博士学位論文審査要旨

２０２０年1月１７日

論 文 題 目：　コーパスにおけるモーラ情報を用いた日本の方言分類分析

学位申請者：　入江　さやか
審 査 委 員：
主　査：　文化情報学研究科　教授　金　　明哲
副　査：　文化情報学研究科　教授　矢野　環
副　査：　文化情報学研究科　教授　沈　　力
副　査：　文化情報学研究科　教授　山内　信幸
副　査：　国立国語研究所　　教授　前川　喜久雄

要　　旨：

本論文は，自然談話を調査対象とし，モーラ n-gram を計量分析することによって各地方言を分類し，東西境界線がどのように考えられるかについて計量分析の結果をまとめたものである。

本論文は，8 章より構成されている。第 1 章では，音声・音韻に関する先行研究および方言分類に関する先行研究についてまとめ，第 2 章では，分析対象としたコーパス，分析データ，および研究の流れについて述べた。第 3 章では，対象としたコーパスから抽出したモーラ unigram を用いて系統樹を作成して各地方言の分類を試み，第 4 章では，先行研究と第 3 章で得られた結果を踏まえて各地方言の分類を明確にするために，線形判別分析を行った。第 5 章と第 6 章では，日本の方言を東西に分けるのに有効なモーラ unigram と bigram について，それぞれ変数選択を行い，各地点が東西に分類される確率の方言地図を示した。第 7 章では，東西の分類に有効なモーラ unigram について，その形態音韻論的特徴について分析した。第 8 章で，各章を総括し，モーラ n-gram の頻度を用いる重要性とその有効性についてまとめ，今後の課題と展望を述べた。

本論文により，自然談話データにおける少数個のモーラだけで東西方言の識別が可能であることが初めて明らかにされた。また，本論文の結果および用いられた計量分析の諸方法は今後方言などの計量分析に大変有益な情報を与えている。よって，本論文は，博士（文化情報学）（同志社大学）の学位論文として十分な価値を有するものと認められる。

総合試験結果の要旨

２０２０年1月１７日

論 文 題 目：　コーパスにおけるモーラ情報を用いた日本の方言分類分析

学位申請者:　入江　さやか
審 査 委 員：
主　査：　文化情報学研究科　教授　金　　明哲
副　査：　文化情報学研究科　教授　矢野　環
副　査：　文化情報学研究科　教授　沈　　力
副　査：　文化情報学研究科　教授　山内　信幸
副　査：　国立国語研究所　　教授　前川　喜久雄

要　　旨：

学位申請者は 2016 年度 4 月より本学大学院文化情報学研究科博士後期課程に在学しており，国内外での研究発表を通じて研究活動を積極的に行い，それらの成果を査読付き論文 2 本として公刊している。また，英語の語学試験にも合格していることから語学（英語）について十分な能力を有していると認定されている。

2020 年 1 月 17 日金曜日 15:30 から 16:30 まで，約 1 時間の公聴会と 30 分の審査会において，種々の質疑応答の結果により博士（文化情報学）（同志社大学）の学位を授与するに十分な学力を有することを確認した。

よって，総合試験の結果は合格であると認める。

博 士 学 位 論 文 要 旨

論 文 題 目：　コーパスにおけるモーラ情報を用いた日本の方言分類分析

氏　　名：　入江　さやか

要　　旨：

本研究は，実際の言語行動の結果である自然談話を調査対象とし，モーラ n-gram の頻度を計量分析することによって，各地方言を分類し，東西境界線がどのように考えられるかを検討したものである。本研究のポイントを次の 3 つにまとめる。

（1）　自然談話におけるモーラ n-gram の出現率のみで，系統樹や線形判別分析などの手法を用いて，方言を東西に分類することができる。
（2）　東西に分類される理由を探るために，東西分類に有効なモーラ n-gram について，統計的手法を用いて抽出し，そのモーラが持つ形態音韻論的特徴を探る。
（3）　東西分類に有効なモーラ n-gram には，形態音韻論的特徴があり，それを東西方言における特徴としてルールで示すことができる。

本研究は，8 章より構成されている。第 1 章では，音声・音韻に関する先行研究，および方言分類に関する先行研究についてまとめ，第 2 章では，分析対象としたコーパス，分析データ，及び研究の流れについて述べる。第 3 章では，対象としたコーパスから抽出したモーラ unigram を用いて系統樹を作成して，各地方言の分類を試みる。第 4 章では，先行研究と第 3 章で得られた結果を踏まえて，各地方言の分類を明確にするために，線形判別分析を行い，それによって，本研究における各地点の東西の所属を決定する。第 5 章では，日本の方言を東西に分けるのに有効なモーラ unigram をいくつかの変数選択の方法を用いて分析する。その結果を比較し，最終的に選んだ変数を総当たり法によって組み合わせて，その変数の組み合わせで線形判別分析を行う。そして正解率の高い変数の組み合わせを用いて，各地点が東西のどちらに分類されるか，判別分析の確率から求め，各地点の東西所属を日本地図に示す。第 6 章では，判別に有効なモーラ unigam について，その特徴を明らかにするために，モーラ bigram についても同様の分析を行う。第 7 章は，東西の分類に有効なモーラ n-gram について，その形態音韻論的特徴について述べ，それらの特徴を用いて，東西方言分類が可能であることを示す。第 8 章で，各章を総括し，モーラ n-gram の頻度を用いる重要性と，その有効性についてまとめ，今後の課題と展望を述べる。以下，各章の要点について述べる。

第 1 章では，これまでの音節，音素の出現頻度に関する研究を概観し，その研究史をまとめる。その研究のほとんどが書き言葉，かつ共通語を対象としたものであり，方言についての研究はないことを指摘する。さらに，方言分類に関する先行研究をまとめ，出現頻度を利用して新しい知見を求めた研究が少ないことを述べ，モーラの出現頻度から方言分類を行うことを提案する。この方言分類は，自然談話を対象とし，かつ頻度を考慮したこれまでの研究にないものである。

第 2 章では，使用する方言コーパス『日本のふるさとことば集成』について説明する。本研究で使用した単位は，音声的単位としての「音節」ではなく，音韻的単位である「モーラ」に相当するものである。文字化されたテキストにおいて，モーラをどのように設定したかについて述べる。

第3章では，モーラ unigram を用いて，系統樹を作成し，日本の各地方言がどのように分類されるのかについて述べる。モーラ unigram の相対頻度を用いて，対称カイ二乗値を求め，近隣結合法と Neighbor-Net アルゴリズムにより，系統樹と系統ネットワークを作成する。その結果，近い地点が同じノードでつながり，モーラ unigram でおおよその方言分類ができることを述べる。

第4章では，先行研究と第3章の結果を踏まえて，各地方言の所属を明確にするために線形判別分析を行う。東西所属に揺れのある愛知・岐阜・石川・福井，および，音韻体系が大きく異なる沖縄の2地点を除いた学習データを用いて判別モデルを構築し，そのモデルに基づき，所属が不明であった4地点が東西のどちらに帰属するか判別する。その結果，愛知は東，岐阜・石川・福井は西に所属するという結果が得られた。なお，本章以下は沖縄2地点を分析から省く。

第5章では，東西に分類する際，どのようなモーラが有効であったのかについて，カイ二乗値，Wilks のラムダを使用した変数増減法，LASSO と Adaptive LASSO による変数選択という3つの方法を用いて分析し，比較する。まず，カイ二乗値の高い上位30のモーラについて，東西方言におけるモーラの比率を示す。ハ行四段活用動詞音便は，西部方言ではウ音便を使用し，東部方言では促音便を使用することや，西部方言においては，促音化や促音挿入語は少ないが，東部方言では多いといった従来の指摘を数値で示す。

Wilks のラムダを使用した変数増減法，LASSO と Adaptive LASSO による変数選択の結果，選ばれた変数のうち，ガ行鼻音と頻度の著しく低いものを除いた「ジャ・ダ・チョ・ネ・ホ・ヤ・レ・(ン)ー」の8つの変数（モーラ）を最終的に選んだ。次に，その8つの変数を用いて，考えられるすべての組み合わせで線形判別分析を行う。総当たり法による線形判別の結果，「ダ」のみで 97.8%の正解率であった。2つの変数を用いた「ダ・チョ」「ダ・ホ」「ダ・ヤ」「ダ・(ン)ー」の組み合わせにおいて，正解率は 100.0%であった。さらに，「ダ」「ダ・チョ」「ダ・ホ」「ダ・ヤ」「ダ・(ン)ー」を用いて，線形判別分析を行い，各地点が東西のどちらに分類されるかをコロプレス地図によって示す。

第6章では，モーラが持つ情報について検討をつけるために，モーラ bigram について，同様の分析を行う。所属に揺れのある地点は，モーラ unigram と同様に，愛知は東，岐阜・石川・福井は西に所属するという結果が得られた。次に，東西分類に寄与するモーラ bigram をいくつかの変数選択の方法を用いて選び，その特徴を分析した。各変数選択の結果を比較分析し，言語学的な特徴を見出すことのできない変数を除き，「(u)ーテ・(o)ーテ・ダナ・ダネ・ダモ・ダヨ・ンダ」の7つのモーラ bigram を最終的に選択した。次に，総当たり法による線形判別分析を用いて，正解率の高いモーラ unigram の組み合わせを求めた。その結果，3つの変数を用いた「(u)ーテ＋ダナ＋ダネ，(u)ーテ＋ダヨ＋ンダ，(o)ーテ＋ダナ＋ダネ，ダナ＋ダネ＋ダモ，ダモ＋ダヨ＋ンダ」の組み合わせにおいて，東西に分ける正解率が 100.0%であった。

第5章と第6章において，正準判別分析を行った結果，モーラ unigram において，西部方言へは，「ヤ・チョ・ホ」，東部方言へは，「ダ・(ン)ー」が分類に寄与し，モーラ bigram において，西部方言へは，「(u)ーテ・(o)ーテ」，東部方言へは，「ダナ・ダネ・ダモ・ダヨ・ンダ」が分類に寄与することがわかった。これらのモーラは，断定の助動詞，ハ行四段活用動詞連用形，および形容詞連用形のウ音便，サ行に交替可能なハ行が関係していることが考えられる。そこで，これらのモーラが持っている形態音韻論的特徴について第7章で詳しく分析する。

第7章では，東部方言，西部方言から全体の三分の二に相当する12地点，18地点の計30地点を対象に，[s]と交替可能な[h]，断定の助動詞「ダ」，「ジャ」「ヤ」，ハ行動詞連用形におけるウ音便・促音便，形容詞連用形ウ音便とウ音便なしという形態音韻論的特徴を持つモーラの頻度を最初の5分間のみ数える。その変数を用いて，線形判別分析をしたところ，東西に分かれることも確認した。

そして，これらの形態音韻論的特徴を東西における方言の差異として，ルールとして示す。ハ

行動詞連用形のウ音便と促音便，形容詞連用形の音便の有無，[s]に交替可能な[h]を多用するか否か，形態素間接続時の[j]の挿入の有無である。これらは，従来の研究においても，項目として挙がっているが，頻度を重視するならば，数ある項目の中でも，特に重要であることが示せた。

第8章では，コーパスにおけるモーラの情報を用いた方言分類についてまとめる。本研究では，日本語方言学において，過去に何度も議論され，種々の案が出されている，「東西分類」というトピックに対して，モーラという理論中立的なデータと統計的手法を用いて，再分析を行い，従来と同様の方言分類ができることを確認した。これは，形態素解析をしなくてもモーラの情報だけで方言分類ができるということを意味すると同時に，モーラには重要な形態音韻論に関する情報が含まれていることを意味する。

最後に，課題と展望を述べる。本研究では，文字化テキストを使用したが，音融合形「リャ」にも「リェァ・レァ・レア・レヤ」など種々の表記が見られた。音声学の観点から見て，どのように異なるのか異ならないのか，1モーラとすべきか2モーラとすべきかなど，Praatや ELANなどの音声分析ソフトウェアを使い，分析すればさらなる知見が得られる。また，形態素間接続時の[j]の挿入は，本研究では，形態素末母音が/e/で，助詞「は」「ば」が下接する場合しか扱えなかったが，他の母音や，他の助詞についても同様に分析する必要がある。形態音韻論のみの特徴から系統樹を作成するなど，今後の課題としたい。本研究では，モーラ n-gram から距離を求めて系統樹を作成したが，他の音融合についても分析し，音変化についての形式状態を定め，系統樹を作成することによって，新たな方言分類ができると考える。