

A Corpus Linguistics Primer

Through AntConc

What is a corpus?

What?

It's so simple.

It's just a lot of words!

Well, not exactly.

What is a corpus?

A machine readable collection of texts from spoken or written sources that were created in a natural expressive context.

Machine readable: text formats that can be loaded, parsed and manipulated independent of platforms. Despite being technically 'open', they can have dense annotations identifying various kinds of descriptive features. Typically these files are stored in the form of plain text files stored in ASCII or UTF encoding or structured XML files.

Natural expressive context: texts that were not created for the purpose of corpus analysis; in other words, texts that partake in an authentic communication.

A machine readable collection of texts from spoken or written sources that were created in a natural expressive context.

It constitutes methodology for studying the nature of language.

In such a collection, it is expected that there is an intention:

- To identify the collection as representative and balanced in the context of a language, variety, register, or genre. It has a purpose, in other words, so it should aim to reference what is typical.
- To analyse the collection linguistically (attention to word frequency, language change, morphemes, and the like), with explicit annotations.

Types of corpora

General: represents a language in a holistic way.

Specific: restricted to a particular variety, register, or purpose.

Raw: contains files of only corpus material (plain text)

Annotated: contains additional descriptive information (usually with metadata), encoded with parts-of-speech tags, or XML tags under the guidelines of the Text Encoding Initiative (TEI) or Corpus Encoding Standard (CES). Annotated corpuses include information *about* the text within *markup*. This kind of corpus can also be *lemmatized* (each word is followed by its lemma—the standard form that you would look up in the dictionary).

More types of corpora

Diachronic: shows language change over time.

Synchronic: shows a snapshot of language in a time.

Monolingual: shows one language.

Parallel: shows the same text in multiple languages.

Static: have a fixed size (e.g. British National Corpus).

Dynamic: can be constantly extended (e.g. Bank of England).

Does this analysis constitute a theory of language?

Strictly speaking, no. These analyses offer loads of information about huge amounts of textual data, but they only offer information about frequencies. There is no straightforward semantic meaning in a corpora; what you are seeing in corpus analysis is:

- Frequencies of items (how often words or morphemes or grammatical structures occur in a text)
- Frequencies of co-occurring things (that is, groups of words or grammatical structures)

The work of interpretation still needs to be done; you need to decide what is meaningful. Remember, though, that ***formal differences reflect functional differences***. Formal qualities illustrate functional regularities in communication.

AntConc <<http://www.laurenceanthony.net/software.html>>

Pros: Free, well-maintained application; can search and analyse multiple texts in a corpus; has impressive key-word-in-context functions.

Cons: It can only perform basic corpus analyses (i.e., it cannot do more complex linguistic analyses).

Getting started: File > OpenDir

The screenshot shows the AntConc 3.5.7 (Macintosh OS X) 2018 application window. The title bar includes the text "AntConc 3.5.7 (Macintosh OS X) 2018" and system status icons for "Stopped", network, and battery. The menu bar contains "AntConc", "File", "Settings", and "Help". The "File" menu is open, and the "Open Dir..." option is highlighted in blue. A red box highlights the "File" menu and the "Open Dir..." option. The main window area is mostly empty, with a "Concordance" tab selected. The bottom panel contains search settings: "Search Term" with a text input field, "Advanced" button, "Search Window Size" set to 50, "Start", "Stop", "Sort", and "Show Every Nth Row" set to 1. The "Kwic Sort" section has "Level 1 1R", "Level 2 2R", and "Level 3 3R" all checked. A "Clone Results" button is in the bottom right corner. The bottom left corner shows "Total No. 0" and "Files Processed".

AntConc 3.5.7 (Macintosh OS X) 2018

Stopped

Wed 21 Nov 11:18 92%

File

Open File(s)... ^F

Open Dir... ^D

Close Selected File(s)

Close All Files

Clear Tool

Clear All Tools

Clear All Tools and Files

Save Output... ^S

Import Settings from File...

Export Settings To File...

Restore Default Settings

Exit

Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

Word List

Keyword List

File

Search Term Words Case Regex

Advanced

Search Window Size 50

Start Stop Sort Show Every Nth Row 1

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Clone Results

Total No. 0

Files Processed

Choose a corpus folder with texts

AntConc 3.5.7 (Macintosh OS X) 2018

File Settings Help

AntConc 3.5.7 (Macintosh OS X) 2018

Concordance Hits 0

Hit KWIC

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

corp

Devices

- Remote Disc
- CleanMyMac 3
- GoogleVoiceAndVideoAcc...
- GoogleVoiceAndVideoAcc...
- Skype
- Screencast-O-Matic v2.0

corp

- is.txt
- sty.JPG
- otes.txt
- 17.key
- 3.pdf
- (1).pdf
- .md
- ph.pdf
- pdf
- proposal

corp

- durham-wor...ta_EDGES.csv
- stylo_config.txt
- stylo_tests.R
- table_with_frequencies.txt
- wordlist.txt

beale-natural-history-sperm-whale.txt

carlyle_french-revolution.txt

chapman-iliad.txt

chapman-odyssey.txt

gibbon_decline-and-fall.txt

israel-potter.txt

livingstone_j...travels_v2.txt

livingstone_j...ournals_v1.txt

livy_history-rome_v1.txt

livy_history-rome_v2.txt

livy_history-rome_v3.txt

milton-complete.txt

shakespeare-complete.txt

wollstonecra...ts-woman.txt

New Folder Cancel Choose

Search Term Words Case Regex

Advanced 50

Start Stop Sort Show Every Nth Row 1

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

Total No. 0 Files Processed

Clone Results

After texts have loaded, click the Word List tab (top right), then click Start (lower left) to generate a word list

The screenshot shows the AntConc 3.5.7 (Macintosh OS X) 2018 interface. The 'Word List' tab is selected and highlighted with a red box. The 'Start' button is also highlighted with a red box. The interface displays the following information:

- Corpus Files:** A list of files including beale-natural-history-spe, carlyle_french-revolution.1, chapman-iliad.txt, gibbon-odyssey.txt, gibbon_decline-and-fall.tx, israel-potter.txt, livingstone_last-journals_v, livingstone_last_travels_v2, livy_history-rome_v1.txt, livy_history-rome_v2.txt, livy_history-rome_v3.txt, milton-complete.txt, shakespeare-complete.txt, and wollstonecraft_vindication.
- Search Statistics:** Word Types: 0, Word Tokens: 0, Search Hits: 0.
- Search Term:** Words (checked), Case, and Regex (unchecked).
- Hit Location:** Search Only, 0.
- Lemma List:** Loaded (unchecked).
- Word List:** Loaded (unchecked).
- Sort by:** Sort by Freq (selected), Invert Order (unchecked).
- Buttons:** Start, Stop, Sort, Advanced, Clone Results.
- Bottom Left:** Total No. 14, Files Processed.

This is what you should get

AntConc 3.5.7 (Macintosh OS X) 2018

Concordance Concordance Plot File View Clusters/N-Grams Collocates **Word List** Keyword List

Word Types: 80653 Word Tokens: 5631007 Search Hits: 0

Rank Freq Word Lemma Word Form(s)

1	464979	the	
2	289862	of	
3	211613	and	
4	129176	to	
5	95682	a	
6	87198	in	
7	66126	his	
8	55213	that	
9	51994	was	
10	51672	by	
11	48763	with	
12	42753	he	
13	40930	their	
14	38288	i	
15	32977	as	
16	31829	were	
17	31572	is	
18	31027	for	
19	30063	but	
20	29683	which	
21	29582	it	
22	29325	from	
--	-----	.	

Search Term Words Case Regex Hit Location

Advanced Search Only 0

Start Stop Sort Lemma List Loaded Word List Loaded

Sort by Invert Order Sort by Freq

Total No. 14
Files Processed

Clone Results

Now click on the Concordances tab and enter a search term

Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 5151

Hit	KWIC	File
1	are revived as to the results of the	war_ * * * * *
2	kes must arbitrate. Towards which advance the	war.
3	As black as Vulcan in the smoke of	war. A baubling vessel was he captain of, For
4	Romans considered their being at liberty to make	war, a certain victory ; while the Samnites supposed t
5	had been reduced to a few by intestine	war, a colony should be sent thither as a
6	vexing them, and breed (to soothe their childish	war) A common ill to many men, since if
7	, when I reflect that, in the first Punic	war, a contest was maintained by the Romans with
8	of Africa, (A.D. 439,) and before Attila's	war, (A.D. 451.)] 78 (return) [The Bagaudae of Spai
9	, for an instant, the operations of an active	war. A dark conspiracy was detected by the penetratio
10	, for an instant, the operations of an active	war. A dark conspiracy was detected by the penetratio
11	always been successful in the event of the	war." A few days after the departure of Narses,
12	always been successful in the event of the	war." A few days after the departure of Narses,
13	allow the state to breathe in time of	war. 27. A fire which broke out in several places
14	housand horse, together with thirty-five ships of	war, a force of no small importance to bring
15	the more improved state of the art of	war, a general is seldom required, or even permitted
16	the more improved state of the art of	war, a general is seldom required, or even permitted
17	enemy equally skilled in all the instruments of	war. A generous emulation inspired the Romans and the
18	enemy equally skilled in all the instruments of	war. A generous emulation inspired the Romans and the
19	, containing food to be used in case of	war. A large cow is kept up there, which
20	had arisen in arms, they were neglecting the	war, a letter was brought from Quintus Minucius, anno
21	might not pass the year entirely exempt from	war, a little expedition was made into Umbria; intell
22	hichte des Osmanischen Reiches. (M.) I. For every	war, a motive of safety or revenge, of honor
23	hichte des Osmanischen Reiches. (M.) I. For every	war, a motive of safety or revenge, of honor
24	as bad as falling; the toil o' th'	war. A pain that only seems to seek out

Search Term Words Case Regex Search Window Size 50

war Show Every Nth Row 1

Kwic Sort Level 1 1R Level 2 2R Level 3 3R

Adjust the concordance results from right to left

AntConc 3.5.7 (Macintosh OS X) 2018

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 831

Hit	KWIC	File
1	life with dignity, and to acquire wisdom and virtue by the exercise of	wollstonecraft_\
2	, paid for their schooling, and acquired wisdom by the favour and mediation of	montaigne_com
3	not complain if he neither acquires wisdom nor respectability of character. Supposin	wollstonecraft_\
4	takes away everything else, it adds wisdom to old age. Even war, that	plutarch_morals
5	study, that we discover the adorable wisdom of God in his works: when	gibbon_decline
6	.] In the civil administration of Alexander, wisdom was enforced by power, and the	gibbon_decline
7	. In the civil administration of Alexander, wisdom was enforced by power, and the	montaigne_com
8	unknown to man. From His all wisdom nothing but good, common; and regular	montaigne_com
9	of the finest thread, and all wisdom is folly that does not accommodate	montaigne_com
10	fine precepts are vanity, and all wisdom is vanity: "Dominus novit cogitationes	gibbon_decline
11	prosecuted the design of extracting allegorical wisdom from the fictions of the Greek	gibbon_decline
12	prosecuted the design of extracting allegorical wisdom from the fictions of the Greek	wollstonecraft_\
13	of the passions is not always wisdom. On the contrary, it should seem,	gibbon_decline
14	visited the memorable scenes of ancient wisdom or glory, have confessed the inspiration	gibbon_decline
15	visited the memorable scenes of ancient wisdom or glory, have confessed the inspiration	gibbon_decline
16	necks of a people, whose ancient wisdom and power ascend beyond the records	gibbon_decline
17	necks of a people, whose ancient wisdom and power ascend beyond the records	gibbon_decline
18	man would possess more virtue and wisdom than a number of men; and	federalist_pape

Search Term Words Case Regex

wisdom Search Window Size 50

Show Every Nth Row 1

Kwic Sort Level 1 1L Level 2 2R Level 3 3R

Corpus Files

- burke_on-sublime-and-be
- burton_anatomy-of-melan
- carlyle_french-revolution.1
- federalist-papers.txt
- gibbon_decline-and-fall-r
- livy_history-of-rome_v1.tx
- livy_history-of-rome_v2.tx
- livy_history-of-rome_v3.tx
- livy_history-of-rome_v4.tx
- montaigne_complete-ess
- plutarch_morals.txt
- schopenhauer_counsels-a
- schopenhauer_studies-in
- thomas-paine_complete.t
- wollstonecraft_vindication

Total No. 15
Files Processed

Use the “Sort” button to arrange alphabetically

The screenshot shows the AntConc 3.5.7 interface on a Macintosh OS X. The main window displays a concordance search for the term "wisdom". The search results are arranged in a table with three columns: Hit, KWIC, and File. The KWIC column shows the search term highlighted in blue. The "Sort" button is highlighted in blue, indicating that the results are sorted alphabetically. The search window size is set to 50, and the search term is "wisdom". The search options are set to "Words", "Case", and "Regex". The search window size is set to 50. The search term is "wisdom". The search options are "Words", "Case", and "Regex". The search window size is 50. The search term is "wisdom". The search options are "Words", "Case", and "Regex".

AntConc 3.5.7 (Macintosh OS X) 2018

Concordance Hits 831

Hit	KWIC	File
1	life with dignity, and to acquire wisdom and virtue by the exercise of	wollstonecraft_\
2	, paid for their schooling, and acquired wisdom by the favour and mediation of	montaigne_com
3	not complain if he neither acquires wisdom nor respectability of character. Supposi	wollstonecraft_\
4	takes away everything else, it adds wisdom to old age. Even war, that	plutarch_morals
5	study, that we discover the adorable wisdom of God in his works: when	gibbon_decline
6	.] In the civil administration of Alexander, wisdom was enforced by power, and the	gibbon_decline
7	. In the civil administration of Alexander, wisdom was enforced by power, and the	montaigne_com
8	unknown to man. From His all wisdom nothing but good, common; and regular	montaigne_com
9	of the finest thread, and all wisdom is folly that does not accommodate	montaigne_com
10	fine precepts are vanity, and all wisdom is vanity: "Dominus novit cogitationes	gibbon_decline
11	prosecuted the design of extracting allegorical wisdom from the fictions of the Greek	gibbon_decline
12	prosecuted the design of extracting allegorical wisdom from the fictions of the Greek	wollstonecraft_\
13	of the passions is not always wisdom. On the contrary, it should seem,	gibbon_decline
14	visited the memorable scenes of ancient wisdom or glory, have confessed the inspiration	gibbon_decline
15	visited the memorable scenes of ancient wisdom or glory, have confessed the inspiration	gibbon_decline
16	necks of a people, whose ancient wisdom and power ascend beyond the records	gibbon_decline
17	necks of a people, whose ancient wisdom and power ascend beyond the records	gibbon_decline
18	man would possess more virtue and wisdom than a number of men and	federalist_pape

Search Term Words Case Regex

wisdom Advanced 50

Start Stop Sort Show Every Nth Row 1

Kwic Sort

Level 1 1L Level 2 2R Level 3 3R

Clone Results

Corpus Files

- burke_on-sublime-and-be
- burton_anatomy-of-melan
- carlyle_french-revolution.1
- federalist-papers.txt
- gibbon_decline-and-fall-ri
- livy_history-of-rome_v1.tx
- livy_history-of-rome_v2.tx
- livy_history-of-rome_v3.tx
- livy_history-of-rome_v4.tx
- montaigne_complete-essa
- plutarch_morals.txt
- schopenhauer_counsels-a
- schopenhauer_studies-in-
- thomas-paine_complete.t
- wollstonecraft_vindication

Total No. 15

Files Processed

Use regular expressions to do flexible searches: wildcard (*), ? and pipe (|)

The screenshot shows the AntConc 3.5.7 interface. The search term is `wise*[wisdom]` and the search window size is 50. The search results are displayed in a concordance table with columns for Hit, KWIC, and File. The search term is highlighted in the original image.

Corpus Files

- 1590 Love's Labour's Los
- 1591 Romeo and Juliet.tx
- 1591 Two Gentlemen of V
- 1592 Titus Andronicus.txt
- 1599 Hamlet.txt
- 1599 Julius Caesar.txt
- 1599 King Henry V.txt
- 1599 Merry Wives of Win
- 1599 Twelfth Night or Wh
- 1600 Much Ado about No
- 1602 Troilus and Cressid
- 1604 Measure for Measur
- 1604 Othello.txt
- 1605 King Lear.txt
- 1607 Antony and Cleopat
- 1607 Timon of Athens.txt
- 1608 Coriolanus.txt
- 1608 Pericles Prince of T
- 1610 Cymbeline.txt
- 1610 Winter's Tale.txt
- 1610 Cymbeline.txt
- 1610 Cymbeline.txt
- 1610 Cymbeline.txt
- 1610 Winter's Tale.txt
- 1608 Pericles Prince of T
- 1610 Cymbeline.txt
- 1610 Winter's Tale.txt
- 1611 Tempest.txt
- 1613 King Henry VIII.txt

Concordance Hits 288

Hit	KWIC	File
13	... was never More covetous of wisdom and fair virtue Than thi	1599 Julius Cai
14	... then the bold and coward, The wise and fool, the artist and unread,	1597 King Heni
15	... REUS 'Tis your noblest course. Wisdom and fortune combating toge	1602 Troilus an
16	... good will, look you: you are wise and full of gibes and vlouting-	1613 King Heni
17	... , yet I'll speak. IAGO Be wise, and get you home.	1613 King Heni
18	... he text is foolish. ALBANY Wisdom and goodness to the vile sei	1602 Troilus an
18	... made them do it: they are wise and honourable, And will,	1613 King Heni
18	... him not; and yet he talked wisely, and in the street too.	1593 King Rich
19	... you love not, for to be wise and love Exceeds man's n	1599 Hamlet.tx
19	... think an English courtier may be wise, And never see the Louvre	1599 Hamlet.tx
20	... gality of nature, Young, valiant, wise, and, no doubt, right royal,	1599 Hamlet.tx
20	... truth: And thus do we of wisdom and of reach, With win	1591 Romeo ar
21	... Romeo! MERCUTIO He is wise; And, on my lie, hath stol'	1599 Merry Wi
21	... Sir Hugh hath shown himself a wise and patient churchman. Y	1598 As You Lil
22	... of flesh indeed! Learn of the wise, and perpend: civet is of a	1590 Love's La
23	... ROSALINE This proves you wise and rich, for in my eye,--	1590 Love's La
23	... ; to converse with him that is wise, and says little; to fear jud	1605 King Lear

Search Term Words Case Regex

`wise*[wisdom]` **Search Window Size** 50

Show Every Nth Row 1

Kwic Sort

- Level 1 1R
- Level 2 2R
- Level 3 3R

Total No. 31
Files Processed

More on wildcards

Wildcards are used for matching patterns. Technically they are characters that can be used as a substitute for any class of characters in a search, which increases the flexibility and efficiency of searches.

For a full list of available wildcard operators and what they mean, go to [Global Settings > Wildcard Settings](#).

The ? operator is “less greedy” than the * operator:

wom?n – both women and woman

m?n – man and men, but also min

m*n is more flexible: you’ll get mean, melon, etc.

The Concordance Plot tab visualises the search results in each file

Corpus Files

beale-natural-history-spe
carlyle_french-revolution.1
chapman-iliad.txt
chapman-odyssey.txt
gibbon_decline-and-fall.tx
israel-potter.txt
livingstone_last-journals_v
livingstone_last-travels_v2
livy_history-rome_v1.txt
livy_history-rome_v2.txt
livy_history-rome_v3.txt
milton-complete.txt
shakespeare-complete.txt
wollstonecraft_vindication

Total No.

14

Files Processed

Concordance **Concordance Plot** File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 5151 **Total Plots (with hits)** 14

Plot: 1 FILE: beale-natural-history-sper-m-whale.txt
Hits: 11
Chars: 604959

Plot: 2 FILE: carlyle_french-revolution.txt
Hits: 190
Chars: 1852341

Plot: 3 FILE: chapman-iliad.txt
Hits: 176
Chars: 984355

Plot: 4 FILE: chapman-odyssey.txt
Hits: 49
Chars: 997702

Plot: 5 FILE: gibbon_decline-and-fall.txt
Hits: 2462
Chars: 16415589

Plot: 6 FILE: israel-potter.txt
Hits: 44
Chars: 377251

Plot: 7 FILE: livingstone_last-journals_v1.txt
Hits: 31
Chars: 686853

Search Term Words Case Regex

Plot Zoom

war x1

Show Every Nth Row 1

Ngram search finds groups of co-occurring words (adjust the “cluster size” to see more context)

AntConc 3.5.7 (Macintosh OS X) 2018

Concordance Concordance Plot File View **Clusters/N-Grams** Collocates Word List Keyword List

Total No. of Cluster Types 3029 Total No. of Cluster Tokens 5151

Rank	Freq	Range	Cluster
1	118	4	war against the
2	77	6	war with the
3	60	3	war; and the
4	52	4	war, and the
5	23	6	war in the
6	22	6	war to the
7	22	3	war. in the
8	20	1	war and government
9	20	6	war of the
10	19	5	war and the
11	19	4	war on the
12	19	5	war should be
13	18	4	war by the
14	17	3	war had been
15	17	2	war. but the
16	15	4	war which he
17	14	3	war with rome
18	14	5	war, in which
19	13	2	war, with the
20	12	1	war and rapine
21	12	1	war; but the
22	11	2	war with philip

Search Term Words Case Regex N-Grams Cluster Size Min. 3 Max. 3

war Advanced Start Stop Sort

Sort by Invert Order Search Term Position Min. Freq. 1 Min. Range 1

Sort by Freq On Left On Right

Total No. 14 Files Processed

Clone Results

Save the output as a txt file

The screenshot displays the AntConc 3.5.7 (Macintosh OS X) 2018 application window. The interface is divided into several sections:

- Corpus Files:** A list of 31 files, including '1590 Love's Labour's Los', '1591 Romeo and Juliet.txt', and '1613 King Henry VIII.txt'. The total number of files processed is 31.
- Search Term:** The search term is 'wise*|wisdom'. The search is configured for 'Words', 'Case', and 'Regex'. The search window size is set to 50.
- Search Results:** A list of search results is shown, with highlighted text in the original image. The results include: 'made them do it: they are wise and honourable, And will, him not; and yet he talked wisely, and in the street too. you love not, for to be wise and love Exceeds man's r think an English courtier may be wise, And never see the Louvre gality of nature, Young, valiant, wise, and, no doubt, right royal, truth: And thus do we of wisdom and of reach, With win Romeo! MERCUTIO He is wise; And, on my lie, hath stol Sir Hugh hath shown himself a wise and patient churchman. Y of flesh indeed! Learn of the wise, and perpend: civet is of a ROSALINE This proves you wise and rich, for in my eye,-- ; to converse with him that is wise, and says little; to fear jud
- Save As Dialog:** A dialog box is open, showing the file name 'antconc_results.txt' and the location 'Dropbox'. The 'Save' button is highlighted.

1 grace has given a precedent of wisdom Above all princes, in
committing freely 1613 King Henry VIII.txt 30 3
2 in nature, As by your safety, wisdom, all things
else, You mainly were 1599 Hamlet.txt 13 13
3 may believe: censure me in your wisdom, and awake your
senses, that you 1599 Julius Caesar.txt 14 6
4 Benedick. LEONATO O, my lord, wisdom and blood
combating in so tender 1600 Much Ado about Nothing.txt 18 4
5 , in her sex, her years, profession, Wisdom and constancy,
hath amazed me more 1598 All's Well That Ends Well.txt 11 3
6 , quarter'd, hath but one part wisdom And ever three parts
coward, I 1599 Hamlet.txt 13 11
7 : Saba was never More covetous of wisdom and fair
virtue Than this pure 1613 King Henry VIII.txt 30 12
8 then the bold and coward, The wise and fool, the artist and
unread, 1602 Troilus and Cressida.txt 19 1
9 THYREUS 'Tis your noblest course. Wisdom and fortune
combating together, If that 1607 Antony and Cleopatra.txt 23 3
10 good will, look you: you are wise and full of gibes and
vlouting- 1599 Merry Wives of Windsor.txt 16 9

Compare corpora

First, click File > Clear all Tools and Files. Then:

Settings > Tool preferences > Keyword List

Under 'Reference Corpus', click "Use raw files"

Add Directory > select the folder containing the text files

Click Load; when it is finished, click Apply.

Click on the Keyword List tab, and click Start. (AntConc will warn that it needs to jump to the Word List—that is fine.) (Note that *keyness* is the frequency of a word compared to its frequency in a reference corpus.)

Exercises

1. Load the poetry corpus and sort the hapax legomena alphabetically.
2. Now sort the poetry results to find the most common function words in the poetry corpus.
3. Load the Shakespeare corpus into the program. You should still have the poetry files. How has the function words changed? Perform a basic word search.
4. In the Shakespeare corpus, generate collocates for $m?n$ and $wom?n$. Now sort them by frequency to 1L.
5. In Whittier's anti-slavery poems, find the names of the poems that talk about female slaves.
6. What are the most commonly negated words (that is, words preceded by "no", "not" and "never") in the c19-20 prose corpus?

BONUS: pos-tagged corpora

Download TagAnt at <<http://www.laurenceanthony.net/software/tagant/>>.

Open the same corpus directory in the same way as you would with AntConc.

Click start and the program will generate new txt files that are annotated with part-of-speech tags.

(We have made the Whittier and *Moby-Dick* pos-tagged files available on the github repository at <https://github.com/cmohge1/riga-text-analysis>.)

Now run a tagged file through AntConc and see how your results differ (hint: adjust the Kwic sort feature to 2L in order to sort by POS tag).

What is the most common verb form in the King James Bible?