

WordSeer

An interface for exploring language use in slave narratives

Digital Text

More and more source text in the humanities gets digitized every day, making it accessible to large scale computational analysis. By contrast, traditional methods of humanistic analysis are based on detailed arguments built upon on close readings of individual texts. How will the field adapt? How do we use statistics and text mining to answer humanistic questions?

Slave Narratives

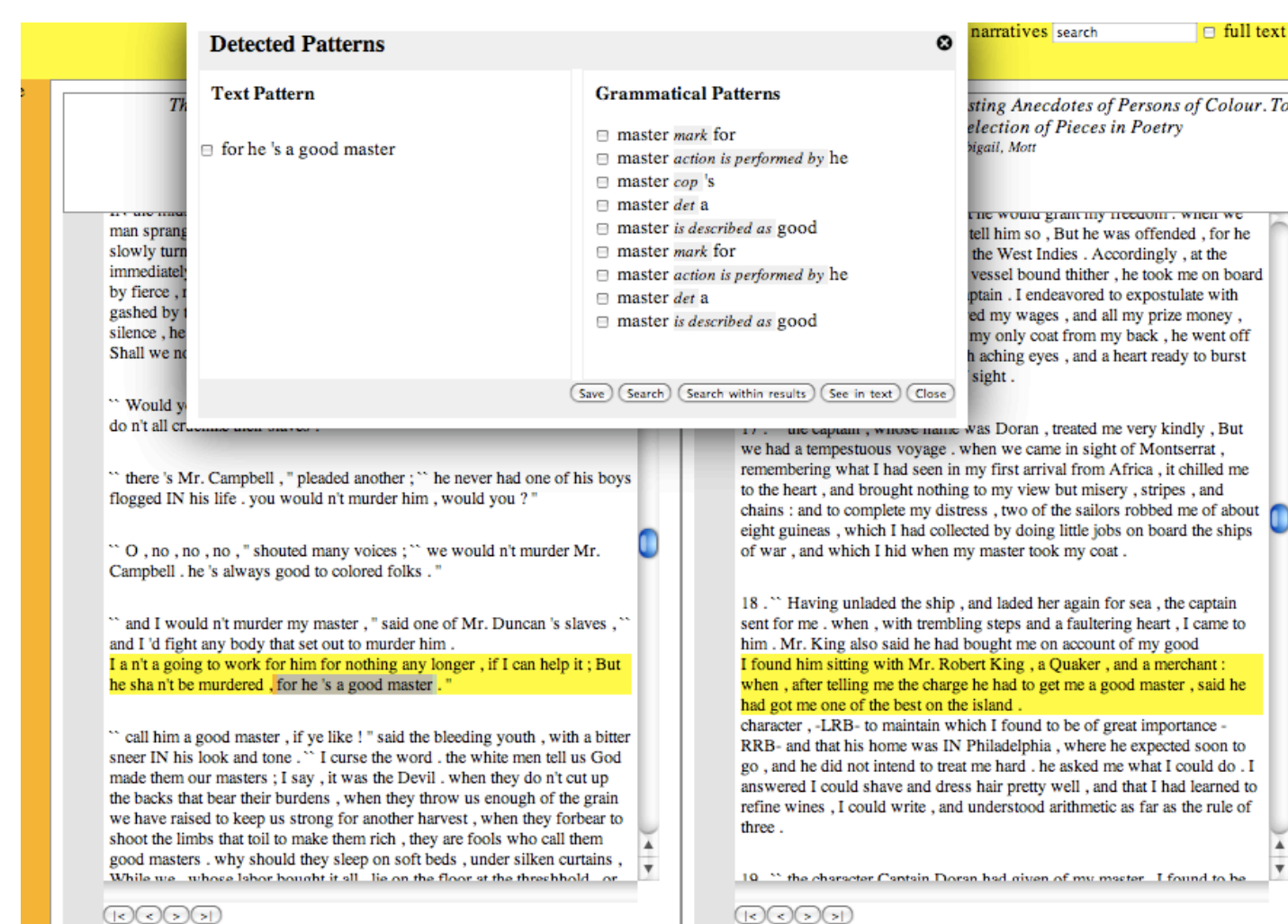
Zoom in to the field of American literature, and further into the realm of studying the (digitized) narratives of escaped former slaves, published by white abolitionists. There are widespread stylistic and thematic similarities among these narratives. How can text mining help literature scholars here?

Related Projects

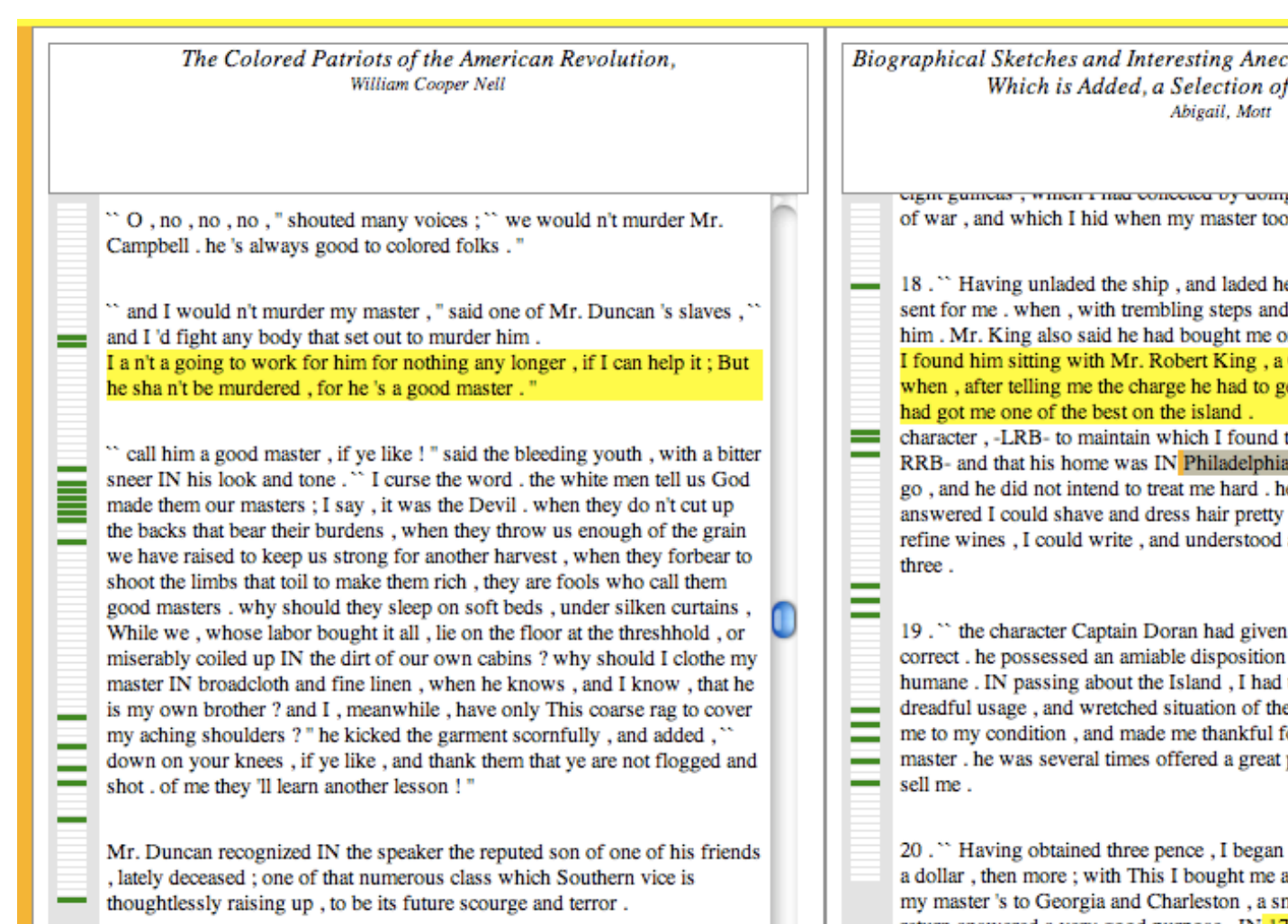
The MONK project at CMU, and the Voyeur project at McMaster University share the same cause as WordSeer. But, when it comes to text analysis, they are essentially search interfaces that show simple statistics about word order, type and frequency. The grammatical relationships within text are neglected.

The Interface

Interviews suggested that a search interface alone would not be enough, so WordSeer supports searching through texts on a large scale and reading them individually.



Reading view. Users can select multiple search results and view them in the narrative reader for closer inspection



Text search

Search narratives full text
search within selected:
 in the same narratives
 in the same sentences

Grammatical search

(any relation to)
is described as
action(s) are performed by
action(s) are performed upon
and/or/but
is possessed by
cruel good kind gene
master masters mistr
search within selected:
 in the same narratives
 in the same sentences

Standard search for keywords or exact matches in the metadata or in the text.

Grammatical search. Using grammatical relationships extracted through natural language processing, users can ask how things were described, what actions were performed upon them and by them, who possessed certain things, or what was possessed by them.

Graphs (right) show the frequencies of different words that match, and can filter results.



Language Processing

Part-of-speech tagging, syntactic parsing, and dependency parsing decompose sentences into their grammatical constituents. For example, the sentence, "The cruel man beat us severely" contains the word "cruel" which is an *adjective modifier* of the word "man", which is a noun. There is *verb object* relation between "beat" and "us", and a *verb subject* relation between "man" and "beat".