでは多肢選択式のテスト項目 j に対する受験者 i の反応そのものを用いて受験者の潜在特性 (本研究では次元の日本語能力) 値 θ_i を推定する。名義反応モデルでは、テスト項目への反応を正誤パタンのような二値型 (正答 = 1、誤答 = 0) にも、順序性をもった評定値のような多値型 (例：優 = 4、良 = 3、可 = 2、不可 = 1) にも採点しない。あくまでも、各選択肢に対して反応 (選択) したかどうかという情報そのものを項目反応として利用するという部分に名義反応モデルの特徴がある。

名義反応モデルでは、潜在特性値 θ の受験者 i がテスト項目 j の選択肢 k ($k=1, 2, \dots, k, \dots, m_j$) を選ぶ確率 $P_{jk}(\theta_i)$ が以下のように定義される。

$$P_{jk}(\theta_i) = \frac{\exp\{b_{jk}^* + a_{jk}^*\theta_i\}}{\sum_{h=1}^{m_j} \exp\{b_{jh}^* + a_{jh}^*\theta_i\}}$$

名義反応モデルにおける項目パラメタ a_{jk} 、 b_{jk} はテスト項目 j における選択肢 k の特徴を定義する母数（パラメタ）であり、二母数ロジスティックモデルにおける項目識別力や項目困難度と同義ではない。本研究におけるモデル母数の推定には SSI 社の IRT 専用ソフトウェアである IRTPRO 2.1 (Cai, Thissen & du Toit, 2011) を用いた。

3 結果および考察

モニター試験757項目について名義反応モデルによる尺度化を行った。その結果得られたテスト項目の各選択肢についての期待選択確率を能力 θ の関数として表現する項目反応カテゴリ特性曲線 (Item Response Category Characteristic Curve; IRCCC) について、テスト項目に対する解答行動として特徴的な形状を示していると思われるテスト項目を15題 (CSAT 4題、JLPT 10題、EJU 1題) 取り上げ、項目内容と IRCCC を提示した。IRCCC の形状に基づいて、各テスト項目における解答行動の特徴について日本語学あるいは日本語教育学的な観点から定性的な項目分析を行った。

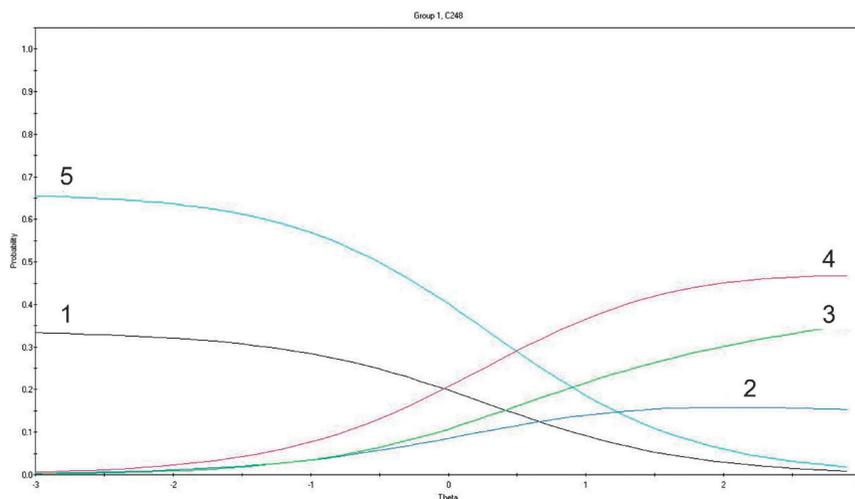
なお、IRCCC の解釈に際しては名義反応モデルの数学的な特性によりグラフの横軸として定義される日本語能力 θ のサイン (+-) に意味はなく、推定ソフトウェア IRTPRO 2.1 のデフォルトの出力結果をそのまま用いた。そのため、本研究においては横軸 (日本語能力軸) の左側 (-側) に行けば行くほど日本語能力が高い水準にあることを示しているので解釈の際に注意する必要がある。縦軸は名義反応モデルによって予測された各選択肢の期待選択確率 (0.0 ~ 1.0) を示しており、横軸の位置 (日本語能力の高低) に応じて各選択肢がどの程度選択されることが期待されるかを定義している。

3.1 大学修学能力試験 (CSAT)

[C248] 下線部に入る言葉を組み合わせて作られる言葉は何か。

- ・窓の 外を 見たら、にわの 花が もう 咲(a)そうでした。
- ・寒いから そとで 待(b)ない ようにと 言われました。

- ① 口 ② 国 ③ 車 ④ 歌 ⑤ 北



動詞の活用を問う問題である。誤答①の期待選択確率がCSAT受験者層の上位層にあっても30%を超えている。この結果より、CSAT受験者層の上位レベルにあっても、一部の受験者にとっては～ソウダという形式が<<伝聞>>であるか<<推量>>であるかを使い分けることが難しいということが分かる。つまり、正答できなかった受験者は、伝聞の～ソウダは現在形で用いられるという理解が不足しており、伝聞と推量の～ソウダの接続の仕方の違いが未習得であったと考えられる。本調査で、CSAT受験者層の上位レベルに至るまで、正答選択肢の期待選択確率の推移と並行した形で、正答選択肢のほぼ二分の一の確率で<<推量>>～ヨウダの用法を<<伝聞>>であると答えてしまったことから、CSAT受験者層の上位レベルでも約半数が～ソウダの用法について未習得あるいは習得過程にあることが分かる。

[C251] 下線部に入る言葉のあるだけ選べ。

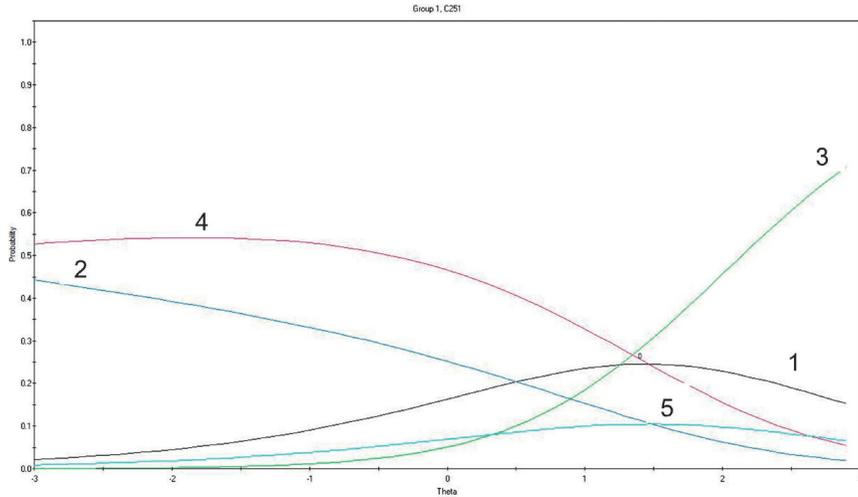
A: この 漢字、_____。

B: いいえ、知りません。

例: a. 知りますか b. 知って いますか c. 知って ありますか

① a ② b ③ c ④ a, b ⑤ a, b, c

大規模言語テストにおける項目解答行動の分析の試み



C251では、誤答であるa「知りますか」が正答であるb「知っていますか」と合わせて正答であると答える、つまり誤答選択肢④の期待選択確率が、正答選択肢②の期待選択確率をどの能力レベルでも上回っており、CSAT受験者の上位レベルに至るにつれて更にその値が上昇し、50%を超えている。これは、明らかに母語（韓国語）の干渉によるものであると考えられる。日本語の「知っていますか」に相当する韓国語が「알아요」であり、これを日本語に直訳すると「知りますか」となる。このように、結果の状態を表す、いわゆる瞬間動詞と呼ばれる日本語の動詞のテイル形の習得は、韓国語母語話者である学習者にとって、やはり難しいことが、本調査結果からも分かる。

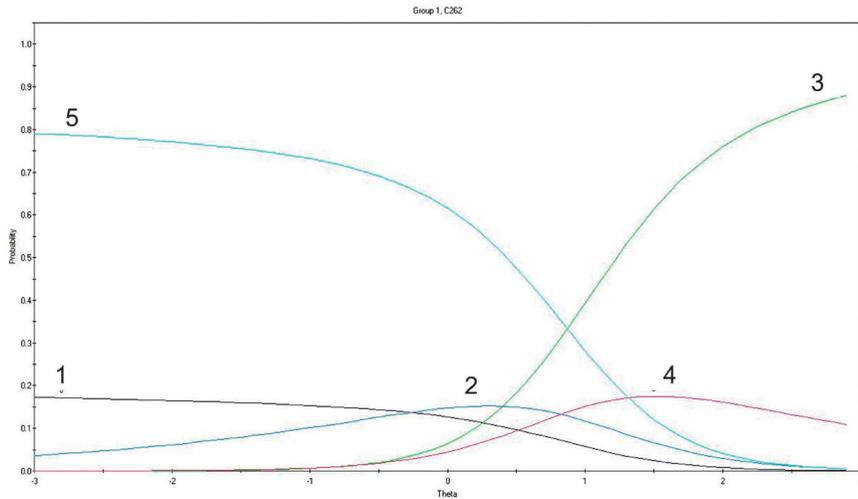
[C262] 下線部の言葉と意味が最も近いものは何か。

A: きのうの 映画、おもしろかったですか。

B: 思ったより よかったですよ。話は かわりますが、きょうは なんじに 帰りますか。

A: 6時ぐらいに なるよ と思いますが。

- ① しかし ② そして ③ あいにく ④ せっかく ⑤ ところで



C262では、《逆接》としての誤答選択肢①の期待選択確率が、日本語能力がCSAT下位レベルの段階から緩く上昇し続け、CSAT最上位レベルでも17%程度あることが分かる。考えられる誤答の原因としては、1.「ところで」の意味がわからなかった、2.「話は かわりませんが」という表現が理解できなかった、3.「話は かわりませんが」の《転換》機能を理解できず《逆接》としか思えなかったことが挙げられる。特に、《転換》機能を習得していれば、すぐに誤答選択肢①を選ぶこともなかったのではないだろうか。「ところで」の習得と合わせて、「話は かわりませんが」の《転換》用法も強化したい学習項目である。

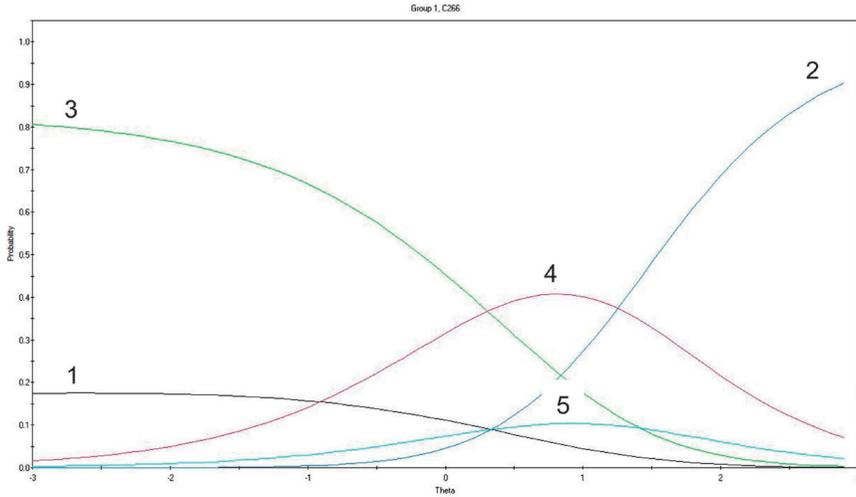
[C266] 下線部に入る言葉のあるだけ選べ。

A: 日本語で メールが_____か。

B: いいえ、まだ メールは ちょっと…。

例: a. 書けます b. 書きます c. 書く ことが できます

- ① a ② b ③ a、c ④ b、c ⑤ a、b、c



C266は可能動詞および可能表現～コトガデキルを答えられるだけでは正答できない問題である。文の《対象》を表す格助詞にガが使われているため、《対象》を表すガが重複することを避けるために～コトガデキルが使われないということに気が付かなければならないのである。例えば、問題文が「日本語で メールを _____ か。」であったら、a「書けます」だけでなくc「書くことができます」も正答になる。結果は、IRCCCが示す通り、a「書けます」とc「書くことができます」の両方を正解とする誤答選択肢③の期待選択確率が、CSAT受験者の最上位レベルでは80%を超えている。これは、シラバス外の内容が試されている問題であると言ってよいだろう。

3.2 日本語能力試験 (JLPT)

[J012] 空所補充

あの男は金をぬすんで、けいさつ _____。

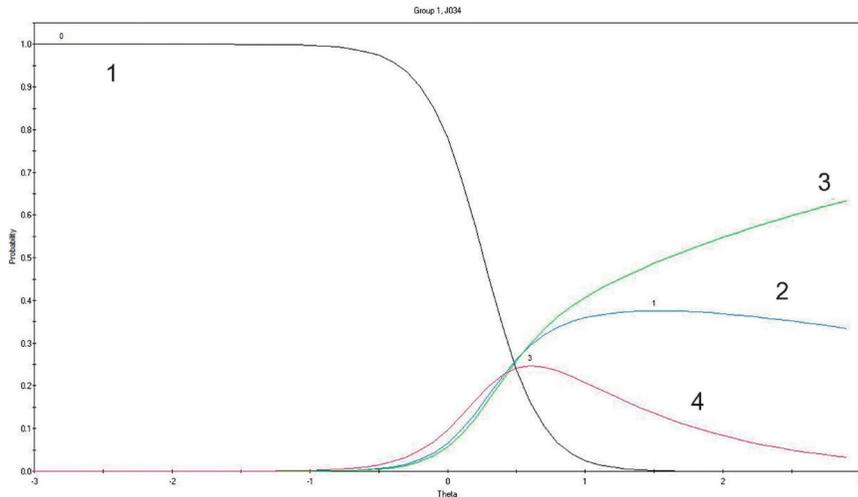
- ① につかまりました
- ② がつかまえてもらいました
- ③ をつかまえさせました
- ④ につかまえました

択確率が上位レベルでも40%程度ある。これは、主節の述語「乗れませんでした」がタ形であることから、従属節の述語について時制の一致を図ろうとしたものと考えられる。こちらも母語（韓国語）の干渉が原因である。本調査結果から、日本語における従属節の中のタについての理解および習得率が、韓国語母語話者の日本語学習者にとって低いということが示された。

[J034] 空所補充

いちねんは _____ です。

- ① じゅうにかげつ
- ② じゅうにしゅうかん
- ③ じゅうににち
- ④ じゅうにじかん

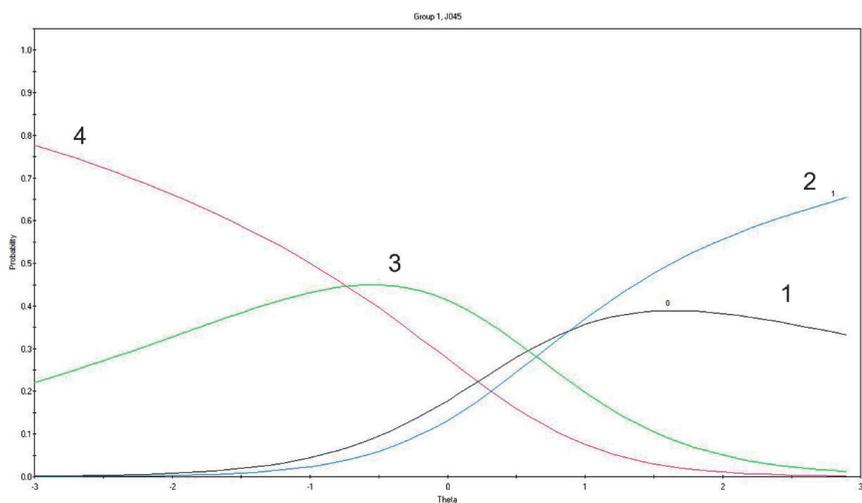


J034の結果をIRCCCで見ると、①から④の選択肢の語彙の習得順序が、学習順・難易度順に従っていることが分かる。受験者は未習得のものが選択肢の中にある場合、既習の項目を選ぶ傾向にあるということが分かる。

[J045] 空所補充

じどうしゃは 1台しか とまって _____ 。

- ① あります ② います
- ③ ありません ④ いません



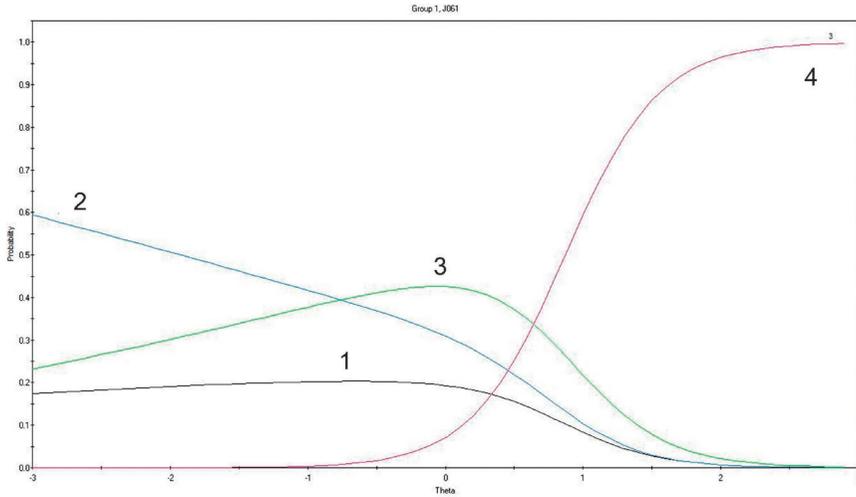
先ず、否定形と呼応するシカの用法については、初級レベル後半から理解しており定着しやすいことが分かる。次に、テイルとテアルの違いについての習得が難しいことが分かる。中級後半までは、テアルを含む誤選択肢③の期待選択確率が、テイルを含む正答④の期待選択確率を上回っており、イルとアルの用法との混合が起きていると考えられる。そして、上級レベルに至るにつれて、正答選択肢の期待選択確率が上昇し、誤答選択肢の期待選択確率が下降する様子がIRCCCからも見て取れる。しかし、上級レベルでもテアルを含む誤選択肢③「ありません」の期待選択確率が20%以上あり、テイルとテアルの用法が、上級レベルでも一部の学習者にとって習得困難であることが分かる。

[J061] 空所補充

わたしは 子どもの ころから おもちゃが すきで、いま _____ いろいろな おもちゃを もって います。

- ① にも ② でも ③ まで ④ から

大規模言語テストにおける項目解答行動の分析の試み

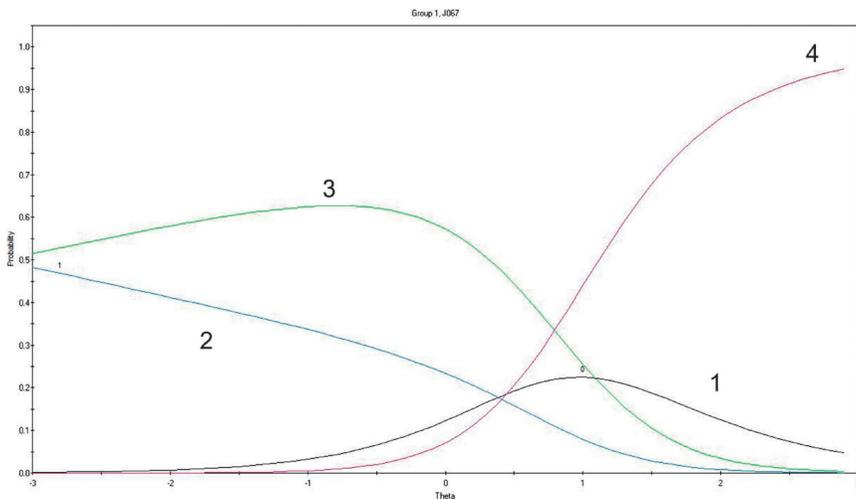


J061では、類似の意味をもった助詞および複合的な助詞の用法の習得が困難であることが示されている。上位レベルでも正答選択肢の期待選択確率は60%に満たない。まったく正反対の意味を表す助詞は初級レベルから排除できており、中上級レベルでは選択されていないことから、類似の助詞ニとデ、ニモとデモ、マデとマデニなどの違いに迷うことが分かる。

[J067] 空所補充

えきから タクシーを _____ きました。

- ① つかえて ② つかんで ③ つかって ④ つかいて



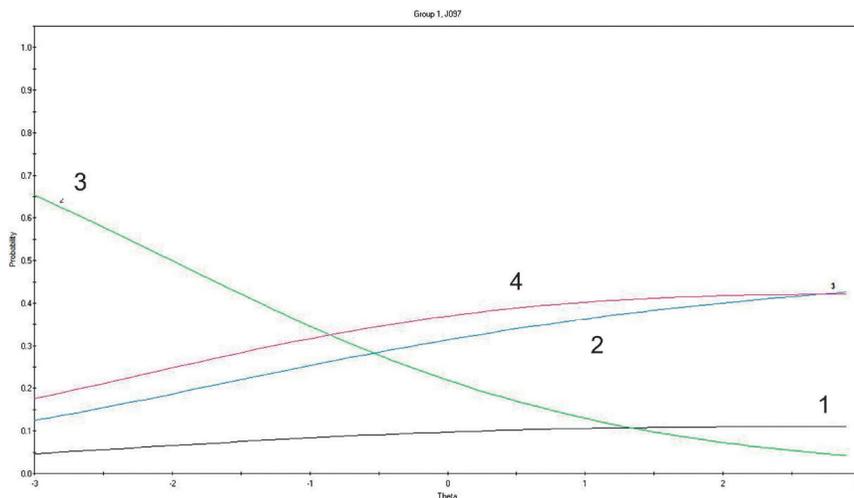
この問題J067では、上位層になればなるほど、「タクシーをつかう」よりも「タクシーをつかまえる」という連語を想起して、つかまえるからの連想でツカムのテ形である誤選択肢①ツ

カンデの期待選択確率が高くなったものと考えられる。動詞の活用の問題もあるが、上級者が知っている語彙や連語によるひっかけ問題には、能力が高ければ高いほど引っかかってしまうという結果が生じることを示している。

[J097] 空所補充

自動車の運転を覚えるには、_____の訓練が必要だ。

- ① 事実 ② 実験 ③ 現実 ④ 実地



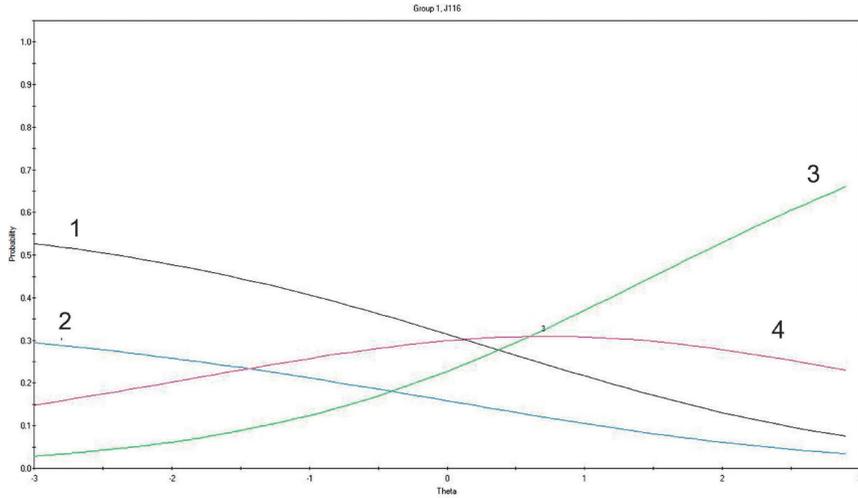
J097では、まず、上級レベルの学習者でも、「実地」という単語を習得している受験者が少なかったと考えられる。受験者は既習の他選択肢の中から、最も無理なく文脈に当てはまる誤答選択肢③「現実」を最も多く選択し、能力最上位層ではその期待選択確率が65%以上に及ぶ。要するに、この問題は難しすぎる問題であったために、識別力の低い問題になってしまったと考えられる。

[J116] 空所補充

午後3時ごろ、おたくで買い物をした者ですけど、買った品物をお店に忘れてきてしまったんです。ちょっとお調べ_____。

- ① いただきませんか ② 願えませんか
 ③ 見ていただきたいんです ④ もらえませんか

大規模言語テストにおける項目解答行動の分析の試み



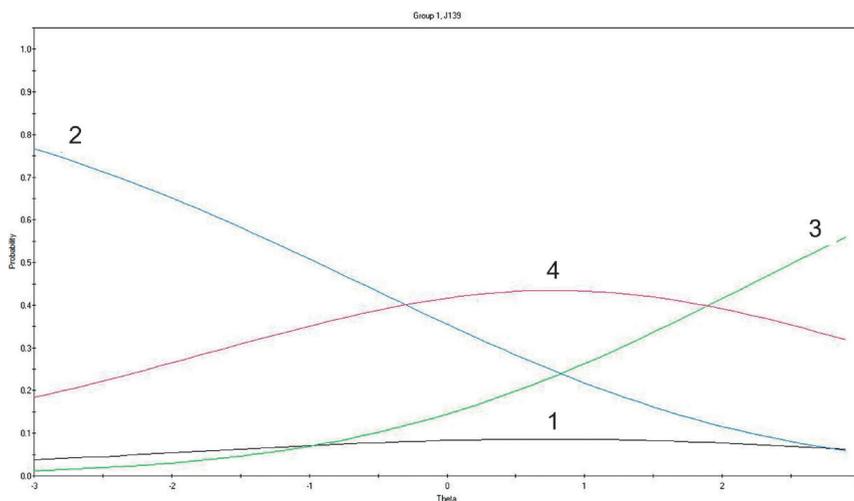
J116は敬語の問題であるが、まず、正答選択肢②の期待正答確率が能力最上位層にあっても約30%に過ぎず、動詞+願エマセンカという表現の習得率が低いことが伺える。次に、誤答選択肢①イタダキマセンカの期待選択確率が上級レベルに進むにつれて増えており、能力最上位層では誤答選択肢①の期待選択確率が50%を超えている。学習時にも気をつけたい敬語の～テイタダケマセンカと～テイタダキマセンカとの混合が上級者に見られることが明らかになった。

一方で、誤答選択肢④モラエマセンカの期待選択確率は中級レベルから徐々に下がっている。このことから、イタダクという語彙は定着しているものの、依頼表現である～テイタダケマセンカの定着は遅く、動詞+願エマセンカの習得は更に待たれることがわかる。習得順序としては自然な流れであるが、～テイタダケマセンカと～テイタダキマセンカの混合については、教育上、更に注意したい。動詞+願エマセンカとともに、敬語表現における習得率の低い項目として見直すことができる。

[J139] 類義語選択

たのまれた仕事を承知する

- | | |
|--------|---------|
| ① 取り込む | ② 受け取る |
| ③ 引き取る | ④ 引き受ける |



J139も、母語（韓国語）の干渉が原因となって誤答選択肢の期待選択確率が異常に伸びている顕著な例である。この問題の場合、「承知する」という語を理解できるようになった日本語能力レベルから、一気に誤答選択肢②「受け取る」の期待選択確率が倍増する。これは明らかに、日本語の「引き受ける」という意味の韓国語「일을 받아들이다」を直訳すると「受け取る」になるためであると思われる。

[J155] 読解

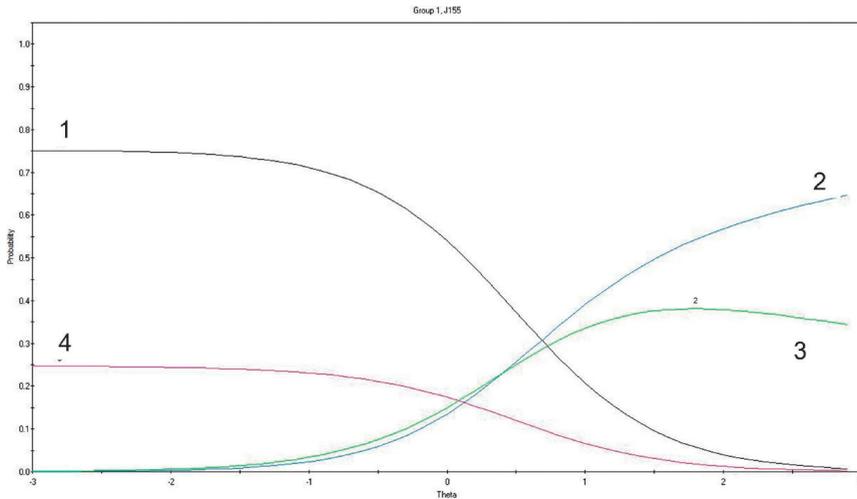
日本には世界平均の約2倍近い雨が降っている。ところが、国の面積が小さく人口が多いため、国民一人当たりの雨の量は世界平均の4分の1程度しかない。

実際たくさんの雨が降っても、日本には流れが急な川が多いので、その雨水はすぐに海へ流れ出てしまう。また、雨は梅雨や台風の季節などに集中して降り、それ以外の時期はあまり降らないし、雨の量は地域によってもかなり差がある。

このように考えると、()。日本人は水をもっと大切な資源として使っていくべきなのではないだろうか。

()にはどんな文が入るか。

- ① 日本は決して水が豊かにある国だとは言えない
- ② 日本は年間を通して降る雨の量が多い国だと言える
- ③ 日本は川などが多いため、水に恵まれていない国だとは言えない
- ④ 日本は地域による違いがあるが、全般的に雨の少ない国だと言える



J155の読解問題では、全体の要点よりも部分的な読み方しかできない受験者が日本語能力の上位レベルに多く、そうした部分的な理解に起因する誤答選択肢④の期待選択確率が能力最上位層で25%程度ある。文法的な部分の理解だけでなく、文章全体から要点を読み取る力を育てる教育が一層大切であることが示唆される。

3.3 日本留学試験 (EJU)

[E150] 読解

次の文章はある病院の健康診断に関する注意書きです。内容と合っているものはどれですか。

健康診断は1か月前から予約を受け付けております。必ず事前にご予約の上、ご来院ください。

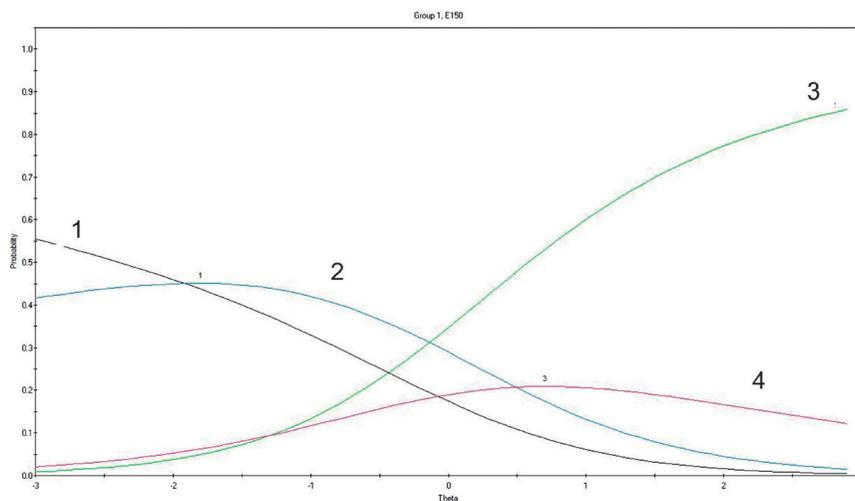
お食事は前日の夜9時まで済ませて、当日は何も食べずにおいでください（多少水を飲む程度なら差し支えありません）。

目の精密検査の際に特殊な目薬を使用しますので、検査後一時的に目がぼんやりします。車やオートバイ、自転車でのご来院はお控えください（目は一時間程度で元の状態に戻ります）。

検査結果は2週間後にご自宅へ郵送致します。検査結果についてご質問等ありましたら、ご来院の上、担当医師とお話してください。

- ① 健康診断の時、車で来ることは避けなければいけない。
- ② 健康診断は1か月前に予約しなければいけない。
- ③ 健康診断の検査結果を知るためには、病院へ行かなければいけない。

④ 健康診断の前日は何も飲んだり食べたりしてはいけない。



誤答選択肢②「健康診断は1か月前に予約しなければいけない。」の期待選択確率が能力中上級レベルまで最も高く、上級者になってようやく選択率が下がり始める。本文中で、正答選択肢①「健康診断の時、車で来ることは避けなければいけない。」と関わる場所は「車やオートバイ、自転車でのご来院はお控えください」である。本文には「避ける」という語が、選択肢には「控える」という語があり、どちらも難易度が高い語であるため、これらの語彙を理解できた受験者から正答にたどりついているものと思われる。

一方、本文中この誤答選択肢②に関わるのが「健康診断は1か月前から予約を受け付けております。必ず事前にご予約の上、ご来院ください。」である。この二つの文を読んで誤答選択肢②「健康診断は1か月前に予約しなければいけない。」を選んだ方略と思われるが、その際に後文が前文の内容を含むと解釈できたのだろうか。上記の「避ける」「控える」等の語彙が未習得で、残りの選択肢の中から最も近い内容を選んだ結果、正答判断が選択肢②となった可能性が高い。しかし、この二つの文章の解釈の仕方については今後も検討の余地があるため、類似問題が集まれば改めて考察を要するものと考えられる。

3.4 今後の展望

本研究では、大規模言語テスト (CSAT, EJU, JLPT) の既出項目を用いて構成された韓国語母語話者の日本語学習者を対象としたモニター試験データのIRT分析に基づいて、韓国語母語話者の日本語学習者が日本語能力を測定する様々なテスト項目に対してどのような解答行動を取るのか、そして特徴的な解答行動を引き起こす諸要因について探索的な考察を行った。

その結果、特徴的な解答行動を引き起こす要因として母語 (韓国語) の干渉や日本語教育課程の影響などが示唆された。これらの分析結果および考察は、公平性の担保が第一義である公

的言語テストにおいてはより妥当性の高いテスト項目を開発するための基礎的な資料となるものである。また、今後の語彙や文法、読解の教授法などの改善といった、個別具体的な日本語教育の現場に有効な示唆を与えるものでもある。今後の展望としては、本稿で取り上げることでできなかった多くのテスト項目についての項目分析の結果をまとめていき精緻化することによって、テスト項目の解答行動特性の体系的なカテゴリ分類を目指していきたい。

引用文献

- Bock, R.D. (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, pp.29-51.
- Cai, L., Thissen, D. & du Toit, S. (2011) IRTPRO 2.1. Scientific Software International. [Computer Software]
- 国際交流基金 (2012) 『2012年度 日本語教育機関調査 結果概要 抜粋』
- Lord, F.M. (1952) A theory of test scores. *Psychometrika Monograph* No. 7.
- Lord, F.M. & Novick, M.R. (1968) *Statistical Theories of Mental Test Scores*. Addison Wesley, Reading, MA.
- Lord, F.M. (1980) *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- 大澤 公一 (2011) 日本語能力における非音声領域の客観的測定および言語運用能力の主観的評定に共通する潜在特性の項目反応モデリング. 『日本語教育』, 150, pp.71-85.

(いとう・りえ 外国語学部准教授)

(おおさわ・こういち 山口大学アドミッションセンター)