Using the Command Line

And Why I Should Care

What Is The Command Line

- Graphical windows and user interfaces are a visual method of interacting and giving instructions for the computer to do something.
- Before computers were powerful enough to waste computation power on a graphical user interface, interactions took place by typing commands.
- You already know how to do this - Google for example

How Do I Get To The Command Line?

- On your Macintosh, open the Applications folder, then open the Utilities folder, then find the application terminal
- On your Windows machine, launch Cygwin
- Type your first commands:
  - pwd - Tell us what you see.
  - ls - what do you see? Can you find the same directory on your desktop
  - cd - navigate into a folder/directory on the command line

Why Do I Care?

- The promise of digital humanities lies in being able to manipulate texts programatically.
- To write programs or scripts that interact with texts, you need to begin to think programmatically rather than visually so you can tell the computer to do things in sequence.

A Few Other Commands

- Download our sample file(s) for the day: Hamlet and the lexicon from Professor Dwyer
- Most commands follow the same basic pattern:
  - COMMAND --OPTIONS FILENAME TO OPERATE ON
- Can you navigate to the directory where they are stored?
- How do you look at them quickly?
  - less
- Find out how many words are in a text
  - wc
- Find out how to use a command
  - man
- How to you copy and move them around
  - cp and mv

Basic Analysis

- How many words are in my file:
  - wc
  - Navigate to our texts directory and tell my how many words are in ham.xml
  - use man to find out how to have wc tell you how many lines and characters are in ham.xml

Find Patterns In A File

- GREP and Regular Expressions
- grep is a command line tool that allows you to search for patterns in a file
- try 'grep ghost ham.xml' (without the quotation marks'
- show a little more context with -A 2 (and note that this is different from -a 2 - commands are case sensitive)

Put It Together

- The | character is called a 'pipe' and it allows you to send the output of one command into another.
- Use grep ghost ham.xml | wc -l to find out how many lines in Hamlet contain the word ghost.
- The > character at the end of a command redirects the output from the screen to a file
- Try grep ghost ham.xml | wc -l > ham.ghost.txt
- How would you look at the contents of this file?

Grep 'wildcards' and Regular Expressions

- . stands for any single character. Grep ".host" ham.xml - what do you see?
  - Try grep "s.n" ham.xml
- * stands for a repetition -Try grep "s.*n" ham.xml
- [ ] lets you define a class of characters such as [1-9]
- ? lets you search for an optional character,

Convert mjg-se_Lex to XML

- Open the file in Excel (or any other spreadsheet), save as Unicode text
- In the Oxygen find and replace panel, search for (.+?)\t(.+?)\t(.+?)\t(.+?)\t(.+?)\t(.+?)\t(.+?)\n
  - Do you understand what this is?
- Replace with <lang>$1</lang><lex_o>$2</lex_o><lex_o>$3</lex_o><POS>$4</POS><gl_eng>$5</gl_eng><etym>$6</etym><source>$7</source>\n