

平成 22 年 6 月 1 日現在

研究種目：基盤研究（C）
 研究期間：2007 年度 ～ 2009 年度
 課題番号：19520455
 研究課題名（和文）日本語口頭能力測定のための試験システムおよび評価法に関する研究
 研究課題名（英文）Development of a web-based scoring system for Japanese Oral Proficiency Test.
 研究代表者
 野口 裕之（NOGUCHI HIROYUKI）
 名古屋大学・大学院教育発達科学研究科・教授
 研究者番号：60114815

研究成果の概要（和文）：

本研究では、外国語としての日本語口頭能力測定にコンピュータを導入する試みのひとつとして、発話者の口頭能力を評価するにあたって、評価者がウェブにアクセスして評価すべき資料音声を手し、即時的に評価をしてウェブにその結果を返送するシステムを開発し、実用可能性を検討した。そのために、予備実験(2008.11.)および本実験(2009.11.)を実施して、評価者が妥当な評価結果を示すか、そして、評価者が評価システムの操作性についてどのように感じるか、などについてデータを収集し分析した。その結果、予め想定した発話者の口頭能力水準に合った評価結果がおおむね得られ、操作性についてもおおむね問題のないことが明らかになった。また、本研究を進める中から解決すべき新たな課題もいくつか明らかになった。

研究成果の概要（英文）：

The purpose of this study is to develop a web-based scoring system for Japanese Oral Proficiency Test. When the rater uses this system, he/she accesses the web-site which presents whole the test materials and controls whole the test process. Prior to the experiment, a pilot experiment was administered in 2008, the result of which applied to improve the system. The raters were instructed how to evaluate the level of four Japanese learners' speaking samples by the check list and the proficiency scale. Data was obtained from the 23 native Japanese teachers who had more than 3 years teaching experience. The results showed a relatively high validity for the raters' evaluation both in the check list and the proficiency scales. According to the raters' respond to our questionnaire, they seem to have felt relatively easy for operating this system. With some improvement, there should be good prospect for the practical implementation of the Follow-Up study.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	1,500,000	450,000	1,950,000
2008 年度	1,500,000	450,000	1,950,000
2009 年度	500,000	150,000	650,000
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：日本語教育・心理測定学

科研費の分科・細目：言語学・日本語教育

キーワード：外国語としての日本語教育・応用言語学・言語テスト・口頭能力測定・評価法・パフォーマンス評価・分析的評価・評価システム

1. 研究開始当初の背景

周知のごとく、国際交流基金と日本国際教育支援協会が実施している現行の日本語能力試験は紙媒体によるテストで、口頭による発話の能力を測定する部門がなく、言語テストとしての不十分性が長らく指摘されているところである。

(1)開発を目指すテストの新規性

申請者は、面接方式でない口頭能力測定法に関する研究を蓄積してきたところで、コンピュータを媒体として用いることによって、受験者が視覚と聴覚双方から入力される課題に回答できるテストの開発を目指している。また、評価についても、評価者がウェブにアクセスして資料音声を手入力し、たちどころに評価をしてウェブに返送するシステムを構想している。これは、最新の言語テスト理論において重視される真正性 (Authenticity) を不完全ながらも実現しようとする試みであると考えている。

今回の実験では、このうち主として評価システムの開発に重点を置いて研究を進める。

(2)海外のテスト開発状況

上記テスト方式は、多数の受験者を対象とすることができない面接方式の弱点に対する大きな利点として関係者の注目するところとなり、米国、韓国をはじめとしてその実用化が着々と進められている。その中では、申請者を中心としたこれまでの研究の成果が利活用されているのに対し、わが国においてはテスト開発が一向に具体化していない。研究の先発性とアイデアにおいていかに優れていようとも、そのアイデアの実現実行性において他国の後塵を拝してはならないと考えるゆえんである。

(3)評定票の改良

本研究では、これまで使用してきた評価用の用紙 (=評定票) を、コンピュータ・ベースのテストに使用しやすいよう簡素化したほか、評価の基準を示す記述 (=ループリック) をより分かり易く書き改めて使用することにした。ただし、発話標本は従来のものを使用する。

2. 研究の目的

外国語としての日本語の口頭能力を測定するテストを CBT (Computer-Based Test) として開発する。単に既存のテストをコンピュータ化するのではなく、タスク (=問題) の提示方法及び能力の評価法を含めコンピュータをテスト実施媒体として用いる利点を最大限生かすことができる総合的な開発研究である。

(1)コンピュータ化の実験

評価者を一堂に集め、標本音声を再生しつつ紙媒体の評定票に記入することによって評価を行う直近過去の実験方法と、コンピュータから流れる標本音声を聞きつつコンピュータ画面に表示した評定票に評価を記入することによって評価を行う方法を比較し、コンピュータ化の実現可能性を検討する。

この目的を達成するため、

- i) 海外における CBT の実施状況を調査する。
- ii) 従来使用してきた評定票をコンピュータに適した形式と内容に改良する。
- iii) 実験対象とする資料音声を決定する。
- iv) 専門業者と共に、ウェブに用意するソフトウェアを作成する。

(2)コンピュータによる標本音声送出現の実現可能性の検証

評価者が自宅などにある自己の端末からウェブにアクセスし、評価対象となる標本音声を原音に近い状態で、かつ大きな困難もなく入手可能かどうか、入手が困難な場合は、いかなる原因によるものかを明らかにする。

この目的を達成するため、

- i) 学会などにおいて上記の試作版を公開し、実施に堪えるかどうか意見を求める。

(3)コンピュータ画面上での評価の実現可能性の検証

紙媒体の評定票の場合必要となる大きな作業空間をコンピュータに置き換えた場合の利便性と作業困難性を明らかにする。また、評価結果が円滑にウェブに集積され、のちの結果分析に有効に活用できるかどうかを検証する。

この目的を達成するため、

- i) 実験遂行中、実験協力者と綿密な連絡を取りつつ、ウェブ作成を担当した専門業者とメンテナンスの方法を調整する。
- ii) 同じく業者と協議して、今後最適の評価システムを模索する。

(4)言語テスト実施に必要な IT 環境

以上の検証過程を実行するためには、言語テストを遂行するための IT 環境が整備されていることが必要であるが、米国、韓国など諸外国に比べてわが国の現状はいまだ十分とは言えない。本実験を通して、言語テスト実施に堪える IT 環境の整備を目指す。

この目的を達成するため、

- i) 実験遂行中、実験協力者から得られたアクセス状況に関する情報を参考に、望ましいデータの管理方法を検討する。
- ii) ウェブ作成を担当した専門業者と協議して、言語テストに必要な環境の整備について検討する。

(5) 評価者アンケート

紙媒体による評価に比べた本方式の利点と問題点に対する意見を求め、今後の改良に資する。

3. 研究の方法

上記の目的のために、以下の2回の調査を行った。

(1) 紙版評定調査 (2008年12月)

評価システム化をする際に使う、紙版の評価票そのものの妥当性・実用性検証を行い、予備調査とした。

調査協力者は、日本語教育歴3年以上、ACTFL-OPI テスター資格保持者、日本語教育専攻大学院生、のいずれかの条件を満たす方を募集した結果、17名となった。

協力者には調査場所に集合してもらい、①口頭能力試験開発の簡単な経緯、②実験の目的、③評定方法について説明した。その後、④練習として一人の発話を聞き、量の評価(チェックリスト)と質の評価(評価基準)の2種類の評定を行った。そして、⑤新たに4人の発話者の発話サンプルを聞いて、本評定を実施した。なお調査時間の制限から、発話者に与えられた課題に関しては、これまで開発した課題のうち、課題特性の点で異なる特徴を持ち、作題意図に近い発話が現れた3課題に絞った。最後に⑥アンケートに記入いただいた。

(2) 本調査：紙版及び PC 版評定調査 (2009年11月～2010年1月)

上記(1)の予備調査を経て、コンピュータ・ベースの日本語発話能力評価システム構築を目指すため、紙版の評価票ではなく、コンピュータ画面での評価票を用いて評価を行う調査を実施した。また、紙版では結果を手入力し、集計するという事後の処理も発生したが、それもシステムに組み込んで得点化を自動的にする試みも行った。

調査協力者は、(1)の調査と同様、以下の、i)日本語教育歴3年以上、ii)OPI テスター資格保有者、iii)日本語教育専攻の大学院生のどれかを条件としたが、さらに、iv)2009年10月に開催された日本語教育学会秋季大会(於九州大学)デモンストレーション発表の会場への来場者にも声をかけた。また、予備調査の協力経験者にも新たにPC版での協力を依頼し、調査協力者は37名となった。最終的な有効な調査票は23であった。

システム開発に関しては、二社からeラーニングにも実績のある一社を選び、開発を開始してからも都度相談しながらシステムを改良した。

調査は、紙版評価票と、システム化したPC版評価票を用いた調査の2種類を行った。協力者とは電子メールでやりとりし、どちらの

調査の協力者にもそれぞれ必要な資料送付を行い、アクセス期間に制限を設け、その期間内に評価を終了するよう依頼した。調査自体の流れは(1)と同様だが、PC版では(1)の④にあたる練習、⑤本評定、⑥アンケートの部分がコンピュータ画面での回答となった。

4. 研究成果

上記の予備調査と本調査から、以下のような成果が得られた。

(1)予備調査では、評価者の評定が発話者4人のレベル通りだったことがわかり、評定票の正当性が証明された。

(2)本調査からは、評価者は自宅、職場、学内、知人宅など、任意の場所で評価が可能となったことがわかった。

(3)評価結果を自動的に集計するようにシステムを組んでいたが、協力者が調査を終了する度に、研究メンバーの元に着々とデータが送信され、かつ、簡単な集計もされていた。

(4)紙版の予備調査でも協力者となり、今回はPC版協力者となった方のデータと、今回初めてPC版協力者となった方のデータでは評価に差があると考えたが、t検定を行い差がないことが証明されたため、両者をまとめてPC版協力者のデータとして扱った。その上で、紙版協力者とPC版協力者の評価結果に差があるかどうかt検定で調べたところ、有意差がないことがわかり、評価媒体の違いによって評価結果に差は生じなかったことが明らかとなった。

また、紙版、PC版とも、4名の発話者について、過去の調査から想定した通りのレベルの結果が得られた。

(5)ある意味での研究成果として、今後の課題も浮上してきた。協力者の環境によっては、通常使用しているインターネットブラウザでは、音声がかえれないケースが散見された。この件については、システム構築をした業者と相当やりとりを重ね、次々に改良を重ねたが、ブラウザの変更以外にはデータとしての音質を高めるなどの余地を残した。しかし音質改良には、予算的にもかなりかかることも判明した。

(6) アンケートの結果

評定に引き続き、PC上で協力者に対して5件法(とても役に立った⇔全く役に立たなかった、など)・自由記述式のアンケートを実施した。有効回答数はPC版が22、紙版が10であった。主な結果は次の通りである。

i) 評定前の練習

本評価システムでは本評定の前に練習用音声で評定の練習を行う機会を設けたが、その練習についてはPC版では22名全員が「とても役に立った」「まあまあ役に立った」のどちらかを選んだ。

また紙版でも「とても役に立った」「まあ

まあ役に立った」が 9 名あり、PC 版・紙版に関わらず、練習は有用であることがわかった。

ii) 操作のしやすさ

紙版ではほとんどの協力者が操作しやすいという評価であった。

PC 版でも多くの協力者が操作性を評価したが、一方、「どちらとも言えない」「とても操作しにくい」が 1 名ずつあり、PC での操作方法をもっと具体的に説明すべきだという指摘があった。

iii) 評価をする際、よかったと思ったこと

PC 版ではクリックひとつで発話の再生、評定、訂正などができ、評定基準が紙でなく PC の画面上で簡単に確認できること、評定結果を紙に書くより時間が短くて済むことなどが挙げられ、主に操作機能の面で PC の優位性が確認された。

紙版では評定基準を前に戻って読み直すことが簡単、PC を開く前に紙で評定基準を読むことができ、後の評定がスムーズにできた、などという点が出た。このことから紙版特有の長所もあることがわかった。

iv) 評価をする際、難しいと思ったこと

PC 版では、音声は 1 回しか聞けないので適切な評価ができていない可能性がある、という意見が出された。紙版でも音声は 1 回しか聞けないのであるが、紙版の協力者からはこのような意見は出なかった。これは、PC 版の方では評価中も PC にアクセスしているため、ボタンひとつで音声を再度聞けるようにすることが容易であると協力者が考えやすかったか、紙版では PC は発話を一方的に流すための道具、評価は紙の冊子で独立に実施すると考え、協力者に音声を再度聴くという発想をもたさなかったか、いずれかの可能性があるが、今回の結果からは結論が得られなかった。

また PC 版では自分がいったん評定したものをさかのぼって確認したり修正したりできないため、その点をできるようにしたほうが良いという意見が示された。

紙版では冊子の頁が多数ありすぎるため、目的のものを探するのが大変、狭い所で評定するときには不便という意見があった。

PC 版・紙版共通して出された難しさには、評定基準が難しい、この評定基準では測れないのではないかなど評定基準そのものに関する疑問があった。

さらに「調査についてのご説明」で「システムは途中から再開することができませんので、長時間の中断は避けてください」と最初に注意を促しているが、このことに対して負担を感じた協力者が数名あった。実験前にシステムは途中から再開できないこと、したがって長時間の中断は避けることを説明したが、実際には中断せざるを得ないことも起

こったので、そういった場合にどう対処するかについても最初に説明されていたらよかったというコメントが見られた。

v) 評価する発話標本の数は負担に思ったか

協力者は問題 A、B、C について各 4 名分の発話の評定を行ったが、この数に対して PC 版では「余裕があった」「少し余裕があった」が 22 名中 6 名、紙版では 10 名中 7 名であった。どちらとも「負担であった」を選んだ協力者はいなかった。

vi) この方式で適切に評価できるか

PC 版では「できる」「ある程度できる」と答えた人が 22 名中 19 名あり、一定の条件を付けた上で適切な評価ができると感じている人が多いことがわかった。

一方、紙版では「ある程度できる」が 5 名、「どちらとも言えない」が 1 名、「できない部分がある」が 4 名であり、PC 版ほど高い評価は得られていない。

5. 主な発表論文等

[学会発表] (計 1 件)

堀川有美・和田晃子・野口裕之 コンピュータ・ベースの日本語発話能力評価システム 日本語教育学会 2009 年度秋季大会 2009 年 10 月 11 日 九州大学

[その他]

ホームページ

<http://www.checkbox.jp/nihongo/>

6. 研究組織

(1) 研究代表者

野口 裕之 (NOGUCHI HIROYUKI)
名古屋大学・大学院教育発達科学研究科・教授
研究者番号：60114815

(2) 研究協力者

庄司 恵雄 (SHOJI YOSHIO)
お茶の水女子大学・国際教育センター前教授
研究者番号：40253017

小林 久美子 (KOBAYASHI KUMIKO)
神田外語大学・留学生別科・非常勤講師

堀川 有美 (HORIKAWA YUMI)
国際交流基金・日本語試験センター・研究員

和田 晃子 (WADA AKIKO)
元国際交流基金・日本語試験センター・研究員