

テキスト分析による地震記録評価の試み

濱野未来(立命館大学文学研究科)

§ 1. はじめに

地震が記録された史料は、武者金吉をはじめとする先達により、発掘・収集・整理がなされ、それらを典拠として、歴史地震研究の蓄積がなされてきた。史料上の地震の記述から被害の検討が行われる一方で、どのように記されているのか、すなわち、地震がどう表現されてきたのかという点については十分な議論がなされてこなかったといえる。そこで、本報告では地震の“記録のなされ方”を対象に分析を行った。

§ 2. 分析手法

本報告では、より客観的な分析を行うために計量テキスト分析という手法をとった。計量テキスト分析とは、テキストデータのなかから言葉を取り出し、出現頻度や相関関係・出現傾向などを解析し、それにより有用な情報を取り出すというものである。本来は、主にインタビュー記事や新聞記事等の分析に用いられる。そのため、歴史史料を対象に活用されていないが、歴史地震の記録表現の分析にどれほど有用といえるのかという点の検討も兼ね、この試みを行った。なお、本報告ではテキストマイニングソフトウェア『KHcoder』を使用した。

§ 3. 分析対象

今回、分析対象としたのは、15~16世紀の地震史料である。なかでも、比較的多数の記録がなされている文安京都地震(1449)・明応地震(1498)・天正地震(1585)・慶長伏見地震(1596)の4つの地震の史料における、本震から一ヶ月間の記録を対象とした。中世から近世へと移り変わる当該期の史料の比較分析を行うことで、その変容を捉える一助とするためである。

§ 4. 地震記録表現の性質

今回は、主に特定の語の前後に、どんな語が多く出現していたかを読み取ることができるコロケーション統計や、頻出語分析によって地震記録表現の性質をみていきたい。

§ 4.1. 「地震」に付随して表れる語

コロケーション統計からは、どんな語が「地震」という語に付随して表れるかを知ることができる。年月日や時刻を表す語が「地震」という語の前に頻出することは、日記史料の性質上からも自明であるが、今回の統計からは、各地震によりこうした時間に関連する語の数にばらつきがみられた。各地震史料における時間に関連する語の数は、文安:11, 明応:17, 天正:14, 慶長:7

(各史料内上位50語中)となった。慶長地震の7語という数字は、孝亮宿禰日次記など、各日に発生した地震を逐次記し、先に起きた地震を後日に記すことが少なく、「今日」や「刻」以外の時間に関する語の使用が少ないことを反映しているといえる。

§ 4.2. 損傷状況の表現

地震による建造物などの損傷状況を示す語としては、「破損」「崩」「壊」などがある。そのうち、「顛倒」は、慶長地震の記録中での出現回数が23回で上位である。文中での「顛倒」の主体をみると、多くは禁裏や伏見、東寺内の建造物であった。また、伝聞を表す「云々」という語とセットで出現している事例も多い。すなわち、人々の関心が向けられる場(禁裏・伏見城)や、著名な場(東寺)における被害についての伝聞記録が多く、そうした場では、諸殿などの「顛倒」被害が顕著であったことの反映ともいえる。

また、慶長地震の記録では、今回比較した他の地震記録には出現していない「大破」という表現が5回出現しており、損傷状況を示す表現の時代偏差が示唆される。

§ 5. おわりに

今回、15~16世紀の地震史料を対象に、計量テキスト分析に基づいた、地震表現の性質の検討を試みた。それにより、本手法を歴史史料に活用していくに際しての課題も浮かび上がってきた。例えば、抽出された語の品詞が文中での使用法と異なる場合や、語の区切り方が不自然になる場合などである。これらは、計量テキスト分析ソフトウェアが、漢文体のテキストを想定していないために発生するものでもあり、漢文体にカナや送り仮名が混合した中近世移行期における史料を扱う上での課題でもある。しかし、これらの点は、分析の前処理段階において、語の強制抽出や品詞指定・複合語指定を丁寧に行うことによって、よりクリアな分析が可能となると考えられる。

以上のように、計量テキスト分析を地震史料の分析に活用していく上での課題は多数あるものの、新たな分析視角をもたらす手法として、今後さらなる検討と試行錯誤を行っていくべきと考えられる。

§. 文献

大野隆造編『シリーズ〈都市地震工学〉7 地震と人間』朝倉書店、2007年 瀬尾和大執筆項
樋口耕一『社会調査のための計量テキスト分析』ナカニシヤ出版、2014年