Prosody is a historically contingent term. Both the branch of linguistics concerning pronunciation (cadence, pitch, tone) and the aesthetic category of poetic form including versification (meter, rhythm, rhyme, verse forms), prosody sits at the border of science and aesthetics and at the invention of linguistics and literary study as disciplines.

**What Is Prosody?**

The PPA is a full-text searchable database of ~5,000 HathiTrust-digitized works on prosody published between 1570 and 1923. It collects historical documents and highlights discourses about the study of language, the study of poetry, and how these intersect and diverge. Rather than a static repository of historical data, it compels users to rethink the past and future of organizing, navigating, conceptualizing, and historicizing large amounts of data.

**What Is the PPA?**

Because duplicate works were skewing search results on our beta-site, we knew we would have to clean our initial dataset. We created a spreadsheet of the metadata, ran a script to cut all duplicate HathiTrust catalog IDs, and painstakingly hand-checked these records for accuracy. This whittled the PPA down to ~5,000 works, showing how many duplicates we were hosting – about 40% of the original file transfer.

With a clean, de-duplicated database, we teamed up with the Princeton CDH to create an intuitive user interface with a robust administrative backend. This is crucial for managing HathiTrust works and allows PPA team members to edit metadata, add new works, and even group documents into filterable collections.

**PPA + HathiTrust**

All material in the PPA comes from HathiTrust, a partnership of academic institutions that offers a collection of millions of titles digitized from libraries around the world. The PPA teamed with Hathi in 2011 to build an archive dedicated to prosody. We used Library of Congress subject headings to curate a dataset of prosodic titles, and Hathi delivered both the plain text and metadata for ~8,500 fair use monographs. We soon discovered that this method led to many duplicate works due to metadata errors and host library mislabeling.