

論文審査の結果の要旨

論文提出者氏名

川崎 義史

古スペイン語の公証文書にはふつう発行年と発行地が記載されているが、それらが欠如している文書も多い。そこで文書に観察される言語的な特徴から年代と場所を推定するための計量言語学的方法の開発が望まれる。スペインの研究グループが作成した約 1500 の古文書の記録を、統計処理が可能な形式に置き換え、統計学研究を参照しながら、自らが開発した言語処理プログラムを使い、実年代・距離と推定年代・距離との差を短縮することができた成果は大きい。このことは、文書の年代推定・場所推定の実用的な価値が高いということだけでなく、言語特徴と年代・場所の関係について定量的に通時的・共時的に記述し考察したことで言語史研究に多大な貢献を成し遂げたことを意味する。

一般に言語史研究で扱われる資料は完全ではなく、文書中に言語特徴が現れないことが、当該の年代・地点でその言語特徴が存在しない、ということにはならない。その欠測値を補うための統計的な手法として有効であるスムージング法を使うことによって絶対頻度を平滑化頻度に変換した。本研究では、この平滑化を時間（年代）と空間（距離）にそれぞれ個別に適用しただけでなく、両者を総合した変数の平滑化も可能にした。

分析方法には 2 文字連続の機械的分析（JS 情報量）と、表記・音韻・形態の文献学的特徴の言語的分析を行い（ナイーブベイズによる文書分類）、両者の分析結果を比較したことによって方法の偏りを避け、研究のスケールを大きくした。2 文字連続分析は言語の違いを超えた言語処理研究で一般に行われる方法であり、文献学的特徴の分析は個別言語の固有な特性を追究する方法である。本研究から得られた知見によれば、前者のほうが後者よりも推定精度が高い、ということである。これは素性数の違い（900: 300）による、と推測しているが、この推測は妥当と思われる。

本研究の価値は、高度な統計学的手法（とくに確率論の適用）と独自の言語データ処理プログラムを膨大な古文書の分析に適用し、年代推定・時代推定の多くの実験によってその精度を高めたことにある（年代の平均誤差は 14 年；場所の平均誤差は 107km）。多くの文献学的特徴(115)の年代推移の平滑曲線によるグラフによる視覚化は注目に値する。また、具体例やグラフ、そして付録による方法論の説明が丁寧であることもよい。分析プログラムに付された適切なコメント文がコードの理解を助けている。

一方、審査委員からは次の疑問点と改善点が挙げられた。

(1) 研究では年代・場所の記載がある文書について、それらを統計的に推定することにより、記載された年代・距離と推定された年代・距離の残差を最小化する方法の開発に力を注いだ。本論文は理論的背景と方法を論じることで完結している。しかし、資料体には年代・場所の記載がない文書が多数あるので、高い推定精度が得られた方法を実際に適用することが望まれる。究極の目的である年代・場所の記載がない文書の年代・場所推定は言語研究に限らず、歴史学や法律学などの他分野にもインパクトがあるはずである。

(2) 年代場所推定において高い精度を示した理由は何か。それに寄与した具体的な言語的特徴が挙げられれば、言語史研究への重要な示唆になる。さらに2文字連続でも言語特徴として考察する価値がある。たとえば、語頭・語末子音は年代差・地域差を示す重要な特徴である。それらが言語的特徴であるならば、公証文書に限らず文学作品の年代推定・場所推定にも応用が可能になると思われる。

(3) 古文書の場所推定の単位として現代の自治州を使う理由は何か？これはたとえば人口密度の高い現代の首都圏 Madrid を中世スペイン語の場所推定の単位として独立させることはよいか？また、Castilla y León という広大な地域を単位とすることはスペイン語史からもスペイン語方言学からも認められない。代わりに旧来の「地方」(Región)を使えば、歴史的にも言語的にも自然な単位となると思われる。

(4) 2文字連続分析で大文字をすべて小文字に変換したことと、「&」(et, e)を除去したことの理由は何か？文字史の観点から両者は古文書の重要な特徴である。

(5) 文献学的特徴の分類に再考が必要ではないか？「形態統語的特徴」とあるが統語的特徴はほとんど扱っていない(例外：que lo non mandó - que non lo mandó；関係節における未来の法)。また、同一語の変異(例：entonces, estonces)と、語の変異(juntamente / ensemble)を区別すべきである。

(6) 他の属性の可能性として文字種の変異(ゴシック体、ユマニスト体の下位区分)を今後の展開に含めるとよい。

(7) 分析結果を単に数字・グラフに語らせるのではなく、自分の言葉によって語るべきである。

(8) 論文全体の章の構成において量的な配分のバランスが悪い。多くの文献学的特徴のリストやグラフは巻末にまとめるべきである。

しかし、今後の研究では、審査委員によるこれらの指摘を考慮した改善と、

次の執筆者自身が述べている課題への取り組みもなされる予定である。

- (1) 社会言語学的属性を考慮する
- (2) 新たな素性を導入する
- (3) 平滑化パラメータを適応的に調整する
- (4) 素性間の相関を考慮する

審査委員全員は、以上の疑問点、改善点、課題が残されているものの、完結した本論文自体の価値をそれらが損なうことはない、と判断し、むしろ本研究が古スペイン語文献学研究において独創的な新分野を開拓したことを高く評価する。

したがって、本審査委員会は博士（学術）の学位を授与するにふさわしいものと認定する。