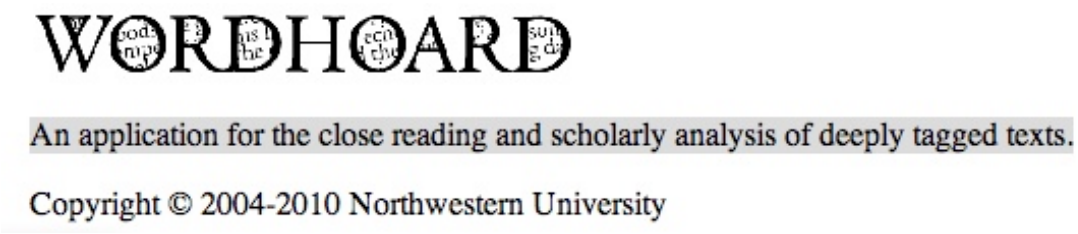
http://wordhoard.northwestern.edu/userman/index.html



WordHoard is an extremely powerful and easy to use application for searching and investigating Shakespeare's vocabulary. Of all the applications I introduce you to on this class, it is the most carefully set-up, and produces the most interesting findings.

It already contains a fully tagged corpus of Shakespeare (and Spenser), so you simply open it and run it - no extra texts required. The downside of this is that you cannot use it to analyse your own corpora.

To use WordHoard, go to its home page:

http://wordhoard.northwestern.edu/userman/index.html

This page links you straight to a series of excellent brief introductions to aspects of both WordHoard, and text analysis: I recommend you to read through them just to get a sense of what the application can do (don't worry too much about things you don't understand):

Table of Contents

- Preface
- Understanding WordHoard
  - What is WordHoard?
  - Metadata and the Query Potential of the Digital Surrogate
  - Working with Very Common and Very Rare Words
  - The Corpora and Tagging Data
- A Hands-On Introduction to WordHoard
  - Getting Started
  - The Basics
- Reading
  - The Table of Contents Window
  - Getting Information about Works
  - Displaying and Reading Works
  - Getting Information about Words
  - Lexicons
  - Getting Information about Lemmas
  - Parts of Speech and Word Classes
  - Translations, Transliterations and Transcriptions
  - The Iliad Scholia and E. K. Annotations
- Searching
  - Searching for Words
  - Concordances
  - Searching for Lemmas
  - Searching for Works

For this week's task you'll need to read at least 'Getting Started' before you download and run WordHoard - please do that now.

Task.

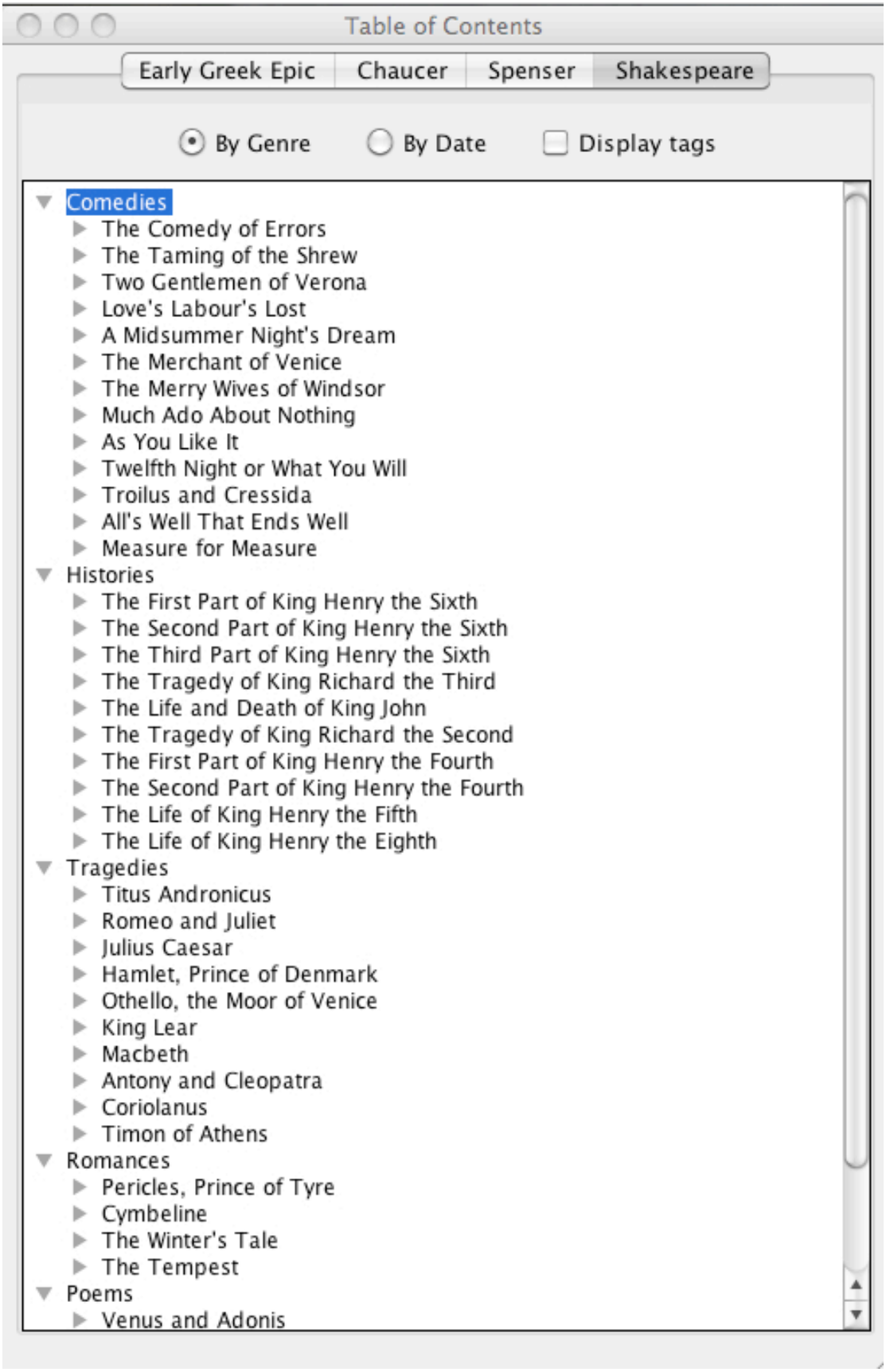
Most concordance programs can give you simple lists of frequency, and things like *TTR* - but in terms of finding anything interesting out, these are not much use. Generally, computer and statistical analysis only starts to produce interesting results when you compare populations. A *population* is a group of observations which are linked in some way - so in our class, a population might be word frequencies in the complete Shakespeare corpus, or in just one play.

It is rarely interesting, for example, just to know what the most frequent word in a play is - for one thing, it will usually be a fairly predictable *function* word (see the discussion of this in the WordHoard materials). Even the most frequent *lexical* words are usually pretty obvious, and simply tell us things we already know about the topic of a play or scene.

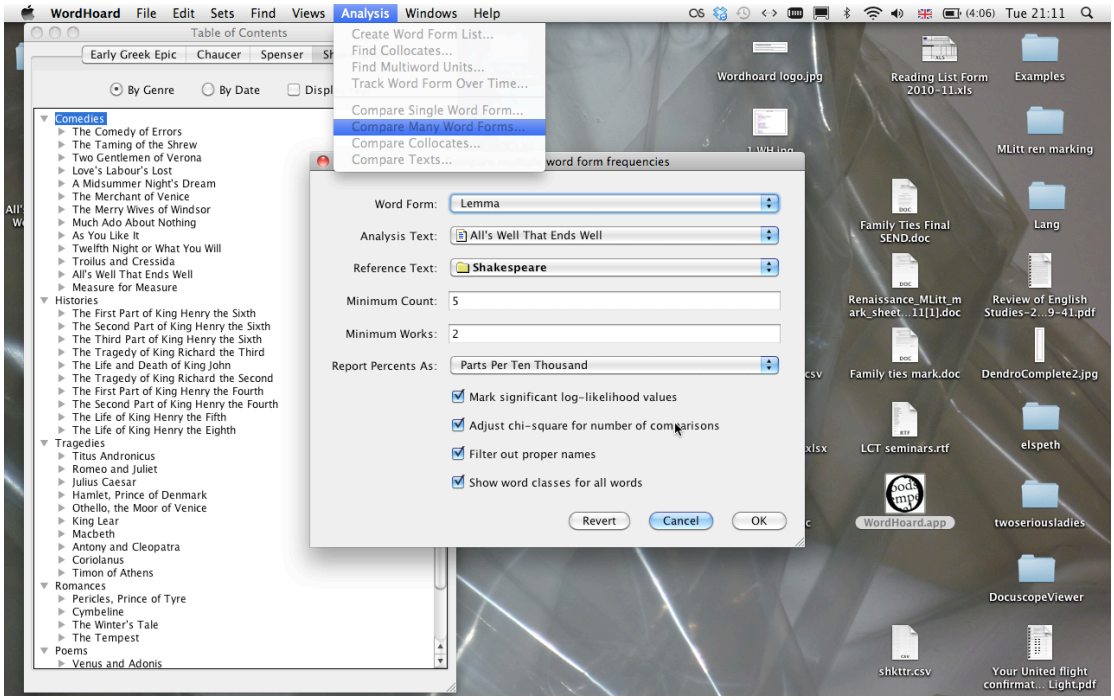
More interesting things start to happen when you compare the vocabulary of a single play with that of the corpus as a whole, effectively asking the questions, 'which words does Shakespeare use more frequently than normal in this play?', and 'which does he avoid compared to his usual practice?'.

WordHoard has a very handy command for finding this out in your assigned play - and the findings will give you a basis for further investigation of the vocabulary of your play.

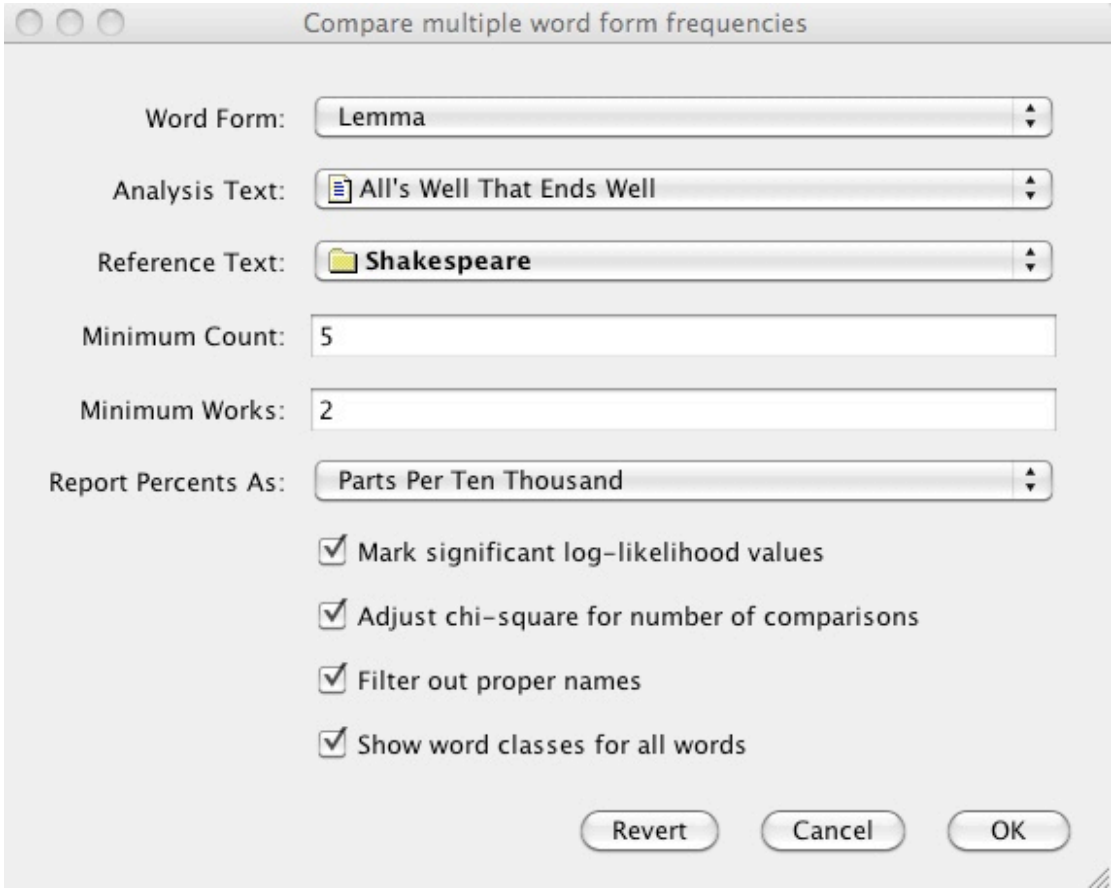
Download WordHoard and, if necessary, open the program. You should see this box (select 'Shakespeare' if you need to):



Now select the 'Analysis' drop-down menu at the top of your window, and 'Compare many word forms' from within that to produce the following screen:



You now need to select certain items in the dialogue box, which should look something like this:



First, under ‘Word Form’, you need to decide if you want to search for ‘spellings’ or ‘lemmas’.

‘Spelling’ will count all occurrences of the same letters as the ‘same’ word - so ‘ring’ (noun - ‘She bought a gold ring’) will be counted in with ‘ring’ (verb - ‘Ring the bell if you want to enter’).

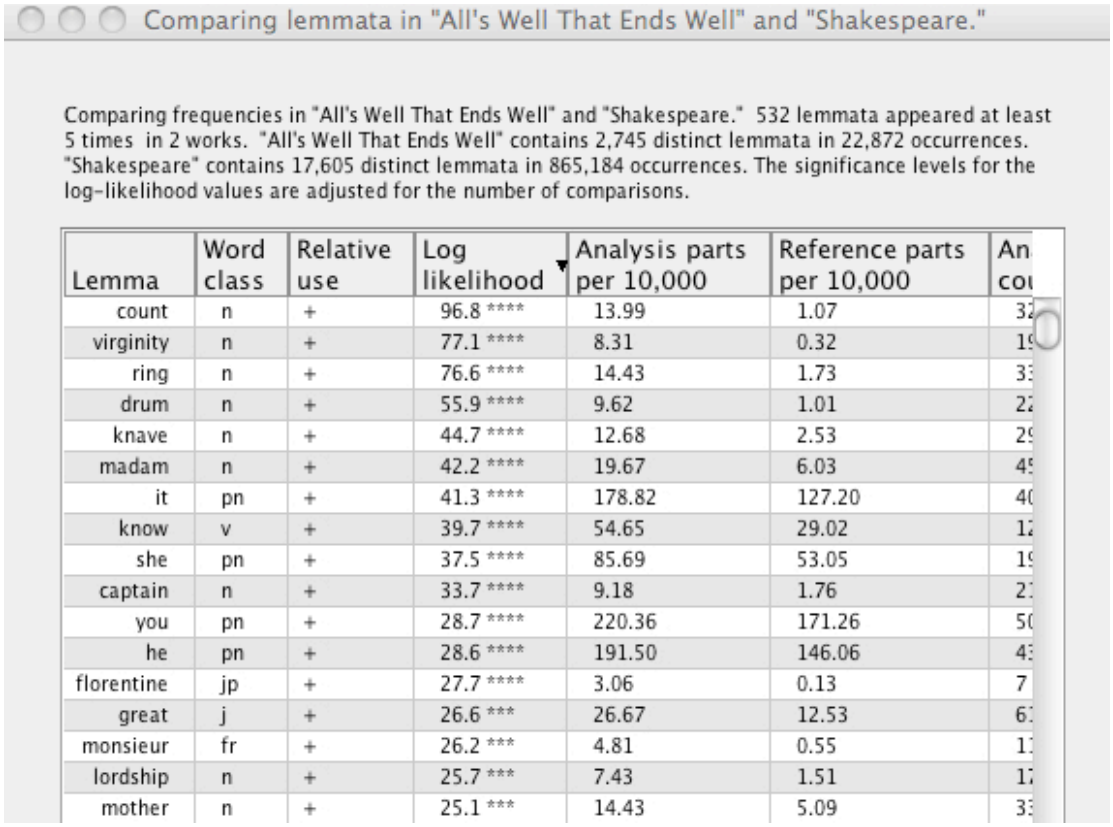
‘Lemma’ uses the grammatical information coded within Wordhoard to distinguish words that are spelled the same, but which have different grammatical functions.

Sometimes this can make a difference to your results, so I would recommend doing your search twice: once on ‘Lemma’, once on ‘Spelling’. ‘Lemma’ results are the most reliable though.

Next, you need to select your ‘Analysis text’. This is the text you are interested in – we suggest you use the text your analysis group has chosen. Do this using the drop-down menu.

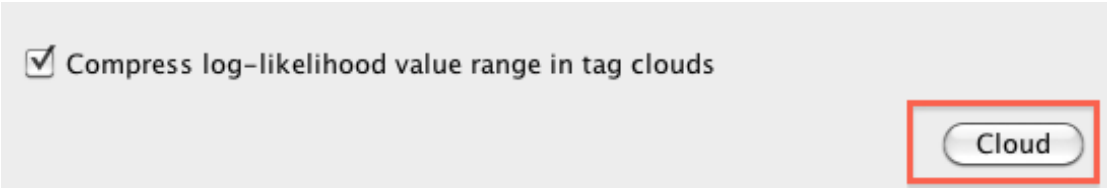
Next you need to select your ‘Reference text’. This is the comparison sample you want to test your text against. You want to find out what your text does that is unusual in relation to the rest of Shakespeare, so for this text you should chose the folder containing all of Shakespeare’s texts (as in the screenshot above).

Now fill in the rest of the boxes as shown above, and hit ‘OK’. You should see the program run the analysis, and report back with a box like this:



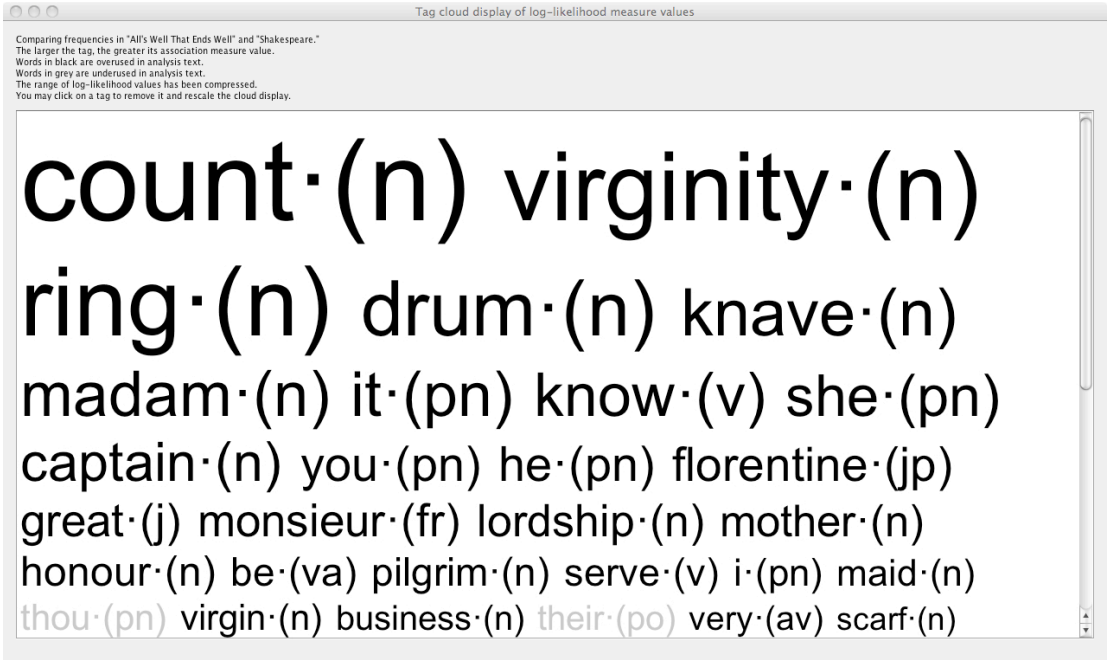
Lemma	Word class	Relative use	Log likelihood	Analysis parts per 10,000	Reference parts per 10,000	An cou
count	n	+	96.8 ****	13.99	1.07	32
virginity	n	+	77.1 ****	8.31	0.32	19
ring	n	+	76.6 ****	14.43	1.73	33
drum	n	+	55.9 ****	9.62	1.01	22
knave	n	+	44.7 ****	12.68	2.53	29
madam	n	+	42.2 ****	19.67	6.03	45
it	pn	+	41.3 ****	178.82	127.20	40
know	v	+	39.7 ****	54.65	29.02	12
she	pn	+	37.5 ****	85.69	53.05	19
captain	n	+	33.7 ****	9.18	1.76	21
you	pn	+	28.7 ****	220.36	171.26	50
he	pn	+	28.6 ****	191.50	146.06	43
florentine	jp	+	27.7 ****	3.06	0.13	7
great	j	+	26.6 ***	26.67	12.53	61
monsieur	fr	+	26.2 ***	4.81	0.55	11
lordship	n	+	25.7 ***	7.43	1.51	17
mother	n	+	25.1 ***	14.43	5.09	33

At the bottom right of the box you should see this button:



Click it.

You should see something like this:



This is a visual representation of your results, which you should use in combination with the initial table to decide which words you want to select for further study.

Reading your results.

Go back to the Table you generated above, and note what each column tells you:

Column 1: the lemma being counted

Column 2: the word class of the lemma (for codes, see WordHoard Help)

Column 3: the 'relative use' column tells you whether the lemma is more ('+') or less ('-') frequent in your play than would be the norm for Shakespeare as a whole. Bear in mind that *absence* is just as important as *presence*. Words that

Shakespeare avoids in a play may be more interesting than those he over-uses (and electronic search tools are just about the only way to discover absences like this).

Column 4: this is the key column - the 'log likelihood' is a measure of the extent to which the frequency of the lemma in this population is similar to that in the reference population. In other words: does this lemma occur in your play strikingly more or less than it does in the Shakespeare corpus as a whole?

For this example, the table is telling us that the noun 'count' is hugely more frequent in *All's Well* than in the rest of Shakespeare (the number of asterixes tells you how significant the result is - the more asterixes, the more significant the result). The table is sorted by this column: the most untypical words, when compared to the Shakespeare corpus, appear first.

Column 5: this gives you a frequency count of the lemma in your 'analysis' text regularised against 10,000 words - so here, we are being told that the lemma 'count' (noun) appears in *All's Well* (the 'Analysis') 13.99 times every 10,000 words. Compare this with column 6:

Column 6: this gives you a frequency count of the lemma in your 'reference' text regularised against 10,000 words - so here, we are being told that the lemma 'count' (noun) appears in the whole of Shakespeare (the 'Reference') 1.07 times every 10,000 words.

Column 7: this gives you a raw total for the number of times your lemma appears in the analysis play.

Column 8: this gives you a raw total (N) for the number of times your lemma appears in the 'Reference' sample - the whole of Shakespeare (remember that this includes your analysis text).

Word Cloud:

The bigger the word, the more shifted its frequency is away from the Shakespearean norm.

Words in black are *raised* in frequency.

Words in grey are *lowered*.

For more on counting words, and WordHoard, read:

Jonathan Hope and Michael Witmore, "The Language of *Macbeth*", chapter in *Macbeth: The State of Play*, edited by Ann Thompson, London, Bloomsbury (Arden), 2014, pp. 183-208 (you can download a version of this from Myplace)

Make sure you understand the paper, and how to use Wordhoard.