

Text Encoding Initiative ガイドラインに基づく古地震史料の電子化 Digitization of Historical Earthquake Recordings Following the Text Encoding Initiative Guidelines

○橋本雄太・加納靖之・大邑潤三

○Yuta HASHIMOTO, Yasuyuki KANO, Junzo OHMURA

Although a number of historical earthquake recordings have been studied and transcribed for years, most of their transcription texts are provided as plain texts. In this paper, we will propose to use the Text Encoding Initiative (TEI) guidelines for encoding these recordings, in order to enable machine processing on these recordings. The TEI guidelines, which have been formulated in Europe since 1980s', specify a standard way to encode historical documents in machine-readable XML formats. We have experimentally encoded 信越震漲録, a recording of the Zenkoji Earthquake (1847), into TEI format, and developed a set of visualization programs that make it possible to grasp the damages described in the recording intuitively.

1. はじめに

歴史地震史料の研究は百年にわたる蓄積を有する分野であり、これまでに多数の史料が翻刻されてきた。近年はこれらの翻刻テキストをデータベース化し、Web上で公開・共有するための動きも進んでいる。他方で、既存の翻刻テキストの多くはプレーンテキスト形式で提供されており、情報の検索や加工に関して難がある。

たとえば、史料中に現れる日時や場所、被害規模に関する情報をピックアップするためには、人間が翻刻テキストの全文に目を通さなければならない。古文書の読解に不慣れな研究者にとって、この作業は大きな負荷となる。地震史料からの効率的な情報抽出を可能にするためには、翻刻テキストが何らかの機械処理可能なフォーマットで記述されていることが望ましい。

2. Text Encoding Initiative ガイドライン

本研究の目的は、人文学テキストの電子化フォーマットである Text Encoding Initiative (TEI) ガイドライン [1] にもとづいて、歴史地震史料の翻刻テキストを機械可読な形式で記述する手法を開発することである。TEI は、1980年代から欧州のアーキビストを中心に策定され、デジタル人文学 (digital humanities) 分野では人文学のテキストを電子化する際の国際標準フォーマットとして流通してきた。

TEI の特徴は、対象文書を汎用マークアップ言語の XML (eXtensible Markup Language) で記述す

ることにある。TEI ガイドラインには、日時や場所など、テキスト中の特定要素をマークアップするための数百種類の XML タグや属性 (attribute) が定義されている。これらを利用することで、翻刻テキストの構造的・意味論的内容を機械処理可能なフォーマットで記述することが可能になる。

3. 『信越震漲録』 翻刻テキストの TEI 化

発表者らは、TEI の試験的な地震史料への適用として、『信越震漲録』の翻刻テキストを TEI 化する作業を実施している。『信越震漲録』は、弘化4年 (1847) に発生した善光寺地震の記録文書である。内容は信濃及び越後の藩、代官所から幕府側に送られた届書が大部分を占めている。発表者らが所属する京都大学古地震研究会では、『信越震漲録』を含む善光寺地震に関わる数点の史料を翻刻してきた。

図1は、TEI ガイドラインにしたがって記述した同史料の一節である。本研究では、TEI が提供する次のようなタグを利用して翻刻テキストをマークアップした：

- <div>: 届書全体
- <head>: 届書の題名
- <date>: 日時に関する記述
- <seg>: 地震被害に関する記述
- <persName>: 人名
- <placeName>: 町や村などの場所名称
- <geogName>: 山、川などの地理名称

さらに、具体的な日時情報や地震の被害種類を

XML タグの属性に記述している。

```
<div type="section" >
<head><date when="1847-05-15">四月朔日</date> 御届</head>
<p>一  我在所<placeName>信州松代</placeName>
<date when="1847-05-08T22:00:00">一昨廿四日亥刻頃</date>
より大地震二而、
<seg type="damage" subtype="building">
城内住居向櫓并垣塙等夥敷破損、
家中屋敷城下町領分村々其外支配所潰家数多、
死失人夥鋪、
</seg>、
<seg type="damage" subtype="fire">
殊二村方二者出火も有之
</seg>
<seg type="damage" subtype="landslide">
其上山中山拔崩、<geogName type="river">犀川</geogName>へ
押埋水堪(湛)追々致充滿、勿論流水一切無之、北国往還
<placeName>丹波島宿</placeName><placeName>渡船場干上り二相成、此上
右溢水押出方二寄如何様之変地も難斗奉存候
</seg>、
只今以折々相震申候、
委細之義者追而可申上候得共此段御届申上候、 以上
</p>
</div>
```

図 1 : TEI を用いて記述した『信越震漲録』の一節

このように翻刻テキストをフォーマットしておくことで、史料から機械的に情報を抽出することが可能になる。例えば、火災被害に関する記述を一括取得するようなプログラムは、Ruby や Python などのプログラミング言語を利用すれば数行で記述可能である。さらに、取得した情報を場所情報や時間情報とリンクさせることにより、様々な分析が可能になる。

4. TEI 化した地震史料の可視化

TEI 化した文書は、様々な方法で機械処理を適用することが可能である。たとえば GIS 技術を利用することで、地震史料中に記述された被害報告のマップを自動生成することができる。

図 2 は、TEI 化した『信越震漲録』から自動抽出した善光寺地震の被害報告を 3D マップ上に投影したものである。3D マップの生成には WebGL ベースの 3 次元全球エンジンである Cesium [3] を利用しており、結果は Web ブラウザ上で閲覧することができる。こうした可視化データを生成するには通常研究者の手作業が必要になるが、史料本体を機械可読形式で記述すること

により、プログラムによる自動生成が可能になる。

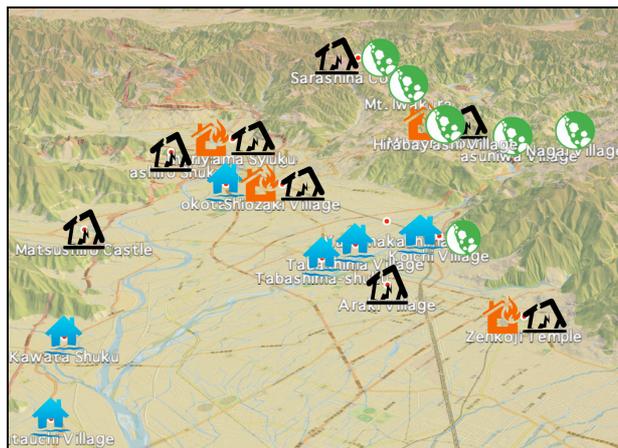


図 2 : 3D マップ上に投影した被害状況

5. まとめと今後の展望

本稿では、TEI を利用して地震史料をマークアップする手法と、TEI 化した史料を可視化する手法を紹介した。ポスター発表では、可視化プログラムのデモンストレーションを併せて行う予定である。

参考文献・URL

- [1] Text Encoding Initiative Guidelines, <http://www.tei-c.org/>
- [2] 五島敏芳ほか (2013) : 弘化四年(1847)善光寺地震について, 京都大学防災研究所年報, 第 56 号 B
- [3] Cesium, <https://cesiumjs.org/>

謝辞

本研究は、平成 27 年度京都大学防災研究所萌芽的共同研究(課題番号 27H-03)の助成を受けたものです。