Shawn Martin
University of Michigan
shawnmar@umich.edu

# Corpus Inter Corpora: The Text Creation Partnership

How should one create a large corpus of text? What standards do you set? How do you manage a diverse group of potential users? These, among other questions are ones the Text Creation Partnership (TCP) project at the University of Michigan grapples with on an almost daily basis. With a collection of currently over 8,000 texts (and growing), the TCP seeks to create a vast collection of texts created between 1470 and 1900, but wants to ensure the usefulness of the corpus for research and teaching for years to come.

This raises many issues. What standard do we use for a corpus of text that we anticipate will be used for many purposes in the future? How do we balance the selection of that text given the many different audiences who have an interest? What role do producers of text have to their users? What role do researchers and teachers have in shaping the corpus? The TCP project, having considered these problems for many years, has constructed a potential model for collaboration and cooperation among text producers and users that hopefully can benefit both parties, and has kept to a model of annotation and tagging that will hopefully insure the usefulness of this collection for many years to come. By continually addressing these issues, it is hoped that this model can continue to evolve and best fit the needs of both users and producers of text collections in the future.