

## コンピュータを使った適応型日本語絶対評価システム : J-CAT 2005 Version\*

Adaptive and Absolute Evaluation System on Japanese Language with a Computer: J-CAT  
2005 Version

今井新悟

### 要旨

拙稿は留学生に対する日本語の目標基準準拠テストであるコンピュータによる日本語アダプティブテスト (Japanese Computerized Adaptive Test、以後 J-CAT と記す) の開発に係る 2005 年度時点での報告と考察を行おうとするものである。J-CAT の開発に着手するに至った経緯については今井 (2005) で述べた。本稿では、その記述を引用しつつアダプティブテスト採用に至った背景を振り返ったのち、開発を通じて明らかになった問題とその対策、プレースメントテストとしての実施状況、今後の課題について述べる。

### キーワード

アダプティブテスト、日本語能力、テスト、絶対評価、目標基準準拠テスト

## 1. J-CAT 開発の理由

### 1-1. 大学における教育の質の保証

国立大学の法人化と少子化による大学全入時代の到来は、過去 1 世紀にわたって連綿と受け継がれてきた大学の「常識」に劇的な変革を迫っている。その流れのなかで民間企業であれば当然の「品質保証」がようやく大学の教育にも必要であると認識されるようになった。

これまでの大学の授業は教員の純然たる自治の場であり、聖域であった。教育の内容も評価も教員個人が基準であった。どの教員が評価が甘く、単位が取りやすいかが学生の授業選択時の指標になりもした。しかし、いまや大学時代をモラトリアムと自他ともに認めた時代は去り、授業に対する意識も変わりつつある。企業も採用後に再教育をする余裕があった時代は去り、採用時での能力や資格を見極めようとする企業の姿勢が強くなってきた。これを受けて学生が資格獲得に敏感になってきた。単位がとれて卒業証書がもらえればよいという認識ではなくなりつつある。大学側も、授業シラバスを公開し、聖域であった授業の内容を開陳するようになった。シラバスには到達目標を記述することになっている。このように学生も大学も教育の品質について意識するようになったが、民間企業レベルの品質保証には至っていない。なぜなら評価の方法、すなわち品質のチェック体制が不十分であるからだ。

評価の方法は多種多様であるが、ここでは数値として表すことのできる、いわゆる「テスト」を取り上げる。テストには集団基準準拠テスト (norm-referenced test) と目標基準準拠テスト (criterion-referenced test) がある。例えば大学入試は前者であり、資格試験は後者である。品質の保証には後者が使われる。到達目標を明らかにし、それを基準にし、絶対的な評価を行うことによって品質が決定される。では、大学の授業の学期末に行われる評価が目標基準準拠テストといえるものになっているかという点極めて心もとない。筆者らの所属する山口大学国際センターでは、留学生に対する評価を目標基準準拠テストで行うための取り組みを続けている。

---

\* J-CAT2005 は山口大学国際センター (宮崎充保、渡辺淳一、今井新悟、杉原道子、赤木弥生、門脇薫) e ラーニングサービス (秋山實) 島根大学 (中園博美) 徳山大学 (重田美咲) の共同研究開発プロジェクトである。

### 1-2. 留学生の日本語プレースメントテスト

日本の大学では、異なった日本語レベルの留学生を受け入れている。学部学生は「日本留学試験」の受験が義務付けられているので、ある程度の日本語能力は確保されている。それでも、以前に課せられていた「日本語能力試験」に比べて、「日本留学試験」で保証される日本語能力は低下しているという指摘もある。一方大学院生には、一律の日本語能力を証明する試験は課せられておらず、当然、日本語能力にもばらつきが生じている。文部科学省奨学生には6ヶ月間の大学院予備教育期間が設けられており、初級日本語の習得は保証されているが、圧倒的多数を占める私費留学生にはそのような機会の保証はない。また、大学院生への前段階としての研究生は身分的にも不安定であり、日本語を自力で学習するしかない者もいる。他方、交換留学生として半年から1年の期間でやってくる留学生のなかには、正規学部留学生以上に日本語ができる学生がいる一方で、まったく日本語ができない学生も混じっている。このように混沌とした状況の中で、各大学では留学生用の授業を設けて日本語能力を伸ばそうとしているが、その際にまず課題となるのが、日本語能力に従ったクラス編成を行うこと、すなわちプレースメントテストを行うことである。ここでの課題を今井(2005)では、以下のように指摘した。

(旧来のペーパーによるプレースメントテストでは)第一に、問題の信頼性が保てないことが最大の問題として指摘された。毎回の素点が、日本語能力の客観的基準にならないということである。例えば、点数が前回のテストよりも10点上がったとしよう。果たして、それは、その個人の能力が10点分伸びたのか、それとも、テストが少し易くなったのか、判断できない。偏差値を使っても、そもそも母集団の能力値が一定であるという保証がないのであるから、信頼性の問題は解決しない。第二に、時間の制限があり、十分な数の課題を解かせることができない。これは、試験全体を例えば2時間と設定したとしても、初級者も上級者も同じペーパーテストを受験させるため、それぞれの能力に合った設問数が限られてしまうことによる。第三に、第二の問題と関連するが、上級者にとっては、初級の問題は意味がなく、逆に、初級者にとっては、全く歯が立たない設問の山をつきつけられて、相当な心理的な負担を強いられる。第四に、研修生が、五月雨式に本学に到着する。それに個々に、ペーパーテストを監督しながら実施することは、実施者にとって負担が多すぎる。(今井 2005; 67)

上記の課題を解決するためには、テストは目標基準準拠テストでなくてはならなかった。そして、テストの結果がそのつどぶれるものではなく、つまりテストの信頼性を得るために古典的な平均点や偏差値にとって替わる新たなテスト理論による能力の測定を必要とした。また、限られた時間のなかでテストを行うことができ、その結果が信用にたるものでなければならなかった。さらに、出題される問題は受験者にとって難し過ぎず、また簡単過ぎず、ちょうどよい難易度のものがふさわしかった。これらのことを可能にしてくれるのが、項目応答理論(Item Response Theory 以下IRT)に基づくアダプティブテストであった。そして我々はこれをコンピュータ上で実施可能にし、インターネットでつなぐことで、時間と場所との制約も取り払い、コンピュータさえあれば、いつでもどこでも受験できるシステムの開発を目指すこととなったのである。

米国の大学入学資格試験であるSATや、外国人が米国に留学する場合に課せられるTOEFLなどは、項目応答理論に基づいた試験の例である。いつ受験してもその点数をもって受験者の能力を示しているといえる。TOEFLにはコンピュータを利用したCBT(Computer Based Test)と昨年からの導入が始まった新たなiBT™(Internet Based Test)も開発されている。ただし各国のテストセンターの運営費に莫大な経費を必要としており、費用対効果の面からは課題を残している。いずれにせよ、これらはすべてEducational Testing Serviceによって開発された国レベルの事業である。翻って日本国内をみると、大学人にとって最大の関心事であろうセンター入試においては、いまだ項目応答理論も導入されておらず、その結果、毎回のよう科目間による点数調整が実施されるか否かが問題となっている。このような状況において、我々自身が項目応答理論によるアダプティブテストに取り組もうとするのは、勇気のい

る決断ではあった。しかし、前節でも述べたように、品質の保証が求められ、客観的な基準による評価が求められる現在、遅かれ早かれ必要となる評価システムであることを自覚し、その構築に実験的に取り組むことになった。

## 2. 開発の方法

### 2-1. 出題基準の設定

まず、出題基準の設定を行うにあたり、「日本語能力試験」および「日本留学試験」の出題基準について検討した。その結果、後者は採用できず、前者も不十分な点があるが、現段階ではこれに替わるものが見当たらず、また、新たに作成する余力もないことから、我々の J-CAT では、前者を踏襲することとした。以下に両試験のシラバスを概観し、前者を踏襲することにした経緯を示す。

#### 2-1-1. 日本留学試験の出題基準

「日本留学試験」は、日本の大学の学部留学をめざす人のために実施される試験で、科目は「日本語」、「理科」、「総合科目」、「数学」の四科目である。そのうちの「日本語」のシラバスには、以下の項目が挙げられている。

I 測定対象能力、II 試験が想定する課題の種類、III 試験で要求される言語技能、IV 試験で要求される言語下位技能、V 試験に含まれる表現類型（テキストタイプ）VI 試験にあらわれる話題（トピック）VII 試験にあらわれる場面（場所、時間、対人関係、対人形態、媒体など）VIII 試験にあらわれる言語的要素（テキスト関連）IX 試験にあらわれる非言語的要素（視覚情報関連）X 試験の課題を達成するために前提となる知識

これらの項目ごとにさらに記述がある。しかし、その記述は以下のようなものであり、概略に過ぎず、具体的な記述にはなっていない。ここでは、「III 試験で要求される言語技能」および「IV 試験で要求される言語下位技能」を例として示すが、他の項目においてもこれらと同様に、いわば常識の範囲を超えない簡単な記述にとどまっている。

#### III. 試験で要求される言語技能

日本での留学生活の中で遭遇する諸課題を解決するために必要となる言語技能を以下に示す（\*「話す」「翻訳する」については部分的ないし間接的測定となる。）

聞く、話す \*、読む、書く、翻訳する \*

#### IV. 試験で要求される言語下位技能

上記の5つの言語技能の下位技能として、以下の類型が考えられる。

情報の全体の流れをとらえる。

情報の全体をある判断や評価をしながらとらえる。

特定の情報を抽出してとらえる。

推測しながら情報をとらえる。

予測しながら情報をとらえる。

その他

この試験の目的としては以下が掲げられており、これは我々の目指すものと軌を一にする。しかしながら、上記のようなシラバスしかない場合、どのようにして問題項目の難易度をコントロールするのか、またそれができるのか疑問である。

#### この試験の目的

この試験は、外国人留学生として日本の高等教育機関、特に大学学部留学を希望する者が、日本の大学での勉強に対応できる日本語力（アカデミック・ジャパニーズ）をどの程度習得しているかをシングルスケールで測定することを目的とする。

なお、試験は、聴解、聴読解、読解及び記述により出題する。

この試験は、項目応答理論に基づく得点等化を行っていることも明らかにしており、(同ホームページ「日本留学試験における得点等化について) シングルスケールで測定するという目的も、まさに我々の目指すところである。しかし、上記のシラバスに従えば、この試験は日本の大学での勉学に対応できるかどうかを識別の焦点としたものであり、前節で紹介したように、各大学に在籍する対象者は初級から上級までと幅広いのであるが、これら日本語学習者をすべてシングルスケールで測定しようとするものではない。

日本留学試験自体はアカデミックジャパニーズの力を測定しようという点で魅力的なものであり、現在の学部留学生に課されている試験でもあることから、これにそった J-CAT を作成したいという希望はあったものの、それは断念せざるを得なかった。現実的な問題として、ここに示されたようなシラバスでは、初級から上級の各レベルに合わせた問題アイテムの作成はできないからである。

### 2-1-2. 日本語能力試験の出題基準

「日本語能力試験」は日本語を母語としない日本語学習者の日本語能力を測定するための試験であり、「文字・語彙」、「聴解」、「読解・文法」の3セクションで構成される。1級から4級までの級別になっており、1級が上級、4級が初級に相当する。問題作成者向けに200ページ以上におよぶ『日本語能力試験出題基準(改訂版)』があり、1級から4級までの出題基準が示されている。各級での語彙、漢字、文法形式のリストが示されている。各レベルでの漢字数と語彙数は次のように定められている。4級漢字103字、3級漢字284字、2級漢字1023字、1級漢字2040字、4級語彙749語、3級語彙1441語、2級5035語、1級8009語。各級の漢字数、語彙数は当該級以下の累積数である。たとえば、3級漢字の284字は4級で習うべき103字と3級での新出漢字181字の計になっている。

語彙のリストは各級の試験で使われる語彙のおよそ80%をカバーするものと規定されている。「文字・語彙」のセクションについては、これらのリストに従って問題が作られているので、量的ガイドラインが明らかになっているといえる。読解問題では語彙リストからの逸脱率として、1級10%以内、2級8%以内、3級および4級5~6%以内と規定されている(『出題基準』p228)。ここでの逸脱率については、以下のように説明されている。

「語彙リスト」そのものが、当該級の語彙数の80%を示すということになっているので、すでに20%の逸脱は許容されていることになる。従って、ここでの逸脱は、出題者が当該級以上と判断する語彙の割合を意味することになる。(『出題基準』p228)

このように読解においても漢字・語彙の量的ガイドラインが示されている。しかし、読解における文法項目の逸脱度やガイドライン、「聴解」における漢字、語彙、文法項目の逸脱度やガイドラインなどの規定はない。さらに、そもそも文法項目については、そのリストが網羅的なガイドラインの体裁をなしているのは3級および4級のみであって、そこでは文法項目の逸脱度を20%程度までと規定してあるのに対し、1級と2級については、当該級で学習すべき文法項目を列挙しているのではなく、それは「サンプル」に過ぎず、「問題作成者の判断」において当該級の水準に相当すると認められる文法については、当該級および当該級より下位級のリスト外から任意に出題することができる。(『出題基準』p.ii)という説明があるに過ぎない。このような規定に従って問題作成を行っていくにあたって直面する問題は「問題作成者の判断」が一様でないことである。「問題作成者の判断」がもともと一様であるならば、各級別のリストは不要なのである。その判断が揺れるからこそ、リストやガイドラインが必要になる。そのガイドラインが、「問題作成者の判断」にまかすという記述では、そもそもガイドラインとして機能することはありえない。

このように、一見周到に用意されているように見える『日本語能力試験出題基準』ではあるが、実際の問題作成の指針としては不十分である。(もっともガイドラインやリストを示すことそのものが根本的に妥当であるかどうかの議論はあろうが、今はそのことには立ち入ら

ない。)そこで我々は、漢字、語彙および3級・4級の文法については、このリストを全面的に踏襲しながら、「問題作成者の判断」をそぎ落として、まずは「読解」の問題作成のガイドラインを作成することにした。

### 2-1-3. J-CAT の出題基準：読解問題作成ガイドライン

J-CAT の読解問題作成基準は『日本語能力試験出題基準(改訂版)』よりも厳しい縛りをかけて、問題作成者による問題のプレを極力抑えようという意図に基づいて作成された。以下に引用する。

#### 【出題基準(級分け)】

原則として国際交流基金・日本国際教育協会『日本語能力試験出題基準(改訂版第2刷)』に従って作成する。

#### 【語彙(選択肢・注・出典を含む)】

注をつければ、すべての級において、当該級外の語彙使用可。ただし、注に使用する語彙も基準に従うこと。

4級：注をつけた4級以外の語彙を一つ使用可。注を付けない4級以外の語彙は使用しない。

3級：注をつけた3・4級以外の語彙を一つ使用可。注を付けない3・4級以外の語彙は使用しない。

2級：注をつけた2・3・4級以外の語彙を二つ使用可。注をつけない2・3・4級以外の語彙は一つ使用可。

1級：注をつけた4～1級以外の語彙を三つ使用可。注のない4～1級以外の語彙は異なり語数で2割を限度に使用可。

#### 【漢字(選択肢・注・出典を含む)】

漢字およびその読みについては、基準の範囲を超えないこと。(漢字の級の判定は、語彙表ではなく、漢字表で行う。例えば、語彙表では「かえず(返す)」が4級になっている。この場合の(返す)は参考のための漢字表記であり、「返」の漢字は、2級の漢字表に載っているため、2級漢字である。)級外の漢字を使う場合はルビを振る。漢字があることによって、漢字圏の受験生に有利になるので不要な級外漢字は原則としてひらがなで表記する。以下の場合には、漢字を使うことも可であるが、なるべく避ける。(1)特に1・2級で、表記の自然さのため。(2)表記のバランスのため。例えば木が4級、金が3級だとする(実際は金も4級)。すると、「木よう日と きんよう日に かいものをしました」となるが、両方とも漢字にして、金にルビを振りバランスをとることも可。(1)文字数を減らす必要のある場合。

#### 【固有名詞】

以下のものは、級を問わず例外的に使用可。ただし、級外漢字にはルビをつけること。東京、大阪、アメリカ、イギリス、日本、日本語、田中、山田、木村、ロン、カム、アン、さくら大学、富士大学。

2級・1級では、固有名詞が問題を解く場合のキーワードにならない限り使用可。問題を解く場合のキーワードになってないとは、固有名詞部分を伏字(×××)にしても解答できる場合を言う。

#### 【字数】

問題文(選択肢・注・出典を含む)の字数は以下に従う。

4級	100字～200字	3級	150字～300字
2級	200字～400字	1級	300字～600字

## 2-2. システムの開発

### 2-2-1. 構成と判定方法

J-CAT では多層構造の項目固定型多段階テスト（芝 1991: 81）を採用している。その理由は以下の通りである。アイテムバンクに蓄積しておく項目数が少なく済むこと。よって、テストの構築が比較的楽にできること。テストの用途が当面プレースメントテストや年2回程度の少ない回数での日本語能力の判定に限られていることから low-stakes なテストであり、レベルをあまり細かく設定する必要がないこと。

現システムの構成と判定方法は以下の通りである。「語彙」、「文法」、「聴解」、「読解」の4セクションからなる。このうち、「語彙」、「文法」、「聴解」は研究目的のため使用許可を得て、日本語能力試験の過去問題の中から、古典的理論による困難度と識別度の値に基づいて良問を選んで使用している。「語彙」については、日本語能力試験には含まれる漢字の問題を除外している。これは漢字の問題が中国語母語話者にとっては必ずしも日本語の能力を測りえないと考えたからである。「読解」問題については、共同研究者を含めたグループで、オリジナルの問題を作って使用している。オリジナルの問題は、「問題作成基準」によって各作成者によって作成されたあと、他2名で互いにチェックしあい、必要に応じて修正したあと集積する。それをプレテストとして編集し、ペーパーでの問題アイテムのセットを作る。それを用いて、国内外の日本語教育機関の協力を得て、プレテストを実施する。解答パターンのデータを集積して、PROX 法による分析を行い、IRT の困難度1パラメータの値を算出する。また、同時に古典的識別力パラメータをも算出している。この分析・パラメータの算出には、TDAP（大友 1996、大友、中村 2002）を改良したバージョンを使用している。こうして得たパラメータ値を参照して適問を選びだして、システムに実装する。

システムには、オープンソフトウェアの e ラーニングシステムである Moodle を利用している。ただし、Moodle そのものには我々のもとめるアダプティブテストの機能がないので、テスト機能の開発・改良を経て、J-CAT を実現させている。

### 2-2-2. アルゴリズム

現 J - CAT は上述のとおり項目固定型多段階テストを採用している。システムの設定により、受験者を任意のレベル数に分けることができる。現行では 20 段階に分けている。レベル当たりの問題数、レベル合格ライン、レベル合格時の移動レベル数 (Mup)、不合格時の移動レベル数 (Mdown) はそれぞれ任意で設定できるようになっており、これらアダプティブテストの実行条件を設定し、以下のようなアルゴリズムで能力を判定している。

- (1) 指定されたレベルから開始する。
- (2) レベルの合否判定が終わるまで同一レベルから出題する。
- (3) レベル N を合格したら  $N + Mup$  のレベルを出題する。
- (4) レベル N が不合格なら  $N - Mdown$  のレベルを出題する。
- (5) 既に合格したレベルより下のレベルは出題しない。
- (6) 既に失格したレベルより上のレベルは出題しない。
- (7) 解答するレベルがなくなったら修了し、合格した最高レベルを受験者のレベルとする。

現行では、3問をひとつの testlet とし、うち2問正解で当該レベルを合格としている。また Mup を + 3、Mdown を - 2 に設定して実施している。

### 2-2-3. 日本語表示の問題と対策

日本語の表示をコンピュータ上で行う場合、まず、Operation System が日本語に対応しているかどうかによってその日本語環境が大きく異なる。国内はともかく、国外からのアクセスをも前提として開発するとき、この点は確実にクリアしておかないと使い物にならないシステムになりかねない。Microsoft の Windows XP からは多言語対応になっているが、個々の

エンドユーザーが日本語を表示できるように設定している確証はない。またブラウザも Microsoft Internet Explorer である保証もない。このように未知数が多い場合、特定の OS やブラウザに依存することは危険であり、避けねばならない。また、日本語能力試験では、教科書体というフォントが使われる。このフォントは最近ようやくワープロソフトに標準装備されるようになりつつあるが、まだ一般的ではない。さらに日本語のテストではルビという特殊な表示形態が必要となる。このルビの HTML タグは Microsoft Internet Explorer ではサポートされているが、それ以外のブラウザには対応していない。このような現状でたいの環境でも不具合が生じないようにするために、J-CAT では問題文を画像として提示することとした。画像であれば、OS やブラウザへの依存は避けられるし、フォントもルビも作成者側で準備すればことが足りる。

このようにして開発方針は決定され、開発されたが、ルビ表示についての課題が残っている。それは、ルビ表示が潰れる現象を解決できないことである。市販のフォントをすべて試してみたが、ルビのサイズとして 6 ポイント程度にすると、ルビの細かなところが潰れてしまう。これについての最終的な解決方法は、ルビ用フォントを自前で作成することであろう。

#### 2-2-4. 画面のコピー・問題の流出防止策

インターネットを介したアクセスを許している限り、問題の流出の心配は払拭できない。物理的に完全な問題の流出を防ぐことは不可能である。たとえばスクリーンをコンピュータ上で、あるいは外部媒体（デジタルカメラなど）でキャプチャーすることは防ぎ得ない。しかし、少なくとも右クリックで簡単に画像をコピーできる、あるいはブラウザのキャッシュにデータを残すということは防ぎたかった。そのため、Macromedia Flash で画像（すなわち問題）を配信し、プログラムに若干の細工を施して簡単なコピーができないようにした。J-CAT のような low-stakes な試験では、問題項目流出防止の満足のいくレベルに達していると考えている。

#### 2-2-5. 音声のストリーミング化

音声問題もインターネットで配信するが、これはストリーミング化し、負荷を抑えている。ただし、この負荷量と質は反比例の関係にあり、現在、どの程度の環境で問題が生じるのか実証している段階であり、ブロードバンド用、ナローバンド用に区別して、ストレスのない配信ができるようにしていく予定である。

### 3. 実施結果：フィードバックと対応策

これまでにプレースメントテストとして、学内で 2 度実施した。2005 年 4 月には 116 名、10 月には 82 名が受験した。大部分の者は学内 2 キャンパスの大規模コンピュータ室で受験した。渡日が遅れるなど、なんらかの事情で指定日に受験できなかった者は、後日、国際センターのコンピュータ室で個々に受験した。受験者個々に User Name と Pass Word を発行して、受験者を識別している。

数例、実施中にシステムのトラブルが生じていたが、その原因は不明である。おそらく、当方の想定していない操作をしたものと思われるが、いかなる場合にも不具合が生じないような堅牢なシステムに向けた改良が必要である。

平均して 30 分から 45 分で全セクションを終了していた。ペーパーテストよりもはるかに短い時間で判定ができる。テスト終了と同時に判定結果がセクションごとの成績ならびに総合成績で示されるのも好評であった。

これまでに寄せられた受験者の感想では、以前のペーパーテストよりも簡単になったというものが多かった。これは、自分に合った難易度の問題が提示されるというアダプティブテストの仕組みがそのような感想を抱かせているのであろう。この点では、受験者の心理的負担を軽減するという当初の目論見のひとつはクリアしているといえる。

このテストを使った教師の感想では、以下の点が指摘された。能力の高い受験者と能力の低い受験者の判定の信頼性が低い。読解の判定が甘く出ている。システムに慣れていず、本

来の実力を発揮できない者がいた。

まず、能力の高い受験者と低い受験者の判定の問題であるが、能力の高い受験者の場合は天井効果が考えられる。問題項目が日本語能力試験の1級を目安に作られていることから、それ以上の実力については測定できない。また、能力の低い受験者に関しては、4級のレベルの問題であっても個々の問題アイテムには相当数の語彙・漢字・文法項目が含まれるのだが、その一部に未習得部分があると、それが致命的な影響を与えて正解にたどりつかないということが考えられる。能力が高くなるにつれて、未習得部分を文脈から推量する能力に長けてくるのだが、初級者にはその能力がないことにより、合否の判定が極端になる傾向があると考えられる。4級レベルのなかでも学習を始めたばかりの者に対する問題作成方法の再検討が課題である。あわせて、1級以上のレベルの者に対してどう対応すべきかも検討が必要である。

次に、読解の判定が初級レベルでも中級あるいは上級となってしまう例があったことについては、以下のように考えられる。このような受験者は中国語母語話者に顕著であった。現システムはデフォルトとして、難易度が中くらいの問題項目から提示するように設定してある。そこには相当数の漢字が使用されている。中国語母語話者にとって漢字表記は内容理解に有利となる。そのせいで、レベルが上がれば上がるほど内容理解がしやすいというパラドックスが発生している。語彙のセクションではその点が既に予想されたので、漢字の問題は排除しているが、読解のセクションで漢字を排除することはできない。対応策として和語やカタカナ表記を増やすことによって、理解過程における漢字依存度を低下させることが提案されているが、どの程度対応可能かは今後の課題である。

操作上の問題については、システムに練習セッションを設けることで、対応できると考えている。致命的な誤操作はこれで避けられるとしても、さらに操作性を向上させるための画面のデザインについては当然考えていかななくてはならない。

## 4. 今後の開発予定

### 4-1. IRT パラメータへの移行

現在は読解のみオリジナル問題を作成し、プレテストを行い、その分析の結果得られるIRTの項目困難度パラメータを利用している。読解はこのIRT困難度パラメータと古典的識別度パラメータの併用である。将来はすべての領域において、オリジナル問題の作成とプレテストの実施を経て、IRT困難度とIRT識別力の2パラメータによるCATに移行させる予定である。

### 4-2. 項目アイテムの拡充と項目可変型CATの導入の検討

項目アイテム数を数千に増やし、現在の項目固定型CATに加えて項目可変型CATを導入したい。ただし、項目可変型CATに比べて項目固定型CATが能力判定に関して必ずしも劣っているというわけではないので、コストパフォーマンスも考慮しながらの今後の課題としたい。

項目固定型か項目可変型かの選択は今後の課題としても、項目数を増やすことは、項目への露出の問題からも避けては通れない。項目数が少なければ、当然、複数回受験すると同じ項目に出くわす確率が高くなるわけだから学習効果が出てしまう。これを避けるためには、項目固定型でパラメータをそろえた上で、項目の入れ替えを行うか、項目可変型で多数の項目アイテムを備えるかしくはなくてはならない。

### 4-3. 2段階推定法の導入

能力推定にかかる時間をより短縮するための2段階推定法を導入する。現システムでは、中レベルから問題が提示されるようになっている。これに対して最初の予備テスト数問でだいたいのレベルを推定し、本テストでの出題レベルを決定する方式を導入することで、より受験者能力に合致した問題の提示を可能にする。なお、予備テスト問題はIRT分析を行う前の新問題を使用し、その解答パターンを記録する。つまり、これら新問題のプレテストとしての機能を持たせる。これにより、受験と同時に項目の拡充のための解答パターンデータを



能力推定値とともにオンラインで収集することができる。

#### 4-4. 開発費用の獲得

今後、問題項目数の飛躍的な増加を図るためにアイテムライターを雇用したい。また、システムの改良も当然必要である。これらを実現するためには、相当の費用が必要となる。本 J-CAT の効用をアピールして、多方面からの資金援助を仰ぐことが必要になるろう。

#### 4-5. 他テストとの相関

J-CAT は日本語能力試験の出題基準にほぼ従っている。当然、両試験の相関は高くなければならない。また、日本留学試験も含めて他の試験との相関にも興味がある。J-CAT と他テスト間の相関については少数のデータを得ているが、本格的な研究は今後の課題である。

#### 4-6. 海外からのアクセス

J-CAT は海外の協定大学からの交換留学生の日本語能力の判定のため、海外からアクセスして使用することが既に予定されている。海外でのインターネットの環境はさまざまで、必ずしもブロードバンドに対応しているわけではない。その点を考慮し、ストリーミングやファイルサイズの縮小などの対策をとってきたのだが、予備実験では大方の国では重大な問題がなかったものの、中国からのアクセスでは、音声流れないという問題が見ついている。中国国内から海外へアクセスする際にデータ処理速度が遅くなる問題が生じているのではないかと思われる。中国国内にミラーサーバーをおくなどして対応することを検討中である。ナローバンド用 J-CAT も開発中である。

#### 4-7. 他教科への応用

本システムは留学生の日本語能力測定のために開発してきたものだが、コンテンツを入れ替えることで、英語や他言語、さらには数学など他教科の CAT としても使用できる。この点も視野に入れて、より汎用性の高いシステムへの改良を目指したい。

### 5. まとめ

J-CAT をインターネット経由で用いることにより、場所と時間の制約がなく、どこからでもいつでも受験が可能になる。比較的短時間で、日本語能力を測定でき、受験者の能力にあった問題が提示されるので受験者の精神的負担も軽減される。項目応答理論に基づく CAT はこれまでのテストの概念・方法に根本的な変革をもたらすだろう。項目応答理論自体は決して新しい理論ではない。ただ、それを実用化するには、現在のレベルのコンピュータの環境が整うのを俟たねばならなかった。今、まさにその環境が整ったのであり、インターネットを介して誰でもその恩恵に浴することが可能になった。新しい時代の新しいテスト開発に少しでも寄与できれば幸いである。

### 参考文献

- 秋山實・今井新悟 2005a 「アイテムバンク構築システムおよび適応型プレースメント テストの開発 - オープンソースの e ラーニングソフトウェア moodle を基盤として - 」『日本テスト学会発表論文抄録集』 pp. 82-83.
- 秋山實・今井新悟 2005b 「オープンソフトウェア moodle をベースにしたアダプティブテストとアイテムバンク構築ツール群の開発」『教育システム情報学会論文集』pp. 391-392.
- 今井新悟 2005 「コンピュータを使った簡易アダプティブテストの開発：J-CAT プロトタイプ 1」『山口大学国際センター紀要』第 1 号 pp.67-71.

今井新悟・秋山實 2005「J-CAT (Japanese Computerized Adaptive Test)」日本語教育学会  
 秋季大会デモンストレーション資料

大友賢二 1996『言語テスト・データの新しい分析法 項目応答理論入門』大修館書店

大友賢二監修・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  
 ～』桐原書店

芝祐順編 1991『項目応答理論 基礎と応用』東京大学出版会

独立行政法人国際交流基金・財団法人日本国際教育協会 2002『日本語能力試験出題基準（改訂版）』凡人社

日本留学試験シラバス:独立行政法人日本学生支援機構ホームページ  
[http://www.jasso.go.jp/examination/efjuafis\\_syllabus.html](http://www.jasso.go.jp/examination/efjuafis_syllabus.html)

日本留学試験における得点等化について:独立行政法人日本学生支援機構ホームページ  
[http://www.jasso.go.jp/examination/efjuafis\\_touka.html](http://www.jasso.go.jp/examination/efjuafis_touka.html)

林規生 2001「英語能力測定における CAT の適応例と効果測定」計測と制御 40(8) pp.572-575

#### あとがき--謝辞に代えて

オリジナル問題項目のパラメータづけと採否のためのプレテストを実施する際にご協力いただいた教育機関は以下の通りである。お礼申し上げます。

天理教語学院 日本語教育センター、ノースウエスト日本語学校、  
 宇部フロンティア大学、広島 YMCA 国際ビジネス専門学校、西日本国際教育学院  
 Nihongo Center (インド)、MOSAI Institute of Japanese Language (インド)  
 The association for the Japanese language education promotion (エジプト)  
 Cairo University (エジプト)、Ain Shams University (エジプト)、世新大学 (台湾)  
 東呉大学 (台湾)

開発予算の一部を山口大学教育研究後援財団より得ている。また、国際交流基金より、研究のために日本語能力試験の問題を使用する許可をいただいた。ここに記して感謝申し上げます。