

言語生活調査の回答データセット Datasets on Sociolinguistics Survey in NINJAL

高田智和 鎌水兼貴

Tomokazu TAKADA Kanetaka YARIMIZU

国立国語研究所, 立川市緑町 10-2

National Institute for Japanese Language and Linguistics,

10-2, Midori-cho, Tachikawa-shi, Tokyo

概要: 言語使用の実態を捉えるため、国立国語研究所は1949年の創設以来、各地で面接法による言語生活調査を実施し、その都度報告書によって調査結果(回答の集計結果)を提供してきた。しかし、統計手法の高度化により、言語研究における統計解析向けのローデータのニーズが高まってきたため、過去の調査のローデータの整備・公開を進めている。現在4種の調査データ(共通語化、敬語運用、国語力観、世界各地での日本語観)を公開している。

Abstract: For the purpose of actual elucidation of word forms and language consciousness in contemporary Japanese, NINJAL has done the sociolinguistics surveys by interviewing for 70 years. We have provided survey results (summarized results of answers) by report. Recently, due to sophistication of statistical methods, the need for raw data for statistical analysis in language research has increased, so we decided to develop and publish raw data of the past surveys. Currently, we have released 4 datasets.

キーワード: 言語生活調査, ローデータ, 言語形式, 言語意識

Keywords: sociolinguistics survey, raw data, word form, language consciousness

1. はじめに

言語使用の実態を捉えるため、国立国語研究所(以下国語研)は1948年の創設以来、各地で面接法による言語生活調査を実施してきた。それまでの言語研究は、文献の書き言葉を対象とした史的研究に偏る傾向があったが、創設時の国語研は、言語とそれをを用いる人間と環境との関わりを重視し、同時代の言語形式、言語意識、言語行動を記述し、性別、世代、職業などによることばの違いや、変化の有り様とその背景を解明しようとする、現代話し言葉の研究を志向した。面接調査には、ランダムサンプリングの手法を用いて調査対象者を選定し、「科学的」とされる社会調査の手法を言語研究に導入した。

国語研の主な社会調査型言語生活調査には、次のようなものがある。

1. 八丈島調査(1949年実施、共通語化調査)
2. 白河調査(1949年実施、共通語化調査)
3. 第1回鶴岡調査(1950年実施、共通語化調査)
4. 伊賀上野調査(1952年実施、敬語調査)

5. 第1回岡崎調査(1953年実施、敬語調査)
6. 第1回北海道調査(1958-61年実施、共通語化調査)
7. 長岡調査(1962-63年実施、文字生活調査)
8. 松江調査(1963年実施、言語行動調査)
9. 第2回鶴岡調査(1971年実施、共通語化調査)
10. 第2回岡崎調査(1972年実施、敬語調査)
11. 大都市調査(1974-75年実施、言語行動調査)
12. 企業敬語調査(1975-77年実施)
13. 豊中・宮津・豊岡調査(1982-84年実施、言語行動調査)
14. 第2回北海道調査(1986-89年実施、共通語化調査)
15. 学校敬語調査(1989-91年実施)
16. 第3回鶴岡調査(1991-92年実施、共通語化調査)
17. 日本語観国際センサス(1997-98年実施、海外の日本語観に関する調査)
18. 「「外来語」言い換え提案」関連の調査(2003-04年実施)
19. 「国語力観」に関する全国調査(2006年実施)

- 20. 第3回岡崎調査(2008年実施、敬語に関する調査)
- 21. 「「病院の言葉」を分かりやすくする提案」関連の調査(2009年実施)
- 22. 第4回鶴岡調査(2011-12年実施、共通語化調査)

これらの調査の結果(回答の集計結果)は、その都度報告書によって、国民と学界に提供してきた。しかし、各種統計ソフトの利便性向上により、個人研究者が大規模調査データを解析する環境が整ってきた。また、統計手法の高度化により、言語データを対象とした文理融合研究も盛んになってきている。このような研究動向をふまえると、国語研が行ってきた従来型の集計結果提供方式では、学界のニーズに応えることができなくなると言わざるを得ない。そこで、過去の言語生活調査のローデータの整備・公開を進めることとした。

本稿では、現在公開している4種の調査データ(鶴岡調査データベース、岡崎敬語調査データベース、日本語観国際センサス、「国語力観」に関する全国調査)を解説するとともに、今後のデータ整備・公開について述べる。

2. 国立国語研究所研究資料室

国語研は、これまで日本語に関する様々な調査研究を行ってきたが、報告書や論文として公刊された研究成果の前段階の中間生成物に相当する情報カードや集計表、さらに、研究の一次資料に当たる調査票、録音、語彙調査の雑誌原本、調査研究運営の記録である調査計画書や会議録も現存し、現在それらは国語研研究資料室に保存して集中管理を行っている。上述の社会調査型言語生活調査の調査票も大部分が現存し、1974-75年実施の大都市調査以降は、調査時の録音が現存するものもある。これらの調査票や録音によって、ローデータの整備を進めている。

国語研研究資料室で保存している研究資料は、概ね1課題のもとで生産された研究資料を1資料群として管理し、現在、資料群は241である。資料群の概要記述は「国立国語研究所研究資料室収蔵資料」(<http://rnr.ninjal.ac.jp/>)として、2017年3月から一般公開を行っている(図1)。研究・教育が目的であれば、申請により研究資料を閲覧することができる。社会調査型言語生活調査の調査票も例外ではない。

また、録音音声や録画映像は、オープンリール、カセットテープ、8mmフィルム、ビデオテープなど、様々な記憶媒体で保存されている。しかし、記憶媒体が経年劣化を起こしたり、再生機器が生産中止になったりと、録音・録画の再生が難しくなってきたものがあるため、

保存と研究プロジェクトでの再利用のために、録音音声・録画映像のデジタル化を進めている。デジタル化した音声・映像は、所内限定利用の「所蔵音声・映像データベース」(図2)に蓄積している。現在、音声21,150点、映像697点であり、収蔵資料の約6割が媒体変換を完了している。「所蔵音声・映像データベース」も、研究・教育が目的であれば、申請により来館利用することができる。



図1 国立国語研究所研究室収蔵資料



図2 所蔵音声・映像データベース

3. 鶴岡調査データベース

鶴岡調査は、山形県鶴岡市において、共通語の浸透具合を測ったものである。1950年、1971年、1991-92年、2011-12年と、20年間隔で計4回の調査を実施した定点経年調査である。第1回調査が行われた1950年は、GHQ占領下で国語の民主化・合理化を強

く推進しようとしていた時期であり、当時の言語政策を強く反映した調査企画であったと考えられる。国語研創設翌年の言語生活調査である八丈島調査・白河調査も共通語化調査であり、喫緊の言語政策的課題であったことがうかがえる。

調査項目は、音声、アクセント、語彙、文法の言語形式に関する項目と社会生活項目であり、4回共通の項目も多い。言語形式に関する項目は、刺激図も用いた、いわゆるなぞなぞ形式で調査対象者から回答を引き出す調査である。

各回ともランダムサンプリングを行い(ランダムサンプリング調査)、集団の言語の推移を捉えることができる実時間データを得ている。調査対象者の人数は、1950年496名、1971年401名、1991-92年399名、2011-12年466名である。また、鶴岡調査では、第1回の調査対象者を、20年後の第2回でも追跡調査(パネル調査)を行い、同一個人の言語変化をとらえるデータも取得している。パネル調査の調査対象者は、第1回～第4回4名、第1回～第3回49名、第1回～第2回54名、第2回～第4回130名、第2回～第3回131名、第3回～第4回199名である。

図3は、ランダムサンプリング調査の音声31項目について、共通語回答を1点、非共通語回答を0点として集計し、調査対象者の生年区間ごとの平均値を図示したものである。世代ごとに、調査回ごとに、共通語化が進行していることがわかる。

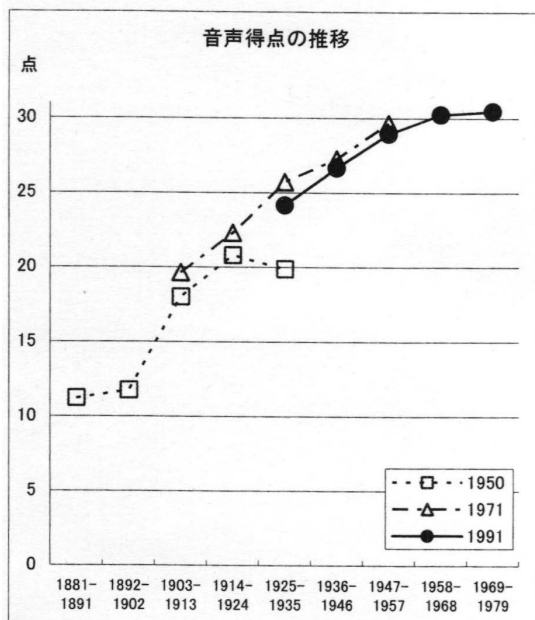


図3 音声得点の推移(文献[14]より)

鶴岡調査データベースは、第1回～第4回調査において継続して実施した調査項目に対し(個人の特定につながる項目を除く)、統一コードを設定して回答を数値化したものである。統計解析向けであることと、各調査回の結果を通覧できるようにするためのデータ設計である。下記 URL にて ver.2.0 を公開している。(データ形式:エクセルファイル)

鶴岡調査データベース

<http://www2.ninjal.ac.jp/longitudinal/tsuruoka.html>

ver.2.0 では、第1回～第3回調査の基礎項目(調査対象者の性別や生年など)と、音声・アクセント項目を収録している。また、ランダムサンプリング調査の回答データに加えて、パネル調査の回答データを増補した。今後、語彙・文法項目と第4回調査の回答データを追加し、鶴岡調査データベースの完成を目指す。

なお、第3回調査の録音音声は、音声資源コンソーシアムにて公開している。

鶴岡調査音声データベース 91-92

<http://research.nii.ac.jp/src/Tsuruoka91-92.html>

4. 岡崎敬語調査データベース

岡崎調査は、愛知県岡崎市において、敬語と敬語意識を測った定点経年調査である。1953年、1972年、2008年の計3回実施した。1952年に国語審議会建議「これからの敬語」が提出され、身分・階級による絶対敬語から、コミュニケーション手段として簡素化された相対敬語への転換が、言語政策として実施されることになった。これを受けて第1回調査が企画されたと考えられる。

鶴岡調査と同様の調査デザインで、ランダムサンプリング調査とパネル調査を行っている。ランダムサンプリング調査の調査対象者は、1953年434名、1972年400名、2008年306名である。また、パネル調査の調査対象者は、第1回～第3回20名、第1回～第2回165名、第2回～第3回62名である。

岡崎調査では、場面を描いた絵を調査対象者に見せて使用表現を尋ねる調査(反応文の取得)と、敬語に対する意識・意見・内省を尋ねる調査を行っている。

図4は、ランダムサンプリング調査の場面12項目の反応文について、「～でございます」など二つの敬語形式の結合による回答を1、「～です」「～ます」など一つの敬語形式の回答を2、「～だ」など敬語形式を含まない回答を3として集計し、調査対象者の生年区間ごとの平均値を図示したものである。3回とも、高年層は

平均値が高く、若年層は平均値が低くなっている。敬語の成人後習得が行われることがわかる。

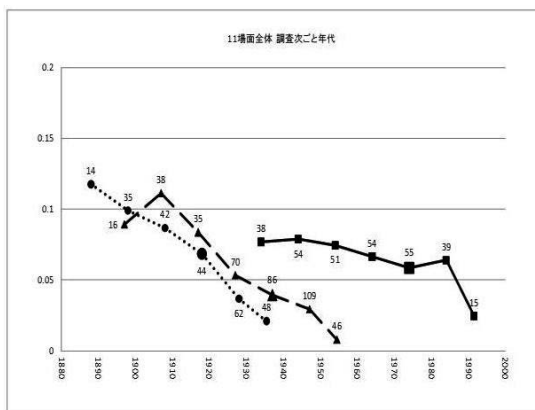


図4 二つの敬語形式の結合の推移(文献[17]より)

岡崎敬語調査データベースは、第1回～第3回調査において継続して実施した調査項目(個人の特定につながる項目を除く)を収録している。基礎項目、敬語意識項目、社会生活意識項目は回答を数値化する一方で、敬語行動項目(上述の場面12項目)は反応文をテキスト化している。データ利用者による自由な分析を可能にするためである。

岡崎敬語調査データベースは、下記URLにて公開している。(データ形式:エクセルファイル)

岡崎敬語調査データベース
<http://www2.ninjal.ac.jp/longitudinal/okazaki.html>

5. 日本語観国際センサス

日本語観国際センサスは、海外の日本語に対するイメージを把握し、国際社会における日本語の役割と位置の予測を目的とした調査研究である。日本の経済力が高かった頃1997-98年実施で、28の国と地域のそれぞれ1,000名の調査対象者をランダムサンプリングした調査である(中国と日本の調査対象者は3,000名)。調査対象者は29,411名である。また、28の国と地域は、アメリカ、ブラジル、アルゼンチン、韓国、オーストラリア、シンガポール、タイ、イギリス、ドイツ、オランダ、ハンガリー、イタリア、スペイン、ポルトガル、ロシア、イスラエル、インド、インドネシア、フィリピン、ベトナム、モンゴル、トルコ、ナイジェリア、エジプト、台湾、中国、日本である。

調査項目は、言語環境、母語意識、言語一般に対する意識、外国語意識、英語意識、日本語学習に対する意識、日本語意識、日本に対する意識、日本人に対

する意識、価値観の全59項目である。調査票は28の国と地域の言語で作成された。

図5は28の国と地域における日本語学習者の割合である。日本からの地理的距離が近いほど割合が高く、距離が遠いほど割合が低くなっている。また、日本との歴史的関わりとの相関も想定される。文献[10]では「地理的隣接効果で説明できる」とする。

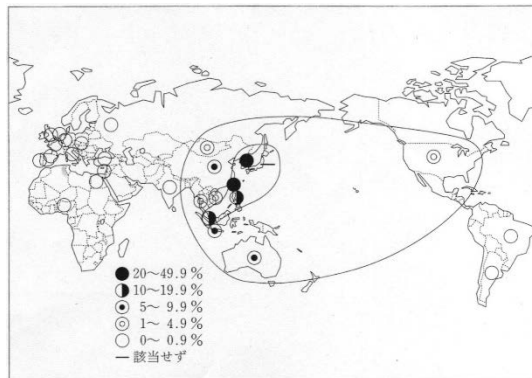


図5 日本語を習った人の割合(文献[10]より)

日本語観国際センサスの回答データセットは、回答を統一コードによって数値化している。公開URLは下記である。(データ形式:エクセルファイル、TAB区切りテキストファイル)

日本語観国際センサス
https://mmsrv.ninjal.ac.jp/n_census/

6. 「国語力観」に関する全国調査

「国語力観」に関する全国調査は、2004年の文化審議会答申「これからの時代に求められる国語力について」を踏まえ、「国語力」を「言語生活力」ととらえて、読み書き能力(いわゆるリテラシー)にとどまらない幅広い言語運用能力の実態解明を目的とした調査研究として、2006年2月と8月に実施した。何を「国語力」と見なすかは、人により立場により考え方に相違があるため、「国語力」をどのようにとらえているか(国民の「国語力観」)を探るための意識調査である。調査対象者は、ランダムサンプリングにより、2008年2月の調査対象者は2,134名、8月は2,129名である。

2006年2月の調査項目は、国語力観、言語活動、国語観、8月の調査項目は、国語力観、国語力の自己評価、「国語力がある人の人物像」など国語力の諸側面である。

図6は、2006年8月調査の「国語力がある人の人物像」の集計結果である。「読む」「書く」「聞く」「話す」の言語能力では、「読む」「書く」の文字言語・書きことば

に対する能力の方が、「聞く」「話す」の音声言語・話しことばに対する能力よりも高い割合である。また、ことばに関する知識では、「漢字」「仮名遣い」の方が、「語句」「慣用句」よりも高い割合である。

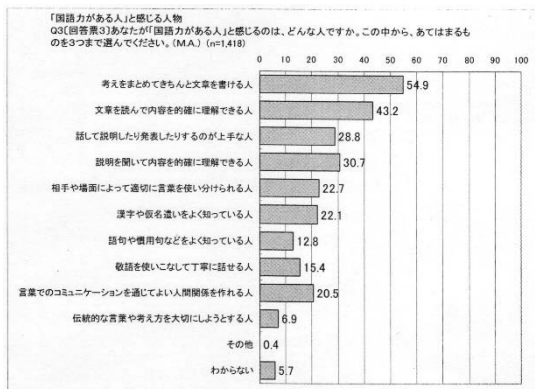


図6 「国語力がある人」と感じる人物(文献[16]より)

「国語力観」に関する全国調査の回答データセットは、回答を統一コードによって数値化している。公開URLは下記である。(データ形式: エクセルファイル、TAB区切りテキストファイル)

「国語力観」に関する全国調査
https://mmsrv.ninjal.ac.jp/krk_survey/

7. おわりに

現在公開している4種の調査データについて述べた。国立国語研究所は他にも多種多様な言語生活調査を実施し、調査票や録音音声、その電子化データを保存している。今後も調査データの整備・公開を進めていく予定である。なお、現在、整備を進めている調査データは以下のものである。

〔北海道調査〕

北海道移住者1世から3世の親子三代を対象に、共通語化の過程をとらえようとした調査。第1回調査は1958～1961年、第2回調査は1986～1988年に実施。鶴岡調査や岡崎調査と同様の経年調査である。報告書: 文献[3][6]

〔外来語に関する意識調査(全国調査)〕

外来語を用いるコミュニケーション場面や、外来語に対する意識を明らかにしようとした調査。公共場面での分かりにくい外来語を分かりやすく伝える言葉遣いの工夫の提案(「外来語」言い換え提案、2002～2006年)を

行うための基礎調査でもある。調査は2004年に実施。有効回収数3,090名。

報告書: 文献[11][13]

〔医療者に対する用語意識調査〕

患者に医療用語を理解してもらう必要性や困難度を把握するため、医療従事者に対して行った意識調査。患者と医療従事者との医療用語の伝達を分かりやすくするための提案(「病院の言葉」を分かりやすくする提案、2009年)の基礎調査でもある。調査は2008年に実施。有効回収数1,652名。

関連書籍: 文献[15]

〔非医療者に対する理解度等の調査〕

医療用語について、非医療者の認知度と理解度を把握するため、非医療者に対して行った意識調査。「病院の言葉」を分かりやすくする提案の基礎調査でもある。調査は2008年に実施。有効回収数4,276名。

関連書籍: 文献[15]

参考文献

- [1] 国立国語研究所編, 地域社会の言語生活—鶴岡における実態調査—, 秀英出版, 1953.
- [2] 国立国語研究所編, 敬語と敬語意識, 秀英出版, 1957.
- [3] 国立国語研究所編, 共通語化の過程—北海道における親子三代のことば—, 秀英出版, 1965.
- [4] 国立国語研究所編, 地域社会の言語生活—鶴岡における20年前との比較—, 秀英出版, 1974.
- [5] 国立国語研究所編, 敬語と敬語意識—岡崎における20年前との比較—, 三省堂, 1983.
- [6] 国立国語研究所編, 北海道における共通語化と言語生活の実態(中間報告), 国立国語研究所, 1997.
- [7] 新プロ「日本語」総括班・研究班1編, 日本語観国際センサス単純集計表(暫定速報版), 国立国語研究所, 1999.
- [8] 林知己夫・鈴木達三・江川 清・米田正人, 日本語観国際センサス—28カ国調査の実施—, 第68回日本統計学会講演報告集2000, 2000, pp.75-76
- [9] 林知己夫・鈴木達三・江川 清・米田正人, 日本語観国際センサス—28カ国調査の2,3の結果—, 第68回日本統計学会講演報告集2000, 2000, pp.77-78
- [10] 井上史雄, 日本語の値段, 大修館書店, 2000.
- [11] 国立国語研究所編, 外来語に関する意識調査, 国立国語研究所, 2004.

- [12] 国立国語研究所編,「国語力観」に関する全国調査〔平成18年2月,8月実施〕,国立国語研究所,2006.
- [13] 国立国語研究所編,公共媒体の外来語―「外来語」言い換え提案を支える調査研究―,国立国語研究所,2007
- [14] 国立国語研究所編,地域社会の言語生活―鶴岡における20年間隔3回の継続調査―,国立国語研究所,2007.
- [15] 国立国語研究所「病院の言葉」委員会編,病院の言葉を分かりやすく―工夫の提案―,勁草書房,2009
- [16] 三井はるみ,「日本語の達人」像から見る「国語力観」の多様性,「国語力観」に関する全国調査 研究発表と分析,2009, pp.58-75.
- [17] 杉戸清樹編,敬語と敬語意識―愛知県岡崎市における第三次調査― 第1分冊【経年調査実施資料編】,国立国語研究所,2010.
- [18] 井上史雄・阿部貴人・鎌水兼貴・柳村裕・丁美貞,敬語表現の成人後採用―岡崎における半世紀の変化―,国立国語研究所,2016.
- [19] 鎌水兼貴,愛知県岡崎市における敬語行動の経年調査, JLVC 2016 発表資料
- [20] 高田智和・大石恵輔・山口亮・石本祐一,国立国語研究所収蔵音源資料と所蔵音源データベース構築,じんもんこん 2017 論文集,2017, pp.259-264.
- [21] 関川雅彦・山口亮,研究資料室中央資料庫収蔵資料の公開に向けての取り組みと課題,国立国語研究所論集,2018, 第14号, pp.231-239.
- [22] 石本祐一・生永匠,国立国語研究所所蔵映像資料のデジタル化と所蔵映像データベース構築,じんもんこん 2018 論文集,2018, pp.217-222.
- [23] 鎌水兼貴,「日本語観国際センサス」データの公開と利用,じんもんこん 2018 論文集,2018, pp.181-186.