

文化的情報の流通路とその滞留

—岐阜県を例とした方言分布の観点から—

同志社大学大学院文化情報学研究科 小野原 彩香（博士前期課程）

日本学術振興会特別研究員 P D（同志社大学大学院文化情報学研究科）/ UCL 連携研究員 藤本 悠

同志社大学 文化情報学部 矢野 環

Keywords : 方言、NCD、GIS、考古学、岐阜県

1 はじめに

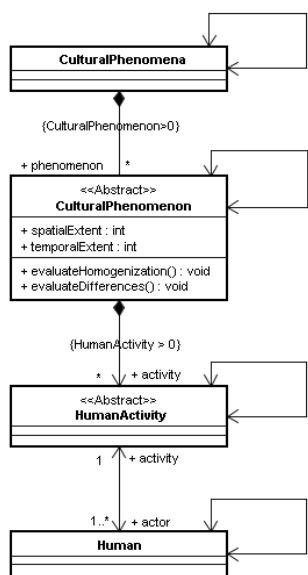


図1 文化現象のメタ・メタモデル

「文化」は、人間活動の時空間的な共通性あるいは相違性として現れる様々な文化現象によって構成される。それゆえに、文化研究は特定の文化現象のみからではなく多角的な視点から行う必要がある。ある特定の要因のみから文化を理解することは難しい。藤本が代表者を務める「CITMAS プロジェクト」では、こうした文化に対する概念を H.Rickert の考え方 [Rickert, 1915] を参考に情報モデル化している(図 1)。本プロジェクトにおいては、これを多様な文化現象を情報として実装するための「メタ・メタモデル」として位置づけている。同プロジェクトでは、このメタ・メタモデルに準じて設計・実装された個々の文化現象を総合することで

多角的な視点での文化研究を目指している。本研究は、同プロジェクトの一環として行われていて、このメタ・メタモデルを元に文化現象としての「方言」のメタモデルを設計・実装し、他の文化現象を総合的に重ね合わせるための基本的な情報構築の方法と分析手法を検討した。

2 研究の背景と先行研究

本研究では、方言分布に注目し仮説的に文化圏を構築した。言語の違いは、コミュニケーション障壁となるため、「方言」のクラスターは文化圏の一側面を写像していると言える。したがって、方言という現象を土台とし、そこに様々な事象を重ねていくことで多様な視点での文化圏の抽出と、異なる「文化圏」の間を流れる「情報」を観察できる可能性がある。従来の地域性研究は、主として「点的な分布」あるいは「面的な領域」の視点によるものが主流であるが、本研究では「線的な関係」として文化間の関係を考える。

本研究の先行研究である『徳山村のあけぼのを求めて』[1984]では、岐阜県旧徳山村において、関東・東海系アクセントを取り巻くように、関西系アクセントが存在し、一部で極めて特殊な中間型アクセントであることが紹介されている。また、同様の調査結果は、山口[2003]にも示されている。

泉ら [2007]は、これらの研究成果を踏まえ、考古学における遺跡情報と方言分布、集落特性の明確化を試みた。この研究からは、方言分布に違いが見られた地域に存在する戸入村平遺跡や塚奥山遺跡などは、集落の規模や土器系統に周辺遺跡とは異なる特徴があることが分かった。

泉らの研究結果から、近現代の方言分布と縄文時代の集落研究の共通性は、文化現象における地理空間的脈絡が強く反映されていると考えることができる。

3 方言モデルの設計とデータの実装

研究では、最初に文化現象としての方言をメタモデルとして設計した。このメタ・モデルは前述のメタ・メタモデルを継承する。図2において、「Dialects (方言)」クラスは「Speaker (話し手)」クラスと、「Vocabulary (語彙)」クラスと関係を持つ「Message (メッセージ)」クラスから成る。これらのクラスは、「地理情報標準」に準じていて、「応用スキーマ (ISO19109) と同様の役割を果たす。

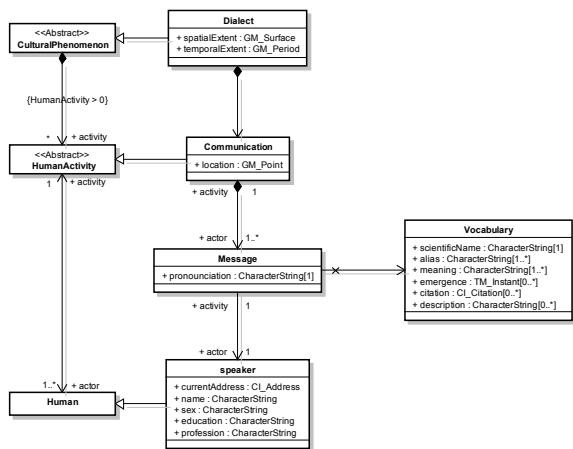


図2 文化現象としての方言の定義（メタモデル）

これらのクラスは既存の資料から収集した情報によって実装し、「Communication」クラスは『岐阜県方言の研究』[奥村, 1976]に記載された岐阜県全域の方言語彙の地域差を示す 10 枚の分布図から実装した。

これらの分布図は、それぞれ「唐がらし」、「唐もろこし」、「薺指」、「旋毛」、「竹馬」、「お手玉」、「粳米」、「馬鈴薯」、「真綿」、「里芋」という 10 の調査語彙の分布を示している。それぞれの分布図では、125 箇所の調査地点について、各語彙をどのように発音するかを記号でプロットしている(図 3)。本研究では、ArcGIS を用いて 10 枚の分布図についてジオリファレンスを行い、調査地点の位置データを作成した。また、それぞれの語彙については各調査

地点（調査地点 1～調査地点 125）で各語彙（「唐がらし」、「唐もろこし」…「里芋」）がどのような発音であるかという語彙データを作成した。



図3 『岐阜県方言の研究』p. 41掲載の「唐がらし」の図に他の語彙の図を参照して全調査地点を書き加えたもの

```
<?xml version='1.0' encoding='UTF-8' standalone='yes'?>
<dataset>
  <communication id='1'>
    <name>LOC_001</name>
    <location>
      <position>
        <coordinates>137.146015 36.338982</coordinates>
      </position>
    </location>
    <message>
      <vocabulary idref='voc-redpepper'/>
      <pronunciation>NANBAN</pronunciation>
    </message>
    <message>
      <vocabulary idref='voc-corn'/>
      <pronunciation>TONAWA</unciation>
    </message>
  <message>
    <vocabulary idref='voc-annularfinger'/>
    <pronunciation>BENISASHIIBI</pronunciation>
  </message>
  <message>
    <vocabulary idref='voc-hairwhorl'/>
    <pronunciation>JIJIMAKI</pronunciation>
  </message>
  <message>
    <vocabulary idref='voc-stilt'/>
    <pronunciation>SAGASHI</pronunciation>
  </message>
  <message>
    <vocabulary idref='voc-otedama'/>
    <pronunciation>OJAMI</pronunciation>
  </message>
  <message>
    <vocabulary idref='voc-rice'/>
    <pronunciation>URICHI</pronunciation>
  </message>
  <message>
    <vocabulary idref='voc-potato'/>
    <pronunciation>JAGAIMO</pronunciation>
  </message>
</dataset>
```

```

<message>
  <vocabulary idref='voc-otedama' />
  <pronunciation>OJAMI</pronunciation>
</message>
<message>
  <vocabulary idref='voc-rice' />
  <pronunciation>URICHI</pronunciation>
</message>
<message>
  <vocabulary idref='voc-potato' />
  <pronunciation>JAGAIMO</pronunciation>
</message>
<message>
  <vocabulary idref='voc-cotton' />
  <pronunciation>NULL</pronunciation>
</message>
<message>
  <vocabulary idref='voc-aroid' />
  <pronunciation>NULL</pronunciation>
</message>
</communication>
</dataset>

```

図4 ISO 標準に準拠した XML データベースファイル(調査地点
1)

4 方言の類似性分析と分類

本研究では、方言の類似性を分析するために、汎用類似度算出技術の一つである NCD を算出し、分類する方法を検討した。NCD は、以下の式が示すような正規化圧縮距離である。

$$NCD = \frac{C_{xy} - \text{Min}\{C_x, C_y\}}{\text{Max}\{C_x, C_y\}}$$

ここで、C はある圧縮アルゴリズム、x および y は任意のファイルである。その x, y それぞれを圧縮したファイルのサイズを C_x, C_y とし、x に y を接続して圧縮したファイルのサイズを C_{xy} とする。この x, y 両ファイルが記号列として一致度が高ければ NCD 値は 0 に近づき、相違度が高いならばその程度に従い、NCD 値は 1 に近づく。

本研究においては、x および y は、XML ドキュメントでインスタンス化された各々のファイルである。この分析方法の一つの利点は、一般的な分析手法と比較して、事前のデータ抽出と整形などが簡略化されることにある。また、本研究のように、地理情報標準に準拠して実装された分析手法としても注目できる。なお、XML 文書から NCD を求めるという方法は、今回、藤本が新たに考え、試みた手法である。

本研究においては、藤本が開発した NCD 算出ソフトウェア「百花壱式 Ver1.0」を用いて距離行列を作成した。NCD は、圧縮アルゴリズムに依存す

る。同ソフトウェアは GZIP 圧縮のみをサポートしている。

算出した NCD は、二つのオブジェクトに対してのみ有効であるため、全ての組み合わせで NCD を計算する必要があり、最終的には距離行列となる。この距離行列から分類を行う手法として、カルテツト法が提案されているが、同手法は、大規模なデータには不向きであるため、効率的かつ高速に計算できる Neighbor Joining 法のアルゴリズムに基づいた Neighbor Net を用いた（図5、図6）。

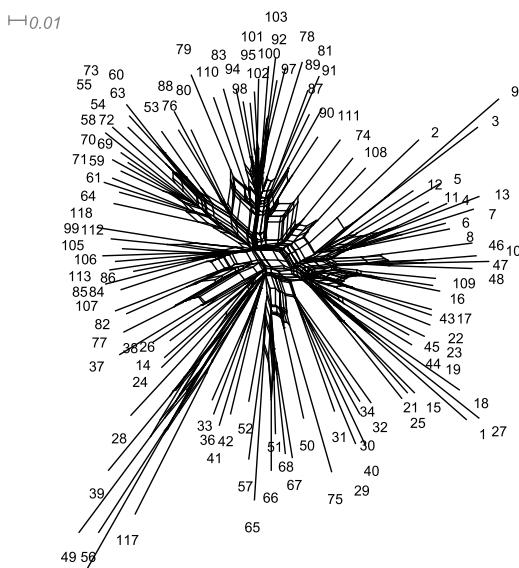


図5 NCD における Neighbor Net 法による系統樹

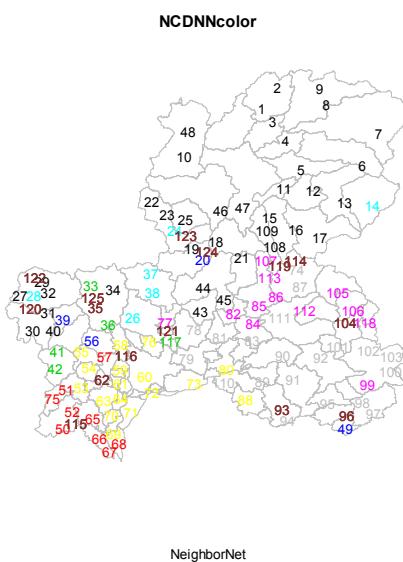


図6 NCD における NeighborNet 系統樹での分類を地理空間上に

投影したもの

なお、図 6 のような地図と調査地点の描画には、R のパッケージである maptoolsなどを用いた。

また、非階層的分類法として PAM (Partitioning Around Medoids) も用いて、分類基数および地理空間的な傾向を観察した。

これらの分析の結果、PAM では 4 分類とするのが適切であると分かった（図 7）。

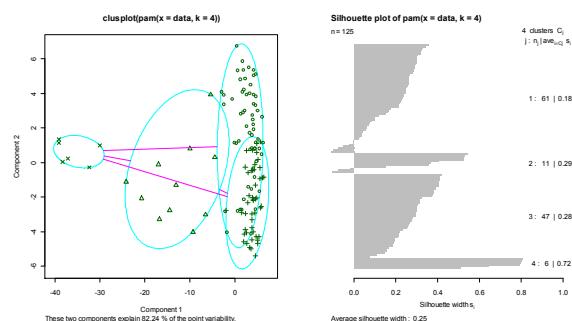


図 7 PAM4 分類の際の主成分 plot と Silhouette

これをまた、R の maptools 等を用いて、地理空間に投影した結果が図 8 である。この結果は実質 3 クラスタと残差 1 グループといえる。NCD は、発音の類似性を適切に反映している可能性が高い。

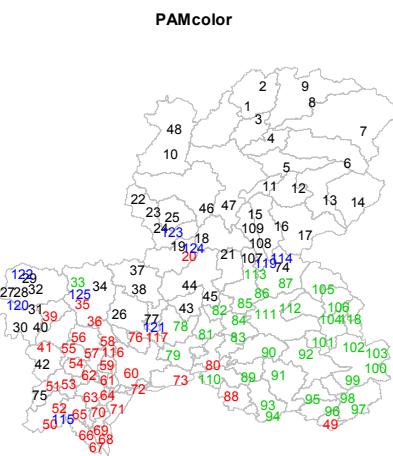


図 8 PAM での 4 分類の結果を地理空間上に投影したもの

5 分析結果の検証

方言データを分析する方法としては様々な手法

が存在し得るが、実際にどの手法が適切であるか、先行研究を通して検証する必要がある。方言学の先行研究では揖斐川が東方アクセントと近畿アクセントの境界として古くから知られている。北の境界は大垣市（調査地点 63）と垂井町（調査地点 51）の間にも走っている。垂井町およびその付近には特殊のアクセントが行われていることも発見された [服部, 1930; 柴田, 1950]。また、徳山村の戸入地区（調査地点 28）は、東方アクセントであり、塚（調査地点 122）は、近畿と東方の中間型とされている。

ここでは、こうした先行研究の成果と照らし合わせながら、他の分析手法と比較し、NCD による分類の有効性を検討する。まず、本研究では検証に用いる分析データセットを準備した。準備したのは次のような二種類のデータである。一つ目だけではなく二つ目の 0,1 データも作成したのは、クラスタリング方法によっては、0,1 データでなければ処理できないものがあるためと、一つの調査地点に複数個の発音がある場合の処理を行うためである。

一つ目は、行を調査語彙、列を調査地点番号にし、要素として各調査地点での発音を入れたもの（表 1）である。例えば、調査地点 1において、「唐がらし」という語彙が「NANBAN」という発音であれば、2 行 2 列目のセルに「NANBAN」と入れる。一つの調査地点に二つの発音がある場合は、横に二つ記号が並んでいれば左側を、縦に二つ記号が並んでいれば上を採用した。

表 1 各調査地点での各語彙の発音

	唐がらし	唐もうこし	薬指	...	里芋
1	NANBAN	TONAWA	BENISASHIBI		NULL
2	NANBAN	TONAWA	KUSURIIBI		MAIMO
3	NANBAN	TONAWA	BENISASHIBI		HATAIMO
4	NANBAN	TOUNA	KUSURIIBI		HATAIMO
5	NANBAN	TOUNA	KUSURIIBI		HATAIMO
6	NANBAN	TOUNA	KUSURIIBI		HATAIMO
7	NANBAN	TOUNA	KUSURIIBI		HATAIMO
8	NANBAN	TONAO	KUSURIYUBI		HATAIMO
9	NANBAN	TONAWA	KUSURIIBI		NULL
10	NANBAN	TONAWA	BENISASHIYUBI		EGOIMO
11	NANBAN	TOUNA	KUSURIIBI		HATAIMO
...					
125	NA				NA

二つ目の分析データセットは、行を調査語彙のすべての発音、列を調査地点番号にしたもので、各調査地点でその発音が話されている場合は 1、話されていない場合は 0とした。

表2 表1を0.1データ化したもの

	唐がらし_TOG	唐がらし_NANBAN	唐がらし_KOSHIO	...	里芋_NANKIIMO
1	0	1	0		0
2	0	1	0		0
3	0	1	0		0
4	0	1	0		0
5	0	1	0		0
6	0	1	0		0
7	0	1	0		0
8	0	1	0		0
9	0	1	0		0
10	0	1	0		0
11	0	1	0		0
...	125	0	0		1

さらに、この分析データセットでの適切なクラスター数を選択するため、最初に多次元尺度構成法を行い、データの概要把握を試みた。この分析の結果、4つに分類にすることが適當であることが分かった（図9）。勿論、孤立する特異値が発生するので、実際はより多くのクラスターに分類することが必要になってくる。

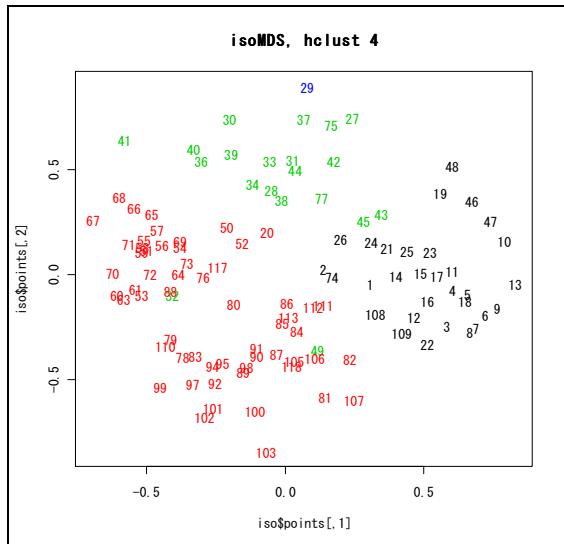


図9 MDSによる距離の散布図

この結果を踏まえて、各々クラスター分析を行った。しかし、4つのクラスタリングでは、どの方法においてもあまり差は見られなかつたため、クラスター数を適宜増やして、分析を行つた。先行研究で明らかとなった地域差を見ようとするならば、どのクラスター分析の方法においても、7から8のクラスターに分ける必要があると分かつた。

これらの分析の結果、二値に対する距離である Jaccard 係数に基づき、DPCLUS を用いた分類が、今回の場合最も適切であると判明した。

Jaccard係数によるDPCLUSを用いたクラスターを地図上に投影したものが図10である。この結果を見てみると、今回は語彙の類似性を用いたのであるが、アクセントについても東方アクセントと近畿アクセントの境界である調査地点63や調査地点51がそれぞれ別のクラスターに属しており、従来の研究成果との間に齟齬は無い。

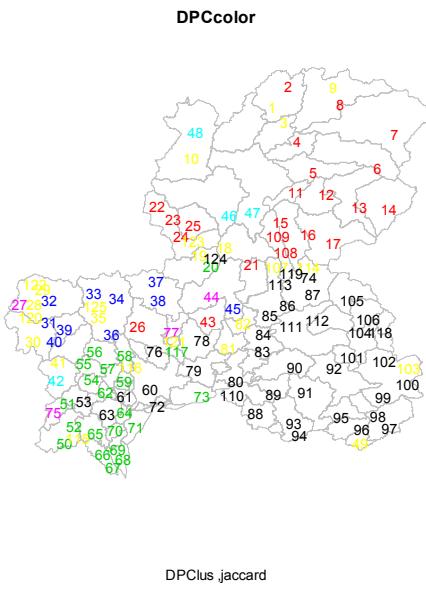


図10 Jaccard係数によるDPCLUSを用いたクラスター分類を地図上に投影したもの

一方、NCDの結果では、Jaccard係数によるDPCLUSを用いたクラスタリングでは判別できなかつた徳山村の戸入集落（調査地点28）の孤立を始めとした西揖斐地方における東方アクセントと近畿アクセントの対立が明確に区別されている。NCDでのNeighbor Netの分類結果は、Jaccard係数によるDPCLUSによる分析データセットの分類よりも、発音の類似性を反映している点が有利かと思われる。また、xmlを元にNCDで行う場合には、分析前の準備作業が簡略化されるので、膨大なデータを扱うことになる方言研究には総じて有効であると言える。

6 まとめ

今回は、多角的な視点による文化研究を行うための基礎となる仮説的文化領域の構築と、構築手法の開発を試みた。一連の分析結果から、PAM、多次

元尺度構成法およびNeighbor Net 法のいずれの方法においても、岐阜県下は大きく 4 つに分類されることが分かった。

これら 4 つのクラスターには「飛び地」が存在し、それは関ヶ原町のような交通の要衝や山間部などの過疎地に位置する集落であることが分かった。さらに、4 つ分類の境界は岐阜市周辺に位置することも明らかとなった。岐阜市周辺において文化の境界が存在し得る可能性については藤本[2010]による縄文土器を用いた分析でも指摘されていることから、土器系統の伝播と方言の伝播メカニズムには何らかの関係があると推測できるが、土器系統の伝播と方言の伝播には、どのような文化的な関係あるのかをこれから解明していく必要がある。

仮説的な文化領域構築として試みた NCD による距離化法および PAM による分類は、地理情報標準で規定されているようなオブジェクト指向の分析手法として大きな可能性を持っていることが分かった。

今回の NCD による分析は、語の共起による系統的分類とも整合しており、信頼できると見なしうる。

今後は、方言から構築した仮説的文化領域に、土器を始めとした様々な現象を重ね、総合的に地域の特色を分析していく必要がある。また、方言における文法的側面、音韻変化の方向性による精密な類似度も見なければならない。本研究で試みた一連の手法はそうした多角的な視野での分析手法にも適用できる可能性が高い。

参考文献

- [1] H.Rickert, “Kulturwissenschaft und Naturwissenschaft”, 3rd ed., 1915. (リッケルト. 佐竹哲雄訳(1939)『文化科學と自然科學』. 岩波書店.)
- [2] 德山村の歴史を語る会. (1984). 『徳山村のあけぼのを求めて—岐阜県揖斐郡徳山村遺跡分布調査中間報告—』.
- [3] 山口幸洋. (2003). 「岐阜県徳山村戸入方言のアクセント」. 『日本語東京アクセントの成立』. 鎌倉：港の人.
- [4] 泉拓良, 藤本悠, 三嶋誠. (2007). 「旧徳山村所在の縄文集落の GIS 研究」. 日本文化財科学会第 24 回大会講演論文集.
- [5] 奥村三雄. (1976). 『岐阜県方言の研究』. 大衆書房.
- [6] D.H. Huson and D. Bryant, Application of Phylogenetic Networks in Evolutionary Studies, Molecular Biology and

Evolution, 23(2):254-267, 2006. software available from www.splitstree.org.

- [7] 服部四郎. (1930). 「近畿アクセントと東方アクセントとの境界線」『音聲の研究』. 第三輯. 文學社.
- [8] 柴田武. (1950). 「揖斐川上流のアクセント」『文字と言葉』. 刀江書院.
- [9] 藤本悠. (2010). 「文化情報学および理念型モデル化分析法に関する基礎的研究—文化情報学における諸概念と方法について—」. (http://www.cis.doshisha.ac.jp/kundaikan/students/doctoral/fujimoto/data/pdf/D_Main_ver.2.0.0.pdf).