

コンピューター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

今井 新悟

1. 開発の背景

国際センターでは、2002年度開設以来、留学生の日本語能力を測るべく、ペーパーによる試験を、4月と10月の年二回実施してきた。この試験は、毎回、本学の全外国人留学生および研修生を対象としている。この試験の目的は、以下のとおりである。第一に、日本語・日本事情の授業を受講する学生のプレースメントテストのため。第二に、留学生の日本語能力を判定し、留学生指導の基礎資料とすること。これは、例えば、学業不振の原因が、専門知識にあるのか、日本語能力が低いことが原因なのかを考える場合の資料として使ったり、指導教官から日本語能力についての問い合わせがあった場合に回答するための資料とすることなどを想定している。第三に、日本語能力の伸長率を、奨学金選考の参考資料とすることである。

以上の目的をもって実施されてきた試験ではあったが、徐々にこの目的を達成するのに障害となる問題が明らかになってきた。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、初級の問題は能力測定には意味がなく、逆に、初級者にとっては、全く歯が立たない上級問題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対応することは、実施者にとって負担が多すぎる。

これらの問題をクリアするための方策を探っていたときに出会ったのが、コンピューターを使ったアダプティブテストという方法であった。

2. CAT開発の基本コンセプト

コンピュータによるアダプティブテストはCAT (Computerized adaptive test) と略されるので、本稿でもこれに従う。CATは以下のような特徴をもつ。受験者は比較的少ない問題を解くことにより、能力が判定される。受験者の能力に合った問題が出題される。受験者個々の能力により、出題される問題が異なる。テスト結果が絶対評価として示され、信頼性が保証される。

現在、最も有名なCATはTOEFLのCBTである。これに使われているシステムやアルゴリズムは米国のETSで開発されたもので、公開されていない。日本ではアダプティブテストの研究は少なく、情報も限られている。

TOEFLのアダプティブテストの理論的背景は項目応答理論であるが、最も精度の高い

3パラメータの場合、これによって信頼できる統計値を得るためには、数千人の受験者のデータが必要になる。1パラメータの場合でも百人程度の受験者データが必要になる。TOEFLのように社会的な信頼度を要求されるテストはhigh-stakeな試験と呼ばれる。大学において日本語能力を測ろうとするとき、TOEFLほどのhigh-stakeな試験は要求されない。

日本語の能力に関するテストの結果は「日本語能力検定試験」で蓄積されている。さらに、古典的テスト理論に基づく統計データもある。しかし、古典的テスト理論に基づくデータをそのまま、項目応答理論を応用したCATには適用できない。そこで、項目応答理論で得られる項目困難度および項目識別力の代わりに古典的テスト理論で算出される正答率および項目識別力を使うことにより、2パラメータを用いたアダプティブテストを開発することにした。コンピュータ上で動く、簡易アダプティブテストである。そしてこれをJ-CAT (Japanese Computerized Adaptive Test) と命名した。本稿は、そのプロトタイプ1の開発に係る報告・考察である。

3. 期待される効果

期待される効果は多々あるが、主に以下の点が挙げられる。留学生に対する日本語能力の絶対評価が可能になる。時間・場所の制約なくテストが実施できる。テスト時間の大幅な短縮ができる。採点をコンピュータが自動で行うので、実施の負担が少ない。結果がテスト終了と同時に本人に通知されるのでフィードバック効果が高い。結果がデータとして残るので、成績管理が効率よくできる。

日本語能力テストの4級から1級のレベル、さらにそのなかでのどの程度の位置にあるのかを知るためには、現在は、年1回行われる日本語能力テストを受験するしかない。さらにその判定には1ヶ月以上を要する。一方J-CATでは、オンラインテストであるため、時間的、経済的な制約がなく、受験者の希望するときに、何回でも受験できるシステムとなる。また、級別4段階をさらに細分化することにより、識別力の高い能力判定を可能とする。

J-CATをプレースメントテストとして利用する場合の価値は以下のようなことである。プレースメントテストは、クラスを出来るだけ能力レベルの合った学生で構成し、各クラスの学習効率を良くするために行うものである。したがって、クラスの数に相応しい確度で対象となる学生の能力を識別することが必要である。また、テストによって測定する範囲は、これから教授しようとする範囲の基礎となっている部分を網羅していなければならない。大学に入学してくる留学生の日本語能力を測定する最も手軽で確実な方法は日本語能力試験の成績を利用することである。しかし、日本語能力試験の成績は大学入学時点の日本語能力を測定した結果ではない。また、日本語能力試験を受験していない学生が存在することを考えれば、入学時点で一斉にプレースメントテストを行うことが望ましい。プレースメントテストを、オンラインのJ-CATで代用することにより、テスト実施にかかる人的、時間的コストを下げることが可能である。

J-CATをアチーブメントテストとして利用する場合の価値は以下の通りである。各授業のアチーブメントテストは各担当者にまかされており、シラバスの違い、教科書の違いなどのため、クラス内相対評価以外に標準化ができないのが現状である。同一テストを開発・利用することにより、クラスをまたがった受講者全員の相対評価および絶対評価を可能とする。

さらに、本システムの開発は、コンテンツを取り替えることにより、英語など他言語にも対応できる。

4．開発の全体計画概略

2004年12月	プロトタイプ1作製
2004年12月	国立国語研究所「コンピュータと新日本語教育2004」で発表
2005年3月	プロトタイプ2作製
2005年4月	山口大学留学生を対象に試行
2005年10月	プロトタイプ3作製
2005年～2007年	アイテムバンク構築予定

5．プロトタイプ1開発経緯

システム開発研究目的のため、日本語能力試験で過去に使用された問題を借用して使用する。しかし、日本語能力試験の問題は複合問題や複雑な解答指定方法をとる問題など、オンラインテストに適さないものもある。その判断の基となる分析を行う。

過去問題の各問題についての統計的データを入手し、CBTテストに適当な問題の抽出を行う。統計的データは、全受験者数を基とした困難度（つまり正答率）および識別力の数値を利用する。（プロトタイプ3以降、項目反応理論によるデータを収集できる方法を考え、その分析結果が問題の選択に反映されるシステムを開発する。）

再編集された問題をオンラインテストシステムに載せ、システムの稼働具合をテストする。（実際の受験者に対してテストを試行するのは、プロトタイプ2以降に予定している。）

オンラインテストシステムは、オープンソースソフトウェアのMoodleを使うことで開発コストを抑える。Moodleの入力形式に合わせてテスト問題を入力する。

6．プロトタイプ1のシステム仕様

6-1 動作環境

Linux、Apache、MySQL、PHPなどmoodleが稼動する環境で動作する。

6-2 設定項目

J-CATシステムの動作を決める下記のパラメータを任意に設定できる。

- ・レベル合格に必要な正答数
- ・レベル合格時および不合格時に移動するレベル数

6-3 出題方式

問題を1問ずつ提示し、その解答の正誤によって出題すべき問題を決定する。また、当該レベルで合格した場合は上のレベルに、不合格の場合は下のレベルに移動した上でそのレベルの最初の問題を、合否判定ができない場合は同一レベルの残りの未出題の問題を出題する。

6-4 アルゴリズム

- ・真ん中のレベルから出題を開始する。レベルが偶数の場合、1レベル上のレベルから開始する。
- ・同一レベルでN問出題し（プロトタイプ1では最大3問）、M問（ $M < N$ ）正解したらレベル合格とする。
- ・一つのレベルで合格したらPレベル（プロトタイプ1では最大2レベル）上のレベルでレベル判定する。
- ・一つのレベルで不合格になったらQレベル（プロトタイプ1では最大2レベル）下のレベルでレベル判定する。
- ・レベル判定を一度でも受けたレベルは2度判定しない。

6-5 判定結果

レベル毎の判定結果をそのまま表示するとともにCSV形式でダウンロードできるようにする。

6-6 レベル判定・終了規則（Stopping rule）

合格したレベルと不合格になったレベルの間に合否判定すべきレベルがなくなったとき、レベル判定を終了し、合格した最高のレベルを受験者のレベルとする。

6-7 問題データベース

使用する問題データベースは、古典的テスト理論の項目困難度と項目識別力が既知の問題を12レベル24題とする。

7. まとめ

2004年12月にプロトタイプ1を作成した。本学サーバーに入れて、動作実験を行ったところ、予測通りに動作した。今後は、扱えるセクションを増やし、また、解答時間制限などを加え、受験生に対する実験を行いたい。（本稿脱稿時には、このプロトタイプ2ができ、受験生に対する試用も行われたが、これについては稿を改めて報告したい。）

8. 謝辞

J-CAT開発は、山口大学国際センターのプロジェクトとして、大勢の人の協力を得ながら進行しているものである。（2005年4月以降は合資会社eラーニングサービスとの共同研究に移行）moodleプログラムの改良においてお世話になったeラーニングサービスの秋山實氏に感謝したい。テスト理論について指導を仰いだ教育測定研究所の林則生氏にも感謝する。そして、山口大学国際センターのプロジェクトメンバーの名前を記して、その献身的な貢献に対して感謝したい。宮崎充保（センター長）、渡辺淳一、杉原道子、赤木彌生、門脇薫（敬称略）。

参考文献

- (1) 大友賢二 (1996) 『項目応答理論入門』大修館書店
- (2) 大友賢二監修 中村洋一著 (2002) 『テストで言語能力は測れるかー言語テストデータ分析入門ー』桐原書店
- (3) 国際交流基金・日本国際教育協会 (1992) 『日本語能力試験の概要 1992年版 (1991年度試験結果の分析)』
- (4) 国際交流基金・日本国際教育協会 (1993) 『日本語能力試験の概要 1993年版 (1992年度試験結果の分析)』
- (5) 国際交流基金・日本国際教育協会 (1994) 『日本語能力試験の概要 1994年版 (1993年度試験結果の分析)』
- (6) 国際交流基金・日本国際教育協会 (1995) 『日本語能力試験の概要 1995年版 (1994年度試験結果の分析)』
- (7) 国際交流基金・日本国際教育協会 (1996) 『日本語能力試験の概要 1996年版 (1995年度試験結果の分析)』
- (8) 国際交流基金・日本国際教育協会 (1997) 『日本語能力試験の概要 1997年版 (1996年度試験結果の分析)』
- (9) 国際交流基金・日本国際教育協会 (1998) 『日本語能力試験の概要 1998年版 (1997年度試験結果の分析)』
- (10) 国際交流基金・日本国際教育協会 (1999) 『日本語能力試験の概要 1999年版 (1998年度試験結果の分析)』
- (11) 国際交流基金・日本国際教育協会 (2000) 『日本語能力試験の概要 2000年版 (1999年度試験結果の分析)』
- (12) 国際交流基金・日本国際教育協会 (2001) 『日本語能力試験の概要 2001年版 (2000年度試験結果の分析)』
- (13) 国際交流基金・日本国際教育協会 (2002) 『日本語能力試験の概要 2002年版 (2001年度試験結果の分析)』
- (14) 国際交流基金・日本国際教育協会 (2003) 『日本語能力試験の概要 2003年版 (2002年度試験結果の分析)』