

国立国語研究所学術情報リポジトリ

はじめに

メタデータ	言語: jpn 出版者: 公開日: 2020-03-18 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00002686

はじめに

熊谷 康雄
(国立国語研究所)

1. プロジェクトの目的と本報告書

本報告書は国立国語研究所の共同研究プロジェクト(独創・発展型)「大規模方言データの多角的分析」(プロジェクトリーダー 熊谷康雄, 研究期間 平成21年10月～平成24年9月, 研究取りまとめ期間 平成24年10月～平成25年3月)の研究成果報告書である。

様々な研究分野において、基本的な情報の電子化、データベース化は重要なものとして推進されている。方言研究の分野においても、海外でも方言のデータベースが構築され、そのデータを用いた研究が行われている。しかし、膨大な量のデータが公開されても、使って初めて意味があるが、我が国の方言研究において、このようなデータを本格的に駆使した研究は、今後に期待するところが大きい。

我々は、かねてから、日本の方言研究の基盤的な資料である『日本言語地図』(調査期間1957～1965, 調査地点数2,400, 調査項目数285)や「各地方言収集緊急調査」(調査期間1977～1985, 全国約200地点)の資料の電子化・データベース化を進めてきた。これらは共に全国レベルの大規模な方言研究の資料であり、日本の方言研究における基礎的な資料である。本プロジェクトでは『日本言語地図』データベースや全国方言談話データベース(「各地方言収集緊急調査」より作成)などの資料・データの整備を研究基盤として進めるとともに、既に電子データが公開されている『方言文法全国地図』やその他の資料等の利用も視野に入れ、新たな情報も加えながら、計量的方言研究, 言語地理学, 日本語史, 談話研究など専門を異にする共同研究者が、共有する方言データをそれぞれの関心のある観点から実践的に使い込み、複数の視点から多角的に分析を行うことを目的とした。研究の基盤となる大規模方言データを整備, 共有し、これが持つ可能性を引き出す多角的な研究を通して、ことばの地域差の実態やその形成の解明に寄与する新たな知見の獲得, 研究方法の開発, 研究基盤となる資料・データの整備・共有や利用法の蓄積などを行い、方言研究の一層の発展に寄与することを目指した。

本報告書には、研究メンバーによる論文と『日本言語地図』データベースの概説および共同研究会発表会の記録を収録した。報告書の副題は「言語地図と方言談話資料」である。大規模方言データといっても様々な考えられる。このプロジェクトで扱うのは、この2種類のものであるという意味の副題であるが、研究メンバーによる研究は、言語地図に関わる研究, 方言談話資料による地域差に関わる研究, そして、言語地図と談話資料の両者を用いた研究という広がりで行われている。

本報告書の論文は、共同研究のメンバーにより、それぞれの視点から、この共同研究のテーマに取り組んだ成果の一部であるが、資料とこれを活用した研究という意味でも、多様な視点から迫っている。方言研究が資料に基づいて行われるのであれば、その資料そのものの性格を知り、また、その上でその資料を生かした分析が必要である。また、資料の

制約を知り、その上で、その制約の中で、資料の中からその資料の持つ可能性を引き出し、新たな研究の発展に結びつけていく観点と研究の実践が重要であると考えます。

本共同プロジェクトでは専門を異にする研究者が集まって、共有のデータを核にして、異なる視点からの分析が、共通のデータの上で交錯することにより、データの持つ可能性を掘り起こし、新たな知見、視野を得たいと考えた。本報告書に収められている共同研究のメンバーの論文は、それぞれに、このような側面を持つものと思う。

プロジェクトの期間が過ぎても、研究は終わるわけではない。分析は途に着いたばかりの研究もある。方言談話資料を用いた研究にしても、分析には多くの時間を要し、全国的な広がりの中で、対象地域を拡大していくにも、まだ、時間が必要と思われる。また、『日本言語地図』データベースも、このプロジェクトを通して、構築を加速し、利用可能な項目数を増やして、実質的な分析に手をつけることができるようになったが、完成までは、まだ、しばらくの時間がかかる段階である。今後の研究の継続によって、一層の分析の深化と発展が期待される。

2. プロジェクトの組織

以下の研究組織でプロジェクトを進めた。

・共同研究者

- 熊谷康雄（国立国語研究所時空間変異系・准教授）
- 井上文子（国立国語研究所時空間変異系・准教授）
- 大西拓一郎（国立国語研究所時空間変異系・教授）
- 沖 裕子（信州大学文学部・教授）
- 小林 隆（東北大学大学院文学研究科・教授）
- 澤木幹栄（信州大学文学部・教授）
- 澤村美幸（和歌山大学教育学部・講師） [2011年4月から]
- 高橋頭志（群馬県立女子大学文学部・教授）
- 日高 水穂（関西大学文学部・教授）
- 三井はるみ（国立国語研究所理論・構造研究系・助教）

構築を推進した『日本言語地図』データベース（LAJDB）の利用に関心の深い方に、共同研究を進めていく過程の中で、随時、研究協力者として加わっていただいた。

- 竹田晃子（国立国語研究所時空間変異研究系特任助教）
- 鎌水兼貴（国立国語研究所時空間変異研究系プロジェクト非常勤研究員）
- 吉田雅子

（所属は報告書刊行時点のもの）

3. データの共有

3. 1 共有データ：方言談話資料

方言談話資料の共有データとしては（1）『全国方言談話データベース』（国立国語研究所）、（2）『方言談話資料』（国立国語研究所）、（3）『方言録音資料シリーズ』

(国立国語研究所)を共有データとしてひとつのハードディスクに入れ、利用しやすい形にしてメンバーに提供し、共有することにした。初年度に共有データのハードディスクは共同研究者全員に配布した。

(1) 『全国方言談話データベース』全20巻

(1-1) 全20巻のCD収録の全音声ファイル (wav ファイル) を収録した。

(1-2) 全20巻の各巻別のディレクトリに全てのデータを収録した。

データの形式：ファイルメーカー、エクセル、PDF、音声ファイル

(1-3) 『全国方言談話データベース』統合版を新たに作成した。

『全国方言談話データベース』には各収録地点別にデータベースや文字化テキストが作成、収録してある。全国的な視野での分析を助けるために、全都道府県の統合版のデータベースを作成した。

(a) 『全国方言談話データベース』のオリジナルのファイルメーカー版の文字化テキストデータの全都道府県の全てを1つに統合したデータベースを作成した。

(DDJD_ALL 統合版 V1)

対訳形式のテキストファイルも作成

(b) 検索、分析のために、DDJD_ALL 統合版 V1 の文字化テキストデータの共通語訳テキスト、方言テキストのそれぞれについて、文節単位の KWIC を作成し、ファイルメーカー上で検索できるようにしたデータベースを作成した。

(ddjd_kwic_kyotugo_all および ddjd_kwic_hogen_all)

対訳形式の KWIC のテキストファイルも作成

(2) 『方言談話資料』 各巻別のデータを収録

01 山形・群馬・長野

06 鳥取・愛媛・宮崎・沖縄

02 奈良・高知・長崎

07 老年層と若年層との会話

03 青森・新潟・愛知

08 老年層と若年層との会話

04 福井・京都・島根

09 場面設定の対話

05 岩手・宮城・千葉・静岡

10 場面設定の対話

データの形式：文字化テキスト、音声ファイル (一部に未整備あり)

(3) 『方言録音資料シリーズ』 各巻別のデータを収録

01 鹿児島県鹿児島市

09 石川県志雄町

02 宮崎県都城市

10 愛知県小牧市

03 鹿児島県笠沙町

11 京都府京都市

04 岐阜県垂井町

12 沖縄・瀬底島

05 高知県高知市

13 静岡市旧大川村-1

06 秋田県男鹿市

14 静岡市旧大川村-2

07 鹿児島県上屋久町

15 沖縄県八重山鳩間島

08 高知県大方町

データの形式：文字化テキスト、音声ファイル (一部に未整備あり)

なお、上の（２），（３）はこれまでに国立国語研究所で HP 上に公開している。

3. 2 共有データ：『日本言語地図』データベース

『日本言語地図』データベース（LAJDB）については、本編を参照することとして、ここでは、データを共有した方法について記しておく。LAJDB の構築は、方言談話資料の共有データと異なり、LAJDB は研究期間を通して整備し、項目を増やしていった。画像データベースの容量も大きいため、当初は共有ハードディスクに最初の項目を入れ、追加項目分がある程度揃うと、DVD-R を作成し、メンバーに郵送する方法をとっていた。その後、この方法では手間と時間がかかり、また、データの追加にもラグが生じるため、ネット上のディスクスペースにアップして、メンバーで共有できる方法に変更した。画像データベースの容量は1項目で200MB ないし300MB 程度のもが多く、容量が大きいことが問題であったが、項目毎に出来たところで随時アップし、連絡するやり方で、効率的に運ぶことができるようになった。その後は、研究期間を通して、この方法によって随時最新の状態の LAJDB のデータを共有した。

3. 3 データの公開

研究成果ならびにデータベース（LAJDB）はホームページより公開する。公開の情報はプロジェクトのホームページ上に掲載する。

URL: <http://www.ninjal.ac.jp/research/project/b/daikibo/>

4. 今後に向けて

本プロジェクト期間を通して、当初の目的に向かって資料の整備、共有やその分析を前に進めることができた。開始当初に予想していた以上に、LAJDB のデータ整備などで、解決しなければならない困難があり、一定の段階に持ってくるまでに多くの時間を費やしたが、ようやく見通しが開けてきたところのように思われる。

プロジェクト期間を経て、研究は見通しが開けてきたところにある。今後、さらに深化、発展させるべき課題が多い。LAJDB の整備も続け、完成を目指していく。今後とも、研究メンバーの研究の進展と、さらに、データを公開することによって、多くの研究者による研究がなされ、蓄積されていくことを期待したい。プロジェクト開始当初にテーマを議論した問題意識を振り返り、今後さらに深化を図っていきたいと思う。