

方言データから方言の類型を探る試み

滋賀大学教育学部 松丸真大

Michio Matsumaru

Faculty of Education, Shiga University

要旨

方言研究における「方言区画論」は、言語的特徴の地理的分布を総合的にみて、「日本語方言をいくつに分類できるか」という問いに答えることを目的とする。しかし、従来の区画論は、特定の研究者の知識と洞察力に依存しており、客観性・再現性の点で限界があった。本研究は、この課題に対しデータ駆動型のアプローチを適用し、日本語諸方言の語彙・文法事象の潜在的な構造(類型)と、その空間的な現れ方を定量的に導くことを目的とする。

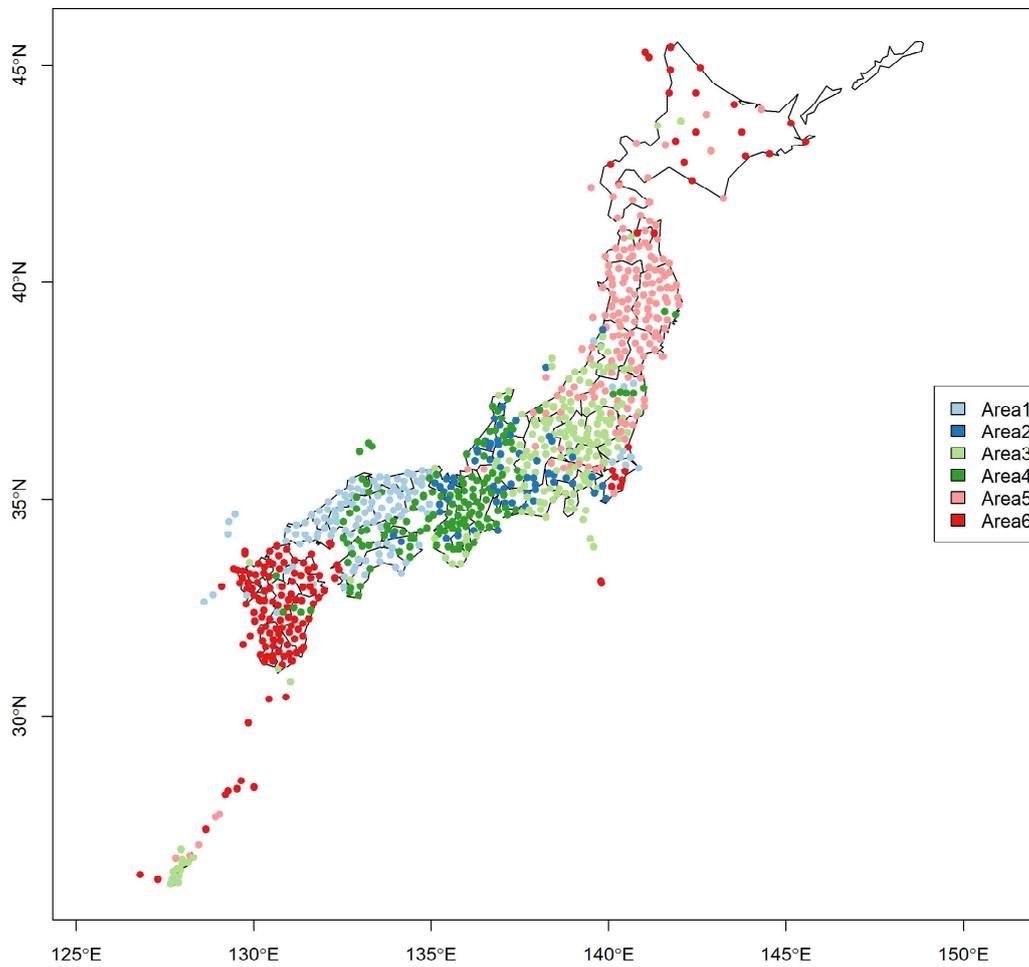
データとして国立国語研究所の『方言文法全国地図』第1～6集の全データを用いた。これには日本全国807地点で行われた語彙・文法調査の回答が収録されている。このデータから頻度数50以上の見出し語形を抽出し、各調査地点の経度・緯度情報とともに「地点×言語的特徴」の地理行列として構造化した。この行列データに対し、非負値行列因子分解(NMF)によるトピックモデルを適用した。NMFは、データセットに内在する潜在的な構造(方言の類型)を抽出し、その構造を非負の重みとして解釈できるため、多変量データの類型化に適している。

NMFの適用結果、データは6分類、あるいは8分類に類型化することが適していることが判明した。6分類の結果は、以下の二つの側面から解釈できる。

(1) **構造的側面**(特徴の束の抽出): 抽出された分類(コンポーネント)は、特定の言語的特徴群が高い寄与率を示す「特徴の束」として現れた。例えば、ある分類では、格助詞「ガ」の使用、自称詞「ワシ」・対照詞「アナタ」の使用、動詞仮定形の縮約(書キヤ)、動詞過去形のウ音便(モロータ)、形容詞ウ音便(タコーテ/メズラシュー)などの言語的特徴が同時に高い寄与率を示した。これは、これらの言語的特徴が地理的に相関して現れる潜在的な方言の類型が存在することを示唆する。

(2) **分布的側面**(空間的類型化): 各分類の空間的な重み(寄与率)を地図上にプロットした結果、同じ構造を持つ方言が地理的にまとまって現れることが確認された(図参照)。このデータ駆動型の区画は、大局的には先行研究(東条操, 1953)による伝統的な方言区画案と一致し、結果の妥当性を裏付ける。さらに、NMFは従来の区画案では見えなかった新たな構造も明らかにした。具体的には、地理的に遠く離れた千葉県房総半島と九州地域が、共通の類型に分類されるという興味深い結果が得られた。

非負値行列因子分解は、多次元的な方言の文法データを客観的に類型化し、その潜在的な構造を抽出する強力な手段である。今後は、より大規模な方言コーパスを用いた分析へと展開し、方言類型とその地理的分布の解明に貢献することを目指したい。



引用文献

東条操編(1953)『日本方言学』吉川弘文館