

中国語および韓国語を母語とする日本語学習者のリーディングを基にした 日本語聴解能力テストの開発と評価¹

早川 杏子 (関西学院大学)²

魏 志 珍 (中華大学)³

初 相 娟 (天津外国語大学)⁴

玉岡 賀津雄 (名古屋大学)⁵

The present study reported the development and evaluation of a listening comprehension test for learners of Japanese as a foreign language. The ability of listening comprehension was subdivided into three abilities; namely, assembly comprehension ability, interference ability, and restructuring ability. The test was divided into three subtests with each subtest corresponding to one of the three abilities. A total of 20 listening comprehension questions were administered to 249 native Chinese speakers and 41 native Korean speakers all of whom were learning Japanese. The mean score of 290 participants of the 20 questions (maximum possible score: 20 points) was 10.99 with a standard deviation of 3.96. Scores ranged from 3 to 20 points across all participants. Cronbach's alpha reliability coefficient was 0.76, indicating a relatively high reliability. Among the three abilities, the subtest for interference ability was the most difficult for both Chinese and Korean speakers. Using the Item Response Theory, each of the 20 test items were further evaluated using the three criteria of item difficulty (DIFF), item discrimination power index (DISC), and actual equivalent number of options (AENO). This data shows that the current test provides a reliable and accurate representation of a Japanese language learner's listening comprehension ability.

キーワード：日本語聴解テスト、集積的理解、推論能力、再構造化能力、テスト評価

1. はじめに

聴解は、言語知識のみならず、語用的知識、話題の背景知識、認知能力などさまざまな下位技能によって駆動する複雑な言語活動である。その素材であるテキストの理

¹ English Title: Developing and Evaluating a Japanese listening comprehension test using the data elicited from native Chinese and Korean speakers learning Japanese

² Kyoko Hayakawa (Lecturer, Center for Japanese Language Education, Kansai Gakuin University, E-mail: hayakawa@kansai.ac.jp)

³ WEI, Chih Chen (Assistant Professor, Department of Applied Japanese, Chung Hua University, Taiwan, E-mail: weichihchen@chu.edu.tw)

⁴ CHU, Xiang Juan (Associate Professor, Department of Japanese Studies, Tianjin Foreign Studies University, E-mail: chuxiangjuan@aliyun.com)

⁵ TAMAOKA, Katsuo, (Professor, Graduate School of Languages and Cultures, Nagoya University, E-mail: ktamaoka@lang.nagoya-u.ac.jp)

解については、テキスト言語学、談話分析、スキーマ理論、情報処理容量理論などの観点からさまざまな説明付けが行われている。

命題間の意味的関連付けの過程を階層的なネットワーク構築として捉えて、文章理解過程を構築-統合モデル (construction-integration model) によって理論化したのは、Kintsch (1998) である。情報の受け手は、言語入力に対応する概念と命題群のミクロ的な結束表象を形成する。すると、そのミクロの結束表象とリンクした背景知識がノードを通じて活性化され、その命題に対する推論が作用する。この時点では、命題に関する候補群が活性化している状態だが、言語入力が進むにしたがって、文脈に合わない背景知識や推論は不要な推論として排除されていく。このようなプロセスが繰り返されることによって、マクロの結束性を作り、最終的に情報の受け手は、テキスト全体の命題テキストベース表象を生成し、そこに読み手の推論を加えた状況モデルを構築するとしている。

聴解は、連続的に音声入力される個々の言語情報を順次に統合させていく漸進的な (incremental) 性格を持つ。そのため、テキスト理解という意味では、その理解過程は読解と重なる部分が多々ある。しかし、最も大きく異なる点は、読解では読み進めてきた理解に矛盾があることに気づけば、途中で後戻りをする読み返しができるが、聴解ではそれができないことである。また、読解における文章のほとんどが事前に十分に練られた計画的なテキストであるのに対して、聴解におけるテキストには、相互交渉の過程で、考えがまとまらずに言い直しをしたり、話が前後したり、頭に浮かんだ概念をまず口に出して、後からそれについて補足説明したり、あるいは辻褄が合わず説明をやり直したりするなど、無計画な談話が見られることがある (Oshs, 1979)。表象の修正と再統合を頻繁に行わなければならないという点において、このようなテキスト理解には、認知的な心的作業に高い負荷がかかるだけでなく、継時的要因による記憶保持の情報容量制約ともかかわってくることから、ワーキングメモリの観点からみても、読解とはまた異なる種類の難しさが存在する。

テキストの理解過程における意味的な関連付けの理論的実証は、聴解を対象にした第2言語習得研究においてはほとんど行われていない。今後は、以上のような観点から詳細に日本語学習者の聴解の理解過程について明らかにしていく必要があると考える。そこで本研究では、第2言語 (あるいは外国語) としての日本語の聴解の理解過程を解明するための一つの基盤として、実施が簡便であり、聴解能力を適切かつ安定的に測定できるテストの開発を試みる。そして、開発したテストを中国語もしくは韓国語を母語とする日本語学習者に対して実施し、そのデータを基にテストの信頼性と妥当性を検討する。

2. 聴解テスト項目の構成

言語テストにおいては、言語能力をどのように捉えるかが重要な問題となる。測定上仮定する能力を構成概念（construct）という。本聴解テストでは、まとまったテキストの理解という観点から、3つの構成概念を想定し、テキストタイプの異なる聴解テストを作成した。また、本聴解テストは、日本語学習者の聴覚入力による一般的な言語理解能力を測定しようとする目的から、作成にあたっては、語用的側面、認知・心理的の要素を調整した。具体的には、語用的側面の要素は談話形式、トピック、場面、対人関係などである。認知・心理的側面の要素とは、言語を認知する際に関わる処理上の難易条件のことである。以下、詳細について述べる。

2.1 設問の構成

本聴解テストでは、Kintsch（1998）の構築-統合モデル、テキスト言語学による一貫性の概念、談話特性やワーキングメモリなどの言語心理的側面の概念を援用し、聴解能力を構成する概念として「集積的理解」（assembly comprehension）、「推論能力」（inference ability）、「再構造化能力」（restructuring ability）の3つを想定した。「集積的理解」は、次の問題のように、テキストの明示的な情報を積み重ねていけば、答えが導き出せるようなテキストである。これは、構築-統合モデルの相互に関連のあるミクロ的な結束表象を統合していくことにより、命題テキストベース表象を作り上げていくという基本的な考えに基づいた構成概念である。例1に挙げた問題では、閉館時間が午後4時であるというアナウンスに続いて、初めて図書の貸し出しを申し出る場合には、閉館30分までに1階のカウンターでの手続きを行わなければならないことが明示的に指示されている。したがって、問いの条件である「初めて本を借りる人」の行動は、午後3時30分までに1階のカウンターで手続きをするという3番が正答となる。

例1 集積的理解

集積的理解（独話）図書館のアナウンス

図書館のアナウンスです。初めて本を借りる人はこれからどこでどうすればいいですか。

女：お知らせいたします。本日曜日の閉館は午後4時、あと1時間ほどで終了いたします。図書の貸し出し手続きは閉館15分前までとなっております。コピーの申し込みは2階受付で3時半までにお願います。また初めて図書の貸し出しを希望される方は、閉館30分前までに1階のカウンターで貸し出しカードの作成をお済ませください。

質問：今日初めて本を借りる人はこれからどうすればいいですか。

- ① 3時45分までに、1階のカウンターで手続きをします。
- ② 3時45分までに、2階の受付で手続きをします。
- ③ 3時30分までに、1階のカウンターで手続きをします。
- ④ 3時30分までに、2階の受付で手続きをします。

次に、「推論能力」とは、談話のやり取りの中に明示的・直接的な言語表現が提示

されていないものの、非明示的に話し手の含意が埋め込まれている種のテキストである。テキスト言語学では、文の意味的連続性を一貫性 (coherence) と呼ぶが、この「推論能力」とは、文間に意味的な連続性を見出し、話し手の意図や含意を適切に把握する能力を指す。例えば、以下の例2では、女性が作成したポスターに対する男性の意見を問う問題であるが、男性の発話には、選択肢のような「～たほうがいい」といったような明示的な助言の表現は使われていない。男性の「留学生の写真を載せたのはいい」、「明るい色が使われていて活発な雰囲気も出ている」、「デザインは悪くない」という好意的な評価を合わせ見れば、修正すべき点は、男性の「一番知りたい情報がやや見づらい」、「いつどこで開催されるのか」、女性の「具体的な情報の部分が小さすぎるかも」という発話群から推測すると、正答は3番の「日時や場所などの字を大きくしたほうがいい」のみであることがわかる。

例2 推論能力

推論能力 (会話) イベントのポスター

女性が留学生交流イベントのポスターについて意見を聞いています。男性の意見は下のどれですか。

女：ねえ、留学生交流イベントのポスターを作ってみたんだけど、どうかな。

男：うん、留学生の写真を載せたのはいいね。明るい色が使われていて、活発な雰囲気も出ているね。

ただ、一番知りたい情報の部分がやや見づらいかなあ。遠くから見たら、いったい何をやるのかとか、いつどこで開催されるのかとか、ぱっと見てわからないんじゃない？

女：たしかに。イベントをやることは前面に出ているけど、具体的な情報の部分が小さすぎるかも。

男：でも、全体的なデザインは悪くないから、少し変更するだけで十分だと思うよ。

質問：男性の意見は下のどれですか。

- ①もっと明るい色を使ったほうがいい。
- ②留学生の写真を載せたほうがいい。
- ③日時や場所などの字を大きくしたほうがいい。
- ④デザインを全体的に変えたほうがいい。

最後に、「再構造化能力」は、継時的な制約の下での活動にある聴解において特に重要となる能力である。この構成概念は、構築-統合モデルにおけるテキストベースの再統合の過程や、記憶容量に制限があるといわれるワーキングメモリの観点に基づくものである。会話の場合には、相互交渉の過程で終了した話題情報が、当該の話題を跨いで提示されることがある。例えば、以下の例3はそのようなテキストタイプの例である。これは、旅行社で客がホテルの紹介を受ける談話で、ベッドタイプの条件が合わなかった「白山ホテル」に続いて他の候補を探す途中で、最初の「白山ホテル」に望み通りの空きができ、客は最終的にそこに予約を決めるという内容である。談話の最後のほうで、初めに取り下げたはずの「白山ホテル」が相互交渉によって再提示されるという事態に、聞き手は流れてくるテキスト情報を処理しながら、同時に、冒頭の話し手の希望条件を再想起し、他の候補であったホテルの情報を却下して速やかに修正と再統合を行わなければならない。談話展開から、それまで結束化された表象群を速やかに且つ柔軟に構造化し直す過程は、記憶や注意の資源配分と深く関わるもの

であり、認知的に高い負荷がかかる作業であるといえよう。すなわち、再構造化能力とは、結束化された表象群を談話展開に応じて、構造的に変容させ、注意すべき情報の取捨選択を行うことができる能力を指す。なお、こうしたテキストは、通常予め練られたテキストにではなく、談話の相互交渉上においてみられることから、このようなテキストは、ほとんどが会話形式のものであった。

例3 再構造化能力

再構造化能力 (会話) ホテルの予約

男の人と女の人が話しています。女の人は、ホテルをどうすることにしましたか。

男：いらっしゃいませ。

女：友人と2人で行く北海道旅行のホテルを探しているんですが、おすすめありますか。予算はだいたいこれくらいで。

男：そうですね。白山ホテルはいかがですか。駅から近いですし、新しいですよ。

女：あら、でもこれダブルしかないわね。

男：ツインがよろしいですか。それでしたら、こちらの雪見ホテルはツインになります。ただ、少し駅から遠くなりますが。

女：ここは駅から1キロくらい遠いのか。うーん、もう少し近いところは。

男：丘上ホテルでしたら、駅から歩いて10分ほどです。

女：いいわね。ツインで、禁煙ルームで予約できますか。

男：あ、申し訳ございません。こちらはもう予約がいっぱいで、お取りできません。喫煙ルームなら空いているんですが。

女：私、たばこのにおいがだめなの。雪見ホテルは禁煙ルームが空いていますか。

男：ここも喫煙ルームしか空いていないようです。あ、最初におすすめしたホテルでツインに空きが出たようです。ちょうど禁煙ルームですよ。

女：よかった。じゃあ、そこにするわ。

質問：女の人は、ホテルをどうすることにしましたか。

①白山ホテルに泊まる ②雪見ホテルに泊まる ③丘上ホテルに泊まる ④どこにも泊まらない

以上のような聴解の構成概念に基づき、本聴解テストは、「集積的理解」に10問、「推論能力」に5問、「再構造化能力」に5問を割り当て、合計20問で構成した。

2.2 テキストの調整

2.2.1 語用的側面：談話形式・トピック・場面・対人関係

聴き取りに少なからず影響を与えることが想定される語用的側面として、談話形式、トピック、場面、対人関係の要素を考慮して、それらにできるだけ偏りが出ないように配慮した。談話形式は、会話ないし独話のテキストで、トピックや場面では、学校・会社・旅行・社会的な話題や場面を広くカバーするようにした。日本語では、対人関係においては、話す相手によって言語スタイルの意識的な選択が行われるため、既知の相手としては、インフォーマルな言語スタイルが選択される友人同士や、フォーマルな言語スタイルが望ましいとされる上司と部下、あるいは初見の相手との対話など、多様な談話テキストを作成した。また、公共放送など不特定多数への聴き手を想定し

表1 聴解テストの概要および分析結果

| 問題番号 | 構成 | 談話形式 | トピック | 場面 | 関係 | 語彙難易度*注1 | 文章難易度*注2 | readability score | 項目困難度 (DIFF) | 項目弁別力指数 (DISC) | 実質選択肢数 (AENO) |
|------|--------|------|-----------|-----------|---------------|----------|----------|-------------------|--------------|----------------|---------------|
| 1 | 集積的理解 | 会話 | 授業の欠席 | 日常 (学校) | 友人/友人 | とてもやさしい | 初級後半 | 5.45 | 0.738* | 0.303 | 2,357 |
| 2 | 集積的理解 | 会話 | 結婚アンケート | 日常 | 友人/友人 | やさしい | 初級後半 | 4.9 | 0.483 | 0.563 | 3,441 |
| 3 | 集積的理解 | 会話 | 銃の問題 | 日常 | 友人/友人 | ふつう | 初級後半 | 4.9 | 0.476 | 0.421 | 3,329 |
| 4 | 集積的理解 | 独話 | 処方箋 | 病院 | 医者/患者 | とてもやさしい | 中級後半 | 4.84 | 0.486 | 0.380 | 3,489 |
| 5 | 集積的理解 | 独話 | 勉強場所 | レポート | 話し手/不特定多数 | とてもやさしい | 中級前半 | 3.77 | 0.810* | 0.392 | 2,000 |
| 6 | 集積的理解 | 独話 | レポートの提出 | 日常 (学校) | 教師/生徒 | ふつう | 中級後半 | 3.56 | 0.734* | 0.432 | 2,293 |
| 7 | 集積的理解 | 独話 | 運動会の延期 | 学校アナウンス | 事務 (教師)/父兄・生徒 | とてもやさしい | 中級後半 | 3.0 | 0.803* | 0.364 | 1,990* |
| 8 | 集積的理解 | 独話 | 図書館のアナウンス | 公共スペース | 話し手/不特定多数 | ふつう | 中級後半 | 2.95 | 0.659 | 0.433 | 2,519 |
| 9 | 集積的理解 | 独話 | 少額硬貨 | レポート | 話し手/不特定多数 | 少し難しい | 中級後半 | 2.89 | 0.241 | 0.390 | 3,902 |
| 10 | 集積的理解 | 独話 | 人気のアパート | レポート | 話し手/不特定多数 | ふつう | 中級後半 | 2.53 | 0.645 | 0.433 | 2,704 |
| 11 | 推論能力 | 会話 | 仕事の依頼 | 職場 | 先輩/部下 | ふつう | 中級前半 | 4.46 | 0.403 | 0.384 | 3,444 |
| 12 | 推論能力 | 会話 | 鳥と電波 | インタビュー | リポーター/研究者 | やさしい | 中級前半 | 4.27 | 0.541 | 0.434 | 3,090 |
| 13 | 推論能力 | 独話 | 留守番電話 | 日常 (ビジネス) | 話し手/取引先 | ふつう | 中級前半 | 4.27 | 0.238 | 0.486 | 3,960 |
| 14 | 推論能力 | 会話 | イベントのポスター | 日常 (学校) | 友人/友人 | ふつう | 中級前半 | 3.59 | 0.452 | 0.613 | 3,472 |
| 15 | 推論能力 | 会話 | 会社の建て直し | 職場 (会議) | 会社役員/会社役員 | ふつう | 中級後半 | 3.3 | 0.469 | 0.278* | 3,421 |
| 16 | 再構造化能力 | 会話 | ホテルの予約 | 日常 | 旅行会社/客 | やさしい | 初級前半 | 5.66 | 0.666 | 0.566 | 2,622 |
| 17 | 再構造化能力 | 独話 | 旅行の日程 | 日常 (電話予約) | 旅行会社/客 | やさしい | 初級後半 | 5.21 | 0.434 | 0.338 | 3,418 |
| 18 | 再構造化能力 | 会話 | 出産祝い | 日常 | 友人/友人 | ふつう | 初級後半 | 4.85 | 0.717* | 0.451 | 2,393 |
| 19 | 再構造化能力 | 会話 | 商品のネーミング | 職場 (会議) | 社員/社員 | 少し難しい | 初級後半 | 4.75 | 0.400 | 0.450 | 3,280 |
| 20 | 再構造化能力 | 独話 | 新人の歓迎会 | 職場 | 話し手/社員 | やさしい | 中級前半 | 3.75 | 0.593 | 0.338 | 3,049 |

注1：語彙難易度は、「リーディングチュウ太」 <http://language.tiu.ac.jp/> より算出。

注2：文章難易度とreadability scoreは、「文章難易度判定システムalpha版」 <http://readability.net/> より算出。

注3：*は項目困難度、項目弁別力指数、実質選択肢数の各指標において、問題の適切さを示す望ましい値を満たしていないことを表す。

たフォーマルな独話形式なども含んだ。したがって、本テストの言語形式には、場面や対人関係に応じて、普通体、丁寧体、敬語などが織り交ぜてあり、特定の言語形式に固定することはしなかった。

2.2.2 認知・心理的側面：テキストの難易度の調整

聴解に影響を与え得る要因には、上述の要素以外に、語彙やテキストそのものの難易度が関わってくる。そこで、各構成概念における語彙やテキスト難易度を調整するために、語彙の難易度判定には、「リーディングチュウ太」（川村よしこ氏が開発した <http://language.tiu.ac.jp/>）を、テキストの難易度判定には、李在鎬・長谷部陽一郎・浅尾仁彦氏らが開発した「日本語文章難易度判定システム alpha 版」（<http://jreadability.net/>）を用いて、文章難易度とリーダビリティスコア (readability score; 母語話者判定) を算出した。その結果は、表 1 に示した。

語彙の難易度判定とテキストの難易度判定は、記述的な指標であるため、難易度が極端に偏ったものではないことを確認するにとどめる。一方、リーダビリティスコアは、文章（テキスト）の難易度が数値で表され、数値が大きいほど易しく、小さいほど難しいテキストであることを示してくれる。それぞれの平均値を見てみると、「集積的理解」3.88（標準偏差 1.06）、「推論能力」が 3.98（標準偏差 0.50）、「再構造化能力」が 4.84（標準偏差 0.71）であった。「再構造化能力」のスコアが他の 2 つよりもやや易しめと示されたが、この種のテキストは、聴解処理の過程上において情報の一定時間の保持や、必要な情報の取捨選択ならびに再統合という複雑な作業が関わってくるため、テキストはむしろ難易度が低いほうが適当だと考えた。リーダビリティスコアを従属変数として、「集積的理解」、「推論能力」、「再構造化能力」の 3 つの変数の間に差があるかどうかを一変量の分散分析によって検討したところ、変数間に有意な違いはなく [$F(2, 17)=2.144, ns$]、どの構成概念のテキストも同程度の文章難易度であった。これにより、本テストの構成概念別のテキスト間における難易度は、等質であることが確認できた。

3. 聴解テストの実施と分析

3.1 調査協力者

3.1.1 中国語を母語とする日本語学習者

調査協力者は、調査協力者は、中国にある 2 つの大学（2～3 年生）および台湾にある 1 つの大学（3～4 年生）で日本語を主専攻とする日本語学習者 249 名（250 名のうち、1 名欠損値があり分析から除外）であった。内訳は、2 年生が 51 名、3 年生が 150 名、4 年生が 48 名である。

3.1.2 韓国語を母語とする日本語学習者

日本の大学で学ぶ韓国人日本語学習者 41 名で、日本語学習歴は、1 年未満が 7 名、

1年～2年未満が10名、2年～3年未満が10名、3年～4年未満が7名、5年以上が6名であった（うち1名は学習歴不明）。

3.2 結果

3.2.1 全体の結果

20問の聴解テストを計290名に実施した。分析の結果、全体の平均点が10.99（標準偏差3.96）点で、最高点は20点、最低点は3点であった。クロンバックの α 係数は.757であり、20問という少ない問題数から考えると、比較的高い値が得られた。なお、母語別で信頼性の検定を行ったところ、中国語母語話者は $\alpha = .709$ 、韓国語母語話者は $\alpha = .844$ となり、韓国語を母語とする日本語学習者においては非常に高い値が得られた。なお、「集積的理解」、「推論能力」、「再構造化能力」の3つに下位カテゴリ化した項目の項目得点は、「集積的理解」（10点満点）の平均が6.06（標準偏差2.14）点で、「推論能力」（5点満点）が1.96（標準偏差1.38）点、「再構造化能力」（5点満点）は、2.81（標準偏差1.35）であった。

3.2.2 項目分析

全体としてのテストの内容的一貫性の指標である信頼性係数からも、本聴解テストは比較的安定していることが分かった。さらに、テストを構成する各項目の問題を詳細に評価するために、TDAP（Test Data Analysis Program）Ver.2.0（大友・中村, 2002）を用いて項目分析を行った。大友（1996）によれば、項目分析では、項目困難度（item difficulty: DIFF）、項目弁別力（item discrimination power index: DISC）、実質選択肢数（actual equivalent number of options: AENO）が主な評価指標となる。この3つの指標による全体の結果は、表1に示した。なお、本研究では中国語あるいは韓国語を母語とする日本語学習者にテストを実施したが、これらの群によって構成概念、項目によってその傾向に異なりが見られたため、以下これらの指標から両群の結果を比較する。

3.2.2.1 項目困難度（DIFF）

項目困難度とは、その項目がどのくらい難しかったのかを示す指標である。この指標の値は、項目困難度＝正答者数／受験者総数で求められるため、正答率ともいえる。項目困難度は、1に近いほど易しく、0に近いほど難しい項目ということになる。しかしながら、受験者が不特定多数である集団基準準拠テストの場合、特定の教育機関で行った教育内容の到達度を測定する目標基準準拠テストとは性質や目的が異なる。そのため、その値がどのような意味を持つのかということを考えなければならない。集団基準準拠テストの項目評価を行おうとするならば、第1に適切さの基準を定める必要がある。項目困難度の適切さは、主に選択肢数によって決められる。選択肢が複

数ある場合は、当て推量によって偶然正答が選ばれる確率も加味され、最適困難度の式は、最適困難度=0.5+0.5（1/選択肢数）で求められる。本テストの選択肢数はすべて4つなので、最適困難度は0.625となる。ここでは概ね0.70以上の値はやや容易な問題とみなすことにする。全体の結果では、「集積的理解」カテゴリでは「授業の欠席」、「勉強場所」、「レポートの提出」、「運動会の延期」の4問が0.70以上（表1に*で示した）であった。「推論能力」カテゴリでは1つもなく、「再構造化能力」カテゴリの項目では、「出産祝い」の1問であり、全体としては0.40~0.60を推移する値の問題が多かった。ただ、「集積的理解」カテゴリの「少額硬貨」（0.241）と「推論能力」カテゴリの「留守番電話」（0.238）は項目困難度の値が小さく、この値を見る限りでは、難しい問題であったといえる。

表2 中国語／韓国語を母語とする日本語学習者別の項目分析の結果

| 問題番号 | 構成 | 談話形式 | トピック | DIFF | | DISC | | AENO | |
|------|--------|------|-----------|--------|--------|--------|-------|-------|--------|
| | | | | 中国 | 韓国 | 中国 | 韓国 | 中国 | 韓国 |
| 1 | 集積的理解 | 会話 | 授業の欠席 | 0.735* | 0.756* | 0.269* | 0.550 | 2.369 | 2.254 |
| 2 | 集積的理解 | 会話 | 結婚アンケート | 0.434 | 0.780* | 0.565 | 0.339 | 3.579 | 2.136 |
| 3 | 集積的理解 | 会話 | 銃の問題 | 0.418 | 0.829* | 0.372 | 0.344 | 3.477 | 1.809* |
| 4 | 集積的理解 | 独話 | 処方箋 | 0.482 | 0.512 | 0.390 | 0.416 | 3.468 | 3.254 |
| 5 | 集積的理解 | 独話 | 勉強場所 | 0.795* | 0.902* | 0.359 | 0.587 | 2.076 | 1.524* |
| 6 | 集積的理解 | 独話 | レポートの提出 | 0.719* | 0.829* | 0.413 | 0.538 | 2.376 | 1.749* |
| 7 | 集積的理解 | 独話 | 運動会の延期 | 0.787* | 0.902* | 0.335 | 0.511 | 2.060 | 1.454* |
| 8 | 集積的理解 | 独話 | 図書館のアナウンス | 0.627 | 0.854* | 0.428 | 0.302 | 2.624 | 1.664* |
| 9 | 集積的理解 | 独話 | 少額硬貨 | 0.213 | 0.415 | 0.318 | 0.548 | 3.887 | 3.550 |
| 10 | 集積的理解 | 独話 | 人気のアパート | 0.622 | 0.780* | 0.387 | 0.638 | 2.817 | 1.901* |
| 11 | 推論能力 | 会話 | 仕事の依頼 | 0.369 | 0.610 | 0.339 | 0.426 | 3.536 | 2.568 |
| 12 | 推論能力 | 会話 | 鳥と電波 | 0.522 | 0.659 | 0.427 | 0.439 | 3.162 | 2.577 |
| 13 | 推論能力 | 独話 | 留守番電話 | 0.217 | 0.366 | 0.461 | 0.537 | 3.923 | 3.800 |
| 14 | 推論能力 | 会話 | イベントのポスター | 0.418 | 0.659 | 0.577 | 0.723 | 3.571 | 2.563 |
| 15 | 推論能力 | 会話 | 会社の建て直し | 0.450 | 0.585 | 0.217* | 0.502 | 3.517 | 2.534 |
| 16 | 再構造化能力 | 会話 | ホテルの予約 | 0.635 | 0.854* | 0.553 | 0.588 | 2.742 | 1.758* |
| 17 | 再構造化能力 | 独話 | 旅行の日程 | 0.418 | 0.537 | 0.270* | 0.631 | 3.470 | 2.769 |
| 18 | 再構造化能力 | 会話 | 出産祝い | 0.711* | 0.756* | 0.468 | 0.445 | 2.420 | 2.194 |
| 19 | 再構造化能力 | 会話 | 商品のネーミング | 0.341 | 0.756* | 0.383 | 0.471 | 3.321 | 2.194 |
| 20 | 再構造化能力 | 独話 | 新人の歓迎会 | 0.582 | 0.659 | 0.294* | 0.581 | 3.109 | 2.563 |

注：*は項目困難度、項目弁別力指数、実質選択指数の各指標において、問題の適切さを示す望ましい値を満たしていないことを表す。

次に、中国語あるいは韓国語を母語とする日本語学習者別に項目困難度の結果を見てみると、「集積的理解」カテゴリにおいて、韓国語を母語とする日本語学習者の場合は0.70を超えた問題が多く、この種の聴解テキストは、韓国人日本語学習者にとって理解しやすいものであったようである。一方、0.70を超えた問題が8項目であった韓国人日本語学習者に対し、中国語を母語とする日本語学習者では、その半分の4項目のみであり、必ずしも易しい問題とはなっていなかったようである。

中国語および韓国語を母語とする日本語学習者で、どちらも値が高い項目が、「集積的理解」カテゴリの「授業の欠席」（中国：0.735、韓国：0.756）、「運動会の延期」（中国：0.789、韓国：0.902）、「勉強場所」（中国：0.795、韓国：0.902）、「レポートの提出」（中国：0.719、韓国：0.829）であった。2群の間に大きな差があった項目は、「集積的理解」カテゴリの「銃の問題」（中国：0.418、韓国：0.829）と「再構造化能力」カテゴリの「商品のネーミング」（中国：0.341、韓国：0.756）であった。以下の例4「銃の問題」の解答傾向を見ると、韓国語母語話者が選んだ選択肢にほとんどばらつきがないのに対し、中国語母語話者は、正答の3番のほか、1番や4番を選んだ受験者も多く、解答の分散度が高いことがわかった。テキストの中では、「獲物の皮を取った後、それをそのまま捨てて」行く人がいることが、「さらにとんでもないのが」という表現で強調され、最後に「腐った匂いがひどい」と放置された皮の腐臭が迷惑であることが述べられている。それ以外の選択肢は、発砲された銃の音をきっかけに説明された問題の一部であり、もっとも重要な問題点とはいえない。答えを導くキューが与えられている集積的理解のテキストであっても、談話の情報の要点を捉え、情報の要不要の判断がつかなければ惑わされてしまう例である。

例4 集積的理解

集積的理解（会話）銃の問題

男の人と女の人が話しています。男の人が一番問題だと言っているのは何ですか。

女：わあ、なーに、今の音！

男：銃の音だよ。この辺りで狩りをする人がいるんだ。シカやウサギを銃で撃ったりして。

女：ええ。誤って登山者や住民に当たったらどうするの？

男：そうなんだよ。さらにとんでもないのが、獲物の皮を取った後、それをそのまま捨ててっちゃう人がいるんだ。

女：信じられない。そもそも何のためにシカやウサギを殺すの、あんなに可愛いのに…。

男：まあ、彼らの目的はともかく。とにかく困るんだよ。腐った匂いがひどくてさー。

質問：男の人が一番問題だと言っているのは何ですか。

- ①シカやウサギを殺すことです。 ②銃を撃つことです。
 ③死んだ動物を放置することです。 ④獲物の皮を取ることです。

そして、特に中国語を母語する日本語学習者の値が低い項目が、「集積的理解」カテゴリの「少額硬貨」（0.213）と「推論能力」カテゴリの「留守番電話」（0.217）であった。「留守番電話」は、ビジネスの内容であり、他の日常生活場面の項目群と大きく異なる点は、敬語が非常に多く使用されていることである（例5）。敬語は、「御社」と「弊社」、「うかがう」、「させていただく」など、ビジネス場面では頻繁に使われるが、敬語が正しく理解できていなければ、この談話を適切に理解することはできない。なぜなら、尊敬語か謙譲語かの区別が主語を決め、誰の行動かを規定するからである。中国語には、敬語にあたる文法形態がない（その代わりに、語や言い回しで敬意を表現する）ことから、韓国語母語話者に比べて言語習慣的にも敬語の使用方法に対する理解が難しく、困難度が高くなったのであろう。

例5 推論能力

推論能力 (独話) 留守番電話

男の人が話しています。留守番電話を聞いた人は、このあと何をしますか。

男：あーもしもし、いつもお世話になっております。三ツ星商事の田中です。

えーと、さきほど御社にうかがったんですけれども、ちょうど席を外されているとのことだったので、携帯にお電話させていただいています。今回ですね、弊社の新製品のご案内をさせていただこうと思ひまして、カタログをお持ちしました。カタログは、デスクのほうに置いておきましたので、戻られましたら、ご覧いただければと思います。また後日、お邪魔させていただきますので、よろしくお願ひ致します。失礼致します。

質問：留守番電話を聞いた人は、このあと何をしますか。

- ① 田中さんの電話を待つ。 ② 田中さんに電話する。
③ 田中さんを案内する。 ④ 特に何もしない。

この項目は、言語形式の理解を測定しているだけではなく、メッセージを受けて、次にどのような行動を取るべきか、という語用的な知識を問う問題でもある。この問題は、営業担当者が取引先の顧客に、自分の会社のカタログを届けにいったが、不在だったので、顧客のデスクにカタログを残し、後日そのカタログについて直接会って話したい、という内容である。したがって、電話の受け手は、後日田中さんの来訪を待てばよく、留守番電話を聞いたあと、特に何かをする必要はない。ビジネス場面では、このようなやり取りが多く交わされるため、この電話の意図や含意を理解できなければ、適切な答え（実際には行動）を導くことはできない。日本語能力が非常に高くても、日本語母語話者のメッセージの含意が理解できず、話し手の意図とは異なる行動を取ってしまう学習者は多い。聴解能力とは、音声から談話内容の概念構築という作業に加え、メッセージの意図に対する理解という面も含め合わせて聴きの理解というものを考慮していく必要がある。

3.2.2.2 項目弁別力指数 (DISC)

項目弁別力指数とは、言語能力が高い受験者と低い受験者を識別するのに用いられる指標である。上述の指標の項目難易度だけで評価しようとした場合に生じる問題は、全受験者が90%以上の正答率だったり、もしくは10%以下など、ほとんどが誤答であったりした場合に、何をもって良い（または良くない）項目とするのかという判断が難しくなることである。そこでテスト理論では、能力ごとの正答率を用いて計算することによって、その正答の選択傾向が能力にみあった項目であるかどうかを項目評価の指標とする。すなわち、上位者ならば適切に答えが選択できるような問題となっているかどうか（下位者の場合は適切に選択することが難しいはずである）を判断するための基準である。この指標を用いると、上位者も下位者も同じように正答が選択できたものは、あまり弁別力が高くない（易しすぎる、もしくは難しすぎる）問題だといえ、修正を要する項目とみなす。項目弁別力は、項目弁別力 = 上位グループの総受験者数に対する正答者数の割合 - 下位グループの総受験者数に対する正答者数

の割合の式で求められる。項目弁別力の適切度は、通例、0.40 以上が最良項目、0.30-0.39 が良好項目、0.20-0.29 が予備項目で改訂の要あり、0.19 以下が不良項目で除外もしくは改訂すべきであるとされている（大友, 1996）。

本聴解テストでは、全体（ $N=290$ ）の項目弁別力指数においてほぼ全ての項目が 0.30 以上の値を示しており、弁別力の高い項目で構成されていることがわかった。唯一「推論能力」カテゴリの「会社の建て直し」が 0.278 とやや低かったが、韓国語母語話者の値を見てみると、0.502 ときわめて高い識別力を持っており、中国語母語話者における値に引きずられたものだと推定される。

例 6 推論能力

推論能力（会話）会社の建て直し

男の人と女の人が話しています。女の人はどんな意見を主張していますか。

男：今の会社の経営状況は、やはり厳しいです。赤字続きでこのままじゃ潰れてしまいます。

女：確かにこのままだと、そうなるに違いありません。しかし、逆に考えれば、今の状況、我々にとっては、ちょうど改革のためのいいチャンスだと思いませんか？

男：と言いますと。

女：現在の赤字を解消することだけを考えると、我々は時代の流れに取り残されてしまいます。社員を全体的に減らせば費用が下がることは事実です。しかし、優れた研究開発ができなければ、将来この会社は生き残れません。

男：つまり、どこを減らしてどこを残すか、何が我々にとって大事かということをよく考えて決めなければなりませんね。

質問：女の人はどんな意見を主張していますか。

① 経営改革を実行すべきではない。

② 研究開発の社員は減らすべきではない。

③ 赤字をすぐに減らすべきだ。

④ 社員全体の数を減らすべきだ。

この問題（例 6）は、ビジネス場面での会議の談話だが、特に話し手の含意を適切に理解することが求められる場面であり、実際のコミュニケーション場面ではごく通常に展開される談話の流れである。中国語では、はっきりと主張を述べる社会文化性を持つことから、日本語を母語としない者にとっては、このような婉曲的な主張の述べ方に慣れておらず、話し手の含意の推論が困難である可能性がある。日本語学習の動機や目的がビジネスでの使用という学習者が多いことから、今後の聴解教育の方向性として、こうした場面での話し手の意図や含意の適切な理解を目指していく教育訓練の必要性を感じる項目である。

3.2.2.3 実質選択肢指数（AENO）

実質選択肢指数とは、各選択肢がどれくらい選ばれたかを示す指標で、言い換えるならば、どれだけ魅力的な選択肢があったかということを表すものである。項目難易度や項目弁別力指数のみで、項目の評価を判断しようとする場合に、錯乱肢が錯乱肢としての機能を果たしていなければ必然的に正答が絞られてしまうため、本来の実力とは別に、方略によって正答率が上がってしまう恐れがある。つまり、その問題の項目難易度や項目弁別力の指標だけで項目の適切さを評価するだけではなく、選択肢が

適切に作成されていたかという点も考慮した上で、その項目の判定を行う必要が出てくるのである。実質選択肢指数は、情報の不確かさを表すエントロピー（entropy）の式 $H = -\sum p_j \log_2 p_j$ を用いて算出される。仮に4つの選択肢 j_1, j_2, j_3, j_4 が10名の受験者に選ばれ、それぞれの選択率が0.00、0.20、0.40、0.40だった場合、そのエントロピーは、 $H = -(0.00 + 0.20 * \log_2 0.20 + 0.40 * \log_2 0.40 + 0.40 * \log_2 0.40) = 1.522$ となる。実質選択肢数は、 $H = 2^H$ で求められるので、この項目の実質選択肢数は $2^{1.522} = 2.872$ と、解答がほぼ3つにばらついていることを示し、錯乱肢が機能をきちんと果たしていた問題であったということが判断できる（大友, 1996）。

本聴解テストでは、全体としては比較的均等に解答がばらついており、錯乱肢がきちんと機能していることから、概ね選択肢は適切であったと考えることができよう。ただし、一つだけ「集積的理解」カテゴリの「運動会の延期」が1.990と、実質的な選択肢が2つのみであった。この問題の選択肢には改訂を加えれば、より全体の信頼性も高くなるであろう。

例7 集積的理解

集積的理解（独話）レポートの提出

先生が話しています。レポートはどのような形で提出しますか。

先生：えーでは、学期末のレポートについてですが一、少子化の社会的影響についてまとめて、自分の意見を述べてください。えー、A4で3枚以内、それ以上は認めません。レポートはパソコンで作成してください。手書きで記述する人はめったにいないでしょうが、念のため言っておきます。それから一、提出方法ですが一、私の研究室の前に提出用のボックスを用意しておきますので、そこに入れておいてください。締め切りは、7月20日の午後4時までとします。それを過ぎたら、いかなる理由でも受け付けません。注意してください。

質問：レポートはどのような形で提出しますか。

- ①手書きで3枚まで書いて、研究室まで渡しに行く。
- ②パソコンで3枚以上書いて、研究室まで渡しに行く。
- ③パソコンで3枚まで書いて、研究室の前にある箱に入れる。
- ④手書きで3枚以上書いて、研究室の前にある箱に入れる。

母語別にこの実質選択肢指数を見てみると、中国語母語話者は全項目において解答が均等に分布しているのに対して、韓国語母語話者においては、「集積的理解」の10項目のうち、6項目が2.0以下となっており、魅力的な選択肢が少なかったようである。この6項目は、項目困難度の値も高く、この群の学習者にとって問題が簡単であったために、解答も絞られたのではないかと考えられる。たとえば、例7「レポートの提出」の問題は、教師の談話中にレポートに関連する情報、さまざまな数字が出てくるため、不要な情報を適切に取捨できるかどうかという能力を測定した問題となっている。重要な情報である提出方法が「手書きか、パソコンか」、「直接渡すか、箱に入れるか」という2×2の問題構成になっているが、重要な情報部分が問題文でおおよそ提示されているため候補が絞られ、選択傾向が安定したと思われる。一方、中国語母語話者の実質選択肢指数が2.376と2～3つに選択肢がばらついていたことがわ

かる。解答のばらつきから、枚数制限に関する部分「3枚以内、それ以上は…」を、「3枚まで」ではなく「3枚以上」と受け取ったり、提出方法の「私の研究室の前に提出用のボックスを…」を「ボックス（箱）に入れる」のではなく「渡しに行く」と聞き誤ったりしているなど、細かい部分を適切に把握せずに、大雑把に聞いている傾向にあることが見て取れた。

3.3 中国語および韓国語を母語とする日本語学習者の比較

中国語を母語とする日本語学習者と比べると、韓国語を母語とする日本語学習者のほうが、得点平均や信頼度係数が高かった。後者は、調査人数が41名と少なかつたにもかかわらず、 $a=0.844$ という非常に高い信頼性が得られた。

また、母語別に項目分析の結果をみると、「集積的理解」は、韓国語母語話者にとっては理解が容易なテキストタイプである可能性が示された。3つのカテゴリの中で、両群ともにもっとも困難度が高く、理解が難しいテキストタイプであったのは、「推論能力」を測定するカテゴリである。これは、話し手の意図が非明示的で、文間から含意を読み取る能力が必要とされるテキストである。日本語の言語知識のほか、社会言語学的な知識や語用論的能力も必要となるため、日本語を外国語として学ぶ者に共通する難しさを持っているのであろう。「再構造化能力」のカテゴリでは、韓国語母語話者のほうがやや困難度が低く、全体的に高い得点に分布が偏っているのに対し、中国語を母語とする日本語学習者は、ほぼ中央に分布が集まっていた。談話中の重要な情報を一定の時間記憶しつつ、談話の初めで一度形成した概念を、談話の流れから再度組み立て直して、適切な情報を取捨選択するというタイプのテキストは、中国語母語話者にとっては、「推論能力」に次いで困難度が高いといえる。

以上のことから、中国語を母語とする日本語学習者の場合は、聴解能力を考慮するとき、特に「推論能力」と「再構造化能力」の構成カテゴリで高い得点が取れるかどうかは弁別の一つの判断材料となるのではないかと考えられる。

4. おわりに

本研究では、Kintsch (1998) の構築-統合モデル、テキスト言語学による一貫性の観点、構築-統合モデルにおけるテキストベースの再統合、ワーキングメモリの観点を導入して、「集積的理解」、「推論能力」、「再構造化能力」の3つの構成概念を設定した聴解テストを開発した。そして、本聴解テストを中国語あるいは韓国語を母語とする日本語学習者に対して実施し、その結果を信頼性の検定および項目分析によって評価を行った。その結果、本聴解テストは3つの側面から概ね適切に日本語学習者の聴解能力を測定していることが確認できた。さらに、母語別の分析から、それぞれに理解度に違いが生じるテキストタイプの傾向があることがわかった。この違いについては、今後何

によって生じたのか要因を慎重に検討していく必要がある。考えられる要因としては、第1に、学習環境である。今回の調査対象者は、韓国語母語話者が第2言語環境下で、中国語母語話者が外国語環境の下での日本語学習であった。第2言語環境下では、インプットの量だけでなく、頻繁に日本人との言語文化的接触を通して、社会言語的な知識を身に付けている可能性がある。「推論能力」にはそうした知識が必要とされる問題があったことから、今後は学習環境別に比較検討をしていくべきであろう。第2に、調査対象者の総体的な日本語能力である。今回は、学習期間を要因に含めず本聴解テストの結果に限定し分析を行ったが、学習段階によって当然理解度も異なってくる。等質な群によって構成概念別の測定を行い、どのようなテキスト特徴が、どのような学習段階において、聴きの理解に影響するのかを詳細に検討していくことが以降の課題である。

今後は、本聴解テストをもとに、第2言語としての日本語の聴解過程や、その他の技能との関わりについてさらに明らかにしていきたい。

謝辞：山口県立大学の林炫情先生、東亜大学の李良姫先生には、韓国人日本語学習者のデータ収集に際し、多大なご尽力をいただきました。ここに感謝の意を申し上げます。

参考文献

- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. New York: Cambridge University Press.
- Ochs, E. (1979). Transcription as theory. In E. Ochs and B. B. Schiefflin (Eds.), *Developmental pragmatics*, 43-72. New York: Academic Press.
- 大友賢二 (1996) 『項目応答理論入門』 大修館書店
- 中村洋一 (2002) 『テストで言語能力は測れるか～言語テストデータ分析入門』 大友賢二監修 桐原書店

オンライン資料：

- 文章難易度判定システム alpha 版 <http://jreadability.net/> (2013年9月1日アクセス)、
李在鎬・長谷部陽一郎・浅尾 仁彦「読解教育支援を目的とする文章難易度判別システムの開発」(課題番号：25370573)
- リーディングチュウ太 <http://language.tiu.ac.jp/> (2013年9月1日アクセス)

使用分析ソフト：

- 2002 Test Data Analysis Program (TDAP) Ver. 2.0 [Windows 版]
(大友賢二監修・中村洋一著『テストで言語能力は測れるか』桐原書店に添付)