

# 日本語方言の系統樹からみるミームの系統進化

田村光平<sup>†</sup>

鈴木麗壘<sup>‡</sup>

有田隆也<sup>‡</sup>

名古屋大学情報文化学部<sup>†</sup> 名古屋大学大学院情報科学研究科<sup>‡</sup> 名古屋大学大学院情報科学研究科<sup>‡</sup>

## 1 はじめに

時間発展とともに生物が変化していく現象を進化と呼び、生物学では重要な研究課題となっている。この進化の原動力が、親から子への形質の伝播、その過程での変異、多様化した個体の中で環境に適応したものの生き残りという複製・変異・淘汰のプロセスである。しかし、親から子へと実際に伝えられるのは DNA そのものではなく、塩基配列の並び、すなわち遺伝情報である。したがって、情報の複製によって系統が生じ、そこに淘汰の下で変異が蓄積されることを進化と呼ぶならば、それは生命以外においても起こりうる。このとき、生物における遺伝子のように、文化において祖先から子孫へと伝えられるものがミームである。ミームの導入により、文化を進化生物学の手法で扱うことが可能になる。

本研究では、題材として日本語方言の系統樹を描くことで、生命・非生命の垣根を越えた一般的な進化現象に関する知見を得ることを目的とする。系統樹を描くことは、進化生物学においては一般的な手法であり、多様化の過程をうまく記述できるため、進化現象の分析に適している。しかし、文化進化においては、一度分岐した後に再結合することが一般的に起きるため、ネットワーク型の系統樹の方がより正確に進化の様子を可視化できる。本研究ではさらに、できあがった系統樹について数理解析を行い、生命進化や他の文化進化と比較した。

## 2 手法

本研究では、NeighborNet[1]を用いて系統ネットワークを作成した。系統樹を描く上での操作単位を OTU (operational taxonomic unit) と呼ぶ。これは通常は種である。行列の要素が各 OTU 間の距離である行列を距離行列と言う。NeighborNet はこの距離行列から以下のようにして系統樹を作成するアルゴリズム (図 1) である。



図 1 左は NeighborNet の基本的なアルゴリズム。右は OTU の置換方法。点線で囲まれた部分が neighbors である OTU。既に neighbors であった OTU の組と新しく neighbors となった OTU は網目構造を形成し、新たに 2 つの OTU に置換される。

i) 生成された距離行列のうち、最も距離の近い OTU の組み合わせを見つける。こうして選ばれた組み合わせを neighbors と呼ぶ。

ii) 選んだ組み合わせに、既に neighbors を持っている OTU があれば、その OTU は網目構造を形成する (図 1)。この網目がネットワークを描いた際、文化の結合を表わす。その後、3 つの OTU を 2 つに置換する。この際、置換によって生まれた新しい OTU も neighbors である。選んだ組み合わせが neighbors を持っていなければ i) に戻る。iii) OTU の数が 2 以下になるまで i), ii) を繰り返す。

本研究では、各県の方言における単語の有無を遺伝情報とみなし、各県の方言を OTU とした系統ネットワークを作成した。具体的には、ある単語が方言に存在していれば 1、していなければ 0 をコードしてビット列を作成した。これを、[2]に基づいて 64 種類 1220 個の単語について行った。ここで、「種類」とは、例えば「手」を表す単語という意味である。なお、ネットワークの描画には SplitsTree4[3]を使用した。

## 3 結果

できあがった系統樹を図 2 に示す。



図 2 日本語方言の系統ネットワーク。鹿児島と沖縄の枝の長さはそれぞれ 1/2 に省略。

密な網状構造になっている部分は形質の水平伝播が盛んなところだと考えられる。その結果として、東北地方、関西地方、四国地方、九州地方の方言がそれぞれ同じ分類群として存在していることが確認できる。それ以外でも、地理的に近い方言が近くに存在する傾向が確認できる。ただし、系統樹上では、東北地方と九州地方が近いことがわかる。これは、都からの距離がほぼ等しく、言語の伝播速度が同程度であることが一因であろう。また、愛知県と広島県の方言が類似している点が興味深い。これは、関ヶ原の合戦の後、広島藩の藩主となった福島氏・浅野氏が愛知県出身であるため、このときに語彙の伝播が起こったのではないかと推測される。

次に、土松・池上[4]らによって提案されている、下流の分岐ノードの偏りと距離空間の歪みの尺度を用いて、数理的な解析を行った。これは、全てのノードについて、その下流に存在する分岐ノードの数と、数の多さの順位をグラフにプロットするものである。系統樹の分岐に偏

A phylogenetic approach to memetic evolution in Japanese dialects

Kohei Tamura<sup>†</sup>, Reiji Suzuki<sup>‡</sup> and Takaya Arita<sup>‡</sup>

School of Informatics and Sciences, Nagoya University<sup>†</sup>,

Graduate School of Information Science, Nagoya University<sup>‡</sup>

りがなければ、下流の分岐ノードの分布は、べき乗則に従う、つまり、両対数グラフでは、分布が直線になるはずである。図3は、方言の系統樹と、OTUの数とビット数を等しくしてランダムに生成した系統樹の分岐ノードの分布である。ランダムに生成した系統樹はほぼ直線に従っている。土松らの研究では、鳥居や雑煮の分岐ノードの分布は、お椀型になることが示されている。方言の系統樹は、そのお椀型と直線の間のような形となった。これは、言語が他の文化と比べて制約が緩く、自由な変異が可能であること、いわゆる「言葉の恣意性」を意味していると考えられる。

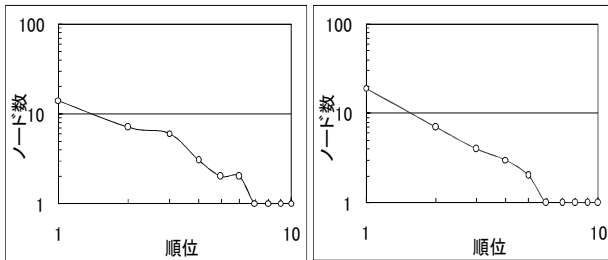


図3 下流の分岐ノードの分布。左は方言の系統樹。右はランダムに生成したコードから作った系統樹。縦軸が下流の分岐ノードの数、横軸がその順位。

次に、距離空間の歪みについて解析を行った。系統樹を描く場合に距離行列を用いるのは一般的な方法の一つだが、特殊な場合にのみ生成される距離行列の形が存在する。それが Additive Metric(親から子へと累積的に進化してきた結果生成される距離空間)と Ultrametric(Additive Metricのうち、進化速度一定の場合に生成される距離空間)である。歪みは元の距離行列 $D(d_{i,j})$ と、最も距離行列に近い Additive Metric・Ultrametric な距離行列 $D^A(d_{i,j}^A)$ 、 $D^U(d_{i,j}^U)$ との差によって求められる。つまり、それぞれの歪みの指標は

$$\frac{\sum_{i,j}(d_{i,j}-d_{i,j}^A)}{\sum_{i,j}d_{i,j}}, \frac{\sum_{i,j}(d_{i,j}-d_{i,j}^U)}{\sum_{i,j}d_{i,j}}$$

である。これにより、対象としている文化の系統性と中立性に関する知見を得ることができる。日本語方言の系統樹についてこの歪みを求めた結果を表1に示す。同表より、日本語方言の系統樹は系統性・中立性ともに大きく歪んでいることがわかる。系統性の歪みは、頻繁に語彙の借用が起こることによるものであろう。系統性は変異が親から子へと一方的に累積することの指標なので、形質が頻繁に水平伝達される場合には歪みが発生しやすい。中立性に関しては、言語の変化には安定した平衡期と分裂が起こる中断期があるとするディクソンの断続平衡モデルで説明できる[5]。つまり、言語は漸進的に一定の速度で進化するのではなく、従って中立性に歪みがある。ディクソンは、系統樹が適用できるのは中断期だけであるとして、言語への系統樹の安易な使用を批判しているが、平衡期の文化伝播による言語の収束を網状構造で表わせると考えれば、ネットワークモデルには言語進化の平衡期と中断期を包括できる可能性がある。

表1 AdditiveMetric と Ultrametric からの歪み

	AdditiveMetric からの歪み	Ultrametric からの歪み
日本語方言	0.184	0.538
ランダム	0.029	0.030
哺乳類 (30種類)	0.014	0.098

ここで、系統性の歪みがどの程度ネットワークの網状構造に反映されるか調べてみた。系統性の歪みが大きいほど形質の水平伝播が頻繁であり、ネットワークの網目が密になると予想される。ランダムに20ビットのビット列を10個生成し、そこから得られた距離行列から系統樹を生成した。これを繰り返して50個の系統樹を生成し、それぞれについて系統性の歪みとネットワークの網目の数を調べた。その結果が図4である。同図より、基本的に、網目の数は系統性の歪みに比例して大きくなっているが、歪みが大きくなると発散していることがわかる。歪みが0のとき、網目の数もまた0である。しかし、例えば、大きく二つの分類群を有する系統樹で、その分類群中 OTU 間では非常に系統性が高いならば、全体として見れば歪みがあったとしても網目の数は少ないだろう。したがって、系統性の歪みは確かにネットワークの網状構造に影響を与えるが、他の要因もまた影響を及ぼしていると考えられる。

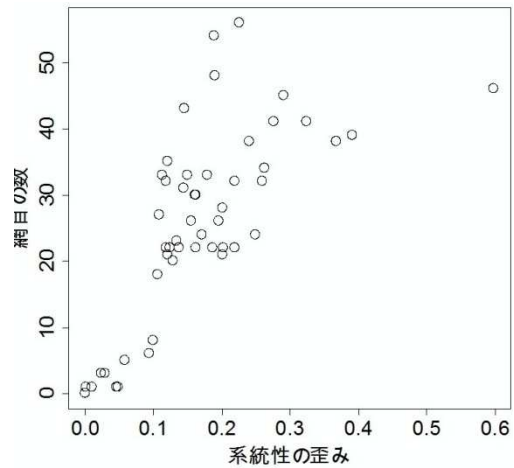


図4 系統性の歪みと網目の数との関係。横軸が系統性の歪み、縦軸が網目の数。

#### 4 まとめ

本研究では、生命・非生命の垣根を越えた進化現象の理解を目指し、NeighborNet を用いて日本語方言の系統ネットワークを描いた。さらに、できあがった系統樹について、数理解析を行い、日本語方言は、他の文化に比べてランダムな変異が許容されること、系統性・中立性ともに大きな歪みがあることを示した。また、系統性の歪みと系統ネットワークの網目の数を比較し、基本的には歪みの大きさに比例して網目の数が増えることを示した。ディクソンの言語進化の断続平衡説の検証などを今後の課題としたい。

#### 参考文献

- [1]D. Bryant and V. Moulton, NeighborNet: An agglomerative method for the construction of planar phylogenetic networks, *Molecular Biology and Evolution*, 21, 255-265, 2004.
- [2]平山輝夫, 全国方言基礎語彙の研究序説, 明治書院, 1979.
- [3]D. H. Huson and D. Bryant, Application of Phylogenetic Networks in Evolutionary Studies, *Molecular Biology and Evolution*, 23(2), 254-267, 2006.
- [4]石山,伊藤,柴田,土松,池上, 系統樹から迫る非生命進化: 鳥居・雑煮・デジタルカメラ, 第7回日本進化学会大会ポスター発表, P2-46, 2004.
- [5]R.M.W.ディクソン,大角翠, 言語の興亡,岩波書店, 2001.