E-news
R3-No. 14 (25/Oct/2021)

研究紹介

琉球諸語発音に対するアラインメント手法と特徴発見

工学科知能情報コース・教授　岡﨑 威生

１．研究背景

言語発音には時間と位置の違いによる変化が見られます。「音」の伝搬は人間行動に伴うものであることから、その情報は単なる言語変化だけでなく生活様式や文化の変遷を含むと言われています。特に琉球諸語は、海により隔たれた離島環境という中で変遷し、単なる位置関係だけで説明できない分布を持つため、多くの言語学者が研究対象としてきました。研究は主に現存する言語における発音の違いに基づいています。様々な地点で多様な言語発音が調査され、発音記号列として蓄積されてきました。従来言語学者の知識によるところが大きかった言語学分野の研究は、モダンな自然言語処理手法の適用が試みられており、単語の発音を発音記号列へ変換、発音記号列間の距離を算出、距離行列に基づいた系統樹や系統ネットワークなどの生成という段階に大別されます。本研究では、発音変化を定量的に解析するためのアラインメント手法と、変化パターンの特異性を抽出するための拡張 n-gram を開発しました。

２．発音アラインメント

同じ言葉に対する地点による発音の差異を明らかにするために、塩基配列などの解析に使用されているシーケンスアラインメントが用いられています。Levenshtein 距離が方言の解析に導入されて以降、広く用いられるようになりました。さらに距離の重み付け方法として、個々の発音の特徴を舌の位置や唇の形などによって表現する音素素性を用いたものや、実際の発音の周波数を利用した手法も提案されています。近年では、自己相互情報量(PMI:pointwise mutual information)を利用し、記号対の重みを更新しながらアラインメントする手法が一般化しつつあります。琉球諸語の発音記号列は 73 種類の子音、28 種類の母音によって構成されています。また琉球諸語の特性として、子音と母音の結びつきが非常に強いということが挙げられます。このような特性を考慮した上で、発音記号を音素素性ベクトルとして表現し、PMI に対してベクトルに対応するする拡張を施した手法を提案したところ、高い正解率を獲得できました。

３．発音変化特徴発見

発音間距離に基づいた系統樹あるいは系統ネットワークの生成は、言語間の類似性を比較するものです。一方で各地点の特異性には、発音変化の特異性抽出が必要となります。そこでアラインメントされた発音記号列をもとに、ある地点を対象とした発音へ他地点の発音が変化したと仮定し、各地点の発音変化パターンの発生頻度から成る特徴ベクトルとして表現しました。ところで発音変化が起こる際に、発音記号列において前後の発音記号の影響が確認されています。そこで対象地点とそれ以外の地点の単語間で対応のある並びの発音記号列を、変化後と変化前の発音変化パターンとしてその頻度を要素とする特徴ベクトルを構成しました。この定量化に対する統計的検定により、特異的な発音記号の変化を抽出することが可能となりました。

L-news

R3-No. 14 (25/Oct/2021)

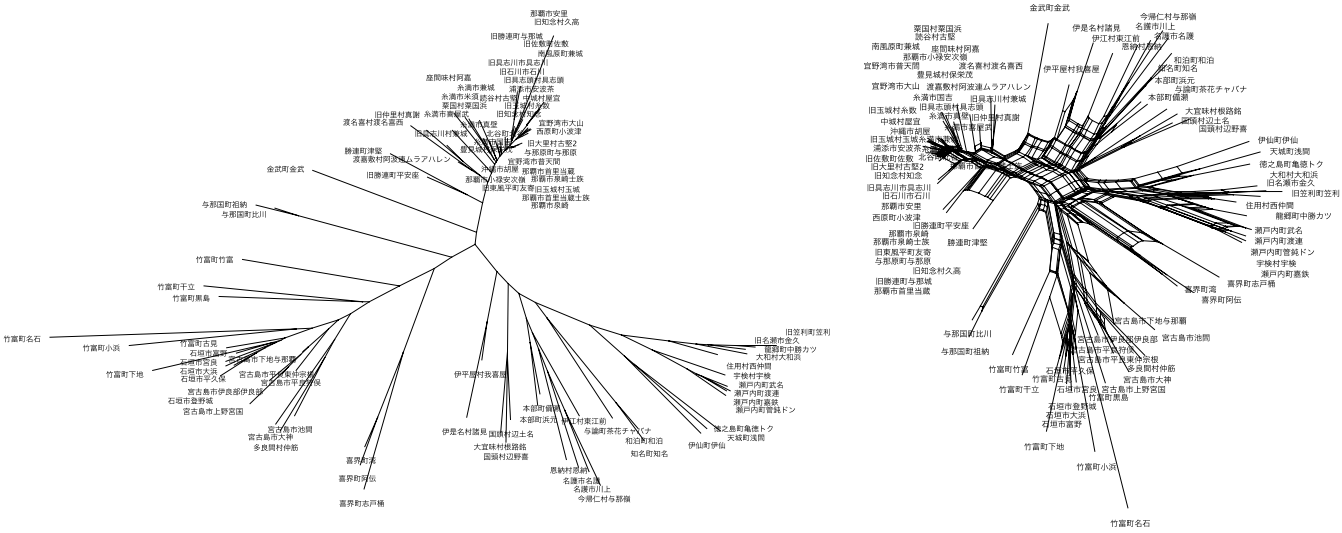

図１　提案アラインメント手法により生成した系統樹と系統ネットワークの例