

京都女子大学現代社会学部の卒業論文題目の テキストマイニングによる分析

Analysis of Thesis Title at Faculty for the Study of Contemporary Society, Kyoto Women's University

奥平 美貴†
Miki Okudaira

丸野 由希‡
Yuki Maruno

1. はじめに

京都女子大学現代社会学部（以下、現代社会学部）では、多様化、多層化、複雑化した「現代社会」に関する幅広い知識を身につけ、社会問題を多角的に認識する方法を学ぶ。ダブル・クラスター制のもと、学びの領域として「人間・文化」「家族・地域社会」「政治・公共政策」「経済・経営」「情報・環境」の 5 つのクラスター（表 1）から興味や関心のある 2 つのクラスターを選び専門分野に学び進め、その集大成として卒業論文に取り組む。

表 1 現代社会学部の 5 つのクラスター

クラスター	説明
人間・文化	人間の精神・こころ・身体・生命のあり方やメディア・宗教・スポーツなどの文化的営みを、多彩な切り口で探求する。
家族・地域社会	家族・ジェンダー・地域社会を社会学的なアプローチで探求し、次代の新しい家族像・地域社会のあり方を探る。
政治・公共政策	政治学・行政学・法学分野の科目をバランスよく学び、良き市民として生きる力を養う。公務員志望者にも対応。
経済・経営	経済学・経営学・会計学を包括的に学ぶことによって、ビジネス社会で必要とされる知識や考え方、提案力を培う。
情報・環境	高度情報社会を支える情報科学・情報工学の知識とスキルを修得し、あわせて学際的知識を活かして環境問題を考える力を養う。

本研究では、現代社会学部創立から平成 26 年度までの計 12 年分の卒業論文題目データを集積し、分析を行った。探索的分析や推測分析などさまざまな分析方法があるなかで、頻度分析とネットワーク分析を、年度ごと、指導教員の専門分野ごとに行った。指導教員の専門分野はクラスターと同様に「人間・文化」「家族・地域社会」「政治・公共政策」「経済・経営」「情報・環境」の 5 つに分類した。専門分野ごとの分析には、ゼミの指導教員の情報が入手できた平成 21 年度～平成 26 年度までのデータのみを使用した。

2. 分析データ・手法

2.1 使用したデータ

本研究では、京都女子大学現代社会学部の学生の卒業論文の題目、副題、指導教員の 3 項目をデータとして使用した。分析対象は、現代社会学部が設置されてからはじめての卒業生となる平成 15 年度から平成 26 年度までの計 12 年分である。指導教員については、はじめの数年間のデータが残っていなかったことから、平成 21 年度から平成 26 年度までの計 6 年分のデータのみを用いた。表 2 は年度ごとの卒業論文提出人数である。

表 2 年度ごとの卒業論文提出人数

年	人数	年	人数
平成 15 年度	274 人	平成 21 年度	269 人
平成 16 年度	253 人	平成 22 年度	293 人
平成 17 年度	232 人	平成 23 年度	271 人
平成 18 年度	253 人	平成 24 年度	256 人
平成 19 年度	257 人	平成 25 年度	254 人
平成 20 年度	256 人	平成 26 年度	274 人

2.2 形態素解析による品詞情報の抽出

本研究では、卒業論文の題目、副題に用いられた名詞に着目した。そこで、題目、副題を形態素解析にかけ、名詞のみを抽出した。形態素解析とは、文を単語単位に切り分け、品詞の情報などを加えることを指す[1]。例えば、「すももももももものうち」という文を形態素解析にかけると以下のようになる（図 1）。

すもも 名詞, 一般,*,,*, すもも, スモモ, スモモ
も 助詞, 係助詞,*,,*, も, モ, モ
もも 名詞, 一般,*,,*, もも, モモ, モモ
も 助詞, 係助詞,*,,*, も, モ, モ
もも 名詞, 一般,*,,*, もも, モモ, モモ
の 助詞, 連体化,*,,*, の, ノ, ノ
うち 名詞, 非自立, 副詞可能,*,,*, うち, ウチ, ウチ

図 1 MeCab の解析

形態素解析エンジンはいくつか公開されているが、本研究では MeCab を用いた。MeCab（和布蕉）は、京都大学情報科学研究科と NTT コミュニケーション科学基礎研究所の共同研究ユニットプロジェクトを通じて開発されたオープンソース形態素解析エンジンである[2]

†株式会社富士通四国インフォテック, FUJITSU SHIKOKU INFORTEC LIMITED

‡京都女子大学, Kyoto Women's University

2.3 頻度分析を可視化するためのワードクラウド

頻度分析は、ある単語がテキストの中でどのくらいの頻度で出現するのかを分析することである。テキスト中の単語の出現頻度は数字を並べた表で表現されることが多いが、全体の傾向を直感的に捉えにくい。そこで、文章中で出現頻度が高い単語を複数選び出し、その頻度に応じた大きさで図示する手法であるワードクラウドを用いた[3]。本研究では、形態素エンジン MeCab を R から使うための R のパッケージである RMeCab と、ワードクラウドの手法を使って単語の頻度分析を視覚的に表現するための R のパッケージ wordcloud を用いて、年度ごと、専門分野ごとに頻度分析を行った。

2.4 ネットワーク分析

ネットワーク分析は、社会学や通信ネットワークなどの分野で多く用いられており、グラフ理論を基礎に置いている。グラフ理論では、ものとの関係を頂点同士結ぶ線分によって表した図形をグラフと呼び、頂点の間に向きがあるグラフを有向グラフ、向きがないものを無向グラフと言う[1]。テキスト解析において、特定の複数の単語の組が同じテキストの中に現れることを「共起」と呼び、共起関係について調べることは重要な分析方法である。ネットワーク分析では、二つの語の共起関係をネットワーク図と呼ばれるグラフを使って視覚的に表現することで、分析対象のテキストについて有用な知見を得ることができる。本研究では、R のパッケージである igraph を用いて、4 年ごとに区切った場合、指導教員の専門分野ごとの共起関係をネットワークマップで表し、ネットワーク分析を行った。MeCab の辞書を用いて形態素解析を行うと、例えば「経済学」などの言葉は「経済」と「学」に分かれるが、このままでネットワーク分析を行うと「学」の文字が「経済」、「経営」、「政治」などを結びつけるハブとなってしまう、正しい関係性が見失われてしまう。そこで「経済学」などを一つの名詞として辞書に登録することで問題を回避した。

3. 頻度分析

3.1 年度ごとにみた頻度の変化

年度に題目、副題を RMeCab で形態素解析し、名詞の出現回数を数えた。このとき、1 文字の単語は意味をなさないものが多かったため除外した。

年度ごとにみた場合、出現回数が上位の語彙にはあまり変化がみられず、ほとんどの年度で「女性」「日本」「社会」「問題」などの言葉が上位を占めた。

3.2 ワードクラウドで表現した頻度の変化

出現頻度の少ない名詞にも着目するために、ワードクラウドを用いて分析を行った結果を図 3.1-3.2 に示す。文字の大きいものほど出現頻度が高くなっている。

平成 25 年度(図 3.1)には、「原発」「被害」「福島」「災害」などの文字がみられる。これらの単語が使われている題目、副題をもとのデータから検索すると、以下のようない題目、副題がみられた。

- 原発災害後の長期汚染地域における家屋環境での空間線量率の低減方策に関する研究
- デマ災害における「情報」の重要性～福島原発災害におけるデマ・風評被害をなくすために～



図 3.1 平成 25 年度の場合



図 3.2 平成 16 年度の場合

- 福島原発事故と原子力損害賠償訴訟法を基礎とする関西での被害予測
- 原発災害後の復興における福島県産品販売の現状に関する研究
- 原発ゼロは可能か正しい選択か
- 世代を超えて伝わる戦争～ベトナム枯葉剤被害者を事例に～
- いじめの発生誘因と被害後の長期的影響について

このことから、福島県の問題を題材に卒業論文を執筆した人が多いことが分かる。ゼミを選択するのは論文執筆のおよそ 2 年前からのため、平成 23 年度近辺の出来事に着目すると、平成 23 年 3 月 11 日に東日本大震災が起きている。この影響で、平成 25 年度に「原発」「被害」「福島」「災害」などの言葉が例年より多くみられるのではないかと考える。

次に平成 16 年度(図 3.2)をみてみると、「食品」という文字がみられる。もとのデータから「食品」という単語を検索してみると以下のような題目、副題がみられた。

- サプリメント(栄養補助食品)、必要性の是非
- 食品添加物の必要性～絶対安全を求める消費者～

- 食品添加物とその表示について～食品添加物から見る食の安全性～
- 食品公害事件とこれからの対策

卒業論文執筆 2 年前の平成 14 年度近辺の出来事に着目すると、平成 13 年に食品偽装事件のきっかけにもなった雪印牛肉偽装事件が起きている。この影響で、食品公害事件や食品添加物に関する題目で執筆した人が平成 16 年度に多く、「食品」という言葉が例年より多くみられるのではないかと考える。

3.3 指導教員の専門分野ごとにみた頻度の変化

次に指導教員の専門分野ごとに頻度解析を行った。年度ごとと同様に、題目、副題を RMcCab で形態素解析し、名詞の出現回数の集計を行った。このとき、単語として分類されたもののうち一文字のものは、名詞として意味をなさないものが多かったため除外している。

指導教員の専門分野ごとにみた場合では、それぞれの専門分野の特長が大きく反映された結果となっている。各専門分野にしか出現していない言葉をみてみると、経営・経済では「戦略」「経済」「業界」「企業」、家族・地域社会では、「家族」「子ども」、人間・文化では「現代」「調査」「関連」「大学生」「女子」、情報・環境では「環境」「京都」「エコ」「提案」「観光」「地域」「ため」「情報」、政治・公共政策では「貧困」「政策」「歴史」「制度」となっており、上位の言葉だけでもこれだけの違いがあることがわかった。

3.4 ワードクラウドで表現した頻度の変化

ワードクラウドで表現したものを図 3.3-3.7 に示す。頻度が少なく小さな文字で書かれた言葉に着目と、経営・経済では「マーケティング」「リーダーシップ」「モチベーション」「マネジメント」など経営に関する言葉が多くあげられている。家族・地域社会では、地域を表す「都道府県」「商店街」や家族に関係する「主婦」「母子」「保育」などの言葉が、人間・文化では「ポジティブ」「いじめ」など人に関する言葉があげられている。情報・環境では、「システム」「アプリケーション」「シミュレーション」、政治・公共政策では「格差」や様々な国の名前があらわれ、強く専門分野を表している。このようにワードクラウドで表した図からも、専門分野ごとに多様性にとんでいることがわかる。

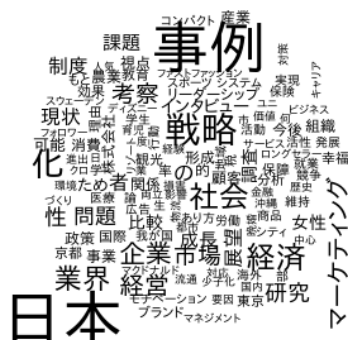


図 3.3 経済・経営



図 3.4 家族・地域社会



図 3.5 人間・文化

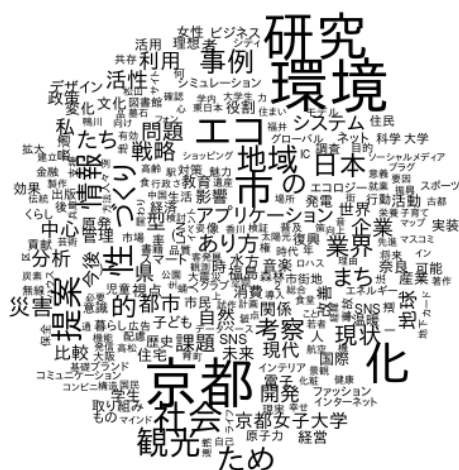


図 3.6 情報・環境



図 3.7 政治・公共政策

4. ネットワーク分析

次に名詞同士の関係性に着目してネットワーク分析を行った

4.1 年度ごとにみた場合

分析結果をより分かりやすくするため 12 年分のデータを，平成 15 年度～平成 18 年度，平成 19 年度～平成 22 年度，平成 23 年度～平成 26 年度の 4 年ごとに区切り，ネットワークマップを作成した。

4 年ごとに出力したネットワークマップから，「女性-就業-継続」というつながりが，平成 15 年度から平成 18 年度が「女性-10-就業-3-継続」に対し，平成 19 年度から平成 22 年度は「女性-15-就業-14-継続」，平成 23 年度から平成 26 年度「女性-17-就業-7-継続」と年々，女性の就業問題について興味をもつ人が増えている。1993 年から 2005 年は就職氷河期とよばれている。また近年では，働く女性も増えてきている。このことから女性の就業状況や就業してから継続していけるのかどうかに関心を持ち，卒業論文で扱っている人が多いのではないかと考える。

4.2 指導教員の専門分野ごとにみた場合

ワードクラウドの時と同様に，ネットワークマップにおいても，それぞれの専門分野に大きく違いがあらわれた。経営・経済では「日本」「経済」「事例」などの言葉が中心になっており，「業界」や「戦略」などさまざまな言葉が飛び交い，経営・経済の特長が顕著にあらわれた。家族・地域社会では「日本」「女性」に多く矢印が向いているが，その次に「ライフ」「子ども」「社会」などに多く矢印が向いており，他の語彙との関係性が高いことを示している。家族の中における子どもや，家庭を持った時の自分のライフコースなどに関心をもつ人が多いことがわかった。また，他の専門分野と比べて「県」や「市」のまわりに様々な地域名があることも，この分野の特長の 1 つといえるだろう。人間・文化では「日本」「社会」「調査」に加え，「医療」の周りにも多く語彙がとりまいており，また全体をみると，「妊娠」「死」「移植」など生殖や生命に関する言葉が多くみられた。このことから人間・文化では，生殖や生命系の分野に関心を持つ人も多いことがわかった。情報・環境では，その名の通り「環境」や情報系の卒業論文を書く上で多い「研究」「提案」の他にも「京都」「観光」などの周りに多く言葉がとりまいており，こちらも顕著に専門分野の特性が表れている。政治・公共

政策では，政治に関する言葉が多くみられた。外国の地名も多いが，「日本」「問題」の周りに多く言葉を取りまいており，日本のことはもちろん，世界のあらゆる問題に関心をもっているのだろう。このことから，同じ現代社会学部であっても，専門分野によって大きな違いが生じていることがわかった。

5. おわりに

現代社会学部の卒業論文の題目の頻度分析，ネットワーク分析を行った結果，現代社会学部の学生は，同じ学部にも関わらず，専門分野ごとに大きな違いがみられ，多様性にとんだ題目で卒業論文を執筆していることがわかった。また，現代社会学部を標榜していることにも見られるとおり，常に新しい問題やその年に大きな問題となったものに関心を持ち，その時代に応じた内容で卒業論文を執筆していることがわかった。今後は，名詞だけではなく他の品詞にも着目し，さらに分析を進めたい。

参考文献

- [1] 金明哲「テキストデータの統計科学入門」岩波書店，2009
- [2] MeCab: Yet Another Part-of-speech and Morphological Analyzer, (mecab.googlecode.com/svn/trunk/mecab/doc/index.html) 2016/01/06 閲覧
- [3] 「ワードクラウド（ワードクラウド）とは- コトバンク」， (<https://kotobank.jp/word/ワードクラウド-674221>) 2016/01/12 閲覧
- [4] 奥平美貴「京都女子大学現代社会学部の卒業論文題目の分析」現代社会学部卒業論文，2015