

平成 21 年 6 月 29 日現在

研究種目：若手研究（スタートアップ）

研究期間：2007～2008

課題番号：19830031

研究課題名（和文） 計量テキスト分析による情報行動の調査研究

研究課題名（英文） Research on Information Behavior using the Method of Quantitative Text Analysis

研究代表者

樋口耕一（HIGUCHI KOICHI）

立命館大学・産業社会学部・准教授

研究者番号：00452384

研究成果の概要： 第一に、コンピュータを用いた内容分析手法「計量テキスト分析」に関する基礎研究を行った。大量の日本語テキスト型データを対象として、事例的な解釈では発見が難しいデータの潜在的な構造を抽出するなど、量的方法の利点を活かした分析手法の提案と、その実現に必要な独自ソフトウェアの開発を行った。第二に、計量テキスト分析の応用研究として、人々が情報を獲得・解釈・発信するプロセス、すなわち情報行動についての調査・分析を行った。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2007 年度	1,320,000	0	1,320,000
2008 年度	1,350,000	405,000	1,755,000
年度			
年度			
年度			
総計	2,670,000	405,000	3,075,000

研究分野：社会学、メディア情報学・データベース、統計科学

科研費の分科・細目：社会学

キーワード：社会調査法、内容分析、情報行動、計量テキスト分析、テキストマイニング

1. 研究開始当初の背景

(1) 計量テキスト分析 コンピュータを用いた内容分析の手法

社会調査によって得られるテキスト型データには、インタビュー記録や会話記録のトランスクリプト、新聞記事、質問紙調査における自由回答項目など様々なものがある。しかし、日本語をコンピュータで扱うことが困難だったこともあって我が国ではこうしたデータをコンピュータで扱ったり、計量的に分析したりする方法についての研究がほとんど進んでいない。そのため、大量にデータ

が得られた場合でも、従来はその分析が事例的な解釈にとどまることが多かった。

それに対して英語圏では、1960年代に P. J. Stone らによって製作された「General Inquirer」以来、コンピュータ利用が始まってからでも半世紀近くにならって内容分析の方法論が研究されてきた。

この内容分析の方法論に依拠しつつ、近年の自然言語処理技術や統計解析を取り入れることで、よりすぐれた分析手法、それも日本語データを扱う手法を提案しうると思われる。

(2) 情報行動についての調査研究 計量テキスト分析の応用

計量テキスト分析の応用研究として、情報行動についての調査研究を行う。インターネット上のブログやチャット、掲示板などにおけるコミュニケーションは、その多くが記録され蓄積されていく。これらの膨大な量のコミュニケーション記録（ログ）を収集し、多数のコミュニケーションに共通する特性を数量的に把握することが、上述の計量テキスト分析によって可能になった。

従来、インターネット上で行われるコミュニケーションの研究においては事例調査・実験が多く行われてきたが、事例調査で得られた知見をインターネット全般に一般化できるのかどうか、実験で得られた知見が実験状況に固有のものでないのかどうか、といった点に検討の余地があった。

計量テキスト分析によって多数のコミュニケーション記録の調査を行うことで、こういった点の検討を行いうると考えられる。

2. 研究の目的

(1) 計量テキスト分析 コンピュータを用いた内容分析の手法

内容分析の方法論に強く依拠しつつ、自然言語処理分野の近年の研究成果を取り入れることを目指す。そして、実際の分析に必要なソフトウェアについても、欧米で質的調査に用いられているものに匹敵するものを開発・公開する。これによって、大量の日本語テキストデータを対象として、事例的な解釈では発見が難しいデータの潜在的な構造を抽出するなど、計量的方法の利点を活かした分析の手法を提案し、必要なソフトウェアを開発することを目指す。

(2) 情報行動についての調査研究 計量テキスト分析の応用

計量テキスト分析を応用することで、インターネット上の掲示板や Chat などで行われたコミュニケーションの特徴を分析したい。ただし現時点では、インターネット上でコミュニケーションに積極的に参加する人々は一部の「先進的」利用者に限られているため、日本人全般の情報行動のあり方と、その中でのインターネットの位置づけを探る必要がある。こうした点について、既に行われた質問紙調査の2次データおよび新聞記事の分析から探索することを目指す。

3. 研究の方法

(1) 計量テキスト分析 コンピュータを用いた内容分析の手法

「計量テキスト分析」の基礎研究に関しては、欧米で質的調査に用いられている手法に

学ぶだけでなく、日本語データの分析に適した分析の方法を追究する。具体的には、複合語の抽出、否定語の処理といった比較的新しい自然言語処理技術や統計手法を積極的に評価・検討し、取り入れていくことで、分析方法と分析用コンピュータ・プログラムを改良する。さらに、こうした新しい技術によって自動化すべき部分と、時間と労力を割いても人間・研究者の判断を活かすべき部分との判別を進める。そして、人間の判断を活かすべき部分については、簡便な操作でデータ中から判断に必要な部分を抽出して表示するといった形で、研究者の判断を支援するインターフェイスを準備する。

(2) 情報行動についての調査研究 計量テキスト分析の応用

計量テキスト分析を応用して行う情報行動の調査研究においては、インターネット上で行われたコミュニケーションの内容分析だけでなく、そうしたインターネット上で行われるコミュニケーションが、人々の日常生活の中でいかに位置づけられ、意味づけられているのかを探ることを目指す。質問紙調査データや新聞記事の分析を行い、ここでは重要な比較対象としてマス・コミュニケーションを取り上げる。そして、その位置づけや意味づけを分析した上で、インターネット上のコミュニケーションと比較する。

なお、コミュニケーション記録の収集・分析にあたっては(i)コミュニケーションへの参加者が匿名で、不特定多数の人に公開されている電子コミュニティを対象とし、(ii)研究発表時には会話記録の些細な引用を行わず、統計的な分析の結果を発表することで調査対象者のプライバシーに配慮する。また質問紙調査の実施にあたっては、統計的にデータを処理し、個々の対象者を特定するような分析を避けることで調査対象者のプライバシーに配慮する。

4. 研究成果

(1) 計量テキスト分析 コンピュータを用いた内容分析の手法

分析手法およびその実現のためのソフトウェアについて、以下のような改良を行った。まず自然言語処理技術に関しては、分析における有用性を検討した上で、複合語を検出する機能（図1）や否定語を取り出す機能を分析用ソフトウェア「KH Coder」に取り入れた。

次に統計解析に関しても、テキスト型データの分析に適した手法を調査・検討した上で、階層的クラスター分析（図2）、多次元尺度構成法（図3および図4）、対応分析/数量化III類（図5）、共起ネットワーク（図6）などの多変量解析手法を選択した。これらの解

析を KH Coder 上で実行できるようになったことで、データ中でどのような語と語、概念と概念の間の関連が強かったのか、また、似通った文書のグループにはどのようなものがあったのかということ、多変量解析を通じて容易に探索できるようになった。



図1 「利用」を含む複合語を検索

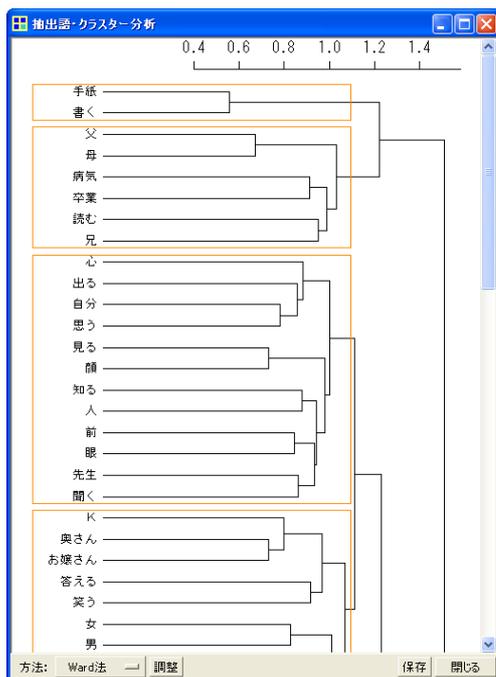


図2 階層的クラスター分析

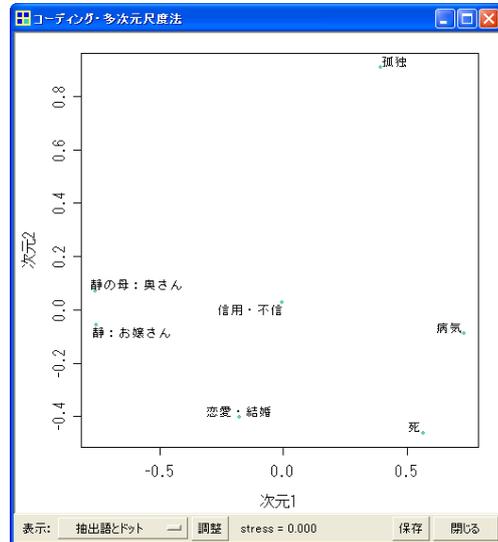


図3 多次元尺度法(2次元の解)

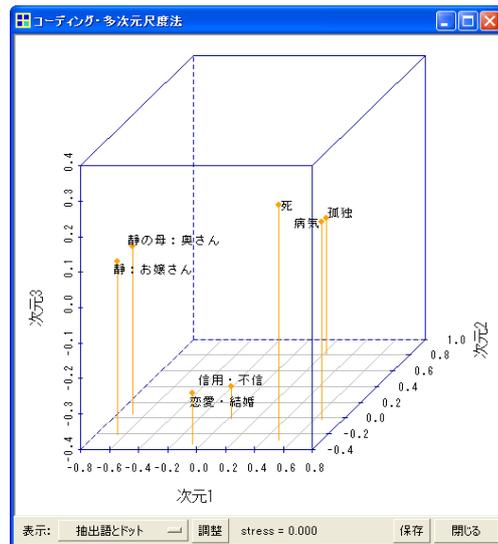


図4 多次元尺度法(3次元の解)

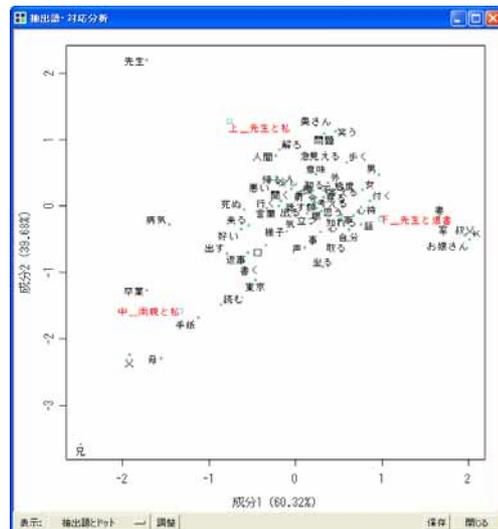


図5 対応分析(数量化III類)

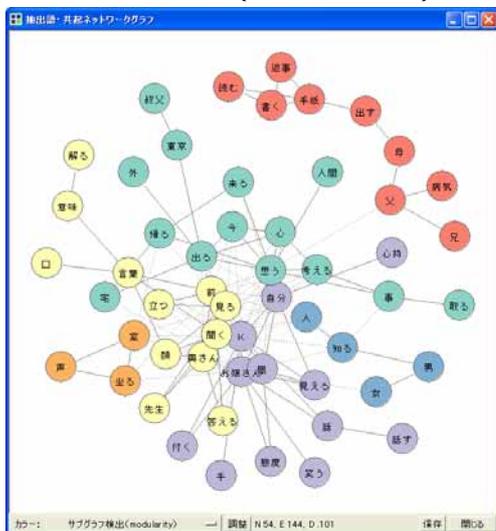


図6 共起ネットワーク

以上に示した図はすべて、本研究で製作したソフトウェア KH Coder の分析画面である。

なお、幅広い分野での試用とフィードバックを仰ぐために、KH Coder の現在の版をインターネットで公開するとともに、日本教育工学会合宿研究会「授業評価・授業改善におけるテキストマイニングの有効性を探る！」(2008年2月22日・兵庫教育大学) 日本質的心理学会研究交流委員会主催シンポジウム「テキストマイニングからみた質的研究のもう一つの未来」(2008年2月2日・立命館大学) 関西学院大学大学院社会学研究科 21世紀 COE プログラム 2007 年度連続シンポジウム「大学院における社会調査教育はどうあるべきか」第3回「社会調査と言語」(2007年10月13日・関西学院大学) 等で講演を行った。

(2) 情報行動についての調査研究 計量テキスト分析の応用

計量テキスト分析の応用研究として、人々が情報を獲得・解釈・発信するプロセス、すなわち情報行動についての調査研究を行った。ここでは、テレビやインターネットの普及が進む中、新聞が人々の現実認識にどの程度の影響を及ぼしているのかを実証的に検討することを試み、結果を第58回関西社会学会大会にて報告した。なお、この報告により、第58回関西社会学会大会優秀報告賞を受賞した。

こうしたマスメディアについての分析を行った上で、次にインターネットおよびイン

ターネット上でのコミュニケーションがいかに普及しつつあるのかを分析した。2001年及び2004年に行われた繰り返し調査「情報化社会に関する全国調査」を用いて探索的な分析を行い、結果を第47回数理社会学会大会で報告した。また、暴力に関連するコミュニケーションの探索的な調査を行い、その結果を「兵庫県こころのケアセンター報告書」に発表した。今後、以上の研究から得られた結果を活かして、インターネット上で行われたコミュニケーションの内容分析に取り組みたいと考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

酒井佐枝子・樋口耕一・稲垣由子・良原誠崇・加藤寛、「児童養護施設内暴力に関する調査研究」『兵庫県こころのケアセンター研究報告書』 査読無し、平成20年度版、2009、pp. 113-143

〔学会発表〕(計3件)

樋口耕一、「イノベーション普及過程の探索——2001年から2004にかけてのインターネットを事例に」第47回数理社会学会大会、2009年3月7日、於・京都産業大学。

樋口耕一、「計量テキスト分析の提案と実践(ワークショップ『テキストマイニングによる内容分析の現状と課題』)」第48回日本社会心理学会大会、2007年9月24日、於・早稲田大学

樋口耕一、「現代の新聞と人々の意識——全国紙の内容分析は社会調査の一環たりうるか」第58回関西社会学会大会、2007年5月26日、於・同志社大学

〔その他〕

樋口耕一、「KH Coder」(計量テキスト分析のためのソフトウェア)、2008-2009、<http://sourceforge.net/projects/khc> 研究成果の紹介・公表用ホームページ <http://khc.sourceforge.net>

6. 研究組織

(1) 研究代表者

樋口 耕一 (HIGUCHI KOICHI)
立命館大学・産業社会学部・准教授
研究者番号：00452384

(2) 研究分担者

(3) 連携研究者