

## JF日本語教育スタンダードに基づく日本語会話テストの開発 —中等教育における日本語学習経験者を受け入れるハノイ国家大学の試み—

ホアン, ティ マイ ホン

### 要旨

本研究は、JF日本語教育スタンダードに基づき、日本語会話テストを開発し、その有効性を検証することを目的とする。インタビュー、独話、ロールプレイという3つのタスクを設け、同じ評価基準に基づく2つの会話テストを開発した。日本語能力はほぼ同程度だが、学習環境が異なる2つの調査対象者にこのテストを行い、3名の評価者によって評価を行った。その結果、今回開発した会話テストは信頼性が高く、調査対象者の日本語能力を適切に測定できる妥当性が高いものであることが分かった。さらに、今回開発したテストは、教師の授業、学習者の日本語学習に肯定的な波及効果があるものであることが明らかになった。

[キーワード] ルーブリック、指標、信頼性、妥当性、波及効果

### 1. 研究の背景

ベトナムでは、2003年に日本語が課外科目として中等教育に導入された。これは中学校から高等学校卒業までの7年間のプログラムである<sup>1</sup>（以下、7年プログラム）。2005年より、日本語は課外科目から正規科目となり、ベトナム全国に広がっている<sup>2</sup>。そして、2012年には、7年プログラムの卒業生が日本語で大学入学試験が受けられるようになる見通しである。

また、2005年には、ハノイ国家大学・外国語大学（以下、ハノイ国家大学）付属の外国語専門高等学校において、高等学校のみの3年間プログラムの日本語教育が開始され（以下、3年プログラム）、このプログラムの学習者数も年々増加している<sup>3</sup>。

2008年7月には、3年プログラムの卒業生が、初めて日本語で大学入学試験を受け、ハノイ国家大学をはじめ、いくつかの大学に入学した。2008年と2009年の2年間で、大学に入学した3年プログラムの卒業生は52名であった。ハノイ国家大学に入学した3年プログラムの卒業生は2008年に5名、2009年に7名であった。入学した人数が少ないため、ハノイ国家大学では、高等学校の日本語教育と継続性のある日本語カリキュラムはまだ作成されていない。しかし、上述のように、2012年からは、7年プログラムの卒業生も誕生し、大学に入学する日本語既習者が増加することが予想されるため、日本語既習者のための継続的な日本語カリキュラムの作成が急務である。

適切なカリキュラムを作成するためには、大学に入学した学習者の日本語能力を把握しなければならない。大学入学試験科目としての日本語試験は、ベトナム教育訓練省規定に基づ

いて作成され、日本語初級を修了した段階のレベルの内容となっている。試験の形式は 90 分の筆記試験で、会話試験と聴解試験はない。そのため、大学入学試験の結果から日本語の文法、文字語彙の能力はある程度把握できるが、口頭表現能力は明らかではない。口頭表現能力を測定するためには、ハノイ国家大学の 1 年生の現行の会話テストを使用することも考えられるが、このテストは、ハノイ国家大学の 1 年生が学習した内容を評価するアチーブメントテストである。したがって、学習内容も学習環境も異なる 3 年プログラム、及び 7 年プログラムの学習者の口頭表現能力を測るには不適切であると考えられる。そこで、これらの対象者の口頭表現能力を測るためには、共通のスタンダードに基づき、会話テストを開発することが必要である。

石田 (1992: 6-8) は、テストの機能として、「教育上の機能」「研究上の機能」「教育管理上の機能」<sup>4</sup>の 3 つがあると述べている。ハノイ国家大学の会話テストでは、結果は全て点数で表される。そのため、「教育上の機能」における「教育目標の設定」と「教授活動の促進」という機能はある程度満たしていたが、「学習活動の促進」、「学習の方向付け」という機能は欠けていた。今後は、学習者への波及効果を考えたテスト開発が必要であると考えられる。

上述したような現状から、本研究は、共通のスタンダードに基づき、多様な学習者に対応可能な会話テストを開発し、その効果を考察することを目的とする。

## 2. 先行研究

口頭能力のプロフィシエンシーを測るテストとしては、OPI (Oral Proficiency Interview) が広く知られている。レベルは、初級、中級、上級、超級の 4 つの主なレベルに分かれている。評価基準は「機能・タスク」、「場面・内容」、「正確さ」、「テキスト・談話の型」という 4 つの観点から構成され、話題や語彙、文法項目は基本的に無限定である。テストは資格を持ったテスターが 30 分間のインタビューを行い、OPI 終了後に録音をもう一度聞いて、評価が下される (牧野: 1991)。

庄司 (1996) は、所属校の日本語研究コースの口頭能力修了テストを開発した。テストの内容は既習項目を対象としたアチーブメントテストで、学習者が要求される口頭表現能力は研究室場面などにおける教官とのフォーマルな会話を構成する基礎的能力である。テストは、実施上の経済的、時間的な問題から、OPI をそのまま採用するのではなく、18 分の口頭テストとした。そして、OPI より「正確さ」に重要性を置く必要があるという判断から、OPI の 4 つの評価の観点を「談話形成」「語彙」「文法」「発音」「流暢さ」「会話運用能力」「社会言語学的能力」の 7 つに拡大した。評価の観点ごとに具体的な内容記述を設けたことで、観点ごとにレベル判定を受けることができ、それによって学習者は自らの口頭表現能力の特徴を分析的に知ることができる。

一方、ベトナムの状況を考えると、会話テストの所要時間は 10 分程度であり、テストの際、

1、2名の評価者がテストターの役割をしながら評価する機関が多い。従って、OPIのように1つの日本語教育機関に訓練された複数のテストターを置くのは難しく、1人の学習者に30分も要するのは現実的ではない。庄司（1996）が開発した会話テストも所要時間が長く、また一般的な日本語口頭表現能力を測るものではないため、そのまま使用できない。また、OPIでは確かなレベル判定を下すために、録音を聞き返す必要があり、多くの時間を要することから、ハノイ国家大学で実施するには困難が多い。庄司のテストでは、評価の観点別に具体的な内容記述があるため、学習者は自分の日本語能力をイメージでき、この点については、ハノイ国家大学のテストにも導入可能である。

日本語教育の共通のスタンダードとして、国際交流基金による JF 日本語教育スタンダード（以下、JF スタンダード）がある<sup>5</sup>。JF スタンダードは CEFR<sup>6</sup>の考え方を基盤にして作られ、日本語教育の多様な現場で参照できる、より具体的な枠組みである。JF スタンダードでは、日本語を使って何がどのようにできるかという能力記述（「Can-do」）により、日本語の熟達度を A1、A2、B1、B2、C1、C2 という6レベルに分けている。

JF スタンダードは1本の木にたとえられ（図1参照）、「コミュニケーション言語能力」（木の根の部分）を基盤として、「コミュニケーション言語活動」（木の枝の部分）が行われると考える。「コミュニケーション言語能力」と「コミュニケーション言語活動」はそれぞれ下位分類され、全部で53の 카테고리<sup>7</sup>がある。そして、各カテゴリは「Can-do」からなっている。

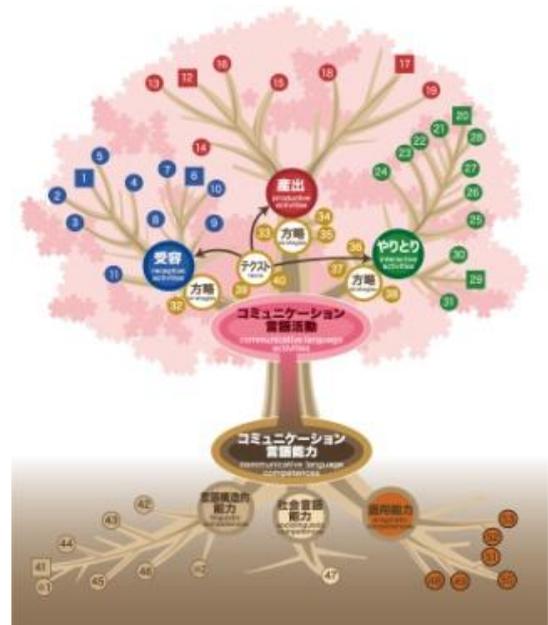


図1 JF スタンダードの木

JF スタンダードは、様々な教育実践に応用できるが、話す力を測るための「評価基準」を決め、「評価シート」を作成するのもそのひとつである（国際交流基金 2010b: 57-68）。「評価シート」作成の流れは、まず、JF スタンダードのカテゴリを参考にして、評価の観点（評価項目）と評価基準で扱うレベルを決める。そして、「みんなの Can-do サイト」（<http://jfstandard.jp/cando>）を利用して、該当するレベルとカテゴリの「Can-do」を選び、分かりやすい文章に書き換えるというものである。

本研究の会話テスト開発では、この「評価シート」の作成方法を参考にし、ルーブリックを用いた評価を行うことにしたい。ルーブリックとは、縦軸に評価項目を、横軸にレベルを配し、レベルごとに評価基準を記述した表である。高浦（2004: 91-148）は小中学校の学習活動を評価するためのルーブリックの作成、事例について紹介しているが、ルーブリックの評価基準・指標作りにおいては、事実に・行動的な記述化を試みる事が重要であると述べ

られている。そこで、本研究では、評価基準を作る際に、JF スタンダードに基づいた「Can-do」の記述（以下レベル記述）に加えて、より具体的な指標を設定したい。

### 3. 研究目的

本稿では、JF スタンダードに基づき、ハノイ国家大学で実施可能な会話テストを開発することを目的とする。

上記の研究目的を達成するために、次の研究課題を設定した。

研究課題 1. 開発したテストの信頼性はどうか。

研究課題 2. 開発したテストの妥当性はどうか。

研究課題 3. 開発したテストは、学習者及び教師にどのような波及効果をもたらすか。

### 4. 研究方法

#### 4.1 テストの概要

##### (1) 調査対象者

調査対象者は、①ハノイ国家大学附属外国語専門高等学校の高校 3 年生（3 年プログラムの学習者）14 名（日本語学習時間は約 492 コマ、1 コマ=45 分）、及び②ハノイ国家大学の日本語学部の 1 年生 14 名（日本語学習時間は約 416 コマ、1 コマ=45 分）、合わせて 28 名である。①と②は日本語能力がほぼ同程度であるが、学習環境が異なる対象者である。

##### (2) 評価者

本研究に協力してもらった評価者は、ハノイ国家大学で教えている日本人ネイティブ教師（以下、NT）1 名とベトナム人ノンネイティブ教師（以下、NNT）の 1 名である。そして、筆者も加わって評価者は 3 名である。

##### (3) テストとレベルの内容

ベトナム教育訓練省が定める中等教育の日本語教育、そして、ハノイ国家大学の大学 1 年生の日本語の時間数や内容を見ると、到達目標とされる能力は JF スタンダードの A2 レベルに当たる。そこで、テストの目標となる日本語レベルを JF スタンダードの A2 レベル（基礎段階の言語使用者）にした。そして、ベトナムの実情に合う内容、実際に日本人と出会う場面、またテストの所要時間を考えた上で、次のようにテストの内容を定めた。

まず、JF スタンダードの「コミュニケーション言語活動」から「産出と方略」、「やりとりと方略」を選んだ。そして、ベトナム学習者にとって有用性が高いと思われる「経験や物語を語る」を「産出と方略」から選び、「独話」のタスクを設定した。また、「やりとりと方略」から「社交的なやりとりをする」「情報交換する」「インタビューする／受ける」を取り出し、テストの「インタビュー」と「ロールプレイ」のタスクを設けた。各タスクの所要時間は「インタビュー」を 3 分、「独話」を 2 分、「ロールプレイ」を 3 分に設定し、タスク間の移行時

間を含めて10分間のテストとした。そして、同じテスト方式でタスクのトピックを変えても安定した結果が得られるかどうかを検証するために、同じレベル、パターンの2つのテストを作成した。テスト①とテスト②の内容を以下の表1にまとめる（詳細は資料1参照）。

表1 テスト①とテスト②の内容

タスク	インタビュー	独話	ロールプレイ
テスト①	日常的な情報について (10問程度)	家族について	パーティーに先生を招待する
テスト②	日常的な情報について (10問程度)	住んでいる町／ふるさとについて	先生に本を貸してもらおう

#### (4) ルーブリックの作成

評価を行うために、ルーブリック（詳細は資料2参照）を作成した。評価項目を決めるために、JFスタンダード「コミュニケーション言語能力」の13のカテゴリーから評価に必要なと判断した「使用語彙の領域」「語彙の使いこなし」「文法的正確さ」「音素の把握」「社会言語的な適切さ」「柔軟性」「発言権」「話題の展開」「一貫性と結束性」「話しことばの流暢さ」の10カテゴリーを選択した。これらを「発音」「文法」「談話形成」「語彙」「流暢さ」「会話運用能力」「社会言語学的能力」の7項目にまとめ、最も測りやすいタスクは何かを考えて、各タスクに評価項目を振り分けた。タスクと評価項目の関係は、以下の表2の通りである。

表2 テストの各タスクの評価項目

タスク	発音 (音素の把握)	文法 (文法的正確さ)	談話形成 (話題の展開・一貫性と結束性)	語彙 (使用語彙領域・語彙の使いこなし)	流暢さ (話しことばの流暢さ)	会話運用能力 (柔軟性・発言権)	社会言語学的能力 (社会言語的な適切さ)
インタビュー	●	●					
独話			●	●	●		
ロールプレイ						●	●

能力レベルとしては、JFスタンダードの「A2レベル」を中心に、「A2レベル未満」「A2レベルより上」という3つのレベルに分け、「Can-do」の記述を参考に記述内容を作成した。そして、客観的に評価できること、テスト実施中に評価を済ませられること、評価者間のずれを抑えること、あまり訓練していない評価者でも評価できることを考慮して、この記述内容に基づいて、具体的な指標を定めた。以下の表3はその一部である（詳細は資料2参照）。

表3 インタビュータスクのルーブリック（一部）

レベル	まだ努力が必要 (A2レベル未満)	目標を達成 (A2レベル)	目標以上を達成 (A2レベルより上)
文法 (JF: 文法的正確さ) ※	<ul style="list-style-type: none"> <li>・単純な文法構造があまり正しく使えない。本人が何を話しているか分かりづらい。</li> <li>1つ～3つの指標ができる</li> </ul>	<ul style="list-style-type: none"> <li>・単純な文法構造がほとんど正しく使える。本人が言いたいことが言いたい分かる。</li> <li>4つ～7つの指標ができる</li> </ul>	<ul style="list-style-type: none"> <li>・単純な文法構造が正しく使える。本人が言いたいことがよく分かる。</li> <li>8つ～9つの指標ができる</li> </ul>
	(指標) <ul style="list-style-type: none"> <li>□基本的な語順 (SOV)</li> <li>□「～か月/時間V」</li> <li>□「～AですがAです/Aです」</li> <li>□「～から～/～ですから」</li> <li>□「Vます」と「Vました」の区別</li> <li>□「～手段でV/歩いて来ます」</li> <li>□「～Vて、Vて～/V～」</li> <li>□「～VたりVたり～」</li> <li>□「～Vたいです」</li> </ul>		

※ (JF: 文法的正確さ) とはJFスタンダードのカテゴリーであることを示す。

総合評価は、7つの評価項目のうち、4つ以上の評価項目のレベル判定結果によって決めることとした。例えば、7つの評価項目のうち、4つが「A2 レベルより上」だと評価されれば、総合評価として「A2 レベルより上」だと判断した。

#### (5) テスト及び評価の実施方法

テストは以下のように行った。

- 1) 1名の調査対象者が2つのテストを受ける。そのために、調査対象者を半々に分け、一方のグループはテスト①、テスト②の順で、もう一方のグループはテスト②、テスト①の順で行う。
- 2) テストは1名の評価者(第1評価者)が調査対象者1名にテストを行いながら評価する。
- 3) テスト実施後、ルーブリックのフィードバック(ベトナム語版)と総合評価シート(詳細は資料4参照)を調査対象者に提供する。
- 4) 今回の実験のため、テスト実施後、各テストの録音を聞いて他の評価者(1名)が再評価を行った。評価の分担は表4の通りである。

表4 評価作業の分担

評価者 テスト	第1評価者	再評価者
テスト①	NNT②	NNT①
テスト②	NNT①(筆者)	NNT②(大学生と高校生の14人分) NT(大学生と高校生の14人分)

※ 高校生はNTとの会話テストを受けた経験がないため、テスト①と②の第1評価者はNNTとした。

### 4.2 分析データと分析方法

#### (1) 分析データ

本研究の分析に用いたデータは、以下の通りである。

- 1) 調査対象者全員のテスト①とテスト②の録音及びルーブリックによる評価結果
- 2) 調査対象者28人及び評価者2人に対する質問紙調査の結果
- 3) 調査対象者8人及び評価者2人に対するフォローアップインタビューの結果
- 4) テスト間で評価結果にずれがある箇所に関する評価者間のディスカッションの内容

#### (2) 分析方法

課題1の「開発したテストの信頼性かどうか」を検証するために、まず、テストの評価結果について評価者間のずれがないかを、次に、テスト①とテスト②の難易度が安定しているかを検討する。

課題2の「開発したテストの妥当性かどうか」に答えるために、まず、テストが調査対象者の日本語能力に合っているかどうかをテストの総合評価、質問紙調査の結果から検討する。次に、ルーブリックのレベル別評価項目の記述内容、及び指標の適切さを質問紙調査の結果とフォローアップインタビューによって検討する。

課題3の「開発したテストが、学習者と教師にどのような波及効果をもたらすか」を明らかにするために、ルーブリックのフィードバックについての調査対象者と評価者に対する質問紙調査の結果、及びテスト全体の内容についてのフォローアップインタビューの結果を分析する。

## 5. 結果と考察

### 5.1 テストの信頼性

#### 5.1.1 評価者間にずれはないか

##### (1) 総合評価の一致率

まず、56件のテスト（テスト①：28件、テスト②：28件）の総合評価について、第1評価者と再評価者の評価の一致率を見る。次の表5から以下のことが明らかである。

表5 総合評価の一致率

テスト	総合評価の一致率 (～件数)	評価者
テスト①	100%(28)	第1評価者はNNT、再評価者はNNT
テスト②	96%(24)	第1評価者はNNT、再評価者はNNTとNT
平均	98%	

全体的に第1評価者と再評価者の評価の一致率が高い。特に、第1評価者も再評価者もNNTであるテスト①では、評価にずれがない。テスト②では、ずれは1件しかない。これはNT評価者とNNT評価者間のものである（NT評価者はNNT評価者より高く評価した）。

以上のことから、第1評価と再評価の作業において、評価者間で評価のずれはほとんど見られず、テストの信頼性は高いと言える。

##### (2) 評価項目別評価の一致率

次に、56件のテストの項目別評価について、第1評価者と再評価者の評価の一致率を見る。結果を示した表6から以下のことが明らかである。

表6 評価項目別評価の一致率

( )内は件数

評価項目	全体の一致率	調査対象者別の一致率	評価者別の一致率	テスト別の一致率
		大学生/高校生	NTとNNT評価者間/ NNT評価者間	テスト①/テスト②
流暢さ	82%(46)	82%(23) / 82%(23)	43%(6) / 98%(41)	96%(27) / 68%(19)
社会言語学的能力	91%(51)	96%(27) / 86%(24)	79%(11) / 95%(40)	93%(26) / 89%(25)
文法	93%(52)	100%(28) / 86%(24)	79%(11) / 98%(41)	100%(28) / 86%(24)
会話運用能力	95%(53)	100%(28) / 89%(25)	79%(11) / 100%(42)	100%(28) / 93%(26)
語彙	96%(54)	96%(27) / 96%(27)	100%(14) / 95%(40)	96%(27) / 96%(27)
発音	98%(55)	100%(28) / 96%(27)	93%(13) / 100%(42)	96%(27) / 100%(28)
談話形成	100%(56)	100%(28) / 100%(28)	100%(14) / 100%(42)	100%(28) / 100%(28)
平均	94%	96% / 90%	81% / 98%	97% / 90%

1) 全体の一致率を見ると、全ての評価項目において一致率が高い。このことから、評価項目別の評価においても、評価者間のずれが少なく、テストはこの点で信頼性が高いと言える。

2) 「流暢さ」においては、NT評価者とNNT評価者間の一致率が低い(43%)。テスト別の一致率では、テスト①は96%、テスト②は68%で、テスト②の方が評価の一致率が低い。

NT評価者とNNT評価者間の一致率が低い要因として、「流暢さ」の評価項目に関する認識が両者では異なることが推測された。

そこで、日本語教師である別のNT2名(第3評価者)に、ずれが生じた典型的なケースに対して改めて評価してもらうことにした。取り上げたのは評価者間で2段階もずれ

があったケース（S27）で、評価者間のディスカッションを経ても、最終的な評価結果が調整できないものであった。第3評価者の評価結果は、NNT評価者の評価結果と一致し、「流暢さ」の評価項目に関する認識は、NNT評価者とNT評価者間においてだけでなく、NT評価者間においても異なることが明らかになった。

NNT評価者とNT評価者間で「流暢さ」の評価項目に関する認識が異なることは、「流暢さ」において、テスト②の一致率がテスト①の一致率より低いことにも影響を与えていると考えられる。なぜなら、テスト①は第1評価も再評価もNNT評価者が行ったのに対し、テスト②では、第1評価はNNTで、再評価はNNT評価者とNT評価者が行ったからである。そこで、このずれが生じた要因を明らかにする必要があるが、これについては、ルーブリックのレベル記述、及び指標の適切さを見る際に、再び検討することにする。

### 5.1.2 テスト①とテスト②の難易度は安定しているか

同じ評価基準に基づいて作成した2つのテストの間で、難易度が安定しているかどうかを見る。そのために、同じ調査対象者に対する同じ評価者によるテスト①とテスト②の評価結果（再評価結果も含める）について、その一致率を検討する。検討対象は、NNT評価者による14人分（高校生7人分と大学生7人分）28件のテスト①、及びテスト②の評価結果である。結果は、次の通りである。

#### (1) テスト①とテスト②の総合評価の一致率

テスト①とテスト②の総合評価は一致率が高く（82%）、テスト①とテスト②の難易度は安定していると言える。この結果は、質問紙調査において、テスト①と②は同じレベルだと回答した調査対象者及び評価者が100%に近いという結果とも一致している（詳細は資料3参照）。

テスト①とテスト②の総合評価でずれのある5件のうち、4件はテスト①よりテスト②の評価が高く、1件はテスト②よりテスト①の評価が高かった。この要因を検討するため、質問紙調査の結果を調べた。まず、テスト①よりテスト②の評価が高い4件について、テスト②はテスト①の後で受けたからよくできたと答えたものが2件（S22とS23）、テスト②の独話のトピックは、テスト①のトピックより話しやすかったと答えたものが1件（S20）あった。残りの1件（S16）は未回答のため、原因が明らかではない。次に、テスト②よりテスト①の評価が高い1件（S14）については、テスト①はテスト②の後で受けたからよくできたと答えていることが分かった。これらのことから、テスト①とテスト②の評価結果にずれが生じたのは、テストのレベルそのものに問題があったからではなく、テスト実施の順番や調査対象者自身の要因が関係したと考えることができる。つまり、テスト①とテスト②の難易度の安定に大きな問題はないと言える。

#### (2) テスト①と②の評価項目別評価の一致率

テスト①とテスト②の評価項目別評価の一致率を検討すると、表7の通りである。

表 7 テスト①と②の評価項目別評価の一致率

( ) 内は件数

評価項目	全体の一致率	調査対象者別の一致率	評価者別の一致率
		大学生 / 高校生	NNT評価者1 / NNT評価者2
流暢さ	57% (16)	43% (6) / 71% (10)	57% (8) / 57% (8)
社会言語学的能力	57% (16)	43% (6) / 71% (10)	57% (8) / 57% (8)
文法	82% (23)	86% (12) / 79% (11)	79% (11) / 86% (12)
会話運用能力	93% (26)	100% (14) / 86% (12)	93% (13) / 93% (13)
語彙	74% (20)	93% (13) / 50% (7)	71% (10) / 71% (10)
発音	100% (28)	100% (14) / 100% (12)	100% (14) / 100% (14)
談話形成	93% (26)	100% (14) / 86% (12)	93% (13) / 93% (13)
平均	79%	81% / 78%	79% / 80%

- 1) 平均を見ると、全体の一致率も、調査対象者別の一致率も、評価者別の一致率も 80% に近く、テスト①と②の難易度は安定していると考えられる。
- 2) 評価項目別に全体の一致率を見ると、「流暢さ」と「社会言語学的能力」以外の評価項目の一致率は高い。
- 3) 調査対象別に見ると、大学生の「流暢さ」の一致率 (43%) と「社会言語学能力」の一致率 (43%)、高校生の「語彙」の一致率 (50%) は低い。以下、具体的に検討する。

まず、大学生の「流暢さ」においてずれのある 8 件のうち、テスト②よりテスト①の評価が高いものが 4 件、テスト①よりテスト②の評価が高いものが 4 件であった。このことから、「流暢さ」が評価項目となる独話タスクのテーマの違い、または、調査対象者個人の経験の違いによって、テスト①あるいはテスト②のテーマの方が流暢に話せたり、話せなかったりしたことが考えられる。しかしながら、このことについては、資料が限定されており、これ以上は分析できない。

次に、大学生の「社会言語学的能力」にずれのある 8 件の全てについては、テスト①よりテスト②の方が評価が高い。そこで、ループリックの指標になっている言語表現と、調査対象者が使用している教科書に見られる言語表現との関連を調べた。「社会言語学的能力」はロールプレイで測られる評価項目である。ロールプレイ① (電話で先生を誘う) では、指標は「もしもし」「～先生」「あのう」「失礼します」「です・ます体を使う」の 5 つである。ロールプレイ② (先生の部屋に行って、本を貸してもらう) でも、指標は「～先生」「あのう」「ありがとうございます/ました」「失礼します」「です・ます体を使う」の 5 つである。ずれのある 8 件を見ると、ロールプレイ①の 5 つの指標のうち、全て「あのう」と「失礼します」ができなかったために、「A2 レベル未満」だと評価されていた。

その原因には次の 2 つが考えられる。まず、大学生が学習している教科書の「電話」場面の会話文には、「あのう」も「失礼します」も出ておらず、また、「失礼します」の使い方について、部屋に出入りする時に使うという説明があるが、電話を切る時に使うという説明はない。一方、教科書の練習帳にはロールプレイ②の場面と近い「テープを

使う許可をもらう」という場面があり、その会話文に「あのう」という声をかける際に使う言語表現が見られる。したがって、大学生調査対象者が「社会言語学的能力」において、テスト①よりテスト②の評価が高い要因は、テスト②の指標の方が教材を通して、より多くのインプットが与えられていたためであると考えられる。

さらに、「語彙」については、高校生の一致率が低い(50%)。「語彙」が評価項目となっている独話タスクについて、そのテーマの違いを見た。ずれのある7件の全てにおいて、テスト①の「家族」のテーマよりテスト②の「町・故郷」の方が評価が高い。その原因は2つ考えられる。まず、調査対象者が使用している教科書のトピックから見ると、「町・故郷」に関するテーマは5つ、「家族」に関するテーマは3つで、「町・故郷」に関するテーマの方が多く扱われている。しかも、「町・故郷」に関するテーマは会話文ではなく、読解文が多いため、語彙の量が多い。したがって、「家族」に関する語彙より、「町・故郷」に関する語彙のほうが多くのインプットが与えられていたと考えられる。また、調査対象者は「町・故郷」に関するテーマを「家族」に関するテーマより最近学習したためにうまく使えたという可能性が、評価者でもある担任の高校教師に対するフォローアップインタビューで分かった。

5.1 をまとめると、評価者間で評価のずれは少なく、テスト①とテスト②自体は、難易度が安定しており、テストの信頼性は高いと言える。しかし、一方で、難易度については教科書や学習時期などの条件が結果に影響を与えていることが分かった。

## 5.2 テストの妥当性

本節では、テストの妥当性について(課題2)検討する。そのために、まず、このテストが調査対象者の日本語能力に合っているかどうかを検討し、次に、評価する際に使用されたルーブリックの各評価項目のレベル記述、及び指標が適切かどうかを検討する。

### 5.2.1 テストは調査対象者の日本語能力に合っているか

テストが調査対象者の日本語能力に合っているかどうかを検討するため、調査対象者28名のテスト①とテスト②の総合評価、調査対象者28名に対して行った質問紙調査、及び評価者2名に対して行った質問紙調査の結果を分析する。

#### (1) テストの総合評価

テスト①とテスト②の総合評価を見ると、表8の通りである。

表8 調査対象者の総合評価

( )内は人数

レベル	A2レベル未満 <sup>a</sup>	A2レベル <sup>a</sup>	A2レベルより上 <sup>a</sup>	レベル未定
テスト				
テスト①	7%(2)	89%(25)	4%(1)	0%
テスト②	11%(3)	68%(19)	18%(5)	3%(1)※
平均	9%	79%	11%	

※ 1件(S24)は評価者間に評価のずれがあり、レベル未定とした(NTはA2レベルより上、NNTはA2レベル)。

調査対象者の学習時間と学習内容から想定される達成レベルを JF スタンドアードの A2 レベルとし、同じレベル及びパターンのテスト①とテスト②を開発したが、表 8 を見ると、2 つのテストにおいて、「A2 レベル」と評価されたものは 80% 近い。このことから、開発されたテストのレベル・内容は、調査対象者の日本語能力に合っていると言える。

テスト②において「A2 レベルより上」と認定された調査対象者数は、テスト①より 4 人多い。この原因については、5.1.2 (1) で述べた通りで、テストの実施順や調査対象者自身の要因によるものと考えられる。

## (2) 調査対象者に対する質問紙調査の結果

「テストは自分のレベルに合っていると思いますか」、「テストの際、自分の会話能力を發揮できたと思いますか」、「テストの内容は話しやすいと思いますか」について、調査対象者に尋ねた結果は、次の図 2、3、4 のようにまとめられる。

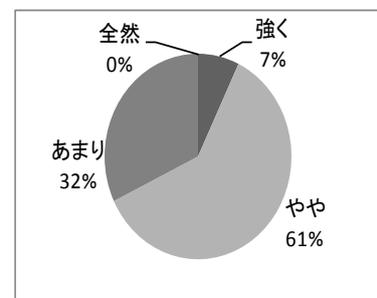
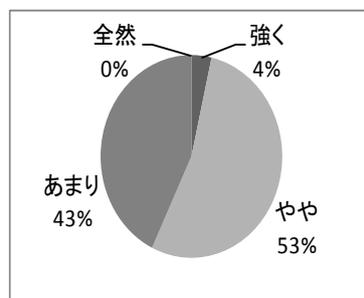
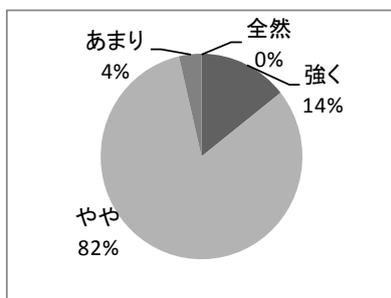


図 2 レベルに合っている

図 3 会話能力を發揮できた

図 4 内容は話しやすい

※ 「強く」 = 「強くそう思います」、「やや」 = 「ややそう思います」、「あまり」 = 「あまりそう思いません」、「全然」 = 「全然そう思いません」

図 2 から、このテストが調査対象者のレベルに合っていたことが分かる。一方、「テストの時、自分の会話能力を發揮できた」と答えたもの (図 3)、及び「テストの内容は話しやすい」と答えたもの (図 4) は、「テストは自分のレベルに合っている」と答えた調査対象者より少ない。その原因について、フォローアップインタビューを行ったところ、多くの調査対象者は、これまでこのようなロールプレイを経験したことがなく、その上、準備時間がないため緊張して力が發揮できなかつたり、話しにくいと感じていたことが分かった。

この結果は、「テストでやったことは授業でやっている活動と似ていると思いますか」という質問に対し、否定的に答えた調査対象者の割合が 50% であることと一致している。つまり、調査対象者は、このテストは自分の日本語能力に合っていると認識しているが、学習経験に照らし合わせると、テストの形式が部分的に新しいものであったことが明らかになった。

## (3) 評価者に対する質問紙調査結果

テストの内容と長さに関して、質問紙調査を行った結果、評価者 2 名もこのテストの内容は調査対象者の日本語能力に合っており、テストの所要時間も適切だと認識していることが明らかである。

## 5.2.2 ルーブリックのレベル記述、及び指標の適切さ

ルーブリックのレベル記述、及び指標が適切だったかどうかについて、質問紙調査の結果とフォローアップインタビューの結果から検討する。そして、前述したNT評価者とNNT評価者間でずれが多かった「流暢さ」の評価指標についても、ここで検討する。

### (1) ルーブリックによる評価の適切さ

ルーブリックによる評価の適切さを検討するために、調査対象者と評価者のそれぞれに「ルーブリックのフィードバックが実際の日本語能力に合っていると思いますか」と尋ねた。結果は次の通りである。

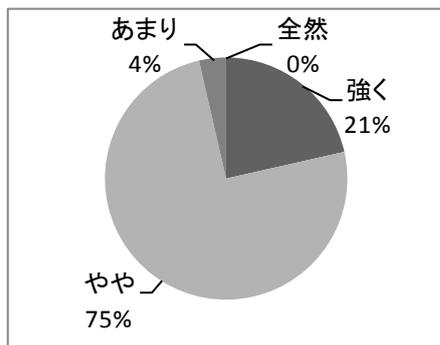


図5 自分の日本語能力に合っている

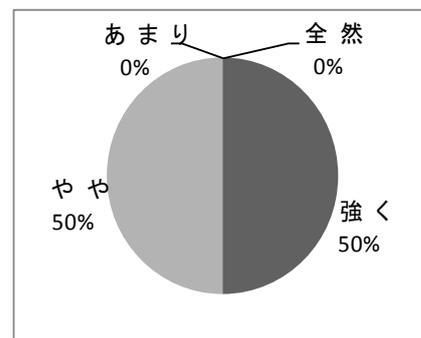


図6 学習者の日本語能力に合っている

図5、6を見ると、調査対象者も評価者もルーブリックのフィードバックが日本語能力を反映しているとする割合が非常に高いことが分かる。このことから、ルーブリックに示されたレベル記述の内容や指標は、調査対象者の実際の日本語能力を正しく評価していると考えられる。

### (2) ルーブリックのレベル記述と指標の適切さ

ルーブリックのレベル記述及び指標の適切を検討するために、評価者にフォローアップインタビューを行った。その結果、ルーブリックの7つの評価項目のうち、問題を指摘されたのは「発音」、「社会言語学的能力」、「会話運用能力」についてであった。

「発音」は3分間のインタビュータスクで測定されるが、指標は「短音と長音の区別」「促音ができる」「ザ、ズ、ゾとジャ、ジュ、ジョ、ヨの区別」「ヤとジャの区別」「ツとチュの区別」「自然なイントネーション」の5つである。指標が細かく定められていたため、テスト実施中に一つ一つ注意して聞くことが難しく、評価しづらかったということが指摘された。

「社会言語学的能力」はロールプレイタスクで測られるが、「A2 レベル未満」と「A2 レベル」において指標は5つで、そのうち4つ以上できたら「A2 レベル」、3つ以下の場合「A2 レベル未満」と判断される。しかし、5つ中3つできている場合に「礼儀正しい言葉遣いがほとんどできない」とするのはレベル記述に合わないという指摘があった。また、「あのう」という表現が、テスト①とテスト②の両方で指標になっているが、必ずしも実際のロールプレイでは使用する必要がなかったため、指標として適切でないのではないかという意見もあ

った。

また、「会話運用能力」についても、ロールプレイタスクの評価項目になっているが、レベルの高い学生の会話能力が正しく測れないという意見があった。これは、「A2 レベルより上」の「上手な聞き返し」という指標に関する問題である。つまり、相手が話したことを全部理解した場合、学生は聞き返しを行わないため、指標にチェックを入れることができない。以下に「A2 レベルより上」と評価された例を示す。

テスト①(前略)S18: あ、8時からです。 T: 8時からですね。え、その日には <u>シンポジウム</u> が入っているんですが。 S18: へえー T: でも、何ですか。(後略)
テスト②(前略)S22: 来週試験がありますが、その本にたくさん必要な知識が書いてありますから… T: そうですね。ええと、今から <u>シンポジウム</u> があるんですけど。 S22: ええと、じゃ、明日、私は戻ってもいいですか。(後略)

上掲の部分で T (第 1 評価者) が「シンポジウム」という言葉を使っているのは、調査対象者が知らない言葉であると想定し、「上手な聞き返し」の能力を試したかったためである。しかし、S18 の場合も S22 の場合も、T の発話内容が全部理解できたため、聞き返しが見られなかった。つまり、この指標は、指標として機能しなかったということになる。今後、指摘された上述のことを参考にし、改善を行う必要があるだろう。

### (3) 「流暢さ」の評価指標について

ここでは、5.1.1 で取り上げたように、「独話」のタスクでの NT 評価者と NNT 評価者間の評価の一致率の低かった「流暢さ」に関して、改めて分析する。NNT 評価者と NT 評価者の「流暢さ」に関する評価には 8 件のずれがあった。そのうち、7 件は NT 評価者が NNT 評価者より低く評価し、残りの 1 件 (S24) は NT 評価者が NNT より高く評価した。その原因を探るため、ここでは「流暢さ」の指標へのチェックの入れ方について、具体例から検討する。

まず、NT 評価者が再評価した 14 件の評価結果を全て見てみる。14 件のうち 10 件は「A2 レベル未満」、2 件は「A2 レベル」、2 件は「A2 レベルより上」と判定され、「A2 レベル未満」のものがほとんどである。指標は「言葉に詰まる」「黙る時間がある」「発話がとぎれとぎれになる」「同じ言葉を繰り返す」「言い直しをする」の 5 つである。「A2 レベル未満」と判定された 10 件を見ると、8 件は 5 つの指標のうち 3 つの指標「言葉に詰まる」、「黙る時間がある」、「発話がとぎれとぎれになる」の全てにチェックが入れていることが分かった。この結果から、NT 評価者は、3 つの指標のうち 1 つにチェックを入れた場合、残りの 2 つの指標にもチェックを入れる傾向があり、そのために「A2 レベル未満」という低い判定が多くなったと考えられる。そして、それは「言葉に詰まる」「黙る時間がある」「発話がとぎれとぎれになる」という 3 つの指標の意味が重複しているためと考えられる。

次に、以下にあげる NT 評価者が NNT より高く評価した「流暢さ」の評価結果を検討する。

### 第1 評価者の NNT による評価

- 「言葉に詰まる」
- 「黙る時間がある」
- 「発話がとぎれとぎれになる」
- 「同じ言葉を繰り返す」
- 「言い直しをする」

### 再評価の NT による評価

- 「言葉に詰まる」
- 「黙る時間がある」
- 「発話がとぎれとぎれになる」
- 「同じ言葉を繰り返す」
- 「言い直しをする」

第1 評価者の NNT は2 つの指標にチェックを入れたのに対し、再評価者の NT は1 つもチェックを入れなかった。NNT がチェックを入れたのは、発話中、2 回ぐらい言葉に詰まったり、2 分程度の時間があるにも関わらず、1 分 10 秒で話終わってしまったりしたからということだった。このことから、どの程度なら「言葉に詰まる」とするか、また「黙る時間がある」とする場合、発話時のみを対象とするか、発話が終わったあとも対象とするか、評価者間で指標の解釈が異なっていたと考えられる。

5.2 をまとめると、今回の会話テストは、調査対象者の日本語能力に合っており、使用されたルーブリックの各評価項目のレベル記述及び指標は概ね適切である。つまり、今回のテストは妥当性が高いものであると言える。しかし、指標の定め方については、「発音」「社会言語学的能力」「会話運用能力」「流暢さ」において多少改善すべきところがあることが明らかになった。

### 5.3 テストの波及効果

本節では、このテストが学習者と教師にどのような波及効果をもたらすか（課題 3）を検討する。そのために、ルーブリックのフィードバックについての調査対象者と評価者に対する質問紙調査の結果、及びテスト全体の内容についてのフォローアップインタビューの結果を分析する。

ルーブリックのフィードバックについて尋ねたところ、調査対象者全員が「今後もこのようなルーブリックのフィードバックをもらいたい」と回答した。図 7、8 は、フィードバックについての調査対象者の意見をまとめたものである。

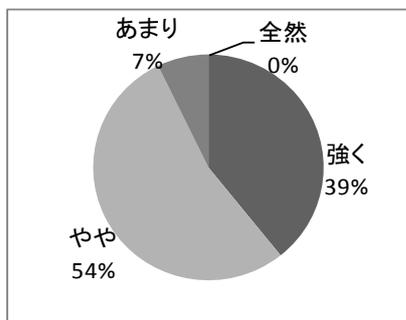


図 7 自分の弱点に気付くようになった

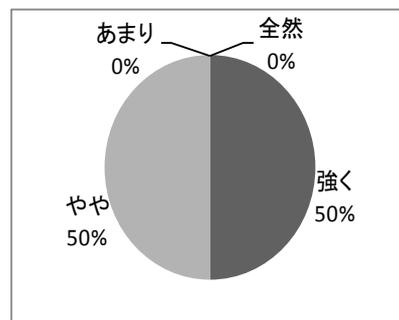


図 8 今後の日本語学習に役に立つ

図 7、8 を見ると、調査対象者は、フィードバックによりこれまで気付かない自分の弱点に気付くようになり、フィードバックが今後の日本語学習に役に立つと考えていることが明らか

かである。また、評価者に「ルーブリックのフィードバックは今後の授業に役に立つと思いますか」、「ルーブリックのフィードバックは学生の日本語の勉強に役に立つと思いますか」と質問したところ、肯定的な回答を得た。さらに、フォローアップインタビューでは、今後このようなテストを受けたい、実施したいという意見が多く見られた。そこで、以下で、学習者と教師にとって、具体的にどんなことに気付くようになったのか、どのように役に立つのかをより詳細に分析することによって、テストの波及効果を明らかにしたい。

### (1) テストが学習者に与える影響

調査対象者 8 人に対してフォローアップインタビューを行った結果、学習者に与える影響は以下の 2 つにまとめられる。

#### 1) 学習活動の促進

テストの内容、及び評価項目の影響について、以下のような意見が多く出された。

- |  |
|--|
| (S1) 「テストを受けて、自分のできた点とできなかった点に分かります。できたことが分かったら安心しますが、できなかったらもっと勉強しなければなりません。」                                   |
| (S4) 「各レベルの記述内容は詳しいです。そして、指標にはどんな場面にどんな表現を使うか分かりました。「いかがですか」と「伺います」という 2 つの表現をよく間違えていますが、テストの後で、区別できました。」        |
| (S5) 「3段階に分けることによって、今回足りない点に分かって、何を練習したらいいのか具体的な方向が分かるようになりました。そして、上のレベルに達成できるようにどうすればいいかという目標が分かったので、自分で頑張ります。」 |
| (S13) 「各評価項目に具体的な指標があるので、自分でも勉強できます。そして、例えば言葉が詰まると繰り返しが多いと評価されたから、もっと家で言葉を勉強したり、たくさん会話を練習したりしなければなりません。」         |

※ フォローアップインタビューはベトナム語で行った。枠内の内容は筆者が日本語に訳したものである

上に引用した意見から、調査対象者が、自分のできたこと、できなかったことについて詳しく知ることができ、テストが今後の自律学習のきっかけになっていることが分かる。

#### 2) 学習活動の方向付け

フォローアップインタビューでは、次のような意見が多く出された。

- |   |
|---|
| (S10) 「テストを受けて、補強しなければならない欠点に分かるようになりました。家へ帰って、習った文型、トピック、場面をもう一度練習しなければなりませんと思います。」                                  |
| (S3) 「今まで殆ど文型と書くことに注目して、あまり話すことに注意しませんでした。しかし、通訳を専攻している私たちにとって、話すことと聞くことは大切です。」                                       |
| (S4) 「語彙項目はまだできていないので、今まで学習の仕方を変えて、もっと語彙の量を増やしたいです。そして、ロールプレイでは、内容だけ注意して、「社会言語学的能力」の表現の挨拶を忘れましたので、これからも注意しなければなりません。」 |
| (S5) 「独話のタスクでは、筋のない文をたくさん話して、全然談話形成を考えませんでしたので、これから注意して練習します。」  |
| (S7) 「具体的な評価項目があって、そして、たくさんの指標が書いてあるので、テストを受けて、自分の学習方法の方向を定めることができ、何をどうふうに勉強すればいいか、ヒントになると思います。」                      |

上に引用した意見から、これまでは主に文型を重視して学ぶだけであったが、テストでロールプレイや独話のタスクを経験し、また、文型以外の観点からも評価を得ることによって、文型を正しく操作するだけでは十分ではないことに気付き、今後の学習の改善点が具体的に把握できたことが分かる。

以上の分析 1) 2) から、調査対象者にとって、今回の会話テストには、日本語学習を促進する、口頭表現能力の重要性を認識する、何を重点的に学ぶべきかについて、学習の方向付

けを得るといった波及効果があったことが分かる。

## (2) テストが教師に与える影響

評価者2人に対してフォローアップインタビューを行った結果、今後の授業に役に立つという以下の意見が出された。

「このテストは内容が詳しいので、自分の学習者の日本語能力が分かって、よく間違えたところが何かも分かります。そして、学習者の間違えたところのリストも作ることができて、学習者の間違いを直すことができます。そして、これを見て、自分の授業を振り返って、付け加えなければならない点、より重視すべき点も分かるようになると思います。」  
「会話というのはやっぱり目で見えないものですから、ルーブリックの中に、例えば問題点、できた点どちらも書いてあると、目で見ても教師も学習者も確認できるので、その点は、授業でこれから何を取り入れようとか、ほかにもやるべきことがあるんじゃないかとか。いろいろ参考になると思います。」

上に引用した意見から、教師にとって、テストの実施は、学習者の口頭表現能力を詳しく把握し、今後の授業を設計し、自分の授業を振り返る良い機会になったことが分かる。つまり、教師も今回のテストを通じて、学習者の能力や問題点を具体的に知ることができ、自分の教授活動の適切さを判断し、今後の教授活動を改善することができると考えられる。

5.3 をまとめると、今回のテストは、学習者にとっては、今後の日本語学習活動の促進、及び学習の方向付けに対しての肯定的な波及効果がある。また、教師にとっても、今後の教授活動についての内省を促進するという効果があると考えられる。

## 6. テストの改善

本章では、5.1 のテストの信頼性、5.2 の妥当性の検証で明らかになった問題点とテストの改善点について調査対象者及び評価者に対するフォローアップインタビューの結果を参考にしながら、今後のテストの改善について検討する。

### 6.1 タスクの内容

5.1.2 で述べたように、教科書や学習時期などの条件が調査対象者の評価結果に影響を与えた点を考慮して、今後多様な対象者に対して公平で信頼性の高いテストにするために、タスクの内容を決める際、留意が必要である。

### 6.2 ルーブリックのレベル記述及び指標について

指標については、評価者からいろいろな指摘があったが、指標があったために短期間で評価が可能となり、学生及び教師に対しても肯定的な波及効果があった。そこで、指標を取りやめるのではなく、変更することで改善案を提示したい。

#### 1) 「発音」の評価指標について

NNT 評価者からは、「発音」はインタビューのタスクにおいてだけ測定するのではなく、3つのタスク全体で測った方が良いという意見があった。これは、タスク全体において、学生の発音の誤りをメモし、一人一人の発音の誤りのリストを作ることができれば良いということである。一方、NT 評価者からは、「ザ、ズ、ゾとジャ、ジュ、ジョ、ヨの区別」「ヤとジャの区別」などというような具体的な音の区別ではなく、「長音ができる」「短音ができる」「促

音ができる」というより総括的な指標にした方が良い、そして「自然なイントネーション」の代わりに「理解できる程度にはっきりしているイントネーション」という指標にしてはどうかという提案があり、以下の表9の原案に対して表10のような改善案を得た。

表9 元の「発音」の指標

<input type="checkbox"/> 長音と短音の区別
<input type="checkbox"/> 促音ができる
<input type="checkbox"/> 「ざ、ず、ぞ」と「じゃ、じゅ、じょ、よ」の区別
<input type="checkbox"/> 「や」と「じゃ」の区別
<input type="checkbox"/> 「つ」と「ちゅ」の区別
<input type="checkbox"/> 自然なイントネーションできる

表10 NTの「発音」の指標の提案

<input type="checkbox"/> 長音
<input type="checkbox"/> 短音
<input type="checkbox"/> 促音
<input type="checkbox"/> 理解できる程度にはっきりしているイントネーション

しかし、「発音」は3つのタスク全体で測るというNNT評価者からの提案は、評価者の負担が重く、実施するのは困難である。そして、表10のNTの提案には、ベトナム人に特有の発音のくせに関する指標がない。テストを実施した際、「よう」を「じょう」、「や」を「ざ、じゃ」などと間違えて発音した調査対象者がいたため、発音のくせに関する指標は必要であると考えられる。そこで、以下の改善案が考えられる。

「短音、長音、促音」などの具体的な拍を区別する元の2つの指標を「拍の区別」という1つの指標に変える。発音のくせに関する元の3つの指標を「発音のくせ」という1つの指標にまとめる。さらに、「自然なイントネーション」の代わりに「理解できる程度にはっきりしているイントネーション」という指標に変更する。具体的には、表11の通りである。

表11 「発音」の評価指標の改善

	まだ努力が必要 (A2レベル未満)	目標を達成 (A2レベル)	目標以上を達成 (A2レベルより上)
発音 (JF: 音素の把握)	<input type="checkbox"/> 発音をはっきり聞き取れない、本人に繰り返しを求めても、分かりづらい。 ・拍の区別の間違いが多い。(短音、長音、促音など) ・発音のくせが多い(しゃ、ざ、や、じょ、ぞなど) ・イントネーションがはっきりせず、理解しづらい。	<input type="checkbox"/> ベトナム人に特有の発音が見られるが、理解できる程度にははっきり発音している。 ・拍の区別の間違いが少ない。(短音、長音、促音など) ・発音のくせがあるが、少ない(しゃ、ざ、や、じょ、ぞなど) ・イントネーションがはっきりしていて、理解できる。	<input type="checkbox"/> 発音を間違えることが少なく、自分自身で言い直すことができる。聞き取りやすい。 ・はっきり拍の区別ができる。(短音、長音、促音など) ・発音のくせがほとんどない。(しゃ、ざ、や、じょ、ぞなど) ・イントネーションがはっきりしていて、大変分かりやすい。

## 2) 「流暢さ」の評価指標について

5.2.2 (3) で検討した「言葉に詰まる」「黙る時間がある」「発話がとぎれとぎれになる」という意味の重複がある3つの指標を1つの指標にする。そして、各レベルについて頻度を表す言葉も付け加え、より正確な測定ができるようにする。具体的には、表12の通りである。

表12 「流暢さ」の評価指標の改善

	まだ努力が必要 (A2レベル未満)	目標を達成 (A2レベル)	目標以上を達成 (A2レベルより上)
流暢さ (JF: 話しことばの流暢さ)	<input type="checkbox"/> 言いたいことを表現するのが困難である。 ・かなり黙る時間がある ・かなり同じ言葉を繰り返す ・かなり言い直しをする	<input type="checkbox"/> あまり困難なく、言いたいことが表現できる。 ・時々黙る時間がある ・時々同じ言葉を繰り返す ・時々言い直しをする	<input type="checkbox"/> 困難なくスムーズに言いたいことが表現できる。 ・ほとんど黙る時間がない ・ほとんど同じ言葉を繰り返さない ・ほとんど言い直しをしない

また、早く話終わってしまうケースに対応するため、時間の管理に対し、どのように対処したらいいのかについて、評価者間で事前に決めておく。

### 3) 「会話運用能力」、及び「社会言語学的能力」の評価指標について

5.2.2 (2) で取り上げたように、「会話運用能力」の「A2 レベルより上」の「上手な聞き返し」の指標は、よくできる学生にとって適切ではなかった。そこで、「聞き返し」の指標を「上手に会話が維持できる」に変え、その具体例として「上手な聞き返し」、「自ら会話を展開し、進める」などを付け加える。具体的には、表 13 の通りである。

目標以上を達成 (A2レベルより上)
会話を始め、続け、また終えることがスムーズにできる。 <b>指標：A2レベルのことに加えて、以下のよう なことが1つ以上できる</b> <input type="checkbox"/> 補足的な情報交換(パーティーの雰囲気、先生が参加することに対する皆の気持ち、行き方)など <input type="checkbox"/> 上手に会話が維持できる(上手な聞き返し、自ら会話を展開し、進める)など <input type="checkbox"/> 多様なあいづち「あのう、ええと、へえ、そうですね、そうですか」

次に、「社会言語学的能力」において、「A2 レベル未満」のレベル記述の「……がほとんどできない」を「あまりできない」と変える。そして、指標となる表現の「あのう」については、ロールプレイ①の場合は「ええと」「お願いがあるんですが」、ロールプレイ②の場合は「すみません」「ごめんください」など、初級レベルで使えるといい表現の例をいくつか加える。具体的には、表 14 と表 15 の通りである。

表 14 テスト①の「社会言語学的能力」の改善

	まだ努力が必要 (A2レベル未満)
社会言語学的能力 (JF: 社会言語的な適切さ)	日常的な挨拶や呼びかけなど、礼儀正しい言葉遣いがあまりできない。 <b>指標：1つ～3つできる</b> <input type="checkbox"/> 「もしもし」と言う <input type="checkbox"/> 「～先生」と呼びかける <input type="checkbox"/> 「あのう～」/「お願いがあるんですが」などと言う <input type="checkbox"/> 「失礼します/ました」と言う <input type="checkbox"/> 「です・ます」体を使う

表 15 テスト②の「社会言語学的能力」の改善

	まだ努力が必要 (A2レベル未満)
社会言語学的能力 (JF: 社会言語的な適切さ)	日常的な挨拶や呼びかけ言葉遣いがあまりできない。 <b>指標：1つ～3つできる</b> <input type="checkbox"/> 「先生」と呼びかける <input type="checkbox"/> 「あのう～」/「すみませ <input type="checkbox"/> 「ありがとうございます/ <input type="checkbox"/> 「失礼します/ました」と <input type="checkbox"/> 「です・ます」体を使う

## 6.3 テストの実施について

テストの実施方法について調査対象者及び評価者に尋ねたところ、ロールプレイ①(パーティーに先生を招待する)のタスクは、先生と向かい合って話すのではなく、本当に電話で話しているような設定にすれば、緊張しないでもっとうまくできるのではないかと、また、ロールプレイ②(先生に本を貸してもら)のタスクでは、座ったままで話すのではなく、実際と同じようにした方が良いという調査対象者の意見があった。ロールプレイのタスクでの場面設定については、実際の場面の雰囲気が出せれば良いという意見が評価者からもあった。そこで、調査対象者の余計な緊張感を減らし、また実際の場面に近づけるため、小道具を用意するなど、ロールプレイのタスクのリアリティを出す工夫が必要である。

## 6.4 テストのフィードバックについて

評価者2名の意見では、より効果的なテストのフィードバックとしては、できればテスト後、録音したテープを学生に渡し、聞かせるか、あるいは、学生と一緒に聞きながら確認したほうがよいという提案があった。しかし、機関の設備の問題もあり、この点はまだ実現す

るのが難しく、今後の課題の1つであると考えられる。

## 7. まとめと今後の課題

本研究では、JF スタンダードに基づき、10分程度の会話テストを開発し、その有効性を検証した。設定した3つの研究課題に対しての結果は、以下の通りである。

研究課題1の結果として、今回開発した会話テストは信頼性が高いことが分かった。ただし、難易度については教科書や学習時期などの条件により、評価結果にずれが生じることがあったため、今後のテスト作成の際は、これらの点に留意すべきである。

さらに、評価者からの意見として、今回のテストはJF スタンダードに基づいて開発されたため、広く使用できる可能性があるということ、また、テストで用いたルーブリックには評価項目ごとに評価の観点が詳しく記述されているため、評価しやすく、学習者の口頭表現能力が正しく測定でき、評価の信頼性が高いということがあげられた。

研究課題2の結果として、作成したルーブリックには多少改善点があるものの、今回開発した会話テストは調査対象者の日本語能力を適切かつ正しく測定でき、妥当性が高いものであるということが分かった。

研究課題3の結果として、今回の会話テストは、学習者にとって、学習活動の促進、学習の方向付けという肯定的な波及効果を与えるものであり、教師にとっても、教授活動についての内省の促進という肯定的な波及効果を及ぼすものであることが分かった。

今後の課題としては、第6章でテストの改善で述べた通り、実際にその改善点をもとにテスト結果を再検討し、よりよい会話テストを開発し、実施していきたいと考えている。

また、本研究で開発したのは、JF スタンダードのA2 レベルの会話テストだけである。今後は、A2 レベルだけではなく他のレベル、また、大学1年生だけではなく各学年の口頭表現能力が測れる会話テストを開発したい。さらに、本研究で得られた結果から、現在の学習者の日本語のカリキュラムを見直し、JF スタンダードを参照して、具体的な到達目標を設定することを検討したい。

## 謝辞

本研究の調査にご協力くださった国際交流基金ベトナム日本文化交流センターの横山直子先生、JICA 元派遣専門家の津村知美先生、ハノイ国家大学外国語大学の先生方や学生の皆さんに深くお礼を申し上げます。さらに、温かく励ましてくれた家族にも感謝します。

## 注

<sup>1</sup> ベトナムでは中学校は4年（6年生から9年生）、高等学校は3年（10年生から12年生）である。

<sup>2</sup> 2009年10月には、日本語教育を実施する中等教育機関数は22校、教師数は21名、学習

者数は 3,301 名である（国際交流基金ベトナム日本文化センターによる）。

- <sup>3</sup> 開始時は 33 名であったが、2009 年 10 月には 122 名となっている。
- <sup>4</sup> 「教育上の機能」は「教育目標の設定」「教授活動の促進」「学習活動の促進」「学習者の自己形成の一助となる」「学習の方向付け」「クラス活動の評価」に、「研究上の機能」は「学習効果の測定」「学習者の持つ諸条件の測定」「学習上の困難点の予測」「よりよいテストの開発」に、「教育管理上の機能」は「コースレベルの設定」「コースレベルの維持」「教育環境の維持」にそれぞれ下位分類されている。
- <sup>5</sup> 2009 年に国際交流基金は JF スタンドアの試行版を出版し、2010 年に『JF 日本語スタンダード 2010』及び『JF 日本語教育スタンダード 2010 利用者ガイドブック』を出版した。
- <sup>6</sup> Common European Framework of Reference for Languages: Learning, teaching, assessment の略で、ヨーロッパの言語教育・学習の場で共有される枠組みである。
- <sup>7</sup> 「コミュニケーション言語活動」は「産出と方略」、「やりとりと方略」、「受容と方略」に分けられ、それらがさらに「経験や物語を語る」、「指示やアナウンスを聞く」「情報交換する」などのカテゴリーに下位分類される。そして、「コミュニケーション言語能力」は「言語構造的な能力」、「社会言語能力」、「語用能力」に分けられ、それらがさらに「文法的正確さ」、「社会言語的な適切さ」、「発言権」などのカテゴリーに下位分類される。

## 参考文献

- (1) 石田敏子 (1992) 『入門日本語テスト法』大修館書店
- (2) 国際交流基金 (2009) 『JF 日本語教育スタンダード試行版』
- (3) 国際交流基金 (2010a) 『JF 日本語教育スタンダード 2010』
- (4) 国際交流基金 (2010b) 『JF 日本語教育スタンダード 2010 利用者ガイドブック』
- (5) 庄司恵雄 (1996) 「日本語研修コースのための口頭能力修了試験」『日本語教育』91 号、日本語学会、108-119.
- (6) 高浦勝義 (2004) 「評価の観点とルーブリックの作成」『絶対評価とルーブリックの理論と実践』黎明書房、91-148.
- (7) 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』1 号、国際交流基金、15-32.

## 資料1-1 テスト①の内容

### インタビュー（3分）の流れ図

1. おはようございます。どうぞ、座ってください。  
↓
2. では、始めましょう。まず、お名前を言ってください  
↓
3. Sさんは今までどのぐらい日本語を勉強しましたか。  
↓  
[~か月/時間V]
4. 日本語はどうですか。  
↓  
[~AですがAです/Aです]
5. どうして、~ですか。  
↓  
[~から~/~ですから]
6. 日本語能力試験を受けたことがありますか。  
↓
7. ところで、(話がかわりますが、) Sさんは今朝何時に起きましたか。  
S: 6時に起きました  
S: 6時です  
↓  
[「Vます」と「Vました」の区別]
8. 朝ごはんは何を食べましたか  
↓  
[「Vます」と「Vました」の区別]
9. 毎日学校までどうやって/何で来ますか。  
↓  
[~手段でV/歩いて来ます]
10. 今日、家に帰って、何をしますか。  
↓  
[~Vて、Vて~/V~]
11. 日曜日はよく何をしますか。  
↓  
[~VたりVたり~]
12. 最後に、1つの質問があります(けど)、Sさんは高校/大学を卒業してから、何をしたいですか。  
↓  
[~Vたいです]
13. そうですか。頑張ってください。

注意：1. テスターは適当で自然な相槌をうつこと。

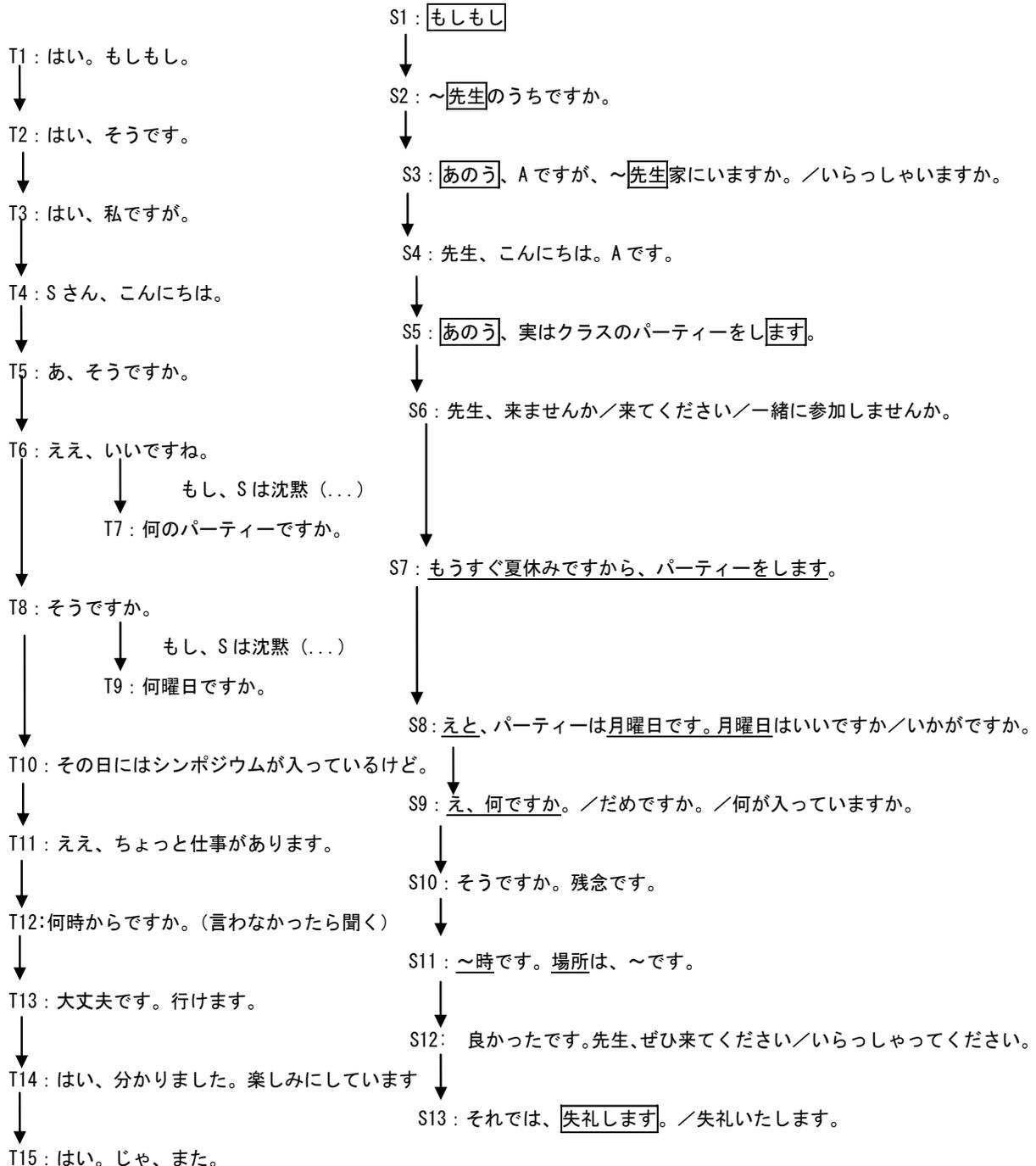
2. 「基本的な語順(SOV)」の指標は、会話全体で判断すること。

### 独話（2分）

- ・タスク：自分の家族について、詳しく教えてください。
- ・内容：家族のメンバー、メンバーの仕事・性格・外見・趣味など（自分と比べて）。家族の楽しい出来事。

## ロールプレイ（3分）の流れ図

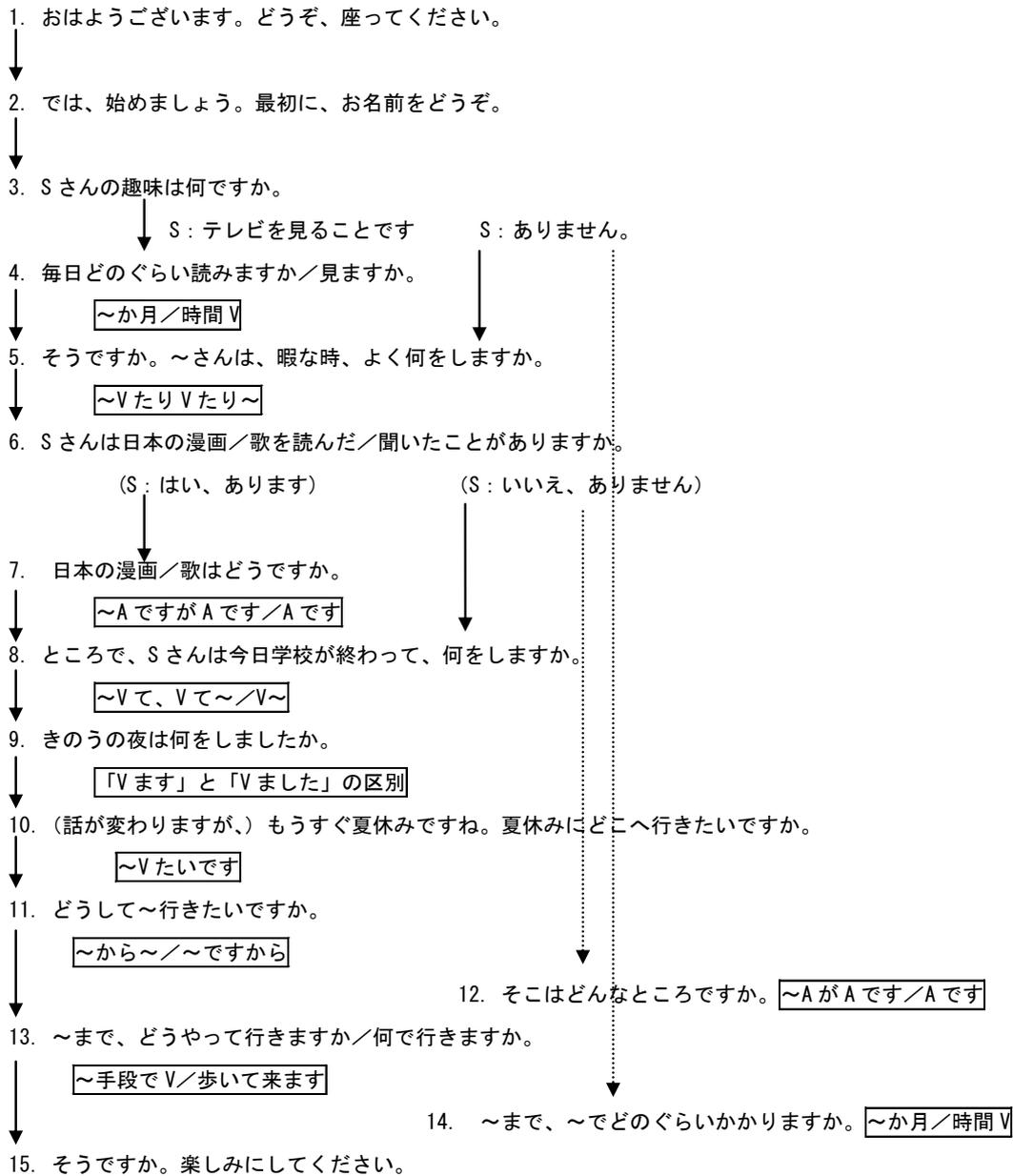
- ・状況：あなたは日本人の先生をクラスのパーティーにさそいたいので、先生の家にかけます。
- ・タスク：先生が家にいるかどうか、確認してから先生を誘ってください。パーティーの（時間、場所、目的）について伝えてください。



- 注意：1. 質問はSが場所・時間・目的を自分から言わなかったらする。なるべく待つこと。  
2. 電話するので必ずお互いの顔を見ないこと。

## 資料1-2 テスト②の内容

### インタビュー（3分）の流れ図



注意：1. テスターは適当で自然な相槌をうつこと。

2. 「基本的な語順 (SOV)」の指標は、会話全体で判断すること。

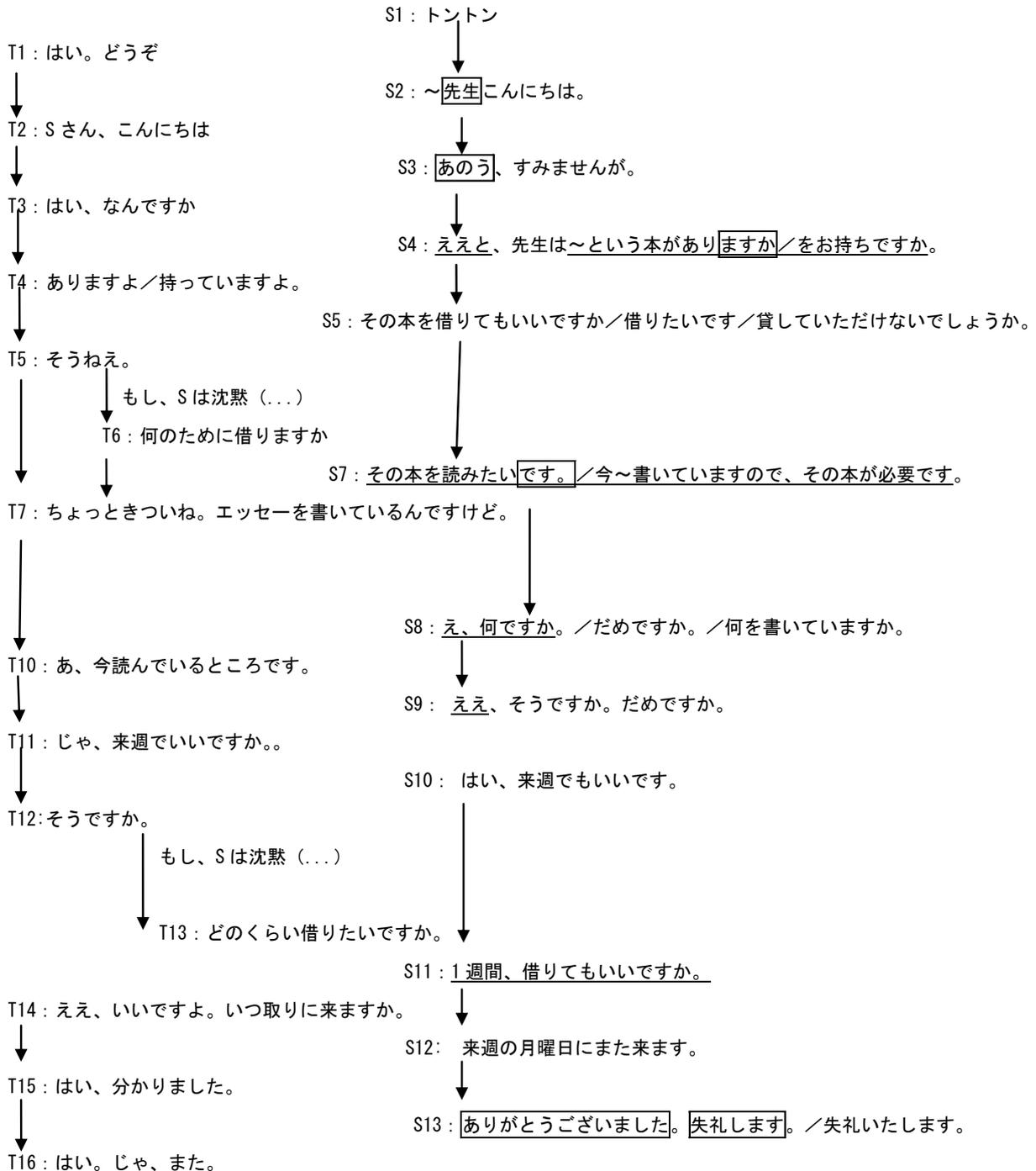
### 独話（2分）

・タスク：住んでいる町／ふるさとについて、詳しく話してください。

・内容：町の風景、何があるか、以前との比較、町で一番好きな所やそこで何をしますか。

## ロールプレイ（3分）の流れ図

- ・状況：あなたは日本人の先生の部屋へ本を借りに行きます。
- ・タスク：ドアをノックして、借りたい本があるかどうか確認してから、借りてください。本のタイトル、借りたい理由、借りる期間について伝えてください。



注意：質問はSが本のタイトル・借りたい理由・借りる期間を自分から言わなかったらする。なるべく待つこと。

資料 2-1 テスト①のルーブリック

評価項目\レベル	まだ努力が必要 (A2レベル未満)	目標を達成 (A2レベル)	目標以上を達成 (A2レベルより上)
発音 (JF: 音素の把握)	<p>・発音がはっきり聞き取れない、当人に繰り返しを求めても、分かりづらい。 <b>1つ～2つの指標ができる</b></p>	<p>・ベトナム人に特有の発音のくせが見られるが、理解できる程度にははっきり発音している。 <b>3つ～4つの指標ができる</b></p>	<p>・発音を間違えることが少なく、自分自身で言い直すことができる。聞き取りやすい。 <b>5つ～6つの指標ができる</b></p>
文法 (JF: 文法的正確さ)	<p>・単純な文法構造があまり正しく使えない。本人が何を話しているのかわかりづらい。 <b>1つ～3つの指標ができる</b></p>	<p>・単純な文法構造がほとんど正しく使える。本人が言いたいことがだいたい分かる。 <b>4つ～7つの指標ができる</b></p>	<p>・単純な文法構造が正しく使える。本人が言いたいことがよく分かる。 <b>8つ～9つの指標ができる</b></p>
談話形成 (JF: 話題の展開・一貫性と結束性)	<p>・文レベルの発話が維持できず、時には単語レベルの発話になる。話の内容がはっきり理解できない。 ・接続表現が少ない。</p>	<p>・簡単に短い文をつなげ、ポイントを並べ上げる形で、物事を語るができる。 ・よく使われている接続表現が使える。(でも、そして、それから)</p>	<p>・簡単なまとまりのある話ができ、分かりやすい。 ・文レベルを維持し、必要に応じて、適切な接続表現が使える。</p>
語彙 (JF: 使用語彙領域・語彙の使いこなし)	<p>・家族に関する必要な語彙がない。 <b>A2レベルの以下のような指標があまり言えない</b></p>	<p>・思い出すのに少し時間がかかるが、家族に関する必要な語彙がある。 <b>A2レベルの以下のような指標が殆ど言える</b></p>	<p>・家族に関する基本的な語彙がだいたいある。基本以上の語彙もある程度使える。 <b>A2レベルの単語に加えて、以下のような指標が2つ以上言える</b></p>
流暢さ (JF: 話しことばの流暢さ)	<p>・言いたいことを表現するのが困難である。 <b>指標が3つ以上ある</b></p>	<p>・あまり困難なく、言いたいことが表現できる。 <b>指標が2つある</b></p>	<p>・困難なくスムーズに言いたいことが表現できる。 <b>指標が1つ以下ある</b></p>
会話運用能力 (CEFR: 発言権、柔軟性)	<p>・簡単な方法を使って、短い会話を始め、続け、また終えることができない。 <b>下線の部分ができない</b></p>	<p>・簡単な方法を使って、短い会話を始め、続け、また終えることができる。 <b>下線の部分をふくんで、2つ以上の指標ができる。</b></p>	<p>・会話を始め、続け、また終えることがスムーズにできる。 <b>A2レベルのことに加えて、以下のような指標が1つ以上できる。</b></p>
社会言語学的能力 (JF: 社会言語的な適切さ)	<p>・日常的な挨拶や呼びかけなど、礼儀正しい言葉遣いが殆どできない。 <b>1つ～3つの指標ができる</b></p>	<p>・日常的な挨拶や呼びかけなど、礼儀正しい言葉遣いが殆どできる。 <b>4つ～5つの指標ができる</b></p>	<p>・尊敬語的、謙讓語的な表現、丁寧な表現などがある程度使える。 <b>A2レベルのことに加えて、以下のような指標が1つ以上できる。</b></p>

資料 2-2 テスト②のルーブリック

評価項目\レベル	まだ努力が必要 (A2レベル未満)	目標を達成 (A2レベル)	目標以上を達成 (A2レベルより上)
発音 (JF: 音素の把握)	<p>・発音がはっきり聞き取れない、本人に繰り返しを求めても、分かりづらい。 <b>1つ～2つの指標ができる</b></p>	<p>・ベトナム人に特有の発音のくせが見られるが、理解できる程度にははっきり発音している。 <b>3つ～4つの指標ができる</b></p> <p>(指標)  <input type="checkbox"/> 長音と短音の区別ができる  <input type="checkbox"/> 促音ができる  <input type="checkbox"/> 「ざ、ず、ぞ」と「じゃ、じゅ、じょ、よ」の区別ができる  <input type="checkbox"/> 「や」と「じゃ」の区別ができる  <input type="checkbox"/> 「つ」と「ちゅ」の区別ができる  <input type="checkbox"/> 自然なイントネーションができる</p>	<p>・発音を間違えることが少なく、自分自身で言い直すことができる。聞き取りやすい。 <b>5つ～6つの指標ができる</b></p>
文法 (JF: 文法的正確さ)	<p>・単純な文法構造があまり正しく使えない。本人が何を話しているのかが分かりづらい。 <b>1つ～3つの指標ができる</b></p>	<p>・単純な文法構造がほとんど正しく使える。本人が言いたいことがだいたい分かる。 <b>4つ～7つの指標ができる</b></p> <p>(指標)  <input type="checkbox"/> 基本的な語順(SOV)  <input type="checkbox"/> 「～か月/時間V」  <input type="checkbox"/> 「～VたりVたり～」  <input type="checkbox"/> 「～AですがAです/Aです」  <input type="checkbox"/> 「～Vて、Vて～/V～」  <input type="checkbox"/> 「Vます」と「Vました」の区別  <input type="checkbox"/> 「～Vたいです」  <input type="checkbox"/> 「～から～/～ですから」  <input type="checkbox"/> 「～手段でV/歩いて来ます」</p>	<p>・単純な文法構造が正しく使える。本人が言いたいことがよく分かる。 <b>8つ～9つの指標ができる</b></p>
談話形成 (JF: 話題の展開・一貫性と結束性)	<p>・文レベルの発話が維持できず、時には単語レベルの発話になる。話の内容がはっきり理解できない。 ・接続表現が少ない。</p>	<p>・簡単に短い文をつなげ、ポイントを並べ上げる形で、物事を語るができる。 ・よく使われている接続表現が使える。(でも、そして、それから)</p>	<p>・簡単なまとまりのある話ができ、分かりやすい。 ・文レベルを維持し、必要に応じて、適切な接続表現が使える。</p>
語彙 (JF: 使用語彙領域・語彙の使いこなし)	<p>・町に関する必要な語彙がない。 <b>A2レベルの以下のような指標があまり言えない</b></p> <p>(指標)  <input type="checkbox"/> 人称(父、母、姉、妹、兄、弟、兄弟)  <input type="checkbox"/> 職業(会社員、銀行員、教師、エンジニア、医者、高校生、学生)  <input type="checkbox"/> 性格(優しい、きびしい、面白い、明るい、親切)  <input type="checkbox"/> その他(好き、嫌い、料理が上手、きれい、背が高い/低い)</p>	<p>・思い出すのに少し時間がかかるが、町に関する必要な語彙がある。 <b>A2レベルの以下のような指標がほとんど言える</b></p>	<p>・町に関する基本的な語彙がだいたいある。基本以上の語彙もある程度使える。 <b>A2レベルの単語に加えて、以下のような指標が2つ以上言える</b></p> <p>(指標)  <input type="checkbox"/> 環境がいい/悪い  <input type="checkbox"/> 自然が恵まれている/豊か  <input type="checkbox"/> ～に接している  <input type="checkbox"/> 住みやすい/にくい  <input type="checkbox"/> 落ち着いた  <input type="checkbox"/> 経済/文化の中心  <input type="checkbox"/> 大都市  <input type="checkbox"/> 郊外</p>
流暢さ (JF: 話しことばの流暢さ)	<p>・言いたいことを表現するのが困難である。 <b>指標が3つ以上ある</b></p> <p>(指標)  <input type="checkbox"/> 言葉に詰まる  <input type="checkbox"/> 黙る時間がある  <input type="checkbox"/> 発話がとぎれとぎれになる  <input type="checkbox"/> 同じ言葉を繰り返す  <input type="checkbox"/> 言い直しをする</p>	<p>・あまり困難なく、言いたいことが表現できる。 <b>指標が2つある</b></p>	<p>・困難なくスムーズに言いたいことが表現できる。 <b>指標が1つ以下ある</b></p>
会話運用能力 (CEFR: 発言権、柔軟性)	<p>・簡単な方法を使って、短い会話を始め、続け、また終えることができない。 <b>下線の部分ができない</b></p> <p>(指標)  <input type="checkbox"/> 必要な情報交換: 本のタイトル、借りた理由、借りる期間  <input type="checkbox"/> 相手の言うことが分からない場合は黙らずに、簡単に聞き返す。「え/もう一度/分かりません/何ですか」  <input type="checkbox"/> 基本的なあいづち「ええ、はい」</p>	<p>・簡単な方法を使って、短い会話を始め、続け、また終えることができる。 <b>下線の部分をふくんで、2つ以上の指標ができる。</b></p>	<p>・会話を始め、続け、また終えることがスムーズにできる。 <b>A2レベルのことに加えて、以下のような指標が1つ以上できる。</b></p> <p>(指標)  <input type="checkbox"/> 補足的な情報交換(パーティーの雰囲気、先生が参加することに対する皆の気持ち、行き方)  <input type="checkbox"/> 上手な聞き返し「すみません何ですか/御都合が悪いのですか/何とおっしゃいましたか」  <input type="checkbox"/> 多様なあいづち「あのう、ええと、へえ、そうですね、そうですね」</p>
社会言語学的能力 (JF: 社会言語的な適切さ)	<p>・日常的な挨拶や呼びかけなど、礼儀正しい言葉遣いが殆どできない。 <b>1つ～3つの指標ができる</b></p> <p>(指標)  <input type="checkbox"/> 「～先生」と呼びかける  <input type="checkbox"/> 「あのう」と言う  <input type="checkbox"/> 「ありがとうございます/ました」  <input type="checkbox"/> 「失礼します/ました」  <input type="checkbox"/> 「です/ます」体を使う</p>	<p>・日常的な挨拶や呼びかけなど、礼儀正しい言葉遣いが殆どできる。 <b>4つ～5つの指標ができる</b></p>	<p>・尊敬語的、謙譲語的な表現、丁寧な表現などがある程度使える。 <b>A2レベルのことに加えて、以下のような指標が1つ以上できる。</b></p> <p>(指標)  <input type="checkbox"/> 「～お持ちですか」と言う。  <input type="checkbox"/> 「～貸していただけないでしょうか」と言う  <input type="checkbox"/> 「伺います/まいります」  <input type="checkbox"/> 「失礼いたします」</p>

資料 3-1 質問紙調査の集計 (調査対象者 28 名)

	質問項目	強く	やや	あまり	全然
<b>テストについて</b>					
1	テストでやったことは授業でやっている活動と似ていると思いますか。	0	14 (7/7)	*13 (6/7)	1 (1/0)
2	テストは自分のレベルに合っていると思いますか。	4 (1/3)	23 (12/11)	1 (1/0)	0
3	テストの時間は適当だと思いますか。	1 (0/1)	24 (11/13)	3 (3/0)	0
4	テストの時、自分の会話能力を発揮できたと思いますか。	1 (1/0)	15 (5/10)	12 (8/4)	0
5	テストの内容は話しやすい内容だと思いますか。	2 (0/2)	17 (7/10)	9 (7/2)	0
<b>ルーブリックのフィードバックについて</b>					
6	ルーブリックのフィードバックは実際の自分の日本語能力に合っていると思いますか。	6 (2/4)	21 (11/10)	1 (1/0)	0
7	ルーブリックのフィードバックを読んで、これまで気がつかなかったことが分かるようになったと思いますか。	11 (6/5)	15 (8/7)	2 (0/2)	0
8	ルーブリックのフィードバックは今後の日本語の勉強に役に立つと思いますか。	14 (8/6)	14 (6/8)	0	0
9	今後も、このようなルーブリックのフィードバックをもらいたいと思いますか。	12 (8/4)	16 (6/10)	0	0
<b>テスト①と②について</b>					
10	テスト①とテスト②は同じレベルだと思いますか。	4 (2/2)	22 (11/11)	2 (1/1)	0

※「強くそう思います」「ややそう思います」「あまりそう思いません」「全然そう思いません」

\*13(6/7): 13は選択した調査対象者の総人数であり、6は大学生の人数、7は高校生の人数である。

(1) 日本語学習期間: 大学生: 6ヶ月半(13人); 3年6ヵ月半(1人)

高校生: 3年程度(14人)

(2) 日本語能力の受験経験: 大学生: 1人、2級

高校生: 14人; 2級(12人); 3級(1人); レベル不明(1人)

資料 3-2 質問紙調査の集計 (評価者 2 名)

	質問項目	強く	やや	あまり	全然
<b>テストについて</b>					
1	テストでやったことは普段授業でも取り入れていると思いますか。	1	1	0	0
2	テストの内容は学生の日本語能力と比べて、適当だと思いますか。	0	2	0	0
3	テストの長さは適当だと思いますか。	0	2	0	0
4	このテストで学生の会話能力が正しく測れると思いますか。	0	1	1	0
<b>ルーブリックについて</b>					
5	テストで使用したルーブリックは評価しやすいと思いますか。	0	1	1	0
6	ルーブリックのフィードバックは今後の授業に役に立つと思いますか。	2	0	0	0
<b>ルーブリックのフィードバックについて</b>					
7	ルーブリックのフィードバックは実際の学生の日本語能力にあっていると思いますか。	1	1	0	0
8	ルーブリックのフィードバックは学生の日本語の勉強に役に立つと思いますか。	2	0	0	0
<b>テスト①とテスト②について</b>					
9	テスト①とテスト②は同じレベルだと思いますか。	2	0	0	0
<b>テスト実施可能性について</b>					
10	今後、このテストを大学に入った日本語既習者に対して、実施する可能性があるといますか。	0	2	0	0

※「強くそう思います」「ややそう思います」「あまりそう思いません」「全然そう思いません」

2010年4月 日

## 口頭テストの結果

——さんは以下のことができました。

1. 毎日の生活、趣味、好き嫌い、学習経験などの身近な話題についての簡単な質問に答えることができる。将来の希望や簡単な意見を述べるができる。

A2レベル未満

A2レベル

A2レベルより上

2. 自分の家族、住んでいる町やふるさとなどの身近な話題について、ある程度まとまりのある話ができる。

A2レベル未満

A2レベル

A2レベルより上

3. 身近な場面で、人を誘ったり、人に何かを頼んだりのように、簡単に日常的な課題を遂行することができる。また、挨拶、別れ、感謝などの社交的な表現を使うこともできる。

A2レベル未満

A2レベル

A2レベルより上

評価者：ホン、バン

A2レベルはJFスタンダードのレベルの1つであり、以下のことができるレベルである。

- ごく基本的な個人的情報や、家族情報、買い物、近所、仕事など、直接的関係がある領域に関する、よく使われる文や表現が理解できる。
- 簡単に日常的な範囲なら、身近で日常の事柄についての情報交換に応ずることができる。
- 自分の背景や身の回りの状況や、直接的な必要性のある領域の事柄を簡単な言葉で説明できる。