

教師の教育経験は学習者の発話評価に どのような影響を及ぼすか

鈴木 秀明

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

Hideaki Suzuki

本研究は教師の教育経験の多少が学習者の発話評価の際にどのような影響を及ぼすかを調査するものである。現職の日本語教師40名に、OPIのサンプルテープを聞いてもらい、全体的評価、分析的評価をしてもらった。分析的評価の際には、テキストタイプ、文法、音声、語彙、流暢さ、機能、内容の7項目についての評価を実施した。また、発話評価の際に何に着目しているのかを自由に記述してもらった。結果として、分析的評価で、中級テープの文法において、教育経験の少ないグループの方が経験の多いグループに比べて評価平均点が高くなるという点で有意差が見られた。しかし、上級テープの評価では7項目いずれにおいてもグループ間に評価平均点の差は見られなかった。また、教師の自己基準に関しては、文法、音声、語彙を経験の多いグループと少ないグループの両方が高い割合で評価項目として挙げていた。この結果を先行研究 (Hadden, 1991; Okamura, 1995) を参考に考慮すると、教師は教育経験を積むにしたがって文法に関する項目内の知識の量が増え、より詳細な評価基準を身につけていくと推測される。ただし、これは中級テープの評価に関してのみに言えることで、教育経験の多い教師が全ての習熟度において詳細な評価基準を身に付けているとは言えない。

*発話能力 *教師評価 *教育経験 *分析的評価 *自己基準

言語科学研究第10号（2004年）

はじめに

教師は、日常的に学習者の能力の測定と評価を行っているが、果たして学習者の能力を適切に評価できているのだろうか。教師評価に関する研究では、学習者の能力を評価する際の教師の評価基準や一般人の評価基準を調査したものが数多く見られるが、教師の教育経験が学習者への評価にどのような影響を及ぼしているかという点はまだ明らかにされていない。教育経験を積めばより適切な評価ができるようになるのか、あるいは教育経験を積んでも評価の仕方は変化しないかが明確になれば、問題点を改善するための対応策も提言できる。もし仮に教育経験を積んでも評価技術が向上しないのであれば、評価に関するトレーニングが教育現場においても必要であることが示唆できる。そこで、本研究では、特にスピーキング能力に対する教師の評価に焦点を当て、教師の教育経験が発話を評価する際にどのような影響を及ぼしているのかについて調査する。

1. 先行研究

1.1 スピーキングテストに関する研究

スピーキング能力を実生活の場面や状況等を考慮に入れ、言語使用の観点から見ると、コミュニケーション能力 (communicative competence) の一部分として捉えることができる。言語教育学の分野ではコミュニケーション能力に関する代表的な理論的研究が1980年代以降いくつも見られる (Canale & Swain, 1980; Hedge, 2000; Bachman, 1990)。各研究者のコミュニケーション能力の分類方法には多少の違いはあるものの、コミュニケーション能力が複数の能力から構成されていることや、コミュニケーション能力が文法的知識や文の構成といった構造的な能力と、社会言語学的能力のような語用論的能力から成り立っている点では研究者間に共通点が見られる。

教育現場で実際に使用されている測定基準に基づいてスピーキング能力を見ると、言語教育の現場で使用されているスピーキング能力の測定基準としては、ACTFL (American Council on the Teaching of Foreign Languages, 全米外国語教育協会) のガイドラインが代表的なものとして挙げられる。ACTFLは教育者の経験主義に基づいて開発された総合的外国語能力基準であるが、現在世界の教育

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

現場で幅広く活用されている。ACTFLの日本語運用力ガイドライン (ACTFL Japanese Proficiency Guidelines) では、スピーキング能力の構成要素を機能・タスク遂行能力、場面・内容処理能力、正確さ生成能力、談話構成能力の四つとしている。そして、スピーキング能力が習熟度別に初級(Novice)、中級(Intermediate)、上級 (Advanced)、超級 (Superior) の4つに分類されている。さらに、初級から上級までは習熟度内で上 (High)、中 (Mid)、下 (Low) の3つに細分化されている。このACTFLの測定基準は非常に実践的な言語運用能力の習得に主眼を置いて作成されたものであると言える。

教育現場におけるスピーキング能力を測定する代表的なテストとしては、ACTFLのOPI (Oral Proficiency Interview) や、CAL (Center for Applied Linguistics) によって開発されたSOPI (Simulated Oral Proficiency Interview) が挙げられる。これらのテストは外国語学習者の口頭運用能力を一般的な能力基準に参照して測定するために開発されたもので、テストの種類としては習熟度テスト (Proficiency Test) に該当する。OPIでは、テスト1人と受験者1人が対面してテストが実施され、インタビュー形式で進められていく。約30分の制限時間内にテストは受験者に様々な質問を行い受験者の口頭運用能力を判断するのに必要な発話を抽出する。SOPIも会話能力の習熟度を測定する点ではOPIと同様であるが、テスト形式がやや異なり、受験者はテープから流れる音声の質問や指示を聞いて一定時間内に解答する方式となっている。

OPIとSOPIに関する先行研究では、それぞれのテストの特徴を分析し長所と短所について述べたものや、習熟度ごとに学習者の発話分析をしたものなどが見られる。Kuo&Jiang (1995) では、OPIとSOPIの内容や構成の特徴について比較分析し、両者の共通点と相違点を述べている。そして、テストの内容や発話の抽出方法においてどちらのテストにも長所と短所があるため、テストの実施者がテストの目的やテストを実施する際の環境に応じて使い分けることが望ましいと提言している。牧野 (1991) では、OPIを実施する際には学習者の言語的習熟度をより正確に測定するためにロールプレイの回数を多くすることを提案している。また、テストを実施する際には、インタビュアーとなるテスト実施者や教師が必要以上に学習者を手助けすることのないように注意すべきであると指摘している。

テストの信頼性についてブラウン (1999) は、環境、実施方法、受験者自身、

言語科学研究第10号（2004年）

採点方法、テスト本体において測定誤差が生じるとしている。そして、テストを実施する際には様々な段階で測定誤差が生じるため、テストの実施者がどの段階でどの程度の測定誤差が生じるかを十分に認識し、測定誤差を最小限に留められるように努力することが必要であると述べている。スピーキングテストに関わる要素について、McNamara (1995) では、評価者 (Rater)、評価基準 (Scale criteria)、発話 (Performance)、タスク (Task)、協力者 (Candidate)、潜在的能力 (Underlying competence)、対話者 (Interactants) を挙げている。一例として、対話者が友人か全くの他人か、同性か異性かによってもテスト結果が異なっていることが報告されている。また、Skehan (1998) ではMcNamara (1995) の要素にタスクの質 (Task qualities)、タスクの条件 (Task conditions)、第1言語と第2言語の使い分け (Ability for use Dual-coding) を加えより詳細に定義している。

スピーキングテストの信頼性に関する先行研究では、評価者間での信頼性に関するもの、テストの内容や構成、テストの長所と短所について述べたもの等がある。評価者間の信頼性について、庄司 (1995) はACTFL-OPIを参照して留学生の口頭能力試験を改訂して、日本語コースの修了試験として実施した結果を報告している。テストは面接とロールプレイによって構成され、テストで抽出された発話は評価者によって談話形成、文法、語彙、流暢さ、発音、会話運用能力、社会言語学的能力の7項目についてそれぞれ5段階で評価された。その結果、改訂前の口頭能力試験では項目によって評価者間に信頼性に大きなバラつきが見られたが、改訂後は全ての項目において評価者間の信頼性で高い結果が得られたとしている。評価者間の信頼性が高まった要因としては、評価表が具体化されたことで評価の際に何に着目するのかが明らかになったことや、評価表の項目内基準を3段階から5段階にしたことにより評価の客観性が増したこと等を挙げている。また、安井 (1994) でも、発話を評価する際に評価方法を検討し、評価表や評価項目内の基準を具体的なものにした結果、評価の信頼性と客観性が高まったと報告している。さらに、測定の際にビデオテープとオーディオテープを併用したことで、音声だけでは判断できない点が明らかになり、受験者の反応や態度も評価できるようになった点で効果があったとしている。以上の各研究から、スピーキングテストの評価の信頼性を高めるためには、評価者に具体的かつ明確な評価基準を提示することや、評価項目数を増やして多方面から総合的に評価させることが

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

有効であると思われる。

Kuo&Jiang (1995) では、OPIとSOPIの二つのスピーキングテストを詳細に比較分析し、それぞれのテストの長所と短所を報告し、OPIはテストターや教師が直接学習者に対面して発話を抽出しているため、音声に回答する形式のSOPIよりも臨機応変に対応できることから柔軟性があるとしている。しかし、OPIはインタビュアーの性格やインタビュー技術の違いが抽出する発話に影響を及ぼす可能性が高いとしている。それに対して、SOPIは録音されたテープを使用するため、インタビュアーの影響が出ることがなく、常に安定した結果を得られる点でのテストの信頼性が高いと述べている。また、Shohamy (1994) では、OPIとSOPIの二つのテスト内容を比較分析して、SOPIに比べOPIは発話者から聞き出している言語機能の数、トピックの数、およびトピックの種類が学習者の習熟度によってバラつきがあることを指摘している。また、両テストから標本抽出した発話サンプル間にはコミュニケーションストラテジーや談話の特徴にも違いが見られたと報告している。

1.2 非母語話者の発話に対する母語話者の印象および評価に関する研究

非母語話者の発話に対する母語話者の印象や評価に関する先行研究には、一般人の評価と教師の評価を比較したものが数多く見られる。Hadden (1991) では、非母語話者の発話に対する英語教師と一般人の評価の違いについて調査している。実験では、ビデオテープに記録された中国語母語話者の発話した英語をネイティブスピーカーの英語教師と一般人に見てもらい、24の質問について5段階で評価してもらった。質問項目は、言語的能力、理解、性格、表現内容、非言語行動の5つに分類された。分析の結果、言語的能力の評価において一般人の方が教師よりも学習者の能力を高く評価する点で有意差が見られたが、他の4項目では両者の評価の違いに有意差は見られなかったと報告している。

Okamura (1995) では、非母語話者の発話に対する日本語教師と一般日本人の評価の違いと教師の教育経験の差による評価の違いを比較分析している。実験では学習者の発話が録音されたテープを両者に聞いてもらい、文法、流暢さ、適切さ、語彙、理解、発音の6項目について5段階で評価してもらった。その結果、全体的に教師の評価は一般の日本人に比べて厳しいものであり、特に文法上の正確

言語科学研究第10号（2004年）

さにおいて教師は厳しい評価を下していた。また、教育経験の差による教師間の評価では、経験の豊富な教師の方が経験の少ない教師よりも評価が厳しくなる傾向が見られたが、統計的には有意差が見られなかったとしている。そして、日本語の会話能力を評価する際には一般日本人も教師も相手に通じるかどうかを最も重要な評価項目として挙げていた点で共通していたことも報告している。

中川・石島（1998）でも、一般日本人と日本語教師が発話を評価する際の評価について調査し、両者の評価の特徴を比較分析している。実験では、両者に学習者の発話ビデオを見てもらった後に自由に印象を話してもらった。調査者はコメント結果を内容、文法・語彙・表現、音声、会話運用能力、社会的言語表現、非言語行動の6つに分類した。その結果、日本語教師は常に文法・語彙・表現等の言語規則に着目して評価していたが、一般日本人はあらゆる評価項目を総合的に加味して評価を下していたとしている。さらに、一般日本人の特徴として、表情やジェスチャー等の非言語的な情報も重要な評価基準として捉えていると報告している。

渡部（2002）では、中川・石島（1998）と同様の調査方法で実験を行ったが、評価者を母語話者教師、非母語話者教師、一般日本人、学習者に広げ、それぞれの評価基準を比較検討している。その結果、母語話者教師は他の評価者に比べて会話の展開に関するコメントが見られたが、非言語に関するコメントが少なかったと報告していて、学習者と非母語話者教師は意志の疎通ができているかどうかに関してのコメントが共通して見られたことや、非言語に関するコメントが多かったとしている。そして、一般日本人は他の評価者に比べて流暢さに関するコメントが多かったとも述べている。

非母語話者の発話評価に関する研究結果から、評価者によって評価基準が異なっていることが明らかになり、教師は発話を評価する際に文法の正確さに焦点を当てて評価しているが、一般人は総合的な観点で捉えて発話を評価していることが判明した。また、一般人は表情やジェスチャーなど非言語行動についても評価項目として捉えている点で教師と違いが見られた。

ここまで、様々な先行研究を概観してきたが、各分野の研究から明らかにされたことをまとめておく。まず、スピーキング能力は様々な要素から構成されていることや、教育現場ではACTFLのガイドラインがスピーキング能力の測定基準

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

として広く活用されていること、さらにOPIが学習者の習熟度を測定する際に使用されていることがわかった。次にスピーキングテストを実施する際には、様々な段階で測定誤差が生じることから、テストの信頼性を高めるためには、テストの実施者が評価基準を明確にしたり、評価項目をより詳細にしたりすることが有効であることも判明した。そして、非母語話者の発話に対する母語話者の評価においては、評価者によって評価の観点が違うことも明らかになった。これらの事実をふまえ、本研究ではスピーキング能力に対する教師の評価に焦点を当て、教師の教育経験の多少は非母語話者の発話を評価する際にどのように影響しているのかを明らかにする。

2. 本研究

2.1 質問

ここでは、先行研究で明らかにされた事実に基づいて、本研究の質問を提示する。先行研究では、スピーキング能力の構成要素や教育現場で実際に使用されているスピーキング能力の測定基準、スピーキングテスト、非母語話者の発話に対する母語話者の評価について述べた。これらの先行研究をもとに、本研究では教師の教育経験の違いが非母語話者の発話を評価する際にどのように影響しているかについて考察する。本研究においては、次の質問を提示する。

質問：教育経験年数の多少は、非母語話者の発話を評価する際にどのように影響を及ぼすか。

なお、本研究では、教育経験5年以上の教師を教授年数の多いグループ、教育経験5年未満の教師を教授年数の少ないグループと定義する。

2.2 調査協力者

調査協力者の選定にあたっては、現在国内で教授活動に携わっている日本語教師という条件を設定した。調査協力者数は40人で、その内訳は女性38人男性2人、年齢は28歳から70歳までであった。また、教師の教授年数は10ヶ月から13年ま

言語科学研究第10号（2004年）

で、平均教授年数は5.6年であった。

2.3 調査で使ったインタビューテープ

今回の調査には、牧野他（2001）のACTFL-OPIの中級と上級のインタビューテープから5つのサンプルインタビューが評価対象のテープとして選択された。インタビューのトピックは、中級テープでは「日本と自分の国での化粧方法の比較」「故郷の紹介」「趣味について」の3つが、上級テープでは「女性の生き方」と「ドラマの内容説明」の2つが使用された。いずれのテープも約3分程度の長さであった。

非母語話者のインタビューテープを選択する上では以下のような点に留意した。まず、ACTFL-OPIはインタビューとロールプレイから構成されているが、今回の調査では発話相手から受ける影響を少なくするために、ロールプレイは除外した。また、実際のOPIでは一人当たりの質問時間が30分程度と長く、その中に様々なトピックが盛り込まれているが、今回はこの中から、なるべく実際の教育現場でスピーキングテストにも使用されているものを選んだ。そして、テストの実施者が発話者を必要以上にリードしたり、発話者を助けたりしていないかという点にも配慮した。さらに、このインタビューテープを選択したもう一つの理由として、ACTFL-OPIのインタビューテープは、既にOPIテスター資格保持者によって発話者の言語習熟度が分析されていて、どのレベルに該当するのかが明らかにされている点を付け加えておく。インタビューテープが既に分析されていることは、本研究において調査協力者に評価してもらう際にも評価項目や評価基準の設定においてその信頼性が高いと言える。

2.4 評価タスク

評価タスクは、全体的評価、分析的評価、教師の自己基準、評価タスクに対する内省コメントの4つを実施した。

全体的評価では、教師が日頃の教授活動において、非母語話者の発話をどのように評価しているのかを調べた。ACTFL-OPIで中級レベルと判定されている非母語話者のインタビューテープを聞いて、その能力を全体的な印象で評価してもらった。全体的な印象で評価してもらったのは、調査協力者が日頃教育現場で

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

行っている評価方法と同一であり抵抗が少ないからである。なお、インタビューテープで用いられたトピックは「日本と自分の国との化粧方法の比較」である。調査協力者はインタビューテープを聞いた後に評価シートに5段階で評価した。評価シートの基準は1が「とても下手」、2が「やや下手」、3が「普通」、4が「やや上手」、そして5が「とても上手」という定義になっている。

分析的評価では、調査協力者に非母語話者のインタビューテープをテキストタイプ、文法、音声、語彙、流暢さ、機能、内容の7項目について項目ごとに評価してもらった。今回の調査では、ACTFL Provisional Revised Speaking Guidelines (1999) に基づいて、これらの評価項目を決定した。ここで使用されたインタビューテープは中級と上級の2種類で、それぞれACTFL-OPIで中級レベルと上級レベルに判定されているものである。中級テープのトピックは「故郷の紹介」と「趣味」、上級テープのトピックは「女性の生き方」と「ドラマの内容説明」であった。調査では、まず調査協力者にテープに録音された中級レベルの発話を聞いてもらい、評価シートに項目ごとに5段階で評価してもらった。この際項目内における5段階評価の基準は、ACTFLのガイドラインの習熟度分類に対応するように設定した。具体的には1と2が初級、3が中級、4が上級、5が超級に対応するように設定された。また、評価の際には5段階評価に加えテープを聞いて何か気がついた点や問題点があればコメント欄に自由に記述してもらった。そして、中級レベルの発話を評価した後に上級テープの発話を同様の方法で評価してもらった。発話を評価する際にテープを聞く回数は中級テープ、上級テープともに1回であった。なお、今回の調査ではトピックの影響を避けるためにカウンターバランスをとり、トピックと評価の順番の組み合わせを考えて実施した。40人のグループを10人ずつ4グループに分け、それぞれ評価するトピックの順番を入れ替えた。そして、4グループに分ける際には、グループ間の教授年数が均等になるように配慮した。

教師の自己基準では、教師が何に着目して発話能力を評価しているのかを調べるために、全体的評価の際に教師内の評価基準を洗い出した。調査協力者にインタビューテープの発話を全体的な印象で評価してもらった際に、どういう基準で5段階評価を決定したのかという理由を評価シートに記述形式で回答してもらった。また、全体的評価で使ったインタビューテープだけでは教師の持っている

言語科学研究第10号（2004年）

自己基準が全て洗い出せない可能性があるため、この質問とは別に日頃の教授活動で発話を評価する際の評価基準についてもあわせてたずね、記述形式で答えてもらった。評価タスクに対する内省コメントでは、今回の調査の最後の一連の評価タスクを通して感じたことや、自分自身の気づきについて記述してもらった。また、発話を評価する際に評価の仕方や意識がどのように変化したのかをできるだけ詳しく述べてもらった。

2.5 調査手順

調査協力者が全員現職の日本語教師であるため、調査者が設定した調査日程の中から協力者に都合の良いものを選択してもらい、調査者が設定した調査場所に出向いてもらう形で行われた。調査場所は神田外語大学大学院研究室及び都内にある語学教育機関の会議室、公共施設の会議室である。調査にあたっては、評価対象となる非母語話者の発話が録音されたテープを再生するためのテープレコーダーが使用された。

調査手順は、全体的評価、教師の自己基準、項目別評価（中級テープ、上級テープ）、評価タスクに対する内省コメントの順に実施された。調査にかかった時間は全体で40分程度であった。

2.6 分析方法

調査で得られたデータを統計処理する際に統計処理ソフトのStat Viewを使用した。

全体的評価においては、テープに録音されたインタビューの発話を全体的印象で5段階評価してもらった。一方、分析的評価のデータは、インタビューテープを項目ごとに評価基準に照らし合わせながら5段階評価してもらった。全体的評価と分析的評価の測定結果はグループ全体および経験の多いグループと少ないグループごとに集計され、平均点と標準偏差が出された。また、測定結果について一般化が図れるかどうかを調べるために統計処理を行った。

教師の自己基準に関する記述データは、ACTFL-OPIの測定基準をもとに調査者によって10項目に分類された。また、評価タスクに対する内省コメントのデータは、調査者がコメントの種類ごとに5種類に分類し、教授年数の多少のグルー

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

プごとにどのような傾向あるのかをまとめた。自己基準に関するデータと内省コメントのデータはグループ全体および教授年数の多いグループと少ないグループごとに集計された。さらに、教師の自己基準については項目別にそれぞれのグループの総数に対する自己基準の割合が出された。

3. 結果

3.1 全体的評価

調査協力者全体の中級テープの評価平均点および標準偏差は表1に示す通りである。教授年数の多いグループの方が年数の少ないグループよりも評価平均点が低かった。この結果について1要因の分散分析を行ったところ、わずかに有意には至らなかったが、やや関係が見られた ($0.05 < P < 0.07$)。

表1. 全体的評価での中級テープの評価平均点および標準偏差

	教授年数		F 値	P 値
	多いグループ	少ないグループ		
平均点	2.83 (0.60)	3.33 (0.72)	3.56	0.067

3.2 分析的評価

分析的評価についての測定結果を以下では中級テープ、上級テープの順に見て行く。まず、グループごとの中級テープ評価平均点および標準偏差は表2に示す通りである。項目ごとに評価平均点の変化を見ると、機能で教授年数の多いグループの方が高かったが、それ以外の6項目では教授年数の少ないグループの方が高かった。特に、文法においてはグループ間の差が非常に大きく、統計的にも有意差が見られた ($p < 0.05$)。

言語科学研究第10号 (2004年)

表2. 教授年数の多少による中級テープ評価平均点および標準偏差

評価項目	教授年数		F 値	P 値
	多いグループ (n=18)	少ないグループ (n=21)		
テキストタイプ	3.50 (0.71)	3.62 (0.92)	0.20	n. s
文 法	3.00 (0.69)	3.48 (0.75)	4.23	0.047
音 声	2.89 (0.90)	3.01 (0.63)	0.71	n. s
語 彙	2.33 (0.91)	2.71 (1.01)	1.52	n. s
流 暢 さ	2.89 (0.83)	3.05 (0.92)	0.31	n. s
機 能	2.39 (0.50)	2.33 (0.66)	0.09	n. s
内 容	2.33 (0.77)	2.48 (0.93)	0.27	n. s

上級テープ評価平均点および標準偏差は(表3)は、機能で教授年数の多いグループの方がやや高く、内容では両グループの評価平均点は同じであった。それ以外の5項目では、教授年数の多いグループよりも年数の少ないグループの方が高かった。しかし、統計的にはどの項目でも有意差は見られなかった。

表3. 教授年数の多少による上級テープ評価平均点および標準偏差

評価項目	教授年数		F 値	P 値
	多いグループ (n=18)	少ないグループ (n=21)		
テキストタイプ	4.74 (0.56)	4.71 (0.46)	0.02	n. s
文 法	4.47 (0.61)	4.33 (0.80)	0.39	n. s
音 声	4.16 (0.50)	4.10 (0.44)	0.18	n. s
語 彙	4.32 (0.67)	4.05 (0.59)	1.81	n. s
流 暢 さ	4.26 (0.56)	4.05 (0.59)	1.40	n. s
機 能	4.11 (0.88)	4.19 (0.68)	0.12	n. s
内 容	4.05 (0.71)	4.05 (0.50)	0	n. s

中級テープの評価平均点と上級テープの評価平均点について、教育経験のグループごとに比較すると、テキストタイプ、文法、音声、語彙、流暢さの5項目

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

に関しては、経験の多いグループは経験の少ないグループに比べ、中級テープをより低く評価し、上級テープをより高く評価していることが判明した。これらの項目の多くは言語的正確さを測定するものであった。しかし、この結果については統計的に有意ではなかった。

3.3 教師の自己基準

発話能力を評価する際にどのような基準を用いて評価を行っているかに関する自由記述データは表4の通りである。これに基づいて、教授年数の多少によるグループごとの測定結果を示したものが表5である。分析的評価で評価項目として使用した7項目について見ると、文法、音声、語彙といった言語的正確さを測定する項目では両グループとも高い割合を示していたが、内容はどちらのグループでも割合が低かった。また、項目ごとのグループ間の差を見ると、文法、音声、語彙、機能ではグループ間であまり違いは見られなかったが、流暢さでは教授年数の少ないグループに比べて教授年数の多いグループの方が割合がかなり低かった。ただし、この調査は記述形式で行われたため、調査協力者の評価基準が全て記述されなかった可能性がある。したがって、教師の自己基準について考えるには、評価タスクに対する内省コメントで記述された内容と照らし合わせてみる必要がある。

表4. 教師の自己基準の項目とその基準内容

1. テキストタイプ	…文の長さ、話のまとまり
2. 文法	…文型、表現、助詞、活用、時制
3. 音声	…発音、アクセント、イントネーション
4. 語彙	…使用語彙の種類および量
5. 流暢さ	…発話のスピード、滑らかさ、言い直し
6. 機能	…言語を使って目的が遂行できているか
7. 内容	…話の内容は具体的か抽象的か
8. 理解	…相手の言っていることを理解しているか
9. コミュニケーションストラテジー	…あいづち、聞き返し
10. 社会言語学的能力	…待遇表現の使用

言語科学研究第10号 (2004年)

表5. 教授年数の多少による教師の自己基準の人数と割合

評価項目	教授年数			
	多いグループ		少ないグループ	
	人数(人)	割合(%)	人数(人)	割合(%)
テキストタイプ	8	44	6	29
文 法	17	89	19	95
音 声	17	89	19	95
語 彙	12	67	16	76
流 暢 さ	7	39	13	62
機 能	11	61	12	57
内 容	4	22	6	29
理 解	11	61	16	76
コミュニケーションストラテジー	9	50	7	33
社会言語学的能力	4	22	9	43

3.4 評価タスクに対する内省コメント

調査の最後に評価タスクを通して感じたことや、自分自身の気づきについて内省してもらったコメントの項目および種類は表6の通りである。これに基づいて、教授年数の多少によるグループごとの測定結果を示したものが表7である。全ての項目において、教授年数の少ないグループの方が年数の多いグループよりもコメント数が多いかまたは同じであった。この結果から、教授年数の少ないグループの教師はこの評価タスクを通じて、より多く気づきがあったことが判明した。

表6. 評価タスクに関する感想・気づきの項目とその内容

1. 評価項目数の増加	評価基準の対象となる項目が増えた
2. 項目内基準の詳細化	項目内の評価基準量が増え詳細になった
3. 評価項目内の基準と習熟度の意識化	項内の基準と発話の習熟度の対応を認識
4. 評価基準の意識化	主観的評価から評価基準に基づいた評価へ
5. 全体的評価から分析的評価へ	全体的評価から項目ごとの分析的評価へ

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

表7. 教授年数の多少による内省コメント数（単位：人）

内省コメントの種類	教授年数	
	多いグループ	少ないグループ
評価項目数の増加	4	7
項目内基準の詳細化	4	6
評価項目内の基準と習熟度の意識化	4	4
評価基準の意識化	8	11
全体的評価から分析的評価へ	1	3

4. 考察

本研究では、教育経験の違いが非母語話者の発話を評価する際にどのような影響を及ぼすかについて、全体的評価、分析的評価、教師の自己基準、評価タスクに対する内省コメントといった4種類のデータを収集して分析した。

全体的評価では中級レベルのインタビューテープを使用して5段階評価を行った。その結果、教授年数の少ないグループの評価平均点が高くなる傾向が見られたが、統計的には、やや関係は見られたが有意には至らなかった。中級レベルと上級レベルのインタビューテープの発話を項目別に評価した分析的評価では、中級テープを評価した際に、全体として教授年数の少ないグループの方が年数の多いグループに比べ高い評価平均点をつける傾向が見られた。特に文法においては、評価平均点が高くなる点で有意差が見られた。一方、上級テープの評価では教授年数の多いグループの方が高い評価平均点をつける傾向が若干見られたが、7項目いずれにおいても評価平均点にグループ間の有意差は見られなかった。教師の自己基準の記述では、文法、音声、語彙といった言語的正確さを測定する項目では経験の多少に関わらずどちらのグループでも高い割合を示していたが、内容は両グループとも割合が低かった。評価タスクに対する内省コメントでは、全ての項目において教授年数の少ないグループの方が年数の多いグループよりもコメント数が多いかまたは同じであった。以下では、これらの分析結果をもとにして、教育経験の多少が発話を評価する際にどのように影響を及ぼしていたかを考察する。

言語科学研究第10号（2004年）

分析的評価を行った際に、文法の評価平均点でグループ間に違いがあったことから、経験の多いグループと少ないグループでは評価の仕方の違いがあったと推測される。しかしながら、教師の自己基準の記述では、教授年数の多いグループで89%、教授年数の少ないグループでも95%の調査協力者が文法を評価基準として挙げていたため、評価の際に文法に注目していた点でグループ間に差はなかったことが明らかになった。また、調査の最後に実施した評価タスクに対する内省コメントを見たところ、教授年数の少ないグループの方が年数の多いグループに比べコメント数が多かったことも分かった。これらの事実から発話を評価する際に文法に着目しているという点では、どちらのグループも同じだったものの、評価する際に使用していた項目内の評価基準の量や質に違いがあったと思われる。すなわち、教授年数の少ないグループは教授年数の多いグループに比べて項目内基準の量も少なく、基準の内容も明確になっていなかったため、年数の少ないグループの方が項目別の評価平均点が高くなった可能性がある。この結果について、発話を評価する際に教師は一般人に比べて文法において評価が厳しくなるという先行研究 (Hadden, 1991 ; Okamura, 1995) を参考に考慮すると、教師は教育経験を積むにしたがって、文法に関する項目内の知識の量が増え、より詳細な評価基準を身につけていくと推測される。その結果として教授年数の多い教師は年数の少ない教師に比べ文法の評価が厳しくなると思われる。しかし、今回の調査ではグループ間の評価平均点差は中級テープでは有意だったが、上級テープでは有意差は見られなかった。したがって、グループ間で評価平均点に差が見られるのは中級レベルに限定される。本研究結果からは、教授年数の多いグループの方が年数の少ないグループよりも評価の際に使用していた項目内の基準がより多く詳細なものとは言えない。

もうひとつ、本研究の分析的評価から経験の多いグループは少ないグループに比べてテキストタイプ、文法、音声、語彙、流暢さの5項目に関して、中級テープをより低く評価し、上級テープをより高く評価する傾向が見られた。この点に関しても教師の自己基準の記述と照らし合わせてみると、文法、音声、語彙といった言語的正確さを測定する項目では、経験の多少に関わらずどちらのグループでも高い割合を示していたため、グループ間に評価の際の着目点に差はなかったことがうかがえる。しかし、評価の傾向に違いがあったということは、グルー

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

プによって評価の仕方に差があったことを意味している。経験の少ないグループは経験の多いグループに比べ全ての項目においてコメント数が多かったという評価タスク後のコメント数の比較を参考にとすると、経験の少ないグループは学習者の発話を漠然と評価していたため、能力の違いを明確に弁別できていなかったと推測される。これに対して経験の多いグループは上手な発話と下手な発話を明確に識別できていたと思われる。以上のことから、教師は教育経験を積むにしたがって、発話を識別する能力を身につけていく傾向があるのではないだろうか。そして、この識別力は言語的正確さを測定する項目に関して言えるのではないかと思われる。さらに、教授年数の多い教師は年数の少ない教師に比べ学習者の発話を聞いた際に単に上手か下手かだけではなく、どの習熟度に当てはまるのかということも考えながら評価しているのではないだろうか。ただし、統計的には有意ではなかったため、これらの可能性については本研究の結果だけでは言えない。

5. 結論

5.1 教育経験による評価の仕方の違い

本研究の調査結果から、教育経験によって特定の項目における評価の仕方に違いが現れていることが明らかになった。中級テープを評価する際に、7項目のうち文法でグループ間に有意差が見られ、教授年数の少ないグループの方が教授年数の多いグループよりも評価平均点が高くなる傾向が確認された。また、全体的評価の際に行った教師の自己基準の洗い出しでは、両グループともほぼ同じ割合で文法を評価項目に挙げていたことから、評価平均点に差が生じたのは、グループ間で項目内の評価基準の量や質に違いがあったためと思われる。すなわち、教授年数の少ないグループは文法を評価する際の着目点や項目内の評価基準が明確になっていない状態で漠然と評価していたものと推測される。しかしながら、上級テープの評価の際には文法においてもグループ間の差は見られなかったことから、必ずしも教育経験を積んだからと言って、評価項目内の知識量が増え、より詳細な評価基準を身に付けていくとは限らないことが分かった。

5.2 本研究の限界と今後の課題

調査で使用したインタビューテープについて振り返ると、今回の調査においては、既に習熟度が判定されている点で信頼性があるとして、OPIのインタビューテープを材料として採用した。しかし、それ以外の点に関しては十分に配慮をしていなかったため、テープの長さやトピック選択において本研究において妥当であったかどうかは疑問が多く残る。次回の調査を行う際には、この点について調査者自身がもっとよく吟味してテープの選定を行うことが必要だと痛感した。次に測定方法についても検討の余地がある。今回の調査では、非母語話者の発話を評価する際に、全体的評価と分析的評価の二つを実施した。評価項目を決定する際にはACTFL-OPIを参考にして7項目を選択したが、教師の自己基準の洗い出しを行ったところ、この7項目以外にも社会言語学的能力やコミュニケーションストラテジー等を日常で発話能力を評価する際に評価項目として挙げていた。したがって、評価項目を設定する際には何を評価項目にするかを再検討する必要があるだろう。また、本来のOPIにおいては、項目ごとに評価せずに受験者の習熟度を全体的評価で下していることから、今後は本来のOPIと同じ測定方法で調査することも必要ではないだろうか。その際には、最終的な判定は全体的評価で下すが、調査協力者が発話をどのように評価したのかを知るために、評価の過程を測定するのに適したthink-aloud法（発語思考法）や、評価後のインタビューを測定の際に併用することが望ましいと思われる。調査協力者に関しては、今回は調査の際に教育経験の違いを要因として用いたが、教師評価の研究としては、これ以外の要因としてOPIテスト資格保持者とOPIの資格の無い日本語教師の評価を比較することも教師評価がどのように進んで行くのかを調べる上では重要であるだろう。

5.3 教育的示唆

今回の調査を通して、発話能力の評価は非常に技術を要するものであることを再認識した。この発話評価に関する技術の向上を教師の自己研鑽だけに依存したままでは評価の信頼性を高めることは困難であると思われる。事実、教育経験を多く積んでも、特定の項目以外は教師間に差がないことが今回の調査で判明し

The Effects of Teaching Experience on the Evaluation of Japanese Learners' Speech

た。このことから、評価の信頼性を向上させるためには、やはり各教育機関内で明確な評価基準を作成したり、OPIテスト資格者を呼んで講義をしてもらったりするなどの形で評価トレーニングを実施する必要があると思われる。本調査の結果が、今後の教師教育において、各教育機関で評価トレーニングを実施することを促すきっかけになれば幸いである。

〔謝辞〕

本稿は2003年1月に神田外語大学大学院に提出した修士論文の一部を加筆修正したものです。執筆にあたり貴重なご助言を下さった堀場裕紀江先生と小林美代子先生に深く御礼申し上げます。

〔参考文献〕

- 庄司恵雄 (1995) 「日本語研修コースのための口頭能力試験」, 『日本語教育』 91号, pp108-119.
- 中川道子・石島満沙子 (1998) 「会話の上達度を計る評価基準」, 北海道大学留学生センター紀要第2号, pp169-184.
- バックマン・F・R (1997) 『言語テスト法の基礎』, みくに出版
- ブラウン・J・D (1999) 『言語テスト法の基礎』, 大修館書店
- 牧野成一 (1991) 「ACTFLの外国語能力基準およびそれに基づく会話能力テストの理念と問題」, 『世界の日本語教育』 第1号, pp15-32.
- 牧野成一・鎌田修・山内博之・斎藤真理子・荻原稚佳子・伊藤とく美・池崎美代子・中島和子 (2001) 『ACTFL - OPI入門—日本語学習者の「話す」能力を客観的に測る』, アルク
- 安井澄江 (1994) 「会話試験の評価に関する試み」 名古屋大学日本語・日本文化論集, pp111-139.
- 米田由貴代 (2000) 「OPIを授業に生かす—受講者から見たOPIワークショップのメリット／デメリット」 『月刊日本語』 4月号, アルク
- 渡部倫子 (2002) 「日本語口頭運用能力をどのように測定するか」, 日本語教育学会春季大会予稿集, pp179-184.
- American Council on the Teaching of Foreign Languages (1986): ACTFL proficiency guidelines. Hastings-on-Hudson, NY: ACTFL.
- American Council on the Teaching of Foreign Languages (1999): ACTFL Provisional Revised Speaking Guidelines. Hastings-on-Hudson, NY: ACTFL.
- Canale, M., & Swain, M. (1980): Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics Vol. 1, 1-47.
- Hadden, L. H. (1991): "Teacher and Nonteacher Perceptions of Second-Language

言語科学研究第10号 (2004年)

- Communication” Language Learning, 41-1,1-24.
- Hedge, T. (2000): Teaching and Learning in the Language Classroom, Oxford : Oxford University Press
- Kuo, J., and Jiang, X. (1995): “Assessing the Assessments: The OPI and the SOPI”, Foreign Language Annals, 30, No. 4, 503-512.
- McNamara, T. (1995) : Modelling Performance : opening Pandora’s box’. Applied Linguistics 16 : 159-179
- Okamura, A. (1995) : Teachers’ and Nonteachers’ Perception of Elementary Learners’ Spoken Japanese, The Modern Language Journal, 79, 29-40.
- Shohamy, E. (1994) : The validity of direct versus semidirect oral tests, Language Testing 11, No. 2, 99-123
- Skehan, P. (1998) : A Cognitive Approach to Language Learning, Oxford : Oxford University Press