

Proust: Extending AI to Translated Modernist Fiction

The field of digital humanities has grown tremendously over the past decade to encompass a variety of research methods and areas that allow scholars to analyze literature quantitatively and computationally. However, such research has been dominated by studies of pre-1900 works originally written in English. Focusing on older texts avoids logistical copyright challenges, but the modernist novel differs greatly in both content and form from its predecessors and demands further study in the digital humanities. In addition, applying AI and computational methods to translated works raises unique challenges for parsing and data cleaning and the question of whether AI can be used to draw meaningful conclusions from translated texts is still open.

I propose Marcel Proust's *In Search of Lost Time* (*À la recherche du temps perdu*) as an excellent case study for research at the edges of the digital humanities. *In Search of Lost Time* was one of the first modernist novels to be published and it embodies uniquely modernist formal and stylistic elements that departed significantly from the novelistic tradition. Proust's novel was also originally written in French and well-known for its different translations, especially the first translation (completed by C. K. Scott Moncrieff and [criticized](#) for being "inaccurate" and "overinterpretive").

My research uses a side-by-side NLP analysis of both the French and English texts of Marcel Proust's *In Search of Lost Time* to assess the value of applying AI to modernist and translated works. I have elected to use what is commonly considered the "definitive" French edition and an English translation that revised Moncrieff's to match this definitive French edition. Conducting a side-by-side NLP analysis has already raised challenges not often seen in the digital humanities. The two texts needed to be cleaned and processed using different algorithms that account for differences between French and English, and common NLP tools — including dictionaries of words, stopwords, and stemming algorithms — vary from language to language, with no easy comparison. The next steps in my research are to examine whether the results of similar but independent NLP analysis on the French and English texts (including analysis of keywords, tenses, and narrative structures) will yield similar and comparable results. As I continue my research, I also hope to identify whether key elements of the modernist novel, including disruptions of time, can be identified via AI and the digital humanities. In conclusion, this project, via a close study of Proust's *In Search of Lost Time*, aims to shed new light on the limits and potential of digital humanities applied to modernist and translated fiction.

Biography

Cindy Zhang is a senior at Harvard University studying English and Computer Science. She is interested in modernism and aestheticism, legal and ethical questions on privacy and technology, and the digital humanities. At Harvard, Cindy is a Teaching Fellow for a course on privacy and technology. She was previously on the teaching staff for computer science courses at the Harvard Law School and Harvard Kennedy School.

Project advised by Felice Cowl Reid Professor of English Philip Fisher, Gordon McKay Professor of the Practice of Computer Science Jim Waldo, and English PhD Candidate Holst Katsma.