

計量テキスト分析による感情抽出

— 国会会議録と感情辞書を用いた分析の課題 —

○関西学院大学

中野康人

1. 報告の目的

本報告の目的は、テキストデータを計量的に分析することによってデータ中に含まれる「感情」を抽出する試みを紹介し、その課題を整理することにある。テキストデータを情報源とする分析の手法や技術は近年大きく発展しており、社会学的な分析の対象として、蓄積されたテキストデータを利用する研究事例も広がっている。しかしながら、記述内容つまりはそこに含まれる意味・意識・感情などの測定は依然として質的な解釈に大きく依存している。特定の感情がテキスト中に含まれる量を集計し、時系列的变化や討議内容・発話者の属性などとの関係を探索的に記述する事例を紹介する。そのうえで、分析において重要な役割を果たす感情辞書の整備を中心に、その課題を整理する。

2. 社会調査データとしてのテキスト：国会会議録の分析

標準的な調査票による社会調査が困難になる一方で、テキストデータを情報源とする分析の手法や技術は近年大きく発展している。近刊の『社会学評論』では「テキストマイニングをめぐる方法論とメタ方法論」と題して特集が組まれている（樋口 2017 など）。報告者も、これまで調査票調査の自由記述や新聞記事といったテキストデータの分析に取り組んできた（中野 2010）。

テキスト内容の要約や文書の分類を志向するもの。予め単語（群）を特定の意味・概念を表象するものとしてコード化し、データに含まれるそれら単語（群）を分析するもの。データ中の単語間の関連から、そこに含まれるトピック・意味内容を抽出するもの。分析の方向性は各種ある。就中、ここで注目するのはテキストデータから、その書き手や内容の発話者が持つ感情や意識を測定する分析である。そこで必要になるのが、言葉と特定の感情・意識を紐付ける「辞書」である。本報告では、既存の複数の辞書を分析に適用し、結果の比較検討をおこなう。

ここで分析の対象となるのは、国会の会議録である。日本における国会の会議録は、1947年の第一回から今日に至るまで、国立国会図書館の検索システムで公開されている (<http://kokkai.ndl.go.jp>)。発話者の発言そのままの記録ではないが、70年にわたる国政の話題が、時間や発話者の記録とともに蓄積・公開されている。社会を通時的に分析する一つの窓口として有用なものである。また、同様の会議録は他国の国会でも公開されている事例が多々あり、国際比較分析の情報源としても利用できる。

3. 課題

本報告では、複数の感情辞書による分析結果を比較し、分析の安定性を確かめる。また、数十年の過去に遡る通時的なデータを分析する際の課題、共通の内容を国際比較する際の課題を提示し、発展的利用の方向性を探る。

【文献】

樋口 耕一, 2017, 「計量テキスト分析および KH Coder の利用状況と展望」, 『社会学評論』 68(3):334-349.

中野康人, 2010, 「読者投稿の記述的計量テキスト分析 — 「声」と「気流」 —」, 『関西学院大学先端社会研究所紀要』, 2:43-57.