台大日本語文研究第 39 期
2020 年 06 月，頁 101-130
2020-03-30 收稿，2020-06-01 通過刊登
DOI: 10.6183/NTUJP. 202006_(39).0005

文本探勘用之於日語相關人文類研究方面的質性研究應用與課題

落合由治*

摘要

目前日語相關研究上使用的計量語言學研究方法，大多偏向計量語言學領域與課題。而應用於其他日本語學、文學相關研究領域上的資料探勘（文本探勘），卻仍然非常有限。由於當今語料庫或電子資料的普及，逐漸形成將之廣泛應用於各領域研究的趨勢。

本論文乃是利用第三代 AI 技術之自然言語處理重要領域的資料探勘手法，特別是善用從龐雜的資料中擷取有意義的資訊之文本探勘技術，嘗試探討應用於日語以及日語教育等日語相關人文類研究的質語。

研究結果顯示：以文本探勘技術的量化分析為依據，反而可以將蘊藏於文本中的多元化以及多層次的意義(建立文脈組織的質化意味的集合)等質性分析結果彰顯出來。並可以確定透過導入文本探勘技術，將多元性且客觀性的成果呈現出來。

關鍵詞：日語、人文類研究、文本探勘、質性研究、應用

* 淡江大學日本語文學系教授

臺灣大學學術期刊資料庫

Qualitative Research in the Humanities on Japanese Language: Applications and challenges of text mining methods

Yuji Ochiai*

Abstract

Up to the present in research related to the Japanese language, the quantitative linguistic method is limited to quantitative linguistic disciplines and issues. In other areas of Japanese language studies, the application of data mining (text mining) techniques has been quite limited. However, with the widespread use of corpora and electronic data, there has been a movement towards research using them.

In this paper, I focus on text mining techniques to extract meaningful material by applying data mining methods to linguistic materials, which is a field of natural language processing that is an important part of the third AI technology. This study is an attempt at text mining into a qualitative study of Japanese language related humanities studies, including Japanese language studies and Japanese language education.

As a result, when performing qualitative analysis, it is possible to trace the layers of diverse and layered semantic spaces (sets of qualitative semantic relations that make up a context) contained in a text by using text mining as a cue. The application of text mining into humanities research in the Japanese language paves the way for more diverse and objective results.

Key words: Japanese, humanities research, text mining, qualitative research, application

* Professor, Department of Japanese, Tamkang University, Taiwan

臺灣大學學術期刊資料庫

日本語関係人文系研究の質的研究における テキストマイニング手法の応用と課題

落合由治*

要旨

現在までの日本語に関わる研究では、計量言語学的手法は計量言語学的分野や課題に限定され、他の日本語学研究領域では、データマイニング（テキストマイニング）技法の応用はかなり限定的なものであったが、コーパスや電子データが普及したことで、それらを利用した研究の動きが始まっている。

本論文では第三次 AI 技術での重要部分である自然言語処理の一分野である、データマイニング手法を言語資料に適用して有意義な資料を取り出すテキストマイニング技法を取り上げて、日本語学および日本語教育等の日本語関係人文系研究の質的研究への応用を試みてみた。

その結果、質的分析は、逆にテキストマイニングを使った量的分析を手掛かりにして、テクストに含まれている多様で重層的な意味空間(文脈のまとまりを作る質的な意味関係の集合)の層をたどることが可能になることがわかり、テキストマイニングの導入で、より多様性と客観性のある成果に繋げる道が開かれることが明らかになった。

キーワード：日本語、人文系研究、テキストマイニング、質的研究、応用

* 淡江大学日本語文学科教授

臺灣大學學術期刊資料庫

日本語関係人文系研究の質的研究における テキストマイニング手法の応用と課題

落合由治

一、はじめに

現在までの日本語に関わる研究では、計量言語学的手法は計量言語学的分野や課題に限定され、他の日本語学研究領域では、データマイニング（テキストマイニング）技法の応用はかなり限定的なものであったが、コーパスや電子データが普及したことで、それらを利用した研究の動きが始まっている。[1]一方、計量言語学的手法は技術的制約から語単位での統計的処理に限定され、言語事象としてそこから捉えられる現象自体に大きな制約を受けていたが、自然言語処理技術と連動した研究の動きが始まっている。[2]言語資料を元にした研究では、社会科学分野でも内容分析などによるメディア研究、

---

1 データマイニングとは、数量的データを統計的に分析して、有意味な資料を取り出す各種の手法の総称で、その中で主に言語資料を対象とした手法はテキストマイニングと呼ばれる。一例として日本語学会の近年の学界展望を見ても、コーパス、言語資料の電子データの利用は広がっているが、計量的方法の応用に相応しい語彙、文字・表記、文章・談話、社会言語・言語生活の分野でもデータマイニング、テキストマイニングを主な方法として用いている研究はまだ多くはないが、遠藤織枝,小林恵美子,佐竹久仁子,髙橋美奈子（編）(2016)『談話資料　日常生活のことば』ひつじ書房、相澤正夫,金澤裕之（編）(2016)『SP 盤演説レコードがひらく日本語研究』笠間書院、小川芳樹,長野明子,菊池朗（編）(2016)『コーパスからわかる言語変化・変異と言語理論』開拓社等の研究が見られる。日本語学会(2018)「特集 2016 年・2017 年における日本語学界の展望」『日本語の研究』14-3 参照。応用例として、博士論文での試みは増えている。清水勇吉(2015)『テキストマイニングによる日本語コーパス分析に関する研究』徳島大学博士論文、山崎誠(2017)『テキストにおける語彙的結束性の計量的研究』和泉書院等を参照。

2 同注 1 日本語学会(2018)「特集 2016 年・2017 年における日本語学界の展望」、また計量国語学会『計量国語学』http://www.math-ling.org/archives/index.html(2020 年 3 月 20 日閲覧)参照。しかし、2010 年以降、自然言語処理技術の発達によって、情報処理分野の研究課題である数理的言語モデルの検証や開発、多変量解析、クラスター分析などデータマイニング手法を応用した言語事象の規則性の解明など、情報処理分野の自然言語処理に接近した研究課題が増えてきている。松田謙次郎(2017)「変異理論における一般化線型混合モデルの導入」『計量国語学』31－6pp.402-416、岩崎拓也(2017)「読点が接続詞の直後に打たれる要因―Elastic Net を使用したモデル構築と評価」『計量国語学』31－6pp.426-442 参照。

臺灣大學日本語文學系

社会心理学的調査による社会意識研究があり、こうした分野では計量言語学と同じデータマイニング（テキストマイニング）技法を応用していたが、研究対象は言語表現をしている元の主体や集団の意識、思想などであり、言語研究の対象とは異なっているため、言語研究としてのテーマの方向性が課題と言える。[3]

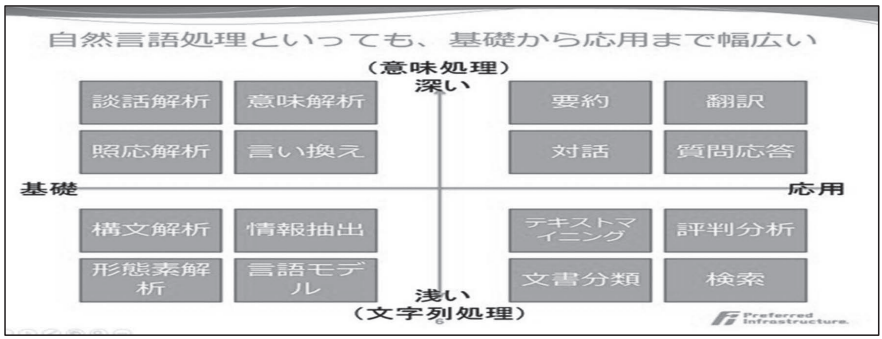


図 1　自然言語処理の分野[4]

その一方、現在、第三期 AI 技術の発展で、非常に大きな資料から自然言語処理を行う手法が開発されつつあり、さまざまな分野への応用が期待されている。また、AI 技術のために新たな視点で言語事象を捉えなおす必要も生まれている。現在、数理的方法でさまざまなレベルの言語の規則性を明らかにしようとする研究は、情報科学分野で活発な研究が進んでいる。[5]自然言語処理といっても、その基

[3] 社会科学分野でのデータマイニング、テキストマイニングの技術と応用は、樋口耕一(2014)『社会調査のための計量テキスト分析――内容分析の継承と発展を目指して』ナカニシヤ出版、今井耕介/粕谷裕子,原田勝孝,久保浩樹訳(2018)『社会科学のためのデータ分析入門上下』岩波書店等を参照。研究展望として、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』68-3pp.334-350 参照。また、テキストマイニングについては、プログラミング言語の R を用いた例は、石田基広(2017)『R によるテキストマイニング入門第 2 版』森北出版、Python を用いた例は山内長承(2017)『Python によるテキストマイニング入門』オーム社参照。

[4] 図版は、LogmiTech(2015)「スマニューにも使われる「自然言語処理」ってなに？」https://logmi.jp/tech/articles/45207(2020 年 3 月 20 日閲覧)参照。

[5] 日本語を中心とした自然言語処理については、言語処理学会『自然言語処理』https://www.anlp.jp/abst/index.html、人工知能学会『人工知能』https://www.ai-gakkai.or.jp/published_books/journals_of_jsai/、情報処理学会『情報処理』

本的な技術の分野は、応用的なものから基礎的なものまで幅広く、処理の方法も多様で、その中から人文科学的な質的方法に適した自然言語処理の技術を見出し、応用していくことで新しい可能性が生まれると言える。

そこで本論文では第三次 AI 技術での重要部分である自然言語処理の一分野である、データマイニング手法を言語資料に適用して有意義な資料を取り出すテキストマイニング技法を取り上げて、日本語学および日本語教育等の日本語関係人文系研究の質的研究への応用を試みてみたい。語レベルにとどまらず、文、文章レベルでの問題について、今までの人文科学的な質的方法と AI による自然言語処理とを組み合わせることで、新しい研究課題と分野が広がっていけば何よりである。

二、研究の対象と目的に応じた研究方法の問題

日本語学、日本語教育学など日本語関係の人文社会系研究で、よくある誤解の一つは、データマイニング、テキストマイニングなどの方法で数理的に処理できるデータに対する量的研究のほうが、研究者が対象、資料との相互作用によって規則性を見出してくる質的研究よりも客観的で、取り出された規則性もより事実に即しているという見方であろう。しかし、具体的場面での人間あるいは集団の履歴、行為、心理等を研究対象に観察、インタビュー、言語資料などによって特徴を捉えようとする教育学、心理学、社会学、経営学、保健衛生学、看護学、介護学等の分野では、すでに早い時期から量的方法の限界と新しい方法の開拓という方法的選択が大きな問題になっていた。久保田賢一(1995)(1997)(2003)は、教育学における客観主義と構成主義の基本的パラダイムの相違を捉え、扱う対象と問題、教育方法等の違いを検討し、量的研究と質的研究の方法選択の問題

https://www.ipsj.or.jp/e-library/digital_library.html 等を参照(リンクは 2020 年 3 月 20 日閲覧)。最新の研究動向の一例として、人工知能学会(2019)「特集：「深層学習による言語生成」」『人工知能学会誌』34-4 参照。

臺灣大學圖書館典藏資料庫

を論じている。[6]久保田賢一(1997)は、客観主義と構成主義の言語観の相違を以下のように論じている。[7]

> 実際上の研究において質的方法論と量的方法論の両方の手法を取り入れることがありえても、パラダイムの基盤となる「暗黙の哲学的前提」は、客観主義と構成主義パラダイムの間では対立するものであり、理論的には二つのパラダイムが融合することはあり得ない（KUBOTA1991）。この対立する前提に立つことは、研究目的が異なることであり、目的を達成する手順や手法も違ってくる（HAMMERSLEY&ATKINSON1983）。量的研究の目的は、一般法則を見つけだすことであり、その法則に従って教育現場の出来事を説明し、さまざまな場面において起こる事象を予測することである。それに対し、質的研究の目的は、ある状況において人々がどのように現実を捉え、その現実との相互作用のなかでどのように生きているか、人々の主観的な立場を尊重し、理解することである。

言語研究で見れば、言語現象の一般的法則を見つけ、その法則に従って個々の言語現象を説明し、さまざまな場面で生じている現象を記述するのが言語研究における量的研究になり、ある状況において個々の言語主体が現実を捉え、その現実の相互作用の中で産み出している言語表現を理解し、言語主体の主観的な立場を尊重して、

---

6 久保田賢一(1995)「教授・学習理論の哲学的前提 ：パラダイム論の視点から」『日本教育工学雑誌』18-3_4pp.219-231、久保田賢一(1997)「質的研究の評価基準に関する一考察 ：パラダイム論からみた研究評価の視点」『日本教育工学雑誌』21-3pp.163-173、久保田賢一(2003)「構成主義が投げかける新しい教育」『コンピュータ＆エデュケーション』15pp.12-18 参照。久保田賢一は量的研究と質的研究のパラダイムを、1980 年代の質的研究の始まりの時期の Lincoin, Y. & Guba, E. (1985) *Naturalistic Inquiry.*Sage Publications, Beverly., GUBA, E. & LINCOLN, Y. (1989) *Fourth Generation Evaluation*. Sage Publication, Newbury Park, GUBA, E. (Ed,) (1990) *The Paradigm Dialog.*Sage Publicatiens, Newbury Park.を元に分析しており、研究パラダイムの相違について早くから認識していた。

7 引用は久保田賢一(1997)「質的研究の評価基準に関する一考察 ：パラダイム論からみた研究評価の視点」『日本教育工学雑誌』21-3pp.163-173、引用文献は Hammersley, M. & Atk-inson, P. (1983) *Ethnography: Principles in Practic.* Routledge, London., Kubota, K (1991) *Developing an alternative Learning Environment: A Constructivist View.* Indiana University, Bloomington.

臺灣大學 版權所有 翻印必究

理解するのが言語研究での質的研究になる。目指すものが元々まったく異なっているため、明らかになるものもまったく意味が違い、研究方法の優劣を論じたり、客観性の程度を批評したりすることは、それぞれの研究の独自性を損なうことになる。日本語研究、日本語教育研究では、1980年代以降、欧米の質的研究が行ってきた研究パラダイムの質的相違に関する議論が基本的に欠けており、何を明らかにするかという点の根本的な違いの認識が共有されておらず、いまだに優劣を論じるような二者択一的観点が目立つと言える。しかし、両者の違いは優劣ではなく、哲学的前提の相違による目的、対象、方法の相違であり、価値的なものではなく、純粋に方法的問題である。久保田賢一(1997)は、両者の研究を以下のように比較している。[8]

表 1 量的研究と質的研究の比較

比較する点	量的研究	質的研究
焦点	量（どれくらい）、発見	質（性質、本質）、意味
哲学的前提	容観主義、論理実証主義	現象学、解釈学、象徴的相互作用
関連する用語	実験、実証、統計、客観、中立的	エスノグラフィック、自然主義、データ対話型、主観的、中立的ではあり得ない
目標	予測、制御、確証	理解、描写、協同で構築
研究デザイン	事前に決定、構造的	柔軟、変化していく、次第に明らかになっていく
状況	人工的、操作的、実験室、状況から独立している	自然的、日常的、状況に依存する
標本	大きい、無作為、代表	小さい、意識的
データ収集	テスト、アンケート、サーベイ調査など	インタビュー、参与観察、目誌など
理論	理論の検証	理論の生成
期間	（一般に）短い	長い
知見	正確、狭い、還元主義的	理解、全体的、広がり

（久保田賢一(1997)による）

[8] 同注 7p.166 参照。引用文献は、Merriam, S.B. (1988) *Case Study Research in Edcation: A Qualitative Approach*. Jossey−Bass Publishers,San Francisco. S.B.メリアム/堀薫夫、久保真人、成島美弥翻訳(2004)『質的調査法入門：教育における調査法とケース・スタディ(叢書・現代社会のフロンティア)』ミネルヴァ書房参照。

臺灣大學學術期刊

日本語学や日本語教育学が問題にしている日本語は、20世紀前半に語、文、文章という質的単位の基準が置かれたことで、20世紀後半は、その基準に従って各単位で言語現象の一般的規則性を明らかにし、また歴史的変遷を明らかにする量的研究が中心に進んだと言える。その一方で、個々の言語主体の言語使用を問題にする文章・談話研究や社会言語学的研究、また語、文が特定の場面で果たす具体的機能を明らかにする文型、用法の研究も発展してきたが、これらは基本的に質的研究の方向性を採っていると言える。言語現象の一般性対言語使用の個別性という、解明したい対象の性質の異なりで、採るべき方法はまったく違ってくる。今回、取り上げるデータマイニング手法中で言語資料を主に扱うテキストマイニング手法は、従来は主に量的研究のための方法であったが、今回は質的研究のために使用し、量的分析で質的データの何が明らかになるのかを提示してみたい。なお、各国での質的研究の情報科学技術との連携にかんする研究現状は、稿を改めて取り上げることとしたい。

三、事例研究としてのテクストの質的研究へのテキストマイニング手法の応用

従来、人文社会系研究の基本的パラダイムの問題は、日本語や日本語学習の研究に関係した分野では近年はほとんど論じられることがなかったため、言語現象の一般性を目指す研究を言語使用の個別性から見て意味がないと批判したり、逆に、言語使用の個別性を明らかにしている研究を一般性がないと非難したりする研究評価の混乱は日本語関係学界の日常的問題だと考えられる。日本語教育でも同じで、学習者を集団として捉え、一般的行動の法則性を知りたい場合と、個々の現場、個々の学習者の特徴を理解したい場合では、それぞれの研究に対して、しばしば相互に対立する評価を下す場面はよく見られる。台湾での日本語研究者、日本語教育研究者自身も自己の研究の位置づけに、苦労することが非常に多い。量的研究と質的研究のパラダイム問題は、社会系研究分野ではすでにかなり明

臺灣人學學術期刊資料庫

確に意識されてきているので、台湾の日本語関係の研究分野である、言語、文学、歴史、教育、社会文化等の人文系関係研究においてもこうした基本的パラダイムの問題を理解して研究を進めることは、今後の日本語教育現場への応用を考える場合にも極めて重大な意味を持っていると言える。[9]

ここでは質的研究の方法のひとつである事例研究を用いて、具体的文学テクストを対象に、テクスト内部の文章構成と作品中での語彙の機能にテキストマイニングを使った量的方法で接近できるかどうかを試して見たい。[10]今まで文学関係テクストは、計量言語学で作者の特定や語彙資料などの量的観点では扱われてきたが、台湾ではもっぱら文学的な内容理解、解釈を中心にした質的研究の対象で、テキストマイニングなど自然言語処理や AI のプログラムを活かした研究などは行われてこなかった。ここでは、テキストマイニングを活かした作品分析の事例として、台湾でも人気のある日本人作家・村上春樹『女のいない男たち』所収の「イエスタデイ」を取り上げる。

（一）「イエスタデイ」の文章構成とその質的分析

「イエスタデイ」の文章構成は、かなり複雑である。文章構成の分類基準は、形態的指標が明確で分析再現性がある永尾章曹(1992)の「事件の話をする文章」と「話し手の思いを述べる文章」を用いた。[11]「イエスタデイ」の文章構成を段落単位で見ていくと、以下の

---

9 教育学の分野では、大谷尚(2008)「質的研究とは何か―教育テクノロジー研究のいっそうの拡張をめざして」『教育システム情報学会誌』25-3pp.340-354、大谷尚(2017)「質的研究とは何か」『藥學雜誌』137-6pp.653-658 参照。日本語教育分野で基本的パラダイムの問題を取り上げたのは、西口光一(2013)『第二言語教育におけるバフチン的視点――第二言語教育学の基盤として』くろしお出版、西口光一(2015)『対話原理と第二言語の習得と教育――第二言語教育におけるバフチン的アプローチ』くろしお出版参照。

10 事例研究の方法については、荘島幸子(2008)「<研究論文>1 事例研究再考：個を理解するということをめぐって」『教育方法の探究』11pp17-24, 村上祐介(2011)「教育学における事例研究の方法論再考：定性的研究における比較の方法(<特集>教育学における新たな研究方法論の構築と創造)」『教育学研究』78-4pp.398-410 を参照。

11 永尾章曹(1992)「第 4 章日本語の文法について」永尾章曹編著『日本語学』和泉書院 pp.103-134 参照。

臺灣大學學術期刊資料庫

表 2 のようになる。

表 2　「イエスタデイ」の文章構成

段落	内容の概要	文章の種類＝「語り」の種類
① 1～2	僕の語る対象である木樽の紹介	▲話し手の思いを述べる文章（大学時代への懐旧）
② 3～11	早稲田大学文学部の二年生のときの僕と木樽との出会い	△話し手の思いを述べる文章（縮約された事件の話をする文章）（東京／関西への思い）
③ 12～17	関西弁を使わなくなった僕と東京育ちで関西弁を使う木樽の説明と会話	△話し手の思いを述べる文章（縮約された事件の話をする文章）（東京／関西への思い）
④ 18～20	木樽の家で風呂に入って関西弁で「イエスタデイ」を歌う木樽を知ったときの対話で出た彼女の話題	△話し手の思いを述べる文章（縮約された事件の話をする文章）（東京／関西への思い）
⑤ 21～34	木樽の語る彼女との小学校からの付き合いの中で、3 人で会いたいと勧められる	■話し手の思いを述べる文章（縮約された事件の話をする文章）（木樽の彼女への思い）
⑥ 35～46	木樽、彼女と僕の 3 人で会い、二人の付き合いについて会話し、彼女とデートの約束をする	○事件の話をする文章
⑦ 47～68	栗谷えりかと僕とのデートで、木樽との関係をえりかが説明し、夢の航海で見る月の話をする	○事件の話をする文章
⑧ 69～77	翌日、デートの様子を木樽に話し、えりかとの関係を大切にするようにアドバイスする	○事件の話をする文章
⑨ 78～81	木樽の失踪	▲話し手の思いを述べる文章（大学時代への懐旧）
⑩ 82～102	16 年後の赤坂でのホテル栗谷えりかとの再会と、木樽についての思い出を話し、えりかの夢について会話する	○事件の話をする文章
⑪ 103～108	20 歳の頃の木樽やえりかとの思い出への感慨	▲話し手の思いを述べる文章（大学時代への懐旧）

（注）説明:段落のまとまりは、作品に入っている区切りによる。会話文は直前の地の文に含めて一段落とした。
(1)「話し手の思いを述べる文章」の類別 1)▲印:僕の大学時代への思いを述べる、2)△:大学二年生の頃の木樽との思い出について述べる、3）■印:木樽のえりかへの思いを述べる。(2)事件の話をする文章:○印:大学 2 年生の時の木樽、えりか、僕のデートでの出来事。および、16 年後のえりかとの再会。

作品は 108 段落あり、11 の小節に句切られている。前文(①～⑤)と後文(⑨～⑪)に当たる内容の中に事件の話をする文章(⑥～⑧)が組み込まれる村上春樹の作品の特徴である額縁構造が見られる作品で、焦点に当たる部分が多元化されて浮かび上がる文章構成になっている。文章構成上の額縁構造とは、本題にあたる部分の前後に導

臺灣大學圖書館期刊資料庫

入部の物語を外枠とし、その内側に、短い物語を埋め込んでいく入れ子構造の物語形式である。質的分析の面から見ると、文章構成の違いによる「語り」の違いは、それぞれ別な方向性を読者に対して持つので、それぞれを別に取り扱うことで、作品の重層的な意味空間(文脈のまとまりを作る質的な意味関係の集合)をたどることができるようになる。[12]

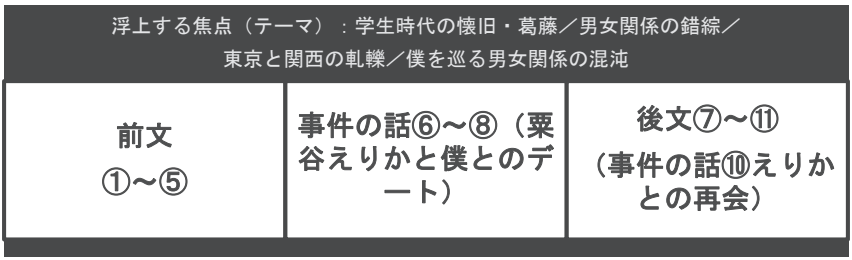


図 2　「イエスタデイ」の文章構成

内容は大きく、まず前文の第①～⑤段落で出てくる現在の僕が想起している＜大学時代(ビートルズの時代)への懐旧・葛藤＞、僕と木樽の＜東京対関西を巡る思いの軋轢＞、木樽の屈折したガールフレンドの栗谷えりかへの思いという＜男女関係の錯綜＞の 3 層の意味空間があり、その中にビートルズなど時代的あるいは社会文化記号性を持つ諸要素が織りこまれて作品の社会文化的後景が設定されている。同時に、その前文の話題の流れを受けて、「イエスタデイ」では中心になるストーリーに当たる、栗谷えりかと僕とのデートでの出来事を語る事件の話をする文章が前文に続く第⑥～⑧段落にあり、僕、木樽、えりかの三者関係を巡る事件と思い出という、微妙な男女関係の意味空間が＜僕を巡る男女関係の混沌＞として焦点化されている。作品の最後には、後文の第⑦～⑪段落に、僕と栗谷えりかとの再開を挟んで、僕、木樽、えりかの関係を巡る事件の顛末

[12] 額縁構造は、枠物語 frame story とも呼ばれ、物語論で議論されている基本的な物語構造の概念である。ジェラルド・プリンス/遠藤健一(1991)『物語論辞典』松柏社 p.69 等を参照。

臺灣大學學術期刊資料庫

とそれを懐旧する内容があり、再度、前文と呼応する意味空間が生まれている。紙数の関係で以上の概要の指摘に留めるが、質的分析を進める場合、こうした作品の言語的構成を基に相互作用的に解読を進め、作品の構造が生み出している重層的な意味空間を析出することになる。どれを主に取り出すかで読解と研究の方向はそれぞれ変わることになり、そうした多様な方向づけが可能なのが質的分析の特徴であり、同時に人文系研究の意義とも言える。

また、事例にした村上春樹の作品では、一つの作品の中に複数のこうした意味空間が形成される文章構成が作られているため、ただ最初から機械的に読んでいけば読解できるようにはなっておらず、また中心になるストーリーが最初から最後まで続いているわけではない文章構造を採っていることで、何を作品から読み出すかは、質的分析の具体的手続きなしには答えを得がたいと言える。実は、こうした文章構成による輻輳した意味空間の形成は、文学、思想、歴史、社会文化などの人文社会系テクストではそれぞれのジャンルの様式でそれぞれの言語非言語表現の固有性記号性を持って作られており、各ジャンルでの質的分析の課題になっている。学習者のインタビューやライフヒストリーなど教育分野での調査資料も、実は同じ多様な意味空間を内部に秘めている。質的分析の基本にあるのは、こうした文章構成の理解に基づいた意味空間の構造的理解なのである。これは、いわゆる一般化された言語を量的に扱う場合の単語の頻度や文型の類別などの上位に形成される言語の意味的カテゴリーであり、語、文、文章という質的統一体が言語研究で区別されている基礎にある問題である。[13]

（二）テキストマイニングによる「イエスタデイ」の量的分析

次に、よく使われるテキストマイニング手法を用いて、「イエスタデイ」に量的分析を行ってみる。使用した手法は、最も基本的な「出

[13] 質的統一体としての言語と言語の基本的な単位については、時枝誠記(1971/1941)『国語学原論―言語過程説の成立とその展開』岩波書店、時枝誠記(1988/1950)『日本文法口語篇』岩波書店を参照。

臺灣大學圖書館資料庫

現頻度順リスト」と「共起ネットワーク」、さらに、第 3 次 AI ブームで開発されたニューラルネットワークを用いる「自己組織化マップ(SOM)」と、最近、よく使われる語彙の重要度を判定する TF-IDF 法による「ワードクラウド」、トピックモデルによる統計的潜在意味解析手法の LDA および Word2vec である。

1. 語彙の「出現頻度順リスト」と「共起ネットワーク」での特徴

		抽出語	品詞細分類1	頻度
	1	僕	タグ	220
⊞	2	言う	動詞	184
	3	彼	タグ	118
	4	木樽	タグ	91
	5	女	名詞C	78
	6	えりか	人名	51
	7	栗谷	人名	46
⊞	8	思う	動詞	44
⊞	9	見る	動詞	34
	10	関西	地名	33
	11	自分	名詞	29
⊞	12	知る	動詞	25
	13	話	サ変名詞	24
	14	人	名詞C	23
⊞	15	考える	動詞	21
	16	アキ	人名	19
⊞	17	出る	動詞	19
	18	東京	地名	19
	19	勉強	サ変名詞	19
	20	ひとつ	副詞可能	17

図 3　自立語の頻出度順リスト

まず、一般的なテキストマイニングの手法である「自立語の頻度順リスト」(図 3)を示し、次に出現した語のうちで「出現パターンの似通った語、すなわち共起の程度が強い語を線で結んだネットワーク」を描く「共起ネットワーク」(図 4)の結果を提示した。[14]

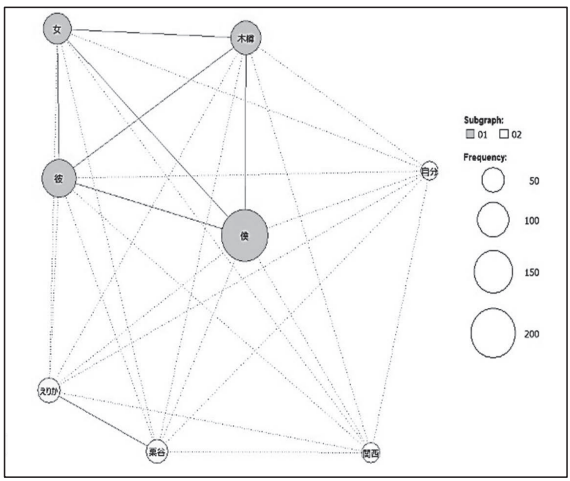


図 4　共起関係ネットワーク

この二つの分析から、この作品の言語的特徴として、図 3 からは人称代名詞「僕」「彼」「女」「自分」と人物名「木樽」「栗谷」「えりか」および、地域名「関西」「東京」がキーワードで、大きな意味を持つことが分かる。質的分析と対応させ

[14] テキストマイニングには「KHCoder」の頻度順リスト、共起ネットワーク、自己組織化マップを使用した。同注 3 樋口耕一(2014)参照。

臺灣大學學術期刊資料庫

ると、「関西」は前文の第②③④段落に対応する「木樽」が東京生まれなのに関西弁を使い、「僕」が関西生まれなのに東京の言葉を使っている前文のひとつのテーマになっている意味空間の内容で、前文では「東京」対「関西」という日本の歴史的文化的二項対立が大きなテーマの一つであることが分かる。また、図 4 の「共起関係ネットワーク」からは「僕」「木樽」対「自分」という対立があることが分かり、図 4 の右側に白く分類された「自分」という人称代名詞が同じグループに分けられた「関西」と関係して、「関西」を強く意識しており、また「僕」「木樽」という左上の青のグループの人名とは異なる機能を文章中ではたしていることが分かる。さらに、図の左上の青のグループの人名の関係を見ると「栗谷えりか」との関係で「僕」と「木樽」が「僕」対「木樽」「彼」、「僕」「木樽」対「女」という奇妙な二重関係、或は四角関係にあることも浮かんでくる。また、図の左下イエローで示される「栗谷えりか」はボーイフレンドのはずの「木樽」よりも中央の「僕」とより近い関係にあることも読み取れるであろう。

先に見た質的分析では、文章構成に基づいて大きくテーマのまとまりを形成する意味空間を取り出すことができたが、具体的にそれを示す言語的な手掛かりを決めるには、用例を抜き出すなどの別な作業が必要であった。しかし、テキストマイニングを用いると、前節で見た、＜学生時代の葛藤／男女関係の錯綜／東京と関西の軋轢／僕を巡る男女関係の混沌＞というような、質的分析で得られた作品の意味の焦点となる意味空間に対して、具体的な手掛かりとなる言語的根拠の一端が示されると言える。後は、たとえば、「関西」と「東京」の関係について、該当する部分の用例を集め、それぞれを比較対照することで、＜東京と関西の軋轢＞は具体的にどのような問題であるのか整理することで、作品に潜在している意味空間を相互作用的に顕在化させることができる。今まで、人文系のテクストの質的分析で、テキストマイニングが応用されることは少なかったが、現在の第三次 AI ブームによる自然言語処理の発達で、従来の

臺灣大學學術期刊資料庫

語、文を対象にした計量的研究からさらに踏み込んで、テクストの意味的領域に踏み込む手掛かりが与えられるようになっていると言える。

2.「自己組織化マップ(SOM)」からわかる特徴

さらに、その他のテキストマイニング手法を見ていくと、第三次AIブームによる機械学習、深層学習の発達で、2000年代では扱いにくかった表示法や分析法が可能になっている。その一つは、「自己組織化マップ(SOM)」で、データの分類をAIが学習する方法の一つであり、類似するデータを自動的に分類できるニューラルネットワークである。[15]

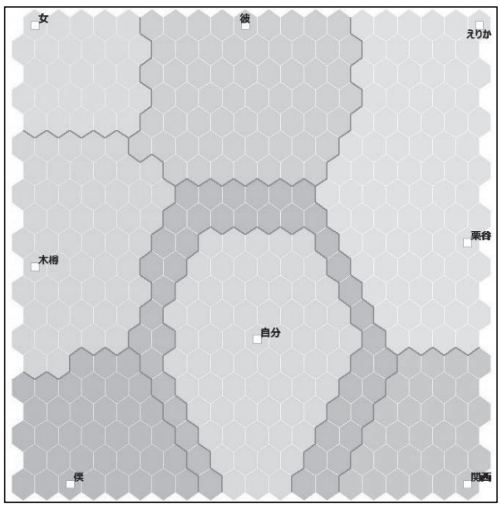


図5　頻出語自己組織化マップ

図5の自己組織化マップを見ると、「自分」を中心に「木樽」「女」「彼」「栗谷」「えりか」が上半分に、下半分に「僕」「関西」が並んでおり、「自分」が大きく二つの関係に分かれていることが分かる。また、「僕」は「木樽」と関係し、「彼」「女」という「栗谷」「えりか」の領域と繋がっている。同時に、「自分」は、「栗谷」「えりか」と「関西」との関係で繋がっており、それぞれ具体的にどのような語彙が各グループに出てくるかを見ることで、作品内の意味空間とその相互関係を取り出していくことができる。ただ、作品を読んだときには、こうした作品内の意味空間を示すキーワー

[15] 「KHCoder」の自己組織化マップを使用した。SOMは様々なデータの分類に使われているが、言語表現に関しても内容の類別に使われている。一例として、東宏一(2011)「Web上の政治的発言の分類とその応用」『人工知能学会全国大会論文集』 JSAI2011 1F4-3、黒澤義明、竹澤寿幸(2011)「自己組織化マップ SOMを用いた擬情語の分類比較」『人工知能学会全国大会論文集』JSAI2011　1C2-OS4b-7参照。

臺灣大學學術期刊論文版權

ドの相互関係を明確に取り出すことは簡単ではないが、この AI による語の共起関係の分類結果を使えば、「自分」という語を中心にして、それぞれの登場者の関係を整理し、作品の社会文化的記号になっている「東京」対「関西」の関係を、テーマ化することが可能である。この作品で言えば、「自分」として作品中に描かれている存在は場合、その意識は、上下左右の関係で分裂しており、多項対立的関係・葛藤が存在していることが浮かび上がる。

3. TF-IDF 法による「ワードクラウド」からわかる特徴

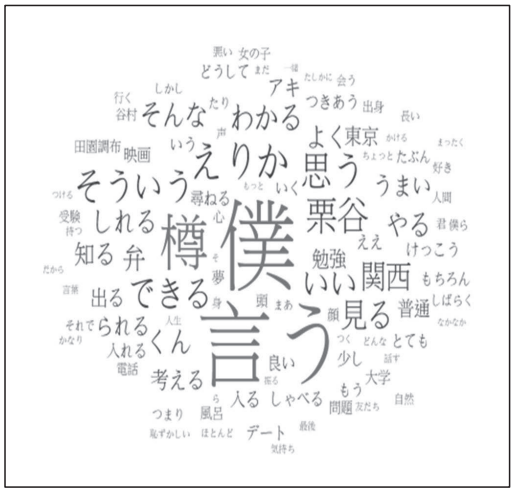


図 6　語彙の重要度によるワードクラウド

一方、図 6 は、現在インターネットでキーワードを示すために使われているデータ中での語彙の出現頻度を語彙の文書での特殊性で重み付けした TF-IDF 法によるワードクラウドである。[16]このような単語の出現頻度の重み付けを基にした場合、「イエスタデイ」のような小説の場合、出現頻度が多い「僕」が中心に出て、「樽（木樽）」「栗谷」「エリカ」「関西」など、他の分析方法でも出てきた語は出ているが、「自分」は出ておらず、共起ネットワークと自己組織化マップのほうが、より作品の意味空間の関係の深い点を取り出していると考えられる。しかし、ワードクラウドで見ると、「東京」／「栗谷えりか」、「田園調布」／「木樽」、「木樽」／「僕」／「栗谷」／「関西(弁)」という別な二項対立、三

[16] TF-IDF 法はインターネットサイトの検索でキーワードのヒットを決める際に用いられたりする内容判定のひとつの方法である。詳細については、奥村学(2010)『自然言語処理の基礎』オーム社 pp.116-119 参照。分析でのワードクラウドは MathLab「Text Analitics Toolbox」https://jp.mathworks.com/help/textanalytics/ug/analyze-japanese-text.html を使用した。

臺灣大學學術期刊資訊

項対立が捉えられている。また、同時に勉強、映画、電話など繰り返し出現した場面に関わる名詞も出ているため、意味空間の中でサブテーマになるような記号的キーワードが取り出されていると言える。動詞では、「言う」「思う」「見る」「できる」「わかる」などがよく出ており、多くが会話文で成り立っていることと、それらについて何かを思ったり、見たりして、何かができたり、わかったりしたストーリー展開があることが浮かび上がる。今まで見てきたテキストマイニングの方法と比べて、ワードクラウドは、特に名詞、代名詞に注目すると、作品内の主要語を取り出して二項対立的な関係や、付随する記号的関係を考察したりする手掛かりになると言える。また、動詞に注目することで作品の主な場面のストーリー展開を理解したりする手掛かりも得られると言えよう。

4. トピックモデルによる統計的潜在意味解析 LDA からわかる特徴

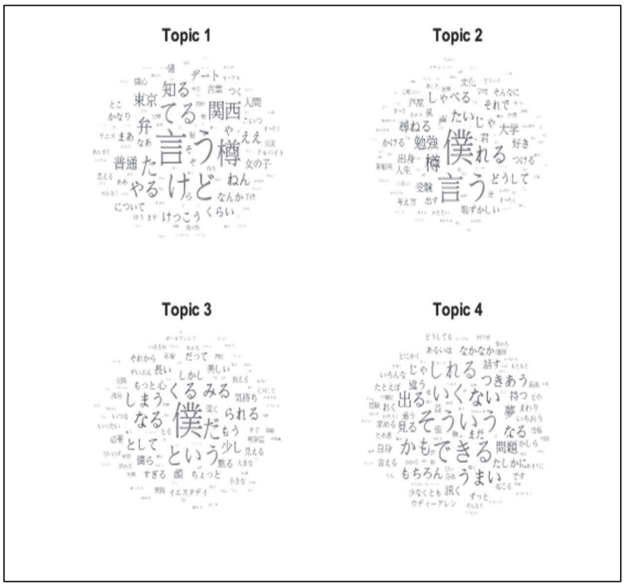


図 7　LDA によるトピック抽出

さらに、第 3 次 AI ブームの中で、テキストマイニングの方法として注目されているトピックモデルによる統計的潜在意味解析として LDA を実施してみる。LDA は、文章を語彙に分け、語彙の共起関係を出現確率として学習して、トピックの分布として文章を捉えることで、文章の持つ意味を統計的に推測する方法である。ここでのトピックとは単語の共起

臺灣大學學術期刊資料庫

関係の分布で表現される。[17]「イエスタデイ」の LDA 分析では、図 7 のように T-OPIC1 と TOPIC2 に木樽と僕との会話と僕の個人的語りが分類され、TOPIC3 には僕の「語り」に見られる動詞、形容詞などが集まっており、TOP-IC4 は「栗谷えりか」との会話での内容が取り上げられている。先にあげた、共起ネットワークや自己組織化マップで取り出された内容とはかなり差があり、作品の別の面からの意味空間の構造を取り出していると言える。ワードクラウドでの単語の比重と較べて見ると、「言う」という動詞の重要度が非常に重くなっており、会話の多いこの作品の中で、前文の段落①～⑤の主要な部分を占めている「僕」と「木樽」の会話の内容が主なテーマとして取り出されたと考えられる。語りの内容は、TOPIC1 では、「関西」「東京」「知る」「デート」「女の子」など、主に「木樽」と「僕」の故郷と言葉の関係、またそれに付随している「栗谷えりか」など「女の子」との関係が出ており、「イエスタデイ」では、「関西」「東京」という都市文化記号が女性との関係に深く結びついていることがわかる。一方、TOPIC2 には、「受験」「勉強」「大学」「文化」など受験体制に関わる語彙が出ており、この作品での意味空間に日本の大学受験体制のもたらす影響というテーマが潜在していることがわかる。TOPIC3 には、「顔」「気持ち」「美しい」など「僕」の「栗谷えりか」への見方が伺える内容が出ており、「僕」の「栗谷えりか」への隠蔽された好意が伺える。TOPIC4 は、「夢」「問題」「出る」「つきあう」など「僕」が「栗谷えりか」とデートしたときに話した内容が出ており、二人の会話に「木樽」と「僕」との関係とは別の、この作品での意味空間の大きな潜在的内容があることがわかる。4 つの TOPIC を比べて見ると、作品の内容は「僕」と「木樽」との語

[17] LDA などのトピックモデルによる統計的潜在意味解析は、単語の共起性に通目して生まれる資料の情報を潜在的「意味」として捉える方法で、現在、言語情報の示す意味を解析する手法として様々な方法が開発されている。LDA は階層的ベイズモデルによる方法で、現在、広く使用されている。詳細は、奥村学監修、佐藤誠一(2015)『トピックモデルによる統計的潜在意味解析』コロナ社参照。同注 15MathLab「Text Analitics Toolbox」を使用した。word2vec は python3 の Gensim を使用した。

臺灣人文與藝術期刊資料庫

りの関係と、「僕」と「栗谷えりか」との語りの関係で、大きく二項対立をなしていることもわかる。その対立をもたらしているのは、「東京」対「関西」のような地域と文化に関わる背景と、「大学」「受験」などに関わる現代社会の構造に関わる背景という二種類の社会文化記号であることも浮かんでくる。

このように LDA を用いると、先に見てきたテキストマイニングとは、かなり異なる記号的語彙や登場者の関係の意味空間に注目することができる。また、TOPIC として取り出される潜在的意味は、集められた語彙を元の作品で確認していくことで。読解の大きな手掛かりになる。

5. ニューラルネットワークによる単語の分散表現 word2vec による特徴

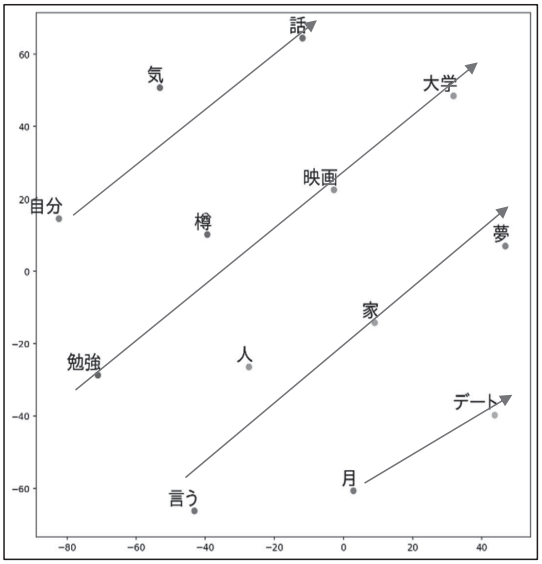


図 8　word2vec による単語ベクトル 1

さらに、word2vec による意味解析を行ってみる。word2vec は、Google によって開発された第三次 AI ブームの深層学習による典型的な自然言語処理プログラムで、ニューラルネットワークによって文章中の単語をその分布に基づいてベクトルで表現しながら学習し、前後の文脈から、ある単語と一緒に使用されやすい単語の確率を推定することで、資料中での語彙の意味関係を表示できる。単語の埋め込み（語彙からの単語またはフレーズが実数のベクトルにマッピングされる）を生成す

臺灣大學學術期刊資料庫

るために使用される[18]上位50語の中で人称代名詞や人名などを除いた名詞について、word2vec による単語ベクトル分析を行った結果、図8のように「自分、気、話」／「勉強、樽（木樽）、映画、大学」／「言う、人、家、夢」／「月、デート」という相関する言語ベクトルのグループにそれぞれ分かれた。「自分、気、話」は、「僕」に関わる語り、「勉強、樽（木樽）、映画、大学」は、大学時代の木樽との思い出、「言う、人、家、夢」は、僕の回想に関わる語り、「月、デート」は栗谷えりかとのデートに関する内容で、LDA の結果と似た、それぞれ「イエスタデイ」の文章構成上の意味空間のまとまりに応じたキーワードが取り出されていると考えられる。「自分、気、話」は、前文①～⑤、後文⑦～⑪に関する内容と考えられ、「自分」というキーワードが作品中での語り手の「気」分、「気」持ち、会「話」としての語りに深く関係していることがわかる。「イエスタデイ」での「自分」という語の特別な用法に注目する必要があると言える。「イエスタデイ」は、「僕」と「木樽」との関係を語っている小説のように見えるが、そこに深く関係しているのは「勉強、樽（木樽）、映画、大学」という「僕」の「東京」での大学生活で、作品の大きなテーマが1970年代の東京と大学生活にあり、「木樽」はその文脈で「僕」に大きな意味があると考えられる。一方、「言う、人、家、夢」のベクトルは、「僕」や「木樽」と「家」族との関係、「月、デート」は「僕」と「栗谷えりか」との関係をさしていると考えられる。「言う、人、家、夢」は「木樽」や「僕」が大学進学時にあった「家」族との軋轢あるいは故郷との関係を示していると見られるが、「栗谷えりか」と関わる「月、デート」は、それとは異なる独立し

[18] Word2vec は、全部のデータをベクトル化するのではなく語彙の共起性を確率的に推論することで語彙の分布を分散表現としてニューラルネットワークで学ぶ方法の最初のモデルで、これに続く分散表現学習（単語や文の埋め込み）の様々な方法が現在も開発されている。注目された成果は、Mikolov, Tomas; et al. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv:1301.3781 [cs.CL] https://arxiv.org/pdf/1301.3781.pdf%C3%AC%E2%80%94%20%C3%AC%E2%80%9E%C5%93(2020年3月20日閲覧)参照。詳細は、斎藤康毅(2018)『ゼロから作る Deep Learning②自然言語処理編』オライリー・ジャパン参照。

臺灣人日語學習與教育資料庫

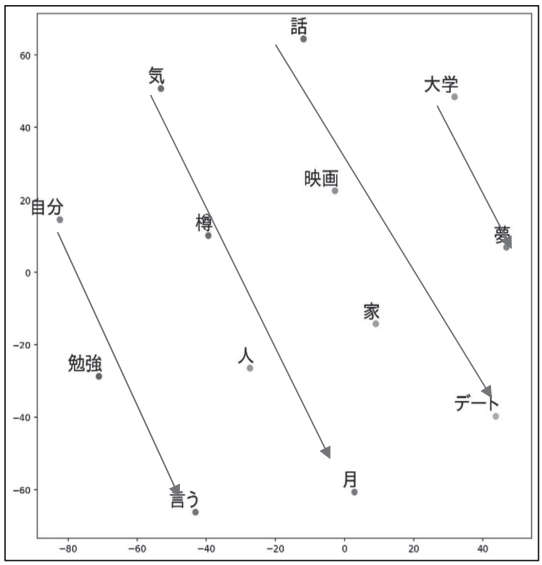


図 9　word2vec による単語ベクトル 2

た意味を「僕」あるいは「自分」に対して持っていると言える。また、ベクトルの方向を変えてみると、図 9 のように読むこともできる。ここでは、「自分、勉強、言う」／「気、木樽、人、月」／「話、映画、家、デート」／「大学、夢」という別な意味空間のまとまりを、読み取ることができる。「自分、勉強、言う」は、「自分」の語りが「勉強」が示す大学受験と深く関わっていることを伺わせ、作品の潜在的なテーマが大学受験に象徴される日本の現在の社会経済制度にあることがわかる。「気、木樽、人、月」は、作品の「気」分、気持ちに関わる部分で、「木樽」、「人」的関係、「月」の夢を話す「栗田えりか」が中心的な役割を果たしていること、また、「話、映画、家、デート」／「大学、夢」からは、こうした作品での登場者の関係が、「木樽」とその家族の「栗谷えりか」との関係、また、「僕」と「栗谷えりか」とのデートそして、「僕」の「大学」での「夢」と関わっていることが浮かんでくる。

このように、word2vec を用いると、作品内に形成されている重層的な意味空間をキーワードを通じて推測でき、読解の方向性を決める参照点にすることができる。また、意味空間の方向性はベクトルの方向に応じて変わると考えられ、作品中で一つの語が持っている重層的な意味関係を浮かび上がらせることができる。取り上げる語数を増やしていけば、さらにその重層性は明確になると言えよう。

臺灣大學學術期刊資料庫

6. 量的分析としてのテキストマイニングからの質的分析への接続

以上、事例の分析を通じて見てきた量的分析としてのテキストマイニングの結果から考えてみると、テキストマイニングで何が質的分析のヒントになるかは、一義的には決定できない。それぞれの方法に応じて、作品中から取り出される言語的な意味空間の構造は異なってくる。従来のようにデータマイニング的数量化による言語分析の一般化を目指す立場から見ると、テキストマイニングによって出てきた結果の意味するものは一義的結論でなければ意味がないと考えられてきた。テキストマイニングの結果が答えであって、研究者の側はその出力に従う必要があった。量的研究での自然言語処理は従来と同じく、プログラムの結果＝研究結果で、その結果に従って資料の意味を決める必要があった。しかし、今回、質的分析に応用するために行った作品の事例分析を通じてわかったことは、テキストマイニング手法によって取り出される出力は、質的分析の手掛かりになる出発点にあたるもので、テキストマイニングの分析結果は結論ではないという点である。今回、事例研究によって示したように、プログラムの処理の方法によって、同じ言語データでもその出力は、それぞれ異なってくる。逆にそのことが、元の資料の持つ多様で重層的な意味空間の存在を浮かび上がらせるのである。一義的結果を求める量的研究の立場から見ると、分析結果がそれぞれ異なることは有意義な研究成果とは見なしがたいことになるが、質的分析の立場から見れば、多様に見えるテキストマイニングの結果は、まるで対話者、アドバイザーがテクストの読みについて自分の意見を述べてくれたようなもので、以降は、それを入り口にして必要な用例を集めて分析し、テクストの中に構成されている意味空間の一端を取り出せれば、研究成果になる。研究のパラダイムが異なれば、出てきた結果の意味はまったく役割が違うのである。

臺灣大學學術期刊資料庫

表 3　テキストマイニングの結果に対するパラダイムの相違

項目	量的研究	質的研究
テキストマイニングの結果に対するスタンス	資料に対する分析結果は、そのまま研究の結論になるか、仮説検証の証明になる。	方法によって、結果が異なる分析結果は、それぞれが資料読解の手掛かりになり、研究の出発点になる。
テキストマイニングの役割	資料に対する分析結果が量的分析として正しい証明になる。	資料に対する分析結果は、元のテキストでの質的分析と対照させて、初めて意味がわかる。
テキストマイニングの使用方法	基本的には資料を量的データとして扱って、統計的検定を実施した結果が、目的に応じた研究の解答になる。	量的データの統計的検定だけでは意味がなく、その結果が、元のテクストで示している意味空間を検証し、質的分析で改めて確定する。
テキストマイニングを使用する価値	統計的検定による数量データの示す関数的客観性を確保する決定者、判断者。	質的分析に対する、テキスト読解のキーワードを提示する助言者、アドバイザー、対話者。

質的分析で応用する場合、テキストマイニングは、それぞれ分析方法によって、取り出されてくる作品中での重層的な意味空間(文脈)の層は異なっており、いずれもが一つの読み方として可能なものである。一義的ではない多様な答えが得られることが逆に、作品を分析、読解する場合の、多元的な視点を与え、読みの多様性をもたらすことになる。テキストマイニングの方法による結果の相違は、テクストの意味空間のいずれかの手掛かりが捉えられていると見ることで、量的研究では曖昧さ、非客観性と評価される点が、質的研究では逆に作品を精読する言語表現上の手掛かり、入り口として有効に機能する可能性が考えられる。言ってみれば、テキストマイニングの結果は質的分析にとって、一部しか本当の姿が見えていない原石に当たるものであり、そこから原石が秘めている多様な意味空間を掘り出していく、浮かび上がらせていく徴表なのである。

テキストマイニングを、作品の個性的特徴を理解する質的分析で応用する場合、その手順は量的研究の場合とは大きく異なり、以下の図 10 のように設定する必要があるであろう。

臺灣大學學術期刊資料庫

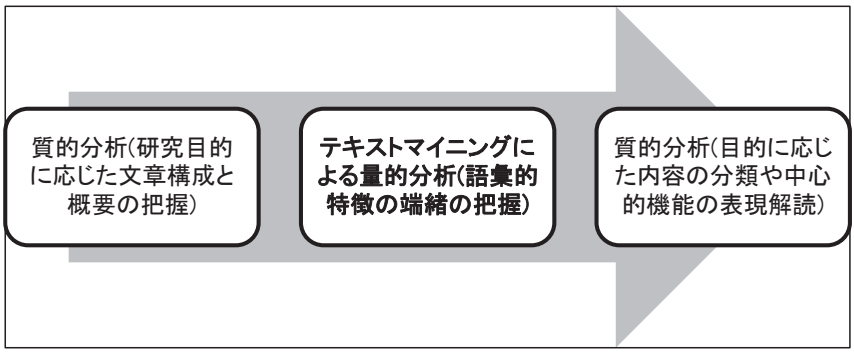


図 10　質的分析でのテキストマイニングの応用

一方、資料について特徴の一般化を目指す量的分析の場合、テキストマイニングの使用は以下の図 11 の手順になると言えよう。

量的分析と質的分析では、研究や読解行為におけるテキストマイニングの位置付けがまったく異なる点を、今後の自然言語処理の応用と発展では認識する必要がある。これは繰り返しになるが、量的分析と質的分析という研究方法の価値に関する問題は、今まで繰り返されてきたような不毛な研究価値の優劣の争いではなく、それぞれ基本的な研究対象やパラダイムがまったく異なっているという純粋に研究遂行の方法選択の問題に過ぎないことを改めて認識することで、テキストマイニングは、今まで応用が難しかった質的分析の分野でも有効な研究サポートを提供するものになる。

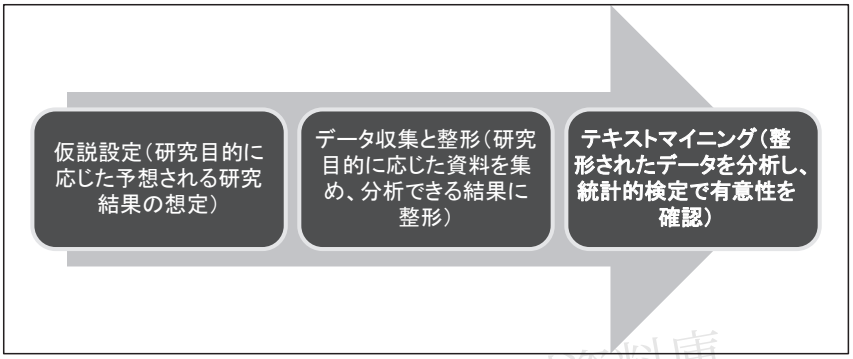


図 11　量的分析でのテキストマイニングの応用

臺灣大學學術期刊資料庫

同時に、今まで人文系の領域で研究者が自分の主観で行っていた資料の質的分析や手作業での分類では、多くの要素、意味空間が同時に見えてしまうため、どれを中心にするか軽重を付けることは容易ではなく、それが研究の困難を生み出し、教育の現場では読解の難しさに繋がっていた。それによって、また研究の客観性が問われ、人文系の研究価値の問題にもなっていた。しかし、量的研究手法であるテキストマイニングを援用する今回の方法を用いると、対象とするテクストの言語的表現な意味空間の軽重の付け方に何通りかの可能性があることが客観的に分かり、それを用例の分析を通じて提示することで複数の読解可能性があることを説得力を保って証明できる。その点で、テキストマイニングは人文系研究にとって研究の客観性を担保する非常に有意義な研究方法を提示している。

四、おわりに

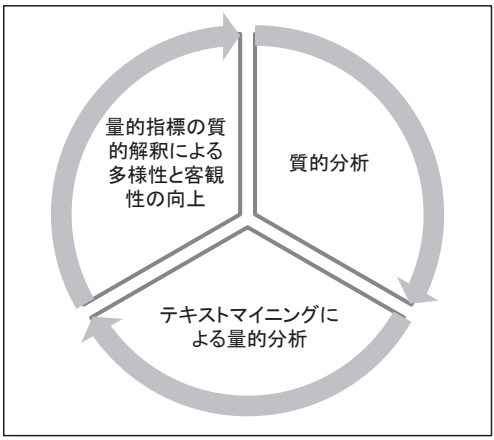


図 12　質的研究における
テキストマイニングの位置

以上、本論文で指摘してきたように、今まで対立的であった量的分析と質的分析は、事象に対する対立的研究パラダイムであり、目指している対象、目的が世界観として元々異なっているだけで、それは優劣や価値の問題ではない。今回は、質的方法で言語作品を研究する事例研究として、テキストマイニングを使った量的分析の結果の質的分析に基づく解釈、読解を取り上げた。量的分析の結果解釈だけでは、データからその出力の意味を導き出すことはできず、元のテクストとの関係で意味的に結果を理解する主観的読解過程と密接不可分な関係がある。テキストマイニングを使った量的研究の結果自体は、量的研究のテ

臺灣大學學術期刊資料庫

キストマイニングの結果と異なり、質的研究の全過程の中では図 12 のように研究対象との関係で中立的中間的な過程である。質的分析は、逆にテキストマイニングを使った量的分析を手掛かりにして、テクストに含まれている多様で重層的な意味空間(文脈のまとまりを作る質的な意味関係の集合)の層をたどることが可能になる。指向する研究対象と目的、課題に応じて、二つの方法を選択し、また、組み合わせることで現代の情報処理技術と人文社会研究の接続の道を具体化し、拡大していくことが可能になると言えよう。

【付記】本論文は、2019 年 12 月の 2019 年台湾日本語文学会年度大会での研究発表等を元に大幅に編集、加筆、訂正をおこなったものである。また、科技部研究案 MOST 107-2410-H-032 -030 -MY2 の研究成果の一部である。論文審査での査読のご意見、研究へのご支援に心からの感謝の意を表するものである。

【参考文献】

相澤正夫,金澤裕之（編）（2016）『SP 盤演説レコードがひらく日本語研究』笠間書院

東宏一(2011)「Web 上の政治的発言の分類とその応用」『人工知能学会全国大会論文集』 JSAI2011 1F4-3

石田基広(2017)『R によるテキストマイニング入門第 2 版』森北出版

今井耕介／粕谷裕子,原田勝孝,久保浩樹訳(2018)『社会科学のためのデータ分析入門上下』岩波書店

岩崎拓也(2017)「読点が接続詞の直後に打たれる要因ーElastic Net を使用したモデル構築と評価」『計量国語学』31－6pp.426-442

遠藤織枝,小林恵美子,佐竹久仁子,髙橋美奈子（編）（2016）『談話資料　日常生活のことば』ひつじ書房

臺灣大學學術期刊資料庫

大谷尚(2008)「質的研究とは何か―教育テクノロジー研究のいっそうの拡張をめざして」『教育システム情報学会誌』25-3pp.340-354
大谷尚(2017)「質的研究とは何か」『薬學雑誌』137-6pp.653-658
小川芳樹,長野明子,菊池朗（編）(2016)『コーパスからわかる言語変化・変異と言語理論』開拓社
奥村学(2010)『自然言語処理の基礎』オーム社 pp.116-119
奥村学監修、佐藤誠一(2015)『トピックモデルによる統計的潜在意味解析』コロナ社
久保田賢一(1995)「教授・学習理論の哲学的前提 : パラダイム論の視点から」『日本教育工学雑誌』18-3_4pp.219-231
久保田賢一(1997)「質的研究の評価基準に関する一考察 : パラダイム論からみた研究評価の視点」『日本教育工学雑誌』21-3pp.163-173
久保田賢一(2003)「構成主義が投げかける新しい教育」『コンピュータ＆エデュケーション』15pp.12-18
黒澤義明、竹澤寿幸(2011)「自己組織化マップ SOM を用いた擬情語の分類比較」『人工知能学会全国大会論文集』JSAI2011　1C2-OS4b-7
計量国語学会『計量国語学』http://www.math-ling.org/archives/index.html (2020 年 3 月 20 日閲覧)
言語処理学会『自然言語処理』https://www.anlp.jp/abst/index.html
斎藤康毅(2018)『ゼロから作る Deep Learning②自然言語処理編』オライリー・ジャパン
清水勇吉(2015)『テキストマイニングによる日本語コーパス分析に関する研究』徳島大学博士論文
ジェラルド・プリンス/遠藤健一訳(1991)『物語論辞典』松柏社
情報処理学会『情報処理』https://www.ipsj.or.jp/e-library/digital_library.html
人工知能学会(2019)「特集 :「深層学習による言語生成」」『人工知能学会誌』34-4

臺灣大學學術期刊資料庫

人工知能学会『人工知能』https://www.ai-gakkai.or.jp/published_books/journals_of_jsai/

荘島幸子(2008)「<研究論文>1 事例研究再考 : 個を理解するということをめぐって」『教育方法の探究』11pp17-24

時枝誠記(1971/1941)『国語学原論ー言語過程説の成立とその展開』岩波書店

時枝誠記(1988/1950)『日本文法口語篇』岩波書店

永尾章曹(1992)「第 4 章日本語の文法について」永尾章曹編著『日本語学』和泉書院 pp.103-134

西口光一(2013)『第二言語教育におけるバフチン的視点――第二言語教育学の基盤として』くろしお出版

西口光一(2015)『対話原理と第二言語の習得と教育――第二言語教育におけるバフチン的アプローチ』くろしお出版

日本語学会(2018)「特集 2016 年・2017 年における日本語学界の展望」『日本語の研究』14-3

樋口耕一(2014)『社会調査のための計量テキスト分析――内容分析の継承と発展を目指して』ナカニシヤ出版

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』68-3pp.334-350

MathLab 「 Text Analitics Toolbox 」 https://jp.mathworks.com/help/textanalytics/ug/analyze-japanese-text.html

松田謙次郎(2017)「変異理論における一般化線型混合モデルの導入」『計量国語学』31－6pp.402-416

村上祐介(2011)「教育学における事例研究の方法論再考 : 定性的研究における比較の方法(<特集>教育学における新たな研究方法論の構築と創造)」『教育学研究』78-4pp.398-410

S.B.メリアム/堀薫夫、久保真人、成島美弥翻訳(2004)『質的調査法入門 : 教育における調査法とケース・スタディ (叢書・現代社会のフロンティア)』ミネルヴァ書房

山内長承(2017)『Python によるテキストマイニング入門』オーム社

臺灣人學學術期刊資料庫

山崎誠(2017)『テキストにおける語彙的結束性の計量的研究』和泉書院

LogmiTech(2015)「スマニューにも使われる「自然言語処理」ってなに？」https://logmi.jp/tech/articles/45207(2020 年 3 月 20 日閲覧)

GUBA,E.&LINCOLN,Y.(1989) *Fourth Generation Evaluation*. Sage Publication,NewburyPark,

GUBA,E.(Ed,)(1990)*The Paradigm Dialog*.Sage Publicatiens,Newbury Park.

Hammersley,M.&Atk-inson,P.(1983)*Ethnography: Principles in Practic*. Routledge,London.

Kubota,K(1991)*Developing an alternative Learning Environment:A Constructivist View*. Indiana University, Bloomington.

Lincoin,Y.&Guba,E.(1985)*Naturalistic Inquiry*. Sage Publications, Beverly.

Merriam,S.B.(1988)*Case Study Research in Edcation: A Qualitative Approach*. Jossey−Bass Publishers, San Francisco.

Mikolov, Tomas; et al. (2013). *Efficient Estimation of Word Representations in Vector Space*. arXiv:1301.3781 [cs.CL] https://arxiv.org/pdf/1301.3781.pdf%C3%AC%E2%80%94%20%C3%AC%E2%80%9E%C5%93(2020 年 3 月 20 日閲覧)

臺灣大學學術期刊資料庫