

テキストマイニング技術を用いた 東日本大震災に関する研究動向の分析

藏 田 典 子

1. はじめに

千年に一度といわれるほどの甚大な被害をもたらした東日本大震災（以下、大震災と略す）の発生から、2021年3月11日で10年を迎えた。死者と行方不明者の数は2万2千人を超え（災害関連死を含む）、いまだに約3万9千人に及ぶ人々が、47都道府県914市区町村で避難生活を続けている（2021年12月現在）。ボランティアによる被災地での支援活動やインフラの復旧に至るまで多くのセクターの関係者が復興のための取り組みに邁進する中で、学界では様々な分野の研究者が多様な知見を蓄積してきた（ここでは一例になるが、乾2016、今井2014、開沼2011、菅野2015、関2018、田中2016、西城戸・原田2014、原田・西城戸2015、日野2014、松井2017、松永2018、水野2017、山根2013など）。

とくに2021年は大震災から10年が経過したことの節目として、様々な研究領域において大震災に関するシンポジウムや特集が組まれ活発な議論の場が設けられるなど、改めて大震災が着目された時期であった。たとえば、日本地理学会2021年春季学術大会シンポジウム「東日本大震災10年——防災概念の変革期における地理学——」では、自然科学から行政、防災教育、災害伝承、文理融合研究など多様なテーマが扱われ（鈴木2021）、大震災に関する研究が極めて学際的な広がりを持つことが示された。また、社会学分野では、早稲田大学総合人文科学研究センター年次フォーラム「東日本大震災10年の軌跡と大規模災害からの復興をめぐって～新たな「日常」への模索～」が開催されるなど、多様な学術的視点から10年間で震災を振り返る動きがみられた。

研究者が被災地において調査を含む大震災に関する研究活動を行うことは、復興や被災者支援に有用な情報を提供するなど、建設的な影響を与える場合がある。例えば、宮城県岩沼市では、地元出身で事前に住民との信頼関係ができていた研究者が大震災被害の調査と復興に関する提言を行ったことで、集落移転などの方針が迅速に決定されたとの報告がある（河村2013）。また、東北地方から北海道へ避難した人々とその受け入れ自治体を調査した三宅ら（2013）は、アンケートや個別の聞き取り調査を用いて自治体の支援と避難者の需要が一致していないことを指摘し、実情に合った避難者支援の必要性を訴えた。

このように研究者の活動がもたらす効果を示されている一方で、多くの研究者が被災地を訪れて聞き取りなどの調査を行ったことで、被災者・避難者が感じる調査疲れもまた指摘されている（朝日新聞2012年11月11日）。被災地での住民参加やその合意形成に関して調査した河村

(2013) は、研究者や研究機関によって連日のように行われる調査が「学者疲れ」をおこし、学者に来て欲しくないという意味を態度で示した地域もあったと述べている。このように、多くの研究が精力的に実施されたことは、その成果の副産物として、負の影響ももたらしたことが示唆される。

アンケートや聞き取りなどの現地調査（一次研究）は被災の現状や避難者の実情を知るために不可欠な調査手法ではあるが、これらがもたらす負の影響を考慮すれば、現地への影響が少ない他の調査方法の導入についても検討されるべきだろう。その一例として、大震災に関する研究を俯瞰的に調査し、研究テーマごとの報告例の多寡などを可視化する文献研究（二次研究）が挙げられる。震災から10年が経た今、これまで実施された多様な大震災に関する研究を対象に、分野を横断し総合的に振り返り、研究者が大震災に関して残してきた研究成果の全体像をまとめる文献研究は重要と思われるが、筆者の知る限り、いまだ研究領域を横断した総合的なレビューはなされていない。

大震災に関連する研究は、分野や手法が多岐にわたることが、総合的なレビューを阻害していた要因ではないかと考えられる。この問題を克服し、多様な研究を網羅的に俯瞰した文献研究を実施するため、テキストマイニングの手法による定量的な解析は効果を持つと期待される。これまで大震災を対象にテキストマイニングが用いられた研究事例には藤森ら（2014）や新美ら（2018）、吉見（2015）などがあり、計算機による自動解析がもたらす成果が実証されている。しかし、大震災に関連して自然科学から人文社会科学までを網羅する研究を対象とし、テキストマイニングを用いて包括的な視点から大震災研究の動向を扱った先行事例は筆者の知るかぎり存在しない。

そこで本研究では、研究データベースから網羅的に取得される多数の学術発表タイトル（以下、研究タイトル）を対象としてテキストマイニングで特徴抽出を行う。さらにクラスター解析で既存研究の自動分類を行い、研究分野ごとの特徴を分析する。特に、大震災に関する研究において、どのような研究クラスターが存在しているのか特定すること、そしてそれらの研究クラスターに含まれる研究の数は過去11年間でどのような動向を示すのかを明らかにすること、この2つを主な目的として解析と議論をすすめる。これにより本研究は、分野横断的に大震災研究を取りまとめた成果を提供する。

2. 分析方法

2.1 解析デザインの全体像

図1は本研究における解析デザインの全体的な流れであり、その過程には、①データ収集、②データの前処理、③データのクラスター分類、④クラスターの定量的解析、という主に4つの段階が含まれる。①データ収集では、インターネット上のデータベースから解析対象となるデータを抽出する。②データの前処理では、研究タイトルの文字情報を単語に分割し、不要な文字情報を

を除外するためのデータの整形を行う。③データのクラスター分類では、統計的クラスタリングの手法を用いて文書データを複数のクラスターに分類する。そして、④ではクラスターごとの時系列的な動向解析やクラスター間の性質の比較分析などを行う。これら解析の各段階における方法論の詳細は以下のとおりである。

本研究における解析の実装は全て R version 4.2.0 上で行った (R Core Team 2021)。また、全体を通じたデータの操作・処理に関しては tidyverse というパッケージ群に準備されているパッケージ (Wickham et al. 2019) と、それらと関連する lubridate, magrittr, tidytext というパッケージ (Grolemund and Wickham 2011; Bache and Wickham 2020) を用いて行った。これら以外に個別の解析に特異的に利用した R のパッケージおよび関数に関しては初出の都度に引用を記す。

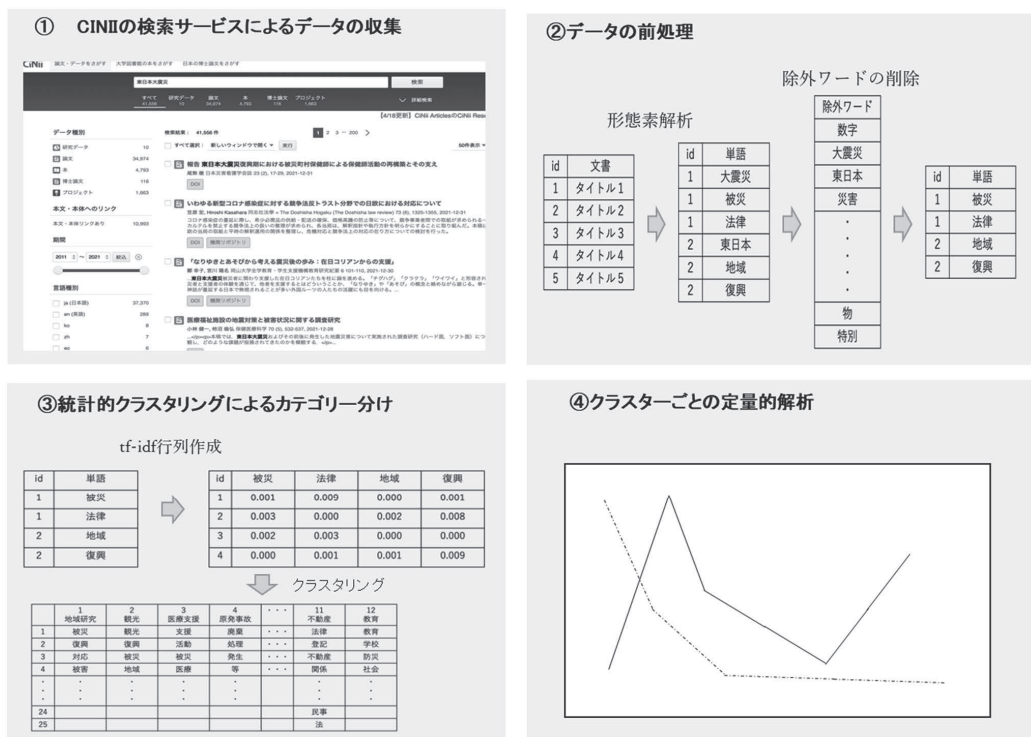


図1 テキストマイニング解析の模式図

2.2 データの収集と整理

本研究では、大震災に関する研究として、研究タイトルに「東日本大震災」を含む日本語で執筆された研究成果物を解析対象のテキストデータとして収集した (図1の①)。データ収集には国立情報学研究所が提供する学術情報データベースである CiNii の検索サービスを利用し、「東日本大震災」という検索ワードで得られた情報を記録した。検索条件から得られた著者名、論文名 (研究タイトル)、雑誌名、ISSN、出版者名、出版日付、巻、号、ページ、URL、URL

(DOI) を TSV (tab-separated values) 形式でエクスポートした後に、そこから著者名、論文名、出版日付のみを抽出し、手動で CSV (comma-separated values) 形式のファイルに変換して解析に供した (n = 26,874)。検索結果として得られた研究タイトルには原著論文だけでなく学会発表の要旨なども含まれるが、本研究ではそれら全ての種類の研究タイトルを解析対象とした。また、時系列解析を行うにあたり、2022年のデータは本研究の解析実施時点(2022年9月現在)でまだ出揃っていないため、解析対象を2011年から2021年までの11年間のデータに限定した。

抽出した研究タイトルには様々な種類の単語が含まれるが、その特徴分析に必要な単語のみを抽出するために形態素解析を行った(図1の②)。形態素解析とは、テキストデータを解析する際に対象となる文書に含まれる単語を識別する処理のことであり、そこには、単語の識別(tokenization)、活用処理(lemmatization, stemminig)、品詞の同定(part-of-speech tagging)が含まれる(松本2000)。要するに、ある文書を単語の単位で分解して、選別を可能とするために単語を品詞ごとに分類し、文書中あるいは文書間での比較を可能とするために単語の活用を統一する処理である。本研究では、RMeCabパッケージに含まれるRMeCabDF関数を用いてタイトルデータの形態素解析を行った(Ishida 2021)。形態素解析の結果、全てのタイトルデータは、記号を除いて、10種類の品詞に属する631,582個の単語に分割された。本研究では、解析対象を名詞と形容詞に制限して、それら以外の品詞(例えば、動詞や助詞など)に属する単語は分析対象から除外し、最終的に残された15,692種類(語彙数)の424,066個の単語を分析に供した。

続いて、名詞と形容詞に絞られた単語データに対してストップワードの除外を行った。ストップワードとは、文書の正しい形式をなす上で必要不可欠ではあるが、タイトルデータを解析する上で本質的には必要ない単語のことである。本稿のタイトル「テキストマイニング技術を用いた東日本大震災に関する研究動向の分析」を例に挙げると、「テキストマイニング」や「動向」などは文章の意味理解に必要ではあるが、「用いた」や「関する」などは不要と考えられる。ストップワードには、多くのテキストマイニングを用いた研究において除外されるような一般的なストップワードと、各研究で特異的に除外されるような独自のストップワードがある。一般的なストップワードは、多くの研究で用いられることからインターネット上にそのリストが掲載されている。本研究ではそのようなストップワードリストの一種である slothlib (<http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>) を用いて、一般的なストップワードを除外した。独自のストップワードとは、文書の属性(文章のテーマや対象など)によって決定される単語である。研究タイトルをテキストマイニングの対象とした場合、「研究」、「目的」、「方法」、「結果」、「考察」といった単語は高頻度で出現することが想定されるが、これらは研究タイトルの構成上必要ではあるがその本質的な内容には関連しない単語であるため除外した方が精度の高い解析が可能となると考えられる。本研究では、次節で示すように tf-idf という文書全体に高頻度で出現する単語の重要度を下げるような指標による工夫を施すため、必ずしも全ての一般語が解析の妨げになるとは限らない。そこで、独自に除外

が必要であるストップワードに関しては、次節で説明するクラスタリング結果の解釈を阻害する単語が生じた場合は除外するという試行錯誤的な方法を採用した。

2.3 テキストマイニング結果によるクラスター解析

クラスター解析を行うためにストップワードを除外した研究タイトルデータおよび各研究タイトルに属する単語のデータから、文書-単語行列 (DTM: Document Term Matrix) を作成した (図 1 の③)。DTM は、行に文書を、列に単語を配置し、要素には文書における単語の出現を表す指標を格納した行列である。DTM の要素には、各文書における単語の存在を示す二値データや出現数のカウントデータなどが用いられるが、本研究では効果的なクラスタリングを実現するために *tf-idf* という指標を用いた。*tf-idf* 指標は *tf* (term frequency) と *idf* (inverse document frequency) という 2 つの指標の積として以下の式によって算出される。

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D)$$

このとき、*tf* は個別の研究タイトル *d* における単語 *t* の出現頻度を表す。*idf* に関しては以下の式によって算出される。

$$idf(t, D) = \log\left(\frac{N}{count(d \in D : t \in d)} + 1\right)$$

tf は個別の文書においてよく出現する単語の出現頻度であり、その単語の重要性を表す指標である。*idf* は逆文書頻度とも称される指標であり、総文書数を任意の単語 *d* が含まれる全ての文書数で除したものの対数をとることで算出される。*D* は本研究で収集した単語のリストである。対数項の中で 1 を足しているのは、値が 0 である場合に対数計算が不可能となることを防ぐためのものである。*idf* はその算出式からもわかるように稀に出現する単語ほど高い値を、一般語のような頻繁に現れる単語ほど低い値を示すような指標である。つまり、*tf-idf* とは文書の特徴抽出に重要である可能性が高い単語に高い *tf* を与え、特徴抽出にあまり影響を与えないと考えられる単語に低い *tf* を与える指標である。これより、単純な出現数や二値データから作成した DTM をよりも *tf-idf* を用いた DTM を用いた方がより現実的なクラスタリングを実現することが期待できる。本研究では、前処理が完了した研究タイトルデータと対応する単語のリストから *tf-idf* による DTM を作成するために *tidytext* パッケージに準備されている *cast_dtm* 関数を使用した (Silge and Robinson 2016)。

tf-idf を用いた DTM に対して統計的クラスタリング手法を適用して研究タイトルを複数のクラスターに分類するために、本研究では *kmeans* 法によるクラスタリングを行った (図 1 の③)。*kmeans* 法とは、まずデータをランダムにクラスターに分けて、クラスター内の DTM の重心を計算し、その重心から各データへの距離が最小になるようクラスター分けを繰り返していくアル

ゴリズムであり、すべてのデータ同士の距離の計算が不要になるため、計算効率の高いクラスタリング手法である。最初に分析者が任意のクラスター数を決定して、それに基づいてクラスタリングを行う。このクラスター数を自動的に求める手法も開発されているが、本研究ではテキストデータを解析対象としており最終的なクラスターの解釈可能性が重要であるため、5~20までクラスター数を変えながらクラスタリングを実施し著者が比較検討することで、最も解釈しやすいクラスター数を決定した。その結果、クラスター数は12となった。さらに議論を進めやすくするために、各クラスターの出現頻度の高い単語を用いてクラスター名を付与した。その際、含まれる研究タイトル数が最小となったクラスターで2回以上出現した単語数をカウントしたところ、その個数は34となったため、すべてのクラスターにおいて上位34個の頻出単語を取得し、クラスター名の付与に用いた。さらに、各クラスターの研究数の時系列変化を探るために、各クラスターに含まれる研究タイトル数の推移や、その相対的な変化についても比較検討を行った。

3. 結果と考察

3.1 研究成果数の変化

本研究では、CiNii データベースで「東日本大震災」という語を研究タイトルに含むものを検索し、解析の対象とした。図2は今回収集した大震災に関する研究タイトル数の経年的な推移である。大震災が発生した2011年の研究タイトル数は11年間において最大値を示した(7,945)。震災翌年の2012年では、前年に比べて減少するものの6,386の研究タイトルが発表された。その後の研究タイトル数は急激に下降し、2013年の発表数は前年の約半数まで減った(3,087)。この減少傾向は2014年以降も続き、2020年に最小値(478)を示した。しかし、震災発生から10年の節目となる2021年には増加に転じ、年間での研究成果数が2014年に次ぐ過去5番目となった(1,645)。

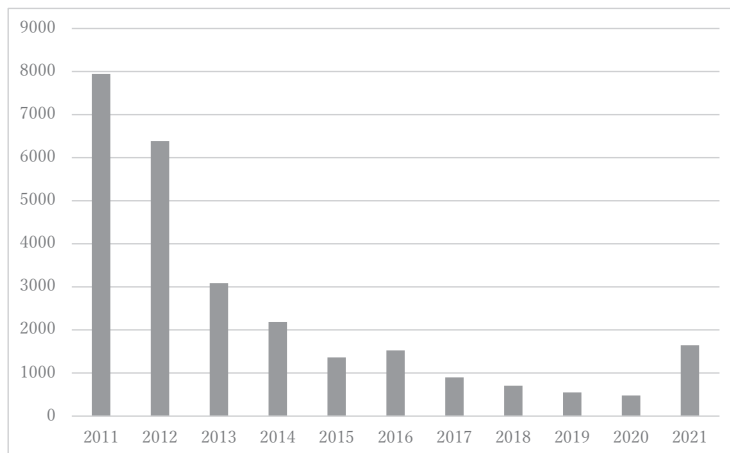


図2 大震災に関する研究タイトル数の経年変化

これは大震災発生直後の2011年や2012年は研究関心が高く、多くの研究者や研究機関が調査結果を迅速に研究成果として公表したものの、それ以降は研究量が減少し大震災研究への関心が時間の経過とともに薄れていったことを示唆している。柴田ら(2019)が指摘するように、被災者や被災者への支援は継続的であるべきで、それらについての研究もやはり継続的に実施されるべきだが、傾向としては先細りの状態であった。

ただし、大震災から10年目にあたる2021年の研究タイトル数は、大震災発生直後(2011-2014年)には及ばないものの大きく増加し、節目に当たる年に大震災への関心が再び高まった可能性を示唆している。2021年には複数の学会や研究会で研究集会(日本地理学会2021年春季学術大会シンポジウムなど)が実施されたことも研究タイトル数の増加に寄与したと考えられる。

3.2 研究クラスターの特徴

大震災に関する研究は多様な研究領域に広がっている。この広がりを客観的に分析し、主な研究領域とその動向を理解するためにクラスター解析を実施した。最も解釈可能性が高まるよう研究クラスターの数を12と設定した。生成された各研究クラスターの性質の直感的な理解を可能にするために、各クラスターにふくまれる研究タイトル内での出現頻度が高かった単語の上位34個を取得し、そこから数字・記号や、「大震災」「東日本」「研究」など本研究で扱う研究タイトルに普遍的に現れる単語を除去し、研究トピックの特徴づけを実施した(表1)。それぞれのクラスターに含まれる頻出単語をもとに筆者がクラスター名を付与した。本研究ではこれらを便宜的に、各研究クラスターの特徴を反映する呼称として扱う。

クラスター1には、被災、復興、対応、被害といった大震災研究に共通する単語が多く含まれ、後述のクラスター2から12のような特徴的な単語が現れていないため「一般研究」とした。実質的にクラスター1は、他の特徴的なクラスターに分類できなかった研究タイトルを包含するグループと考えることができる。クラスター2は、観光、復興、被災、ツーリズムといった単語を含んでいたことから「観光」とした。クラスター3は、支援、活動、医療、保健といった単語を含んでいたことから「医療支援」とした。クラスター4は、廃棄、処理、課題、放射といった単語を含んでいたことから「原発事故」とした。クラスター5は、住宅、被災、支援、活動といった単語を含んでいたことから「被災者生活」とした。クラスター6は、協同、組合、ビジョン、座談といった単語を含んでいたことから「協同組合」とした。クラスター7は、体験、外傷、語り、心といった単語を含んでいたことから「被災者心理」とした。クラスター8は、被災、住宅、調査、地域といった単語を含んでいたことから「地域調査」とした。クラスター9は、建築、住宅、計画、デザインといった単語を含んでいたことから「建築」とした。クラスター10は、復興、復旧、まち、支援といった単語を含んでいたことから「まちづくり」とした。クラスター11は、法律、登記、不動産、法務省といった単語を含んでいたことから「不動産」とした。クラスター12は、教育、学校、防災、子どもといった単語を含んでいたことから「教育」とした。

各クラスターの研究タイトル数に着目すると、クラスター1「一般研究」が最も多く（14,885）、総数の半数以上がクラスター1「一般研究」に分類されていた。つづいて、クラスター10「まちづくり」（5,085）、クラスター3「医療支援」（3,071）、クラスター8「地域研究」（2,288）であった。このように2,000以上の研究タイトルを含むクラスターがある一方で、研究タイトル数が100以下のクラスターは、クラスター2「観光」（81）、クラスター11「不動産」（70）、クラスター6「協同組合」（45）、クラスター5「被災者生活」（25）、クラスター7「被災者心理」（4）である。このように、クラスター分析から、関係する研究領域と、その研究数の多寡を類推することができた。

個別のクラスターについての検討を行う。クラスター10「まちづくり」に着目すると、復興、原発、事故、福島といった単語が上位に頻出したことから、被災地のなかでも原発事故における福島を中心とした街づくりや地域支援に関する研究報告、復旧や支援の状況、今後の対策などの研究成果が集められたと推測できる。クラスター8「地域研究」には、宮城、岩手、福島という地名が頻出し、被災地3県における状況調査や支援活動、仮設住宅などの報告が集められたといえる。また、研究成果数が100以下のクラスターでは、被災者生活や被災者心理といった被災者への支援や被災者が体験した困難などを中心に据えたクラスターに分類され、研究数としては少ないものの、被災者の心や体験に焦点を当てた研究者も実施されていることがわかる。

以上から、大震災研究として、「まちづくり」「地域研究」といった総合的な視点で課題を検討する研究クラスターの存在が明らかになるとともに、「医療支援」「建築」「教育」といった専門性の高いクラスターも存在していたことがわかった。研究タイトル数が多いクラスターがある一方で、数としては少ないながらも、協同組合や不動産といった被災者の日常生活に直結した課題に関する研究クラスターや、被災者生活や被災者心理といった被災者のケアを対象としたクラスターも出現した。以上から、大震災に関する研究は幅広いテーマを対象に行われ、全体的な問題関心に着目したものから個別具体的な事例に関するものまで実施されてきたと考えられる。

本研究はデータ収集の規模を、多様な研究領域を包含する大震災研究全般まで拡張したことが特徴である。原発事故から医療、建築、被災者ケアに至るまでの多様な研究領域が存在していたが、それらを定量的に評価することで、首尾よく解釈可能性の高い情報集約を達成することができた。

3.3 研究動向の変化

つづいて、上述の12のクラスターが、2011年（大震災発生）から2021年までの11年間に於いてどのような時系列変化を示したのかを考える。形成された研究クラスターごとの研究タイトル数の時系列的なトレンドを分析するために可視化を行った（図3）。クラスター1「一般研究」、クラスター3「医療支援」、クラスター8「地域研究」、クラスター10「まちづくり」は比較的研究タイトル数の大きなクラスターである。やや減少傾向はあるものの、11年間に於いて継続的に研究が報告され続けたトピックであるといえる。これらと比較して研究タイトル数はやや少ない

表1 各クラスターにおける出現頻度が高かった単語（記号・数字・一般的な単語は除外）

出現順位	1	2	3	4	5	6	7	8	9	10	11	12
1	被災	観光	医療支援	原発事故	被災者生活	協同組合	被災者心理	地域研究	建築	まちづくり	不動産	教育
2	復興	観光	支援	廃棄	住宅	協同	成長	県	会	復興	法律	教育
3	対応	復興	活動	処理	市	組合	法	市	建築	原発	登記	学校
4	被害	被災	被災	発生	県	復興	混合	復興	住宅	事故	不動産	防災
5	調査	地域	医療	等	被災	原発	体験	被災	被害	福島	関係	社会
6	津波	ASEAN	報告	復興	宮城	事故	分析	宮城	計画	被災	号	科
7	復旧	性	復興	課題	時	報告	学	岩手	アザイン	課題	法務省	課題
8	支援	取り組み	シンポジウム	性	対応	ビジョン	Helper	住宅	大会	復旧	免許	科学
9	状況	可能	学会	県	復興	TPP	外傷	福島	調査	地域	民	復興
10	地震	ツォーリズム	子ども	放射	仮設	座談	援助	町	発表	づくり	登録	シンポジウム
11	課題	状況	対応	環境	医療	緊急	被災	仮設	学術	現状	税	地域
12	地域	今	対策	復旧	事例	消防	テキスト	支援	講演	支援	事務	実践
13	対策	市場	保健	解説	影響	大会	マイニング	調査	東海	東海	特例	報告
14	影響	今後	状況	状況	役割	提言	語り	地域	被災	被災	通達	被災
15	報告	想い	物質	市	支援	ネットワーク	Experience	避難	学会	被災	被災	家庭
16	教訓	東北	処分	物質	活動	運動	PTG	課題	構造	回答	回答	学会
17		宮城	現状	現状	活動	シンポジウム	Research	活動	仮設	長	長	
18						社会	ナラティブ	事例	避難	平成	平成	
19						課題	プロジェクト	報告	都市	国税	国税	
20						人間	人間	津波	環境	臨時	臨時	
21						心	心	状況		取扱い	取扱い	
22						心的	心的			規定	規定	
23						構築	構築			民事	民事	
24						焦点	焦点			法	法	
25						質	質					

が、クラスター12「教育」でも、継続的に報告がつづいている。一方で、クラスター4「原発事故」、クラスター9「建築」は、2011年から2015年までの期間は安定した研究タイトル数を保持していたが、2016年以降大きく減少していることがわかる。

つづいて各クラスターの経年変化を明らかにするため、研究タイトル数の相対的な変化を可視化した(図4)。クラスターごとに、全期間(2011-2021年)の研究タイトル数を1とした場合の各年の相対値を表示している。12クラスターの11年間の変動を読み取ると、2011年に最大値をとり、時間の経過とともに減少を続け、2021年に微増するグループの存在がわかる。これには、クラスター1「一般研究」やクラスター6「協同組合」である。これらのクラスターには、比較的研究結果を公表しやすいものや迅速な公表が求められるものが含まれていた可能性がある。たとえば、クラスター1「一般研究」では、2011年に全期間の36%もの研究タイトルが発表されている。

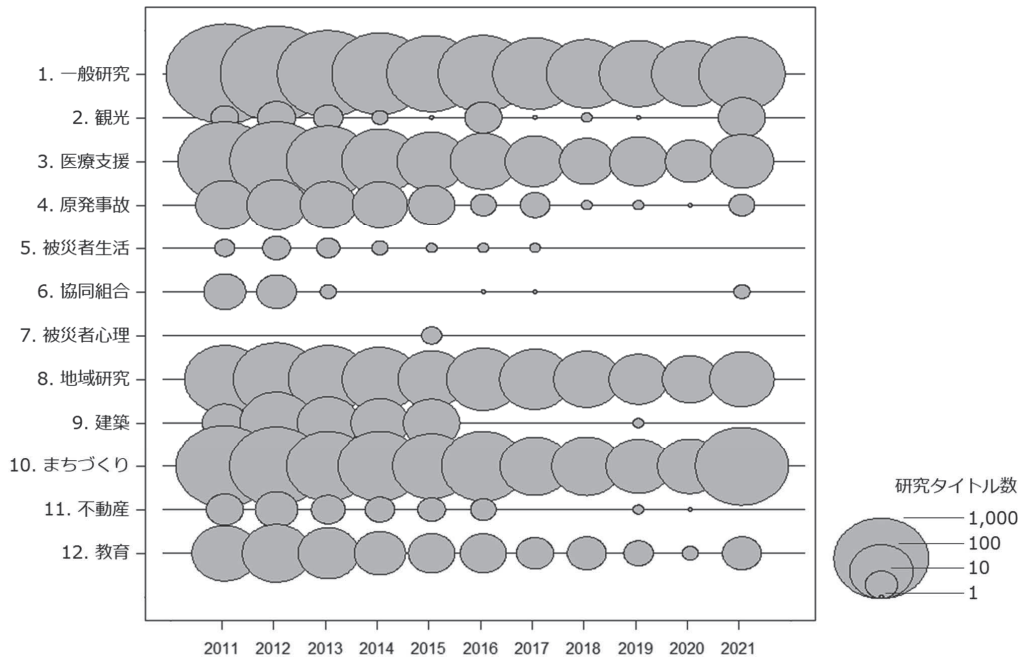


図3 クラスターごとの研究タイトル数(絶対値)の経年変化

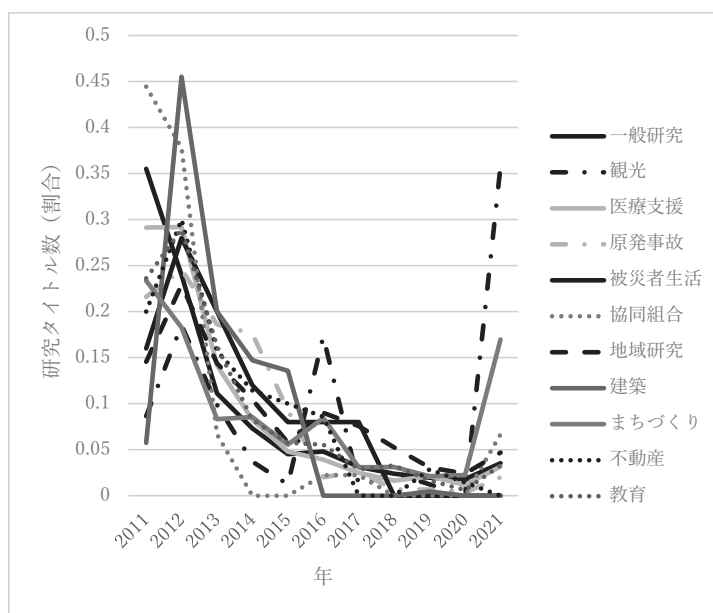


図4 クラスターごとの研究タイトル数(相対値)の経年変化
全期間(2011-2021年)の研究タイトル数を1とした場合の各年の相対値を表示

次に、2012年に最大値をとり、その後は減少するものの、2021年には微増するグループが存在していた。これには、クラスター5「被災者生活」、クラスター8「地域研究」、クラスター9「建築」、クラスター11「不動産」、クラスター12「教育」が含まれる。また、クラスター3「医療支援」は、2011年と2012年という震災直後の2年間は継続的に発表されており、医療支援や現場での医療報告は最初の2年に集中したといえる。これらは、調査研究や公表までにある程度の期間を要したため、ピークが2012年に訪れたことが示唆される。

注目すべきこととして、クラスター2「観光」は、大震災発生直後の3年間(2011-2013年)には比較的研究タイトル数が少なかったが、2021年の研究タイトル数は、2011-2021年の全期間を通じて最大数の研究タイトルを含んでいた。これは、震災直後はインフラの復旧や整備などに追われていて観光についての研究を実施する余裕がなかったが、震災から10年を経て、この節目の年に被災地の観光を本格的に検討することが行われ始めたことを示唆している。また、クラスター10「まちづくり」においても、2021年に著しく研究タイトル数が増加した。これは、節目の年にこれまでの復興都市計画などが振り返られ、研究として取りまとめられたことを示唆している。その一方で、クラスター9「建築」やクラスター11「不動産」においては、2021年における有意な研究タイトル数の増加は見られなかった。これらの研究分野においては、大震災発生後の数年間に分析と研究発表が実施されその後議論が下火となり、研究発表という形には現れなくなったことを示唆している。このように、研究クラスターごとの時系列変化には特徴があることが判明した。近年下火になっている研究分野だからといって、大震災研究での重要性が低いと

は限らない。この機会に、大震災関連研究を俯瞰し、研究報告数が低下している分野についての検討も行われるべきだろう。

4. おわりに

上述の結果と考察をふまえて本研究の結論を述べる。大震災に関連する大量かつ多様な研究タイトルをテキストマイニングによって分析しクラスター解析を行ったことで、研究を12のクラスターに分類し、それぞれの特徴を検討することに成功した。クラスターごとの研究タイトルの時間変化を示すことで、学術的関心の推移についての示唆を得ることができた。全体として、大震災に関連する研究タイトルの減少を提示することで、学術的関心の低下に警鐘を鳴らすことができた。12の研究クラスターに分類した結果、まちづくりや地域研究といった公的・俯瞰的な視点から大震災を捉える研究と、被災者心理のように被災者個人に寄り添った視点からの研究が存在することが示唆された。また、まちづくりや観光といったトピックは、震災によって被害の大きかった被災地が復興する過程において、現在注目されているテーマといえる。

研究数で比較した場合、公的な事業に関する研究タイトルが圧倒的多数を占めており、被災者生活や被災者心理を取り扱った研究タイトルは少数であった。ただし、研究タイトル数が少ないことは必ずしも研究関心が低いことを表してはいない可能性に着目すべきである。というのは、クラスター5「被災者生活」やクラスター7「被災者心理」など、より個人的な感情や生活に直結するテーマは、研究結果を提示するまでに時間がかかることが想定できる。とりわけ、被災者のプライベートな心理や体験に踏み込む必要もあり、研究発表や成果というわかりやすい形で公表することに、心理的や物理的なハードルが高かったともいえる。筆者も大震災の広域避難者への聞き取り調査を5年ほど実施しているが、プライベートな情報を多く含むその結果を研究発表することへの葛藤を抱えている。

本稿では、大震災に関する膨大な研究を客観的に分析することで、学術領域における大震災への研究関心の濃淡や時系列変化を示すことができた。大震災全体を扱う研究領域全体を俯瞰することで研究関心の低いテーマを浮き彫りにしたことで、研究エフォートの配分についてあらためて学界が一考する価値を示したといえる。

本研究で用いた分析手法では大震災でどのような研究が実施されてきたのかをクラスターごとのトレンドとして示すことはできたが、各研究における成果や課題まで踏み込んで分析することはできなかった。これは残された課題であり、今後は個々の成果をより細かく捉えることで、研究の意義や、今後求められる研究領域や研究テーマを具体的に論じることが可能になるであろう。また、本稿では日本語で発表された研究を対象にテキストマイニングを実施したが、英語論文を対象外としたことは課題として挙げられる。今後の展望としては、英語をはじめとする日本語以外で公表された研究を対象に、大震災に関する調査が海外向けにどのように発信されたのかという視点から分析することで、日本語で公表された国内向けの研究成果と英語をはじめとする日本

語を除く言語で発表された海外向け研究成果を比較検討することができるであろう。

大震災は過去に類例のない複合的な大災害であったが、時間の経過とともに研究領域における関心は減少しており、風化していくことへの懸念がある。大震災の経験や災害を記憶に留めるためにも、事象を長期的な目線で捉える研究に関心を寄せ続けることが重要であろう。また、研究関心の低いテーマにも光をあてるための努力が求められているだろう。大震災が発生してから2021年末までに26,874の研究報告がなされていたことは、多くの研究者が大震災に対し、何らかの研究目的をもって調査し、その結果を報告した証である。その研究報告数の背景には、悲しみの土台があることを忘れてはならない。

謝辞

東日本大震災におきまして亡くなられた方々に対し、心から哀悼の意を捧げますとともに被災された方々に対し、心からお見舞い申し上げます。本研究は、JSPS 科研費 19J22325 の助成を受けたものです。ここに記して御礼申し上げます。本研究の実施にあたり、研究の全体像の立案、データ取得、統計解析などで貴重な知見をご提供いただいた、京都大学フィールド科学研究センターの伊勢武史准教授および総合地球環境学研究所の友尻大幹研究員に大いなる感謝を申し上げます。

参考文献

- Bache SM, Wickham H. (2020). magrittr: A Forward-Pipe Operator for R. <https://CRAN.R-project.org/package=magrittr>.
- Grolemund G, Wickham H. (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software* 40, 1–25.
- Ishida M. (2021). RMeCab: interface to MeCab. R package version 1.07. <https://github.com/IshidaMotohiro/RMeCab>.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Silge J, Robinson D. (2016) tidytext: text mining and analysis using tidy data principles in R. *Journal of Open Source Software* 1, 37.
- Wickham H. et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software* 4, 1686.
- 朝日新聞 (2012) 「総選挙 復興の施策見極める」2012年11月11日.
- 乾康代 (2016) 「避難者受け入れ自治体と被災者自治体による県外避難者支援」日本建築学会計画論文集 81–726, 1851–1858.
- 今井照 (2014) 『自治体再建 — 原発避難と「移動する村」』筑摩書房.
- 開沼博 (2011) 『「フクシマ論」』青土社.
- 河村和徳 (2013) 「被災地における住民参加：被災地の調査結果から」横浜市立大学論叢. 社会科学系列

64, 59-81.

柴田理瑛・足立智昭・平野幹雄他 (2019)「東日本大震災の長期的影響と今求められる支援者支援：一般社団法人東日本大震災子ども・若者支援センター2018年度活動報告」宮城学院女子大学発達科学研究 19, 8-16.

菅野拓 (2015)「社会問題への対応からみるサードセクターの携帯と地域的展開——東日本大震災後の復興支援を事例として——」人文地理 67 (5), 1-394.

鈴木康弘 (2021)「東日本大震災 10 年——防災概念の変革期における地理学——」E-journal GEO 16, 187-190.

関礼子 (2018)『被災と避難の社会学』東信堂.

田中正人 (2016)「原発被災地における居住者の避難プロセスと帰還／移住選択困難性の背景——福島県川内村・貝の坂地区事例——」地域安全学会論文集, 29, 155-164.

戸田典樹 (2016)『福島原発事故 漂流する自主避難者たち 実態調査からみた課題と社会的支援のあり方』赤石書店.

新美綾子・山本克彦・佐藤大介・横山由香里・上山崎悦代・野尻紀恵・原田正樹 (2018)「東日本大震災を経験した福祉施設職員の震災前から現在までの体験——テキストマイニングによる分析から——」日本福祉大学全学教育センター紀要 6, 47-58.

西城戸誠・原田峻 (2014)「埼玉県における県外避難者とその支援の現状と課題」人間環境論集, 15, 169-103.

原田峻・西城戸誠 (2015)「東日本大震災・福島原発事故から5年目を迎えた県外避難の現状と課題——埼玉県における自治体・避難者調査の知見から——」立教大学コミュニティ福祉研究所紀要第3号, 59-78.

日野行介 (2014)『福島原発事故 被災者支援政策の欺瞞』岩波新書.

藤森崇浩・小山真紀・清野純史 (2014)「災害報道からの個人属性別の状況の抽出～東日本大震災後の全国紙と地方紙のテキストマイニングから～」地域安全学会論文集 23, 55-64.

松井克浩 (2017)『故郷喪失と再生への時間 新潟県への原発避難と支援の社会学』東信堂.

松永妃都美 (2018)「福島第一原子力発電所の事故を契機に自主避難をした母親が避難先地域との関わりの中で母子避難を継続していくプロセス」日本看護学会誌 38, 107-114.

松本裕治 (2000)「形態素解析システム「茶釜」」情報処理 41, 1208-1214.

水野勲 (2017)「地理的カストロフとしての原発の過酷事故——エクメーネの再概念化——」理論地理学ノート, 19, 71-89.

三宅理一・池田綾香・松下瑛美 (2013)「北海道ならびに石狩市における東日本大震災避者の居住施設に関する研究——住宅施策との関連から見る被災者受け入れ体制とその実態——」人間生活学研究 20, 1-12.

山根純佳 (2013)「原発事故による「母子避難」問題とその支援——山形県における避難者調査のデータから——」山形大学人文学部研究年報 10, 37-51.

吉見憲二 (2015)「東日本大震災後のテレワークに関連した報道内容の減少に関する研究」日本テレワーク学会誌 13, 39-46.

URL

早稲田大学総合人文科学研究センター年次フォーラム「東日本大震災 10 年の軌跡と大規模災害からの復興をめぐる～新たな「日常」への模索～」 <https://www.waseda.jp/flas/rilas/news/2020/12/18/7627/>