

計量テキスト分析による災害対応課題の構造化に関する研究

自治体における災害対応検証を事例として

A Study on the Structure of Disaster response issues using the Text Analysis

A Case Study of the Disaster response verification

○坪井 壺太郎^{*1}
Sotaro Tsuboi^{*1}

*1 公益財団法人ひょうご震災記念 21 世紀研究機構 人と防災未来センター リサーチフェロー 博士 (都市科学)
Research Fellow, Disaster Reduction and Human Renovation Institution Ph.D. (Urban Science)

キーワード：計量テキスト分析；自由回答記述；形態素；災害対応

Keywords: Quantitative Text analysis; structure; Free answer; Morphology; Disaster response

1. はじめに

近年、特に企業ソリューションの現場において、アンケートの自由回答やコールセンター、ウェブサイトなどに寄せられるユーザーからの意見や書き込み等の定性的な文字情報を解析し、自社製品へのニーズや不満点等を分析する「テキストマイニング」手法が注目されつつある¹⁾。テキストマイニングとは、定型化されていない文章の集まりを自然言語解析の手法を援用して単語やフレーズに分割し、それらの出現頻度や相関関係を分析して情報を抽出するテキストデータ分析手法やシステムの総称である。

テキストマイニングでは、定型化されていないデータから知識・情報を見つけ出すことに特徴をもち、その利点は、その発話者・発信者・書き手のもつ思考世界への接近が期待できることにある。一般に、質問紙（アンケート）を用いた調査の多くは、あらかじめ設定された尺度（選択肢回答法）への回答をもとにデータの解析が行われる場合が多く、これまで「自由回答」は、そのデータ解釈を補完するものとして位置付けられてきた。そのため、自由回答が分析対象そのものになることは少なかったことが指摘できる。

一方、ワークショップ等において広く実践されてきている「KJ 法」²⁾は、あるテーマに対する意見等を付箋等へ書き出し、類似した内容をグループ化することにより、最終的に複数の意見を構造化、明確化する手法である。KJ 法における作業の重要な点は「直感」であり、配置やグループ化に際してはあらかじめ仮定した理論に基づくものではなく、情報に対する直感的な判断が重視される。そのため、調査担当者の経験や分類された項目に対する理解度などが影響する限界点も一定程度存在するほか、集約された意見の方向性や階層性を示すことは可能ではあっても、構成された集塊の「意見」同士がもつ意味の構造を論理的に解釈することは必ずしも容易で

はない。

本研究では、上述の課題に着目し、得られた自由回答記述内容（テキスト）を可能な限り客観的に解析するために、日本語を形態素（意味を持つ最小の言語単位）に分類した上で、「自由回答」のもつ意味の構成や構造を明らかにする。日本語を形態素に分類するための手法は、単語ごとに「分かち書き」をする英語に比べて、日本語文法が有する特殊性から、文章を単語ごとに切り分けるための処理技術（形態素解析）の遅れが指摘されてきた。しかし、近年では言語解析プログラム技術の進化に伴い、行政・政策分野での事例研究³⁾等も徐々に蓄積されてきており、本研究には一定の意義があるものと考えられる。

本研究で分析対象とする行政における「災害対応検証」記録については、被災後において庁内の各部局職員の見解等が集約され、その後の災害対応マニュアルの改訂等に反映する取り組みが行われている。しかし、記述内容の集約においては、必ずしも法則性が確立されておらず、客観的な分析視座確保の観点から、構成される意味空間の分析方法を検討していく必要であると考えられる。

そこで、本研究では、調査における自由回答の記述内容について客観的かつ計量的に分析を試みることで、論拠に基づく政策立案（EBPM）に資する方法論を検討し、行政における活用手法を提案することを目的とする。

本研究は、調査対象者の「自由回答記述（テキスト）」を主体的に扱うことを主眼に置き、定型化されていない文章の集まりを自然言語解析の手法を用いて単語やフレーズの形態素に分解したうえで、集合的意味を考究（マイニング）し、あわせて、分析にあたり統計学的解析を援用した「計量テキスト分析」を行う。本研究による論及方法の特徴は、記述された用語の傾向から課題への接近を図る「質的性格」と、文法と出現法則に基づいた用語間の集塊性から課題への接近を図る「量的性格」を併せ持つことにある。

2. 研究対象と方法

2.1. 令和元年台風第 19 号における災害対応検証

本研究では、2019年10月12日～13日にかけて東日本地域に甚大な被害をもたらした「令和元年台風第19号」を対象とし、同災害に対する長野県A市における行政職員対象の「災害対応・避難行動等に関する対応記録」をもとに分析を行う。A市においては人命にかかわる被害はなかったものの、ダムの緊急放流に伴う避難所の拡大や、複数回の緊急速報メールの発信のほか、住民の避難対応においては、市内8か所の避難所を開設し668人(250世帯)の避難者の受け入れが行われた。A市では、同災害直後の10月15日から23日にかけて庁内職員を対象とした災害対応課題に関する自由記述方式の質問紙調査が実施された。本研究では同調査によって記述された内容をもとに分析を行う。

2.2. 研究方法

分析にあたっては、対応記録においてあらかじめ項目として集約された「情報」「本部」「避難所」に関する記述内容を対象とする(Table1)。同項目は、災害の「初動期」から「応急期」にかけて発生する共通の課題であり、本分析を通して、災害対応課題の検討を試みる。

形態素を対象とする形態論研究(morphology)は、言語学分野における研究対象のひとつであり、単語や品詞といった言語が意味を持つ最小の単位の語形変化や並び方などの考察、理論から発達してきたものである。この理論枠組みを引き継ぎ、日本語の品詞間の接合規則性を分類するための解析プログラム「Juman」、「Chasen」、「KAKASI」、「すもも」、「Breakfast」、「Sudachi」等が開発されている。本研究では、基礎的な記述内容の構造解析に重点を置くと同時に、品詞の細分類化に利点を持つ「Mecab」によるプログラムを用いて解析を行った。日本語形態素は、品詞とその前後の接続関係をもとに分類され、分析においてどの品詞を採用するかについては分析者の判断に依拠しており、全用語を含む形態素を投入したうえで探索的に品詞を決定する方法や、あらかじめ投入品詞を設定する方法がある。

本研究では後者の方法を採用し、先行研究⁴⁾での知見をふまえ、名詞と形容詞を中心に、行動や意思や要望を指す「-する」、「-したい」等のサ行変格活用動詞に結節する「名詞-サ変接続」のほか、「名詞-形容動詞語幹」、「名詞-一般」、「名詞-複合名詞」および、「形容詞-自立」「形容詞-接尾」「形容詞-非自立」の7品詞を分析に採用した(Table2)。解析の結果、2,060用語が抽出された。分析にあたっては、あらかじめ表記のゆれの確認を行ったうえで、出現頻度により意味空間が決定されるという形態素分析の前提に基づき、3項目の合計出現回数10回以上の33用語を抽出した(Table3)。

Table1. Number of descriptions and total words

項目	記述概要	件数	単語総数
情報	住民への情報伝達・庁内の情報共有等	84	891
本部	本部運営・指揮調整・庁内調整等	159	1,810
避難所	避難所開設・避難所運営・福祉避難所	78	798

Table2.Types of morpheme and Acceptance / rejection

品詞	品詞細分類	採否	品詞	品詞細分類	採否
連体詞	—	△			
接頭詞	形容詞接続	×	動詞	自立	△
	数接続	×		接尾	×
	動詞接続	×	非自立	×	
	名詞接続	×	特殊	×	
名詞	サ変接続	◎	形容詞	自立	◎
	ナイ形容詞語幹	△		接尾	◎
	形容動詞語幹	◎		非自立	◎
	動詞非自立的	×	副詞	一般	△
	副詞可能	×		助詞接続類	△
	一般	◎	助詞	格助詞	×
	数	×		係助詞	×
	接続詞的	×		終助詞	×
	固有名詞	△		接続助詞	×
	接尾	×		副詞化	×
	代名詞	×		副助詞	×
	非自立	×		並立助詞	×
	特殊	×	助動詞	—	
	複合名詞	◎	感動詞	—	
未知語	—	×	記号	—	
					×

注：◎=採用，△=検討（本研究では不採用），×=不採用

Table3. Number of terms by morpheme

語	品詞	品詞細分類	情報	本部	避難所	合計
必要	名詞	形容動詞語幹	22	35	19	76
職員	名詞	一般	16	25	8	49
避難 所	複合名詞	複合名詞	4	19	25	48
対応	名詞	サ変接続	11	24	12	47
ない	形容詞	自立	7	28	11	46
避難	名詞	サ変接続	6	17	6	29
検討	名詞	サ変接続	6	10	8	24
情報	名詞	一般	7	11	3	21
良い	形容詞	自立	7	9	5	21
状況	名詞	一般	3	13	4	20
避難 者	複合名詞	複合名詞	1	10	8	19
場合	名詞	副詞可能	4	10	5	19
指示	名詞	サ変接続	7	7	5	19
多い	形容詞	自立	3	11	2	16
住民	名詞	一般	5	6	4	15
今回	名詞	副詞可能	3	7	4	14
情報 連絡 室	複合名詞	複合名詞	13	0	0	13
使用	名詞	サ変接続	0	5	8	13
想定	名詞	サ変接続	3	9	1	13
体育館	名詞	一般	0	8	5	13
場所	名詞	一般	3	8	2	13
参集 レベル	複合名詞	複合名詞	5	4	3	12
周知	名詞	サ変接続	3	6	3	12
設置	名詞	サ変接続	6	4	2	12
体制	名詞	一般	4	8	0	12
避難 指示	複合名詞	複合名詞	2	6	3	11
浸水	名詞	サ変接続	3	7	1	11
確認	名詞	サ変接続	3	4	4	11
本部	名詞	一般	0	9	1	10
危険	名詞	形容動詞語幹	3	3	4	10
事前	名詞	一般	4	3	3	10
業務	名詞	一般	2	5	3	10
帰宅	名詞	サ変接続	0	3	7	10

3. 分析結果

3.1. 出現用語の意味の空間構造

既往の自然言語処理・解析研究においては、文章中に同時に出現する用語を、ベクトル空間において依存関係を明示的にモデル化する LSI 分析が知られている。本研究では、この理論枠組みのもと、用語の出現の有無により行列を作成し、数量化理論Ⅲ類による解析を行った。本手法は、外的基準を持たないカテゴリ変数行列における変数相互間の関連から、データ間の類似性を明示することができ、固有値と相関比より適合基準を判断することが可能である。分析で得られる固有値は、個体の数量とカテゴリ数量との相関の二乗あるいは個体間のカテゴリ数量の相関比を表わしているため、0 から 1 の間をとる固有値の大きさにより、各次元でどの程度個体とカテゴリの関係が説明されているかを評価することが可能である。本分析の結果、第 1 軸の固有値 (00.600)・相関 (0.7746)・寄与率 3.34%および第 2 軸の固有値 (0.5807)・相関比 (0.7620)・寄与率 6.57%が示され、高いモデル適合度が得られた (Figure1)。

本分析の結果、相関適合度の高い 2 軸を抽出し、これを採用した (Table4・Table5)。各々の軸の解釈にあたっては、採用した単語の出現頻度から算出したカテゴリスコアの数値を対象としてクラスター分析を補完的にを行い用語の集塊性を考慮して実施した。

第 1 軸では、正 (+) 側に「備品」、「体育館」「確保」といった実際の避難所開設に係る用語が出現しており、それらを含む原文では、「必要な備品等はあらかじめ集約しておきたい」や「避難所備品に有効期限を経過したものが更新等が必要である」などの意見が挙げられている。

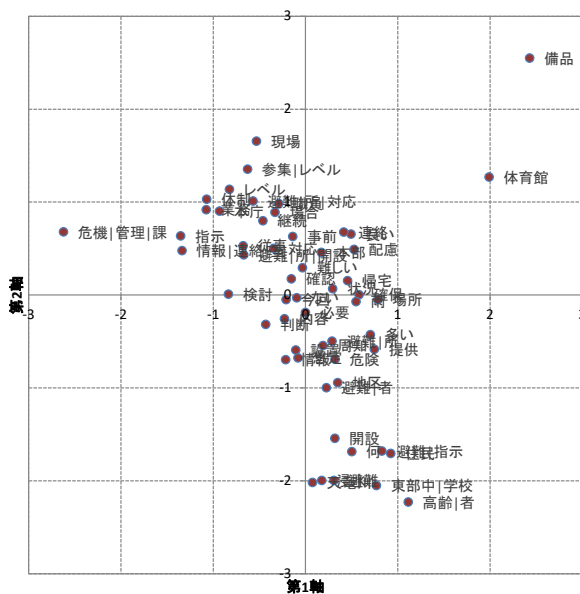


Figure1. Placement of Category score

一方、負 (-) 側には、警戒や参集に関する「レベル」の用語が出現しているほか、「対応」(原文：本部の対応が伝わりにくかった)や、「判断」(原文：退避に係る施設ごとの判断が困難である)など職員の災害対応における課題が提示されている。このことから、第 1 軸を「庁外課題-庁内課題」軸であると解釈する。

第 2 軸では、正 (+) 側に「避難所対応」(原文：避難所対応が長期化した場合の体制構築が必要)等の課題が提示されたほか、負 (-) 側に、「避難者」への配慮や「危険」に関する用語がみられた。このことから、第 2 軸を「避難所課題-避難者課題」と解釈する。以上より A 市における災害対応に関して職員から挙げられた課題や意見は、両軸の持つ意味空間で構成されることが示された。

Table 4. The First axis category score

用語	SCORE (+)	用語	SCORE (-)
学校	3.4843	内容	-0.2275
使用	3.0565	職員	-0.2872
備品	2.4313	場合	-0.3298
体育館	1.9904	対応	-0.3445
高齢 者	1.1143	判断	-0.4303
住民	0.9253	継続	-0.4612
避難 指示	0.8289	現場	-0.5304
場所	0.7860	避難 所 対応	-0.5675
東部中 学校	0.7687	参集 レベル	-0.6270
提供	0.7470	避難 所 開設	-0.6674
多い	0.7049	従事	-0.6751
確保	0.5817	レベル	-0.8234
雨	0.5492	検討	-0.8337
配慮	0.5267	本庁	-0.9300
何	0.5034	体制	-1.0693

Table5. The Second axis category score

用語	SCORE (+)	用語	SCORE (-)
学校	3.4171	判断	-0.3177
備品	2.5489	多い	-0.4291
使用	2.1805	避難 所	-0.4983
現場	1.6550	周知	-0.5450
参集 レベル	1.3543	提供	-0.5855
体育館	1.2683	設置	-0.5929
レベル	1.1363	想定	-0.6770
体制	1.0294	危険	-0.6918
避難 所 対応	1.0114	情報	-0.6977
職員	0.9827	地区	-0.9456
業務	0.9161	避難 者	-0.9980
本庁	0.9019	開設	-1.5428
場合	0.8884	避難 指示	-1.6804
継続	0.7978	何	-1.6867
危機 管理 課	0.6775	住民	-1.7067

3.2 対応分析による属性と出現用語の可視化

本研究で取りあげた「災害対応・避難行動等に関する対応記録」における意見を、あらかじめ分類された『本部』、『情報』、『避難所』の3カテゴリに基づき、出現用語との近接性をもとにその特性を検討する。分析にあたっては、出現回数8回以上の55用語のクロス集計表を用い、行要素と列要素の相関関係が最大になるよう数量化を行った。本分析の特徴は、項目間の類似性と同時に、カテゴリの近くに布置されることで、集塊性を二次元グラフで可視化できるほか、偏りの小さい項目（全体において共通している項目）は原点付近に、偏りの大きい項目（特徴的な項目）は原点から遠くに布置される特徴を持つ。本分析の結果、『避難所』には、「開設」に係る用語のほか、避難所運営において長時間の「継続」従事などが課題として挙げられたほか、『情報』には、避難所での「住民」の名簿情報提供の基準の明確化や、「停電」による情報入手手段の断絶対応などが課題として挙げられたほか、庁内課題として、「情報連絡室と危機管理課における情報共有ができなかった」等が挙げられた。

また、『本部』に近接する「想定」の用語は、A市役所の立地が、浸水想定域内であることから、本部移転による対応の必要性が課題として挙げられたほか、本部が設定している「体制」のレベルと、避難所等の現場での対応の齟齬（原文：「本部から支所に現地確認の要請があっても、体制確保できず行動が制約される」）などが課題として挙げられた。このように、対応分析では、各属性と出現用語の近接性を明示することができることから、自由回答記述内容を、属性で分析を行うことが可能であり、各意見を構造的に可視化できるため、課題を共有したうえでの議論展開に向けた有用性のある技術であると考えられる。

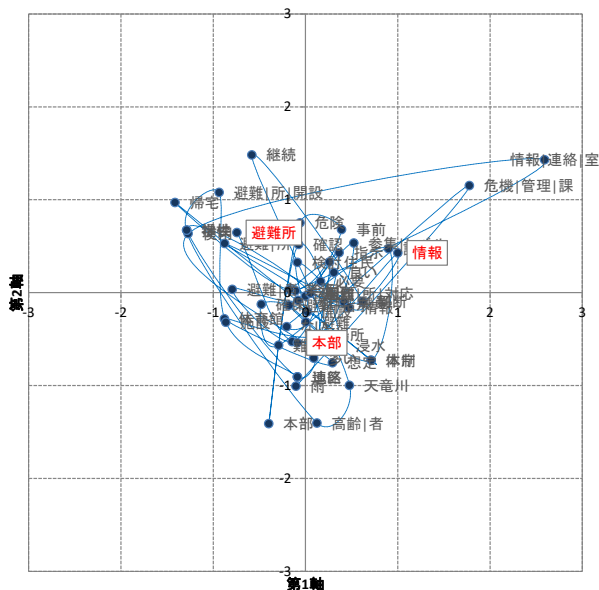


Figure2 Correspondence analysis

3.3 係り受け分析による出現用語の検討

形態素分析では、文章の分かち書きから「単語」を対象として検討が行われるが、語と語の修辞関係や主従関係を明示することができない。例えば「必要」の用語に係る単語が「ない」と「ある」では意味が異なることがあげられる。そこで、本節では、文法規則によって、文の構造を「句」を単位として検討する係り受け分析を併せて実施した。句とは2つ以上の語が集まって1つの品詞と同様の働きをしながら文を構成する語の集塊を指し、名詞句・形容詞句・副詞句・動詞句がある。日本語は語順が比較的自由的な言語であるが、文節内の構造および語順には強い制限を有している。そのため、日本語は文節の連続として解析するためのさまざまな解析手法が開発されてきており、本研究ではこのなかで代表的なシステム Cabocha を用いて解析を行った。分析にあたっては、名詞句を対象とし、「検討する+必要+ある」「周知+必要+ある」「確認+必要+ある」等、市としての災害対応課題を提示する内容が主として抽出された。本研究では行政における災害対応の初動期から応急期にかけての内容であるため、限定性を持つが、「係り受け」による単語単位の形態素と組み合わせることで、より文章内容の全体構造や特性を把握することが可能となる。

4. 結論と課題

本研究で行った自由回答の形態素に基づく計量テキスト分析では、分析担当者の主観に依拠せず、客観的に解析が可能である点において利点をもち各種調査の自由回答記述の分析において、有効な手法となることが示唆された。しかし、本手法ではさまざまな意見を内包している自由回答に対し、統計学的な作業上、出現頻度の少ない用語を排除して集約をしているため、「少数ではあっても重要な意見」を捨象してしまう危険性を内包している。そのため実際の解釈や政策決定への本手法の援用に際しては、原文を参照することで慎重に解釈を行っていくことに留意する必要がある。

【参考文献】

- 1) 那須野哲哉, 2001, 「コールセンターにおけるテキストマイニング」, 『人工知能学会誌』16 (2), 219-225
- 2) 川喜田二郎, 1967, 「発想法」, 中公新書
- 3) 岩見麻子・大野智彦・木村道徳, 2011, 「公共事業計画策定過程の議事録分析のための変動係数を用いた対象語選定手法の開発」, 『環境情報科学』25, 55-60
- 4) 林俊克, 2003, 「テキストマイニングの現在」, 『マーケティングリサーチ』94, 16-25