国立国語研究所学術情報リポジトリ

New Perspectives Enabled by the Linguistic Atlas of Japan Database

メタデータ	言語: jpn 出版者: 公開日: 2015-10-30 キーワード (Ja): キーワード (En): 作成者: 熊谷, 康雄, KUMAGAI, Yasuo メールアドレス: 所属:
URL	https://doi.org/10.15084/00000725

国語研プロジェクトレビュー
NINJAL Project Review　Vol.4 No.1 pp.1–9（June 2013）

〈共同研究プロジェクト紹介〉
独創・発展型：大規模方言データの多角的分析

『日本言語地図』のデータベース化が開く新たな研究

New Perspectives Enabled by the Linguistic Atlas of Japan Database

熊谷 康雄（KUMAGAI Yasuo）

1. 全国の方言を一望できる『日本言語地図』

国立国語研究所編『日本言語地図』（全 6 巻）は 1966 年（昭和 41 年）から 1974 年（昭和 49 年）にかけて刊行された。調査地点数 2400，収録地図 300 枚の言語地図である。調査は 1957 年（昭和 32 年）から 1965 年（昭和 40 年）にかけて臨地調査により行われた。言語地図（方言地図とも言う）とは，どこでどのような言葉が使われているかを地図上に表したものである。図 1 は『日本言語地図』（Linguistic Atlas of Japan, 略称 LAJ）所載の地図の例である。回答された語形を記号で表し，それぞれ当該の調査地点の位置にその記号が押されている。

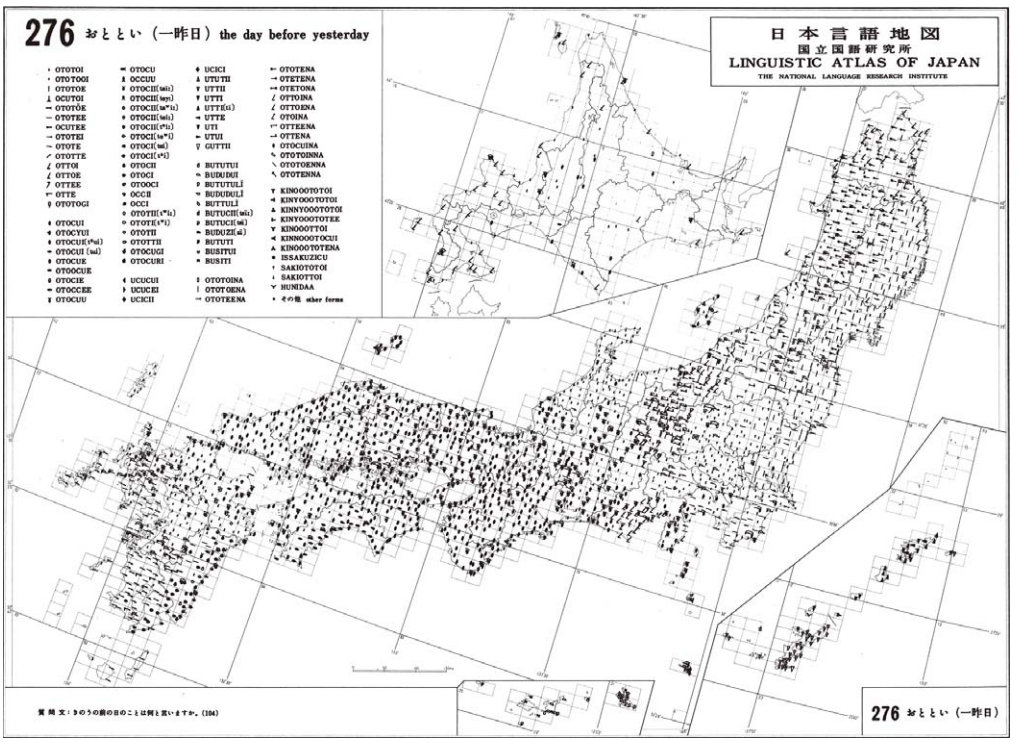


図 1 『日本言語地図』第 6 集　276 図（国立国語研究所 1974）

日本全国のどこでどのような語形が使われているのか，一目で見渡せる。

『日本言語地図』は本格的な言語地理学的調査による日本で最初の全国的な言語地図である。また，語彙に関しては，言語地理学的調査により作成された全国的な規模の言語地図は現在でもほとんど『日本言語地図』しかない。さらに，『日本言語地図』の調査の行われた昭和 30 年代の時期は，テレビの普及が始まり，進みつつあったころであり，交通網も今のように発達する以前の時代である。『日本言語地図』は，そのような，昭和 30 年代の日本の方言の全国的な分布の姿を捉えた，方言研究の基礎的資料である。

2. 『日本言語地図』：地図の編集は紙と手

『日本言語地図』が刊行されたのは，1966 年から 1974 年である。当時はまだコンピュータが一般的に使われるようになる以前であり，編集作業はすべて紙と手で行われた。多色刷りの地図の作成，印刷もすべてコンピュータは介在していない。日本全国の地方調査員のフィールド調査の結果は各調査地点の 1 項目ごとに紙のカードに書き写され，国立国語研究所の調査本部に寄せられた。このカードが『日本言語地図』の原資料である。調査項目数 285，カードの枚数は合計 54 万枚になる。図 2，図 3 にカードケースとカードケースに収め

図 2　原カードを収めたカードケース

図 3　カードケース中の原カード

図 4　白地図にスタンプを押す

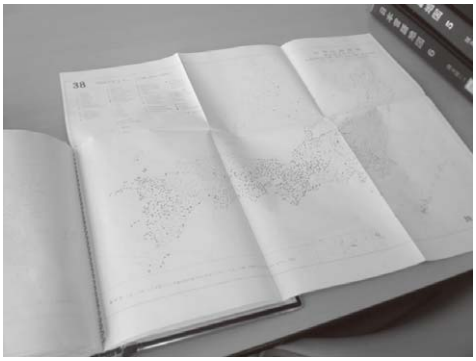


図 5　バインダーから広げた地図

られた原カードを示す。

『日本言語地図』の編集はこのカードをもとに行われた。『日本言語地図』では，どの調査地点でどのような言葉が回答されたのかが地図上に描かれているが，回答をそのまま地図に描くのではなく，回答を分類整理，統合して凡例語形のもとにまとめた上で，語形に記号を与え，それぞれの語形を回答した地点の位置に記号をスタンプして地図を描いている。原資料の 54 万枚のカードを相手に，このような編集，作図が手作業で行われた。図 4 は，ゴム印のスタンプを白地図に押しているところである。このように，『日本言語地図』は手作業で編集，作図され，印刷刊行された。なお，『日本言語地図』の調査と編集については，『日本言語地図』第 1 集の『解説』（国立国語研究所 1966）に詳細が記載されている。

3. 『日本言語地図』をデータベース化する

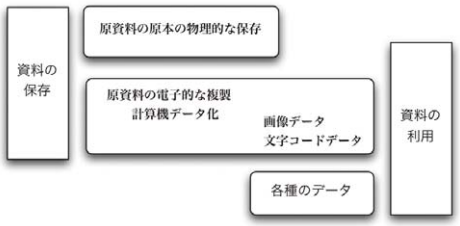


図 6 『日本言語地図』データベース（LAJDB）

図 7 原カードのスキャン作業風景

図 8 LAJDB 画像データベース：画面例

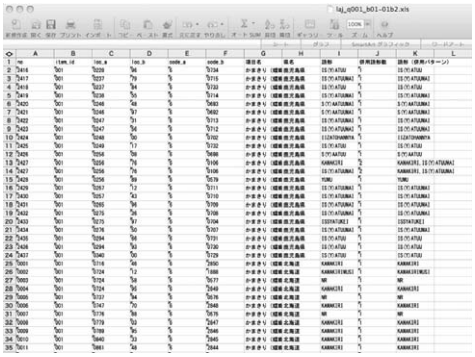


図 9 LAJDB 文字データ（Excel 形式 2）

『日本言語地図』は多くの研究に使われてきたが，既に上でも述べたように，コンピュータの利用が一般化する以前に刊行されたものであった。本堂（1980），河西（1981）そして井上（2001）にまとめられた研究をはじめ方言の全国的な分布を計量的に扱う研究の資料としてさまざまに使われてきているが，これまでになされた研究は，ほとんどの場合，個々の

研究者がそれぞれの研究目的に応じて，紙の印刷物である『日本言語地図』から必要な部分を目で見て直接数えて集計したり，必要な範囲をそれぞれが必要な形で個別のデータを作成したりして分析している。

電子化されたデータがなく，紙の資料からそれぞれが必要な範囲でデータを作成して分析するという状況が持つ制約は，研究方法にも影響を与えていると考えられる。例えば，計量的な分析に関しては，全国的な分析では県を集計の単位としたものが多く，地点を単位として集計したものは地域が限定されていることなどにも影響していると思われる。

『日本言語地図』の作図は，上に述べたように手作業なので，例えば，1 地点で複数の回答があった場合やスタンプした記号が重なるときなど，記号をスタンプする位置をずらして地図を見やすくするなどの調整もされているが，このような人手ならではの側面は，地図上の記号を読み取りやすくするという面がある一方，機械的な側面からは地図上での記号の位置がずれることにより，記号に対応する調査地点の位置の読み取りが容易でなくなる場合も生んでいる。地点数の多さと手作業で作成された言語地図という表現媒体であることにより，印刷された『日本言語地図』をもとにデータを作成する作業は単純ではない。19 都道府県で標準語形の分布の数を数えた本堂（1980）も，県境にスタンプされた記号がどちらの県に属するのか不明である場合や，多数の地点で，手作業で語形を一つ一つチェックする困難さに触れている。全国規模の等質的な方言分布データとして『日本言語地図』の持つ可能性を発揮させるには，『日本言語地図』の全体に渡る一貫したデータ化が必要である。

このような『日本言語地図』の資料の保存と利用の高度化を目指して，『日本言語地図』データベース（Linguistic Atlas of Japan Database，略称 LAJDB）の構築を始めた（図 6）。1999 年頃より本格的にデータベースの構築に取り組み始め，データベース科研の補助を受けて構築を進めてきた（熊谷 2007）。LAJDB には原カード画像にアクセスできる画像データベース（図 8）と，語形の分布情報の文字データ（図 9）のデータベースがあり，『日本言語地図』の原資料と地図情報のすべてが電子化される。この 3 年間の共同研究プロジェクト「大規模方言データの多角的研究」の期間では，プロジェクトの柱の一つとして，LAJDB の整備と分析を進めてきた。LAJDB の構築では，既に全項目数の 9 割近くまでカードをスキャンし画像データ化（図 7）を進めていたが，このプロジェクトの 3 年間では，特に，既に原カードの画像化の済んでいる項目をデータベースとして使えるようにするための作業の部分に注力し，利用可能な項目を増やすことに努力した。『日本言語地図』として地図化されている 240 項目のほぼ半数である 119 項目を公開できる段階に来たところである。なお，公開に関する情報はプロジェクトのホームページ（http://www.ninjal.ac.jp/research/project/b/daikibo/）に掲載する。

4.　標準語形使用数の分布を「高解像度（2400 地点）」で観察する

『日本言語地図』を使った計量的な研究では，県別の標準語の使用率のデータを使った井上史雄氏の研究（井上 2001 他）が代表的なものとしてある。この県別標準語使用率のデータは，河西秀早子氏が『日本言語地図』の中から選んだ 82 項目をもとに作成したデータ（河西 1981）で，標準語形の分布率の平均値を都道府県別に計算したものである。河西（1981）は，

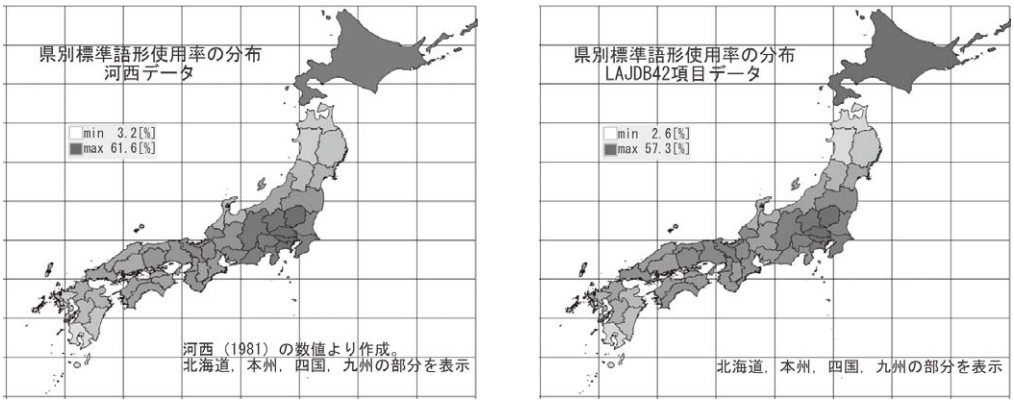


図 10　標準語形県別使用率の分布（河西データと LAJDB 42 項目）

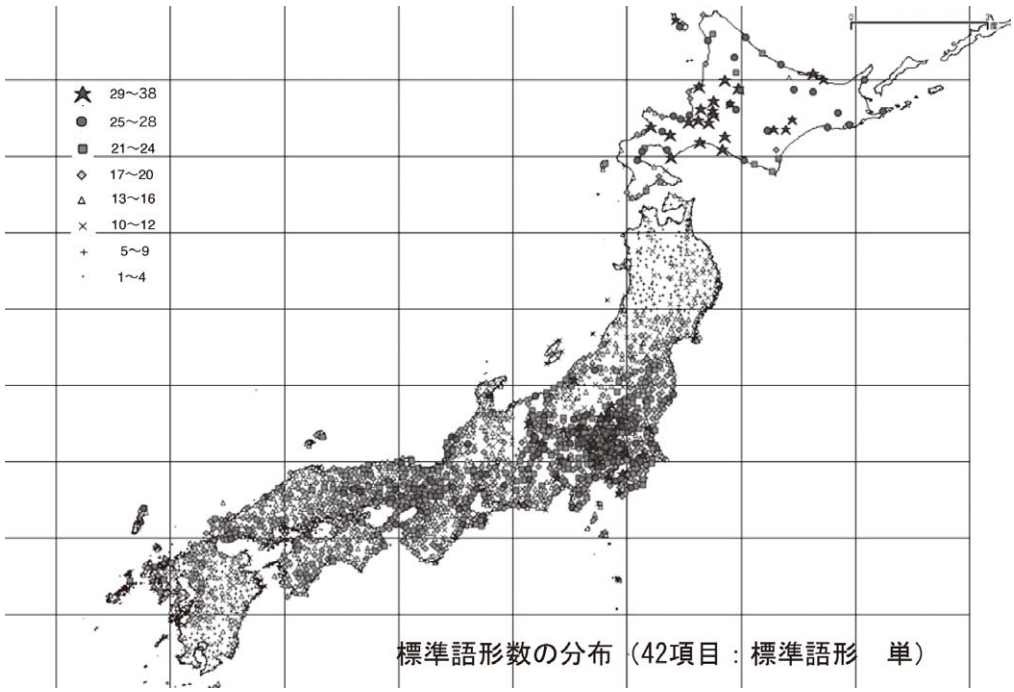


図 11　標準語形数の分布

日本全国に標準語が分布している模様を各県毎の標準語使用率の地図として示した。関東を中心に使用率が高く，（北海道を除く）周辺地域に向かうにつれて低くなる，まさに周圏的な分布の様相を示すものとして概括されている（河西 1981）。図 10（左）には，この河西（1981）の表に示された数値をもとに，作成しなおした地図を示す。

さて，河西データでは 47 都道府県の枠を通して全国的な分布の模様を捉えている。一方，『日本言語地図』データベース（LAJDB）では，2400 地点のデータが使える。これまで，県

別でしか見ることのできなかった，全国の分布を見るとどのような姿が現れるだろうか。

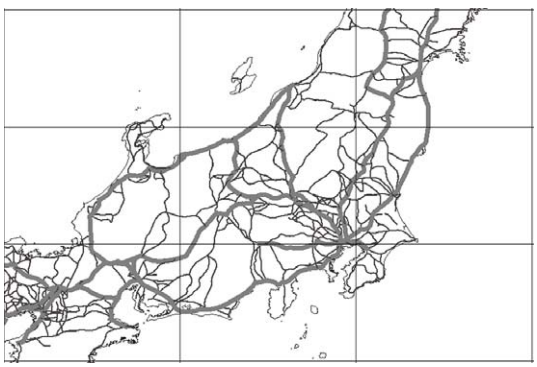
LAJDB はまだ構築途上であるが，この分析を行った時点で利用可能だった項目中から，調査地点数が 2400 ある 55 項目を選び，さらに『日本言語地図』の解説で標準語形が一つに特定されている 42 項目を選んだ（熊谷 2013）。この 42 項目の標準語形の地点別の数を集計して地図を作成したものが図 11 である。

河西データによる集計と比較するために，この 42 項目について，河西（1981）と同じように県別使用率を計算して，地図化したものが図 10（右）である。今回の 42 項目と河西 82 項目とで共通する項目は，27 項目であるが，非常によく似た全体的なパターンを見ることができる。今回の 42 項目の結果が河西（1981）の結果からそう外れてはいないことが分かると同時に，両者のデータが捉えているパターンが『日本言語地図』のデータ全体の中の安定したものであることを示唆していると考えられる。

図 11 では，これまで県別データとして 47 都道府県の枠を通して目にしてきた分布が 2400 の地点別の集計の分布として見ることができる。その分布の模様は，図 10（右）で確認したように，大括りにはこれまでの河西データによる傾向を確認できるとともに，さらに詳細な観察が可能になる。都道府県の枠を外した姿は，都道府県の枠の中で分布が分散してぼやけるというようなことはなく，全体として明瞭な地理的な分布パターンが観察できる。

LAJDB のデータでは，『日本言語地図』の 2400 地点の細かさで捉えることができる。このことは，47 都道府県の枠と比べると，言わば 51 倍の「高解像度」で分布事象を捉えていることになる。県別の枠では得られないこの解像度では何が見えてくるだろうか。

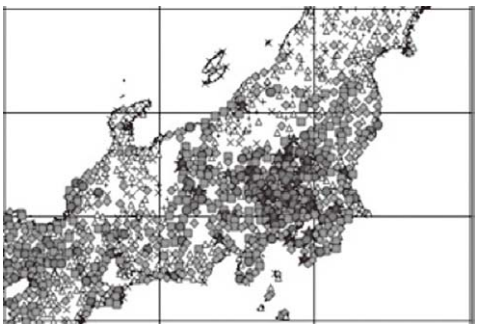
言語地理学では，言葉は地理的に伝播し，言葉の地理的な分布はこの伝播によって形成されたものと捉える。この伝播の過程で，重要な役割を果たす道路とこの標準語形の使用数の分布との関係はどのようになっているだろうか。図 12 は『日本言語地図』参考図 V の「近代道路図」より作成した電子地図の部分である。太線は国道，細線は県道である。この道路図と標準語形数の分布（図 11）を重ね合わせて観察する（図 13）と，道路網と標準語の分布の興味深い重なりを観察することができる。



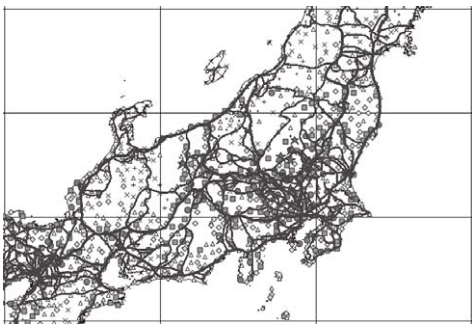
太線：国道，細線：県道

図 12　近代道路図（部分）（『日本言語地図』参考図 V「近代道路図」より作成）

例えば，関東から近畿にかけての範囲を見てみると，標準語形数の多い地点の連なりが示す道筋は，東海道や，そして，それ以上に中山道が大きな度数を示す地点の連なりではっきり浮かび上がっている様子が観察できる。図 13 の（1）と（2）を見比べると，標準語形数の分布にこれらの街道がよく重なっている様子を見ることができる。また，脇街道など，より仔細に追うこともできる。街道筋のネットワークに沿った分析も可能性がある。脇街道の事例を一つあげれば，中山道の脇往還である三州街道（塩尻と岡崎を結ぶ。別名伊那街道。江戸時代は中馬の通商の路。）は中山道と同等あるいはそれ以上に大きな度数を示す地点の連なりで浮かび上がっている様子が観察できる。LAJDB により，県別の枠では不可能だった伝播のルートや過程をより詳細に追跡する可能性が現実的なものとなってくる。



（1）標準語形数の分布（図 11 の部分）



（2）標準語形数分布上への「近代道路図」の重ね合わせ（道路網の線幅を調整）

図 13　標準語形数の分布（図 11 の部分表示）と近代道路図の比較

5.　おわりに：『日本言語地図』のデータベース化で広がる研究

『日本言語地図』データベース（LAJDB）により,『日本言語地図』2400 地点の項目別のデータが使え，また，注記や原表記が記載されている原資料であるカードに簡単にアクセスできる。このような LAJDB を利用した研究には，これまでには見えなかったより詳細な現象の観察，新たな観点による分析，これまでになされて来た研究の再評価や位置付けなど，さまざまな可能性がある。

LAJ2400 地点のデータの描く詳細な模様は県別のデータでは得られなかったより詳しい観察を可能にする。伝搬経路の追跡，語の伝播の分析にもつながる。言語地図上では調査地点の回答から面としての広がりを持つ語形の分布パターンが観察されるが，これは地点間の接触（交通網，人的交流等）を背景に形成されたものである。LAJ の 2400 地点の細かさで，多数の項目を扱うことで，逆に，その面的な分布の背後にある交通網などネットワーク的な関係，その関係の上での語の伝播の様相が全国的な広がりの中で浮かび上がってくる。

LAJDB は構築途上にあり，本稿に示したものも暫定的な段階ではあるが，これまでは見ることのできなかった，あるいは，ぼんやりと見えていた事象に対して，新たな視界が開けてきた。データの整備と分析は研究を進める上で，車の両輪である。今後も LAJDB の整備

を継続し，詳細な分析につなげたい。

なお，LAJDB には科学研究費研究成果公開促進費（データベース）〔平成 13, 14, 15, 16, 17, 20 年，『日本言語地図』データベース（研究代表者：熊谷康雄）〕の補助を得た。地図の描画には QGIS 1.6.0 を利用した。本稿は，共同研究の成果報告書にまとめた熊谷（2013）の内容の一部をまとめ直したものである。

●参照文献●

本堂寛（1980）「現代日本標準語の分布―『日本言語地図』で見て―」佐藤茂教授退官記念論集刊行会（編）『佐藤茂教授退官記念　論集国語学』479–498．東京：桜楓社．

井上史雄（2001）『計量的方言区画論』東京：明治書院．

河西秀早子（1981）「標準語形の全国分布」『言語生活』354: 52–55．

国立国語研究所（1966–1974）『日本言語地図』（全 6 巻）．東京：大蔵省印刷局．（縮刷版 1981–1985）．

熊谷康雄（2007）「『日本言語地図』のデータベース化」『日本方言研究会第 85 回研究発表会発表原稿集』27–34．

熊谷康雄（2013）「『日本言語地図』のデータベース化と計量的分析―併用現象，標準語形の分布と交通網，方言類似度の観察―」『大規模方言データの多角的分析　成果報告書―言語地図と方言談話資料―』111–128．

《要旨》『日本言語地図』のデータベース化（『日本言語地図』データベース，LAJDB）の概略を説明し，3 年間の本プロジェクト期間中に整備を進め，利用可能となった項目（119 項目）の一部を利用した計量的な分析の事例として，標準語形の使用数の地理的な分布を示した。これにより，『日本言語地図』がデータベース化されることの意味とこれが生み出す新しい研究の広がりの一端に触れた。

Abstract: The *Linguistic Atlas of Japan*（LAJ, 6 vols.）was published from 1966 to 1974 by the National Language Research Institute, the present NINJAL. In 1999, we started constructing the *Linguistic Atlas of Japan* Database（LAJDB）. In this three-year project, we accelerated the construction process. Various quantitative studies have been done on LAJ since its publication, but the lack of data in digital form has restricted the development of such quantitative research. As a case study of quantitative analysis using LAJDB, we made some observations on the geographical distribution of the frequency（GDF）of standard forms. The GDF of standard forms for each locality showed clear patterns which could not be obtained from the prefecture-unit calculations of former studies. Such “high resolution” patterns obtained from LAJDB enable us to trace diffusion routes more precisely. We compared these geographical distributions with the network of roads and so on, and found interesting relationships among them.

熊谷 康雄（くまがい・やすお）

国立国語研究所時空間変異研究系准教授。文学修士（埼玉大学）。国立国語研究所情報資料研究部主任研究官，室長，情報資料部門部門長を経て，2009 年 10 月より現職。

主な著書・論文：『シリーズ方言学 1 方言の形成』（共著，岩波書店，2008），「計量的方言区画と方言地理学―計量的方言区画のためのネットワーク法の開発を通して―」（『方言地理学の課題』，明治書院，2002），「方言区画の方法における地理的な次元」（『言語学林 1995−1996』三省堂，1996），The S&K network method: Processing procedures for dividing dialect areas（共著, *Zeitschrift für Dialectologie und Linguistik* 74, 1993），「言語的特徴による地域分割のための「ネットワーク法」―特に NT-1（r）について―」（共著，『国語学』140，1985）．

独創・発展型共同研究プロジェクト「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
（国立国語研究所　時空間変異研究系　准教授）

プロジェクトの概要

研究の基盤となる大規模方言データの整備，ならびに，データが持つ可能性を引き出す多角的な研究を通して，言葉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，研究方法の開発を目指している。『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）などのデータ整備を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資料の利用や，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複数の視点から多角的に分析する。