

論文の内容の要旨

論文題目 内容と文体の情報を用いた計量テキスト分析
政治テキストへの応用を中心として

氏名 鈴木 崇史

本研究では、データベースの整備と自然言語処理技術の進展を利用し、体系的と繊細さ、両者を実現する、「計量テキスト分析」を人文科学、社会科学分野に提出した。これを、政治テキストに適用することで、政治研究に新たな方法論を提示するとともに、日本政治像の構築に資する実証的知見と含意を得た。

現在、入手可能な情報の爆発的な増大と大規模データ処理技術の進展のなかで、学術研究は大きなパラダイム転換期を迎えている。Webの発達とデータベースの整備により、大量の情報が容易に入手可能となり、新たな情報資源の活用方法を検討することが、あらゆる学術分野で急務の課題となっている。テキストデータは、そのような情報資源の中核的位置を占め、大量かつ多様なテキストの蓄積と自然言語処理技術の進展は、テキスト研究の射程を大幅に拡張しつつある。

このような中、人文科学、社会科学分野においても、より大量のテキストデータを、より体系的な分析手法とともに、積極的に活用することが重要な課題となりつつある。同時に、目的に即した分析上の配慮、専門知識の適切な付与、繊細なテキスト情報の利用と解釈、これらを実現する、計量テキスト分析の分析枠組を定式化し、またこれを実践することが要請される。

一方、政治研究においては、これまで世論調査、選挙結果などが実証分析のためのデータの中心であり、テキストデータは周辺的な位置にあった。日本の政治研究においては、そもそも、データ

に基づく実証分析自体が、相対的に軽視されてきており、本研究で対象とする、日本外交研究、総理大臣研究の領域では、とりわけ、データにもとづく実証分析は、従来、試みられることが少なかった。

以上の背景のもと、本研究では、人文科学、社会科学分野での応用を念頭におき、体系性と繊細さ、両者を実現する、「計量テキスト分析」の分析枠組を定式化するとともに、これを、政治テキストに適用し、政治研究に対して新たなデータと分析手法を提供、日本政治研究に資する実証的知見と含意を得ることとした。Webの発達により入力、出力関係が不明確、かつ、インタラクションを伴う、さまざまな種類のテキストが産出されており、このような人文科学、社会科学分野での応用研究は、新たなテキスト資源利用のためのケーススタディとしても意義をもつものである。

本研究は、IV部8章から構成される。第I部が序論、第II部と第III部が本論、第IV部が結論にあたる。第I部では、研究の背景・目的、関連研究、分析枠組に関する事項を記述した。第1章では、入手可能な情報の増大の中、人文科学、社会科学分野に、体系性と繊細さ、両者を実現する「計量テキスト分析」を提出する必要性について論じ、また、これを政治テキストに応用する意義について記述した。

第2章では、本研究の対象、すなわち、日本政治研究と、手法、すなわち、計量テキスト分析にわけて関連研究を概略した。とりわけ、人文科学、社会科学分野における既存の計量テキスト分析が、内容分析、文体分析の両者に二分されていることを指摘し、両者を統合する分析枠組の必要性を論じた。

これをうけ、第3章では、まず、内容分析、文体分析、両者を統合する「計量テキスト分析」の一般的分析枠組を提出した。既存のテキスト理論との関係を整理し、計量テキスト分析の定義を行なった後、その対象、目的、焦点となる情報、計量される特徴、方法論、以上の点を整理した。つづいて、これをもとに、本研究第4章から第7章における実質的な分析枠組を記述した。第4章から第7章で用いるデータは、1945年から2008年まで、29人30人155本の総理大臣国会演説であり、全語彙の延べ語数は、541944語、異なり語数は、11797語である。ただし、形態素解析にはChaSenを用いた。

第II部では、日本政府の対外認識を検討した。まず、第4章では、東西問題、南北問題に関する日本政府の認識を検討するため、仮説—演繹的分析と内容語および共起カテゴリーの計量という既存の内容分析研究の枠組を踏襲した。結果、日本政府の対外認識が東西問題に関しては世界史の展開と先進国の政策理念に従う一方、南北問題に関してはより独自性の高い特徴を示すことを指摘した。これをもとに、既存の日本外交研究がもつ認識枠組を指摘し、多角的な観点から多用な要素を分析することの重要性を指摘した。

第5章では、より積極的に、計量テキスト分析によって、どのような日本外交像が描けるのか、この点を検討した。より具体的には、戦後日本外交における地域的関心の全体像、その相関構造と最重要の二国間関係である日米関係について、地域名／国名のカテゴリー化と多変量解析、複合名詞のカテゴリー化、係り受け解析と機械学習による特徴量選択を利用し、多角的な観点から、多様な要素を考慮にいたした分析を実現した。結果、戦後日本外交が対欧米外交の軸から対アジア外交の

軸へと漸進的に変化しており、また、冷戦崩壊後、日米関係が「実体化」している、すなわち、文化／象徴的価値を含む包括的な関係から、安全保障、経済に特化し、日米関係と国際協調が併置され、さらに、より実務的なレトリックとともに言及される、以上の知見を提出した。これらの知見は、近年の日本外交モデルと関心を共有しつつ、これに新たな含意を付け加えるものである。本研究第Ⅱ部は、長期的な枠組と多角的な観点から、多様な要素を考慮にいられて、日本外交の特徴を論ずることの重要性を指摘し、また、このような観点から実証的知見と含意を積み重ねる上で、総理演説の計量テキスト分析が有効であることを指摘するものである。

第Ⅲ部では、総理大臣のスピーチスタイル、政治スタイルを検討した。まず、第6章では、名詞の分布特徴量、すなわち、語彙量、語彙の多様性、語彙の偏りに注目することで、施政方針演説と所信表明演説の差異、中曽根、小泉がもたらした演説の変革を検討した。結果、施政方針演説と所信表明演説は、その役割の差異を反映して、語彙量は異なるものの、語彙多様性や語彙の偏りにおいては差異がなく、また、中曽根は、演説の量を増やすことで、小泉は多様な語彙を用い、多様なエピソードを演説に盛り込むことで、独自のスピーチスタイル、政治スタイルを示していたことを指摘した。

第7章では、機能的特徴量に注目し、社会言語学的観点を積極的に導入することで、総理大臣のスピーチスタイル、政治スタイルをより詳細に分析した。まず、第7章前半では、文長の計量、機能語の相対頻度分布を用いた多変量解析という既存の文体分析の枠組を踏襲し、総理演説に影響を与える主要な要因を探索的に分析した。結果、時代による日本語表現一般の変化が、総理演説における機能的特徴量に強く影響しているものの、総理大臣個人の影響も観察されることが示された。そこで、第7章後半では、時代要因を一定程度統制した後、機能語の相対頻度分布を用い、機械学習による特徴量選択を利用することで、中曽根、小泉に特徴的な機能語を抽出した。これにより、まさにそれぞれの総理大臣のスピーチスタイルの特徴から、まさにそれぞれの総理大臣の政治スタイルの特徴を分析することを可能とした。より具体的には、中曽根は、規範表現、問いかけ表現、一人称複数の多用によって特徴づけられ、「国民説得型」政治スタイルをもつ一方、小泉は、簡潔で短い文、日常的な文末表現によって特徴づけられ、「国民対話型」政治スタイルをもつという知見を提出した。一連の分析手法は、未だ十分な展開例のない計算社会言語学の領域を切り開くものでもある。本研究第Ⅲ部は、分布特徴量、機能的特徴量に注目することで、総理演説から、総理大臣個人のスピーチスタイル、政治スタイルを分析可能であることを示すものであり、新たな総理大臣像の構築に資する実証的知見と含意を得る上で、総理演説の計量テキスト分析が有効であることを指摘するものである。

第Ⅳ部では、本研究全体の総括を行った。第8章で、本研究全体を要約するとともに、今後の課題を、個別論点に関する課題、計量テキスト分析の今後の展望、以上二点にわけて記述した。