

# 方言分布・言語地図データベース — 時空間情報を持つ言語データ —

## Database on Dialectal Distributions and Dialectal Maps: Linguistic Data Including Spatial and Temporal Information

大西拓一郎

Takuichiro Onishi

国立国語研究所 言語変化研究領域, 東京都立川市緑町 10-2

National Institute for Japanese Language, 10-2, Midori-cho, Tachikawa, Tokyo

**概要:** 方言は、土地によることばの異なりである。ことばとしての本質において、標準語との差異はない。ただし、方言である以上、使われる場所についての情報は欠かすことができない。したがって、方言情報には通常の言語情報に加えて、空間情報が必須である。空間情報を緻密にすると、分布が求められ、地図として表現が可能になる。日本の方言学は分布や地図の情報を豊富に持ち、われわれはそのデータベース化を進めている。その現状と展開を報告する。

**Abstract:** Dialects are variations of language decided by space. There are no differences between dialects and standard Japanese in the nature of language; however, dialectal data require spatial information besides linguistic information. Adding enough spatial information on dialectal data clarifies dialectal distributions and makes it possible to express them on maps. Japanese dialectology has rich data of distributions and maps allowing for the creation of a database of these dialectal distributions. In this paper, I report the present and future works with regard to the database.

**キーワード:** 方言、方言分布、方言地図、GIS

**Keywords:** dialect, dialectal distribution, dialectal map, GIS

### 1. はじめに

方言分布と方言地図のデータベースの公開を開始してから、すでに約 15 年になる。その概要(沿革・現状)を報告し、諸分野との連携を展望する。

### 2. 方言地図とは

方言は、歴史的な同系統言語の場所による異なりであり、言語としての本質は標準語と異ならない。そのような方言がどのように存在しているのか、つまり分布を視覚化し、地図として表現したのが方言地図である(図 1)。方言地図は、言語地図と呼ばれることもある。

方言地図は、方言という特定の対象を表現することから、主題図であり、地形図のような一般図とは区別される。したがって、方言地図において、どのような表現方法を採用するかには、地形図の地図記号のような一般図はない。

分布の表し方は、語形を地図上に書き込む語形記入法、分布域を色や塗りつぶし方で表現する塗りつぶし法、語形を記号(シンボル)に置き換えて示す記号法の 3 方法がある。語形記入法は、世界最初期の『フランス言語図巻』が採用したが、煩瑣で分布が把握しづらく、地図としての効果が低い。塗りつぶし法は対象

となる方言形式の種類が多いと限界があるとともに複雑に入り組んだ分布に対応しきれない。多くの方言地図は、語形や分布の多様性に対応可能な記号法を採用している。地図要素の一つであるポイントで扱う記号法は、対応するデータの性質(各地点の話者)との親和性も高い。

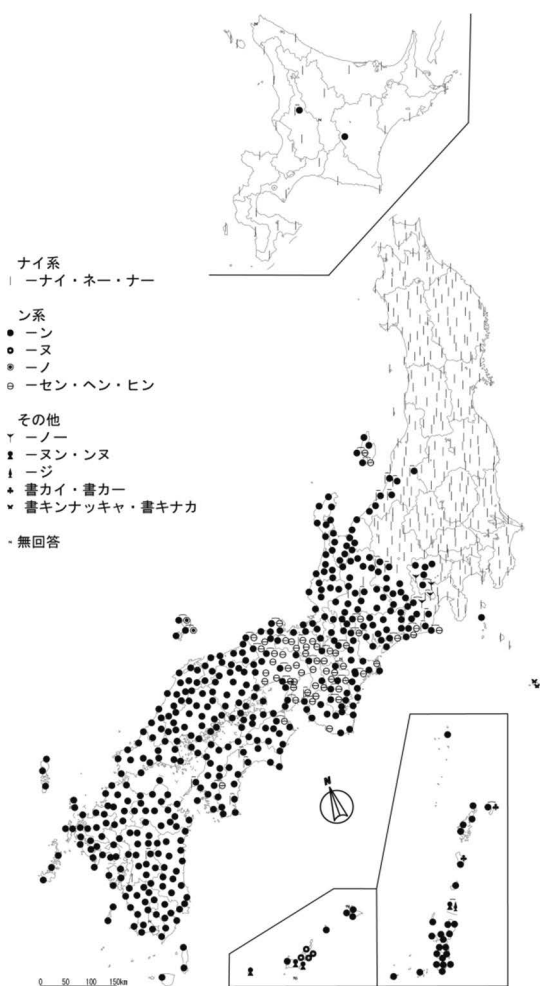


図1 書かない(GAJ2-80に基づく)

### 3. データ収集調査

方言地図のためのデータが、事前に用意されていることはほぼ皆無なので、一般にデータ収集のためのフィールド調査を実施し、それをもとに地図が作成される[1]。各地図集の内容はそれぞれ異なるが、以下に一般的なモデルをあげ、おおまかなイメージを示す。

#### ・対象領域

全国を対象とするような大規模な調査研究は、国立国語研究所が行うことがほとんどである。それに対し、

単一もしくは複数の市町村、あるいは都道府県が対象とされることもあり、河川の流域が選ばれることも少なくない。このような狭域を対象とする場合は、大学の研究室が中心になって実施されることが多い。

#### ・調査項目・方法

何を調べるかであるが、対象地域内に異なり(方言差)が想定される対象を選定するのが基本である。多くの場合は、動植物名などの対象(「かたつむり」「桑の実」等)や標準語形式(「行かない」「行かなかった」等)を絵も含め、意味として提示し、それを表す語形を求める。項目数は200程度が多い。

#### ・調査地点

対象地域の広さによるが、地図にする場合、図郭内に最低200地点程度は含まれていないと、見栄えが良くない。市区町村程度を対象にする場合は、基本的に全集落を目標にすることが多い。全国レベルの場合は、やや幅があり、500~2000地点程度を対象に実施されてきた。

#### ・話者

言語は変化する。また、使用者の移動は言語に影響を及ぼす。したがって、話者の条件をある程度そろえなければ、場所による異なり以外の属性が地図上の言語変異に投影されてしまう可能性が否定できなくなる。

そこで、話者の年齢層や居住歴などの条件を一定にすることが必要である。多くの場合は、高齢(例えば、70歳以上)で基本的に生え抜き(もしくは移動期間を特定年数以内に納める)といった条件が設定される。

### 4. 日本の方言地図

日本では、これまで400冊以上(地図枚数30,000枚以上)の方言地図集が作成されてきた。この数字は、世界最高レベルである。

#### ・広域地図

最初に刊行された方言地図集は、国語調査委員会による『音韻分布図』(1905年)と『口語法分布図』(1906年)であり[2]、全国を対象としている。先に挙げた『フラ

『方言地理学の方法』の刊行(1902-1909年)からそれほど遅れてはいない。

日本全国を対象とした方言地図集は、その後、国立国語研究所による『日本言語地図』(LAJ:2400地点)[3]と『方言文法全国地図』(GAJ:807地点)[4]があり、21世紀に入ってからLAJ・GAJとの経年比較をおもな目的とする『新日本言語地図』(NLJ:554地点)[5]が刊行されている。

#### ・狭域地図

1970年代から、地域ごとの方言地図集が盛んに刊行されるようになる。これは方言地図を用いた研究分野＝言語地理学の展開と連動する。柴田武・徳川宗賢・W.A.グロータース・馬瀬良雄による新潟県糸魚川地方の方言分布調査が1957～1961年に実施された。この調査を基盤として、柴田武は『言語地理学の方法』により言語地理学を整備した[6]。それは柳田国男による方言周圏論[7]を近代化し、目的と方法を明確化するものであった。そこでは、地域を対象にした調査、調査結果の地図化、地図に現れた分布の分析など、それぞれの方法が具体的に明示された。

『言語地理学の方法』の刊行と同時期にLAJの刊行が開始された。これが日本の方言学の中での方言分布研究・言語地理学の隆盛を促し、狭域を中心に多数の方言地図集ならびに研究文献が生産された。その推移を図2に示した。

## 5. 方言地図の作成と方言分布データ

LAJとGAJの場合、調査時は調査票に回答を記載し、そこから各地点の各項目を1枚ずつ転記したカードが、国立国語研究所に報告された。調査票は調査者の手元に残され、国立国語研究所は報告されたカードを保管する。すなわち、調査結果は二重化されることになる。

この背景には、国語調査委員会の苦い経験がある。『音韻分布図』『口語法分布図』の刊行後、第二次全国調査が実施され、その結果は当時の文部省に送付・保管されていたが、関東大震災ですべて焼失した。そこで、このような二重化により、完全消失の危機回避をはかった。

方言分布の調査データは、そのままでは地図にはな

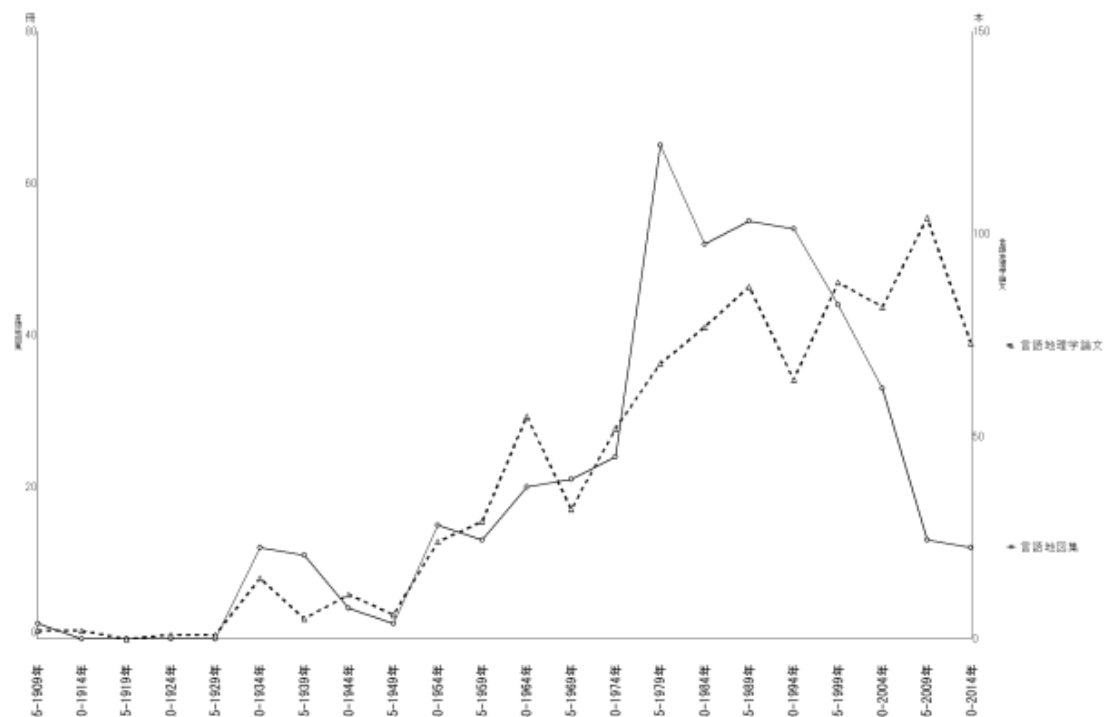


図2 方言地図集の刊行と方言地理学研究文献数の推移[8]

らない。報告された回答の中には、地図として、採用されないものもある。転記ミスによるエラー(例:項目 A の回答なのに B の回答として転記)、誤答(例:動詞否定過去形を回答すべきところを動詞否定現在形で回答)などがその例である。また、採用すべきところを抽出したり(例:動詞否定過去形として報告されたアイツワイカナンダゾからイカナンダを抜き出す)、各地図のねらいに応じて、地図見出しに統合する(例:動詞否定過去形がねらいなので、イカネカッタ・エカネカッタ・イガネカッタ・エガネカッタ・イガネガッタ・エガネガッタは、ひとつの地図見出し「ネカッタ」にまとめる)等の過程が必要である。

この作業を経て、各地図で扱う地図見出しの方言語形が決まる。ここで確定した地図見出しと地図記号の対応が凡例である。

LAJとGAJでは、地図見出しをもとにカードを整理・分類し、LAJとGAJの第4集まではこのカードをもとに手作業で地図化が行われた。地図見出しと地図上の記号の対応を示す凡例に基づき、白地図上に地図記号を専用のハンコで押印する作業である。これを原稿として、印刷所に入稿した。

GAJの第5集以降は、コンピュータによる地図作成に移行した。GAJでは、当初より、報告された全カードのデータ入力を行い、資料一覧として刊行(GAJ付録の解説書に収録)・公開していた。

このデータも活用することで、調査地点・回答・地図見出しの一覧表が作成できる。調査地点は、空間上の場所を指定する空間データであり、回答・地図見出しといった言語(方言)情報は、それぞれの空間データの属性情報である。つまり、この一覧表は、地理行列にはほかならない。これに地図見出しと地図記号の対応表をリンクするなら、空間データと地図記号がつながることになる。このようにして、方言地図はコンピュータで作図可能になった。

GAJは全6巻刊行後、第4集以前にもさかのぼって、コンピュータ作図を行い、エラーを修正した上で、画像ファイル(PDF)の公開を行っている。

NLJのための調査(FPJD)は、2010~2015年に実施した。GAJの作業実績を基盤として、当初から、データベース化とコンピュータ作図を想定して計画を進めた。その結果、調査終了の翌年には方言地図集として、調査結果を公開できた。

## 6. 方言分布データベース

GAJは、調査地点・回答・地図見出しの全データ388,659件を各集に分けて(第1集:59,042件、第2-3集:89,231件、第4集:58,878件、第5集:83,574件、第6集:97,934件)、エクセルファイルで公開している。当初はシフトJISを用いたテキストファイルで公開していたため、音声記号はアスキー文字の組み合わせで表現することで対応していたが、現在はユニコードで音声記号を表現している。

NLJは、FPJDのデータベースを元にするものであり、FPJDのデータベースにNLJの地図見出しデータを付与し、エクセルファイルで公開している(全データ:145,929件)。

なお、LAJについては、別途、国立国語研究所の熊谷康雄氏がデータベース化を進めている[9]。

## 7. 方言地図データベース

### ・目録

前述したように400冊超の地図集が刊行されており、その書誌情報の把握には、データベース化が必須である。代表的なものにしぼった目録が出されることもあった[10]。刊行地図の全体は、国立国語研究所図書館に所収されているもの(日本の方言学においては、刊行物を国立国語研究所図書館に寄贈する慣習があり、日本の方言学の学会である日本方言研究会ではそれをもとに文献リストを作成している[11])をもとに大西が言語地理学の動向を探る[12]作業の過程で集約した。その内容は、当初、日本方言研究会のサイトで公開し、その後、主要なものを選定して目録化した[13]。

このような地図集の目録整備に続き、各地図集に収録された30,000枚以上の地図項目をデータベース化

した。このデータベースは、FPJDを実施する際に、調査項目作成の基盤とすることをねらって作成された。そのため、各地図のキーワード的な分類指標が含まれており、それをもとに地図を検索することも可能である。

以上の地図集書誌目録、地図目録は、FPJD実施途上の2014年2月に公開した。

・画像

目録を手がかりにすることで、原図の存在が確認できる。もとの地図集が手元がない場合は、国立国語研究所図書館(研究図書室)に足を運べば、原図にたどり着くことができる。問題は、狭域地図集の多くが研究室を刊行元とする私家版で、国立国語研究所まで行かないと見られないことである。

そこで、2017年度から、地図を画像にして公開することを開始した。

公開している地図画像には、画像ファイルの中に地図図郭の経度緯度を格納したジオタグ付き画像ファイルが含まれる。ジオタグ付きファイルは、代表的なGISであるArcGIS・SIS・QGISにおいて、地図上の正しい位置に直接表示・オーバーレイできるようになっている。したがって、同じ項目について、異なる地図集の地図をGIS上で表示することにより、対象域を広げる形で分布を確認するようなことも可能である(図3)。

ファイル形式はTIFFを採用している。これは、事前に検討した結果、これらのGISに共通してワールドファイルとして、ジオタグが有効なのがTIFFであったことによる。また、TIFFなので、多少サイズが大きいものの、

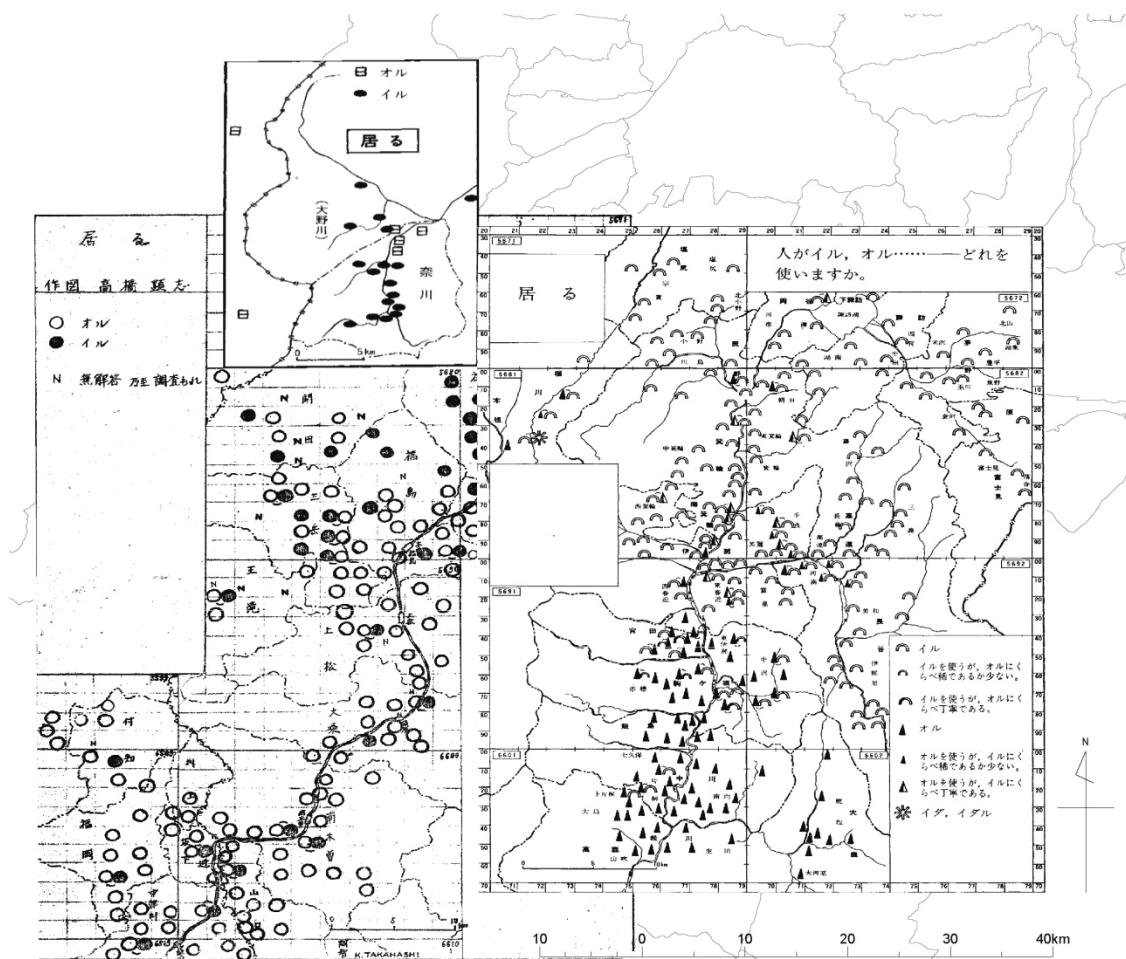


図3 長野県木曾・上伊那・諏訪地方の「いる」

(信州大学方言研究会編 1976『木曾及びその周辺地方の言語地図 語彙篇I』、馬瀬良雄 1980『上伊那の方言』、馬瀬良雄 1997『信飛国境地帯 奈川村の方言』を行政界図上に重ねて表示している。)

単に閲覧する場合も問題は生じない。

なお、初期の狭域地図集は、1960～1970年代に作成されたものが多く、刊行からすでに50年を経ている。用紙やインク類の品質に起因すると考えられる劣化が生じているものも少なくない。画像ファイル化は、資料の保存も目的に含んでいる。

以上のデータベースは、ウェブサイト「方言研究の部屋」[14]から閲覧やダウンロードができる。

## 8. GISでの活用

GAJ・LAJの方言分布データベースは、調査地点と方言情報の組み合わせにより構成されている。このデータベースにおける調査地点は、「地点番号」により6桁の数字でコード化されている。これは、日本全体を、5万分の1地形図が縦横10枚ずつ(計100枚)入るメッシュで縦横10等分し、左上から00、01、02と数字を与え、右下を99とする。そのメッシュの中の地形図に対し、同様に00から99までの数字を与える。さらに地形図の中を縦横10等分して、同じく00から99までの数字を与える。以上の2桁の数字を3回組み合わせることで地点を特定する(地点が収まったメッシュの番号を与える)方法で、国立国語研究所地点番号システムと呼ばれている(図4)。

地形図を基盤にしているため、各図郭の経度緯度は確定でき、それを基にして地点番号が与えられたメッシュの中心の経度緯度を算出することが可能である。すなわち、GAJ・LAJは調査地点データに対し、空間位置を指定する属性を与えることができる。

なお、このシステムは図4の(AとBの枠組みがある)上段から理解されるように、AとBの間で最初の2桁が重複する可能性がある(境界は北緯29度)。実際には、海域と陸(島)域の関係で、ほとんどの場合、重複が回避される(正確には、小笠原で重複が発生するが、LAJ・GAJは小笠原を調査しなかった)。

小地域を対象にした狭域地図は、さらにもう1段メッシュ区切りを追加し、8桁にすることでこの方式を踏襲することが多く、6桁の場合と同様に与えられた地点番

号から経度緯度が求められる。なお、地点番号から経度緯度を求める計算式は、大西のウェブサイト[15]に挙げられている。

NLJでは、最初から調査地点を経緯度の数値の組み合わせでコード化しており、地点の位置指定は、さらに簡便に求められるようになっている。

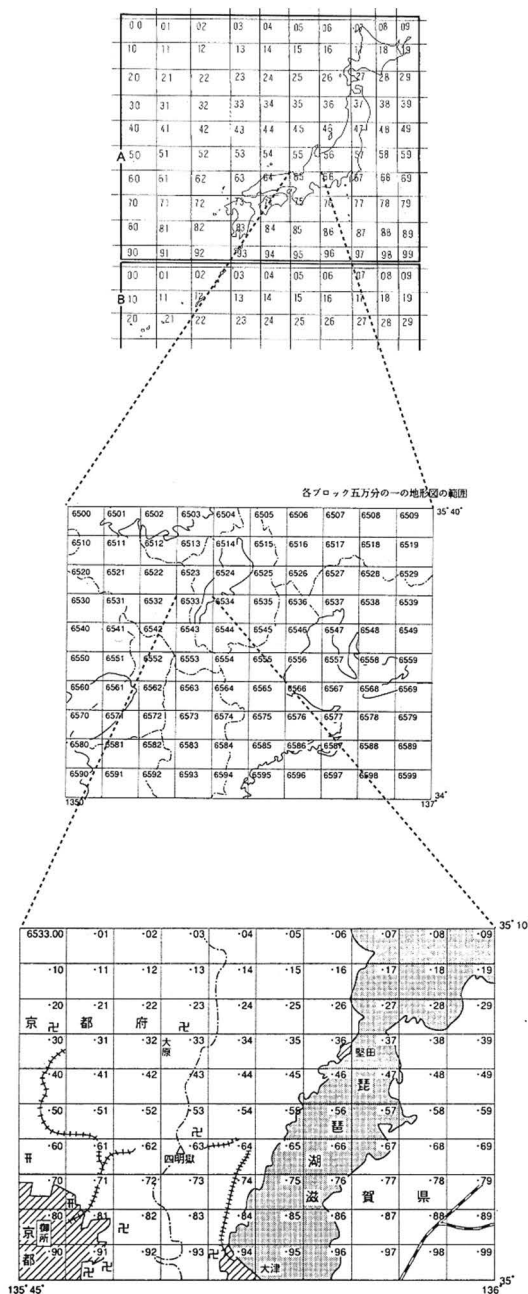


図4 国立国語研究所地点番号システム

以上のように、公開している方言分布データは、GISによる利用の便が高い。

例えば、方言分布と河川や標高を基にした地形との関係を把握することが可能である(図5)。

また、GAJとNLJを較べることで分布の経年変化を把握し、それと人口密度に基づく社会状況との関係を考察することもできる(図6)。

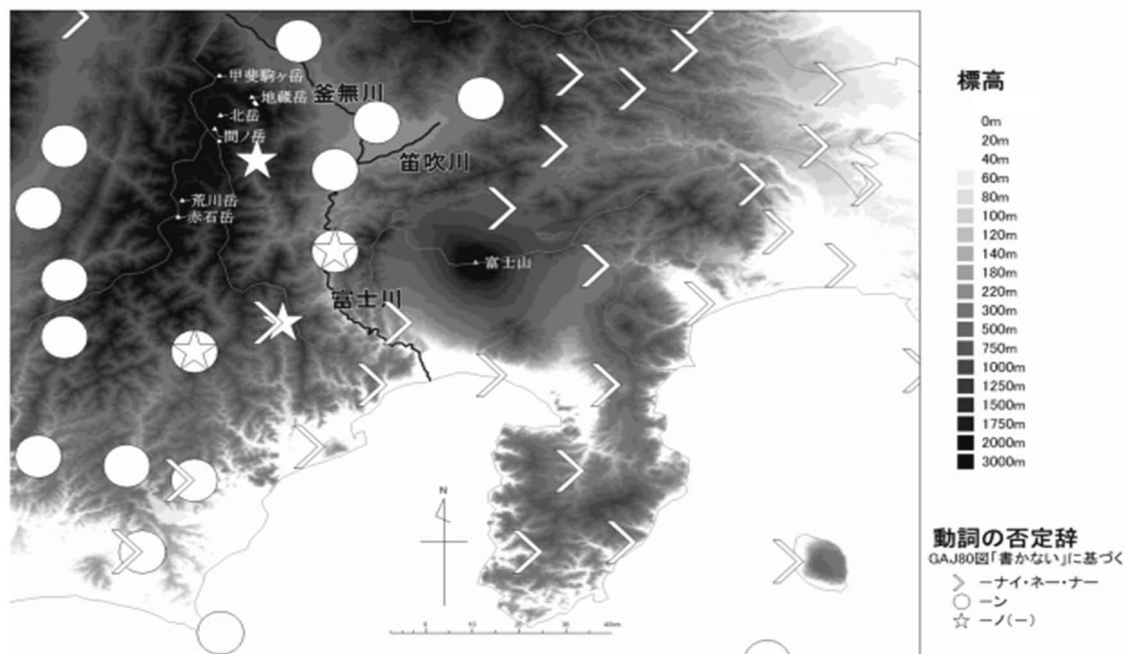


図5 富士川流域の動詞否定辞[16]

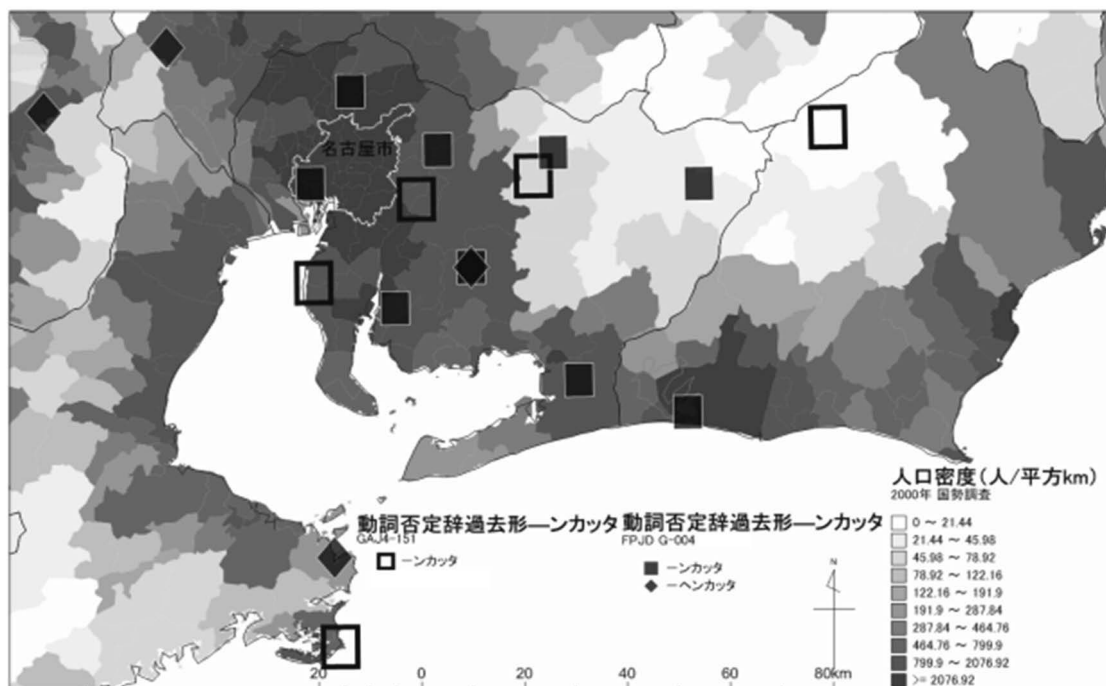


図6 東海地方のンカッタと人口密度[17]

## 9. むすび

方言分布・方言地図のデータベースについて、沿革と現状を述べた。方言学・言語地理学は、人文科学の中でも、保有・公開しているデータ量においては、かなり多い分野として、位置付けられるのではないかと思われる。

ただし、強調したいのは、そのような位置付けにあるわけではない。方言のデータには、場所を定義する空間情報が含まれている点である。空間情報をキーとすることで、分野間をリンクし、データの比較・対照・照合が可能になる。

その際に大きな力を発揮するツールは地図(そしてそれを実現するための GIS)である。ことば(方言)の情報と、例えば、歴史・地理・民俗などの情報の関係について、地図を用いて、また、地図の上で、客観的な裏付けに基づく分析・研究が展開できる。

方言学・言語地理学の中でも、方言の起源と形成過程の解明を目標とする方言形成論は、究極の課題であり、かつ言語に関する研究以外の分野との連携が必須である[18]。ここでは十分に触れることができなかったが、時間データも活用することで領域形成論[19]の再活性化も射程に収まるはずだ。ぜひ、諸分野からの意見や希望をお寄せいただきたい。

## 参考文献

- [1] 徳川宗賢. 架空教室「言語地理学演習」第一日. 言語, 1978, vol. 7, no. 9, p. 12-21.
- [2] 国語調査委員会. 音韻分布図. 日本書籍, 1905. 国語調査委員会. 口語法分布図. 国定教科書共同販売所, 1906.
- [3] 国立国語研究所編. 日本言語地図(全6巻). 大蔵省印刷局, 1966-1974.
- [4] 国立国語研究所編. 方言文法全国地図(全6巻). 大蔵省印刷局・財務省印刷局・国立印刷局, 1989-2006.
- [5] 大西拓一郎編. 新日本言語地図—分布図で見渡す方言の世界—. 朝倉書店, 2016.
- [6] 柴田武. 言語地理学の方法. 筑摩書房, 1969.
- [7] 柳田国男. 蝸牛考. 刀江書院, 1930.
- [8] 大西拓一郎. 方言地理学の研究動向. 方言の研究, 2016, no. 2, p. 83-97.
- [9] “『日本言語地図』データベース(LAJdb)”. <http://www.lajdb.org/>, (参照 2018-01-31).
- [10] 三石泰子. 言語地図集の目録. 言語生活, 1975, no. 284, p. 50-56.

- [11] 三石泰子. 言語地図集の目録. 国語学, 1979, no. 119, p. 87-91.
- [12] “日本方言研究会”. <http://dialectology-jp.org/>, (参照 2018-01-31).
- [13] 大西拓一郎. 日本の方言地図と方言文法地図の動向. 日本語学, 1992, vol. 11, no. 6, p. 188-200.
- [14] 大西拓一郎. 主要方言地図目録. 馬瀬良雄監修. 方言地理学の課題. 明治書院, 2002, p. 446-465.
- [15] “方言研究の部屋”. <http://www2.ninjal.ac.jp/hogen/>, (参照 2018-01-31).
- [16] “方言の宇宙”. <http://www2.ninjal.ac.jp/takoni/>, (参照 2018-01-31).
- [17] 大西拓一郎. ことばの地理学—方言はなぜそこにあるのか—. 大修館書店, 2016.
- [18] 大西拓一郎. 言語変化と方言分布—方言分布形成の理論と経年比較に基づく検証—. 大西拓一郎編. 空間と時間の中の方言—ことばの変化は方言地図にどう現れるか—. 朝倉書店, 2017, p. 1-20.
- [19] 大西拓一郎. 方言形成論序説—言語地理学の再興—. 方言の研究, 2017, no. 3, p. 5-28.
- [20] 大林太良. 東と西 海と山—日本の文化領域. 小学館, 1990.