報　告

学生によるテキストマイニングの実践

星井 進介[1]

[1] 教育研究技術支援センター（Technical Support Center for Education and Research, National Institute of Technology, Nagaoka College）

Practice of text mining by students

Shinsuke HOSHII[1]

要旨

本校における独自制度であるプレラボは，低学年からの研究活動によって学習意欲を高めることを主目的として平成 27 年度から始まった制度であり，萌芽的テーマやセミナーなどを学科横断・学年横断的に全学生および教職員に向けて周知・提案し，取り組みに参加するメンバーを募集して活動が行われることに大きな特徴がある．プレラボは年間 10 件以上のテーマが提起され，活発な取り組みが進められている．関心のあるテーマであれば学科の枠を越えて他学科の教員らが提案する研究活動にも参加でき，学科横断的な活動が展開されている．プレラボ活動の成果は各種コンテストや学会などでの受賞歴もあり，多くの学生が活躍している．筆者は令和 4 年度に「テキストマイニングの実践　ー社会の中のことばを調べようー」というテーマのもとで 2 名の学生とともにプレラボ活動を実施した．テキストマイニングとは，文章データなどを対象として，テキスト同士の関連性や連鎖を見いだすための技術の総称であり，このプレラボ活動では，参加した学生自身が気になる事柄や好きなテーマを選んで，テキストマイニングの手法の一つであるテキスト分析を取り上げ，実践した．本報では，当該テーマのプレラボ活動の実施状況ならびにテキストマイニングによる計量テキスト分析の結果について報告する．

Key Words : *Pre-laboratory System, Text mining, Quantitative text analysis, Koshi-dai gakuen dayori, Novel*

１．はじめに

プレラボは長岡高専における研究教育活動の活性化を目的として平成 27 年度から運用が始まった制度であり，テーマを持った教職員が全学的に参加者を募集して活動するものである [1),2)]．例年，プレラボは多くの活動提案がなされ，令和 4 年度は 17 件，令和 5 年度は 12 件のテーマが提案され，活動が行われている（令和 5 年 8 月 18 日現在）．

筆者は，令和 4 年度に「テキストマイニングの実践　ー社会の中のことばを調べようー」というテーマでプレラボを提起し，参加学生とともにテキストマイニングの一手法である計量テキスト分析を取り上げて，選定した課題における文字データの分析を目的とする活動に取り組んだ．

テキストマイニングとは，文章データなどを対象として，テキスト同士の関連性や連鎖を見いだすための技術の総称である [3),4)]．ここで，マイニング（mining）とは採掘を意味しており，鉱山から有用な鉱物資源を採掘して取り出すように，テキストデータの中から有用な情報を取り出すことがテキストマイニングの由来となっている．テキストマイニン

グは，(1)手動で解析することが困難な大量のデータを対象とした分析処理が可能，(2)言葉を統計的・数値的に処理することによってデータの客観的な解釈ができること，(3)文章データの数量化と可視化が可能なこと，といった点が特徴として示され，社会学や経営学，都市工学，教育科学，医療・看護などの幅広い分野で活用されている．また，映画のセリフやテレビ番組の字幕情報，SNS の投稿情報を解析した内容も報告されており，分析結果の活用成果を目にする機会も増えている．

令和 4 年 1 月 1 日の新潟日報に掲載された記事[5]は，新型コロナウイルスの流行が人々の生活や価値観の変化に及ぼす影響を調査した結果をテキストマイニングで分析した内容が報告されている．291 人による回答の自由記述の結果からは，「コロナ禍」，「大切さ」，「生活スタイル」，「気付かす」，「思い知らす」といった言葉が特徴的なものとして挙げられている．これは，コロナ禍によって人とのふれあいの大切さや，当たり前な日常生活の尊さに気付いたという内容によるものである．

日本経済新聞では，活性化した職場のあり方を探るために社員のクチコミに着目し，テキスト分析をした結果が示されている[6]．評価の高い企業におけるクチコミは，「共感」，「フラット」，「自由闊達」といった頻出ワードが浮かび上がり，社員間で壁のない職場環境の姿が映し出されている．一方で評価の低い企業では「ワンマン」，「イエスマン」などの言葉が示され，閉塞感を感じさせる．

また，インターネットのウエブサイト上にもテキストマイニングを活用した分析事例がある．例えば，映画「刀剣乱舞」のセリフをもとにしたテキストマイニング[7]，「秘密のケンミン SHOW 極！」というテレビ番組の字幕情報についてのテキストマイニング[8]，アニメ「プリキュア」シリーズのサブタイトルのテキストマイニング[9]，ゲームソフト「ガンハザード」内の文章を対象にしたテキストマイニング[10]といった様々な分析成果がウエブ上で述べられている．

さらに，本校 1 年生の科目「基礎情報処理」の教科書「高等学校　情報 I 」[11]においても，データ分析に関する章の中でテキストマイニングが取り上げられており，文字データから有用な価値を引き出すテキストマイニング技術について概説している．ここでは，自由なフォーマットで記述されたテキストデータを調べて有用な情報を取り出すことをテキストマイニングと称するとして，テキストデータは情報が整理されていないためにデータを分析する前処理として，解析ツールを用いて言語データを最小単位で区切る形態素解析という作業を行う必要があること，そして，その形態素解析で区切られた形態素に対する様々な統計データを計測して，対象データ中の単語の出現頻度の分析などがなされることが説明されている．

このように社会の中の身近なところでテキストマイニング技術が適用されている．本報は，テキストマイニングをテーマとしたプレラボ活動の実施状況ならびに参加学生が取り組んだテキストマイニングの分析結果について報告する．当プレラボは，様々な言葉を含むテキスト文章を対象として，潜在的な課題や今まで捉えることのできなかった言葉同士の関連性を見いだすことを目指して，テキストマイニングの一分野である計量テキスト分析の手法を用いた実践的な調査・分析を行うことを目的とするものである．プレラボ活動にあたっては，テキスト分析ツールである KH Coder の分析操作技術の習得を進めながら，プレラボ参加学生自身が興味があって調べてみたいことをテーマとして取り上げて，対象のテキストの分析作業を進める計画である．これらの取り組みを通じてテキストマイニングについての理解を深めるとともに，参加学生自身が選んだテーマを対象としたテキストの分析を試みることで，テキストマイニングの実践的活動を執り行う．

本報の構成は以下のとおりである．第 1 章「はじめに」では，プレラボ活動とテキストマイニングについて概説した．第 2 章「テキストマイニングの現在」では，これまでのテキストマイニングに関する分析事例を挙げて，社会の中の幅広い分野で適用されているテキストマイニングの概要について述べた．第 3 章「テキストマイニングの実践」では，令和 4 年度に筆者が代表者として実施したプレラボ活動「テキストマイニングの実践　ー社会の中のことばを調べようー」における活動内容報告として，当該プレラボの実施状況ならびに参加学生が行ったテキストマイニングの実践例である本校の「学園だより」に記載された文章のテキスト分析結果と小説『羊と鋼の森』の分析結果を報告した．第 4 章「おわりに」では，本報のまとめについて述べた．

２．テキストマイニングの現在

第 1 章では，新聞記事やウエブサイトに掲載されていた一般社会で広く利用されているテキストマイニングの具体例を見てきたが，ここでは，論文や学

会発表などの資料を基に学術分野におけるテキストマイニングの分析事例についての調査結果を示す．これら既往の研究成果の報告事例を検討することで，テキストマイニングの手法を活用した分析アプローチについての知見を深めることができると思われる．

【社会学分野】

住民を巻き込んだ地域づくりのあり方を構築するには，自治体と住民との相互理解が求められる．山下は，首長の言説と地域構想を規定した総合計画との関連に着目し，どのような内容が広報を通して自治体から発信されているのか，また，その内容は地域計画とどのように結びついているのかを問題意識として捉え，テキストマイニングの手法を適用した分析により検証した[12]．首長の言説と総合計画の文脈を比較すると，当該地域における自然災害や社会的事象，国レベルの政策変動に対する首長の見解が述べられ，それらがのちの総合計画の内容に反映されていることが認められた．首長の言葉は総合計画のように網羅的ではないが，その時期に重要視していること，将来の総合計画に重点的に組み込まれるようなキーワードが散在していることが確認できた．

ソーシャルワークにおけるニーズ把握やソーシャルワーカーの業務実態を明らかにする実証研究の場合には，量的研究に基づくアンケート調査の実施などの手法が取られるが，特に自由記述の場合などは回答データ量が膨大になることが多い．このような場合，テキストデータ分析ソフトを用いたアプローチがなされる．日和は，研究法における信頼性や妥当性の意味を問い直し，ソーシャルワーク研究における質的研究のあり方について考察することを目的としてテキストマイニングによる分析手法を検証し，その可能性を述べるとともに，他の質的研究法の併用による多角的な研究アプローチの必要性を説いた[13]．

【経営学分野】

近年，企業は様々な環境活動や社会貢献活動を行っており，それらに関する情報は環境報告書で公表されている．中邨らは，この環境報告書に注目し，報告書の使用単語がどのように変化したのかをテキストマイニングで分析した[14]．環境報告書の文言から出力される言葉の特徴と傾向を見ると，東日本大震災の発生や，社会的責任に関する ISO26000 の発行といった事柄が，企業の環境側面の方向性に影響を与えていたことが読み取れた．

【都市工学分野】

都市に関する市民の意識や都市のイメージについては，全国の自治体において市民アンケート調査が実施されており，その内容は都市計画分野の研究者や自治体の担当者の重要な検討課題の一つとなっている．森田らは，市民アンケート調査の自由記述データに着目し，テキストマイニングをはじめとする自然言語処理技術を用いた定量的で再現性のある分析を試みた[15]．複数の選択肢の中から回答を選ぶプリコードデータと自由記述データとの関連性分析から，居住地の生活の満足度に応じ，自由記述回答にどのような語を用いて都市イメージを記述しているかが把握された．そして，単語の出現頻度と共起性により，生活の満足度との関係において，都市イメージを定量的に捉える一つの方法を示したと結論付けた．

福岡市は過去に大規模な渇水に見舞われたことから，市民の節水意識は全国的に見ても高いと言える．市民の水に対する関心を調べるため，福岡市役所はアンケート調査を実施した．横田らは，その調査結果について自己組織化マップを用いて検証した[16]．その結果，(1) 20～30 代の若年層および市の中心部に住む人々は節水への意識が低い，(2) 70 代以上の高齢者は節水に積極的で，水道水をそのまま飲んでいる，といったことが明らかになり，世代と居住地域による属性の違いが節水意識や行動に深く関係していることがわかった．

【教育科学分野】

平成 27 年の公職選挙法改正に伴って選挙権年齢が引き下げられ，18 歳以上となった．これにより主権者教育のあり方が問われるとともに，高等教育機関を中心とした取り組みが進展している．林は，主権者教育に関するアンケート調査を実施し，その結果について計量テキスト分析による分析を試みた[17]．回答者である学生の意見から見た主権者教育の方向性について探索し，先行研究として行われた投票参加モデルの有効性を確認した．併せて，選挙への関心については，家族や親からの影響に関する要因も見られたことから，周囲の人々による投票行動の重要性についても指摘している．

教育機関の授業評価アンケートで学生の意見として自由記述の回答を求めてきたが，大量のテキストデータについて，分析者の恣意的・主観的な解釈を避けて，客観的に全体的な傾向を把握することは極めて難しい．そこで越中らはテキストマイニングの手法を用いて分析を行い，自由記述回答の可視化を

試みた[18]．分析は，頻出語の確認や単語の共起性の探索などを行い，自由記述の要約に取り組んだ．その結果，「この授業はここがいい」については，体験活動・グループ活動の実施，参加・発表の機会が多い，といった意見が，そして，「こうすればもっとよくなる」では，スライドの使用法や板書の仕方，授業の進め方に関する意見が挙げられた．これらの結果は，今後のアンケート質問項目改定の際の参考としても有用であると示している．

【医療・看護分野】

岩佐は，国内の女性の健康課題に関する研究パラダイムの変化をとらえ，今後の示唆を得ることを目的として，1980 年から 2014 年の医学系論文のタイトルの語をテキストマイニングで分析した[19]．「女」「婦」「母」をキーワードとした検索の結果，29,082 論文が抽出された．分析対象のうち，妊産婦に分類された論文が全体の 30.9%の割合であり，特に 1990～94 年における割合は 40.0%を占めていた．また 2000 年以降は，心理，育児，睡眠疲労に関するテーマが増加した．これらのことから，女性の健康課題に関する研究パラダイムは，産み育てるというテーマから変遷し，2000 年代に女性の身体についてのライフサイクルや生活習慣に関する事柄に，そして 2010 年頃に介護，育児等の役割の困難さに関わるテーマにシフトしたとしている．

今井らは，新人看護婦が抱く現在の職場を去りたいと思った理由に関するテーマについて，定量的なテキストマイニングと定性的な質的帰納的分析とを併用した混合分析法によりアプローチし，両者の分析法の共通点と相違点を検討するなどして混合分析法の有用性を明らかにすることを目指した[20]．質問紙調査における自由回答文において，現在の職場を去りたいと思った理由について回答した 39 名分のデータを対象とした結果から，テキストマイニングと質的帰納的分析による分析を比較検討したところ，表現上の違いはあるものの，両者の分析法はともに研究ターゲットとなるトピックを抽出できることが判明した．分析結果では，勤務条件の不満，理想と現実のギャップによる精神的葛藤，職場における人間関係の悩みなどの概念が読み取られ，質的データを検討するにあたって，テキストマイニングと質的帰納的分析を用いることは分析結果をより堅固なものに導くものと考えられると述べている．

【観光分野】

石井らは，観光地でのアンケート調査を対象として，個々の項目ごとに内容の詳細を基礎統計的に分析し，それぞれの質問項目と回答者の属性情報とのクロス集計や統計的処理を実施した．さらに，観光客の全体評価構造を可視化するためにテキストマイニングによる分析を試みた[21]．テキストマイニングでは，単語の出現頻度と共起関係をもとにグラフ化し，議論の鍵となる言葉を抽出する解析作業を行った．その結果，観光客の性別や年齢，職業と観光形態との関係，そして，海や温泉，自然といった項目の観光魅力度との関連性についての検証から，今後のプロモーション活動に資する有用な知見を導出した．

旅行ガイドブックやテレビの情報番組，小説，ドラマなどのメディアは，消費者である観光客がもつ観光地のイメージを提示し，制御しうる存在として捉えることができる．そのため，観光産業に関わる研究においては，それらのメディアが示す様々な言葉が分析対象となることから，テキスト分析が有効なアプローチの一つとなる．有馬は富士山観光を題材として，旅行ガイドブックにおける富士山観光のイメージの変化をガイドブックの目次のテキスト分析を通して明らかにすることを目指した[22]．分析の結果から，国内のリゾートブームに牽引された富士山観光のイメージ形成という時代から，観光と地域振興とを結びつけた時代へと変化していき，その後は「登山」という言葉とともに「聖地」や「B 級」，「グルメ」といった単語に表わされるように，近年の多様性を背景とした様々なレジャーや楽しみを包含した富士山観光へと認識が変わっていったことが見て取れた．

【SNS などに関する分野】

川合らは，様々な世代の多様性を考慮した街づくりのため，人々が日常的に情報発信を行っている Twitter に注目し，複数の都市に関する Twitter 上の特徴的な語句を抽出して単語の経時的変動と地域特性について分析を実施した[23]．収集したツイートは 2019 年の 1 年間分で 2,977,437 ツイートの大規模テキストデータとなった．これらのツイートについて，出現頻度の高い言葉に対する Jaccard 係数を用いた階層的クラスター分析，言葉同士の共起性分析，対応分析などの解析による分析結果から，5 つの市の中では，平塚と茅ケ崎，藤沢は類似していること，鎌倉と逗子が特徴的な傾向を持つことがわかった．クラスター分析や対応分析によって全体傾向の視点から単語抽出を行ったが，ツイートが肯定的なのか，否定的なのかの評価を行うといった個別のツイート

の抽出と評価をより細かく実施する必要があり，詳細な街づくり要望についての具体化の作業が求められると論じている．

中野らは，企業 Twitter アカウントの現状と運営における意識を分析し，今後の企業の Twitter アカウントの在り方を考察することを目的として，企業の Twitter 担当者に対する意識調査のアンケートとツイートデータのテキストマイニングを行った[24]．現代は，マーケティングにおいて共感が重視される時代であり，企業情報の拡散と顧客との交流の場として Twitter が有効活用されている．テキストマイニングを用いて頻出語の抽出と階層的クラスター分析を行った結果，顧客への感謝の言葉が頻出語の上位に示された．これは，顧客との日常に大きな関わりがあることが認められ，Twitter を有効活用している企業と消費者との関係性を捉えることが可能であると示唆された．また，分析結果からは，企業の Twitter 担当者，いわゆる「中の人」同士のコミュニティーが存在していることがわかった．

ここでは学術領域におけるテキストマイニングの既往研究事例について検討した．その結果，テキストマイニングの手法が様々な分野において有効な分析手段の一つとなっていることが見て取れた．これらの事例を知ることにより，テキストマイニング学習の意義と重要性を改めて認識することができた．

３．テキストマイニングの実践

３．１　プレラボ活動の推進

令和 4 年度に実施したプレラボ「テキストマイニングの実践　－社会の中のことばを調べようー」では，テキストマイニングの一手法である計量テキスト分析という手法を用いて，プレラボ参加学生が自ら選んだテーマを分析対象としてテキストマイニングを試みた．

このプレラボは，文系的な課題領域である言葉や文章といった資料を対象としつつ，パソコンを用いたデータ管理や分析，統計的処理といった理系的スキルを通じた解析手法をとることで，文系・理系の枠にとらわれない分野横断的な視野で課題にアプローチする文理融合的な取り組みと言える内容になっている．

今回は，計量テキスト分析ソフトの KH Coder を用いて，本校の「学園だより」と小説『羊と鋼の森』を題材としてテキストマイニングを行った結果を報告する．

この節では，当プレラボの進展状況を説明するとともに，計量テキスト分析ソフト KH Coder を用いた分析の流れについて概説する．次の 3-2 節では，毎年 3 月と 4 月に発行されている「学園だより」を題材とした分析を行った結果を報告する．新入生に向けた歓迎の言葉や，教員から卒業生に送る言葉など，高専に通う人々ならではの個性ある文章が並ぶ「学園だより」における学生と教員の書いた文章に注目し，3 年分のテキストを対象として分析を実施した．3-3 節では，宮下奈都著『羊と鋼の森』を対象としたテキストマイニングを実施した内容を述べる．主人公であるピアノ調律師の成長の物語で，作品に出てくる言葉から想像される作品像に焦点を置き，プレラボメンバーである学生自身が当該作品を読んで感じた作品像との比較・分析を試みた．

当プレラボ活動の取り組み状況について，以下のとおり報告する．

年度始めにプレラボ提案書を作成し，当該テーマでのプレラボ活動の実施を申請した．5 月と 6 月にそれぞれプレラボ参加希望の学生からの申し出があったので，学生向けに活動内容を説明し，合わせて 2 名の学生が当プレラボに参加することになった．その後，テキストマイニングの分析対象課題などについて話し合いを行い，各自が取り組むテーマを決定した．テキストマイニングの実施のために計量テキスト分析ソフト KH Coder のダウンロードを行い，学生向けに使用法を説明した．そして，各自の分析対象となるテキストデータを作成して分析作業に取り掛かった．テキスト分析にあたっては，データの調整作業や各種分析方法の条件設定など，進捗状況の確認を進めるとともに，Teams を用いたデータ情報共有も活用してテキストマイニングによる分析作業の進展を図った．プレラボ活動で得られた成果は，令和 4 年 11 月 5 日（土）と 6 日（日）に開催された本校の学園祭である未工祭での活動成果報告においてポスター発表を行った．作成したポスターを図 1 に，成果報告の様子を写真 1 に示す．未工祭での発表後も，引き続きテキストマイニングを用いてテーマに対する分析作業を進めた．

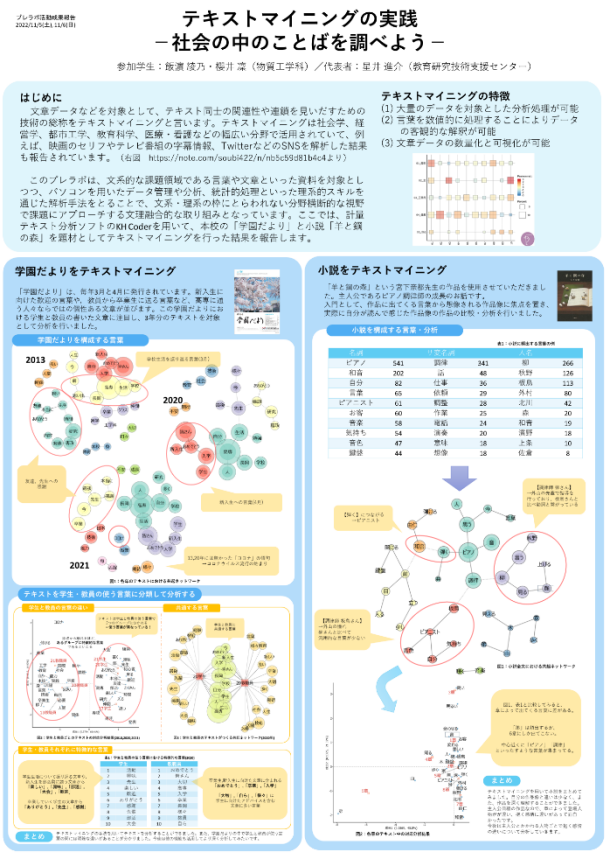


図-1　未工祭でのプレラボ活動報告（ポスター）

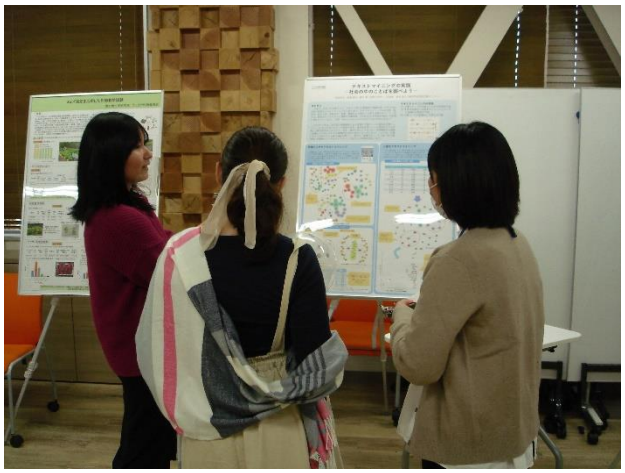


写真-1　未工祭でのプレラボ活動報告の様子

続いて，計量テキスト分析ソフト KH Coder を用いた分析の流れについて述べる．KH Coder は，アンケートの自由記述回答，インタビュー記録，新聞記事などの様々なテキスト文章の分析に使用可能なテキストマイニングのためのフリーソフトウェアであり，令和 5 年 9 月現在，KH Coder 3.Beta.07e 版が提供されている．KH Coder の使用法については，ウエブサイト（http://khcoder.net/）にチュートリアルのファイルがあるほかに，ソフトの設計者による解説資料[25]がある．ここでは一連の作業フローを以下のとおり説明する．今回，学生はこのような分析手順に従ってテキストマイニングの分析作業を進めた．

（1） 分析ファイルの設定

分析対象のファイルはテキスト形式で作成，準備する．本文中には，半角の「’」，「”」，「\」，「<」，「>」，「|」などの語を含まないようにする．KH Coder では，「．」によって文を，改行によって段落を認識する．必要に応じて，見出し部分を括るための H1 から H5 までのタグを設定できる．例えば，H1 を部，H2 を章，H3 を節，H4 を項として設定すれば，実際の分析の際に，これらのタブごとの分析が可能になる．

（2）分析ファイルの選択

KH Coder を起動して，メニュー，プロジェクトから「新規」を選択して分析するファイルを開く．2 回目以降は，「開く」を選択すればよい．ここで，分析対象ファイルの総抽出語数や文，段落の集計数などを確認する．

（3）前処理の実施

メニューの前処理から前処理の実行を選択し，ファイルの前処理を行う．前処理を実施して問題がなければ分析が可能となる．

（4）頻出語リストなどの作成

メニューのツール，抽出語，抽出語リストをクリックして，頻出語リスト（よく出現する語），品詞別の抽出語リストを作成し，分析ファイルに含まれるテキストの確認を行う．

・地名や人名など，一般的ではない特殊な語がきちんと抽出されているか．

・同じ言葉なのに，大文字－小文字の違いで異なる語として認識されていないか．全角－半角の違いで異なる語として認識されていないか．

・一つの言葉として認識されるべき語が複数に分割されて抽出されるなど，不適切な抽出がなされていないか．

問題があった場合は，前処理，語の取捨選択をクリックして，正しく言葉が抽出されるように語の強制抽出を設定するなどして対処する．設定を変更した場合は，もう一度前処理を実行して，改めてテキストの確認を行う．

（5）KWIC コンコーダンスによる確認

注目した語がどんな文脈で使われているのかを KWIC コンコーダンスで確認する．メニュー，ツール，抽出語，KWIC コンコーダンスをクリックして，検索したい語を入力する．抽出語リストで抽出され

た語をクリックしても KWIC コンコーダンスが開くので，そこでも確認が可能である．

（6） 共起ネットワークの作成

メニュー，ツール，抽出語，共起ネットワークをクリックして，共起ネットワークの設定画面を開く．集計単位などの条件を設定して OK をクリックすると共起ネットワークが作成される．共起ネットワーク下部の調整ボタンなどによって表示設定を調整する．最初の共起ネットワーク設定画面の条件を変更して，様々な共起ネットワークを作成して，語句の結びつきを確認する．

（7） 特徴語の分析

メニュー，ツール，外部変数と見出し，リストをクリックする．外部変数と見出しの設定画面から条件を選択して，▽特徴語，一覧（Excel 形式）をクリックすると，Excel ファイル上に分析対象条件の特徴語が表示される．

（8）対応分析の実施

メニュー，ツール，対応分析をクリックすると，対応分析の設定画面が開く．各条件を設定，選択して OK をクリックすると対応分析の結果が図示される．分析対象ファイルの語とタブで区切られた属性との関係が一つの図の中に位置付けられるので，全体の語同士の関連性を確認する．

これらの各分析の実施を通して，分析対象ファイルのテキスト同士の結びつきや文書全体の文脈の流れ，タブで区切られた部分ごとの特徴などを考察，探究する．

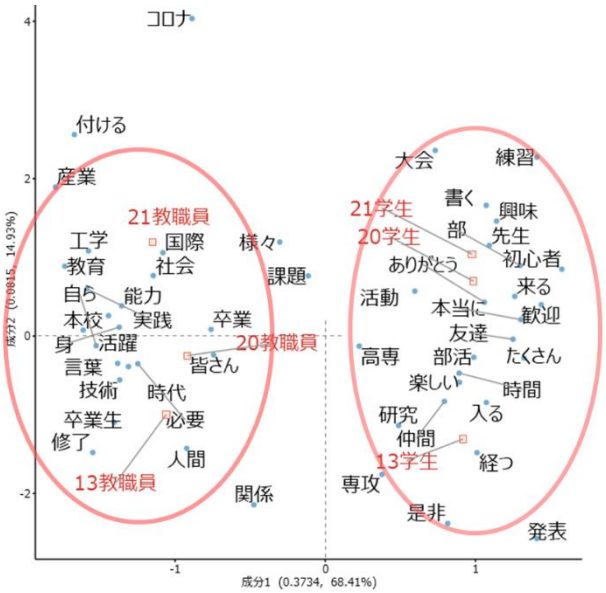
３．２　「学園だより」のテキストマイニング

本校では，年度初めの 4 月と年度末の 3 月の年 2 回「学園だより」を発行している．ここでは，「学園だより」に掲載されている記事を対象としてテキストマイニングを行い，文章中で学生や教員の使う言葉と言葉の間にどのような相関があるのか，学生と教員の使う言葉に違いはあるのかについて，KH Coder による分析から明らかにすることを目指した．分析対象としたのは，2013 年，2020 年，2021 年の 3 年分の「学園だより」である．

KH Coder の分析作業手順に従い，「学園だより」の文章をテキストデータとして KH Coder に読み込み，語句をチェックした後，前処理を実行して分析可能なデータとして取り込んだ．それぞれの年ごとのテキストの抽出語から共起ネットワークを作成し，どのような言葉が使われているのか視覚化した．「学園だより」は主に年度末（3 月）と新学期のはじめ（4 月）に発行されるため，卒業や入学を祝う「おめでとう」，「ありがとう」，「思い出」，「生活」などの言葉が含まれているのが確認できた．また，2021 年には今までの年には無かった「コロナ」という言葉が新しく含まれていることから，その年に特徴的な言葉も確認することができた．より詳細に見ると，作成した共起ネットワークからは，「長岡－高専－生活－思い出」といった学校生活を振り返る言葉のつながりや，「先生－最後－感謝－卒業」や「友達－本当に－ありがとう－先生」などの先生や友達への感謝の言葉，そして，「新入生－皆さん－おめでとう－入学」という新入生に向けた言葉が見て取れた．また，2021 年には「コロナ－授業」という共起性が見られ，新型コロナウイルスによる学校生活への影響が示された．

次に，各年のテキストを学生と教員ごとに分類し，抽出語を対応分析により分析した（図 2）．

図の左側には主に教員に特徴的な言葉が，右側には学生に特徴的な言葉がまとまっていることから，学生と教員の使う言葉には明らかな違いがあると読み取れる．抽出語リストからは，特徴語を出現数順に詳しく確認することができた．学生は新入生歓迎や卒業など，学校生活について話していることが多いため，特徴的な言葉には「楽しい」，「部活」，「興味」，「先生」，「ありがとう」などが含まれているのが認められた．それに対して教員は学生に向けて発する言葉が多いことから，「おめでとう」，「皆さん」，「大切」，「自ら」などの言葉が含まれていた．また，学生と教員に共通する言葉は共起



図−2　「学園だより」の対応分析結果

ネットワークの条件を設定を変えることにより確認できた（図 3）．

このように，対応分析と特徴語の抽出結果からは，「学園だより」の文章中で学生が使う言葉と教員が使う言葉が明らかに異なっていることがわかった．学生は，学生生活を振り返る内容や後輩である在校生に向けた文章から「活動」，「興味」，「部活」，「楽しい」，「先輩」，「友達」，「研究」，「仲間」といった言葉が挙げられており，卒業していく学生からは，「ありがとう」，「感謝」，「先生」などの言葉が挙がった．教員のほうは，卒業生や新入生に向けた「おめでとう」，「卒業」，「入学」とともに，学生に向けたアドバイスとして「大切」，「自ら」，「活躍」，「実践」などの言葉が示されていた．そして，「工学」，「国際」，「技術」，「産業」，「時代」，「必要」といった，これまでの学修成果と今後に向けた視点についても言及していた．

今回の分析では，抽出語に焦点を置き，共起ネットワークや対応分析を用いて言葉同士や分類ごとの相関を分析することができた．

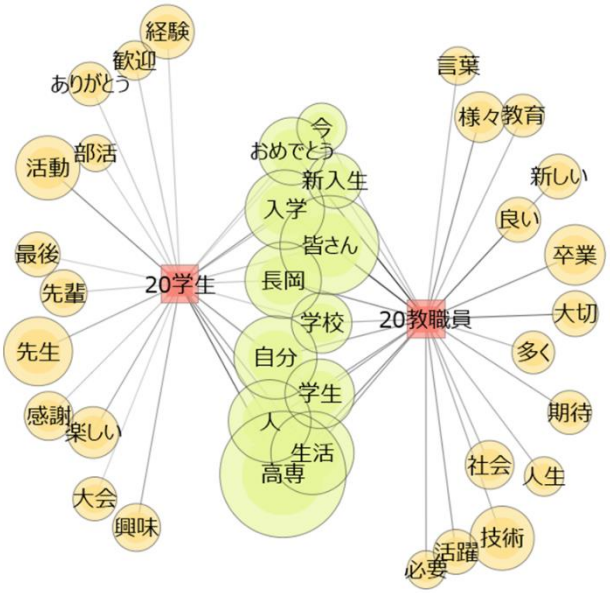


図-3　「学園だより」の共起ネットワーク

３．３　小説のテキストマイニング

2015 年に文藝春秋社より発行された宮下奈都の小説『羊と鋼の森』を題材としたテキストマイニングを試みた．2016 年に本屋大賞を受賞した同書は，243 ページに及ぶ長編小説である．主人公である外村は，高校生の時にピアノ調律師の板鳥と出会ってピアノの調律に魅せられた．調律師として働き始めた外村は，個性豊かな先輩たちや双子の姉妹に囲まれながら音と向き合い，人として成長する姿をつづった内容となっている．

頻出語の分析からは，小説のテーマに即して「ピアノ」という語が 541 回，「調律」が 341 回出現している．人名では，主人公である外村の先輩調律師である柳と秋野という名前が多く現れており，主人公がピアノ調律師に魅せられたきっかけとなった存在である板鳥の名前が続いている．そして，双子の姉妹である由仁と和音の名前も多く出現する．

小説全文における共起ネットワークを作成すると，それぞれの語句同士のつながりが明らかになった（図 4）．

この小説の中心的な意味を表す「人－思う－ピアノ－音－弾く－調律」という語句の関連性が中央付近に表わされている．そして，この共起ネットワーク中の「調律」という語は，柳と板鳥という二人の調律師とゆるやかにつながっている．柳については，「柳－言う－秋野」，「柳－見る－顔－上げる」というかたちで表されており，主人公の先輩として指導を行っていることから動詞の言葉でつながっており，主人公と直接コミュニケーションをとりながら関係性を築いていることが見て取れる．一方で板鳥については「板鳥－ピアニスト－音色」，「板鳥－ピアニスト－自分－気持ち」という語句の関わりが見られた．これは，主人公・外村と板取の関係性が，先輩と後輩という直接的な関係ではなく，自分自身が調律師としての道を歩むきっかけとなった憧れの存在として板鳥を捉えていることを示しているものと思われる．その他にも，「鍵盤－開ける－前－

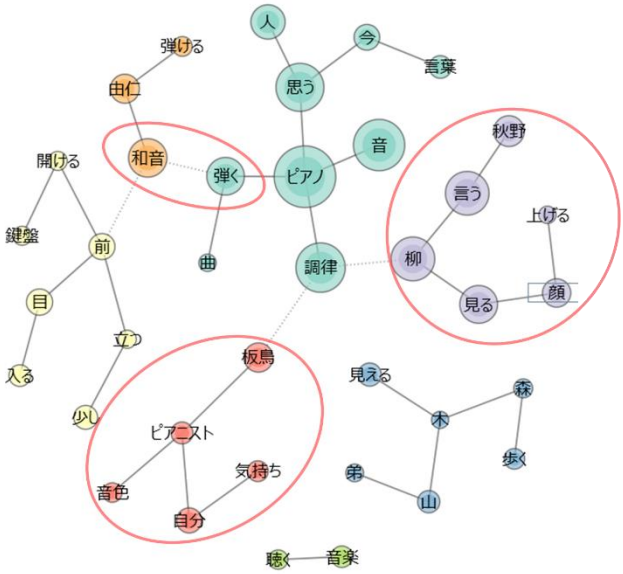


図-4　小説『羊と鋼の森』の共起ネットワーク

立つ」や「聴くー音楽」といったピアノに関する言葉，「見えるー木ー山ー森ー歩く」という，この小説の舞台となっている北海道の自然豊かな情景を表す言葉の共起性が見られた．

続いて対応分析の分析結果を示す．分析対象とした『羊と鋼の森』は章や節などがなく，いわば全体が一つの章として構成されているが，ここでは分析者がテキストマイニングにあたって，内容をふまえた上で 8 つの章に便宜的に分割し，各章における語句の関係性を考察した．対応分析の結果からは，多くの章で使用される言葉が図の中央部分に集まっており，各章で使用される言葉が似通っていることを表している．一方，3 章と 6 章は中央から離れた場所に位置している．これは，両章で使われた言葉が小説全体で見たときに特徴的であることを示しており，3 章では「高い」，「明るい」，6 章では「弟」という言葉が他の章と比して特徴的であることを示している．図の中央付近には「ピアノ」，「調律」，「鍵盤」，「弾く」，「ピアニスト」といった言葉とともに，登場人物である「板鳥」，「柳」，「外村」といった名前も見られる．双子の姉妹である「由仁」と「和音」は 7 章付近に位置しており，小説の後半で登場して主人公と強く関わりを持つことを読みとることができる．

最後に，今回，小説をテキスト分析した学生は次のように感想を述べている．

（以下，学生の感想）自分自身で事前に読んで抱いていた作品像との違いは少なく，より深く作品を理解することができた．主人公目線の作品なので，各章によって登場人物が違って，抱く感情に違いがあって面白かった．今後は，主人公と関わる人物ごとで抱く感情の違いについて分析していきたい．

４．おわりに

プレラボ制度はテーマを持った教職員が全学的に参加メンバーを募集して活動するもので，学科・学年を越えて低学年から高学年までの研究活動をシームレスにつなぐ他に例を見ない制度である．萌芽的研究・学生教育支援・低学年からの研究活動の推進を目指すプレラボ制度のもと，テキストマイニングの一手法である計量テキスト分析の実践をテーマとして掲げ，活動に取り組んだ．

本報では，テキストマイニングの現在地を既往の研究事例を挙げて報告するとともに，令和 4 年度のプレラボ活動「テキストマイニングの実践　ー社会の中のことばを調べようー」で実施した学生によるテキストマイニングの分析例を報告した．「学園だより」のテキストマイニングでは，教員や学生が記した記事の中でどのような言葉が使われているのかを見える化し，テキスト間の関連性を見いだすことができた．小説のテキストマイニングでは，243 ページで 118,000 字を越える内容のテキスト分析を試み，主人公と周囲の人々との関わりや語句同士のつながりが明らかになった．

今回のプレラボにおける取り組みは，参加学生が初めての経験ということもあり，一部のデータ解析に関して検討不足という面が見られたが，学生自身が興味，関心があるテーマを題材として選んでテキスト分析を行うという一連のプロセスに主体的に取り組むことによって，分析テーマに対して，これまでと異なる新たな視点を獲得できた．それにより学生自身の自発性，実践性，探究心の向上に寄与することができたと思われる．テキストマイニングをテーマとしたプレラボについては，令和 5 年度も引き続き実施している．今後も同様の仕組みのもとで活動の進展を図りたいと考えている．

参考文献

1) 赤澤真一，田原喜宏，桐生　拓，土田泰子，床井良徳，村上祐貴，池田富士雄，井山徹郎，外山茂浩：プレラボ制度を活用した全学的な教育研究活動の推進，長岡工業高等専門学校研究紀要，第 52 巻，pp.78-82，2016.
2) 桐生 拓，赤澤真一，田原喜宏，冨樫（新藤）瑠美：プレラボ制度による学科横断型教育・研究活動の推進と波及効果の検証，工学教育研究講演会第 65 回年次大会，講演番号 2A13，要旨集 pp.154-155，2017.
3) 小林雄一郎：ことばのデータサイエンス，朝倉書店，2019.
4) 那須川哲哉，吉田一星，宅間大介，鈴木祥子，村岡雅康，小比田涼介：テキストマイニングの基礎技術と応用，岩波書店，2020.
5) 新潟日報：感染禍　生活や価値観　本社アンケート，2022 年 1 月 1 日
6) 日本経済新聞：カイシャの未来　社員の声　聞こえてますか，2023 年 1 月 23 日
7) “「映画 刀剣乱舞」の全セリフをテキストマイニングを使って分析してみた．”，https://note.com/soubi422/n/nb5c59d81b4c4，2022 年 5 月 10 日確認
8) “秘密のケンミン SHOW 極！兵庫姫路えきそばがちょっと変！？高知絶品いも天[字][デ]…の番組内容解析まとめ”，https://dnptxt.com/television-show/variety/post-

26882/，2022 年 5 月 10 日確認
9) “プリキュア 600 話分のサブタイトルを分析したら見えてきた事．テキストマイニング分析より”，https://prehyou2015.hatenablog.com/entry/2016/05/16/102809，2023 年 5 月 18 日確認
10) “海/が/汚染/さ/れる/ぞ/！ゲーム内文章を対象にしたテキストマイニングの試み”，https://god-bird.net/research/gunhazardmining.html，2023 年 5 月 18 日確認
11) 坂村　健　他：高等学校　情報 I，数研出版，2022.
12) 山下良平：自治体が発信する情報の構造分析に対するテキストマイニングの可能性，農村計画学会誌，31 巻，論文特集号，pp.267-272，2012.
13) 日和恭世：ソーシャルワーク研究におけるテキストデータ分析に関する一考察，評論・社会科学，106 号，pp.141-155，2013.
14) 中邨良樹，高林直樹，大場允晶，山本久志，丸山友希夫：テキストマイニングを用いた企業・業種分析の一指標，横幹，vol.9，No.2，pp.95-103, 2015.
15) 森田哲夫，入澤 覚，長塩彩夏，野村和広，塚田伸也，大塚裕子，杉田 浩：自由記述データを用いたテキストマイニングによる都市のイメージ分析，土木学会論文集 D3（土木計画学），vol.68，No.5，pp.I_315-I_323，2012.
16) 横田いずみ，井料隆太，井芹慶彦，広城吉成，神野健二：自己組織化マップを用いた福岡市民の水に関するアンケート調査結果分析，水工学論文集，第 53 巻，pp.553-558，2009.
17) 林 健一：主権者教育に関するアンケート調査結果からみた大学教育機関の課題，中央学院大学社会システム研究所紀要，第 20 巻，第 2 号，pp.41-61，2020.
18) 越中康治，高田淑子，木下英俊，安藤明伸，高橋 潔，田幡憲一，岡 正明，石澤公明：テキストマイニングによる授業評価アンケートの分析　―共起ネットワークによる自由記述の可視化の試み―，宮城教育大学情報処理センター研究紀要，第 22 号，pp.67-74，2015.
19) 岩佐由美：女性の健康課題に関する研究パラダイムの変化 ―1980～2014 年　医学中央雑誌収録論文のタイトル分析から―，日本保健科学学会誌，Vol.21，No.4，pp.181-191，2019.
20) 今井多樹子，高瀬美由紀，佐藤健一：質的データにおけるテキストマイニングを併用した混合分析法の有用性　―新人看護師が「現在の職場を去りたいと思った理由」に関する自由回答文の解析例から―，日本看護研究学会雑誌，Vol.41，No.4，pp.685-700，2018.
21) 石井康夫，大久保あかね，鈴木大介：観光マーケティングにおける新たな分析手法の提案－伊豆半島の観光魅力度に関するテキストマイニング分析を事例として－，知能と情報，Vol.31，No.4，pp.745-753，2019.
22) 有馬貴之：旅行ガイドブックにみる富士山観光のイメージ変化 −『るるぶ富士山』の目次を対象としたテキスト分析−，地学雑誌，Vol.124，No.6，pp.1033-1045，2015.
23) 川合康央，池辺正典：SNS を用いた街づくり要望の分析 －湘南地区 5 市の Twitter を対象として－，デザイン学研究特集号，Vol.28，No.2，pp.36-39，2020.
24) 中野健秀，松尾爽世：企業 Twitter アカウントによる共感を生み出す呟きの分析，日本マーケティング学会カンファレンス・プロシーディングス，Vol.8，pp.182-189，2019.
25) 樋口耕一：社会調査のための計量テキスト分析【第 2 版】，ナカニシヤ出版，2020.

（2023. 9. 28 受付）