Quantitative Content Analysis on Historical Earthquake Documents Transcribed by Online Transcription Project "Minna De Honkoku"

*Yasuyuki Kano[1], Yuta Hashimoto[2]

1. Research Institute for Earthquake Prediction, Disaster Prevention Research Institute, Kyoto University, 2. National Museum of Japanese History

We have launched Web-based transcription project "Minna de Honkoku" (https://honkoku.org/) in January, 2017. "Minna de Honkoku" is also the name for Web application to realize this online transcription project. "Minna de Honkoku" has transcribed 386 documents out of total 421 documents included in the corrections of Earthquake Research Institute Library, the University of Tokyo. Total number of inputted character is about 3.56 million. The study of historical earthquake is based on historical documents. In Japan, almost all of the documents are written in Kuzushi-ji. Kujzushi-ji is writing style used before ˜1900. Since the style is different from that of modern Japanese, transcription is necessary to use the historical documents as data for earthquake research. Here we made quantitative content analysis or text mining on a large set of texts produced by "Minna de Honkoku." We used KH Coder (http://khc.sourceforge.net/en/), software for quantitative content analysis or text mining. We counted the word frequency words that appears. The most frequent words are those to show: earthquake, collapse, water, people, mountain, fire, town, temple, lodge, river, damage. It is natural to have those frequent words, since the theme of texts analyzed are basically earthquake and its damage. This result is considered to be consistent with qualitative impression from texts Catalogs of historical records such as "New collection of materials for the history of Japanese earthquakes" that has published and used for previous earthquake research. The impression is quantitatively confirmed by the analysis using the set of texts. We tested combination of words with similar appearance patterns, co-occurrence. The word "Earthquake" frequently co-occur with the words indicating directions and places as well as causalities and damage of buildings. For further analyses, we need appropriate dictionary optimized for Kuzushi-ji writing and earthquake documents. Using the result of quantitative content analysis, we can recursively revise the dictionary. Acknowledgement: "Minna de Honkoku" is launced and operated by Research Group for Historical Earthquake, Kyoto University. "Minna de Honkoku" used documents included in the corrections of Earthquake Research Institute Library, the University of Tokyo. Transcription on "Minna de Honkoku" is made by voluntary users including anonymous one.

Keywords: transcription, Japanese Kuzushi-ji, open science, quantitative content analysis, text mining