Application of TEI to Engishiki and Japanese History Text Data Sharing

KOKAZE Naoki and GOTO Makoto

In this paper, we have examined the digitalization of the text information of Engishiki and strategies for its distribution. In particular, we have discussed the data obtained through the application of an international standard known as TEI (Text Encoding Initiative), and introduced ways to more broadly circulate the text data of Japanese historical materials.

This paper we do not discuss in detail the technical aspects of our project, but we rather focus on its significance from the standpoint of historical research. Aiming to facilitate the practical use of the text data for international circulation and research, we have applied TEI markup to the Engishiki. TEI (Text Encoding Initiative) is a project intended to promote the international circulation and use of textual data in the humanities, and it’s also called the standards developed within these project. Since TEI is a tool specifically developed for the digital humanities, it offers the advantage of allowing users to record information on one’s understanding of historical materials through the use of “elements” (pieces of information to which a tag has been attached). Bearing such merits in mind, the authors have proposed a digitalization model based on TEI.

In particular, we have applied TEI markup to quantitative data of Engishiki and ventured into an examination of the possibility of parsing an overall picture of the text through the use of a method known as “transactionography.” In addition, we have also discussed ways of creating a manual on how to use these markups as information bases.

The development of methods of visualization for historical materials and research methodologies is necessary in order to more active the research on Japanese history. At a time in which the humanities and history are said to be in “corner”, it is all the more urgent to construct base information, distribute it freely, and devise new methods that can lead to new avenues for inquiry.

Keywords: TEI (Text Encoding Initiative), markup to the Engishiki, Information infrastructure construction, Historical informatics, the digital humanities