

国立国語研究所学術情報リポジトリ

『方言文法全国地図』における共通語化の状況：
多変量解析を用いた分析

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 鑓水, 兼貴 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002967

『方言文法全国地図』における共通語化の状況

—多変量解析を用いた分析—

鎌水兼貴(国立国語研究所・研究開発部門)

1. はじめに

『方言文法全国地図(以下 GAJ)』は、回答語形がすべて電子データの形で公開されているという点で非常に貴重なデータベースであるといえる。利用者が、さまざまな観点から地図を作りなおすことができるだけでなく、公開された一定の基準によって整理されたデータという点で数量的解析に利用することも可能である。数量的手法を使うことによる発見もあるだろう。本発表では、GAJ 各地点の「共通語度」を測定し、GAJ における共通語化の状況を数量的に概観する。

2. 「共通語度」の測定

GAJ の調査は、共通語の表現を提示して、話者が方言形で答える形式をとっている。そのため本発表における「共通語」とは、GAJ における「見出し語」を指す。また、本発表で使用した GAJ のデータは、共通語の確定が容易な GAJ 第 1~3 集に限った。

まず各項目において、各調査地点の回答語形に「共通語度」を与えて行列を作成する。「共通語度」の手法としては、以下のいくつかが考えられる。

(1) 共通語形と一致しているかどうか

共通語形と一致していれば 1、不一致であれば 0 という、もっともシンプルな方法である。一般に使用率を求めるときはこの方法に基づく。河西(1981)による『日本言語地図(LAJ)』の都道府県別標準語使用率や、井上(1983)における GAJ 予備調査データの分析がこれにあたる。

(2) 非共通語形にいくつかの段階を設ける

同じ非共通語形の中にも、音声の一部が変異しただけと思われるものもあれば、語の系統が異なるものもある。そこでこれらを区別して得点を与える方法である。YARIMIZU et al. (2005) はフランスのパリ周辺の方言地図データを用いて共通語化について考察した。ここでは、共通語形に 1、音声変異形に 2、語形が異なるものを 3 を与えた。KAWAGUCHI (2006) は、語形が異なるものについてはもっと大きい値を与えるべきだとしている。

(3) 語形を音素ごとに比較する

語形を音素単位で分解して共通語形との類似度を測定する方法である。語形間の類似度の測定法は、主に比較言語学の計量的手法や(安本・野崎 1976)、音声認識研究において発展してきた。特に動的計画法を用いて語形間距離を測定する方法は古くから行われている(横山・板橋 1978)。

音素列の比較方法においても、一致・不一致だけとするのではなく、音素間の類似性を考慮されることが多い。簡易的に五十音図の列・段の一致関係で測定を試みた例(水谷 1990)や、詳細に当該言語の音響解析結果から音素間の距離データを用いた研究(HEERINGA 2004)もみられる。

本発表では、(3)の音素間距離に基づく動的計画法によって、単語間類似度を計算し、

そのあとで(2)のように非共通語形を分類した^{注1}。

GAJ で使用する各音素の弁別素性行列について単相関を求め、相関係数(1~-1)を変換して音素間の類似度(0~1)とみなした。そしてこれを利用して動的計画法によって回答語形と共通語形の距離(=類似度)を求めた。例を表1に示す。この距離が0の場合は共通語形とみなし、0.5未満、1.0未満を音声変異形、1.0以上は別の語形と定義し、4段階に分類した^{注2}。

こうして GAJ807 地点における「共通語度」すなわち、回答語形と共通語形との距離についての144項目分(GAJ第1~3集150項目より選択)の行列が出来上がった。

距離	第130図 来れば
0.0000	k u r e b a
0.0000	k ü r e b a
0.0661	k u r e b a
0.0661	k ü r e b a
0.0969	k u l e b a
0.1312	k i r e b a
0.1544	k u r a b a
0.1646	k u j e b a
0.1818	k i r e b a
0.2625	k o r e b a
0.2856	k u r e j a
0.3149	k u r e b a
0.3149	k ü r e b a
0.3509	k u r u b a
0.4030	h u r i b a
0.4293	k u e b a
0.4293	k ü e b a
0.4633	k u r e a
0.4941	k u r i w a
0.5044	k i j o b a
0.5186	k i d a r a
0.5481	k o r e j a a

表1・語形間距離の例

3. クラスタ分析

つづいて「共通語度」の行列に対してクラスタ分析を適用した。距離の定義には市街化距離(Manhattan Distance)を、クラスタ間距離の計算方法は最遠隣法(Complete Linkage)を用いた。図1は、807地点すべてについてクラスタ分析を行った結果のデンドログラム(樹状図)である。細かいため上位8クラスタにまとめている。図2は各調査地点の所属クラスタを地図化したものである。

左右を対照させると、琉球列島が共通語形からもっとも遠い存在であることがよくわかる。さらに本土は、東北・九州の周辺部と、中央部にかかれている。これも共通語化の進展の遅い地域であることが予想される。中央部は、関東・関西とそれ以外に分かれる。すなわち、関東と関西の両方が共通語の中核をなしていることがわかる。

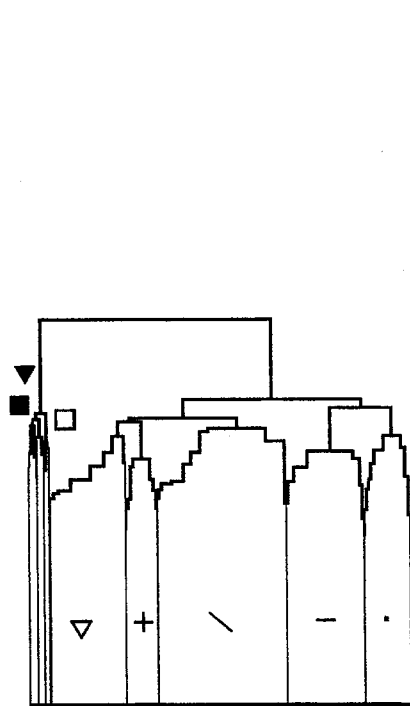


図1・GAJ第1~3集のデンドログラム



図2・GAJ第1~3集のクラスタの分布

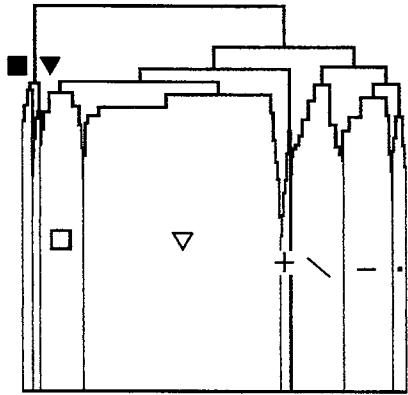


図3・GAJ 第 1 集のデンドログラム

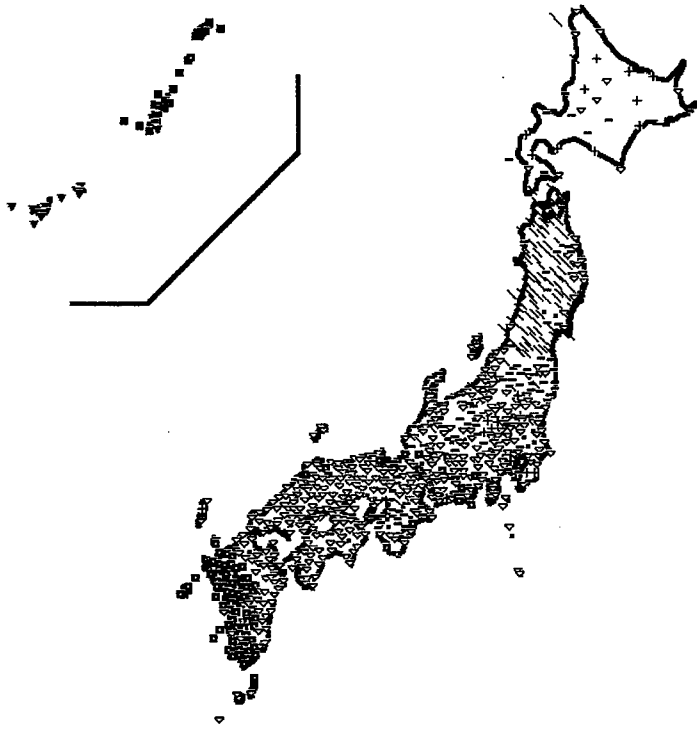


図4・GAJ 第 1 集のクラスターの分布

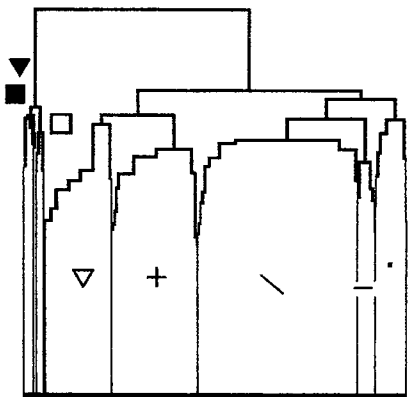


図5・GAJ 第 2・3 集のデンドログラム

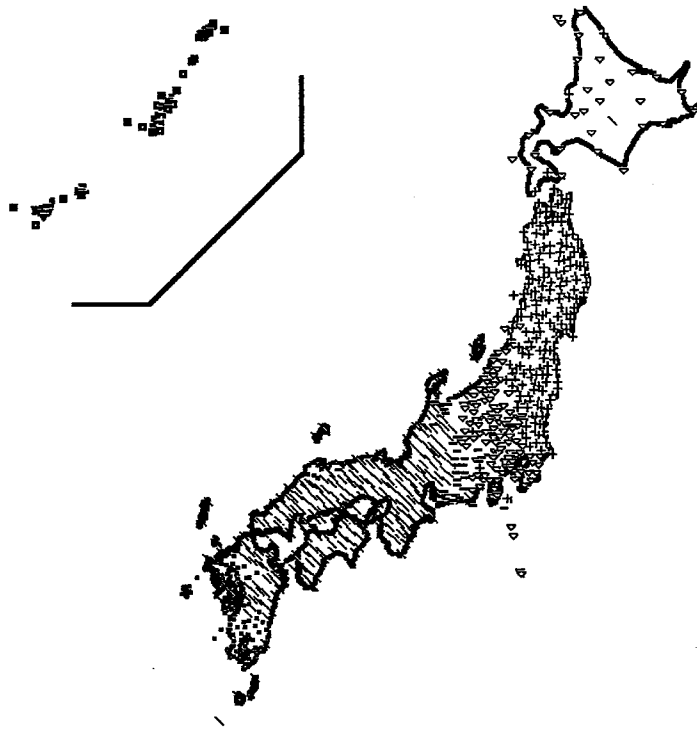


図6・GAJ 第 2・3 集のクラスターの分布

さらに第1集(助詞項目)と第2,3集(活用項目)にわけて違いをみる。図3・4が第1集、図5・6が第2・3集のクラスター分析の結果である。助詞項目である図5・6をみると関東・関西の両方を含む大きなクラスターが広がっていることがわかる。その大きなクラスターと東北を中心とするクラスターにわかれている。

一方で図5・6の活用形項目は、東西差が明確である。東側は関東・新潟と東北に、西側は、ちょうど関西地方を中心として、その周辺の東西境界域と西九州が単独のクラスターとして分かれている。境界域が1つのクラスターになっているのが興味深い。

共通語度からの分析であるため非共通語形(=方言形)間の距離はわからないが、助詞項目においてはすでに関東・関西の両方での共通語化が進んでいるため広い勢力となったと推測される。また活用項目の場合には東西差がそのまま残った状態であり、まだ関東的な共通語要素は広がっていないと思われる。

4. おわりに

以上「共通語度」を計算し、多変量解析を適用することによって、日本語の文法項目の計量的区画を考察する例を示した。動的計画法やクラスター分析は古くから使用される手法であり、決して新しいものではないが、数量処理においてGAJのような大規模データが利用できることは意義深いと思われる。

最後に課題点を挙げる。まず、動詞などの活用体系について語形の類似関係を測定することにどれだけ意味があるかという問題がある。本発表では語形の類似性も言語変化の要因になるため一定の意味はあると考えた。このほか、語形間類似度や音素間距離の計算方法についても、さらに深く考える必要もあるだろう。あわせて今後の課題としたい。

注1 紙面の都合で具体的な手続きについては、ほとんどすべて割愛した。当日の発表では示す予定である。

注2 距離という連続値であったものを間隔尺度になおしたのは、「その他」「無効回答」などとの距離が算出できないためである。これらは「別の語形」という扱いにした。しかしこの点はさらに考える必要があるだろう。

参考文献

井上史雄(1983)「共通語的文法表現の地理的分布パターン」『国語学』133

河西秀早子(1981)「標準語形の全国的分布」『言語生活』354

中川聖一・義永洋士(1985)「誤りを含んだ音素系列からの候補単語の検索」『計量国語学』14-8, pp.327-334

水谷静夫(1990)「距離に基づく語形類似指数」『計量国語学』17-5

安本美典・野崎明弘(1976)『言語の数理』(筑摩書房)

横山晶一・板橋秀一(1978)「弁別的特徴と2次系モデルに基づいた日本語単語間の距離」『計量国語学』11-4

HEERINGA, Wilbert (2004) Measuring Dialect Pronunciation Differences using Levenshtein Distance, Rijksuniversiteit Groningen.

KAWAGUCHI Yuji (2006) Is It Possible to Measure the Distance between Near Language? – A Case Study of French Dialects –, Near Language Conference, Limerick.

YARIMIZU Kanetaka, KAWAGUCHI Yuji and ICHIKAWA Masanori (2005) Multivariate Analysis in Dialectology - A Case Study of the Standardization in the Environs of Paris, Linguistic Informatics 3, Tokyo University of Foreign Studies.