

みんなで翻刻「賀茂社関係文書翻刻プロジェクト」から得られた地震記事

加納 靖之*(東京大学地震研究所/地震火山史料連携研究機構)・山本 宗尚(賀茂県主同族会)・

橋本雄太(国立歴史民俗博物館)・青池 亨(国立国会図書館)・中西 一郎(京都大学名誉教授)・

大邑 潤三(東京大学地震研究所/地震火山史料連携研究機構)・濱野 未来(立命館大学大学院文学研究科)

§ 1. はじめに

「みんなで翻刻」の「賀茂社関係文書翻刻プロジェクト」(<https://honkoku.org/app/#/projects/kamosha/info>)は、国立国会図書館が所蔵する『賀茂社記録』など、デジタルアーカイブとして公開されている賀茂社に関する記録を翻刻するプロジェクトである。『賀茂社記録』には地震の記事も含まれており、その抽出と整理をおこなった。

§ 2. みんなで翻刻による地震史料の発見

2017年1月に開始した「みんなで翻刻」(バージョン1)では、東京大学地震研究所図書室が所蔵する「石本文庫」をはじめとする和古書類を翻刻してきた。2019年7月には地震史料だけでなく幅広く歴史史料の翻刻を行うためのプラットフォームとしてリニューアルし、そのなかのプロジェクトとして地震史料を翻刻するものを登録してきた。

2023年5月に開始した「賀茂社関係文書翻刻プロジェクト」に登録した『賀茂社記録』(全97冊、第42冊のみ未デジタル化)には、氏人の日記なども含まれ、地震や天気に関する記述がある。

『賀茂社記録』は、国立国会図書館の次世代デジタルライブラリー(<https://lab.ndl.go.jp/dl/>)において全文検索することができる。「地震」をキーワードとして検索したところ48件が抽出された。そのうち実際に地震についての記述であるものが21件あった。そのうち、祈禱文のなかに記載され実際に起きた地震ではないと思われるものが1件、付箋の有無を別コマとして撮影したために重複したものが1件ある。残る19件の地震発生日の内訳は、1400年代1件(ただし、別の日記からの写し)、1600年代1件、1700年代3件、1800年代14件である。別の日記からの写し1件を除いた18件のうち、16件は上賀茂神社周辺で感じられた地震であると考えられる。2件は葵使の道中日記のもので、1件は箱根付近、もう1件は諏訪付

近の有感記録と考えられる。既刊の地震史料集の網文に挙がっていない日付のものが3件ある。

当初、地震に関する記述があるとは予想していなかった史料群であるが、「みんなで翻刻」への登録をきっかけに地震の記述を発見することができた。「みんなで翻刻」に多様な史資料を登録することで、地震史料の発見にもつながる可能性がある。

§ 3. みんなで翻刻のテキストデータ活用とAI翻刻

次世代デジタルライブラリーで公開されているテキストデータは、AI(機械学習)を用いたNDL古典籍OCRによって生成されたものである。このNDL古典籍OCRは、みんなで翻刻が成果物として公開しているテキストデータを教師データとして用いている。

前述のように「地震」でキーワード検索した結果を分析したが、史料上では「地震」となっていないものが半数以上を占めていた。「社地拾四坪」という記述の「地拾」を「地震」と判読しているケースは、誤認識といえる。また、日付のあとに誤って「地震」と挿入されているケースがあるが、これは教師データとして利用した「みんなで翻刻」の資料の性質に影響を受けたのかもしれない。

「みんなで翻刻」ではNDL古典籍OCRによっていわば下読みしたものを人手で修正することができる。こうして生成されたデータを追加して学習することで、よりよい認識モデルを構築することができるだろう。

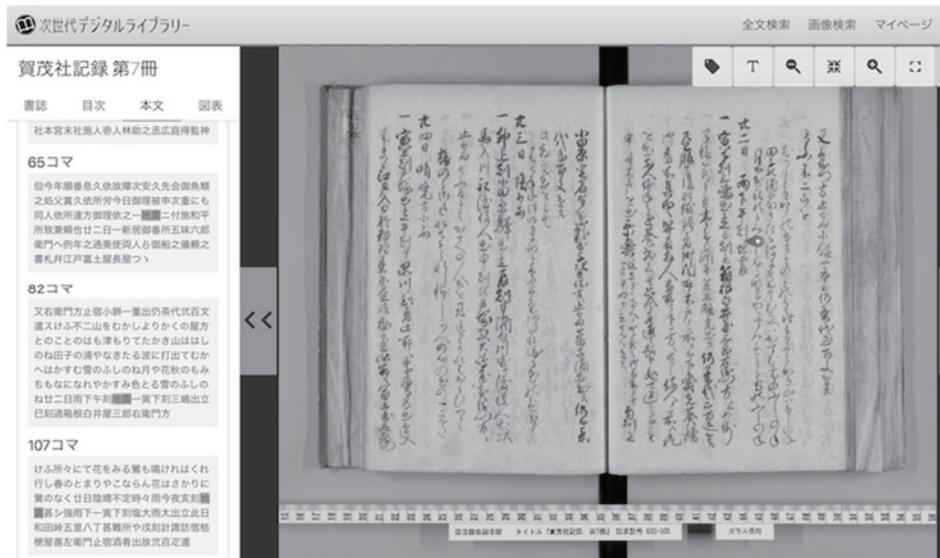


図. 次世代デジタルライブラリーでの『賀茂社記録』の「地震」での検索結果の例。