関西学院大学日本語教育センター紀要　第５号　抜刷
2016年３月31日　発行

中国語および韓国語を母語とする日本語学習者のデータを基にした日本語聴解能力テストの開発と評価

早　川　杏　子
魏　志　珍
初　相　娟
玉　岡　賀津雄

関　西　学　院　大　学
日本語教育センター

中国語および韓国語を母語とする日本語学習者のデータ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, Kwansai Gakuin University, E-mail: hayakawa@kwansei.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)

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」（interference 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 時半までにお願いします。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

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

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・淺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、

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。

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いまして、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、よろしくお願い致します。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。そこでテスティング理論では、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

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 j1、j2、j3、j4 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、実質選択肢数=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ー、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。それからー、提出方法ですがー、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

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・淺尾 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）