

| | |
|--------------|---|
| Title | 計量テキスト分析の方法と実践 |
| Author(s) | 樋口, 耕一 |
| Citation | 大阪大学, 2005, 博士論文 |
| Version Type | |
| URL | https://hdl.handle.net/11094/45737 |
| rights | |
| Note | 著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。 |

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏 名 **樋 口 耕 一**

博士の専攻分野の名称 **博 士 (人間科学)**

学 位 記 番 号 **第 1 9 1 4 5 号**

学 位 授 与 年 月 日 **平成 17 年 3 月 25 日**

学 位 授 与 の 要 件 **学位規則第 4 条第 1 項該当**

人間科学研究科人間科学専攻

学 位 論 文 名 **計量テキスト分析の方法と実践**

論 文 審 査 委 員 **(主査)**

教 授 川 端 亮

(副査)

教 授 直 井 優 助 教授 吉 川 徹

論 文 内 容 の 要 旨

社会調査が扱うデータには大きく分けて量的データと質的データがある。このうち文章・画像・音声・映像といった質的データを分析する方法については、標準的な分析の方法が十分に確立されているとは言い難い。そのため、これらの質的データを扱う場合には、分析方法を検討する段階において、量的データの場合よりも長い時間をかけて、注意深く、慎重に方針を探らねばならないのが現状ではないだろうか。また実際に分析を行う段階においても、量的データの場合には必要な機能が一通り詰め込まれた統計パッケージと呼ばれる分析用プログラムが普及し、容易に分析が行えるようになったが、質的データの場合にはそうはいかない。こういった状況に、僅かなりとも、変化を引き起こすことができればという希望のもとに、本稿では質的データの中でもテキスト型（文章型）データに注目し、これを分析するための新たな方法を提案する。

以上のような目的のもとに第 1 章では先行研究を振り返り、本稿全体を通して達成すべき課題を明らかにする。ここでは内容分析（content analysis）の歴史を振り返り、内容分析に学ぶべき点を検討する。内容分析の方法に関する議論の中でもとりわけ、テキスト型データのような質的データを分析する際に、量的方法と質的方法とをいかに併用すべきか、またコンピュータをいかに利用すべきかといった点は、新たな方法を考える際にも重要である。次に、こういった点に関する内容分析の考え方に良く合致したものとして、計量テキスト分析という方法を本稿の直接の先行研究として取り上げ、現在の計量テキスト分析に欠けている点を検討する。この結果として、(i) 量的分析の結果が重視される場合により適し、(ii) 大量のテキスト型データを扱うことができ、(iii) 内容分析の分野における蓄積を活かした分析の方法・手順を提案することで、計量テキスト分析を拡充することを本稿の課題とする。

この課題に沿って第 2 章では、テキスト型データを分析するための新たな分析の方法・手順を提案する。従来の内容分析においては、テキスト型データを計量的に分析する際、Correlational アプローチか Dictionary-based アプローチのうち、いずれかが用いられることが多かった。前者は多変量解析の応用、例えば、クラスター分析を用いて頻繁に同じ文書の中にあらわれる言葉のグループを見つけ出すといった方法で、データ中の主題を探索するアプローチである。それに対して後者のアプローチでは、分析者の指定した基準にそって言葉や文書が分類され、計量的な分析が行われる。この章ではこれらのアプローチを検討して、それぞれに一長一短を持つこれら 2 つを互いに補い合う形で接合したアプローチを提案し、その実現に必要な分析用プログラム「KH Coder」を作製する。そして、この接合アプローチならびに KH Coder を用いた実際の分析例を示した上で、従来のアプローチに対する本アプローチの有効

性について検討を加える。

第3章では、KH Coderを利用して接合アプローチにもとづく分析を行う手順を、チュートリアルで示す。ここで言うチュートリアルとは、読者のパソコン上で実際に同じ手順を追えるような記述のことを指す。これによって、前章で提案した接合アプローチの詳細な手順を示すとともに、KH Coderの利用法の概略を示すことを目指す。ここでは夏目漱石の「こころ」を題材として用い、Correlationalアプローチに似通った方法で作品の上・中・下それぞれの部分に特徴的に出現する言葉から、作品の構成や特徴を探る。次にDictionary-basedアプローチによって、人の死やその原因となりうる事柄を表す言葉が、作品全体のどの部分で頻出しているかを集計し、人の死が作品中でいかに描かれているのかを探索する。これらの手順によって、「先生」という登場人物が持つ死への執着心が明らかになることを示し、「先生」の自殺が突然・不自然になされているという従来の指摘が、必ずしも妥当でないことを再確認する。

なおKH Coderは、LinuxまたはWindows上で動作するプログラムで、接合アプローチにもとづく分析のために必要な計算・検索を、マウス操作によって容易に実行することができる。KH Coderはインターネットを通じて誰でも自由に利用できる形で公開し、また、チュートリアルの手順を追うために必要な「こころ」を入力したデータも同様に公開する。KH Coderの詳細については、プルダウンメニューから呼び出すことができる38の機能すべてと、分析用データ準備などについて記述したマニュアルを資料Aとして添付する。

次に第4章では、伝統的な内容分析において頻繁に取り上げられてきた新聞記事に注目する。同様のデータを同様の目的や観点から分析する場合には、新たな方法であれ従来の方法であれ、同じ結論が出ることが望ましい。また、従来の方法で頻繁に分析対象として取り上げられてきたような種類のデータは、当然、新たな方法でも無理なく分析できなければならない。これらの点に関する確認・検討を行うために、1991年以降の『毎日新聞』から「サラリーマン」に言及した記事を取り出し、KH Coderを用いて接合アプローチにもとづく分析を行う。そして、その結果を既に手作業で行われた分析の結果と比較し、不況期における「サラリーマン」の語られ方について、伝統的な手作業による分析と同様の分析結果が得られたことを示す。さらに、手作業による分析では明らかにすることが難しかった事柄として、例えば映画やドラマに取り上げられる「サラリーマン」の特徴が、KH Coderおよび接合アプローチを用いることで新たに明らかになることを示す。

最後に第5章では、計量テキスト分析という方法を拡充するための以上4章にわたる試みを総括し、社会調査における計量テキスト分析の役割を明らかにすることを目指す。まず従来の計量テキスト分析において確立していた方法、すなわち計量的手法によるデータ要約の結果を参考にしながら主に質的な分析・記述を行う方法と、本稿で新たに提案した量的分析に重きをおく方法との関連を整理する。これら2つの方法は決して排他的なもの・分離したものではなく、共通した概念的基礎・フィロソフィーの上に成り立ち、さらに互いの手順・方法を部分的に共有していることを示す。その上で、これら2つをあわせた総体としての計量テキスト分析をいかに社会調査に応用しうるかという展望と、その際に、コンピュータがデータの文脈に敏感でないという問題をいかに扱うべきかという課題について論じる。

論文審査の結果の要旨

本論文は、質問紙調査の自由回答、小説、新聞記事などの日本語テキストデータを分析する方法の改良を提案し、その提案に従ったプログラムの作成、並びにその分析結果の検討を内容とするものである。

そこで提案される計量テキスト分析は、近年のコンピュータ技術の発達を利用する一方で、従来から研究の進んでいる内容分析の成果を検討して、一段と進んだ量的な方法と質的な方法を接合させたものである。それは、第一段階では、自己組織化マップやクラスター分析を利用し、データそれ自体の関連を重視するCorrelationalアプローチを用いた上で、第二段階でコーディングルールを用いるDictionary-basedアプローチを用いる方法である。この二つの方法を併用できる日本語のデータを分析できるプログラムは存在しないため、KH Coderと名付けたWindowsでも利用できるプログラムが作成された。このKH Coderによって、第一段階では、分析者の予断を排してデータの全体

像を理解することができ、その上で第二段階として、分析者の仮説や問題意識を生かしたコードをデータに与え、目的とする分析が可能となることが示された。

データの分析を重視する社会調査の分野では、コンピュータの操作、利用の能力は不可欠である。さらに目的に即して高度にコンピュータを使えること、すなわち他の多くの社会調査に携わる研究者は作ることもできないが、一方でそれらの人でも使うことが可能な使いやすいプログラムを作成したことも非常に高く評価できる。

以上により本論文は、博士（人間科学）の学位に十分に値すると判定した。