Beer: Digital Transcription of a Medieval Manuscript

The benefits of digitally transcribing Medieval manuscripts go beyond the ease of publication online and extend into pedagogy. Digital transcription allows scholars to precisely identify each word and put forward a normalized text quickly for non-epigraphers to work with. The methods of digital transcription are rather easily taught to classes and projects similar to the one here make for an effective learning experience for students. The methods of transcription in this project are readily applied to other source materials, such as inscriptions.

MS 13 is a 16th century German manuscript written in Latin in Tufts University's Tisch Library by an unknown author. The topic of this short, six page, manuscript is a treatise on the physical and philosophical qualities of beer. The treatise explores the origins of beer, going as far back as Egypt, and also the opinions of Ancients, including references to ancient writers and texts: such as a quote from Galen. The treatise also discusses the effects of various ingredients in beer including barley, wheat, and hops. Asides from showing the author's thoughts on beer, the text is indicative not only on the technical knowledge of beer available to people in the 16th century, but also about what ancient texts were available to those people.

The transcription of MS-13 was performed on the Perseids editing platform with the goal to make a scientific transcription available in a user-friendly display (http://sosol.perseids.org/sosol/publications/22807). The transcription used the Epidoc rules for XML coding. XML or Extensible Markup Langauge "is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable" (https://en.wikipedia.org/wiki/XML). Using Perseids allowed the transcription of each word along with an exact image of the individual word to be referenced and to grant quick access to a mode of presentation. The benefit of using Epidoc was to use a set of rules established by the

Epigraphic community to transcribe MS 13 into and XML document

(https://sourceforge.net/p/epidoc/wiki/Guidelines/). By digitally transcribing the text, ligatures, absent text, and misspellings could all be coded and rationalized into a user-friendly interface at once without losing the evidence that they all existed in the text. By transcribing onto and XML document, XSLT documents for transforms and presentation could be written and applied easily to emphasize different parts of the text, including but not limited to, the presence of other languages, ligatures, and recurring words.

The process of transcribing MS13 began with making high quality digital images of the manuscript and uploading them to Perseids. Next was the delimiting of each word, including all ligatures, accents, and serifs on the image. After marking a box around each word in a line, a preliminary transcription of each word can be made by typing what is seen on the page and then copying the codes of each word's content and mapping coordinates into the XML document. If there are any irregularities in the words, such as ligatures or absent text, the instances were noted using the rules of Epidoc transcription. This method of transcribing each word individually preserves not only the image of the word in the text, but all of the irregularities of the words while still putting forward an understandable text to readers.

After transcribing the whole text with this method, a normalized text can be made quickly on Perseids. The normalized text on Perseids can also be used to detect any irregularities in the XML document. From a combination of the normalized text and digital images of the manuscript, a translation was written and attached to the transcription on Perseids. Outside of using Perseids, once the XML transcription is complete, an XSLT transform can be written to create HTML websites to emphasize different aspects of the text, such as other languages,

ligatures, or missing text. The benefit of writing an XSLT transform is that it can be customized

not only to each text, but also to emphasize different aspects of the work at different times.