

地域災害文献を対象とした災害語彙抽出手法の予備的検討

— 文献調査と計量的整理を組み合わせた方法論整理 —

三谷直哉¹

¹国立文化財機構文化財防災センター

mitani-n9t@nich. go. jp

概要

本研究は、地域災害文献を対象とした災害語彙抽出における処理手順と課題を、言語処理の観点から整理するものである。奈良県十津川村に関する方言語彙集、地域史資料、文芸作品を対象に、文献精読に基づく語彙抽出と簡易的な計量的整理を行った。その結果、災害や被害を直接指示する語彙と、文脈依存的な状況描写語・比喩的表現が混在していることが確認された。本研究では、災害語彙抽出の各工程において、機械的処理の適用可能性と人手判断を要する範囲を明示し、地域文献を対象とした語彙抽出手法の検討に資する知見を提示する。

1 はじめに

本章では、本研究の背景および目的を整理し、地域災害文献を対象とした災害語彙抽出という課題が、言語処理の観点からどのように位置づけられるのかを示す。まず、地域資源と防災の関係、ならびに方言と災害記憶との関係について概観する。そのうえで、本研究が着目する問題意識と研究目的を明確化する。

1.1 研究背景

近年、災害対応や防災教育において、地域に蓄積された知識や経験といった「地域資源」を活用する重要性が指摘されている。災害時の判断や行動は、必ずしも画一的なマニュアルのみで説明できるものではない。地域固有の自然環境や、過去の災害経験に強く依存する側面を有している。そのため、地域社会に内在する災害経験をどのように把握し、将来に継承していくかは重要な課題である。

地域資源の一つとして、方言や地域特有の語彙が挙げられる。方言は、地域の自然環境や生活様式を反映して形成されてきた言語表現である。災害に関する経験も、語彙や表現として残されている可能性

がある。本研究では、このような災害を示す方言語彙を「災害語彙」と呼ぶ。災害語彙を収集・整理することで、地域の災害経験を明らかにするとともに、防災への取り組みを再認識する手がかりを得られると考えられる。とりわけ、被害状況や異変の兆候、危険の認識などは、比喩的で状況依存的な形で表現されることが多い。そのため、語彙単体では意味の把握が難しく、文脈を考慮した解釈が必要となる場合が多い。

一方で、地域文献に含まれる災害語彙は、一般的な自然言語処理で想定される大規模かつ均質なコーパスとは大きく性質が異なる。資料規模が小さく、文体や表記が多様であること、方言や比喩的表現が多く含まれることから、既存の自動処理手法をそのまま適用することは困難である。したがって、地域災害文献を対象とした災害語彙抽出においては、資料特性を踏まえた方法論の整理が必要である。

1.2 研究目的

本研究の目的は、地域災害文献を対象として、災害語彙を抽出・整理するための予備的な方法論を整理することである。あわせて、その過程で生じる課題を明らかにすることを目的とする。本研究では、文献精読に基づく人手作業と、簡易的な計量的整理を組み合わせた手法を試行する。その結果として、災害語彙抽出の各工程において、どの部分が自動化可能であり、どの部分に人手判断が不可欠であるかを明示する。

1.3 本研究の位置づけ

本研究は、地域災害文献を対象として、災害語彙をどのような手順で抽出・整理できるのかを整理することを目的とする。文献精読に基づく人手作業と、簡易的な計量的整理を組み合わせた分析を通じて、災害語彙抽出の各工程において生じる制約や判断の必要性を明らかにする。本研究で得られる知見は、

今後、災害語彙の体系化や自動抽出手法を検討する際の前提条件を整理するための基礎的な位置づけを持つ。

2 対象地域および調査資料

本章では、本研究における分析対象を明確にするため、対象地域の概要および調査に用いた文献資料について説明する。地域特性および資料の性質は、災害語彙の抽出結果や方法論上の制約に直接影響するため、本章でその前提条件を整理する。

2.1 対象地域の概要

本研究では、奈良県十津川村を対象地域とした。十津川村は山間部に位置し、地形条件の影響から、水害や土砂災害を繰り返し経験してきた地域である。過去には、大規模な風水害による被害も記録されている。このような背景から、災害経験が地域社会に深く刻まれていると考えられる。

また、十津川村に関しては、方言語彙集、地域史資料、文芸作品などの文献資料が比較的まとまって残されている。これらの点から、本研究において災害語彙を検討するための資料的条件を備えた地域であると判断した。

2.2 調査資料

調査対象とした資料は以下の三種である。

- 方言語彙集（1点）[1]
- 地域史資料（4点）[2, 3, 4, 5]
- 文芸作品・昔話資料（4点）[6, 7, 8, 9]

方言語彙集は、地域で用いられてきた語彙を体系的に収録した資料である。本研究では、災害や被害を明示的に指示する語彙を把握するための基盤資料として用いた。一方、地域史資料や文芸作品・昔話資料には、語彙集には収録されていない状況描写や比喩的表現が多く含まれている。これらの資料からは、災害経験が文脈的に表現されている可能性がある。

本研究では、分析の再現性および資料入手性を考慮し、文献資料に限定して調査を行った。方言は本来話し言葉であり、語り部の聞き取りなどの口語資料を用いることで、より多様な表現が得られる可能性がある。しかし、分析単位や記録条件の整理が必要であるため、これらについては今後の課題とする。

3 方法

本章では、本研究における災害語彙抽出および整理の手順について説明する。本研究では、地域災害文献を対象とし、文献精読に基づく人手作業と簡易的な定量的整理を組み合わせた方法を採用した。図1に、本研究で用いた処理フローの概要を示す。以下では、各工程の内容と判断基準を順に述べる。

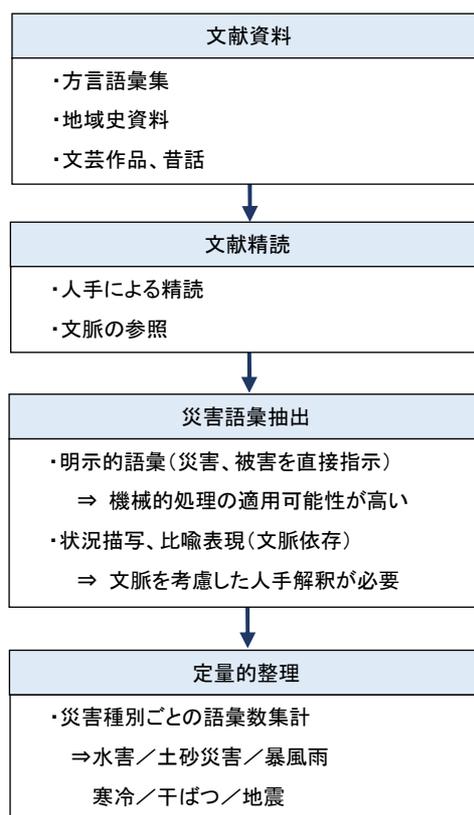


図1 災害語彙抽出・整理フロー

3.1 災害語彙抽出手順

本研究における災害語彙の抽出は、調査対象文献の精読を基本とした。まず、方言語彙集を基盤資料として、災害や被害を明示的に指示する語彙を抽出した。これにより、災害語彙の最小単位として、語彙単体で把握可能な語彙群を整理した。

次に、地域史資料および文芸作品・昔話資料を精読し、災害発生時の状況や環境の変化を示唆する表現を抽出対象に加えた。これらの表現は、語彙単体では災害との対応付けが困難である。多くの場合、前後の文脈を参照することで初めて、災害との関連性が判断できる。

文献精読の結果、災害関連語彙として計 42 語を抽出した。このうち、災害や被害を明示的に指示する語彙が 40 語であった。一方で、状況描写語や比喩的表現は 2 語であった。割合としては少ないものの、災害語彙抽出の過程において、文脈を参照した人手判断を前提とする語彙が一定数含まれることが確認された。

3.2 災害種別分類

抽出した災害語彙について、災害種別による分類を行った。分類カテゴリは、代表的な災害として、水害、土砂災害、暴風雨、寒冷、干ばつ、地震の 6 種とした。

複数の解釈が可能な語彙については、語彙単体ではなく出現文脈を優先し、人手判断により分類を行った。この工程は、災害語彙分類における曖昧性の所在を確認する目的も併せ持つ。

3.3 定量的整理

分類結果に基づき、各災害種別に含まれる語彙数の集計を行った。本整理は、統計的検定や有意差の評価を目的とするものではない。災害語彙の分布傾向を把握するための分析として位置づける。

このような定量的整理を行うことで、語彙分布と対象地域の災害特性との関係を議論するための基礎的な情報を得ることを目的とした。

4 結果と考察

4.1 災害語彙の分布

語彙集計の結果、土砂災害および水害に関連する語彙が相対的に多い傾向が確認された。集計結果を表 1 に示す。例えば、土砂災害関連語彙が 16 語、水害関連語彙が 14 語であり、全体に占める割合も高かった。一方で、暴風雨、寒冷、干ばつに関連する語彙は少数であった。また、地震に関連する語彙は、今回の調査では確認されなかった。

表 1：災害語彙の集計結果

災害種別	語彙数 (語)
土砂災害	16
水害	14
暴風雨	5
寒冷	4
干ばつ	3
地震	0(該当なし)

続いて、抽出された災害語彙の例を表 2 に示す。抽出された語彙の多くは、災害現象そのものや、災害によって生じた環境変化を示すものであった。

例えば、「くえ (崩え)」「山津波」「山抜け」は、土砂災害そのものや、災害によって山の地形が変化した様子を示す語彙である。また、「まくれ水」は川の氾濫を、「ささにごり」は水の濁りの状態を描写している。

一方で、「山の腹が噴き出す」「川の音が死ぬ」といった比喩的表現も見られた。これらは、共通語の語彙を用いているが、一般的な用法とは異なる形で使用されている。災害による衝撃や異常性を強調する語感を持つ表現であると考えられる。

表 1：災害語彙の分類結果

災害種別	災害語彙の例
土砂災害	くえ、山津波、山潮、山抜け、など
水害	まくれ水、ささにごり、ゴタ、など
暴風雨	しけ、大しけ、みしろもち
寒冷	いてる、さぶがり、ゆきおこし、など
干ばつ	雨休み、雨よるこび、しげしげ、など
地震	(該当なし)

4.2 災害史との照合

抽出語彙の分布を既存の災害史資料[10, 11]と照合したところ、頻出する災害種別と歴史的な被災傾向との間に一定の対応関係が見られた。一方で、台風のように語彙数が少ない災害種別についても、災害史上は被害が確認されている事例が存在した (表 3)。このことから、災害語彙の分布は災害履歴を直接反映するものではなく、文献の性質や記述文脈の影響を強く受けることが示唆される。

表 3：十津川村における過去 100 年の主な災害

発生前	災害名称
昭和 9 年 (1934 年)	室戸台風
昭和 25 年 (1950 年)	ジェーン台風
昭和 27 年 (1952 年)	ダイナ台風
昭和 28 年 (1953 年)	紀和水害
昭和 34 年 (1959 年)	伊勢湾台風
昭和 36 年 (1961 年)	第 2 室戸台風
平成 2 年 (1990 年)	9 月豪雨
平成 15 年 (2003 年)	8 月豪雨
平成 23 年 (2011 年)	紀伊半島大水害
令和 2 年 (2020 年)	上湯川斜面災害を伴う風水害

4.3 方法論上の課題

抽出語彙のうち、約5% (2件) は状況描写語や比喩的表現であり、語彙単体では災害との対応付けが困難であった。これらは災害や被害を直接指示せず、文脈を通じて災害状況を示す表現であるため、表層的な語彙一致に基づく抽出手法では十分に捉えられない。困難性は、主に次の三点に整理できる。

- 比喩的表現による意味の間接性：
⇒災害を直接指示せず、状況や印象として表現される
- 文脈依存性の高さ：前後の記述を参照しなければ意味解釈が成立しない
- 方言特有表現による語形・意味の多様性：表記や語義の揺れが大きい

これらの結果は、既存のキーワードベース手法や頻度分析のみでは、地域災害文献に含まれる災害語彙を十分に扱えない可能性を示している。一方で、人手判断を前提とした整理によって、災害語彙の特徴を把握できる余地があることも確認された。

5 おわりに

本研究では、地域災害文献を対象として、災害語彙抽出のための予備的な方法論を整理した。文献精読と簡易的な定量的整理を組み合わせることで、災害語彙の分布傾向とともに、処理工程ごとの制約条件を明示することができた。

今後の課題として、資料の拡充、抽出基準の明確化、人手判断と機械処理を補完的に組み合わせた分析手法の検討が挙げられる。これらを通じて、地域災害文献に対する言語処理技術の接続可能性について、引き続き検討を行いたい。

参考文献

- [1] 榑本利清. 十津川言葉. 十津川村. 1975.
- [2] 内野吾郎. 方言基礎語彙の研究—奈良県十津川方言を中心として—. 国学院大学日本文化研究所. 1979.
- [3] 十津川村教育委員会. 千葉政清氏遺稿 十津川方言集. 十津川村教育委員会. 1996.
- [4] 十津川村史編さん委員会. 十津川村史 歴史資料編（古代中世・文化財）. 十津川村. 2025.
- [5] 西田正俊. 十津川郷. 奈良県吉野郡十津川村役場. 1994.
- [6] 川村たかし. 北へ行く旅人たち—新十津川物語. 偕成社. 1989.
- [7] 川村たかし. 十津川出国記. 北海道新聞社. 1987.
- [8] 十津川村教育委員会. 十津川郷の昔話 一集. 第一法規出版. 1989.
- [9] 十津川村教育委員会. 十津川郷の昔話 二集. 第一法規出版. 1989.
- [10] 国立研究開発法人 防災科学技術研究所. 災害年表マップ, (閲覧日 2026-01-05) . <https://dstr.mhr.bosai.go.jp/saigai2016/>
- [11] 歴史から学ぶ 奈良の災害史. 奈良県総務部知事公室 防災統括室. 2014.