

Renaissance and Reformation Renaissance et Réforme



Anthony, Laurence, project dir. AntConc (Version 3.5.8).
Software

Scott, Michael, project dir. WordSmith Tools (Version 8).
Software

Emily Louisa Smith

Volume 44, numéro 1, hiver 2021

URI : <https://id.erudit.org/iderudit/1081152ar>

DOI : <https://doi.org/10.33137/rr.v44i1.37062>

[Aller au sommaire du numéro](#)

Éditeur(s)

Iter Press

ISSN

0034-429X (imprimé)

2293-7374 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Smith, E. (2021). Compte rendu de [Anthony, Laurence, project dir. AntConc (Version 3.5.8). Software / Scott, Michael, project dir. WordSmith Tools (Version 8). Software]. *Renaissance and Reformation / Renaissance et Réforme*, 44(1), 200–214. <https://doi.org/10.33137/rr.v44i1.37062>

better resource for the study of eighteenth-century English performance and social life.

RENAE SATTERLEY

The Honourable Society of the Middle Temple

<https://doi.org/10.33137/rr.v44i1.37061>

Anthony, Laurence, project dir.

AntConc (Version 3.5.8). Software.

Tokyo: Waseda University, 2019. Accessed 30 August 2020.

laurenceanthony.net/software.

Scott, Michael, project dir.

WordSmith Tools (Version 8). Software.

Stroud: Lexical Analysis Software, 2020. Accessed 5 September 2020.

lexically.net/wordsmith.

Introduction

Concordancing is far from novel: our early modern counterparts would have been very familiar with the concept of verbal indexes, as the popularity of biblical concordances indicates. Nevertheless, both the usability and potential applications of concordances have evidently developed enormously with their digitization. In digital formats, the speed and convenience of concordance consultation is naturally improved, meaning that much larger textual corpora become viable. Perhaps most revolutionary, however, is that concordancing software enables users to investigate and visualize their text in qualitative terms—for instance, by calculating the relative frequency of a word within a text, or a word's dispersion throughout a corpus.

AntConc and WordSmith Tools are two programs that enable users to create such enhanced concordances. AntConc, developed by Laurence Anthony, is available to download on Windows, Macintosh, and Linux; WordSmith Tools, developed by Michael Scott, is available for Windows. As corpus analysis programs, it is unsurprising that they both appear to be implicitly geared towards linguists; as the index of “Research Using WordSmith” illustrates, the tool has

primarily been used for this purpose, while Anthony, the developer of AntConc, is a linguist by training. Although neither was specifically developed for use with early modern texts, their functions do nevertheless facilitate this usage—particularly given the increasing availability of digitized texts and tools for pre-processing tailored to the period. Furthermore, and perhaps most importantly, both tools are highly accessible even to those without prior experience; neither requires programming skills at all, and both are intuitive enough to allow exploration of corpuses without extensive preparation or training. That said, however, to use the tools in an informed manner it is essential to have a basic comprehension of what the statistical operations measure, and how they do so, in addition to having an awareness of the factors that affect concordances (such as lemmatization and spelling variations).

AntConc is free to download, while a single-user licence for WordSmith 8.0 costs £50 (site licenses are also available to organizations at an incrementally cheaper cost per additional user). AntConc does not state any specific system requirements, while WordSmith only suggests that it will be “happiest on a fairly modern laptop” (“System Requirements,” lexically.net/wordsmith). From personal experience, both run without any issues on a 2.5GHz Windows 10 laptop.

Guides and tutorials

Anthony’s website provides links to multiple tutorials for AntConc, both written and on YouTube, in multiple languages (currently English, Urdu/Hindi, Japanese, Arabic, Chinese, German, Korean, Portuguese, and Spanish); it advertises an active Google Groups discussion space and links to the tool’s citation index to view projects that have used the software. While not explicitly mentioned by Anthony, AntConc has inspired multiple other online guides, including a hands-on directed tutorial by the project *The Programming Historian* (available in English, French, and Spanish).¹

WordSmith Tools hosts an extensive online manual on its website, including three videos that introduce and demonstrate the main functions

1. Heather Froehlich, “Corpus Analysis for AntConc,” *The Programming Historian* (2015, updated 2020), programminghistorian.org/en/lessons/corpus-analysis-with-antconc. The *Programming Historian* offers tutorials to assist in the use of digital tools within academic research—aimed especially at those without prior training in computational research.

(Concord, WordList, and Keywords), and a guide to processing the British National Corpus. While Version 8 currently has guides in English, Catalan, Chinese, and Vietnamese, Version 7 also has manuals available in Albanian, Arabic, French, Farsi, German, Italian, Portuguese, Spanish, and Ukrainian. Again, there is an active Google Groups discussion space for WordSmith Tools, which Scott himself frequents, and a database of publications that have used the software.²

Both tools are intuitive enough to be attempted without following a tutorial, provided that you have already prepared (or downloaded) a corpus for analysis. In all cases, however, a word of warning is necessary: digital processing carries some limitations for any early modern texts, as has been noted countless times. In creating concordances, the two main technical issues faced are inaccurate OCR transcriptions (OCR stands for “optical character recognition,” the process whereby text is extracted from a physical document into a machine-encoded, computer-readable electronic form; this technique often struggles with early modern print, thus producing errors) and the lack of standardized spelling; a good analysis, of course, requires acute alertness to historical linguistic variations. It is prudent to consider how you will manage these issues before beginning a concordance to ensure accuracy of results. The results you will get out of a concordance, in other words, will only be as good as the files you put in.

Getting started with WordSmith Tools

Presuming you have already validated a license, WordSmith Tools opens by providing three large icons for access to the functions Concord, KeyWords, and WordList (fig. 1). From the launch window, it is also possible to directly launch Aligner (enabling the comparison of two or more texts by sentence or paragraph), Character Profiler, Chagrams, Corpus Checker, File Utilities, File Viewer, Languages and Fonts, Minimal Pairs, Registration, Text Converter, WordSmith Version Checker, and WSCongram, while the left-hand column, below all of the utility icons, enables quick access to the settings for each of these features. Clicking on any of these utilities will open a new window.

2. “Support,” WordSmith Tools, lexically.net/wordsmith/support.

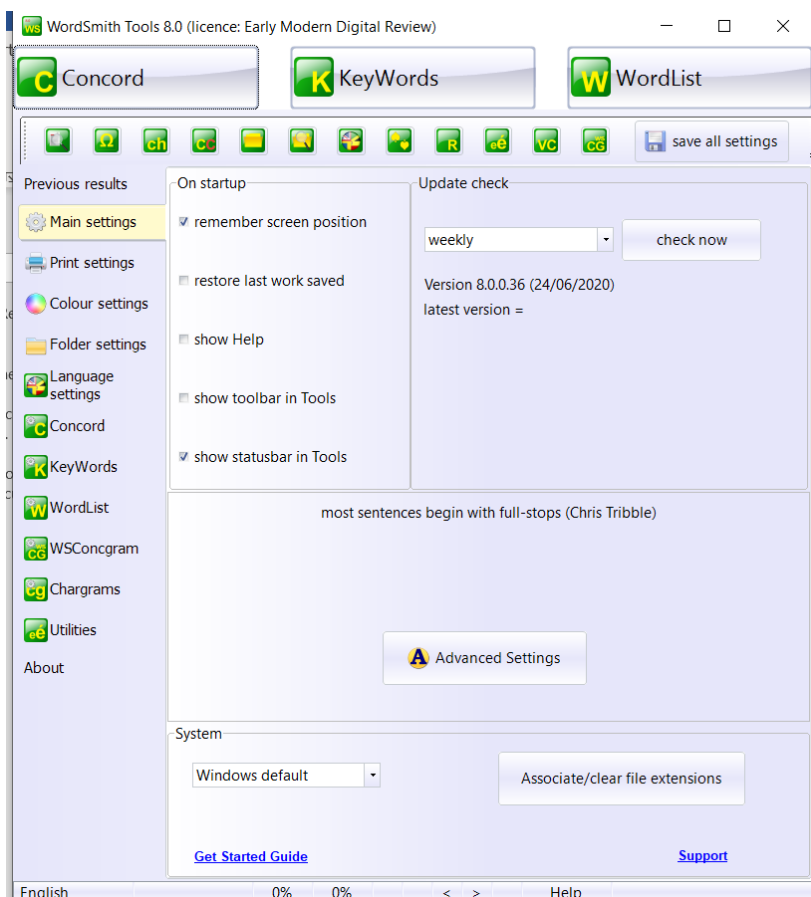


Figure 1. WordSmith Tools launch window.

While the number of options may at first seem overwhelming, analyses with Concord, KeyWords, or Wordlist all begin in the same way: assembling your corpus by choosing files to open for analysis. This process is intuitive enough: you must choose a folder to open, which will then show the .txt files available on the left-hand side of the window. It is possible—albeit time consuming—to choose multiple files by ctrl+clicking each one individually, but more convenient to select an entire folder to upload at once if you wish to analyze multiple files simultaneously.

Figure 2. Exemplar of concordance results in WordSmith Tools.

The concordance window (fig. 2) shows each result in its immediate context, in addition to giving its sequential position in the text, the number of the sentence and paragraph in which it appears, and its section number and position; if the user has defined sets, tags, or headings, these are also visible here. It is possible to sort the concordance results by any of these features, in either ascending or descending order. From Concord, other functions can be reached through the tabs at the bottom of the window. First, the “collocates” function enables computation of the words that most frequently appear near the search word(s), and their location relative to the search word. The “plot” tab shows how many “hits” are in each file returned by the search, the search word’s normalized frequency per thousand words, and its dispersion. It also provides

a visualization of the search word's position(s) in the text(s) on a plot. Next, the "patterns" tab shows the words that most frequently collate with the search term and their relative position to it (within the range of five to the left or right).

The remaining features of Concord are simpler, though no less useful. The "clusters" function identifies words that frequently occur together, giving the frequency of their co-occurrence in the corpus and their length (counted in words). If the corpus' texts are dated, chronological variations in the search word's frequency can be viewed along a "timeline." "Filenames," unsurprisingly, gives a list of files along with their number of tokens and, if assigned, its date. Double-clicking on a hit in the "concordance" tab will take you to the tab "source text," which allows you to view the entire text of a file, thus enabling greater contextualization of where a concordance hit appears within its source text. Finally, the "notes" tab provides a convenient summary of the present settings and actions performed by the user, which can be copied and pasted to an external text file to be saved as a record.

The other two main functions, KeyWords and WordList, are slightly more straightforward. KeyWords compares a given text to a (user-defined) word list in order to identify the "most unusually frequent" words appearing in the text—in other words, those that have a higher frequency than the norm provided by the word list. The WordList function is able to generate these lists (which, incidentally, are useful in themselves for studying the vocabulary used, word use frequency, common word clusters, and so forth).

Concord, KeyWords, and WordList each allow results to be saved and exported as plain text, SML, Excel, Rich Text (RTF), or a concordance list within WordSmith Tools itself. The latter option is a particularly welcome feature, as it enables resumption of an interrupted analysis.

Research possibilities with WordSmith Tools

The functions of WordSmith Tools provide a simple and rapid means to survey the use of a word or words within a corpus. The program's capabilities in this respect could be used as a point from which to begin enquires into a text or set of texts. Once key words are identified within a research topic, a search using WordSmith Tools would point to where the words of interest are used in a given

corpus—producing results more quickly than a manual search and with fewer omissions due to human error.

For example, if a researcher were interested in the representation of monarchy in early modern English drama, a search for the word “king” in Shakespeare’s First Folio (fig. 2) would provide a full list of instances on a single screen. The user’s at-a-glance access to the search results is potentially timesaving in comparison to searches conducted on a platform such as EEBO. Likewise, the display of a search term in its immediate context can help a researcher identify whether a given item is of interest. Moreover, when points to be taken forward for future analysis are identified, the information provided about each search word’s position in a text ensures that it can be found easily in a physical version if needed.

In addition to facilitating qualitative research, WordSmith Tools can also provide quantitative information that can illuminate and contextualize points of textual evidence (or, indeed, promote new lines of enquiry within a research topic). For instance, the Concordance Plot function provides a visual representation of the concentration (or, conversely, absence) of a search word in a given text, in addition to its frequency per thousand words. In combination with a frequency list of the words that appear in the corpus (generated through the WordList and KeyWords functions), this information gives quantitative insight into how widespread a word is in a given corpus, and can thereby inform analysis by suggesting the local and/or global prevalence of a topic, theme, or concept.

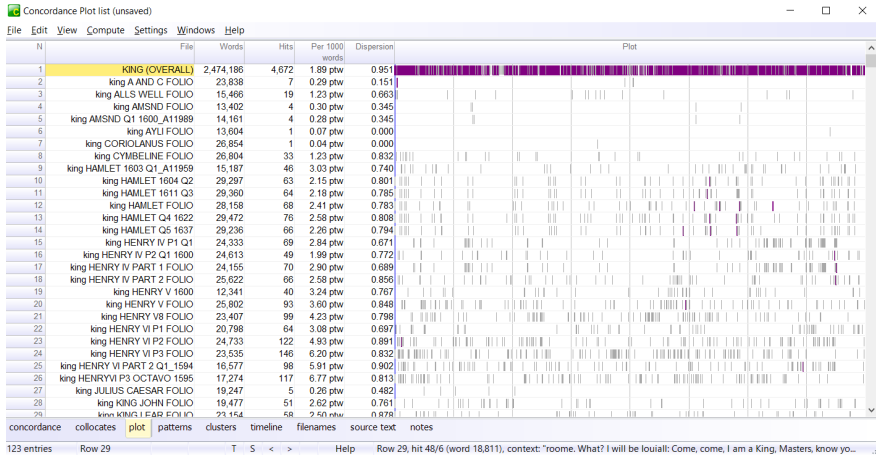


Figure 3. A Concordance Plot of “king” for the Quartos, Octavos, and First Folio texts of Shakespeare in WordSmith Tools.

For example, a researcher comparing the representation of monarchy in Shakespeare’s texts may be interested in the dispersion of the word “king.” A higher dispersion value indicates that the word occurs more evenly throughout a text, whereas a lower value means that the frequency of its use fluctuates to a greater degree. This information allows various comparisons, both within and between texts in a corpus. To give an example from the Shakespearean corpus sampled above (fig. 3), the higher concentration of the word “king” towards the end of the Folio text of the play *King Henry IV Part 1* could inform examinations of this text that are alert to, and consider, this relative difference in its content. Comparing the dispersion values of a word in different items within a corpus, or between corpora, could equally lead to further research questions and highlight areas for further qualitative examination.

Producing a Concordance Cluster list, through returning the most common collocations in a text, may also inform research by pointing to frequent, or infrequent, word combinations. Knowing whether a word combination is recurrent or unusual in a corpus provokes questions about the use of language by a certain text, and could also provide a numeric basis to support any claims that are made later in the research process. For instance, the prevalence of the definite article before the word “king” in Shakespeare’s First Folio (fig. 4)—as

opposed to the indeterminate “a”—might raise queries about the importance of this construction in Shakespearean drama, or, conversely, provoke examination of the instances where the article “a” is used instead, due to a new awareness of its relative deviance from the form usually found in Shakespeare’s corpus.

Given its broad range of functions, WordSmith Tools offers insight into the structure of a text or texts in quantitative terms. The overview thereby gained allows linguistic phenomena to be contextualized within a corpus on various levels: locally, within an entire text, or within a corpus. With multiple searches, of course, it is also possible to compare the results given by different corpora. The program can support various stages of the research process, such as the initial identification of points of interest and the discovery of patterns within or between texts, and provide quantitative insights into corpora that extend or challenge its analysis.

Concordance Cluster list (unsaved)

File	Edit	View	Compute	Settings	Windows	Help
N	Cluster	Freq.	Set	Length	Related	
1	TO THE KING	208		3		
2	OF THE KING	163		3		
3	THE KING AND	156		3		
4	THE KING IS	152		3		
5	THE KING OF	98		3		
6	THE KING MY	85		3		
7	THE KING HATH	82		3		
8	OF A KING	74		3		
9	WITH THE KING	68		3		
10	KING OF ENGLAND	62		3		
11	FROM THE KING	56		3		
12	A KING AND	56		3		
13	THE KING YOUR	53		3		
14	THE KING SHALL	52		3		
15	KING YOUR FATHER	47		3		
16	IS THE KING	44		3		
17	THE KING THAT	42		3		
18	LORD THE KING	40		3		
19	LIKE THE KING	40		3		
20	THE KING WILL	38		3		
21	THE KING THE	36		3		
22	MY LORD THE	36		3		
23	THAT THE KING	36		3		
24	KING MY FATHER	35		3		
25	THE KING HIMSELFE	34		3		
26	MY KING AND	34		3		
27	IF THE KING	34		3		
28	SHALL BE KING	34		3		
29	THE KING DOTH	34		3		

concordance

collocates

plot

patterns

clusters

timeline

filenames

source text

notes

788 entries

Row 1

T

S

<

>

Help

TO THE KING

Figure 4. A Concordance Cluster list of “king” for the Quartos, Octavos, and First Folio plays of Shakespeare in WordSmith Tools.

Getting started with AntConc

To begin analysis in AntConc, you first open the desired file or files (mercifully, AntConc does have the ability to drag and click to choose multiple files at once, as well as the ability to upload entire folders). A single search term may then be entered; it is possible to enter multiple terms by clicking on “advanced settings,” from which it is also possible to import a list of terms. “Advanced settings” allows you to define a context window (that is, to search for a term only where it appears within the designated proximity to another term).

Searching, again, for the word “king” in the texts of Shakespeare’s First Folio (in original spelling) results in the following output (fig. 5):

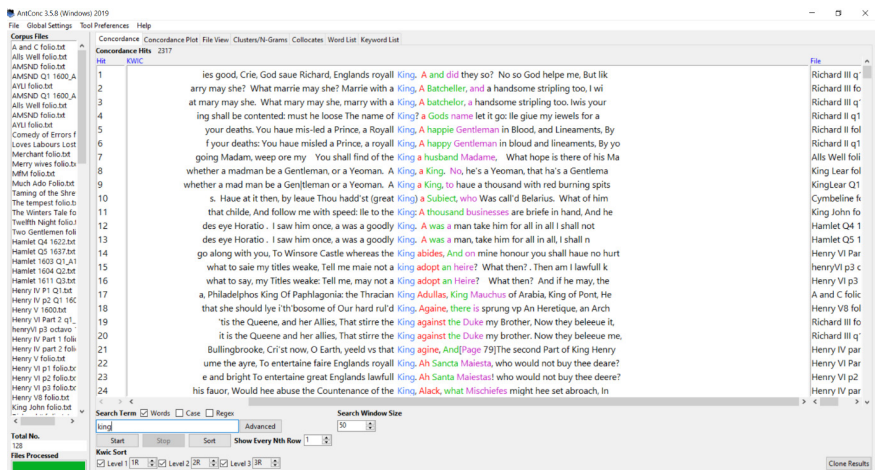


Figure 5. Exemplar of concordance results in AntConc.

Changing the ‘search window size’ adjusts the number of characters shown either side of the search, while “Kwic sort” (keywords in context) enables the search returns to be reordered according to the words surrounding the search term. For instance, in the following case (fig. 6) the search has been reordered by the first term to the left of the search word, 1L (rather than, as above, the first to the right, 1R):

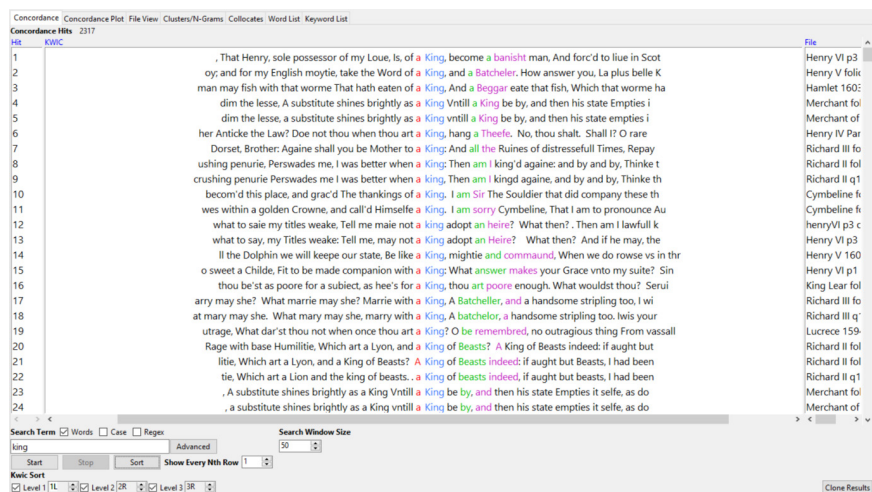


Figure 6. KWIC sorting in AntConc.

The next tab, “Concordance Plot,” provides a visualization of where the hits in a file occur. “File View” enables one to view an entire file (again, accessed by double-clicking on a returned hit on the Concordance tab). The “Clusters/N-grams” tag identifies terms that frequently appear with the search term(s), according to user-defined parameters. The “Collocates” tab identifies the words that appear near the search word most frequently and their locations, while the WordList tool lists the words in the corpus by their frequency (and can incorporate lemma forms if desired). The Keyword List tool compares the present corpus to a reference corpus, which must be provided by the user, in order to find the words that are most unusually typical of a given text(s); it is possible to change the statistical operation by which this is calculated, and its parameters, in the Tool Preferences.

The results of all of the operations in AntConc can be saved and exported as a .txt, html, xml, AntConc, or PostScript file. It is also possible to save one’s chosen settings for use in the next session.

Research possibilities with AntConc

AntConc also provides a useful overview of a corpus, which saves time and increases accuracy in comparison to manually searching for a word. These results may be helpful when beginning enquiries into the use of key terms in a corpus. The colour-coding of the words that surround the search word (which can be modified in the program settings) and the ability to easily change the order in which these concordances are sorted according to these adjacent words (see figs. 5 and 6) are advantages that AntConc has over WordSmith Tools for those primarily interested in reviewing concordance lists. For example, if I were interested specifically in the use of determiners before the word “king” in Shakespeare’s plays, the ability to filter the initial results by the first word to the left (“1L”; see fig. 6) would allow me to group my results together more efficiently. Likewise, the colour coding makes the position of other words in relation to the search word more immediately apparent.

The Concordance Plot function of AntConc does not provide quite as much information as that of WordSmith Tools but could still be useful as a general indicator of the frequency and prevalence of a key term throughout a corpus. Unlike WordSmith Tools, no numerical analysis of dispersion is available, nor is a clear indication of the word’s location in a given text provided. Likewise, AntConc does not provide a measure of the relative frequency of a search word within a text, making comparisons between texts within a corpus (which are highly likely to be of varying lengths) more problematic. Nevertheless, the visualization below (fig. 7) is useful to get a general sense of where, and how often, a word appears within a corpus.

The “Clusters/N-grams” function is also helpful to identify words that frequently appear with a search word. The ability to manipulate the size of collocation searched for from within the search interface in AntConc is an advantage, as it allows diverse queries to be made quickly and easily without having to go into the program settings to change search parameters.

For instance, if you are interested only in 2-grams (two words that appear immediately next to one another) of the word “king,” AntConc can be used to provide a list of these pairs of words in addition to noting their frequency throughout the corpus (fig. 8); or, conversely, the search could be tailored to return longer phrases. The data generated about the statistical probability of the search term preceding the word that follows may also be of interest.

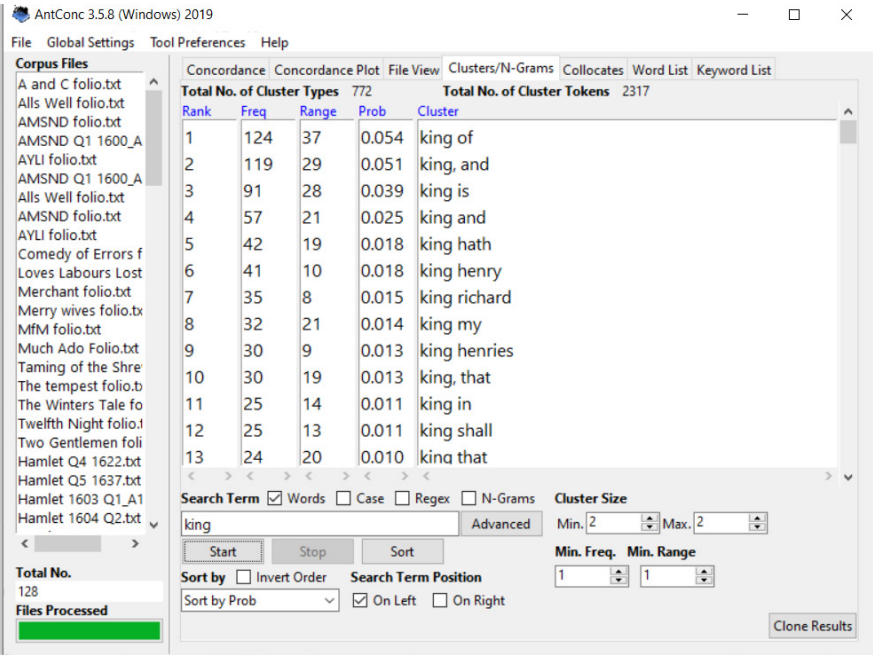


Figure 7. Concordance Plot of “king” in the Quartos, Octavos, and First Folio texts of Shakespeare in AntConc.

Much like WordSmith Tools, AntConc could be used to inform and shape research questions and is equally promising for its ability to support textual analysis with numerical data. As AntConc is free and openly available, it may well be a very attractive option; however, the absence of statistical information about the dispersion and relative frequency of a work means that the tool might be more suited to preliminary exploration of a corpus than to its qualitative analysis in some cases.

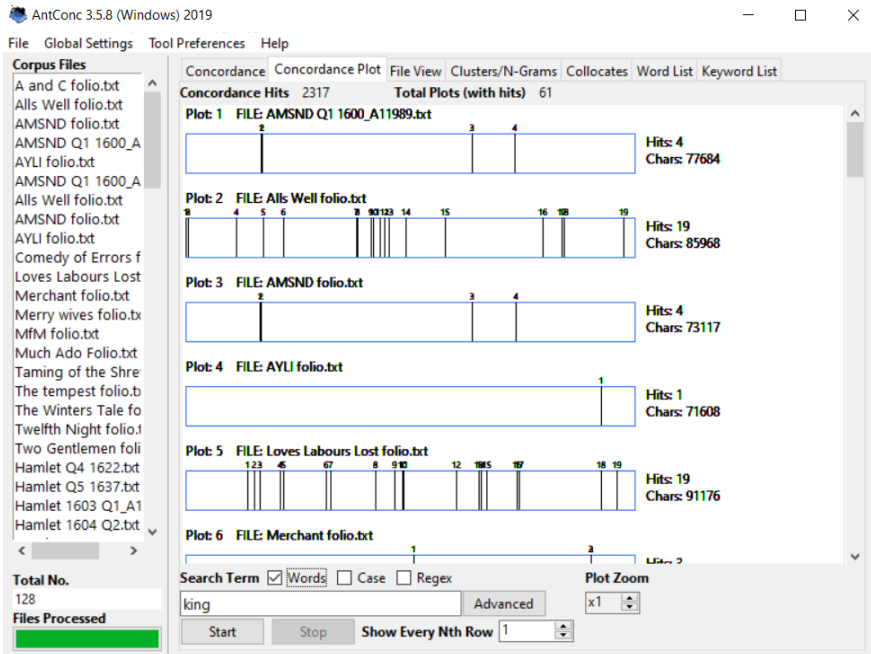


Figure 8. ‘Clusters/N-Grams’ of “king” in the Quartos, Octavos, and First Folio texts of Shakespeare in AntConc.

Advanced settings and options

Although very similar in terms of basic functions, the range of advanced settings and options means that WordSmith Tools begins to become a better option for more extensive use. WordSmith Tools accepts any plain text file for processing, including HTML or XML; conveniently, it has an inbuilt Text Converter for non-compatible files, which also enables the batch conversion of files. When converting such texts, the advanced settings allow the user to choose how markup is processed, including tags to include or exclude, entities to translate, and automatic conversions of tags (such as ‘*Eacute’ to ‘É’, for instance). In contrast, AntConc only works with .txt files, and does not have an inbuilt converter; there is, however, the option to either show or hide tags.

In both WordSmith Tools and AntConc, there is also an option to upload a stop list file (the former also allows you to choose whether to remove or retain these stop words from statistical calculations made upon the file) and a

lemmas file. While lemma lists for AntConc are available on Anthony's website for modern English, French, and Spanish, these will not necessarily produce desirable results when used with early modern language. Although the use of a reference lemma file would cause similar issues in WordSmith Tools, its option to interact manually with the text—clicking and dragging entries together in order to link them as lemmas—might be particularly useful in an early modern context, with its characteristically unpredictable and idiosyncratic spelling (and oddities even when modernization has been applied).

There are marginally more options for visual customization of the program's output—for aesthetic or accessibility reasons—in WordSmith Tools, which enables the user to change the colours and fonts of most elements; however, AntConc does allow such customization for text. The interfaces of both WordSmith Tools and AntConc are, however, admittedly rather clunky-looking.

In terms of future development, AntConc lists some "Plans for 2020 and beyond" ("AntConc Homepage," laurenceanthony.net/software/antconc) including increasing the types of file import supported and improving tag handling, but states that nothing is currently under development. WordSmith Tools reached Version 8.0 in 2020 and appears to remain in active development with monthly updates. It is also worth noting that Anthony develops numerous other tools, some of which both overlap with and improve upon functions within AntConc (such as AntGram for the identification of n-grams, for instance). However, if you really like keeping to a single program, the built-in corpus-checking and conversion facilities of WordSmith Tools might just win you over. Overall, there may not be enough of a difference to justify the cost of WordSmith Tools for those beginning to explore concordancing and who are not sure that they will use it long-term; for those lucky enough to obtain access through their institution or who discover a new love of concordancing, though, the advanced settings and/or the ability to save sessions may justify the purchase price of WordSmith Tools.

Both tools have the potential to be very welcome additions for humanities researchers: the ability to explore textual patterns encourages a liberating, and enlightening, new perspective.

EMILY LOUISA SMITH

University of Geneva

<https://doi.org/10.33137/rr.v44i1.37062>