

# 『日本語学習者による日本語／母語発話の対照言語データベース』の設計

宇佐美 洋・籠宮隆之・梶本総子

## 1. はじめに

国立国語研究所 日本語教育部門では、日本語教育に関する各種研究データとして、また日本語教育実践・教師研修等における「素材」として活用するため、日本語学習者の日本語データと、それと内容的にはほぼ対応する母語データを大量に収集し、データベースとして公開する事業をおこなっている。

これまでに書きことばに関するデータベースとして、『日本語学習者による日本語作文と、その母語訳との対訳データベース』(略称：作文対訳 DB)を 2001 年に公開した。このデータベースには、日本語学習者による日本語作文(1 編あたり 400～800 字程度)と、執筆者本人によるその母語訳約 1,000 名分が収録されており、収集データに基づく各種調査研究もおこなってきた(前田(宇佐美)2001)。

2002 年からはこれに平行して、音声言語でも同様のデータベースを作成する事業を始めている。具体的には、

- 1) 課題に基づき、日本語学習者がおこなった日本語発話データ
- 2) 1)と同一の発話者が、同一の課題に基づき母語でおこなった発話データ

を大量に収集し、相互参照が可能な形で電子化する、という計画である。現時点では、中国語・韓国語・タイ語を母語とする日本語学習者にデータ提供をお願いするとともに、日本語を母語とする人々にも、学習者と同じ課題に基づく発話データの収録をおこなっている。

本論では、現在作成中の『日本語学習者による日本語／母語発話の対照言語データベース』(略称：発話対照 DB)について、その作成意義と、データベース設計について述べる。

## 2. データベース作成の目的

### 2.1. 「教育現場への応用」を目指すデータベース

国立国語研究所で作成している発話対照 DB の目的は、第一義的には「教育現場への応用」である。この「教育現場」には、学習者に対する教育現場だけでなく、日本語教師に対する研修の現場も含まれる。

この目的のため、本データベースでは、

- 1) 学習者の日本語発話と、それとほぼ同内容の母語発話との対照

## 2) 学習者の話しことばと書きことばとの対照

という2種類の対照が可能となるよう、データベースを設計することとした。以下、この2種類の対照が、日本語教育においてどのような意義を持つかについて述べる。

### 2.2. 「日本語データ」「母語データ」の対照

教師が学習者に対し、発話(会話・スピーチなどを含む)の指導をおこなう際、教師はまず学習者本人が何を話したいのかを正確に理解する必要がある。しかし実際には、学習者の発話を聞くだけでは、本人がいったい何を意図しているのか理解しがたいことが多い。

また、誤用や不自然さを生み出す要因にはさまざまなものがある。言語内的要因(音声・語彙・文法など)だけでなく、語用論的ストラテジー(話の切り出し方や収め方、特定の言語機能を実現するために用いられる言語形式の選択など)の違いが、コミュニケーションにおける違和感の原因となることも少なくない。しかし、誤用・不自然さがどのような要因に由来しているのか、日本語発話を聞くだけでは特定することは難しい。

そこで本データベースでは、学習者の日本語発話だけでなく、ほぼ同じ内容での母語発話も収録し、両者を対照できるような形式をとった。これにより、以下のようなメリットが得られる。

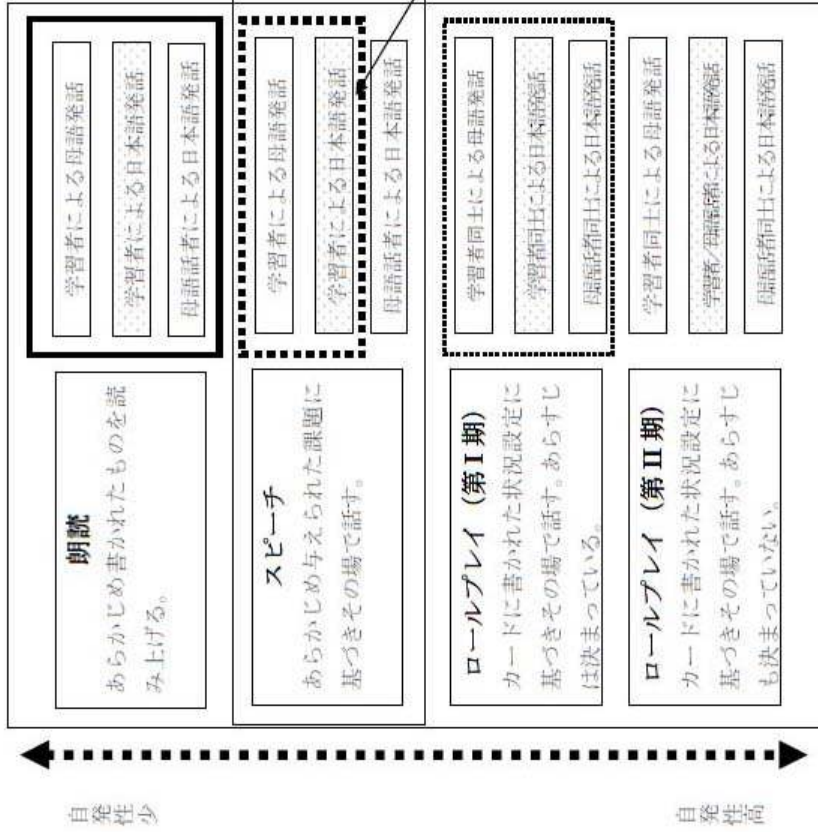
- 1) 学習者本人による母語発話を参照することで、本人が何を意図していたのかを知り、学習者の発話の誤用・不自然さの要因が特定可能となる。
- 2) 語用論的な問題に関して、学習者と日本語母語話者との相違を知る手がかりを提供する。
- 3) 日本語としての誤用・不自然箇所と、母語訳の対応箇所の表現を参照することで、母語干渉の有無を検証できる。また、その誤りや不自然さがなぜ生じたのか、その理由をある程度推測することができる。さらにそうした考察を通じ、日本語とX語との対照言語学的研究へと発展させることができる(宇佐美 2002)。

### 2.3. 「作文対訳 DB」「発話対照 DB」の対照

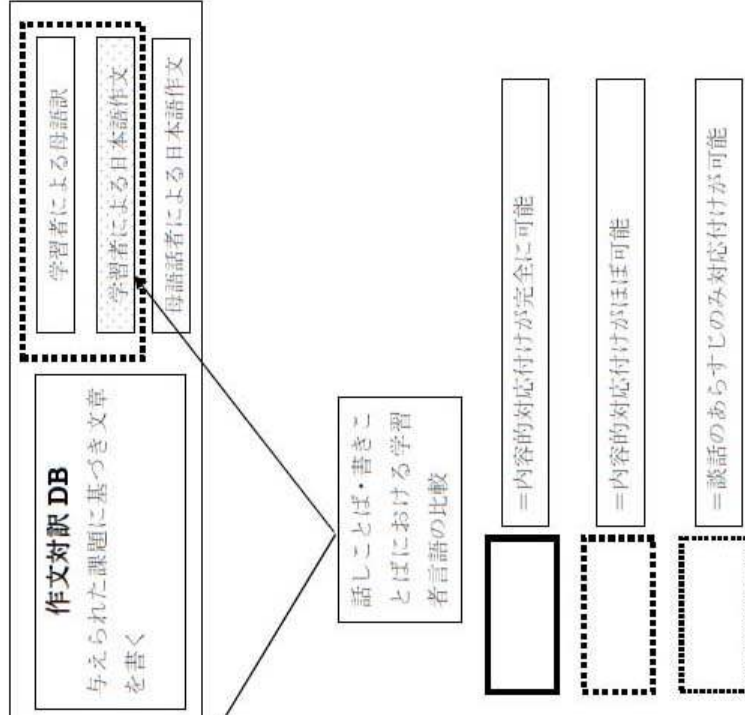
従来日本語教育の分野で公開された学習者データベースは、話しことばのみ、または書きことばのみを対象とするものがほとんどであり、話しことばと書きことばを直接比較対照するための手段がなかった。しかし経験的に言って、話しことばと書きことばとは誤用・不自然表現等の種類・頻度がかかなり違っている。話しことば教育・書きことば教育をより効率的に進めていくためには、学習者の話しことば・書きことば双方の特徴の異同を明らかにしておく必要がある。

このため国立国語研究所の学習者データベースでは、「作文対訳 DB」と「発話対照 DB」との相互対照を目的にデータ収集をおこなっている。この両者を使用することで、学習者の話しことばと書きことばとの比較研究という、研究の新しい視野が開かれることになる。

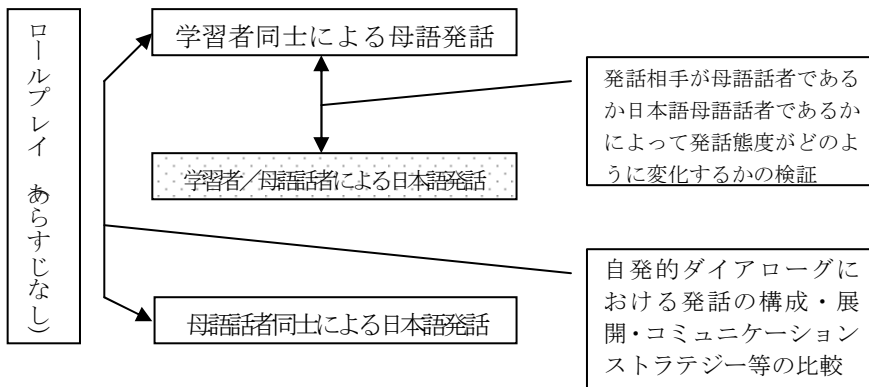
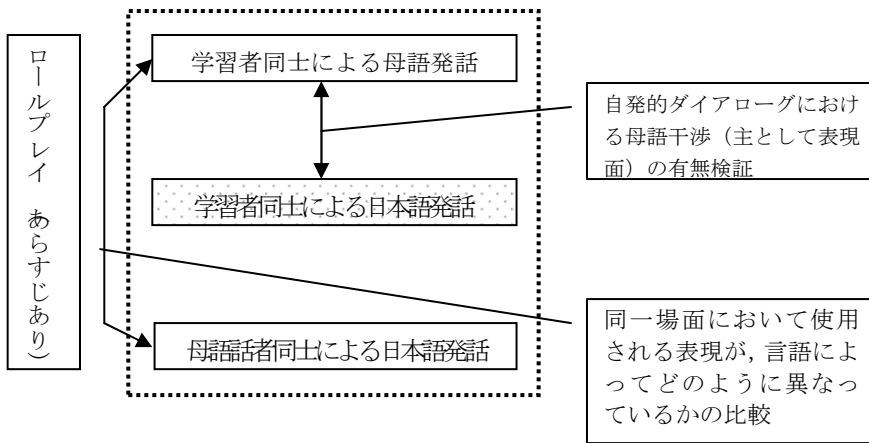
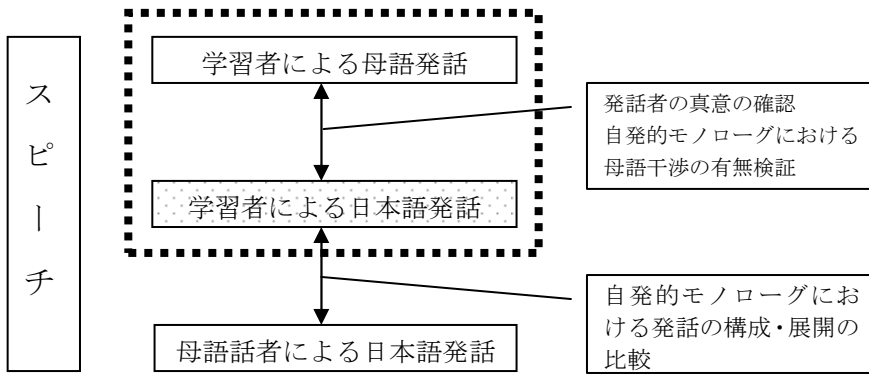
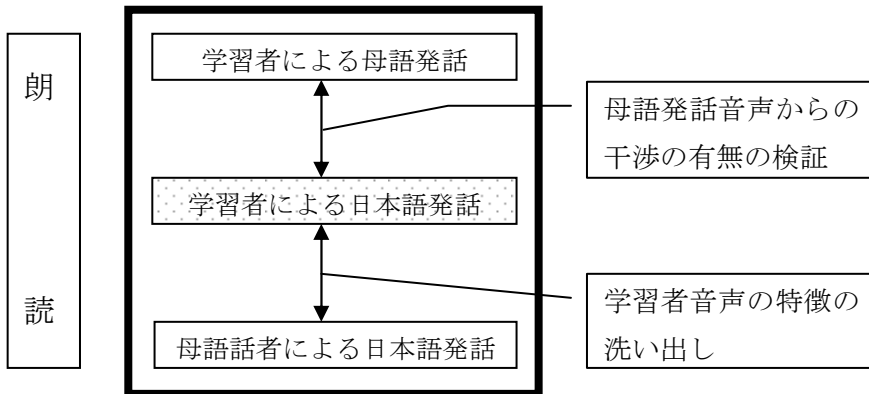
## 話しことば



## 書きことば



データベース各部門と、それぞれの部門を使用してできること



### 3. 発話課題

発話のためのタスクとしては、

- A. 朗読：1人あたり2または3課題
- B. スピーチ：1人あたり4課題
- C1. あらすじありロールプレイ：1ペア(2人)あたり4課題
- C2. あらすじなしロールプレイ：1ペア(2人)あたり4課題

を準備した。

データ収集は、第1期(02年～03年)、第2期(03年～04年)の2回に分けておこなう、それぞれ異なる課題のセットを使用した。データ提供者には朗読、スピーチ、ロールプレイという3種類の課題すべてについて発話をお願いしたが、ロールプレイに関しては、第1期においては「C1. あらすじなし」の課題を、第2期では「C2. あらすじあり」の課題を使用した(詳細は「5.3.ロールプレイ」を参照)。

学習者は、与えられた全課題に対し日本語・母語の双方でデータ提供をおこなったが、日本語母語話者は日本語のみのデータ提供をおこなっている。

### 4. データベース規模

現在までに中・韓・タイ・日合わせて約220名、約160時間分<sup>1</sup>のデータを収集済みであるが、最終的には約240名、約175時間分の収録を目標としている。

	データ種類	時間数
A. 朗読	NNSの日本語発話	39時間
	NNSの母語発話	28時間
	NSの日本語発話	7時間
B. スピーチ	NNSの日本語発話	36時間
	NNSの母語発話	28時間
	NSの日本語発話	7時間
C1 RP あらすじあり	NNSの日本語発話	8時間
	NNSの母語発話	7時間
	NSの日本語発話	4時間
C2 RP あらすじなし	NNSの日本語発話	5時間
	NNSの母語発話	4時間
	NSの日本語発話	1時間

表1 データベース規模(最終目標)

<sup>1</sup> 電子情報通信学会技術研究報告 Vol.104, No.148に掲載した論文では総時間数の計算に誤りがあったため、本稿では改めた。

## 5. 課題の詳細

### 5.1. 朗読

朗読は、紙に書かれた日本語の原稿・母語の原稿をそのまま読み上げるというタスクである。朗読の課題としては、純粋な書きことば（意見文・説明文）と、一部話しことばも混在するもの(物語文)の両方を含めることとした。

また「中国語／韓国語／タイ語を母語とする学習者にとって、特に発音しにくいと思われる音を含んだ日本語短文のリスト<sup>2)</sup>」を示し、それを読み上げてもらう、ということもおこなった（この課題は日本語の読み上げのみであり、母語訳はない。またこの課題は、第2期のデータ収集時に使用した）。

読み上げ原稿は、日本語・母語訳(中国語・韓国語・タイ語)ともにデータベース作成者が作成した。日本語の原稿にはすべての漢字にルビが振られている。

#### <日本語>

赤ずきんは森の近くの村に住んでいる女の子です。ある日赤ずきんはお母さんに、お使いを頼まれました。

「赤ずきん、おばあちゃんにお菓子を持っていてくれる？」

「お菓子？ いいわよ、お母さん」

赤ずきんはさっそく、かごにお菓子を入れて出かけました。

(以下略。データ提供者用の原稿はルビつき)

#### <韓国語訳>

빨간두건은 숲속 마을에 살고 있는 여자 아이입니다. 어느날 빨간두건은 엄마의 심부름을 하게 되었습니다.

“빨간두건아, 할머니께 과자 좀 갖다주겠니?”

“과자? 좋아요, 엄마”

빨간두건은 바구니에 과자를 넣고 집을 나섰습니다.

(以下略)

### 例 1 朗読課題 物語文例(一部)

<sup>2)</sup> この短文リストは、国立国語研究所 日本語教育部門第二領域で作成している「発語発声訓練システム用例文集」から一部を抜粋して作成した。この例文集は、「ITを活用した日本語学習環境の整備」事業の中の、「対照言語の研究成果による言語資料の発信」プロジェクトにより作成されたものである。例文は以下 URL を参照。

<http://jweb.kokken.go.jp/eJapan/contents/sound/toppage.htm>

<韓国語母語話者用短文リスト>

1. 現在は会社員です。
2. 餃子が食べたいです。
3. 冗談ではありません。
4. バスが来ました。
5. 貿易会社につとめています。

(以下略)

<タイ語母語話者用短文リスト>

1. おじさんもおじいさんも元気です。
2. クーラーが故障しました。
3. プレゼントを選んでいきます。
4. 自分でやってみてください。
5. 単位がたりません。

(以下略)

(詳細は pp.16-34, pp.52-70 の「朗読原稿」参照)

スピーチやロールプレイなどの自由発話においては、個別の単音の音声的な実現形にか  
なりの変異が予想されるため、音声上の母語干渉の有無を検証する際にはやや使いにくい。  
そこで「朗読」という、単音の実現形に大きなゆれが出にくい状況を設定し、そこで得られ  
るデータを音声上の基本データとして使用することとした。

## 5.2. スピーチ

スピーチは、あらかじめ指定されたテーマに基づき、学習者が「何も見ないで」話をする  
というタスクである。データ提供者には、当方で用意したスピーチ課題をひとりにつき計 4  
課題提示し、各課題につき、目安として「3 分程度」話すようお願いした。まず日本語の  
スピーチを先に収録し、その直後に母語で、同内容のことを話すように求めた。与えられ  
たテーマに基づき、即興で 3 分間スピーチをすることはかなり難しいため、少なくとも録  
音の前日まではスピーチ課題をデータ提供者に渡し、録音時に話す内容についてあらか  
じめ考えてきてもらった。ただ、「書きことばの朗読」にはならぬよう、読み上げ原稿を文  
章の形で書き出すことはしないよう強くお願いした。また録音時には、基本的には何も見  
ないで話をするよう求めたが、日本語スピーチを母語スピーチの内容が大きく食い違わな  
いようにするために、簡単な「あらすじ」や「キーワード」のみを書いたメモを適宜参照する  
ことは許可した。

1. 日本とあなたの国とをくらべて、いちばん違うと思うことはどのようなことですか？ 実際の例を挙げて説明してください。
2. 大学受験について、ある人は、入学試験を難しくして成績の優秀な学生を選ぶのがいいと言います。また、ある人は、入学試験をやさしくして卒業を難しくするほうがいいと言います。あなたはどちらの意見に賛成しますか？ 理由を挙げてあなたの考えを述べてください。
3. 外国語の学習について、一つの言語をできる限り深く専門的に学ぶのがいい、という考え方があります。また、不完全でもいいから、できるだけたくさん言語を学ぶほうがいい、という考え方もあります。あなた自身はどう思いますか。理由を挙げて、あなた自身の考えを述べてください。
4. あなたの国の歴史のなかから大きな事件をひとつえらんで、それがいつ起きた、どのようなできごとであったかを説明してください。

#### 例 2 スピーチ課題例

(詳細は p.35, 71 の「スピーチ課題」参照)

本データベースの中では、スピーチの部門がもっとも中核的な地位を占めている。なぜなら、「与えられたテーマに基づき、意見や経験、事実を述べる」という課題は、「作文対訳 DB」における作文課題と相似したものであるからである。話しことばと書きことばとの直接比較をおこなうのであれば、「発話対照 DB」の中ではこの部門を使用するのがもっとも適切である。

### 5.3. ロールプレイ

ロールプレイは、2人のデータ提供者がペアになり、それぞれに渡された「ロールカード」の状況設定にしたがって話をするタスクである。ロールカードも、日本語のものと母語のものが用意されており、日本語での対話は日本語カードを見ながら、母語での対話は母語カードを見ながらおこなうようお願いした。ロールカードは、データ提供者同士が話の進め方についてあらかじめ相談することをできるだけ防ぐため、録音の直前に渡すものとした。

それぞれのロールカードには、「依頼と断り、最終的な受諾」「抗議と謝罪、許し」など、さまざまな言語機能を織り込んだ場面が設定されており、データ提供者は4種類のロールカードに基づいて(すなわち4種類の場面設定にしたがって)対話することが求められる。それぞれの言語機能を実現するためにどのような言語形式を使用するか、どのようなストラテジーをとるかはデータ提供者に任されている。



前述の通り、ロールプレイ課題には「あらすじあり」と「あらすじなし」があり、データ提供者は双方のうちいずれかのみを与えられた。以下、これら 2 種類のタスクについて説明する。

### 5.3.1.あらすじありロールプレイ（第 1 期）

「あらすじありロールプレイ」は、母語でのロールプレイと日本語でのロールプレイとの直接対照を可能にするために収集された。この形式のロールプレイは、第 1 期に録音をおこなった。

このロールプレイは、同じ母語をもち、日本語レベルもほぼ同じと考えられる 2 人の学習者の間でおこなわれた。ロールカードにはやりとりの「あらすじ」が示されており、母語・日本語のいずれで対話をおこなうときも、対話の流れは同じになるようになっていた。

日本語で対話をおこなう際、データ提供には「日本人になったつもりで」、つまり、日本人であればカードに示された状況でどのような行動をとるかを想像して対話をおこなうよう求めた。これに対して母語で対話をおこなう際には、(中国・韓国・タイ)人としての「素の自分」になって話をするよう求めた。したがって日本語での対話の際と母語での対話の際とでは、話の細部が多少変わってもかまわない、と伝えた。

次ページにロールカードの例を示す。ロールカードの A と B は対になっており、データ提供者のうちひとりが A を、もうひとりが B を見ながら会話をおこなう。

<ロールカード A>

あなたは日本人の大学生です。

ある日、同じクラスの友達があなたの部屋を訪ねてきました。あまり親しい友達ではありませんが、悪い友達でもありません。話を聞いてみたら、お金を貸してほしいという話です。お父さんが急に入院し、手術することになったので、10万円ほどのお金がすぐに必要だということです。ちょうどあなたには、夏休みに海外旅行をするために貯めた10万円があります。長い間かかって貯めたお金なので、最初は貸すことを断わりますが、しかし、来月には必ず返してもらえることを聞いて、結局は友達にお金を貸すことにします。

<ロールカード B>

あなたは日本人の大学生です。

お父さんが急に入院し手術することになったので、10万円ほどのお金がすぐに必要になりました。何人かの友達から断われ、最後にAさんの部屋を訪ねます。Aさんは、あまり親しい友達ではありませんが、悪い人ではありません。事情を説明し、10万円を貸してくれよう頼みます。最初は断われますが、今やっているアルバイトの給料が来月初めにもらえるので、そのときには必ず返せることを説明して、やっと、10万円を借りることができます。

例3 日本語ロールカード例(あらすじあり)

(ロールカード母語訳は省略。詳細は pp.36-51 の「第1期ロールプレイ・ロールカード」参照)

**5.3.2.あらすじなしロールプレイ(第2期)**

あらすじのあるロールプレイは、日本語発話と母語発話との間でアラインメント(対応箇所同士の関連付け)がとりやすいという利点がある。しかし日本語ロールプレイと母語ロールプレイとを比較する際、単に表現の対照だけでなく、話の進め方も含めた各種ストラテジーも研究の対象に含めようとする場合、「あらすじのないロールプレイ」も必要である。

そこで第2期では別途、「あらすじなしロールプレイ」という課題を設定し、こちらの枠組みにより録音をお願いした。このロールプレイで用いるロールカードには状況設定のみを記述し、詳細なあらすじは記載しなかった。また話しことばデータとしての自然さを高めるため、日本語ロールプレイは、学習者1名と、日本語母語話者1名との間でおこなう

こととした。このことにより、日本語母語話者と学習者との「接触場面」に関する研究にも使用することが可能となる（なお、母語によるロールプレイは同じ母語を持つ学習者同士の間でおこなわれた）。

<ロールカード A>

来週の水曜日は先生の誕生日です。それで、なにかいいプレゼントを贈ろうと思います。あなたは、プレゼントを一人で贈るより、ほかの友達と一緒に贈ったほうがいいと思っています。そこで友達のところに行って、一緒にプレゼントを贈ろうと誘い、どんなプレゼントがいいか、相談して決めてください。あなたは、プレゼントはあとにずっと残るものがあると思っています。

<ロールカード B>

友達が訪ねてきて、来週の水曜日は先生の誕生日なので、一緒にプレゼントを贈ろうと誘いました。あなたも、それはとてもいいことだと思いました。

そこで、友達と相談して、どんなプレゼントにするか決めてください。あなたは、プレゼントには先生の好きな花か食べ物がいいと思っています。

例 4 日本語ロールカード例(あらすじなし)

(ロールカード母語訳は省略。詳細は pp.72-87 の「第 2 期ロールプレイ・ロールカード」参照)

## 6. データ収録

### 6.1. データ提供者の属性

前述の通りデータ提供者は、「中国語・韓国語・タイ語のいずれかを母語とする日本語学習者」にお願いしている。また学習者データとの比較のため、日本語母語話者からも同一課題によるデータ提供をお願いしている。

データは日本国内・国外、両方で収集をおこなった。

最初、日本国内の大学・日本語学校等で日本語を学んでいる学習者に、国立国語研究所内のスタジオにて録音を依頼していたが、日本国内の学習者にデータ提供を依頼する際、出身地・学習歴・性別などの条件を統制することは現実問題としてかなり困難であった。しかしロールプレイの際、出身地や日本語レベルが違うデータ提供者同士がペアとなって

発話をおこなうと、研究目的によっては不都合が生じることがある。

また、日本国内で生活している学習者は、日本人との直接接触の機会をすでに多く持っているものと考えられる。文化・言語背景による会話のストラテジーの相違を見るためには、「日本での生活経験を持たない学習者」の発話データを取っておくことも必要である。

そこで一部のデータについては、ソウル・バンコクの日本語教育機関にデータ提供者の手配を依頼し、現地にて収録をおこなった(中国については次期研究期間内に現地での録音をおこなう予定)。データ提供者を手配する際の統制条件は以下の3つである。

- 1) 当方で指定する発話課題に基づき、何も見ないで3分以上日本語で発話を続ける程度の日本語能力を持っていること(大まかな目安としては学習歴2年以上)
- 2) 首都(ソウル・バンコク)またはその隣接県において、言語形成期(3~12歳ごろ)を過ごし、そこで第一言語(韓国語またはタイ語)を習得したこと
- 3) 1ヶ月以上日本に滞在した経験をもたないこと

また、ロールプレイのペアを作る際には、男・男、女・女、男・女のペアができるだけ偏りなく含まれるようお願いしたが、国によっては日本語学習者の性別に大きな偏りがある場合があるため(タイの大学で日本語を学ぶ学生には女性が極めて多い)、この点に関しては必ずしも当方の希望通りとはならなかった。

データ提供者情報としては、出身地・性別・年齢のほか、学習歴(教育機関・学習期間・週当たりの学習時間数)、日本での滞在歴をたずねた。また、第2期に収録をおこなったデータ提供者については、日本語能力を簡便に測定できるテストである“SPOT”(小林他1996)を実施し、その点数を記録した。データ提供者情報も、音声データ、書き起こしデータとともに公開されるが、個人を特定する手がかりとなる個人名などの情報は公開されない。

## 6.2. データ収録条件

録音は、デジタルテープレコーダ(DAT)を用いておこなった。録音時のサンプリング周波数は48kHz、16bitである。

日本国内での録音は、防音処置の施された「対面録音室」でおこなった。この部屋は、ガラス窓で仕切られた2つの小部屋が向かい合うような形で隣接しており、それぞれの部屋にいる者同士がお互いの顔を見ながら会話をすることができるが、部屋の間には完全な防音が施されており、音声のやりとりはマイクとヘッドホンを通じてのみおこなわれるようになっている。2つの部屋の音声を完全に別トラックに分けて録音することができるため、ロールプレイの収録には特に有効である。

一方海外での収録では、録音専用のスタジオを利用することはできず、一般の教室にて収録をおこなったため、録音には時として雑音が入っていることがある。ロールプレイに

おいては、二人の発話者が同じ部屋の中でそれぞれヘッドセットを装着して並び、お互いの顔が直接見える状態で録音をおこなった（発話者間の距離は 30～50cm 程度である）。このためロールプレイデータにおいて両トラックが十分に分離できていない場合もある。こうした録音条件についてもデータの付帯情報として公開する。

### 6.3. データ公開についての許諾

データ提供者には以下の点について十分な説明をし、承諾書にサインをしていただいた。

- 1) 日本語・母語の発話データおよびデータ提供者情報は、教育・研究目的のために公開されるが、その他の目的のためには使用されないこと
- 2) 個人名・所属など、個人を特定する手がかりとなる情報は公開しないが、文字化データだけでなく音声データもそのまま公開されるので、個人を特定される可能性がまったくないとはいえないこと

## 7. データベース公開時の仕様

### 7.1. 書き起こしテキストと音声との同期

本データベースの書き起こし方法は、国立国語研究所で作成している『日本語話し言葉コーパス(CSJ)』でのやり方に準じている(前川他 2000, 小磯他 2001, 小磯他 2004)。

発話中、原則として 200ms 以上のポーズが見られる位置で発話を分節した。こうして得られる単位を「発話単位」と呼ぶ。書き起こしは、この単位ごとにおこなわれる。

書き起こされたテキストは、「発話単位の開始・終了についての時間情報」と関連付けられているため、個々の「発話単位」の書き起こしテキストと、音声ファイルの対応箇所との間に同期がとれることになる。

書き起こしは、日本語データについては CSJ と同様、「表記のゆれを統一した、漢字かな混じりのテキスト」と、「実際の発音をカタカナでできるだけ正確に書き起こしたテキスト」の 2 種類でおこなった。一方母語データの書き起こしについては、その言語の正書法によるテキストのみにとどめた(書き起こしの実例については例 5 参照)。

### 7.2. 付加情報

書き起こしと同時に、言い直し、言い間違い、フィラー、音節の引き伸ばしなど、談話に伴って現れる各種現象を表現するタグを、テキスト中に挿入している。日本語データに挿入するタグ種も、前述 CSJ での使用タグに準じている(前川他 2000, 小磯他 2001)が、一部独自に変更・拡張した点もある。

母語データに関しても、基本的な談話現象(言い直し、言い間違いなど)に関するタグは日本語データ用と共通のものを使用している。しかし言語特有の音声現象もありうるため、

今後も書き起こし作業を進めながら必要なタグを定義していく予定である。

このデータベースを構築するに当たって使用したタグ種とその付与基準については、p.89以下の「書き起こし・タグ付けマニュアル」を参照。

### 7.3 モニター版の公開

上記のようなやり方で整理したデータのうち、第1期収集分<sup>3</sup>の一部を、2005年3月、『日本語学習者による日本語／母語発話の対照言語データベース・モニター版』として公開することとした。

モニター版に含まれるデータは以下の通りである（なお以下で「データ」というのは、「1つの電子ファイルとして保存されている、ひとまとまりの発話」のことである）。

#### <日本語発話データ>

朗読	1人あたり 3データ×10名×4(中・韓・タイ・日)=120データ
スピーチ	1人あたり 4データ×10名×4=160データ
ロールプレイ (あらすじあり)	1組(2人)あたり 4データ×5組×4=80データ

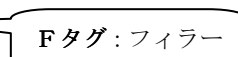
計：360データ

#### <母語発話データ>

朗読	1人あたり 3データ×10名×3(中・韓・タイ)=90データ
スピーチ	1人あたり 4データ×10名×3=120データ
ロールプレイ (あらすじあり)	1組(2人)あたり 4データ×5組×3=60データ

計：270データ

モニター版に収録されている全データサイズは約29時間分（約4GB）になる。また音声データは、48kHz, 16bitで収録したものを、16kHzにダウンサンプリングしてデータベースに収録している（朗読・スピーチはモノラル、ロールプレイはステレオ）。

-----  
--  
0002 00000.857-00002.573 L:  
タイの & タイノ  
歴史の中から & レキシノナカカラ  
0003 00003.412-00004.707 L:  
(F あの) & (F アノ) 

<sup>3</sup> 今回のモニター版には第2期収集分のデータは収録されていない。今回未公開のデータについても、整理が付き次第公開の予定である。

の中で	&	ノナカデ
0004 00004.907-00007.738 L: .		
色々な	&	イロイロナ
大きな事件が	&	オーキナ(W チケン;ジケン)ガ
あると思います	&	アルトオモイマス
0005 00008.040-00008.615 L:<gulp>		

W タグ：言い誤り・発音の忘れ

例 5 日本語スピーチ(タイ語母語話者)の書き起こしテキスト例

0001 00000.165-00002.370 L:  
 สำหรับใน(RP ประหวัดสาด;ประวัติศาสตร์)ไทยนี้  
 0002 00002.403-00002.422 L:<click>  
 0003 00003.097-00007.518 L:  
 มีเหตุการณ์หลายๆเหตุการณ์ที่ถือว่าเป็นเหตุการณ์ที่ยิ่งใหญ่

RP(real pronunciation)タグ：正書法では1種類の表記で書かれる語が自由変異として複数の読みを持ち得る場合、「実際にどう読まれたか」を“;”の前に、正書法どおりの表記を“;”の後ろに記す。

例 6 タイ語スピーチの書き起こしテキスト例(上記データと同一話者, 同一内容)

## 文 献

- [1] 宇佐美洋, “「対訳作文データベース」と日本語教育——対照言語学を教育に生かすために,” 国立国語研究所(編), 日本語と外国語との対照研究 X『対照研究と日本語教育』, pp.81-94, 国立国語研究所, 2002.
- [2] 小磯花絵, 土屋菜穂子, 間淵洋子, 斉藤美紀, 籠宮隆之, 菊池英明, 前川喜久雄, “「日本語話し言葉コーパス」における書き起こしの方法とその基準について,” 日本語科学, 9, pp.43-58, April. 2001.
- [3] 小磯花絵, 間淵洋子, 西川賢哉, 斉藤美紀, 前川喜久雄, “転記テキストの仕様,” 『日本語話し言葉コーパス』添付電子文書
- [4] 小林典子, フォード丹羽順子, 山元啓史, “日本語能力の新しい測定法[SPOT],” 世界の日本語教育, vol.6, pp.201-218, 1996.
- [5] 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明, “日本語話し言葉コーパスの設計,” 音声研究, 第4巻第2号, pp.51-61, Aug.2000.
- [6] 日本語教育のためのアジア諸言語の対訳作文データの収集とコーパスの構築, 前田(宇佐美)洋 (編), 課題番号国 11691041 平成 11-12 年度科学研究費補助金基盤研究(B)(2)研究成果報告書, 国立国語研究所, 2001.

※本稿は, 2004年6月24日, 25日に開催された電子情報通信学会/日本音響学会 音声研究会 研究例会(日本音声学会・日本音響学会聴覚研究会と共催)における発表原稿(電子情報通信学会技術研究報告 Vol.104, No.148, pp.29-34 に掲載)をもとに, 発表後の変更事項などを踏まえ, 加筆修正を施したものである。なお, この発表は宇佐美洋・籠宮隆之・相本総子の連名でおこなわれたが, 本報告書に採録するに当たっての加筆修正は宇佐美の責任の下におこなった。