Joel Goldschlag

Joyce & Interdisciplinary Modernism

Greg Erickson

05/15/2016

**The Ulysses Text Interpreter**

For my creative project, I wrote a generalizable program that can take a text and roughly sort it chapter by chapter into parts of speech. To do so, I downloaded a part of speech dictionary from the Moby Project that contains ~250K words followed by some strange character, and then letters which designate part of speech. For example, the entry “abate◊Vti” would indicate the word “abate” can be a participle, transitive, or intransitive verb. Using this library, I was able to sort the words in each chapter of Ulysses (which I downloaded in plaintext from project Guttenberg—those guys are amazing for this sort of project!) in to lists of words, aggregated by part of speech. I then wrote a function that took those lists, which had some words repeated 10’s or 100’s of times, and condensed them down to a shorter list of “tuples” which stored a word alongside how many times it appeared, in one list element. Doing so, I was able to then make comparisons of word use in the chapters that would otherwise have been impossible. For example, there were 40% more pronouns used proportionally in chapter 18 than chapter 1. More interestingly, the pronoun “I” made up for 16% of all pronoun usage in chapter 18, 13.5% of pronoun usage in chapter 15, and 10% of pronoun usage in chapter 1, and just 4% of chapter 4. Comparing usage of a word like “I” in chapters mostly in Stephen’s or Bloom’s or Molly’s voices can give some insight into these characters. I think its particularly interesting that Molly uses the word “I” at 4X the rate of Bloom (though this is no doubt effected by one being a monologue – enough of the book consists of internal monologues that I feel such comparisons are still important).

However, there were some important limitations to my project. Natural language processing is very difficult for computers. One can both book it to the library and check a book from the library, and there is no simple way to tell a program to differentiate the two. Subsequently, if the Moby POS library I downloaded had multiple possible parts of speech attached to a word, when that word came up in Ulysses it would be added to all part of speech lists that were designated by the Moby Library. Additionally, there are many words in Ulysses that are spelled in a way (or entirely made up) such that they do not exist in my reference libraries. These words were not added to any lists, and therefore I felt most comfortable making arguments about usage of words like pronouns that would more likely than not been caught and sorted with high precision.

In keeping with these pronoun readings of the text, I was able to make some fun broad claims about the text (I was kind of inspired to apply Jacob’s ridiculously broad reading to Sofie’s obscenely close one). For example, though chapters 1 and 4 used roughly the consistent ratios of pronouns/word (12% vs 13% -- finally proof[[1]](#footnote-1) that Bloom and Stephen are father and son!) chapter 15 was 9% pronouns (which makes sense given the chapter’s proclivity to name people in ever-changing hilarious ways that do not use formal pronouns) and chapter 18 18% pronouns (how fitting!).

For future analysis I would love to use professional natural language processing (such as NLTK), as opposed to one I made myself to try and see if I can make this more intuitive. I could also write functions that look for suffixes that may betray a part of speech for made up words, as I already have a list assembled of words not found in my custom dictionary.[[2]](#footnote-2) This was a ridiculously fun project that gave me a new way of looking at the text!

1. Read: not at all proof [↑](#footnote-ref-1)
2. read: all the fun Joycian words [↑](#footnote-ref-2)