Preparatory Assignment: *Moby Dick vs Jane Austen Homework Assignment*

***Use data file “Melville\_vs\_Austen”***

* Post information about method (how did you scrub the data, what strings did you search with, how big was the search window, etc.)
* Post screen shots of output graphs.
* Class Discussion about findings:
	+ What students searched for, and why?
	+ Are there any common patterns across our searches?
	+ What are the advantages and disadvantages of an approach like this to gain a better understanding of these texts?

*Language Analysis Project*: Test hypotheses about language using your ***own*** data

***Come up with a personally interesting question about your language, and learn how to:***

* …curate a data set that can help you to answer that question
* …approach the analysis of the data set with feasible hypotheses/predictions about the question you are testing (vs. unbiased fishing)
* …think about your findings, draw conclusions, and integrate your result with the published literature

***Ideas for research questions:***

1. Differences in language that result from differences between speakers
* Men vs women
* Developmental differences across the lifespan
* Individual variability
* Developmental differences across college (first year vs seniors)
* bilinguals vs monolinguals
* differences in emotional state
	+ before/after study abroad
	+ before/after romantic relationship
	+ before/after illness or injury
1. Differences in Language that result from differences in the audience (purpose of communication/expectations/context)
* e-mails to teachers vs internship supervisors vs coaches vs employers
* different professors in one department
* written work across disciplines?
	+ Analytical vs. emotional language
	+ Formal vs casual language
* text to one person vs group text
* spoken vs written communication
* emoji use based on audience

***Analysis Tools: Lexos***

* Upload data, in manage tab, group text according to class (what groups are you comparing?)
* Analyze Menu: Top Words & Similarity Query
* Top words that are statistically different between groups (classes)
	+ Z-scores: If A to B and score is highly negative, much less common in A than B
* Similarity Query
* Visualize Menu: Rolling Windows & Word Cloud
* Rolling Windows: Use top words to determine where in the text the differences lie.
	+ Want 10-100 windows – maybe 50 word window for 800 word file?
	+ Where in the text file the words appear -- this will matter if the texts are strung together with a time component (could milestone by month?)
	+ What is the density with which words appear, relative to each other?
* Word Cloud

***Rubric for Powerpoint Presentation:***

* Slide 1: Introduce the research question
	+ Why did you choose this question?
	+ What do you expect to find and why? (the “why” here is most effective when it relies on the published literature as well as your personal insights and suspicions about how you communicate)
* Slide 2: Data set
	+ What data did you collect for this comparison (provide some numbers – 100 texts in each comparison group, 50 characters on average, sent between date 1 and date 2)
	+ Explain what you decided to include and exclude from the data set, and why?
	+ How did you scrub the data before analysis?
* Slide 3: Analysis approaches
	+ Describe and rationalize the ways you analyzed your data set
* Slide 4: Results
	+ Show visuals of your results (graphs, word clouds, lists of top words, etc)
	+ Can you quantify your visuals in order to describe what you see, and do statistics to compare groups?
* Slide 5: Conclusions
	+ Revisit the research question and present the conclusions of your analysis
* Slide 6: Implications and future research
	+ Use the findings of a published research study to contextualize your findings.
	+ Describe the questions these results have raised for you and what you would do next if you were continuing to work on this project.

***Complete Student Survey to provide feedback about this assignment:*** <https://www.surveymonkey.co.uk/r/dlapilot>