

読解テストの難度に関する研究

齊山弥生

読解能力テスト 難度の安定化 タスク分類 選択肢 マッチング

1. はじめに

日本語教育において学習者の日本語力を測る能力テストの難易を安定させることは、テストの信頼性のために不可欠であるが、このことはテストの目的が学習者の日本語力の伸びを測定することである場合、とりわけ大きな意味を持ってくる。しかし、測定する能力が「読解能力」のように複数の下位能力の複合結果と考えられ、直接観察が不可能な能力であるとき、それを測定するテストの難易を客観的に規定するのはきわめて難しい。

産能大学でも1995年度から、留学生を対象に1年次から2年次にかけての2年間に計5回、読解の能力テストを行っているが、その難易の安定化は当初より大きな課題として認識されていた。しかし、今回、1995～1997年に実施された5回のテスト結果の平均点を比較した結果、テスト間に大きな難易の揺れがあることが明らかとなり、全問題の見直し作業を行うことになった。本稿は、この見直し作業のために行われたこれまでのテスト問題の分析方法と分析結果について報告するものである。

分析は、『日本語能力試験出題基準』（国際交流基金・日本国際教育協会1994年）の「読解問題出題基準」（以下、『出題基準』と略す）で提示されている理論的枠組みに則って行った。『出題基準』では、読解テストの性質・難易を決定する要因として、テキスト（読みの対象）とタスク（読みの目標）を取り上げ、この2つを規定することによって、読みの能力モデルを構成する試みが示されている。このような理論的枠組みの提示はテストの作問に携わる者にとってきわめて意義深いだが、今のところ、報告されている実証データは乏しく、作問の枠組みとして応用するには各現場でのデータとの照合が必要となる。今回のデータ分析の目的は、『出題基準』で示されたタスクの規定を参考に、5回のテストで出題された50問のタスクを分類・比較することによって、タスクの難度を左右していると考えられる要因を探り出し、今後の作問の安定化のために具体的な足がかりを得ることにある。

分析の焦点としたのは、次の3つの点である。

- (1) テスト1～5のテキスト間に明らかな難易の差はないか？
- (2) タスクの種類によって難易に何らかの傾向が見られるか？
- (3) ほぼ同じ難易のテキスト、同じ種類のタスクで、大きく難易が異なる場合、その要因は何か？

本稿の興味は主に(2)と(3)にあり、(1)の確認は(2)と(3)の分析を行うための前提条件として取り上げた。

2. テストの実施と結果

分析対象とした読解テストは、産能大学で留学生を対象に定期的実施している「日本語能力測定」の1種目として出題されたものである。「日本語能力測定」は、大学レベルの留学生に必要とされる日本語力の測定を目的としたテストで、種目には読解の他に聴解、作文、クローズの計4種目があり、読解では論理的な文章の読み取りに主眼を置いている¹⁾。

5回の読解テストは、産能大学1995年度入学の留学生を対象に、1年次に3回(4月、7月、1月)、2年次に2回(7月、1月)実施した。全テストを受験した被験者数は21名で、国籍の内訳は、台湾11、中国5、韓国4、マレーシア1、また、大学入学前の日本での日本語学習歴は1～2年だった。

5回のテストのためのテキストは、すべて朝日新聞の社説からとった。テストの問題形式は四肢択一で、各テキストの問題数は10問とした。5回のテストの21人の平均点は、表1に示す通りである。

表1 読解テスト1-5の結果

テスト名	テスト1	テスト2	テスト3	テスト4	テスト5
実施時期	95年4月	95年7月	96年1月	96年7月	97年1月
満点	10	10	10	10	10
平均点	4.8	5.2	7.1	4.3	5.1
中央値	5	5	7	4	5
標準偏差	1.97	1.79	1.75	1.58	1.78
最高点	9	9	10	7	9
最低点	2	2	4	2	2

表1に見るように、テストの得点はテスト3で大きく上がり、テスト4で最低になり、またテスト5でやや上がるというこぼこの推移を示している。本来、大学入学後の留学生の読解力は個人差はあれ、ゆるやかにでも右肩上がりに変化していくことが期待され、大学という環境で学習している留学生の読解力が一度

上がって、短期間にまた下がるというのは、考えにくい。また、5回のテストはすべて同一の集団に対してほぼ同じ条件で実施されたので、テスト環境や動機に大きな変化があったとも思われない。したがって、この得点の推移のいびつさは、テスト間の難易の差によるもの（あるテストに易しい問題が集中し、あるテストに難しい問題が集中した結果）と考えられる。

3. テキストの難度

もし、テスト間の難度の差が、それぞれのテストのテキスト間の難度の差だけで説明できるなら、それ以上の分析にはあまり意味がないことになる。ここでは、テスト1～5の難易の差がそれぞれのテキスト間の難易の差と直結していると言えるかどうかを確認するため、テキスト1～5の難度の比較を行う。

『出題基準』で、問題文としてのテキストの難易を測る目安として取り上げられているチェック項目は、テキストの数量的基準に関するもの（テキストの長さ、語彙の逸脱率²⁾、1文の平均の長さ、漢字含有率）と、テキストタイプ³⁾である。今回の分析では、語彙の逸脱率の調査は行わず、その代わりにクローズ・テストの得点からテキストの難易を測る方法をとることにした。つまり、5つのテキストをクローズ・テスト化し、これをある同一の留学生の集団に解答してもらって5つの得点結果を比較することによって、テキストの難易の差を見ようとしたのである。この方法をとったのは、逸脱語彙の特定に主観が入るのを恐れたことと、留学生にとってのテキストの難度は同じ留学生をフィルターとして測定した方がより正確な値に近づけると考えたことによる。クローズ・テストとは、何らかの規則に則って設けられたテキストの空白箇所を埋めていくテストのことで、テキストの難易を判定するのに有効な方法と言われている（Oller1979）が、日本語についても、クローズの得点がテキストの難易判定の基準になり得ることが報告されている（小出1992）。今回の調査では、クローズ単位は字とし、第2文の第1文字から7字ごとに空白箇所を設けた。また、採点は原文一致法をとった。クローズ・テストの被験者は、同じく産能大学の1996年度入学の留学生29名で、国籍の内訳は台湾7、韓国11、中国9、香港1、イラン1だった。テストの難度は受ける集団によって異なるわけであるから、このグループの得点結果がそのままテスト1～5を受けた1995年度入学のグループにとってのテキストの難度を反映しているとは限らないが、この2つのグループはきわめて似通った性質を持っていることから、参考値として十分意味を持つのではないかと思う。テキストタイプについては、2. で述べた通り、5回のテストとも社説からとったため、すべて共通のテキストタイプ（意見表出型の機能を持ち、論理を展開軸とするテキスト）であり、今回の分析対象とはならなかった。

表2は、今回行ったテキスト調査の結果をまとめたものである。

2. で見た通り、5回の読解テストのなかでテスト3の平均点(71%)とテスト4の平均点(43%)の差が最も顕著だったが、この差がテキストそのものの難度によるものの可能性があるかどうか、このテキスト調査で最も確認したいことだった。表2の結果を見ると、テスト3は1文の長ささと漢字含有率は最も高く、テスト4は1文の長さも漢字含有率も全体のなかでは低いほうであることがわかる。テキストの長さについては、テスト3が目立って短い、テキストの短さが易しさの大きな要因になっていたのだとすれば、制限時間(25分)の統一されていたクローズ・テストの結果にも影響するはずである。しかし、実際はテスト3とテスト4のクローズ得点は59%と61%でほとんど差がない。つまり、5回のテストの難易の差は、少なくともテキストそのものの難易では説明できないということが言えるだろう。したがって、テスト間の難易の差は、タスクの性質によるものである可能性が高いということになる。

表2 読解テスト1-5のテキストの難易に関する調査結果

	テキスト名	テキスト長 (字)	1文長 (字)	漢字含有率(%)	クローズ (%)	読解 (%)
テスト1	1990.12.15 朝日新聞社説 電話と上手につきあうには	1289	40	36	52	48
テスト2	1989.11.11 朝日新聞社説 いい時間を過ごしたい	1277	41	37	64	52
テスト3	1990.12.21 朝日新聞社説 子育て支援は三点セットで	1021	46	47	59	71
テスト4	1990.7.17 朝日新聞社説 アダム・スミスの警告	1282	37	36	61	43
テスト5	1992.7.7 朝日新聞社説 転職の時代と日本的経営	1288	34	40	54	51

4. タスクの難度

ここでは、『出題基準』で示されているタスクの分類基準を参考にテスト1～5の50問のタスクを分類し、これらの正答率を比較することによってタスクの種類と難度の関係、また、同じ種類のタスクで難度が大きく異なる場合の要因について探っていく。前述した通り、テスト1～5は同時期に行ったテストではないので、50問の正答率は厳密に言えば本来は横並びで扱うべきものではないが、現実問題としてタスクの分類比較ができるほどたくさん問題を同時期に同一集団に出題するのは現状では無理があり、ここでは、50問をあえて同じ土俵に並べてみた。したがってここでの考察の目的は、データに基づいた結論を得ることではなく、今後の作問の安定化のための示唆的情報を引き出すことである。

4. 1. タスクの分類

『出題基準』では、読解のタスクを「問題の性質」と「問題にする単位」という2つの視点から表3に示すような分類を行っている。また、難度についてはこの表の左上から右下にいくにしたがって難度が上がるとしている。

表3 タスクの種類と難易（『出題基準』p. 228より）

	問題にする単位 問題の性質	表記（音）	語句	文単位	段落・文間	文章・段落間
①	テキスト内の 言語形式について問うもの	文脈での読み方・語の区切り	同一内容語 指示先行詞			
②	テキスト内の 事実関係について問うもの			発話者の特定・行為者、理由、場所、時、方法、対象、目的等を問う・文中人物の意見・意見と事実の区別・時間関係		
③	意味解釈に関するもの		多義表現・比喻表現 同義表現・文脈的意味 省略の復元・意味の推測			
④	展開予測に関するもの			省略復元	省略復元 接続詞挿入 次の文	段落順序 接続詞挿入 次の段落
⑤	推論的なもの		含意・言外の意味・人物の心理・前提			
⑥	テキスト全体からの情報に基づくもの				因果関係把握・事柄の順序 筆者の意見・理由や根拠の 取り出し・要約・内容正誤 小見出し・タイトル付け	
⑦	テキスト機能全体を対象とするタスク					主張・態度 意見・読後 処理・話題

50問の分類はこの『出題基準』の分類に則って行ったが、「問題にする単位」の特定にはいくつか未解決の問題があったため⁴⁾ 今回の分析からははずし、タスクの性質による分類のみを行った。その結果、50問のタスクは上位分類で6種類、下位分類も入れると14種類のタスクタイプに分類された。上位分類が『出題基準』で示されている7種類より1つ少ない6種類になったのは「推論的なもの」（表3の⑤）に該当するタスクがなかったためで、これは社説という明確な主張を持

つテキストの性質からこの種のタスクができにくかったためと思われる。この分類は当該データの50問を『出題基準』に則って振り分けた結果であり、網羅的なタスク分類を示すものではない。また、『出題基準』には実例は示されていないので、分類基準の解釈や下位分類の項目の立て方で『出題基準』の意図とずれているところがある可能性もある。その場合はすべて筆者の責任である。

以下に本稿での分類基準とタイプごとの問題例を記す。問題例についている番号は問題番号（たとえば3-1はテスト3の問1）を示す。

①形式：テキスト中の特定の言語形式について問うもので、特定の語句や文構造知識の有無が解答のキーとなる問題

- ・語句「22行目の「じたばたする」の意味に最も近いのはどれですか。(1-4)」
- ・文型「33行目の「どれだけ気持ちが和らぐことか」の意味に最も近い表現はどれですか。(1-7)」
- ・指示詞「28行目の「その傾向」というのは何を指していますか。(1-5)」

②事実関係：テキスト中で明示されている人・もの・事についての5W1Hで代表されるような事実関係を問うもの

- ・理由「中年層の人があまり休暇をとらない理由としてあがっているのは何ですか。(2-6)」
- ・時「2行目に「百年になる」とありますが、これはいつからいつまでの百年を指していますか。(1-1)」
- ・誰「11行目に「見えざる手による社会全体の調和、発展など仮説にすぎない」とありますが、これはだれの考えですか。(4-2)」

③意味解釈：テキスト中の特定の語句や文の表す意味を文脈の中でどう解釈すべきかについて問うもの

- ・文脈的意味「5行目の「施策として最も安易なもの」とは何ですか。(3-2)」
- ・比喩表現「筆者の言う、「雨」と「雨靴」と「傘」にあたるものは、それぞれ何ですか。(3-7)」
- ・意味の推測「35行目の「過剰サービス」というのはだれに対するサービスのことですか。(5-6)」

④展開予測：文・段落間の関係や展開の予測等について問うもの

- ・接続詞挿入「30行目の□に入れるのに最も適当な言葉はどれですか。(5-5)」
- ・段落順序「下の文章は本文の一つの段落を抜き取ったものです。もともと文中のどの部分にあった文章でしょうか。(1-10)」

⑤情報総合：テキスト全体からの情報について問うもので、テキスト中の複数の箇所的情報を総合したり、比較したりすることが求められる問題

- ・段落話題「市場経済に関する疑問が述べられているのは1～14のどの段落ですか。(4-8)」
- ・複数情報「勤労者の生活にゆとりをもたらすために必要な取り組みとして筆者が挙げているのは、次のどれですか。a～dの中から正しい組み合わせを選んでください。(5-9)」

⑥テキスト機能：テキスト機能（このテキストが書かれた目的）に関わることについて問うもの

- ・筆者の主張「この文章の中で電話について筆者が最も言いたい事は何ですか。(1-9)」

上記の分類基準に則った分類の結果、各種類のタスクの数は表4に示す通りとなった。

表4 読解テスト1-5のタスクの種類と数

タスクの種類	下位分類	テスト1	テスト2	テスト3	テスト4	テスト5	計	*
①形式	語句	2	2	3	2	1	10	(45%)
	文型	1	0	0	1	1	3	(32%)
	指示詞	1	1	0	1	1	4	(45%)
	計	4	3	3	4	3	17	(42%)
②事実関係	理由	0	2	1	0	0	3	(74%)
	時	1	0	0	0	0	1	(71%)
	誰	0	1	0	1	0	2	(60%)
	計	1	3	1	1	0	6	(69%)
③意味解釈	文脈的意味	1	1	3	0	1	6	(76%)
	比喩表現	1	1	1	1	0	4	(64%)
	意味の推測	0	0	0	0	1	1	(10%)
	計	2	2	4	1	2	11	(66%)
④展開予測	接続詞挿入	0	0	1	1	1	3	(40%)
	段落順序	1	1	0	1	0	3	(32%)
	計	1	1	1	2	1	6	(36%)
⑤情報総合	段落話題	0	0	0	1	2	3	(67%)
	複数情報	1	0	0	0	1	2	(48%)
	計	1	0	0	1	3	5	(59%)
⑥テキスト機能	筆者の主張	1	1	1	1	1	5	(59%)
総計		10	10	10	10	10	50	

(*右端のカッコ内の数字は各タスク・タイプ^oの平均正答率を表す)

この結果から、タスクの数は50問中「形式」に関するものが17問で、目立って多いこと、また、「事実関係」「意味解釈」「情報総合」の問題数はテストによってかなりの差があること等がわかった。これらの偏りはテストの難度だけでなく妥当性そのものにも関わってくる可能性のある問題なので要注意と言えよう。

4. 2. タスクの性質と難度

表4の右端の()内の数字は各種類のタスクの平均正答率を示したもので、図1は個々のタスクの正答率を種類別にまとめて正答率の高い順に並べたものである。

これを見ると、同じ種類のタスクでも個々の問題にはかなりの難度差があることがわかるので、一概には言えないが、あえて傾向について見るなら、「形式」と「展開予測」の問題に難しいものが多く、「事実関係」と「意味解釈」に易しいものが多いこと、また、「テキスト機能」のタスクは比較的難度が安定しているということが挙げられるかと思う。

次に、各タスクタイプごとに、下位分類のグループ間に難度の差があるかどうか、また、同じグループ内のタスク間に大きな難度差がある場合、それが何によって生じているのかについて細かく見ていく。

①形式

全体に正答率の低かった形式の問題については、特に30%以下の正答率のもの6問に的を絞ってその要因を探ってみた。

まず、語句の問題の中で最も正答率の低かった問題(1-4:29%, 2-4:24%, 2-8:19%)と比較的高かった問題(3-9:71%, 3-6:67%)を比べてみると、文脈の助けで正解に行き着ける問題かどうかで難度が違っていることがわかった。たとえば1-4は本文中に「・・ときにもじたばたしないで済むような・・」とある中の「じたばたする」の意味に近いものを、選択肢の{aあわてる bこまる cいそぐ dおどろく}から選ばせる問題である。正解は「あわてる」だが、他の選択肢でも文脈には適合するので、もともと「じたばたする」の意味を知っていたのでない限り、正解がわからない。同じ事が2-4と2-8についても言える。これに対してたとえば3-6は本文中に「・・が、雨はしのげない」とある「しのぐ」の意味に近いものを{aまさる bふせぐ cこえる dまもる}から選ばせる問題で、受験者は「しのぐ」という言葉を知らなくても文中の「雨」との組み合わせを見て正解の「ふせぐ」に行き着くことができる。3-9についても同様の文脈上のヒントが見られた。文脈理解が正解につながらない例としては文型の問題の5-3(29%)もあてはまる。「女性のふえ方が急速なことも忘れては□」の□に入るものを{aいかぬ bいかない

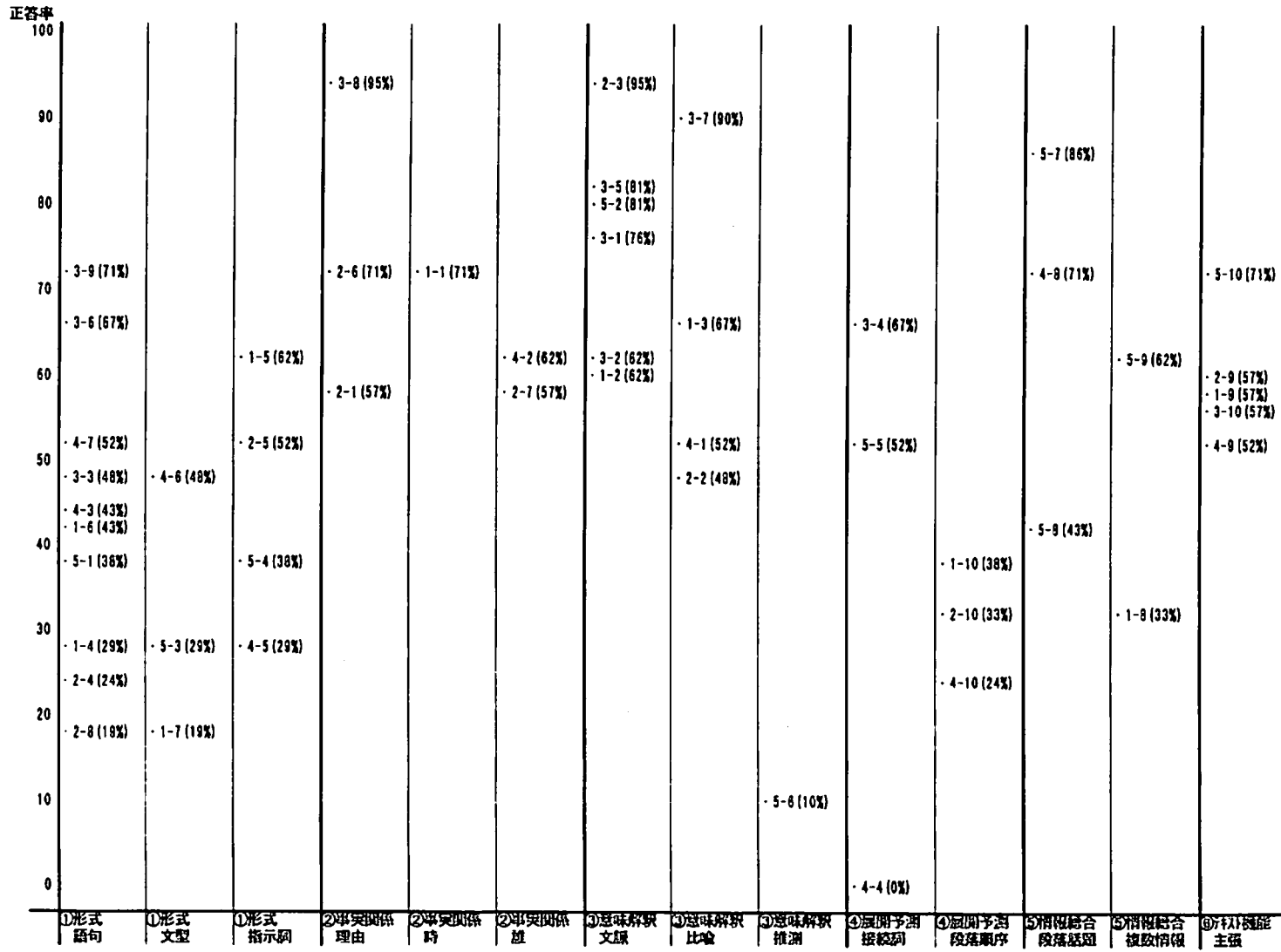


図1 タスク・タイプ別正答率の分布

cなかろう dなるまい)の中から選ばせる問題で、もし文脈を正しく理解し、□には意味のうえで否定がくることがわかっているにもかかわらず、ここで与えられている選択肢はすべて否定を含むものであるから、「～てはなるまい」という表現が文法的に正しいことを知識として知らなければ正解には行き着かない。これら文脈の助けのない問題、言い替えれば本文を読まなくてもできる人にはできる問題の難度の高さは読解の問題として「妥当性を欠いた難しさ」と言えるだろう。

30%以下の正答率の残り2問(文型1-7:19%, 指示詞4-5:29%)の難度の高さは、1-7(4章1節のタスク分類の①問題例参照)については疑似疑問文による詠嘆表現という文型の留学生にとっての難しさがそのまま反映されたもの、4-5については、正解以外の選択肢に正解に非常に近い、いわゆる「まぎらわしい選択肢」があったことによるもの(21人中9人がこちらを選択している)と思われる。

②事実関係

事実関係に関する問題は全体に正答率が高いが、その中で、どちらも理由の問題なのに正答率の目立って高い3-8(95%)と比較的低い2-1(57%)を比較してみた。その結果、3-8は正解の選択肢だけに本文中の表現が使われているのに対して、2-1は正解以外の選択肢にも本文中の表現が使われているという違いがあることがわかった。つまり、3-8は本文との単純なマッチングで正解に行き着けるが、2-1ではそれはできないことが難度の差となっていると思われる。

③意味解釈

意味解釈の問題も全体に正答率が高い方だが、特に高い2問(文脈的意味2-3:95%, 比喩表現3-7:90%)はどちらも本文とのマッチングで正解に行き着ける問題になっていた。他方、意味解釈の問題の中で極端に正答率の低い5-6(10%)は唯一、意味の推測に分類したタスクで、本文中の「過剰サービス」という言葉がだれに対するサービスを指すのかを{a客 b傷害者 c働く人 d雇う側}から選ばせるものである。本文の中では雇用環境の改善のために必要なポイントの1つとして「過剰サービスの見直し」が挙げられているだけで、この言葉の意味についての説明はいっさいない。意味の推測のタスクで求められるのはテキストに明示されていない情報を読み手の世界知識(テキスト外知識)で補って文脈に即した解釈をすることであるから、読み手の既存知識の中に必要な情報がない時は難度が非常に高くなる。5-6の場合は、「過剰サービス」と言えば一般に客や消費者に対するものを指すということが受験者にとって自明の理ではなかったために難度が高くなったものと思われる。当然持っているべきテキスト外知識として読み手に何を期待するかは、かなり慎重に判断すべきだろう。

④展開予測

展開予測の問題は接続詞挿入(段落と段落の間に接続詞を挿入するタスク)と

段落順序（もともと本文の一部だった段落を1つ別に取り出しておき、もとの場所にもどすタスク）の2種類がそれぞれ3問ずつあった。段落順序の方は3問とも正答率が低く(1-10:38%, 2-10:33%, 4-10:24%)、タスクの種類としてかなり難しいと言えそうである。段落順序のタスクが、もともとはそこにあった段落を抜いておいて元に戻させるという通常の読解では求められないスキルを要求していることを考えれば、この難度の高さは納得できる。この種のタスクの有無がテストの難度に与える影響も大きいと見られ、特に高度な読解能力を測定したいのでなければ出題は控えた方が無難かもしれない。一方、接続詞挿入のタスクは3問中2問の正答率は50%を越えていた(3-4:67%, 5-5:52%)が、1問(4-4)は何と、正答率0%だった。これは50問中最低である。ここまで難度が上がってしまった理由としては2つのことが考えられる。1つは前の段落で留学生にとっては難しい文法項目である反語や倒置が使われているため、段落と段落の関係がつかみにくくなっていること、そしてもう1つは正解の選択肢以外に正解と同じ逆説の接続詞があったこと（正解は「しかし」だが、他の選択肢に「ところが」があったため、逆説であることまではわかった受験者も「ところが」の方を選んでしまっている。接続詞の問題で最も正答率の高かった3-4は、4つの選択肢の中で逆説は正解の「だが」1つだけだった）ことである。2つ目の問題は①で触れた「まぎらわしい選択肢」の問題に通じるものがあると思われる。

⑤情報総合

段落話題（ある話題について述べている段落はどれか、探させるもの）の3問を比較してみると、キーワードとのマッチングで誤答を選んでしまっているケースが見られた。最も正答率の低かった5-8(43%)は「今後、労働力がどのように構造的に変化するかについて述べている段落」を選ばせる問題だったが、単に「構造的」「変化」という言葉が出ている段落を選んでいる学生が多数いた。正解以外の選択肢がキーワードや本文とマッチングできる場合は誤答に導かれやすくなり、その結果、タスクの難度は上がると考えられる。

複数情報（テキストの中の複数箇所の情報を総合して答えることが求められるタスク）のタスク2問の難度の差(1-8:33%, 5-9:62%)も、マッチングの問題がからんでいるように思う。5-9が「・・・に必要な取り組みとして筆者が挙げているのはどれですか」という問題であるのに対し、1-8は「筆者が最近の電話の問題として批判していないものはどれですか」という否定形を含む質問になっている。どちらも選択肢の情報と本文とを丹念にマッチングしていけば正解に行き着けるはずだが、そこにはないものを探す方があるものを探すより煩雑さが増し、難度が上がると思われる。

⑥テキスト機能

この種類のタスクは全テストに1問ずつ筆者の主張に関する問題が出題されているが、タスク間の難度幅は①～⑥の中では狭い。5問中4問の難度が50%台で5-10だけが71%になっているが、問題の設定のしかたや選択肢の内容からは、特に難度差の理由になっているものは見あたらなかった。

以上、見てきたことをテストの作成、改訂の際の注意点としてまとめると、次のようなことが言えるかと思う。

- (1)形式の「語句」「文型」に関する問題は、文脈情報が解答に生かせない場合、難度が高くなりやすい。この種の問題は文脈理解が正解につながらないという点で読解問題として妥当性を欠くので極力出題すべきでない。
- (2)意味解釈の「意味の推測」のようにテキスト外知識の重要性の高い問題は難度が高くなりやすい。受験者に期待できる既存知識を考慮する必要がある。
- (3)展開予測の「接続詞挿入」の問題の難度は挿入する接続詞そのものの難度ではなく、接続詞によって結ばれる段落と段落の関係のわかりやすさによって左右される。接続詞の挿入箇所の前後で使われている文型の難度にはよく注意しなければならない。
- (4)展開予測の「段落順序」の問題はかなり難度が高く、また、通常の読みでは求められないスキルであるから、出題には慎重を要する。
- (5)情報総合の「複数情報」は、「・・・ていないものはどれか」のように否定を含む問題の方が難度が高くなる。
- (6)正解以外の選択肢に、正解に近いまぎらわしい選択肢がある場合は、難度が高くなる。
- (7)本文や問題文中のキーワードとのマッチングで正解が得られる問題は難度が低く、逆に正解以外の選択肢でマッチングができる場合は難度が高くなる。

5. おわりに

今回の分析は、先にも述べた通り、ごく限られた、しかも条件的に均一とは言えないデータを対象としたものだったので、分析結果を一般化可能な法則として結論づけることはできない。しかし、今後の作問・改訂の足がかりとしては、妥当性を保つために排除すべき問題の種類や、難度を安定させるために選択肢で均一にすべき条件等について、有益な情報を得ることができた。今回のような、タスク分類を軸としたタスク間の難度の比較分析は、テスト得点というデータを作問・改訂に活用可能な参考値として残していくための枠組みとして非常に有用であると思う。今後は、まずは今回の分析結果に基づいたテストの改訂が急務であ

るが、その後も長期的に改訂・分析をくりかえしながらデータを蓄積し、タスクの具体的な分類基準やタスクの難度要因についてさらに検討を重ねていく必要がある。

付記：『出題基準』の著者のお一人である小出慶一先生には、タスクの分類についての質問にお答えいただき、また、同僚の鬼木和子先生にはテストの統計処理で助力をいただきました。この場を借りてお礼を申し上げます。

註

- (1)「日本語能力測定」についての詳細は参考文献(3)参照。
- (2)語彙の逸脱率とは『出題基準』の語彙リストからの逸脱率のことで、テキストの全文字数に対するリスト外字数の割合をいう。ただし、語彙リストに示されている語彙は当該レベルの80%ということになっているので、リストにない語が当該レベルを越えているかどうかの判断は当事者にまかされることになる。
- (3)テキスト・タイプとはそのテキストの持っている機能と展開の中心になる展開軸によって分けられたテキストのタイプのこと。詳しくは参考文献(9)参照。
- (4)問題にする単位は、それを①出題の直接の対象とする単位（テキスト本文に下線で示されているようなところ）をとるか、②解答のため処理が必要と想定される単位をとるかで変わってくる。また、②の場合は正解にたどり着く方法が複数考えられる場合は単位も複数になる可能性がある。

参考文献

- (1)青木昭六編(1985)『英語の評価論』大修館書店
- (2)石田敏子(1992)『入門日本語テスト法』大修館書店
- (3)鬼木和子・斉山弥生(1994)「大学生のための日本語テスト」『平成6年度日本語教育学会秋季大会予稿集』p. 9-13
- (4)海保博之編著(1985)『心理教育データの解析方10講基礎編』福村出版
- (5)垣田直巳監修・松村幹男編集(1984)『英語のリーディング』大修館書店
- (6)カーマイン,E・ツェラー,R 著(1979)、水野欽司・野嶋栄一郎 訳(1983)『テストの信頼性と妥当性』朝倉書店
- (7)小出慶一(1991)「読解能力の操作的規定と読解テスト・シラバスの骨格について」『産能短期大学第24号』p. 181-193
- (8)小出慶一(1992)「クローズ・テストの得点とテキストの難易判定」
The Language Teacher 16-6, p. 11-15
- (9)国際交流基金・(財)日本国際教育協会(1994)『日本語能力試験出題基準』凡人

社

- (10)日本語教育学会(1991)『日本語テストハンドブック』大修館書店
- (11)羽島博愛編(1979)『英語指導法ハンドブック4 評価編』大修館書店
- (12)Bachman, L. (1990). *Fundamental Considerations in Language Testing*. O. U. P.
- (13)Bachman, L. and Palmer, A. (1996). *Language Testing in Practice*. O. U. P.
- (14)Dubin, F., Eskey, E. and Grabe, W. (1986). *Teaching Second Language Reading for Academic Purposes*. Addison-Wesley Publishing Company, Inc.
- (15)Feedle, R. and Kostin, I. (1993). "The prediction of TOEFL reading item difficulty: implications for construct validity" . *Language Testing* 10-2, pp. 133-170. Edward Arnold.
- (16)Grellet, F. (1981). *Developing Reading Skills*. C. U. P.
- (17)Huges, A. (1989). *Testing for Language Teachers*. C. U. P.
- (18)Lumley, T. (1993) "The notion of subskills in reading comprehension tests: an EAP example" . *Language Testing* 10-3, pp. 211-234. Edward Arnold.
- (19)Oller, J. W. Jr. (1979). *Language Tests at School*. Longman.
- (20)Perkins, K. (1992). "The effect of passage topical structure types on ESL reading comprehension difficulty" . *Language Testing* 9-2, pp. 164-172. Edward Arnold.
- (21)Rost, D. (1993). "Assessing different components of reading comprehension: fact or fiction ?" . *Language Testing* 10-1, pp. 79-92. Edward Arnold.

(産能大学)

〈資料〉 読解テスト問題用紙（テスト 3 の抜粋）

下の文章を読んで、後ろの間に答えてください。

1. 政府と野党の「子育て支援策」が相次いでまとめの段階に入った。表向きはともかく、現実には、出生率を上向きにするための施策である。だがこれまでの経過を見ると期待通りの効果が得られるか、危ぶまれる。

子を産む当事者である女性たちや保育の現場の人々の切実な声に耳を傾けることを怠り施策として最も安易なものから手をつけようとしているからだ。

出生率激減を示す厚生省人口問題研究所の報告以来、子育てに関する委員会が続々誕生した。

（中略）

その先頭を切って自民党小委が児童手当を第一子から支給する改革案をまとめ、厚生省は児童手当法改正の準備に入った。

（後略）

〔朝日新聞 社説 1990年 12月21日〕

(1) 3行目の「期待通りの効果」とはどんなことを指していますか。

- a 女性や保育の現場で働いている人の声が反映されること
- b 「子育て支援策」がまとまること
- c 出生率が上がること
- d 子育てに関する委員会ができること

(2) 5行目の「施策として最も安易なもの」とは何ですか。

- a 出生率の報告書を作成すること
- b 児童手当を増やすこと
- c 保育の専門家を育てること
- d 労働時間の短縮を行うこと