

日本語教育データベースの構築：

その課題と可能性について

野山 広（国立国語研究所）

wisen@kokken.go.jp

1. はじめにー日本語教育データベース構築の背景

1990年の入管法の改正・施行以来、我が国に在留する外国人登録者数は増加の一途を辿り、家族を伴った外国人定住者の数も増えてきている。現在、多様な言語・文化背景を持った幅広い年齢層の人々が、地域社会のさまざまな場面で、日本人と多様なかかわりを持ちながら生活しているが、今後もこの状況は続くものと推測される。こうした多様な価値観を持った人々が共存する社会を多文化共生社会とするならば、社会状況に応じた外国人受入れのための適切な方策が必要となってくるとともに、日本語を母語としない言語生活者の需要に応じた言語教育の展開も期待される。

そこで、国立国語研究所では、本中期計画（2006～2010年度）の中で、こうした展開の充実へ向けた基礎資料・基盤情報を提供し、関係者の方々に出来る限り活用していただけるように、日本語教育データベース（以下、日本語教育DB）を構築しようと考えた。

本稿では、まず、整備普及グループの目標について触れるとともに、日本語教育DB構築の目的と内容、基盤情報の提供方法等について概観する。そして、そのコンテンツの中核的存在の一つと考えられる日本語会話データベースを事例として取り上げ、その構築の過程から見えてきたデータベース作りの課題と可能性について考察したい。

2. 整備普及グループの目標

整備普及グループは、先述の日本語教育データベースの構築を主要な業務として携わっており、グループの目標は以下の通りである。

（1）日本語教育情報の作成基盤（基盤情報の資源としてのDB等の）整備・成果の普及

1）日本語教育DBの構築

日本語使用実態に関する所内外のデータ、日本語教育研究の成果や日本語教育現場からの情報収集から得られた誤用例や習得難易度情報などの日本語の教育・学習データなどの基盤情報により構成されるDBの構築。

（2）成果の効果的・効率的な普及

日本語教育の基盤情報や資料を提供、普及させるとともに、状況に応じて、関連した研究の成果に関する情報も提供する。

3. 日本語教育DB構築の目的と内容：基盤情報の概要

(1) 目的

このデータベース作りは、徐々に多言語・多文化化する社会状況の中、地域で生活する日本語を母語としない人々の需要や要望に、日本語教育研究という分野から基盤情報を提供するべく、以下を目標として応えていこうとするものである。

- 1) 多文化共生社会に対応した日本語教育・学習内容・方法，評価の基準などの再構築に向けた基礎資料として提供すること
- 2) 既存の情報資源や日本語研究や日本語教育研究の成果を基盤としたデータベース作りを試み，提供することによって，例えば，生活日本語に焦点を当てた教材の作成，複合領域としての日本語コミュニケーション・会話研究，日本語教師教育・養成コースの新たな展開などに貢献すること
- 3) 新たな情報資源の一つとして，日本語会話データベースの構築，整備・普及，流通促進の充実を図る。そのことを通して，例えば，口頭能力テストの重要性や活用方法について注意を喚起するとともに，その評価の過程やフィードバック等を通じて得られる情報の有意義性に関する認識の深化を促すこと

(2) 内容：基盤情報の概要

1) 多文化共生社会に対応した基礎資料

学習項目，評価基準グループ，用例用法グループ，整備普及グループの新たな成果を，基盤情報としてデータベース化して提供する。

2) 既存の情報資源

①日本語教育ネットワーク，e-Japan，J-Web等のデータの活用・統合

先述の背景でも述べたような状況変化の中，かねてより，国内外の日本語教育や日本語学習に関わる関係者から，情報の集積や自由に使える素材を求める声があった。そこで，文化庁と独立行政法人国立国語研究所（以下，国研）は，多くの機関，関係者の連携・協力を得て，日本語教育支援のための環境を整備する Web サイト：日本語教育支援総合ネットワーク・システムを開設し，様々な日本語教育に関する資料や情報を提供することとなり，平成 13 年度から，本システムは国研において運用することとなった。この DBは，素材の提供を中心とした素材情報ネットワークと情報流通ネットワークで構成されていた。日本語教育DBには，この中で提供されていたコンテンツが活用されることとなる。

また，この日本語教育ネットワークのコンテンツ以外に，さまざまな関係者の協力を経て作成してきた e-Japan 関連のコンテンツ（高度情報通信ネットワーク社会の形成に関する重点計画の一環で，政府の IT 戦略本部の計画のもと作成されてきたもの）が入る。さらには，平成 5(1993)年に開始され，さまざまな情報や素材を公開していた J-Web（日本語教育の世界）のコンテンツも，できる限り統合されることとなる。

②日本語研究や日本語教育研究の成果：その整備と普及

「中国帰国者教材開発調査」「コミュニケーション・ストラテジー調査」の成果を基盤情報としてデータベース化し提供する。その他には、コースデザイン調査、擬態語・擬声語データ、日本語教育映像教材語彙表、教育基本語彙調査、漢字頻度表、日本語能力試験出題基準、学年配当表、分類語彙表などがデータベース化される。

③所内外の既存情報

CASTEL/J¹の関係者との連携・協働を積極的に図りながら、これまでに収集・整備された日本語教育用のコンテンツ（約 1 万点のイラスト集、見出し（漢字、ローマ字）、用例、要素、動作、感情、状況、音声等の情報付き）の活用方法や、これからの情報提供方法や流通促進の在り方について考える。

これまでの科研費研究の資料の中で、移管可能なものについては、関連した作業を行い、移管ができないものとはリンクをはるなどの工夫を行うことで、情報流通の促進を図る。

その他、日本語教育関連機関・団体との連携・協働を図りながら、既存及び新規の情報・データ、HP、サイトなどのコンテンツ（情報・素材等）の流通促進を図っていく。

3) 新たな情報資源

①日本語学びネット等の提供

新たな学習支援用基盤情報として、日本語学びネットを提供するとともに、漫画における表現意図データ（小学館発行の漫画の表現を機能別、場面別、話題転換等の方略別にデータ化したもの）などを提供する。

②日本語会話データベース（DB）の提供

OPI（Oral Proficiency Interview）テストを活用して、日本語学習者の会話データの収集・整理を行う。そして、発話資料を文字化してデータベース化するとともに、会話者（インフォーマント）の基本情報、評価結果（レベル）などの付加情報をつけて提供する。

4. 基盤情報の提供方法（案）：項目と内容

基盤情報については、以下の（1）の項目、（2）の内容が検索できるようなWebサイトとして公開・提供する予定である。

（1）項目

項目としては、例えば、研究用データ、研究成果・刊行物、研究動向、日本語教育の動向、学習と教育（ツール・素材等）、催し物などが想定される。

¹ 「日本語教育支援システム研究会」の通称（CASTEL/J: Computer Assisted System for TEaching & Learning / Japanese）のことで、コンピュータを利用した日本語教育に関心の高い人々が集まって、不定期に国際的な研究会を開催している。

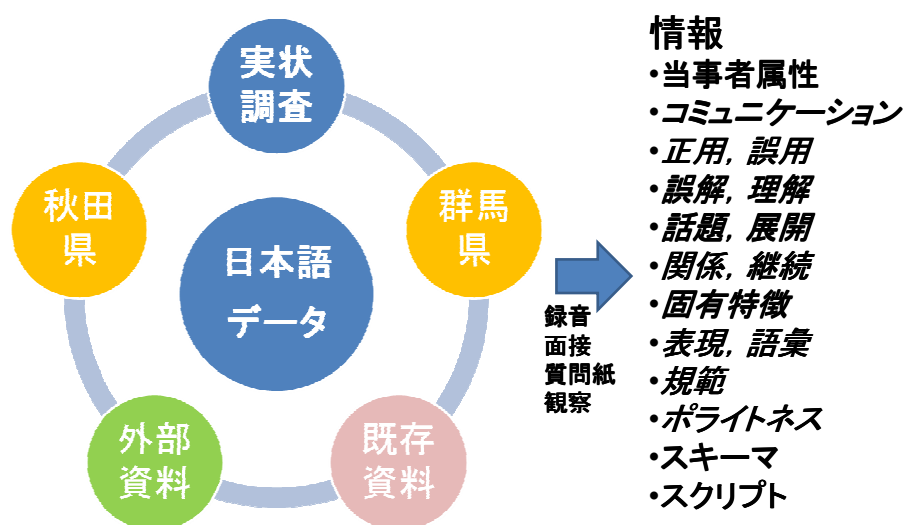
(2) 内容

項目ごとの内容としては、以下のものが考えられる。

- 1) 研究用データ：会話データ，作文データ，語彙データ，調査データなど
- 2) 研究成果・刊行物：報告書や逐次刊行物など
- 3) 研究動向：文献情報，科研研究課題情報，学位情報など
- 4) 日本語教育の動向：機関動向，政策・施策・報告書などの関連情報，関連記事など
- 5) 学習と教育（ツール・素材等）：ツール・素材，教材・教案など
- 6) 催し物：学会，研究会，シンポジウム，パネル開催日程，概要など

なお，この中で，特に中核的な基盤情報となる1) 研究用データに焦点を当てると，例えば，日本人対外国人に関する日本語のデータに関しては，以下のような観点から情報を収集・整備していく予定である。

日本人VS外国人



5. 日本語教育データベースの構築：今後の課題について

ここでは，基盤情報として提供されるデータの中で，特に中核的データの一つと考えている日本語会話DBを事例として取り上げ，そのデータ収集やDB構築の過程から見えてきた課題について考察したい。

(1) 会話データ収集の過程：これまでの経緯，現状，予定

平成18年度は，プレ調査として，90名分（約2700分）のデータを収集した。平成19年度は，約300名分（9000分）のデータを収集中である。その内訳は，以下の通りである。

1) 単年度調査

- ①日本語学校の学生：180名、
- ②大学・大学院などの学生：100名、

2) 3年間の縦断調査

- ①日本語学校の学生に対する縦断調査：20～24名（韓国語：12名・その他：12名）

なお、平成20年度は、約180～200名分のデータ収集を予定しており、合わせて、約500名のデータに達した時点で、第一段階の区切りとする。そして、会話データの整備・普及に、より力を注ぐこととする予定である。

3) 今後の課題

約500名の中の120名分は、音声、タグ情報付きデータとして提供し、特定目的に活用できる基盤情報としたい。そのため、今後、以下のようなデータ整備が課題となる。

*120名分のデータ整備

インフォーマント（被験者・学習者）の言語的背景としては、母語（第一言語）が、中国語、韓国語、英語、ポルトガル語の者で、それぞれ30名ずつを収集して文字化する。レベルごとの比率（人数）は、超級1（5名）：上級2（10名）：中級2（10名）：初級1（5名）で、合計30名とする。

全部のデータに、特定のタグ付けをする。具体的には、導入、突き上げ、ロールプレイ、終結部分などがわかるようなタグ情報を付ける。また、音声データとのリンクはりを行う。

(2) 秋田県能代市、群馬県大泉市での縦断調査

1) 調査の概要

①インフォーマントの背景と数

能代市では、中国、フィリピン、韓国、ロシアなどから来日した国際結婚の配偶者、日本語を第二言語として習得してきた高校生などが中心で、合計10～12名である。また、大泉町では、日系ブラジル人の高校生、親世代の成人が中心で、合計10～12名である。

②調査方法

OPIを活用した口頭会話テストと半構造化インタビューにより、会話データの収集と、言語生活に関する情報の収集を行っている。

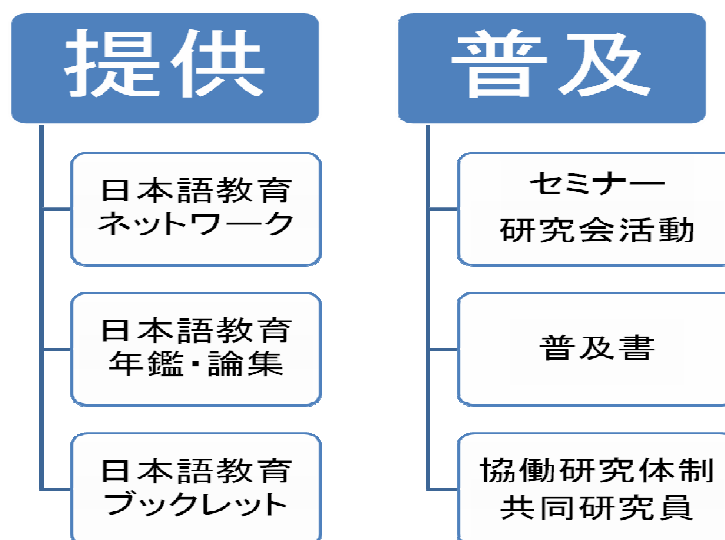
2) 今後の課題

インフォーマントとの接し方、インタビュー内容・方法、OPI そのものに関する理解促進、承諾書の署名をいただくタイミングなど、総合的な観点からのフィールド調査の在り方に関する確認・認識や、調査者・非調査者間の共通理解促進が肝要であり、課題である。

(3) データベース構築の課題

音声データの文字化の工夫を図っていくこと。また、日本語教育の基盤情報を、日本語教育DBを通じて提供する過程で、研究（開発）や素材・教材（開発）の基礎資料として活用されるこ

とを目指して、その流通を促進するための工夫・方法等についての検討を行うこと。また、基本的には以下のような枠組で、基盤情報の提供・普及に関する体制を作っていくことが肝要である。



6. 日本語教育データベースの構築：その可能性について

5で挙げた課題を解決し、収集・整備されたデータが活用されることによって、例えば、以下のような可能性が広がっていくことが期待されよう。

- (1) 日本語教育・学習内容・方法，評価の基準などの再構築に向けた基礎資料となること。
- (2) 日本語コミュニケーション・会話研究，日本語教師教育の新たな展開などに貢献すること。
- (3) 口頭能力テストの重要性や活用方法について喚起すること。
- (4) 日本語の会話や作文の評価の過程や，フィードバック等を通じて得られる情報の有意義性に関する認識の深化を促すこと。

また，アメリカ，イギリスなど，言語教育データベース構築の先進国の事例も参考としながら，情報の流通を促進していくことで，こうした可能性は多様な広がりを見せてくるであろう。

参考文献

国立国語研究所日本語教育基盤情報センター(2007a)『海外主要国におけるデータベースづくり及び自国語普及政策に関する調査研究 I－アメリカ合衆国，イギリス－報告書』(SL-ERG-06-01)

国立国語研究所日本語教育基盤情報センター(2007b)『海外主要国におけるデータベースづくり及び自国語普及政策に関する調査研究 II－韓国－報告書』(SL-ERG-06-02)