

音声認識を用いた完全自動の日本語リスニング・スピーキング試験

原田康也 (harada@waseda.jp)
早稲田大学法学学術院教授・情報教育研究所所長

1. まえがき

近年の日本の国際化に伴い、国内に在住する日本語を母語としない人々と海外において日本企業や関連組織に就労する日本語を母語としない人々が急激に増加している。学習や就労にはスピーキングも含めて一定以上の日本語運用力が必要であるが、スピーキングについては書く・読む・聞くの3技能の試験結果から間接的に予測したり、面接試験等を実施するというのが一般的であった。しかし、他の3技能のスコアが良くても、入学・採用後に話す力が不十分であることが判明することも少なくない。一方、面接試験の実施には、多くの人手とコストを必要とし、大多数の受験者を一定期間内に試験する際には現実的な選択肢ではない。

米国 Ordinate Corporation と早稲田大学総合研究機構情報教育研究所では完全自動の日本語リスニング・スピーキング試験を開発中である。この試験では HMM ベースの音声認識エンジンとコンピュータによる自動採点システムにより、日本語学習者の口頭言語処理能力を迅速かつ正確に測定することを目指し、音声認識エンジンと自動採点システムの開発のためのデータ収集が現在進行中である。本稿では、本試験の開発過程と現状、自動テストシステムの仕組み、並びにデータ収集の状況について報告する。

2. SJT 開発背景

IT 技術・情報処理技術が多目的に様々な分野で利用されている。CALL (Computer-Assisted Language Learning) に代表されるように、教師の語学指導や学習者の語学学習の支援を目的に、語学教育の分野でもマルチメディアなどを用い、積極的に新しい試み・開発が行われている。言語テストもまた情報処理技術の発展の恩恵を受けつつある。インターネットを使ったオンライン中心の CBT (Computer-Based Testing) が最たる例であろう。そうした恩恵の一つに自動採点を挙げることができよう。

しかし、これまでの CBT ではリーディング・ライティング・リスニングの3技能の測定や語彙・文法などの知識の判定が主流であり、スピーキングについてはこれらのスコアから推定するのが一般的であった。しかし、「3技能でのスコアが良くても会話能力が乏しい」というのはよく聞くケースである。

こういった状況を避けるためには直接口頭能力を測定することが一番であるが、口頭能力を自動測定できる試験は数少ない。

日本語教育を始め外国語教育ではコミュニケーション力養成が授業の中心になって久しいが、従来では学習者の口頭能力を測定する場合には OPI (Oral Proficiency Interview) などに代表されるインタビューテストを用いるのが一般的であった。通常 OPI では、試験開始から終了まで受験者一人につき 20-40 分程度を要し、訓練を受けた試験者・採点者を必要とする。このため、一定期間に多数の学習者の口頭能力を測定したい場合には、時間・費用・採点者の確保に多大な負担がかかり、運用上の制約がある。また、一定の訓練を受け認定を受けた採点者も定期的に再訓練を受けねばならないなど、採点要員の確保には多大なコストを要する。

アメリカ・カリフォルニア州にある Ordinate Corporation (www.ordinate.com) は、計算言語学・音声認識・音声処理・テスト理論の研究成果を活用し、電話で受験可能な完全自動の口頭英語試験 SET-10 (Spoken English Test - 10 minutes) とスペイン語の口頭試験 SST (Spoken Spanish Test) を開発し、運営を行っている。SET-10・SST とともに試験者・採点者を介さず、試験開始から終了後の採点までが完全自動で行われる。Ordinate の試験はコンピュータにより自動採点され、SET-10 と SST では、それぞれ 0.97[1] と 0.96[2] という高い信頼係数が得られている。

近年の国際化の流れの中で、日本語学習者数が増加している。国際交流基金[3]の調べによると、海外での日本語学習者の数は2003年現在2,356,745人となり、1979年に比べ約18.5倍に増加した。また、企業においても非日本語母語話者を採用するケースが増加している。留学や就労には一定の日本語運用能力が必要となるが、国際交流基金[4]が実施している2004年度の日本語能力試験では2004年度の受験者は30万人を突破し、この増加傾向を如実に反映している。学習者の増加に伴い、インタビューテスト以外に、迅速で正確、かつ客観的な日本語の口頭能力測定試験への需要が高まっている。

Ordinate Corporation と早稲田大学総合研究機構情報教育研究所では、SET-10・SST と同様に Ordinate テストシステム上で運用可能で、電話により受験でき、音声回答を音声認識システムと自動採

点システムを利用し自動採点が可能な、客観的で完全自動な日本語リスニング・スピーキング自動試験 SJT (Spoken Japanese Test) を共同開発する運びとなった。現状ではプロトタイプ型 SJT テストが電話を用い受験できる段階にあるが、まだ自動採点システムは構築されていない。現在ではこのプロトタイプ型 SJT を用い、自動採点システム構築と妥当性検証に必要なデータ収集を行っている。

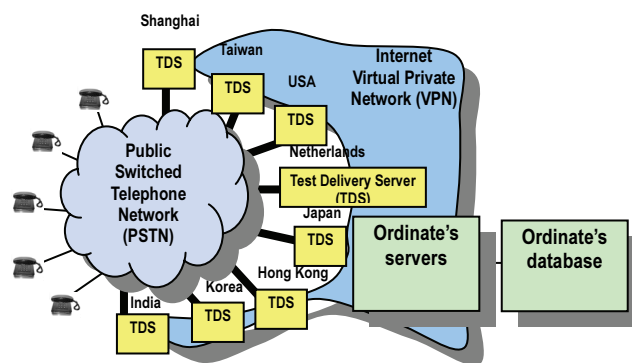
3. Ordinate テストシステム

SJT では SET-10 と SST 同様、Ordinate テストシステムを使用する。受験にあたって、各受験者は個別化された試験用紙と固定電話（携帯電話・ワイヤレス子機は不可）が必要である。試験用紙の表面には受験方法が説明されており、裏面が試験用紙となる。準備ができた段階で、試験用紙に印刷されてある電話番号に電話をかけ、プッシュボタンを使い、印刷されてある個別のテスト番号を入力するとテストが開始する。試験は約 15 分で終了する。現状の SJT はデータ収集の段階にあり、音声認識エンジン、自動採点モデルが構築されていないため、テストは採点されない。SET-10 や SST などの通常の Ordinate のテストでは、試験終了後数分以内に採点結果が Ordinate のウェブサイトからスコアレポートとして入手可能になる。スコアレポートでは、総合スコアのほか、文章構成・語彙・流暢さ・発音の 4 つのサブスコアも報告される。スコアは全て 20 点から 80 点の範囲で報告され、試験担当者（教員・人事担当者）は、自分の担当する受験者のスコアを一覧として閲覧・入手できる。また、回答音声の一部を実際に聞くこともできる。SJT でも最終的にはこのようなスコアレポートシステムになる予定である。

コンピュータを使用した Ordinate テストシステムは、大きく分けて、テスト配信システムと採点システムの 2 つの構成要素から成る。テスト配信システムは電話を通して受験者にテスト項目を音声で提示する。これらのテスト項目は予め対象言語の母語話者により録音される。SJT の項目は日本語母語話者により録音されている。

受験者が試験用紙に印刷されている電話番号に電話をかけると、TDS (Test Delivery System) というテスト配信システムがテストを配信する。日本を含めいくつかの国ではローカル TDS が設置されているため、受験者は国際電話をかける必要がなく、国内フリーダイヤルの番号で受験できる。このローカル TDS システムが Ordinate 本社にあるテストシステムと交信し、テストの実施を行う。試験終了後、回答音声ファイルはデジタル化され、インターネット経由で本社テストシステムに転送される。図 1 はこのシステム体系を図式化したものである。

図 1. Ordinate テストシステム



Ordinate テストシステムに送信された受験者の回答は、第二の要素となる音声認識エンジンとコンピュータ化された自動採点システムにより自動採点される。Ordinate では HMM (Hidden Markov Model) ベースの音声認識エンジンを使用する。上述の通り、現状では日本語用の音声認識エンジンと自動採点システムがまだ開発されていないため、現在母語話者だけでなく、非母語話者からの音声データを収集している最中である。

音声認識エンジンでの認識結果は自動採点システムに送られ、受験者の応答にどの単語が使用されたかとともに、応答のペース、なめらかさ、語句や文中で使われた単語の発音の正確さに基づきスコアが算出される。つまり、Ordinate の自動採点システムでは受験者が「何」を「どのように」に言ったか、という 2 つの側面が採点される。採点結果は、総合スコアのほか、文章構成、語彙、流暢さ、発音といったサブスコアがスコアレポートとして Ordinate のホームページから入手可能であり、SJT も最終的には同様のスコアを受験者・試験利用者に報告する予定である。

4. 開発過程

Ordinate テストシステムで SJT を運用可能にするには、大きく分けて以下の 3 つの作業が必要である。

- 1) テストデザイン・問題作成
- 2) データ収集
- 3) データ分析・妥当性の検証

4.1 テストデザイン・問題作成

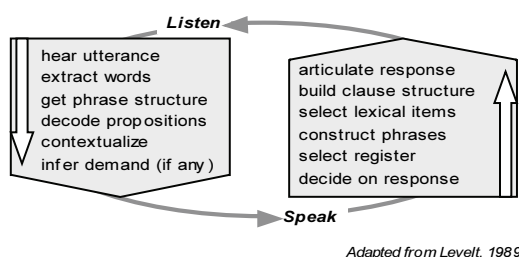
テスト構成概念

SJT では受験者の "Facility in spoken Japanese" の測定を目的にしている。"Facility in spoken Japanese" とは「日常生活一般の話題について、日本語母語話者（もしくは上級の日本語学習者）と同様なペースで日本語の会話を理解し、明瞭に受け答えすることができる能力」のことである。日常会話に効率良く参加するためには「聞いたこと

を「実時間で理解し、応答する」という心理言語学的な基礎的言語処理能力が必要となる。具体的には、話者は発声された単語・文を聞き取り、瞬時に意味を抽出し、そして聞かれていることに関連した内容で明瞭に応答する、ということである。これら一連の心理言語的発話作業をモデル化したのが Levelt[5]であり、その発話モデルを改良し、図式化したものが図2である。

SJT ではすべての設問において「音声による指示または提示とそれに対する回答」という課題になっている。受験者はこうした指示または提示を聞き、理解し、答える必要がある。この「聞いて→話す」という課題に対処するためには、心理言語学の視点からの速やかな「実時間での言語処理能力」が必要になる。

図 2. 口頭言語処理モデル



Adapted from Levelt, 1989

テスト構成

7つの設問項目が SJT 用に作成された。データ収集の段階で使用されているプロトタイプ型 SJT では、作成した7つ全てが使用されている。データ収集後のデータ分析の結果を基に最終的な設問項目数、各設問項目内における問題数などを決定していく。表1に、データ収集用の試験用紙に含まれる問題をまとめた。パート A からパート G で構成され、音読、復唱、反対語、質問、文の構築、話の要約、そして自由回答式質問という設問項目で構成される。パート A の音読の問題だけ、実際の項目がテスト用紙に印刷される。全部で 12 文印刷され、各文には番号が振られ、受験者は電話からの音声で指示された番号の文を順次読み上げていく。各問題文では漢字が使用されているため、全ての漢字にはひらがなで振り仮名をつけ、漢字をあまり学習していない受験者でも文を読み上げられるように配慮してある。パート B からパート G までは全ての問題が完全に音声のみで提示され、試験用紙には各設問項目の説明と例題だけが印刷されている。

表 1 データ収集用試験問題

設問項目	問題数
パート A: 音読	8
パート B: 復唱	16
パート C: 反対語	8
パート D: 質問	16
パート E: 文の構築	10
パート F: 話の要約	2
パート G: 自由回答式質問	2
合計	62

問題作成

テスト構成概念にあるように、SJT では受験者の日常生活一般の話題についての口頭言語処理能力の測定を目的としている。このため、各項目で使用される単語、文体、話題は日常会話で一般的によく使用されるものを用いている。また、日本語の日常での言語会話を反映させるため、敬語を用いた項目も作成した。日本語での日常言語活動を反映するようなバランスの良い項目作成を心がけた。

項目作成後には、日本語言語学・日本語教育を専門とする日本在住の言語学者、米国で外国語としての日本語を教えている日本人日本語教師などに項目レビューを依頼した。基本としている言葉はいわゆる標準語である。使用されている単語、表現が日本語母語話者が日常生活で使用するものかどうか、日本語として適切で正しい表現かどうか、特定地域に固有の表現や日本文化に関する単語・表現がないかどうか、などの確認を依頼した。専門家からのコメントを基に、必要に応じ項目を修正し、約 1,800 から成る項目バンクが出来上がった。項目作成後は、日本語母語話者複数名により録音された。男性、女性両方の声とともに、東京出身だけではなく、東京以外の出身者の声も使用し録音された。

データ収集

現在の開発状況は冒頭でも述べた通り、電話上で受験可能となっているが、自動採点はされていない。現在 Ordinate と早稲田大学情報教育研究所では音声認識エンジンと自動採点システムの開発に必要なデータ、また妥当性検証のためのデータなど、一連のデータ収集を行っている。日本語非母語話者の発話をできるだけ正確に認識、採点できる音声認識エンジンと自動採点システムを開発するためには、母語話者だけではなく非母語話者から大量にデータを収集することが重要である。母語話者に関してはできるだけ東京だけでなく、様々な出身地の人たち

からデータを集めることを心がけている。現在のところ、早稲田大学在籍中の学部生、大学院生、留学生などからデータ収集を行っている。また、日本のその他の大学に在籍する日本人学生や留学生にも現在協力を要請中である。母語話者からのデータは現時点で約 150 名程度に達している。

現時点では非母語話者のデータは日本在住の非母語話者から主に収集しているが、韓国、アメリカなど、日本国外にいる日本語学習者からもデータ収集を行う予定である。また、年齢、性別、母語、語学力などを考慮に入れ、様々な非母語話者からデータを収集していくことも重要な点である。

音声認識エンジンの開発のため、収集された音声ファイルの大部分は日本語母語話者により書き起こされる。書き起こし作業はインターネットベースの書き起こしインターフェイスを用いて行われる。書き起こし作業では母語話者並びに非母語話者の音声データを書き起こしていく。

自動採点システムの開発のためには、回答データ一部を評定者によりまず人為採点される。評定者は「流暢さ」と「発音」に関して評定していく。この採点結果を利用し、自動採点アルゴリズムを構築し、自動採点システムが人間の評定者と同様の評価ができるように調整していく。

一部の非母語話者には SJT の他、口頭能力試験として一般的に広く認知されている ACTFL-OPI (OPI: Oral Proficiency Interview) といったインタビューテストを受験してもらい、併存的妥当性 (Concurrent Validity) も検証していく。

データ分析・妥当性検証

最後のステップでは、得られたデータを様々な分析していく。まず、項目応答理論 (IRT) を用いて、各項目の項目困難度を算出していく。また、各項目では母語話者が正解できたかどうかを審査する。回答データを基に、不適切な問題を明らかにし、最終的な項目バンクから排除していく。最終的な項目バンクに残る項目は 80-90% 以上の母語話者が正解できたもののみとする。また、総合スコアと 4 つのサブスコア (文章構成、語彙、流暢さ、発音) に関して個々に信頼係数を算出する。また、人間の評定者と自動採点システムとの採点結果を比較し、自動採点システムにより算出される客観的なスコアが、人間の評定者が与えるスコアとどのような相関関係があるのかも検証していく。上述の通り、データ収集の段階で一部の非母語話者には ACTFL-OPI などのインタビューテストを同時に受験してもらおうが、そのインタビュースコアと SJT のスコアを比較し、SJT とインタビューテストの間になどどのような関係があるのか、つまり併存的妥当性の検証も行っていく。

5. まとめ

本稿では、アメリカにある Ordinate Corporation と早稲田大学情報教育研究所で共同研究・開発・検証を進めている完全自動の日本語口頭能力測定試験 SJT の開発背景と開発過程、現状を中心に述べてきた。従来の「読む・書く・聞く」の 3 技能ではなく、Ordinate テストシステムでは完全自動で口頭能力を測定できる。その中核にはインターネットなどの IT 技術に加え、音声認識を利用した音声処理技術がある。試験に利用可能な高精度な音声認識システムと自動採点システムを開発するためには日本語の母語話者からだけではなく、初級者から上級者まで様々なレベルの非母語学習者からも大量に音声データを集めることが必要であり、現在そのデータ収集を実施している段階である。ここで収集されたデータを基にこのテスト自体の信頼性と妥当性を検証していくことが今後の課題である。

また、コミュニカティブな授業をすることが語学教育での傾向であるが、簡便に実施できる口頭能力試験がないことから口頭能力についての授業効果を客観的に測定し、評価しているところは少ない。SJT のような客観的で完全自動の試験ができることで様々な Positive Washback も期待できよう。

6. 謝辞

本試験の開発・検証にかかわる共同研究は Ordinate Corporation と早稲田大学情報教育研究所が協力して進めている。本研究を遂行し、本稿を取りまとめるにあたっては、同社の鈴木正紀氏 (msuzuki@ordinate.com) より多大な協力を得ている。本研究の実施に際しては、科学研究費補助金萌芽研究 (課題番号 16652040) 『電話による日本語リスニング・スピーキング自動試験の運用性・信頼性・妥当性の検証』の助成を受けている。

文 献

- [1] Ordinate Corporation, *SET-10 Test Description & Validation Summary*, Menlo Park, CA, 2004.
- [2] Ordinate Corporation, *Spoken Spanish Test: Description & Validation Summary*, Menlo Park, CA, 2004.
- [3] 国際交流基金, “海外の日本語教育の現状—日本語教育機関調査・2003 年—概要”, 凡人社, 2004.
- [4] 国際交流基金, 日本国際教育協会, “日本語能力試験結果の概要, 2004 (平成 16 年度)”, 2005.
- [5] W. J. M. Levelt, *Speaking: from intention to articulation*, MIT Press, Cambridge, MA, 1990.