

# 他領域との相関から見た会話試験

— 集中6か月コースの総括的評価としての試験作成過程から —

庄司 恵雄 岡山大学留学生センター

## 〔要 旨〕

評価は、教育課程において避けて通れない活動である。斯界には教授法、学習理論、教材などの研究に比べるとあまり研究されていない分野がいくつかあるが、評価はその一つと言え、口頭能力の測定・評価に関する先行研究は特に少ない。

本発表では、ある日本語集中コースのために開発、改良した総括的評価としての口頭能力試験を紹介し、その実施結果について、他領域の試験結果との相関を見るという角度から分析した。分析の結果、本試験は、聴解試験との間に高い相関を示したが、文法、文字、読解、作文の4分野の試験結果とは顕著な相関を示さなかった。このことは、口頭能力が聴解能力と深い関係を持つのに比べて、聴解以外の能力の習得とは関係が薄い可能性を示唆しているものと考えられる。

## 1. はじめに — 研究の背景

わが国の国立大学に正規生として留学する方法としては、自国の日本大使館で推薦を受ける(日本国政府奨学金留学生)か、自国政府もしくは所属大学の派遣を受ける(自国政府奨学金留学生、大学推薦留学生)か、さもなければ一般留学生選抜試験に合格する(一般私費留学生)かの三つの道がある。これらのうち、大使館推薦留学生に対しては、学部・大学院における留学生活に必要な基礎的な日本語力を養うことを目的に、1年または半年の集中予備教育を施すことになっている。東京外国語大学と大阪外国語大学の留学生日本語教育センターにおける学部留学生のための1年コースと、このほか全国15国立大学の留学生センターにおける大学院留学生のための6か月コース(以下「日本語研修コース」と呼ぶ)がそれである。

本研究は、日本語研修コースのための総括的評価として、口頭能力測定試験開発を主たる目的として出発した。現在は開発した試験方法及び評定方法の改良を継続課題としている。本発表は、この試験方法及び評定方法の改良研究の一環である<sup>(1)</sup>。

## 2 試験の概要

日本語研修コースの修了生がコース修了後開始する研究活動において最も一般的に直面すると想定されるコミュニケーション場面は、指導教官及びその周辺の日本人たちとの研究室などにおける会話である<sup>(2)</sup>。その場면을仮想して試験内容とした。

日本語の口頭能力は、従来日本語教師の直観に基づき、相対的に評価する規範参照(norm-referenced)方式によることが多かった。しかし、より分析的かつ客観的な評価法を確立するため、本研究では、基準参照(criterion-referenced)による評定方法を開発することを目標とした。開発に当たっては、ACTFL OPI(全米外国語教育協会 口頭能力面接試験、以下「OPI」と呼ぶ)からアイデアの多くを借用した。

基準参照による評定方法を開発することにした目的は次の3点にある。

- (1) 評定者の恣意性を排除する。
- (2) コース受講者の継続的なデータ蓄積結果を分析しやすくする。
- (3) コース及び学習者に対する分析的なフィードバック資料を得る。

### 2.1 試験の性質

本試験を到達度試験(Achievement Test)と能力試験(Proficiency Test)の折衷形態ともいふべき到達度能力試験(Pro-chievement Test)と性格づけた。理由は以下による。

本試験は集中教育課程の修了試験という総括的評価であるから、その基本的な性格は到達度試験であらざるを得ない。しかし、被験者は6か月の教育期間中に教室内のみで純粹培養的に口頭能力を獲得するのではない。むしろ教室内作業では言語知識の注入に重きが置かれ、運用練習は学習者個々人の日常的努力に委ねられる傾向がある。さらに、コース受講者には、日本語に関するなにかの知識を既に習得している者が含まれることが多い。このような背景から、修了試験の中でもとりわけ言語運用能力を測定する傾向の強い口頭能力試験では、能力試験の性格をも加味する必要がある。

こうして、試験内容においては初級のシラバスを意識しながら、試験方法においては現実に似せた場面を設定し、被験者が教室外で習得した能力の出現をも予想した試験づくりになった。

### 2.2 試験内容

試験方法は面接形式とした。被験者は1人ずつ試験室に入り、面接者である教師と面接する。試験室には被験者と面接者以外にはだれもおらず、面接の経過はすべて録音・録画される。

本試験の面接がOPIと異なる点は、レベル・チェック(level checks)も突き上げ(probes)も行わないことである。全被験者に対して一様に、いわば定食メニューの流れで面接を行う。面接者の技術差によって、抽出されるサンプルが不均一になることを最小限に抑えるためである。

試験内容は、面接者との対話及びロールプレイで構成される。

面接者との対話においては、易から難へと配列された四つのタスクが課され、各タスクにおいて評定者が観察すべき能力が評定基準に記述されている(表-1)。

表-1 面接者との対話内容

タスク	トピック	観察対象となる能力
1	一日の生活	基本的な語彙を用いた陳述能力
2	旅行	テンス・アスペクト・ムードを織り込んだ対話能力
3	私の専門	自己の専門について平易かつ手短かに説明する能力
4	私の国と日本	二つの文化の異同について比較説明する能力

ロールプレイにおけるタスクは2種ある。どちらの場面でも、被験者のロールは特別な負荷のない「近い未来の被験者自身」とした(表-2)

表-2 ロールプレイの内容

タスク	場面	面接者のロール	観察対象となる能力
1	指導教官主催のパーティ	指導教官夫人 (初対面)	・会話の開始→維持→終結 ・自己紹介 ・敬語使用
2	指導教官の研究室	指導教官	・依頼のディスコースを展開して欠席の許可をとる

## 2.3 評定方法

評定は、録音と録画により採集したサンプルを複数の教師が所定の用紙に記入する。評定用紙には、以下に述べる基準を記述し、評定者はその基準に照らして評定結果を用紙に記入するという仕組みである。

### 2.3.1 評定基準

OPIの評定規準(Assessment Criteria)は、「機能・タスク」(Global Tasks or Functions)、「場面・話題」(Context/Content)、「正確さ」(Accuracy)、「テキスト・談話の型」(Text Type)の4項目からなる。このうち「正確さ」には「文法」、「語彙」、「発音」、「社会言語学的法則の運用力」、「流暢さ」、「語用論的運用能力」の

6項目の正確さが包含される。したがって、「正確さ」に含まれる六つの下位項目は、評定全体の24分の1の比重しか占めないことになる。2.1に述べた理由から、到達度試験の性格をもつ本試験では、正確さをより重視することにした。そこで、OPIの「正確さ」に含まれる六つの下位項目を個々に取り出して、「機能・タスク」、「場面・話題」、「テキスト・談話の型」の3項目と並列することにした。こうして9項目の評定基準が浮かび上がった。

この9項目の基準で評定を試行したところ、「機能・タスク」と「場面・話題」については、能力判定の対象となるサンプルが均一にとれないことが判明した。そこでこの2項目を問題に組み込むことにし、評定項目を7に絞った。被験者からランダムにサンプルが出るのを待つのでなく、強制的にサンプルが抽出できる出題方法を採用したほうが内容的妥当性が高まるからである。また、項目のラベル名を牧野(1991)に従って改めた(表-3)。

表-3 評定基準

試行基準	改良新基準
①機能(できること)	}タスクに組み込む
②内容(話題)	
③談話構成	(1)談話形成
④発音	(5)発音
⑤語彙	(3)語彙
⑥文レベルの文法	(2)文法
⑦会話の管理力	(6)会話運用能力
⑧流暢さ 声の調子	(4)流暢さ
⑨社会言語学的処理能力	(7)社会言語学的能力

### 2.3.2 評定用紙

試行段階では各項目に下位の要素項目を掲げ、その達成度を数値化したものを合算するという方法を採用した。その結果、下位の要素項目が学習過程に即して配列された基準では評定者の評定のばらつきが比較的少ないのに対して、下位の要素項目を単純に合算する項目では評定者間の評定値のばらつきが大きいこと、また下位小項目が多いほどばらつきが著しくなることが判明した。そこで、小項目合算方式をやめ、すべての基準について学習過程に則した5段階のガットマンスケールを採用することにした。また、評定者の恣意を排除するため、評定基準の段階記述を精緻化した。

次に、各項目を $\alpha$ と $\beta$ の2段階の能力で構成した。 $\alpha$ 能力は予備教育の成果として、また、研究活動に入るに当たって、最低限習得しておくことが望ましい能力を意味し、

$\beta$ 能力は到達目標を上回る能力を意味する<sup>(3)</sup>。 $\alpha$ 能力がすべて習得されていれば「3」と評定され、そうでなければ習得の程度により「2」または「1」と評定される。「3」を得た者については、さらに $\beta$ 能力の有無を観察する。 $\beta$ 能力がなければそのまま「3」、ある程度習得されていれば「4」、十分習得されていれば「5」の評定が与えられる(表-4)。

表-4 7基準の内容

基準 \ 内容	$\alpha$	$\beta$
談話形成	文レベルの発話の維持	段落の形成
文法	文形成上の必須要素(助詞・述語・語順)の適切な使用	文形成上必須要素でない要素(文に意味を添える語・文と文を繋ぐ語・連体修飾語)の適切な資料
語彙	日常生活で使われる最小限の語彙の正確な使用	$\alpha$ 語彙以外の語彙の正確な使用
流暢さ	文レベルの発話に関する流暢さ	複文・段落レベルの発話に関する流暢さ
発音	文節単位の調音・アクセントの正確さ	文末イントネーションの自然さ・一文全体の音調的な日本語らしさ
会話運用能力	ねじれない状況における運用能力(自発性の維持)	ねじれた状況における運用能力(発話権の奪取・維持・譲与に関する談話管理)
社会言語学的処理能力	語レベルの待遇表現(挨拶・呼称・親族語)の適切な使用	敬体・敬語の適切な使用

この評定方法のポイントは、 $\alpha$   $\beta$ という2段構えの構造とガットマンスケールの組み合わせにある。被験者に $\alpha$ 能力がないと判断された場合は、 $\beta$ 能力の有無について評定する手間が省ける。また、かりに評定5に相当する能力の片鱗がうかがえても、より下位の能力が習得されていなければ、習得されている低いほうの評定値を与える。こうすることによって、評定者の主観が侵入する可能性が減殺できるわけである(図-1)。

図-1 評定用紙の一例

1. 談話形成  $\alpha$  : 文レベルの発話の維持  
 $\beta$  : 段落の形成

評定	$\alpha$	$\beta$
1	発話の大部分が単語レベルであり、文単位の発話が少ない。	
2	文レベルの発話がかなり現れるが、不自然に単語レベルに落ちることもあり、文レベルの発話を維持することができない。	
3	文レベルの発話が維持できる。	段落が形成できる兆しが全くない。
4		幾つかの文を並べて説明・描写という機能を果たそうとするが、結束性(cohesion)が十分でない。構造が非常に単純ではあるが、複文や連体修飾節も幾つか現れる。
5		十分な結束性を持った段落の形成ができ、かつ構造がやや複雑化した複文や連体修飾語が現れる。

### 3. 実施の結果

試験の開発を開始して以来、1994年 9月及び1995年 3月の2期にわたって実施データを得た。以下では、その双方を比較しつつ検討を進める。

#### 3.1. 被験者及び面接者・評定者

両回の被験者となった学習者のデータを表-5に示す。

表-5 被験者のデータ

	1994年 9月	1995年 3月
被験者数	男14人、女3人	男9人、女4人
平均年齢	29.9歳	30.8歳
出身地域	アジア 8、中南米 4、他 5	アジア10、欧州 2、中米 1
日本語学習歴のある者	8人 / 平均学習期間 3か月	2人 / 1年(自習)、2か月
日本滞在歴のある者	3人 / 半年 2、2年半 1	3人 / 2月 2、1年半 1
専門分野	農 5、工 5、医 3、他 4	教 8、農 4、経 1

面接は、機関所属の女性日本語教師3人(1994年 9月)または2人(1995年 3月)で担当した。日本語教師歴は4年から8年である。うち1人はOP1のテスター資格を持

つ。

評定には両回とも7人が当たった<sup>(4)</sup>。評定者の年齢は26歳から63歳までで、男性4、女性3である。日本語教師歴は4年から8年までである。OPIのテスター資格を持つ者1人とテスター講習受講経験者3人が含まれる。被験者の授業を担当したことがある者は4人、ない者は3人である。

### 3.2 信頼性係数

1994年9月実施の結果について、評定者をテスト項目と見なしたクロンバックの $\alpha$ 係数を求めた。テストの信頼性を検定する一方法として、評定者間の評定の一致度を測定するためである。

同一のサンプルについて、同一の評定者が新旧二つの方法で評定した。その結果、2.3.2で述べた評定方法の改良の前後で大きな差があることがわかった(表-6)。

表-6 信頼性係数 —— 改良前と改良後(1994年9月)

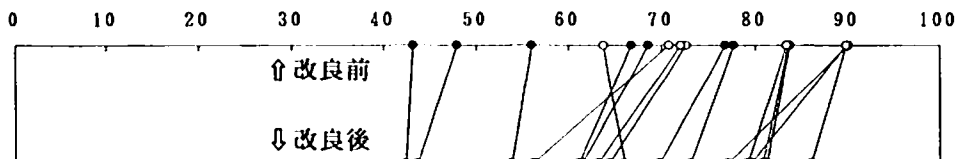
項目	機能	内容	談話	発音	語彙	文法	管理	流暢	社言
改良前	0.92	0.91	0.93	0.13	0.91	0.76	0.89	0.92	0.84
改良後	タスク組み込み		0.93	0.80	0.93	0.91	0.94	0.93	0.88

改良前には、「発音」の項目で問題外に低い数値が出たほか、「文法」と「社会言語学的処理能力」でも数値がかなり低い。改良後はこれら3項目のうち、「文法」と「社会言語学的能力」で一定の改善を見た。「発音」では大幅な改善を見たが、信頼性係数としては依然として低い水準と言わなければならない。

### 3.3 集団内評定分布

被験者集団内の評定分布を見ると、改良前には同じか非常に近似した評定を受けた被験者群が、改良後はさらに細かく差別化された。また、順位の入替わりが少数認められた(図-2)。

図-2 評定方法改良前後の評定結果の変化(1994年9月)



### 3.4 他領域との相関関係

会話試験と同時に修了試験として実施されたのは、1994年、1995年とも文法、文字、

読解、作文、聴解の5分野である。文法、文字、聴解は全く同一の試験問題であるが、読解は内容がやや異なり、作文は問題と採点方法がやや異なる。会話は上述の改良が施された。

これらの領域におけるテスト結果の相関を見た(表-7、表-8)<sup>(6)</sup>。

表-7 6領域の試験結果の相関係数(1994年 9月)

	文法	文字	読解	作文	聴解	会話	平均
文法	1.00	0.79	0.61	0.54	0.71	0.54	0.64
文字		1.00	0.62	0.70	0.76	0.65	0.71
読解			1.00	0.73	0.64	0.66	0.65
作文				1.00	0.54	0.58	0.62
聴解					1.00	0.85	0.70
会話						1.00	0.66

表-8 6領域の試験結果の相関係数(1995年 3月)

	文法	文字	読解	作文	聴解	会話	平均
文法	1.00	0.92	0.47	0.76	0.63	0.63	0.68
文字		1.00	0.70	0.68	0.70	0.74	0.75
読解			1.00	0.74	0.39	0.33	0.58
作文				1.00	0.79	0.63	0.63
聴解					1.00	0.79	0.62
会話						1.00	0.65

表-7から、会話と聴解、文法と文字、読解と作文との間に高い相関があることがわかる<sup>(6)</sup>。また表-8から、文法と文字との間に非常に高い相関があること、聴解と会話及び作文との間に高い相関があることがわかる。両回を通じて文法—文字、読解—作文、聴解—会話という組み合わせが浮かび上る。

### 3.5 評定所要時間

1995年 3月実施の試験において、評定に要した時間を測定した。本研究のメンバー5人と本研究にかかわらなかった一般教師4人の合わせて9人に評定を依頼した<sup>(7)</sup>。評定対象のサンプル時間(秒単位)を分母とし、申告評定所要時間(分単位)を分子として比を求めたところ、最短1.33倍、最長3.15倍、平均1.89倍という数値が得られた。

評定者を個別に検討したところ、共同研究者(平均1.82倍)が特に短時間で評定した傾向も、一般教師(平均1.78倍)が特に長時間を要した傾向もない。つまり、試験の開



発にかかわった予備知識ないしは習熟度と評定所要時間の間に顕著な因果関係は認められなかった。予備知識や熟練がなくとも、日本語教師ならサンプル時間のほぼ2倍の時間で評定が可能であることが明らかになったわけである(表-9)。

表-9 評定所要時間(倍数)比較

<習熟群>			<非習熟群>		
評定者 (性別)	所要時間 (倍)	備考	評定者 (性別)	所要時間 (倍)	備考
A (M)	2. 0 2	12被験者の平均	H (F)	1. 1 5	13被験者の平均
B (F)	1. 5 4	13被験者の平均	I (F)	1. 3 3	4被験者の平均
C (F)	1. 2 5	〃	J (F)	1. 7 1	〃
D (M)	1. 7 2	〃	K (F)	2. 0 3	5被験者の平均
E (M)	1. 2 5	〃	L (M)	2. 8 2	〃
F (F)	2. 0 9	〃	平均	1. 8 1	—————
G (M)	1. 1 5	〃			
平均	1. 5 7	—————			

#### 4. おわりに —— 分析から見える問題点

前章で明らかになった結果は、きわめて常識的な見方を証明したものと言える。すなわち、文法と文字は言語知識として深い関係にあり、読解と作文も補完しあう関係にある。聴解は会話に従属する関係にある可能性があるが、両者はやはり不離一体の関係にあるようである。ここでは、個々のテスト問題の内容的妥当性と採点もしくは評定方法の信頼性には問題がないことが前提である。もっとも、読解及び作文は、出題及び評価法において試行錯誤の過程にあり、内容的妥当性と信頼性に一抹の疑問があることを付言しておかなければならない。会話試験を含めて、各領域の試験のあり方については、今後もデータを蓄積することによって検証を継続する必要がある。

本試験には、上に論じた点以外にも多くの解決課題がある。OPIの場合は、テストターの訓練によってテストの信頼性が担保される。しかし、OPIのような訓練を一般の日本語教育の現場に望むことは現実的ではない。そこで試験のシステムを簡易化し、評定に習熟しなくとも容易に口頭能力が測定できることが必要になる。その意味では、本試験はまだまだ関係者に習熟を要求している面があることは否定できない。「だれにでも・どこでも・簡単に」という本研究課題の目標を達成するためには、面接方法の簡素化、評定方法の容易化という二つの大きな課題を解決しなければならない

い。

また、5段階のガットマンスケールが真に一般的な学習過程に即しているかどうか定かでない部分がある。そのような項目では、高い能力があっても低い評定値に記述されている能力が観察できないという理由で、低い評定値を与えざるを得ないというようなケースが出現している。現行の評定システムは、7項目において等し並みに一定の普遍的学習過程が存在するという仮定を前提としているが、それは果たして真か、真ならばそのヒエラルキーはどうなっているのか、という問題が必ずしもクリアされていない。とりわけ「発音」と「社会言語」にそれが当てはまる。このことは今後課題せられたきわめて基本的な研究課題である。

次に、現行の七つの基準で口頭能力の全容を覆い得ているのかという疑問がある。たとえば、現行法式では、聞き取りが障害になって生じるコミュニケーションの行き違いを評定することができないという指摘がなされているが、そのことをどのように克服するのかという問題がある。今後本試験が多くの機関で採用されることによって、さらに問題点が指摘されるとともに、改善意見が寄せられることを期待している。

#### 〔注〕

- (1) 本発表は、平成6年度文部省科学研究費補助(総合研究(B))研究「日本語研修コース標準口頭表現力測定検査法開発のための企画調査」(研究代表者 庄司恵雄)の研究成果の一部に基づいている。
- (2) 日本語研修コース修了者の日本語使用状況に関しては、黒野(1994)を初めいくつかの調査結果が報告されている。
- (3) 日本語研修コースの到達目標については、先行研究がない。仁科ほか(1991)や庄司(1993)を参考に一応の目安を定めた。なお、本年3月22、23両日、岡山大学の主催で国立大学日本語研修コース担当者連絡会議が開催された際明らかになった各機関の到達目標もほぼこれに一致するものであった。
- (4) 本研究のために、本科研共同研究者の岡本淑明、金澤眞智子、児島協子、酒井峰男、光元聰江、山内博之の諸氏(あいうえお順、いずれも岡山大学所属)の全面的な協力を得た。
- (5) 本発表のために、共同研究者の金澤眞智子氏に相関係数の算出作業などの面で協力を仰いだ。
- (6) 石田敏子(1992)pp.181-182によると、「+ .90～+1.00 非常に高い相関、+ .70～+ .90 高い相関」とされている。

- (7) このデータは本研究の遂行上得たものではなく、コースの成績処理を急いで進める必要上、他の一般教師の協力を得て求められたものに基づいている。

〔主な引用・参考文献〕

- (1) ACTFL (1989) "THE ACTFL ORAL PROFICIENCY INTERVIEW TESTER TRAINING MANUAL" New York
- (2) Alice C. Omaggio (1986) "TEACHING LANGUAGE IN CONTEXT" Heinle & Heile, Boston
- (3) Bachman, Lyle F. (1988) 'PROBLEMS IN EXAMINING THE VALIDITY OF THE ACTFL ORAL PROFICIENCY INTERVIEW' "SSLA" 10. Cambridge University Press New York
- (4) Hatch & Lazaraton (1991) "The research Manual: Design and Statistics for Applied Linguistics" Heinle Publishers
- (5) 石田敏子 (1992) 『入門日本語テスト法』 大修館書店 東京
- (6) 黒野敦子 (1994) 「研究生活で必要とされる日本語能力と留学生の実態」 『日本語研修コース修了生追跡調査報告書』 pp. 39-49 名古屋大学留学生センター
- (7) 庄司恵雄 (1993) 「日本語研修コースにおける会話試験に関するアンケート —— 集計の結果及び若干の感想」 『岡山大学留学生センター紀要』 第1号 pp. 67-78 岡山大学
- (8) —— (1994) 『大学院研究留学生の日本語使用実態に関する調査報告書』 岡山大学
- (9) —— 編 (1995) 『平成6年度科学研究費補助(総合研究(B))研究 日本語研修コース標準口頭表現力測定検査法開発のための企画調査 研究成果報告書』 岡山大学
- (10) 仁科喜久子・武田明子 (1991) 「理工系大学における外国人留学生の日本語能力に関する調査分析 —— 東京工業大学大学院課程を中心に —— 」 『日本語教育』 75号 pp. 113-123 日本語教育学会
- (11) 日本語教育学会編 (1991) 『日本語テストハンドブック』 大修館
- (12) 日本語教育学会調査研究第2小委員会 (1992) 『日本語口頭運用能力の評価に関する考察 —— 最終報告書 —— 』 日本語教育学会
- (13) 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 pp. 15-32 国際交流基金