**These instructions were used at the beginning of the project to guide research assistants in the reformatting of Shakespeare's texts into comma-separated words, stripped of character names, stage directions, and punctuation. For the most part, we sourced texts from the electronic *Riverside*, but for details regarding a specific text's provenance please consult the "Text Provenance and Parameters" file in the "misc-files" folder. It is not necessary for anyone wishing to use our code to follow these directions–they serve as documentation of our research process.**

Text Formatting Instructions

These are the instructions for assembling a file of Shakespeare’s vocabulary by saving *only* the dialogue words in each text, deleting the apparatus he did not certainly write: titles, dramatis personae (= cast list), stage directions, speech prefixes, etc.

1. Open text. Begin scrolling through the play. Highlight and delete the dramatis personae (cast list) at the beginning of the play before Act I, Scene I, feeling free to scroll through the opening act and scene designation, and stage directions, cutting those out too.
2. CTRL/CMMND F: “SCENE” (keep all letters in caps), and cut the words “SCENE,” “ACT,” the (roman) numerals, and any stage directions immediately following the introduction of a new scene. Thus, for example:

SCENE IV

Flourish. Enter KING [DUNCAN], LENNOX, MALCOLM,

DONALBAIN, and ATTENDANTS.

Some plays have as few as nine or ten scenes, some over thirty. You’ll want to do these “by hand” rather than a global search and replace, to make sure you’re getting only the apparatus.

1. Next, replace the line numbers in the right margin. **To access Advanced Find and Replace**, on the Home tab, hit the Editing group. Select Replace. Replace every number from 0 to 9 with a blank space. For example, to replace all 1’s in a text, type “1” into the Find box and insert a space in the Replace box. It is crucial that a space is inserted in the Replace box because if the space is omitted, it will lead to the formation of unwanted and erroneous compound words. [Remember: the window may retain the space inserted here, as well as any other commands—so always check]
2. To get rid of the [….] left behind, highlight, cut and paste one of them in the Advanced Find and Replace pan, inserting a space in the Replace box. Do a global find and replace, and survey the result. Remember that you can always undo a procedure if it goes haywire.
3. To eliminate character names and other stage directions, access Advanced Find and Replace. Select the “More >>” function of a simplified Find and Replace browser pops up. Click the “Find what” box with the cursor. At the bottom left hand corner under the header “Find”, hit the “Format” drop-down menu and select “Font.” Under the option of “Font Style,” select “Italic.” Click “OK.” Hit the “Replace” tab at the top of the Find and Replace browser. In the “Replace with” box, insert a space. [remember that your search window may retain the commands you’ve given it before]. After eliminating the italics, it is critical that the next time Access Advanced Find and Replace is performed that the “No Formatting” option in the bottom middle of the Find and Replace screen is selected.
4. Access Advanced Find and Replace. Click “No Formatting” to reset the Advanced Find and Replace font settings. Replace [ ] (bracket space bracket) with an inserted space in the “Replace with” box.
5. Access Advanced Find and Replace. Replace all **small capitals** by selecting the “Format” drop-down menu located at the bottom left of the screen under the header of “Replace.” (Please note: “small capitals” are a special kind of font employed for stage directions, etc., in this text). Select “Font,” and check the box next to “small caps.” Select “OK.”
6. Access Advanced Find and Replace. Replace all punctuation separately (Ex: , . : ; ? ! ( ) “ ) **with an inserted space** in the “Replace with” box. A list:

comma = ,

period = .

colon = :

semi-colon = ;

question mark = ?

exclamation mark = !

open parens = (

close parens = )

open bracket = [

close bracket = ]

quotation mark = “

Please note: Do **NOT** remove any apostrophes, as these are part of a word = ‘tis, nurse’s (etc.)

1. Scroll through your text until you find an “em dash”. This is a longer dash rather than a hyphen. Thus: “I cannot tell—“ or “To make them kings—the seeds of Banquo kings”. Select one of these em dashes, and replace them all with an inserted space.
2. To verify the word count of the text, select the “’Review” tab at the top of the menu for Word. Select “Word Count.” Verify that this number matches the established word count of the text (see file). If the number is inaccurate, double-check that all of the above steps have been completed. If the word count is still off, consider carriage returns and access Advanced Find and Replace. Type in “-” into the Find box. Manually go through the text, and evaluate whether the hyphens included are stylistically Shakespearean or if they were included in the text for formatting purposes (these will usually be found at the end of a line). For example, in cases of obvious nouns and verbs such as “to-morrow,” the hyphen will be kept because in these cases, the hyphen is used in a manner that is stylistically Shakespearean. The hyphen will also be kept in cases of descriptive adjectives “new-fangled” and “hot-blooded

Here are some examples of compound words in Shakespeare (which we will not touch, obviously):

Might be the **be-all** and the **end-all** here

**even-handed** justice

**trumpet-tongu’d**

Here are some examples of hyphens to remove because they split up a word purely for the purpose of a line break (that is, they were introduced by a modern editor):

swear in both the scales against either scale who **com-**

**mitted** treason enough for God’s sake [where “committed” should be one word]

What three things does drink especially **pro-**

**voke**

Marry sir nose-painting sleep and urine

Lechery sir it provokes and unprovokes it provokes

the desire but it takes away the performance **There-**

**fore** much drink may be said to be an equivocator

with lechery it makes him and it mars him it sets him

on and it takes him off it persuades him and **dis-**

**heartens** [the **bolded** words here should be restored by removing the hyphens]

For the next stage of this data-assembly, we'll perform a word count, comparing it with the official concordance count that was recorded for each play when the *Riverside* was originally printed. If there's a discrepancy (and there usually is), we'll find out where we're in error before transforming the still lineated file we've got into something that the computer can read:

**when, shall, we, three, meet, again, in, thunder, lightning, or, in, rain, when, the, hurly-burly's, done, when, the, battle's, lost, and, won, that, will, be, ere, the, set, of, sun, where, the, place, upon, the, heath, there, to, meet, with, macbeth, i, come, graymalkin, paddock, calls, anon, fair, is, foul, and, foul, is, fair, hover, through, the, fog, and, filthy, air, what, bloody, man, is, that, he, can, report, as, seemeth, by, his, plight, of, the, revolt, the, newest, state, this, is, the, sergeant, who, like, a, good, and, hardy, soldier, fought, 'gainst, my, captivity, hail, brave, friend, say, to, the, king, the, knowledge, of, the, broil, as, thou, didst, leave, it, doubtful, it, stood, as, two, spent, swimmers, that, do, cling, together, and, choke, their, art, the, merciless, macdonwald, worthy, to, be, a, rebel, for, to, that, the, multiplying, villainies, of, nature, do, swarm, upon, him, from, the, western, isles, of, kerns, and, gallowglasses, is, supplied, and, fortune, on, his, damned, quarrel, smiling,**

When the text is checked and corrected against the *Riverside* concordance/source text, save it as a .txt file.