

WordCruncher

User Guide

**For Corpus
Users**

Last updated August 24, 2020

**For the Microsoft
Windows Application**

Copyright © 1991–2020 Brigham Young University Provo, Utah, USA. All rights reserved.

WordCruncher® Copyright © 1991–2020 Brigham Young University Provo, Utah, USA. All rights reserved.

WordCruncher User Guide for Corpus Users © First Edition March 2020

WordCruncher Version 7.1.100, Build 100 as of March 24, 2020

WordCruncher® and the WordCruncher logo are registered trademarks of Brigham Young University.

WordCruncher is a closed source program. While WordCruncher Development tries to give the best resources to users, we reserve the right to make changes to any of the features, reports, and datasets within WordCruncher. Suggestions for improvements are welcome but will be evaluated based on current priorities.

The use of WordCruncher is free. Books that are purchased through the bookstore are connected to a user account. The WordCruncher Support team keeps a record of all book purchases. If there is any loss of data, please contact wordcruncher@byu.edu with your user information.

Licensors make no warranties, whether express or implied, including the implied warranties of merchantability and fitness for a particular purpose, and specifically no warranties or representations are made with respect to the accuracy or completeness of the data.

wordcruncher@byu.edu

Table of Contents

About WordCruncher	4
WordCruncher Features	4
Specifications	4
Terminology of the Program	4
WordCruncher Indexer	5
WordCruncher Search Tool	6
Analysis Tools in WordCruncher	7
Using WordCruncher for Corpus Analysis	7
Indexing Files to Create Books on WordCruncher	8
Indexing Files with WordCruncher Indexer	8
Opening Indexed Files as Books on WordCruncher	9
Searching using Part of Speech	10
Opening the Part of Speech Option	10
Searching for All Words in a Single Part of Speech	11
Search for a Word with a Specific Part of Speech.	12
Using Attribute Search	13
Opening the Attribute Bounds tool	13
Using the Attribute Bounds Dialogue Box	13
Using Neighborhood Report	14
Opening the Neighborhood Report Window	14
Using the Tabs in the Neighborhood Reports Window	15
Creating a Phrase Compare Report	16
Index	18

About WordCruncher

WordCruncher is a free software program used for text analysis, annotation, and comparison. Developed by a team of researchers at Brigham Young University, this program was originally designed for the study of religious documents of The Church of Jesus Christ of Latter-day Saints. WordCruncher is currently overseen by the Digital Humanities Department of BYU, which has since expanded the software's scope to enable users to process any large digitized text.

This manual will aid corpus users in utilizing the WordCruncher search, study, and analysis tools related to the corpus functions of the software. Using this software, users will be able to create their own corpora and search those texts for words, lemmas, or tagged parts of speech, much like in other corpus programs. Additionally, users will be able to compare individual texts, run complex phrase or list searches, and generate statistics. These functions, when used with searchable corpora, will help users find meaningful patterns in texts.

System requirements and installation information can be found on <http://wordcruncher.com/install/>.

WordCruncher Features

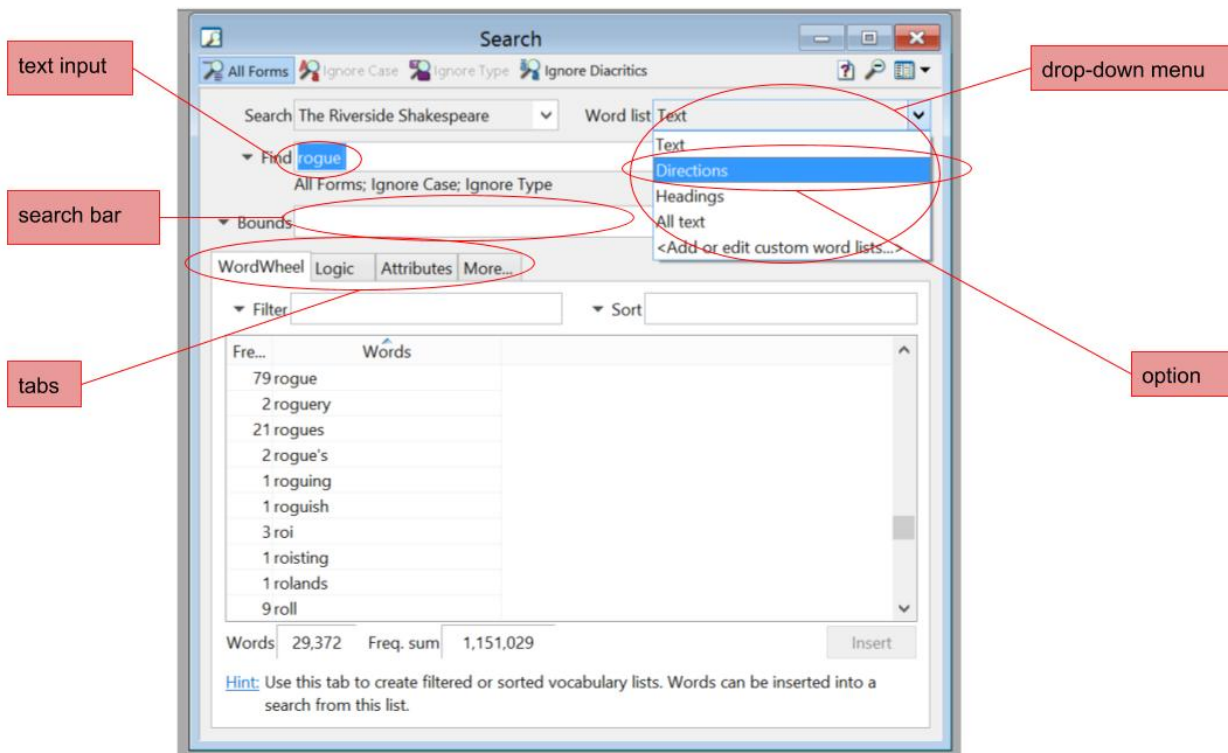
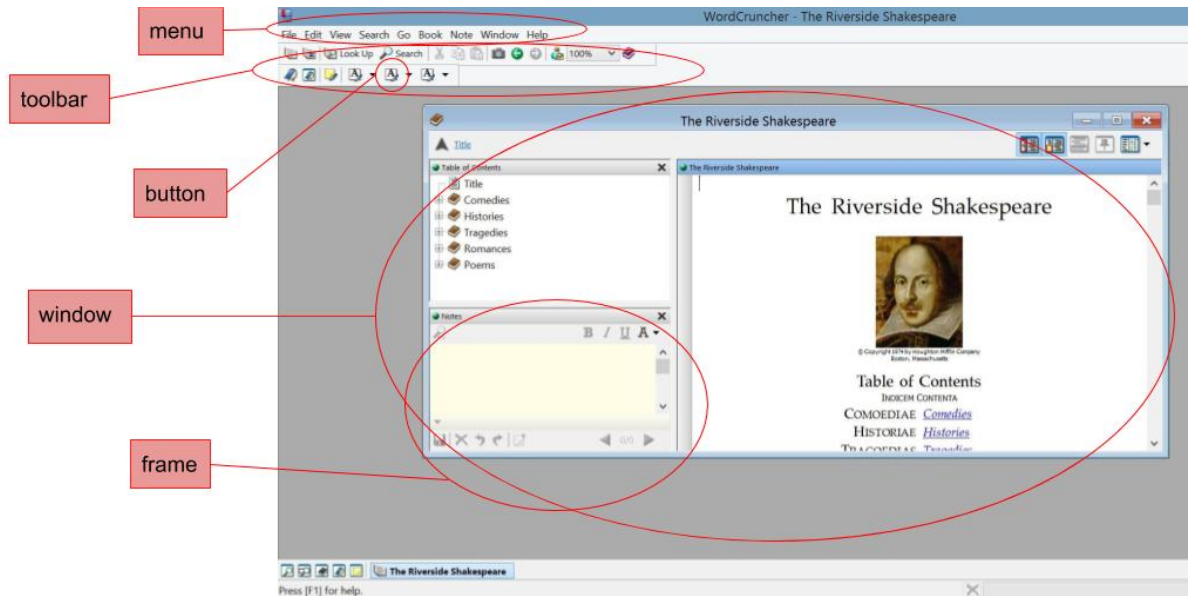
This section will give an overview of WordCruncher's specifications, define the functions of the different tools used to create and use corpora, and illustrate key aspects of the WordCruncher program needed to carry out the following instruction sets..

Specifications

The following section details the basic parts of the program you need to be familiar with to use WordCruncher's corpus tools.

Terminology of the Program

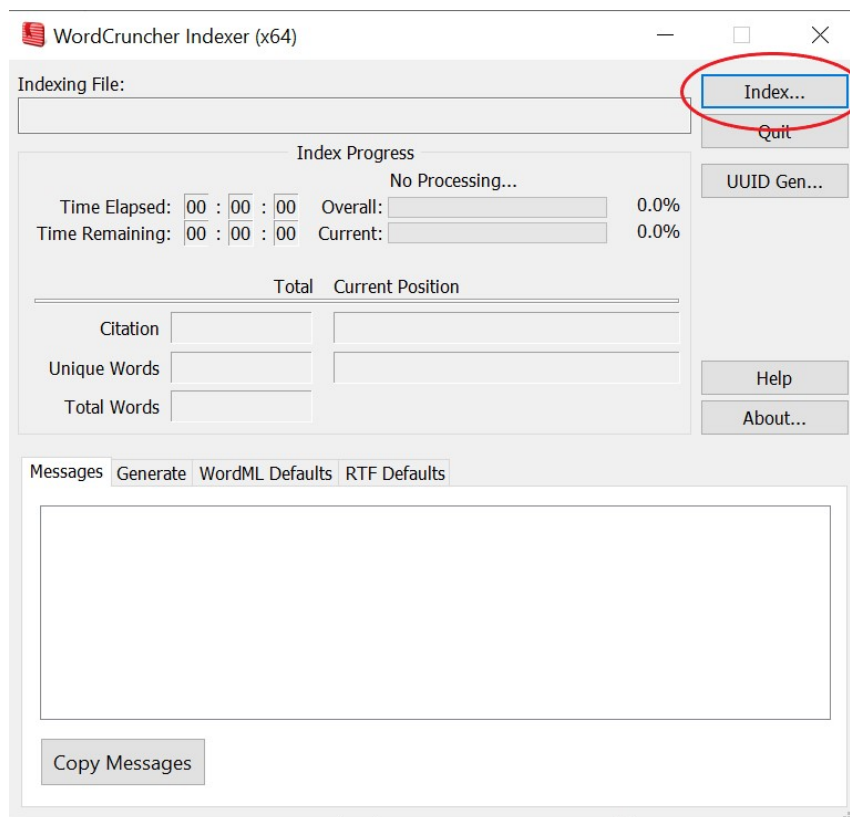
The following pictures illustrate examples of what the user guide means when it refers to "menu," "window," and other general features.



WordCruncher Indexer

The WordCruncher Indexer Program enables users to convert non-compatible documents into a compatible format to download as books in WordCruncher. Users can upload text files into the WordCruncher Indexer and use the program to convert .xml (2003), .txt, .rtf, and .etax text files into .etax and .etbu files that can be analyzed using the WordCruncher program.

The **Index** button enables users to select text files to convert to the .etbu and .etax file formats that are compatible with WordCruncher. For more details on indexing text files, see [Indexing Files to Create Books on WordCruncher](#).



WordCruncher Search Tool

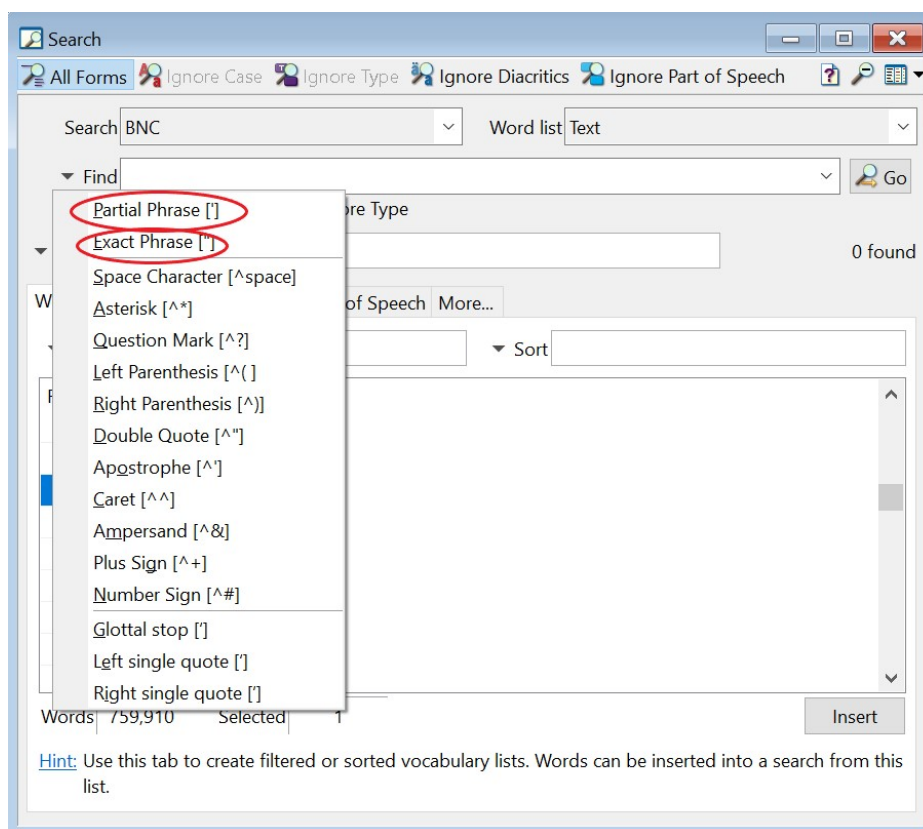
The WordCruncher search function enables users to search through a single document or multiple documents for specific words or phrases. The search function in WordCruncher can filter search results to allow users to search a document for the following items: specific keywords, partial matches to longer searches, exact matches to longer searches, all the semantic forms of a specific word, words belonging to a specific part of speech, and all the instances of individual words contained within a phrase.

The parameters enable users to place certain limitations on the search results by either searching for all semantic forms of a word ("All Forms" tab), or choosing to ignore the capitalization ("Ignore Case" tab) or embedded type ("Ignore Type" tab) of a word.

The search bar enables users to perform searches of certain keywords, exact words or phrases, and partial words or phrases in the text.

The Part of Speech Search tool (only accessible in files tagged for part of speech) enables users to search a text for specific words within a part of speech, or for all words belonging to a

specific part of speech. With this tool, users can break down the text into parts of speech and look for patterns.



Analysis Tools in WordCruncher

The analysis tools within WordCruncher will also benefit corpus creators. These tools enable users to selectively analyze search results.

The Neighborhood Report shows users the words surrounding a searched word or phrase and can show the most common neighbors to the searched word. For more details on accessing and interpreting this tool, see [Using Neighborhood Report](#).

Using WordCruncher for Corpus Analysis

This section includes five instruction sets for helpful actions in creating and using corpora. The first set is about using the Indexer program to create “books” that can be opened and used as corpora on WordCruncher. This step must be completed before any of the others can be used, but the rest of the instruction sets can be applied in any order as needed in your research. The other instruction sets include searching by part of speech, applying restrictions on certain

attributes to a search, and carrying out analyses by creating the Neighborhood Reports and Phrase Compare Reports.

Indexing Files to Create Books on WordCruncher

The WordCruncher Indexer program converts .xml (2003), .txt, .rtf, and .etax files into both .etbu and .etax formats, which are compatible with the main WordCruncher software. Files must be run through the indexer before they can be opened as searchable corpora, or books, in WordCruncher. These instructions will explain how to prepare and convert text files with the WordCruncher Indexer and how to add and open the converted files as books in WordCruncher.

Indexing Files with WordCruncher Indexer

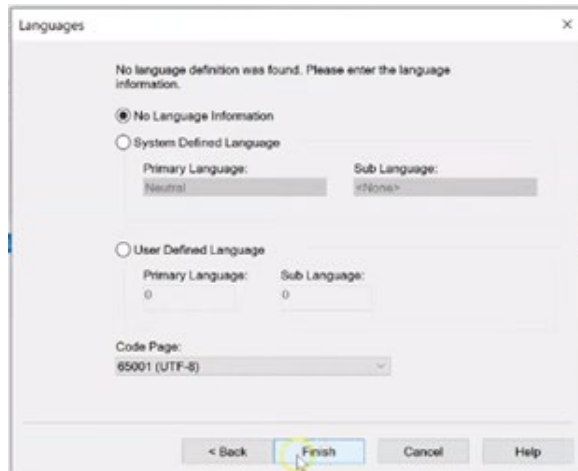
1. Save your document as a .xml (2003), .txt, .rtf, or .etax file using a text editing program, such as Microsoft Word.

Note: For easy navigation if you have many chapters or sections within WordCruncher, follow step 1 for each chapter or section in its own text file.

2. Convert the file to .etbu and .etax files using the WordCruncher Indexer.
 - a. Open the WordCruncher Indexer.
 - b. Click on the **Index** button and then select your saved .xml (2003), .txt, .rtf, or .etax file. The Indexing Information dialogue box will pop up.
 - c. Click the **Next** button.
 - d. If your text is in the same language as your computer's default settings, leave the "No Language Information" bubble as your selection. If the text is *not* in the same language as your computer's default settings, click the bubble for "System Defined Language" and choose the correct language from the **Primary Language** drop-down menu.

Note: If you are indexing multiple chapters for easy navigation as per the note in step 1, when indexing the files (step 2), select all of the text files by clicking the last chronological file, holding the Shift key on your keyboard, and selecting the first chronological file. This will select those two files and all that come in-between in the proper order. Select the **Yes** button on the dialogue box asking "Would you like to append the output to the same file?" On the next dialogue box, check the box for "Use these settings for all files that will be indexed." All selected files will save together as one .etbu and one .etax file called "AppendedOutput." Follow the rest of the steps as stated below.

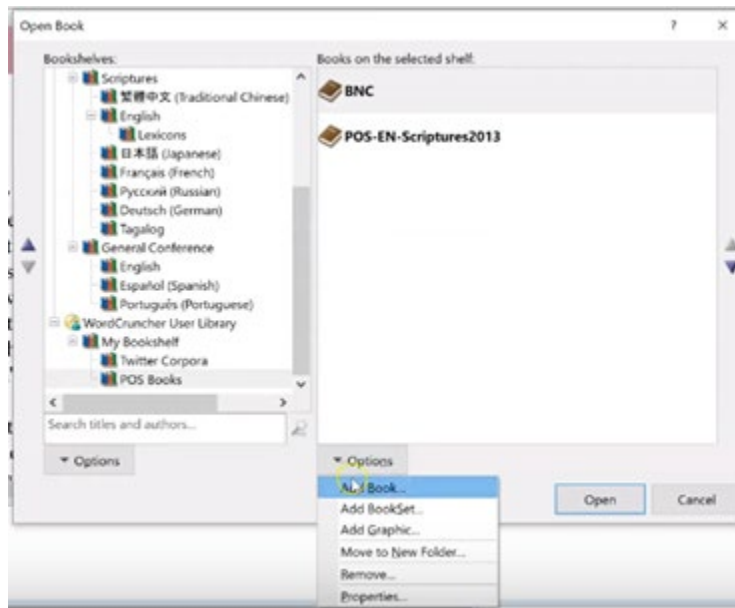
- e. Select the **UTF-8** option from the **Code Page** drop-down menu unless you know you need a different language's character code, in which case, select the appropriate option.



- f. Click the **Finish** button. There will now be an .etbu and an .etax file saved to the same place on your computer as the original text file.

Opening Indexed Files as Books on WordCruncher

1. Add the new .etbu file as a book on WordCruncher.
 - a. Open WordCruncher.
 - b. Click on the **File** menu > **Open Book**. The **Open Book** window will pop up.
 - c. Click **Options** > **Add Book** under the “Books on the selected shelf” selection box on the right half of the window.



- d. Select the .etbu file of your text. You can either double click on the file name or click on the file name and click the **Open** button
2. Open your file as a book in WordCruncher by selecting your text on the default **Open Book** window. You can either double click on the file name or click on the file name and click the **Open** button.

You have now added your original text as a book on WordCruncher. Each text you convert and download as a book can be used as a searchable corpus on your computer. Now that this initial step is completed, the following instruction sets for possible types of analysis and study can be applied as desired.

Searching using Part of Speech

The Part of Speech tool in WordCruncher is an addition to the Search bar that enables you to narrow your search based on parts of speech. This set of instructions outlines how to open the search bar, search for all words in a specific part of speech, and search for a word with a specific part of speech.

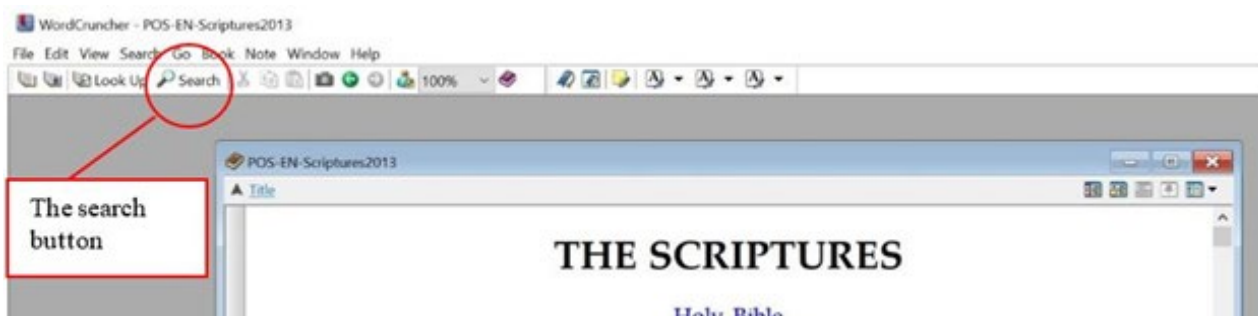
Note: WordCruncher supplies two books in its library that are already tagged for parts of speech: the BNC (British National Corpus) and the POS-EN-Scriptures2013 (Scriptures tagged with parts of speech). WordCruncher does not yet have the capability to tag other documents for parts of speech. This must be done in a separate program.

Opening the Part of Speech Option

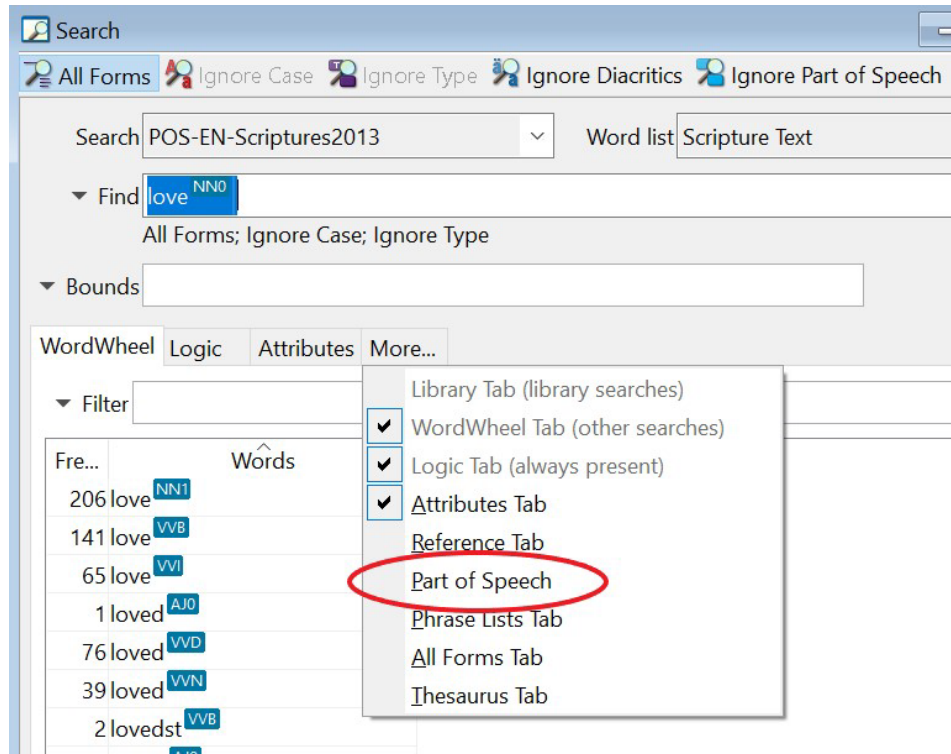
1. Open WordCruncher and ensure that the book you have open to analyze is already tagged for parts of speech.

Note: If you don't know how to open or create your text file as a book in WordCruncher, see the section [Indexing Files to Create Books on WordCruncher](#) in this user guide.

2. Press on the spacebar or click the **Search** button on the toolbar to open the search window.



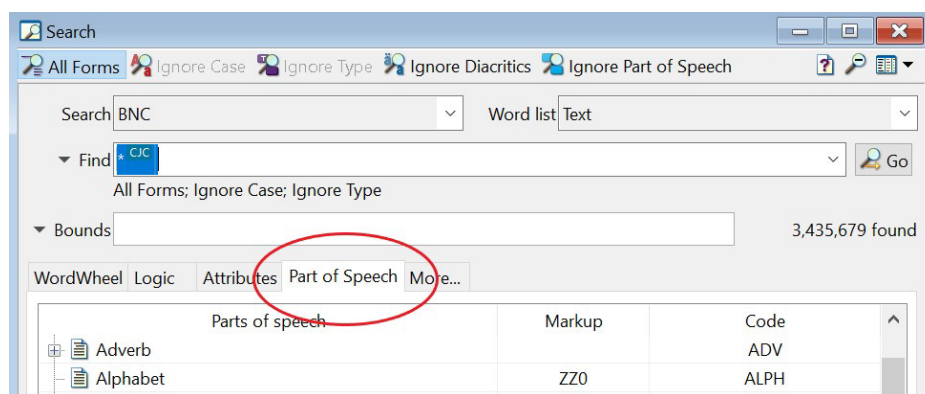
3. Click on the “More” tab and select the **Part of Speech** option from the drop-down menu. This step can be skipped if the “Part of Speech” tab already appears in your menu.
4. Click on the “WordWheel” tab to see a list of all the words contained in the document. A blue part-of-speech tag will appear to the right of each word. You can now search for all



words in a specific part of speech or search for a single word with a specific part of speech.

Searching for All Words in a Single Part of Speech

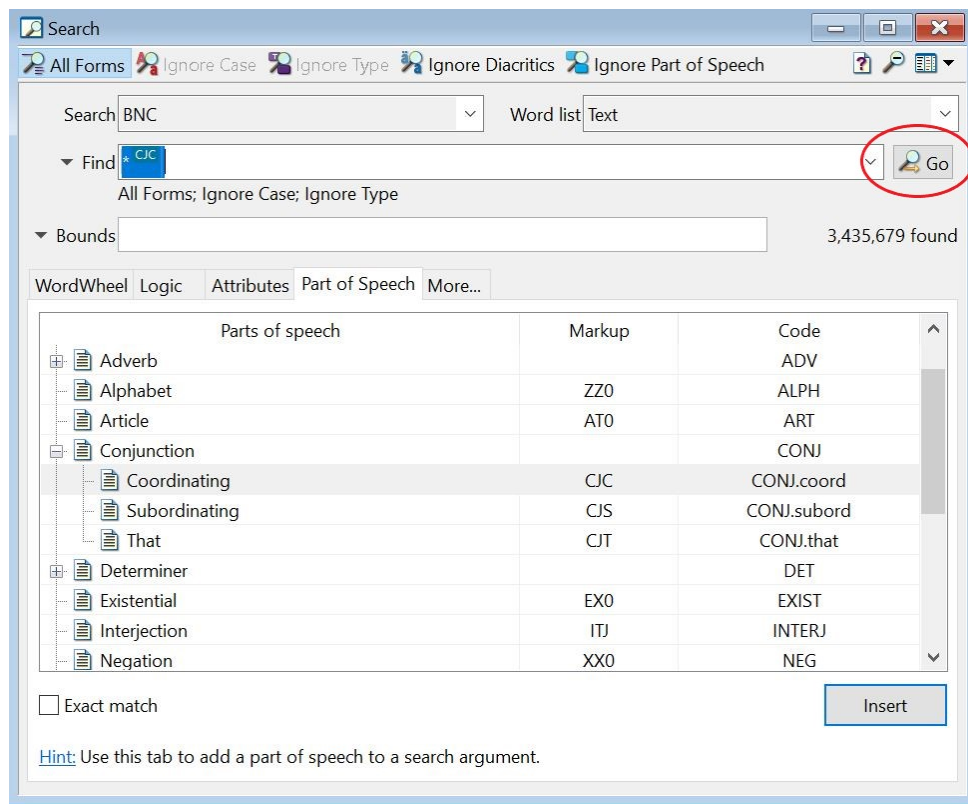
1. Click on the “Part of Speech” tab within the Search window. This will both allow you to select the part of speech to search for and provide a key to what each abbreviation and corresponding part of speech mean.



2. Click on the part of speech you want to search (e.g. Coordinating Conjunctions).
3. Click the **Insert** button in the bottom right corner of the window.

Note: You can add as many parts of speech as you want by repeating steps 1–3 (above).

4. Click the **Go** button to the right of the Find bar. The Search Results window will open with all instances of that part of speech in the book.



5. Search for a different part of speech by clicking on the blue tag in the search bar to select it and then click on the desired part of speech in the “Part of Speech” tab. This will replace the old part of speech with the new one. Or, click on the Find bar and press the backspace button on your keyboard to delete the blue tag in the Find bar. Repeat steps 1–4 to search for a new part of speech.

Search for a Word with a Specific Part of Speech.

1. Type the desired word into the Find bar. A drop-down menu will open with the typed word highlighted. A blue part-of-speech tag will appear to the right of each word.
2. Click the highlighted word with the desired part-of-speech tag beside it in the drop-down menu.

Note: Be careful to select the correct word from the drop-down menu. The same word with a different part-of-speech tag may also be an option on the drop-down menu.

3. Click on the **Go** button to search the document to open the Search Results window with each instance of the searched-for word or part of speech.

If you have correctly followed these instructions, you should now have successfully searched for a part of speech in the WordCruncher application. You should also be able to open the search

bar, search for all words in a specific part of speech, and search for a word with a specific part of speech.

Using Attribute Search

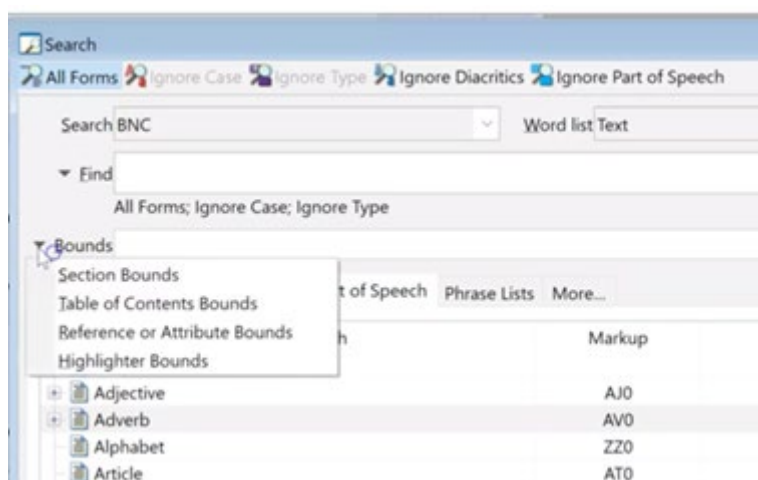
The Reference or Attribute Bounds tool allows you to limit your search results. These instructions will tell you how to locate the Attribute Bounds tool and apply different bounds to your search.

Opening the Attribute Bounds tool

1. Open your selected book or library.

Note: If you don't know how to open or create your text file as a book in WordCruncher, see the section [Indexing Files to Create Books on WordCruncher](#) in this user guide.

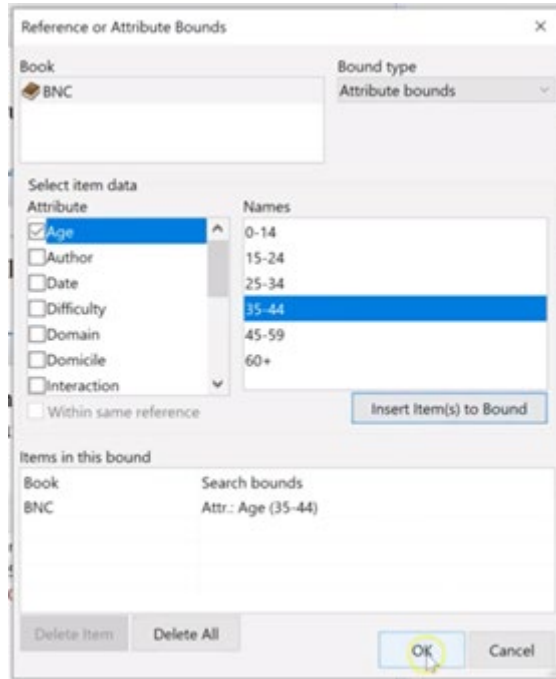
2. Click the **Search** button found in the toolbar. The Search window will appear.
3. Click the drop-down arrow beside the Bounds bar.



4. Select the **Reference or Attribute Bounds** option to open the Reference or Attributes Bounds dialogue box.

Using the Attribute Bounds Dialogue Box

1. Select an attribute and choose from the qualifiers that appear in the box in the middle-right of the window.
2. Select one or more of the qualifiers and click the **Insert Item(s) to Bound** button.
3. Click the **OK** button.



4. Complete any text search; the results will be bound by the selected attributes.

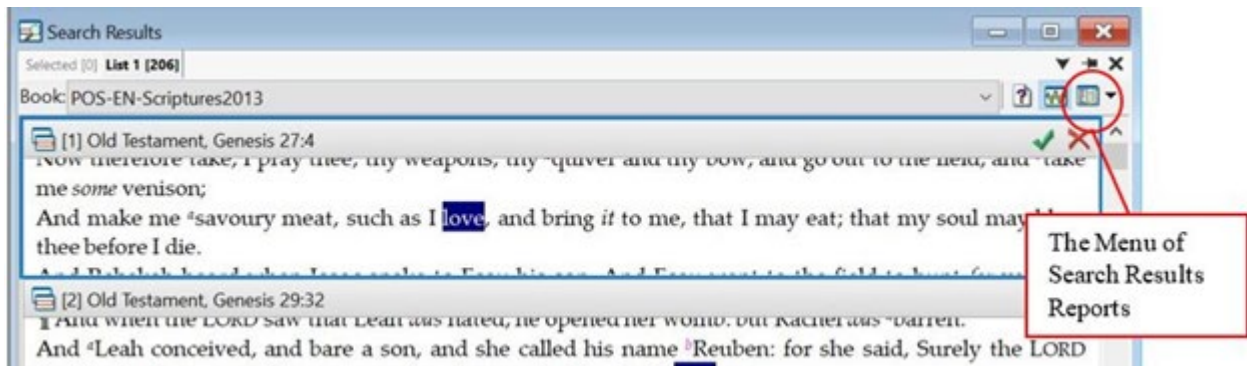
You have now finished using the Attribute Bounds function for your search. You can use a variety of attribute bounds to experiment with the search results.

Using Neighborhood Report

The Neighborhood Report tool is an analytical tool attached to the Search Results window. Clicking on the **Neighborhood Report** button will open a new window for analyzing “neighbors” (words that appear close to a searched word). To look at the Neighborhood Report, first you must search a word in a book (see the User Guide for General Users or [Searching Using Part of Speech](#) in this user guide). These instructions will guide you through the steps of finding and opening the neighborhood report for your search.

Opening the Neighborhood Report Window

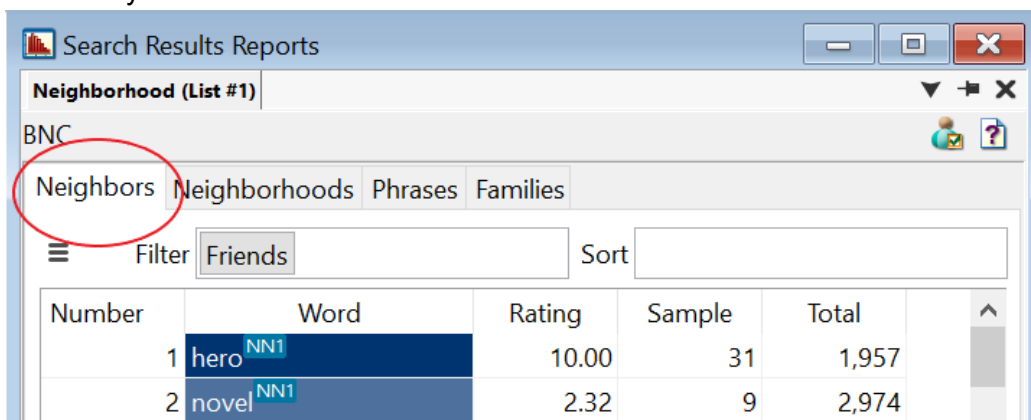
1. Search for a word in a document and open the Search Results window.
2. Click on the **Menu of Search Results Reports** button in the right corner of the Search Results window. This button will open the drop-down menu.



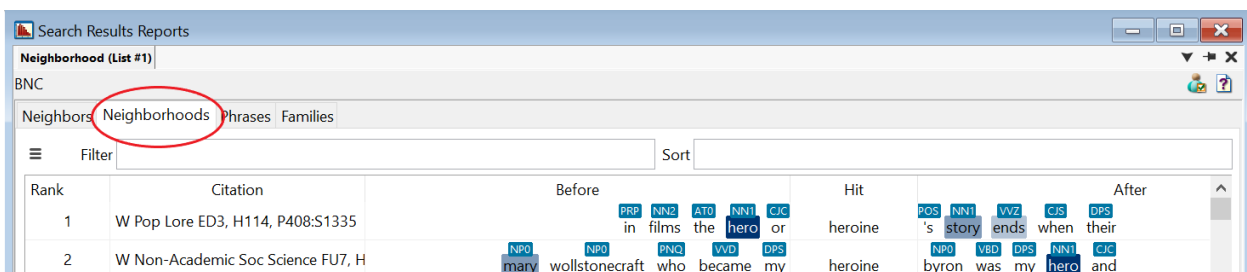
3. Click on the **Neighborhood Reports** option in the drop-down menu. This will open the Neighborhood Reports window.

Using the Tabs in the Neighborhood Reports Window

1. Click on the “Neighbors” tab. This will show you the words that commonly appear beside the word you searched for.



2. Click on the “Neighborhoods” tab. This will show you common words that appear up to ten words before and after the word that you searched for.



3. Click on the “Phrases” tab. This will show you common phrases that contain the word or phrase you searched.



4. Click on the “Families” tab. This will show you a list of words that commonly appear in the neighborhood of the word you searched for.



You have now opened the Neighborhood Reports window and accessed each of the tabs. You can use these tabs for analyzing individual words for a corpus.

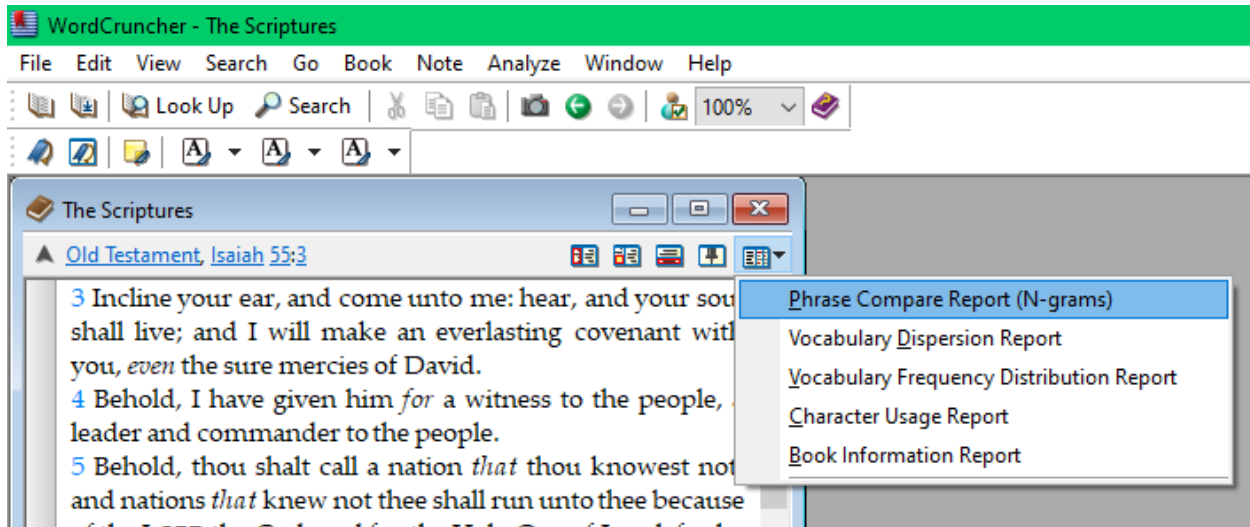
Creating a Phrase Compare Report

The Phrase Compare Report tool allows you to discover which phrases are the most common in a document, compare common phrases within or between documents, and search for phrases of specific lengths in a document. To use this tool, you will need to have at least one book loaded into your WordCruncher library. These instructions will go over opening the Phrase Compare Report window, running the Phrase Compare Report tool, and sorting and interpreting the tool’s results.

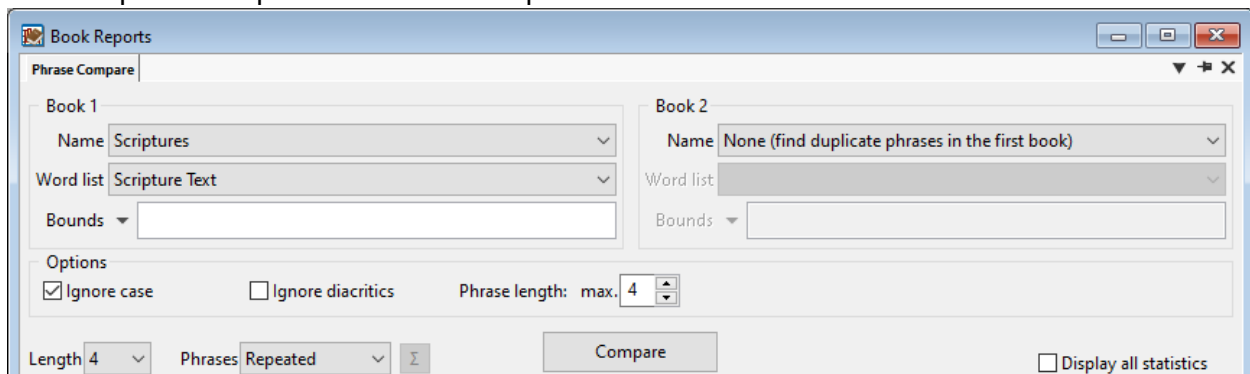
1. Open the books you want to compare from your library.

Note: If you don’t know how to open or create your text file as a book in WordCruncher, see the section [Indexing Files to Create Books on WordCruncher](#) in this user guide.

- Click on the **Menu of Book Reports** button in the upper right-hand corner of the Book window.



- Select the **Phrase Compare Report** option at the bottom of the drop-down menu. This option will open the Phrase Compare window.



- Choose the books to be compared in the Phrase Compare window.
- Click on the **Name** drop-down menu under “The first book.”
- Select the first book to be used in the comparison.
- Click on the **Name** drop-down menu under “The second book.” If you want to compare parts of the same book, select the same book for both “The first book” and “The second book.”
- Select the **Table of Contents** option in the **Bounds** drop-down menu to choose a specific section of the text for comparison.

Note: Leave this box as is if you want to compare the full text.

- Set your desired maximum and minimum phrase length using the arrow boxes above the **Compare** button.
- Click the **Compare** button to start the program.

You should now have several columns of results. You can sort these results by clicking on the button header of the desired output column. The **Freq.[1]** and **Freq.[2]** columns show how often each phrase appears in the first and second book, respectively. Use this to discover the most

common phrases in each book. The **Freq./M[1]** and **Freq./M[2]** columns normalize the counts in the first and second book, providing a better idea of each phrase's relative frequency. Use this to compare the density of a specific phrase between documents or sections. The **Len.** column sorts results by phrase length.

Book Reports

Phrase Compare

Book 1
 Name: Scriptures
 Word list: Scripture Text
 Bounds:

Book 2
 Name: None (find duplicate phrases in the first book)
 Word list:
 Bounds:

Options
 Ignore case Ignore diacritics Phrase length: max. 4

Length: 4 Phrases: Repeated Display all statistics

Num	Len	Phrase	1.Freq	2.Freq	1.RFreq	2.RFreq	1.Exp	2.Exp	LL	BIC
1	4	"it came to pass"	1,871	0						
2	4	"and it came to"	1,581	0						
3	4	"came to pass that"	1,368	0						
4	4	"i say unto you"	725	0						
5	4	"the children of israel"	651	0						
6	4	"thus saith the lord"	517	0						
7	4	"of the lord and"	458	0						
8	4	"of the children of"	414	0						
9	4	"verily i say unto"	375	0						
10	4	"the lord thy god"	338	0						
11	4	"in the land of"	328	0						
12	4	"the word of the"	313	0						
13	4	"of the house of"	307	0						
14	4	"the lord of hosts"	300	0						
15	4	"word of the lord"	300	0						

Phrases: 123,088

Index

attributes, 12–13

BYU (Brigham Young University), 4

books

adding, 4–5, 7, 8

opening, 8, 9

corpus (corpora), 4, 6, 7, 9, 16

documents

preparation of, 4–5, 7–8

conversion of, 4–5, 7–8

.etax file, 5, 7–8

.etbu file, 5, 7–8

find bar, 11–12

Indexer, WordCruncher, 4–5, 7

indexing files, 7–8

neighborhood report, 6, 13–16

part of speech

search tool, 5, 9–12

tab, 6

tag, 6

phrase compare report, 16–17

search, 5–6

bar 5, 9, 11, 12

keyword, 5

part of speech, 5–6, 7–12

results window, 5, 6, 11, 12, 13, 14

WordWheel, 10

WordCruncher

© 1991–2020 Brigham Young University
Provo, Utah, USA. All rights reserved.