

氏 名 上田 澄江

学位（専攻分野） 博士（学術）

学位記番号 総研大乙第 199 号

学位授与の日付 平成 22 年 3 月 24 日

学位授与の要件 学位規則第 6 条第 2 項該当

学位論文題目 人文社会データへの統計数理的アプローチ—古代社会の家系
図の復元と人口推定および語順規則による言語の分類—

論文審査委員 主 査 教授 土谷 隆
教授 種村 正美
教授 中村 隆
准教授 牧野 久実（鎌倉女子大学）
名誉教授 伊藤 栄明
名誉教授 田邊 國士

論文内容の要旨

本論文では(1) 古代メソポタミアの遺跡に関連した考古学データ、(2) 世界の130言語の19項目にわたる語順表という2つの人文社会データを解析している。これらの人文社会データを最適化や多変量解析などの統計数理的手法を用いて新たな視点から解析し、従前の手法では得られなかった有用な知見を得ることが研究の主たる目的である。その過程を通じて統計数理的方法論が、人文社会データの解析において今後さらに重要な役割を果たしうることが示されている。

第I部の土台となっているヌジ人名史料は、古代メソポタミアのヌジ遺跡から出土した粘土板文書に記載されている個人名をまとめた書物である。この粘土板文書は主に土地など財産に関する契約文書であり、その数は約4000枚である。本研究では人名史料に基づき内容を再現するデータベースを作成し、このデータベースを用いて、契約文書に記載される個人名とその親族関係の情報から家系図を自動的に作成する計算アルゴリズムを提案している。復元された家系図と同一文書にどの人物と一緒に現れるかという情報を組み合わせることにより、文書に記載される個人名の生存期間、文書の成立年を推定することができ、これらを手がかりにヌジの人口の推定を試みている。

1章では、ヌジ人名史料の出典であるヌジ遺跡とヌジ文書、そしてヌジ人名史料の特徴について説明している。ヌジ人名史料は、ヌジ遺跡の粘土板文書から登場する人物だけを抜き出した史料であり、本研究ではまず人名史料をコンピュータで扱えるようにデータベース化した。ヌジ文書には契約の当事者の名前を親の名前と共に記載するという特徴があり、親子2代の家系図を容易に想定できる場合が多い。この段階で想定される家系図を基本家系図と呼ぶことにする。2章ではこの基本家系図に基づくより広範な家系図の再構築が試みられている。

ヌジ人名史料を読み解いて家系図を復元したことによって、文書に記載される人物に限定されるが、同時期に生存していた人数を推定したり、文書の成立年を推定することが可能になった。3章では、得られた最大の家系図6代16人のTehip-tillaの家系図を基準として、他の家系図と照らし合わせて、各々の家系が存在した時代を推定することを試みている。具体的には、Tehip-tillaの家系図のメンバーと文書を介して関連する人物をたどり、その人物の属する家系図を特定することにより、それぞれの家系が存在した時代を推定する。

4章では、凸最適化問題の一つである凸2次計画問題を用いて個々の人物の生誕年と死亡年、文書の成立年を推定している。家系図から得られる親子関係を用いて親子の年齢差を仮定し、また契約に関わるのに相応しい年齢を想定し、文書に記載される人物の年齢を与えて制約条件とする。また、予め仮定した寿命と個々の人物の寿命との差が最小になるように目的関数を設定する。結果として22,301変数と39,311制約式をもつ凸2次計画問題が得られ、主双対内点法によりこの問題を解いて個々の人物

の生誕年、死亡年、文書の成立年の推定値を求めることができる。得られた推定値は、3章で得られた結果が世代間の比較であるのに対して、人物一人ひとりの年齢を推定するものであり、より詳細な情報を含む。これらに基づき文書に記載されるヌジ名の存在した年代や人口動態、契約文書の変遷等の様々な情報が得られる。特に文書に登場する人物は8,000～9000人が同時期に生きていたと思われることが明らかになる。この結果は3章で得られた結果とも整合性のあるものである。

第5章では第4章までに得られた結果を検討してヌジには最低20000人程度が居住していたと結論し、そして考古学的に得られている知見とも合わせて集落の個数や領域の面積等についても考察している。さらに、推定された契約文書の成立年代による契約文書の種類の変遷についても検討を行っている。

第II部でとり上げた世界の130言語の語順の表の解析では、言語学者の角田太作博士によって調べられた語順表が土台となっている。言語類型論は世界の諸言語を比較することにより、諸言語間で普遍的な法則性と異同を調べ、各言語に固有の特徴をとらえようとするものである。角田は多くの言語の比較に努め、世界の130言語の語順を19項目に亘って語順表にまとめた。語順表には、1つの語順項目に対して2つ以上の語順がある言語が多々存在する。その場合には、その言語でより普通と思われる語順が先に示されている。複数の語順が使用される割合は個々の言語に特有なものであると思われるが、それぞれの言語に対して適切な数値を割り当てることは困難である。そこで本論文では、すべての言語に共通のある妥当と思われる基準を設け、それに沿った機械的に語順の数値化を試みている。語順規則による言語の解析はこのように様々な言語に対して個々に斟酌することなく語順を数値化して得られた結果である。

6章では、語順表の数値化の方法について述べ、続いて130言語のうちアジア・ヨーロッパ地域に限定した50言語の語順に対して階層クラスタ分析を行った結果について述べている。

7章では、世界の130言語の語順に対して階層クラスタ分析と主成分分析を試み、階層的クラスタ分析の結果に基づいてAICによる分割表における変数選択の方法を用いて言語を7つのグループに分類し、さらに言語の分類に最適な項目について検討している。解析の結果として、言語類型論の分野で世界の言語を2分割する指標として重要視されてきた“名詞と側置詞”の語順について、従来は、前置詞言語と後置詞言語と側置詞を持たない言語の3つに分類されてきたが、後者2つが一つのグループにまとめられることが明らかになった。さらに“名詞と側置詞”の語順に次ぐ、言語分類上の第2の指標を見出すことを試みている。そのために階層クラスタ分析で得られた便宜的なグループ分割を用い、それらのグループを構成するために必要な項目を選択することを試み、また逆に選択された項目を用いて構成されたグループの特徴について考察している。これらには分割表データにおける変数選択の方法を用いた。結果は、第2の指標として“数詞と名詞”の語順の項目が選ばれる。“名詞と側置詞”および“数詞と名詞”の語順項目による4グループ分割は、日本語、タイ語、英語を明確に分けるものになっている。

博士論文の審査結果の要旨

○論文の概要

申請者の論文は人文社会データへの統計数理的アプローチを試みたもので、2部7章からなる。第1部では古代メソポタミアの遺跡に関連した考古学データを解析し、第2部では世界の言語を語順規則の観点から分類している。これら2つのデータを統計や最適化などを用いて新たな視点から解析し、統計数理的方法論が人文社会データ解析において有用な役割を果たしうることを示している。

第1部では、古代メソポタミア、紀元前15世紀から14世紀に存在した都市ヌジ（現在のイラク・キルクーク南方）から出土した、4000枚余りの粘土板に記載されたデータを解析している。これらの粘土板は、当時の社会契約に関する記録であるが、特に現れる人名の観点からまとめたものが、ヌジ人名史料として刊行されている。本研究は、ヌジ人名史料をデータベース化して解析したものであり、このデータベースに断片的に現れる家系(family tree)を統合して家系を復元するアルゴリズムを考案し、さらに、数千の復元家系を、各家系が登場する文書の共有関係を解析することと、最適化アルゴリズムの適用により時間軸上に矛盾なく配置することによって、文書に現れる各人物の共時的関係の推移や社会のあり方について推測し考察した。

第1章では、基礎となるヌジ人名資料や人名資料の元となる粘土板に記された契約、そしてヌジをとりまく社会情勢等の背景が紹介される。

第2章では、家系の復元アルゴリズムを提案し、それを適用した結果、人名資料には10740人が6187家系に現れること、全6世代16人からなる家系が一番大きいこと、そして、2世代2人の家系および1世代1人の単独家系がその内大半を占めることなどが述べられている。

第3章では、家系間の文書の共有関係を解析し、文書に現れる各人物の共時的関係（どの人物が同じ時代に生存していたか）の推定を行っている。2家系で世代別に共有する文書の数からコーホートのマッチングをはかり、既存研究から6世代とされる各世代の人数を見積もり、第3世代の人数が一番多く、約8000人から1万人弱であるという結果を得た。

第4章では、凸2次計画法を活用して、家系を時間軸上に矛盾なく配置することが試みられている。各個人の生誕年と死亡年、そして文書の成立年を変数として、復元された家系の親子関係等から成り立つべき不等式関係を制約として組み込み、家系の配置問題を約30000変数の凸2次最適化問題として定式化し、各個人の生誕年と死亡年を矛盾なく時間軸上に配置することに成功している。これにより、文書の成立年代や文書に現れる人物の各時点での生存数について、時間的な推移を定量的に検討することが可能となる。その結果、文書に現れる人物の中で、最大で8000人強が同時に生存していた時期があること、そして契約文書が一番多く作成された時期はその後であることなどが知見として得られた。

これは、第3章で得られた結果と整合性のあるものである。

第5章では、これらの結果を検討して、ヌジには最低 20000 人程度が居住していたと結論し、そして考古学的に得られている知見とも合わせて、集落の個数や領域の面積等についても考察している。さらに、推定された契約文書の成立年代による契約文書の種類の変遷についても簡単な検討を行っている。

第2部では語順規則による世界の言語の分類を取り扱っており、第6章、第7章の2章からなる。解析データの基礎となるのは、角田によって作成された、世界の130の言語の語順表である。本論文では、語順表をコーディングした上で、クラスタ分析や、主成分分析、AICによる分割表選択といった統計数理的手法を用いて解析し、新しい知見を得ている。

第6章では、角田が作成した19項目からなる言語の語順表について紹介し、そのコーディングの方法を述べ、階層的クラスタ分析による予備的な解析と検討が行われている。

第7章では、まず、コーディングされた語順表に主成分分析を適用し、日本語、英語、タイ語が典型的な言語として挙げられることを観察している。次に、階層的クラスタ分析の結果に基づいてAICによる分割表における変数選択の方法を用いて、言語を7つのグループに分類し、さらに言語の分類に最適な項目について検討している。そして、選択された項目である「名詞と側置詞の順序」と「数詞と名詞の順序」を用い、一番適切な言語の分類を求めた。これらの解析の結果として、(1)従来は、言語が前置詞言語と後置詞言語と側置詞を持たない言語の3つに分類されてきたが、後者2つが一つのグループにまとめられること、(2)「名詞と側置詞の順序」と「数詞と名詞の順序」をキーとする分類項目とすることにより、言語を(北米、豪の言語を多く含むグループ)、(日本語、南米地域の言語を含むグループ)、(タイ語を含むグループ)、(英語を含むグループ、中米地域の言語を含むグループ)の4つのグループに分類することができることが明らかにされている。

○論文の評価

第1部で得られた成果はヌジ社会についてさらに考察する上で重要な手がかりを与える基礎的データとしての価値があるばかりではなく、この種のデータを解析する上での新しい方法論を与えたものとして評価できる。また第2部では、世界の言語を、語順規則という視点から統計数理的方法に基づいて解析し、興味深い知見を得ている。この成果は、言語分類という問題への新しい接近法としても十分に評価できる。これらの成果は、統計数理の人文社会データ解析における有効性を実証するものとして説得力がある。なお、第1部に関連しては申請者を筆頭著者とする学術論文が1編出版されており、また、新しい成果について投稿準備中である。また、第2部に関連しては5編の論文(うち3編が筆頭著者)が出版されている。