

# WordCruncher Tools Overview

Office of Digital Humanities

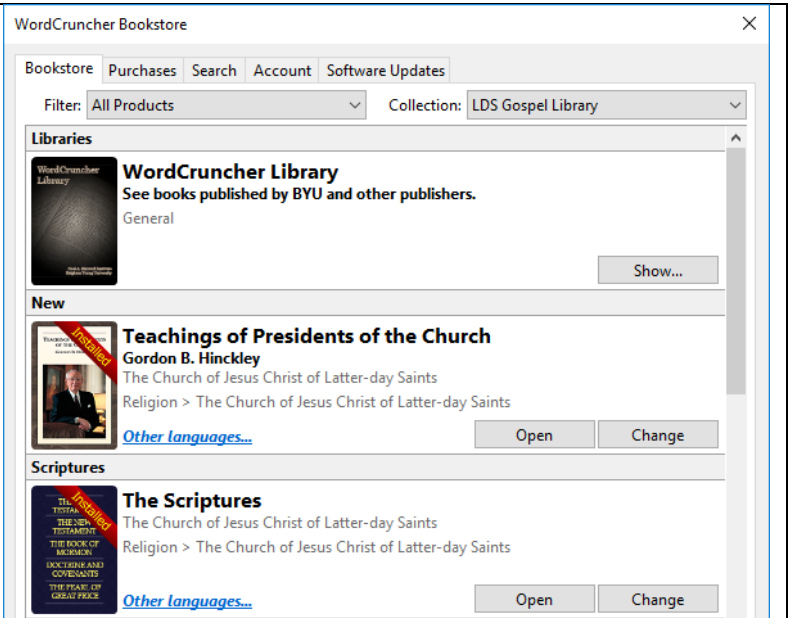
5 December 2017

WordCruncher is like a digital toolbox with tools to facilitate faculty research and student learning.

- Red text in small caps (e.g., **FILE VIEW**) refers to the names of tools found in *AntConc* (an app for corpus linguistic research).
- The tools described below are available in the Windows version for PCs or Macs. The iOS version currently lacks many of the analysis tools described below. A BYU team that is now in the Office of Digital Humanities developed WordCruncher. WordCruncher is a free download. Go to [www.WordCruncher.com](http://www.WordCruncher.com) for instructions.

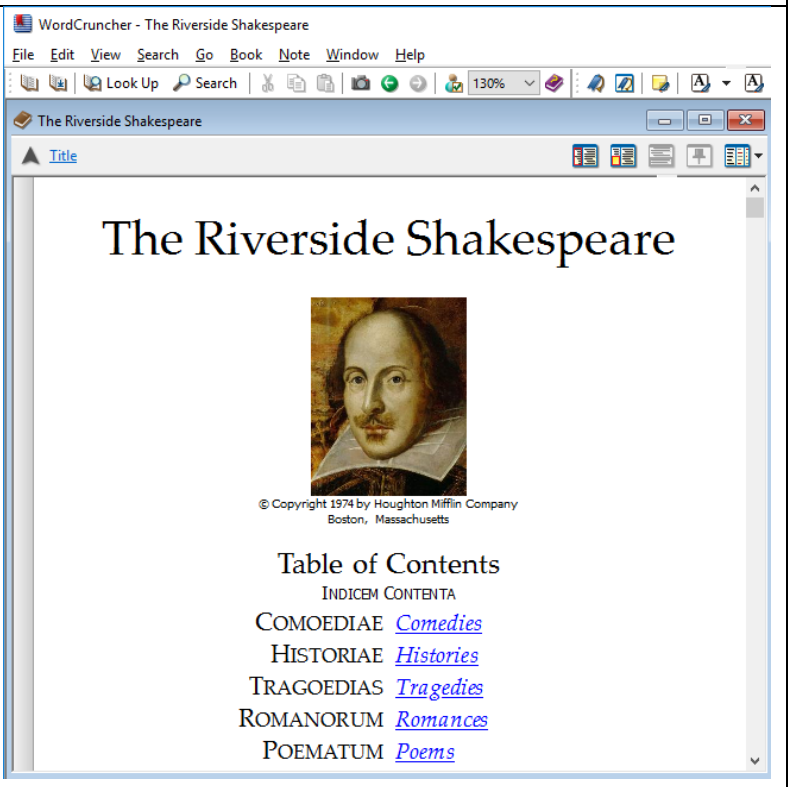
## WordCruncher Library

- **Download an eBook or corpus** from the WordCruncher Bookstore.
  - LDS Gospel Library (Scriptures in 30 languages including Asian and Near Eastern languages, conference talks, and some manuals published by the Church)
  - WordCruncher Library (Religious and academic books published by BYU and other publishers)
- **Create your own WordCruncher eBook or corpus** from txt, docx, or other formats with the Publishing Toolkit.  
(Note: Our team can help prepare eBooks and corpora for your research or teaching needs.)
- **Share your eBooks or notes** with students, colleagues, or the public. Bookstore eBooks are free, rented, or sold to the public or to authorized users.



## Reading and Presentation Tools

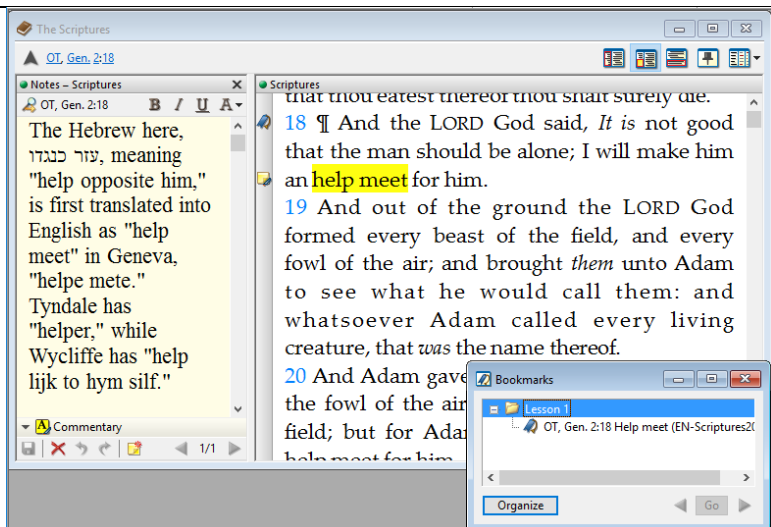
- **Text Window (FILE VIEW)**
  - Formatted text
  - Images (e.g., maps, photos, charts, illustrations) can be enlarged.
  - Hyperlinks (internal and external)
  - Citation bar to see and jump to a location (e.g., 1 Ne. 3:7).
  - Copy and paste selected text.
- **Zoom or Font size** lets you increase or decrease the text to suit the audience. Larger text helps when using a projector for a group presentation.
- **Reader Bar** highlights line with the cursor to facilitate reading and discussion.
- **Night mode** makes the screen less glaring
- **Go > Bookmarks**: Jump to bookmarks in one or more books during a talk, lesson, or presentation.
- **Go > Snapshots** will reopen eBooks, resize windows, and reposition windows to saved locations.



## Annotation Tools

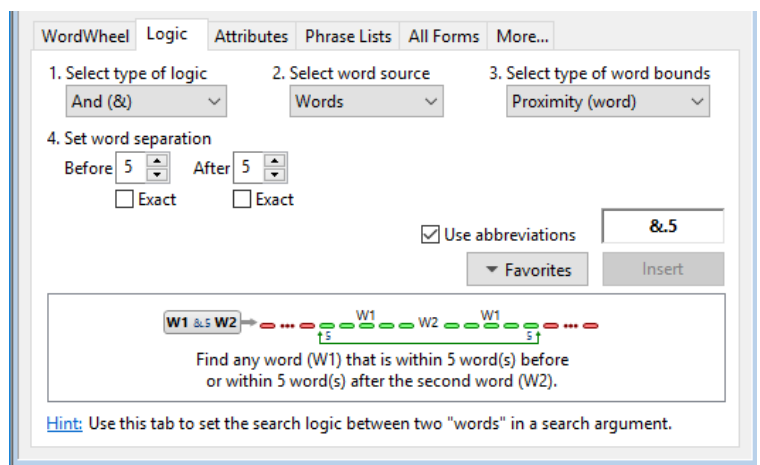
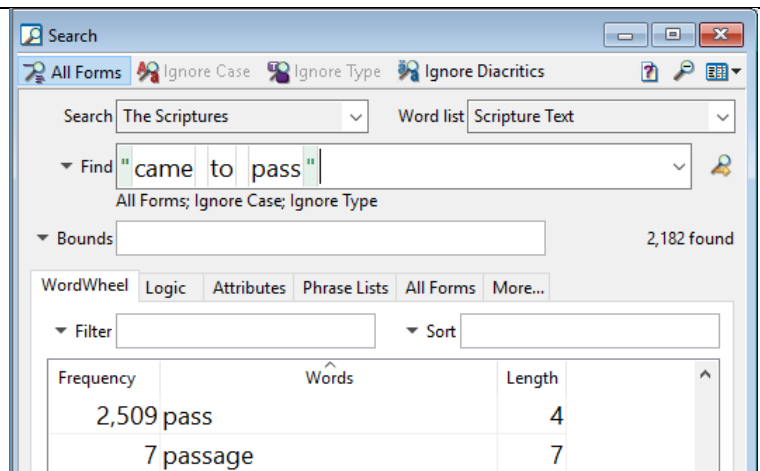
You can highlight, assign topics, and create notes for selected text. You can also add multiple bookmarks.

- **Highlight Text** lets you highlight and assign topics to selected text with several colors.
  - Topics (e.g., commentary, quote, variant) can have different highlighting.
- **Create Note** lets you can type or paste multi-lingual text into a note, and assign topics.
  - You can search your notes and share them with colleagues, friends, and students.
- **Add Bookmark** lets you add a bookmark at a location.
  - You can organize bookmarks using folders (e.g., lesson 1).
  - You can jump to bookmarks in one or more books during a presentation.



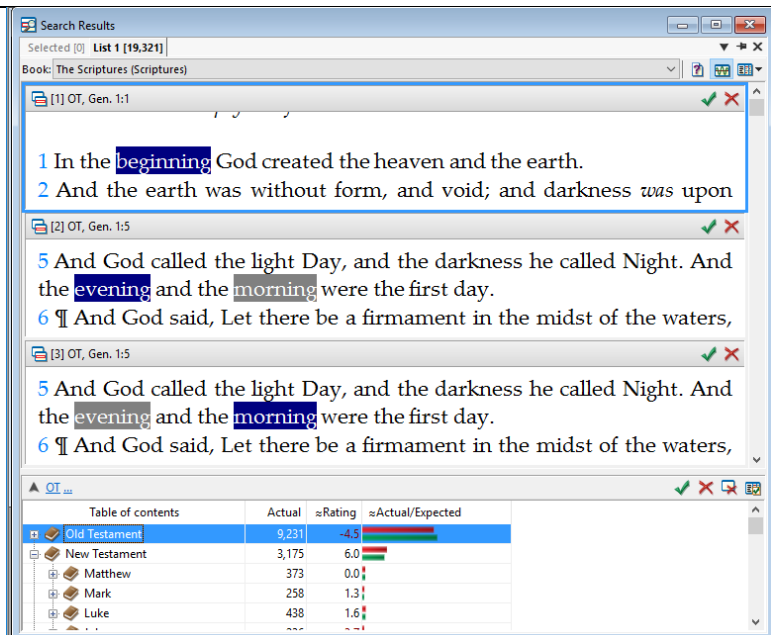
## Search Tools

- Search options
  - All Forms,<sup>1</sup> or
  - Ignore Case, Ignore Type, Ignore Diacritics
- **Search** identifies the book or books to search.
- **Word list** identifies the type of content (e.g., text, headings) to search.
- **Find**<sup>2</sup> lets you do simple or advanced searches.
  - *came pass*
  - "came to pass"
  - 'came pass'
  - \*ing [e.g., being, beginning, morning]
  - ??ing [e.g., being, bring, cling, thing]
  - *came &.5 pass*
  - '(keep + fulfill) commandments'
- **Bounds** limits the search to part of a book (e.g., New Testament).
- **WordWheel Tab (WORD LIST)** shows a list of words and their frequencies in a book.
  - Sort by clicking on a column heading.
  - Filter the list (e.g., \*ing, frequency > 30).
  - Search for selected words in the list.
- **Logic Tab** shows advanced searching options.
  - Select options from menus.
  - See an image showing the selected logic.
  - See how to insert or type the logic.
  - Create shortcut keys to favorite or frequently used logic.



## Search Result Tools (CONCORDANCE)

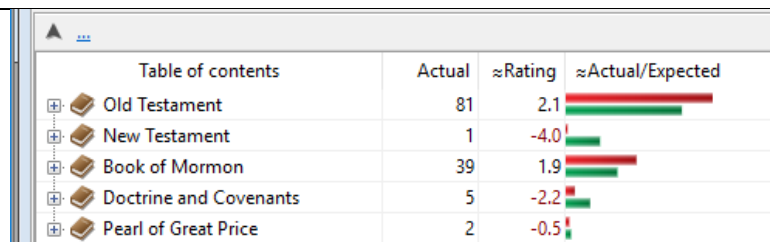
- View all search hits in context
  - Change the window size to see more or less context.
- Double-click on any context to see more context in a Text Window.
- Select or delete individual search hits in the list.
  - The Selected Tab shows the selected contexts.
- Frequency Distribution** shows where the hits occur (e.g., New Testament, Genesis).
  - Click on a book (e.g., Matthew) to see the first search hit in that book.
  - Select or delete all hits in a book or chapter.



## Analysis Tools: Search Results Window

**Frequency Distribution (CONCORDANCE PLOT<sup>3</sup>)** shows how many hits are in each book and chapter.

- The green line indicates the relative size of each book.
- If the Rating is positive and the red line is longer than the green line, there are more hits than would be expected in a chance distribution.



## Search Vocabulary Report

This report shows all search words generated by search (e.g., forms, wildcards).

- Select this report from the Search menu or from report menu button in upper right corner of the Search Results window.
- Sort by one or more columns.

Num.	Word	Frequency
1	atone	5
2	atoned	2
3	atonement	112
4	Atonement	1
5	atonements	1

## Neighborhood Report

This report shows words that occur near the search word(s) and a 'frequent neighbor' rating.

- Top section** shows search words in the center of neighboring words. Highlighting indicates rating in bottom section. (CONCORDANCE, CLUSTERS<sup>4</sup>)
  - Sort by one or more columns.
  - Look for patterns.
- Middle section** shows a scrollable text window for the selected reference.
- Bottom section** shows the neighboring words (COLLOCATES) sorted by a rating from 10 to -10.
  - Sort by one or more columns.
  - Double-click on a neighbor to see the contexts in the top section.
  - Family Tab shows sets of frequent neighbors.

Rank	Citation	-5	-4	-3	-2	-1	Hit	+1	+2	+3	+4	+5
9	OT, Lev. 15:15	the	priest	shall	make	an	atonement	for	him	before	the	LORD
10	OT, Lev. 4:31	the	priest	shall	make	an	atonement	for	him	and	it	shall
11	OT, Lev. 14:31	the	priest	shall	make	an	atonement	for	him	that	is	to
12	OT, Lev. 4:26	the	priest	shall	make	an	atonement	for	him	as	concern	his
13	OT, Lev. 5:13	the	priest	shall	make	an	atonement	for	him	as	touchin	his
14	OT, Lev. 5:18	the	priest	shall	make	an	atonement	for	him	concern	his	ignoranc

Old Testament, Leviticus 15:15

15 And the priest shall offer them, the one for a sin offering, and the other for a burnt offering; and the priest shall make an atonement for him before the LORD for his issue.

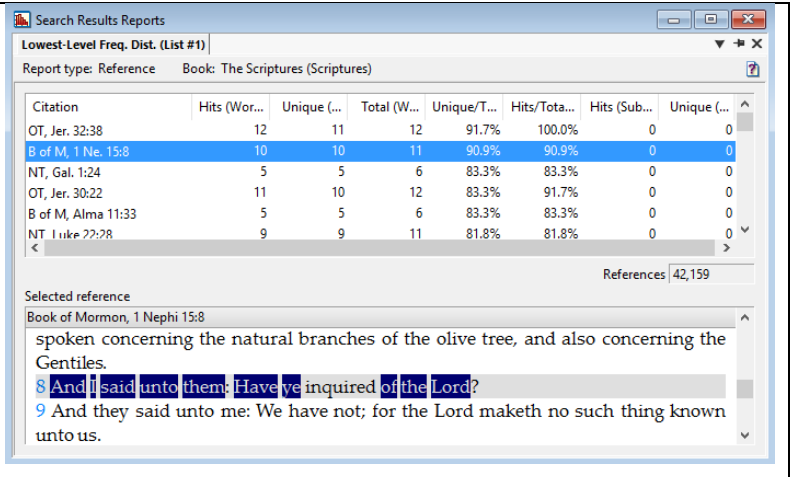
16 And if any man's seed of copulation go out from him, then he shall wash all his flesh in water,

Word	Word List	Rating	Sample	Total
make	Scripture Text	10.00	66	1,261
an	Scripture Text	10.00	72	2,257
infinite	Scripture Text	8.11	4	16
priest	Scripture Text	7.16	21	563

## Lowest Level Frequency Distribution Report

If you search for the 50 most frequently used words or all words ending in *ing*, you can find verses or paragraphs that contain a high percentage of these search words. This has been used to find examples for language learning.

- **Top section** shows several columns (e.g., number of hits per verse or paragraph).
  - Sort by one or more columns.
- **Bottom section** shows a scrollable text window for the selected reference.



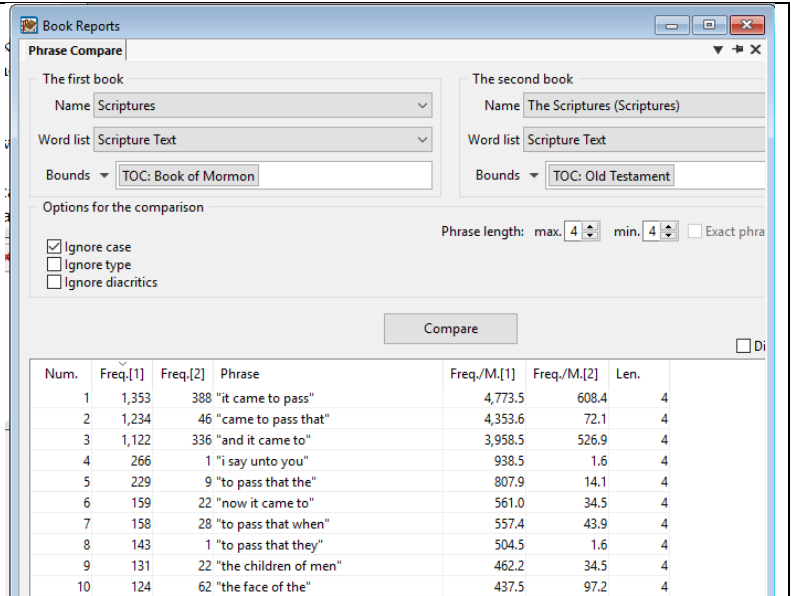
Citation	Hits (Wor...)	Unique (...)	Total (W...)	Unique/T...	Hits/Tota...	Hits (Sub...)	Unique (...)
OT, Jer. 32:38	12	11	12	91.7%	100.0%	0	0
B of M, 1 Ne. 15:8	10	10	11	90.9%	90.9%	0	0
NT, Gal. 1:24	5	5	6	83.3%	83.3%	0	0
OT, Jer. 30:22	11	10	12	83.3%	91.7%	0	0
B of M, Alma 11:33	5	5	6	83.3%	83.3%	0	0
NT Luke 22:28	9	9	11	81.8%	81.8%	0	0

## Analysis Tools: Text Window

### Phrase Compare Report (N-GRAMS)

This report helps you find common expressions or phrases in a book, in two books, or in two different sections of a book.

- Select this report from the Book menu or from the report button in the upper right corner of a Text window.
- A normalized frequency lets you compare frequencies in different books or sections of different sizes.
- You can set the minimum and maximum phrase length.
- Sort by one or more columns by clicking on a column heading.

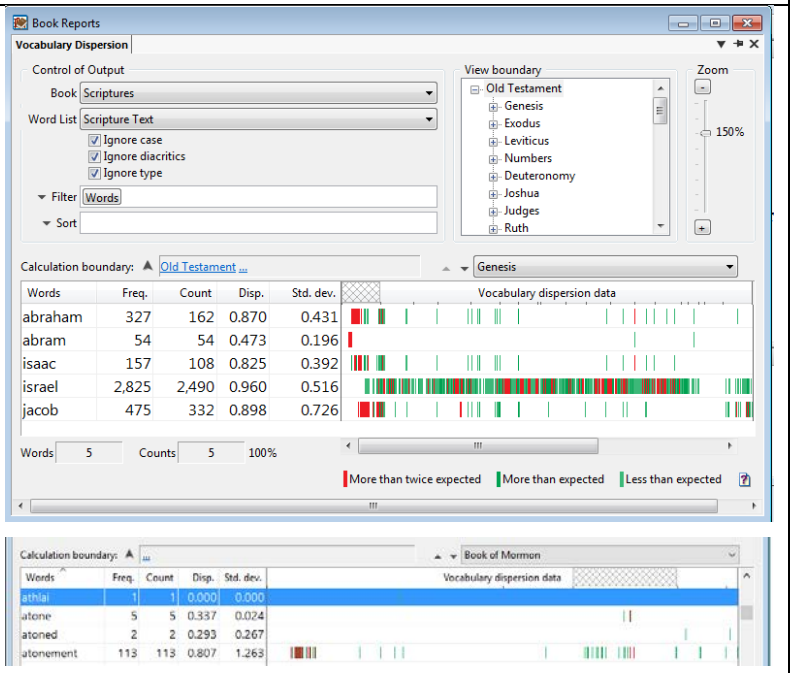


Num.	Freq.[1]	Freq.[2]	Phrase	Freq./M.[1]	Freq./M.[2]	Len.
1	1,353	388	"it came to pass"	4,773.5	608.4	4
2	1,234	46	"came to pass that"	4,353.6	72.1	4
3	1,122	336	"and it came to"	3,958.5	526.9	4
4	266	1	"i say unto you"	938.5	1.6	4
5	229	9	"to pass that the"	807.9	14.1	4
6	159	22	"now it came to"	561.0	34.5	4
7	158	28	"to pass that when"	557.4	43.9	4
8	143	1	"to pass that they"	504.5	1.6	4
9	131	22	"the children of men"	462.2	34.5	4
10	124	62	"the face of the"	437.5	97.2	4

### Vocabulary Dispersion Report (CONCORDANCE PLOT<sup>5</sup>)

This report shows a 'barcode' plot for all words in a book. This shows where words cluster in the text.

- Sort by one or more columns.
- Filter words in the list to see specific words or types of words (e.g., *ing* words).



Words	Freq.	Count	Disp.	Std. dev.
abraham	327	162	0.870	0.431
abram	54	54	0.473	0.196
isaac	157	108	0.825	0.392
israel	2,825	2,490	0.960	0.516
jacob	475	332	0.898	0.726

## Vocabulary Frequency Distribution Report

This report shows how many times each word occurs in a book and in sections of a book.

- A normalized frequency for each section lets you compare frequencies as if the sections were all the same length.
- Sort by one or more columns.
- Filter words in the list to see specific words or types of words (e.g., *ing* words).

Num.	Words	Frequency	A: Book of Mormon	Ab ratio	B: Old Testament
1	lamanites	710	2,463	698	2,463.0
2	nephi	484	1,672	474	1,672.0
3	christ	1,083	1,358	385	1,358.0
4	nephites	381	1,319	374	1,319.0
5	has	472	1,034	293	1,034.0
6	alma	265	935	265	935.0
7	—s	863	1,619	459	809.5
8	church	743	801	227	801.0

## Character Usage Report

This report shows how many times each character occurs in a book.

- Find unmatched pairs of parentheses or brackets.
- Characters that occur rarely may be errors.
- Double-click on a character to see where it occurs in the text.
- Sort by columns.

Character	Code	Frequency	Percent
&	U+0026	8	< 0.01%
(	U+0028	449	< 0.01%
)	U+0029	449	< 0.01%
*	U+002A	22	< 0.01%
,	U+002C	108,719	1.59%
-	U+002D	1,517	0.02%
.	U+002E	40,409	0.59%
/	U+002F	1	< 0.01%
0	U+0030	649	< 0.01%

## Analysis Tools: Search Window

### WordWheel (WORD LIST)

This is a list of words in a book or set of books based on the selected search options (e.g., Ignore Case, Word list).

- It can be filtered to see only part of the list (e.g., high frequency words, exclude punctuation).
- Sort by frequency, length, or Z-score as you click on column headings.
- See unique and total word counts for the selected search options and filters.

Frequency	Words	Length
93,579	the	3
76,346	and	3
52,165	of	2

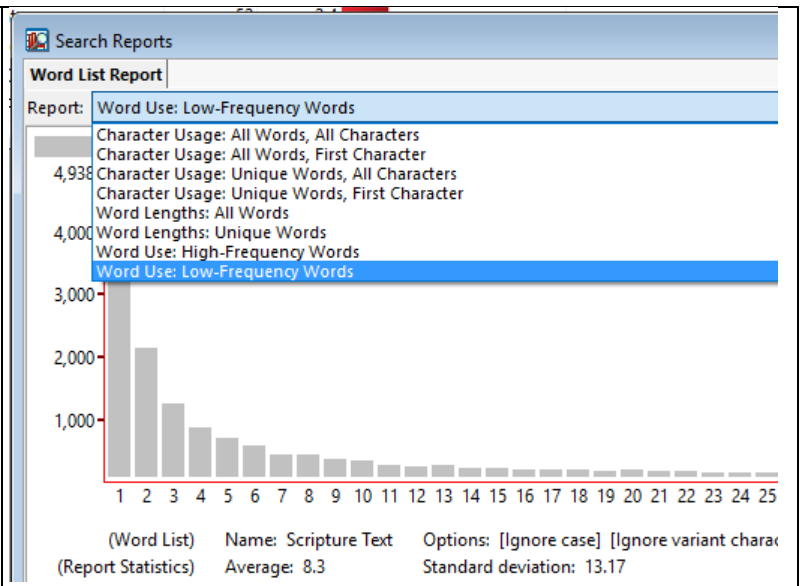
  

Frequency	Words	Length
4	maher-shalal-hash-baz	21
1	jonath-elem-rechokim	20
1	bashan-havoth-jair	18
4	chushan-rishathaim	18
1	spiritually-minded	18

## Word List Reports

These reports show graphs of word frequencies or length, and of character frequencies in words.

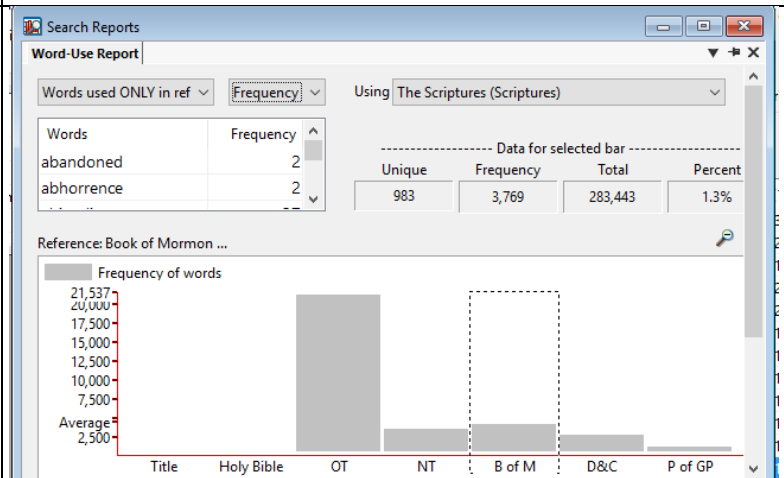
- Select this report from the Search menu or from the report button in the upper right corner of a Search window.
- Select one of 8 charts to see character and word patterns.



## Word Use Reports

Select one of four reports to see a graph and a list of words in the selected column.

- See words used only in or used first in a section.
- See results by reference (e.g., book, chapter) or by decile (e.g., 10 equal size sections).
- Click on a column to see the corresponding words.
- Double-click on a column to expand it.



## Paired Cognates Report

Look for words (e.g., covered, covering) with the same *lemma* or dictionary form (e.g., cover) close together (e.g., “they were *covered* with a shallow *covering*”).

- Select All Forms to include a lemma lexicon.
- Set WordWheel filters and search bounds.
- **Top section:** Set report options, see search filters and options, and click on Find cognates button.
- **Middle section:** See the cognates and frequency.
- **Bottom section:** See a scrollable text window for the selected cognates.

Book name: The Scriptures

Cognate search logic: Word separation: Before 4, After 0, Exact  Exact

Do not show identical pairs

Find cognates

First word	Cognate	Frequency
offer	offered	1
offer	offering	53
offered	offering	18

Paired cognates count: 1,169

Selected cognates pair context (1 / 53)

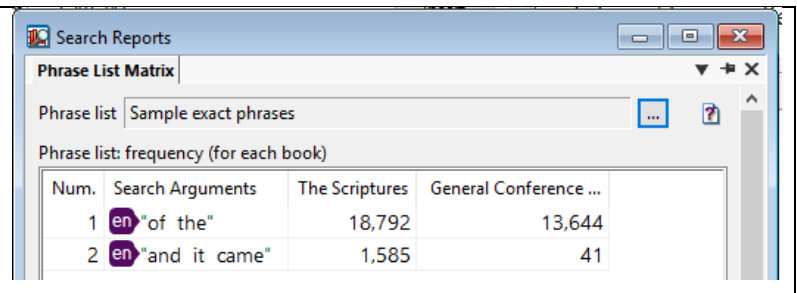
Old Testament, Exodus 35:24

them.  
24 Every one that did offer an offering of silver and brass brought offering: and every man, with whom was found shittim wood

## Phrase List Matrix Report

A phrase list can include words or phrases you want to find in more than one book.

- Compare how often each ‘phrase’ in a saved list occurs in one or more books.
- You can see actual raw frequencies or normalized frequencies.



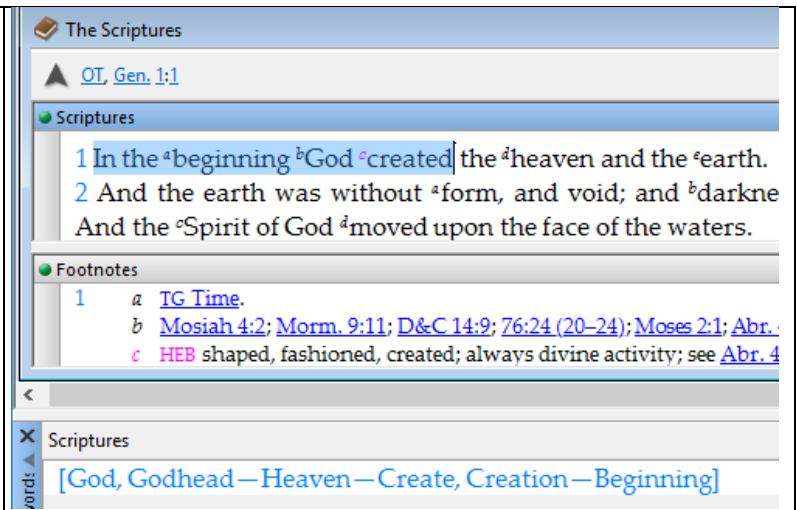
The screenshot shows a window titled "Search Reports" with a tab for "Phrase List Matrix". The "Phrase list" is "Sample exact phrases". Below, a table shows the frequency of phrases in "The Scriptures" and "General Conference ...".

Num.	Search Arguments	The Scriptures	General Conference ...
1	en "of the"	18,792	13,644
2	en "and it came"	1,585	41

## Study Resources and Tools

Study helps can make it easy to find related information to help users understand a text.

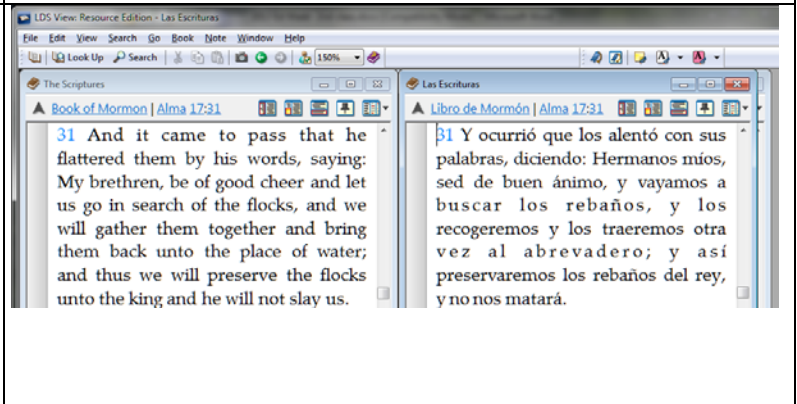
- **Footnotes** might include information or links to maps, photos, or other resources.
- **Images** like maps, photos, or charts can be enlarged to see more detail.
- Hidden **topics** can be added like those shown in the bottom part of the window. Topics can be searched separately.



The screenshot shows the "The Scriptures" window with the text: "1 In the <sup>a</sup>beginning <sup>b</sup>God <sup>c</sup>created the <sup>d</sup>heaven and the <sup>e</sup>earth. 2 And the earth was without <sup>a</sup>form, and void; and <sup>b</sup>darkness. And the <sup>c</sup>Spirit of God <sup>d</sup>moved upon the face of the waters." Below the text are footnotes: "1 a **TG Time.** b **Mosiah 4:2; Morm. 9:11; D&C 14:9; 76:24 (20-24); Moses 2:1; Abr. 4** c **HEB shaped, fashioned, created; always divine activity; see Abr. 4**". At the bottom, a "Topics" list includes "[God, Godhead—Heaven—Create, Creation—Beginning]".

## Synchronize translations or commentaries

You can open the LDS Scriptures in about 30 languages including some Asian and Near Eastern languages. If you open the scriptures in two or more languages, you can look up a verse in one language and the other language windows will automatically synchronize or show the same verse (unless you pin a window). This facilitates comparisons. For example, when you see *LORD* or *host of heaven* in the English Bible, you will see *Jehová* or *ejército* [army] *del cielo* in Spanish.

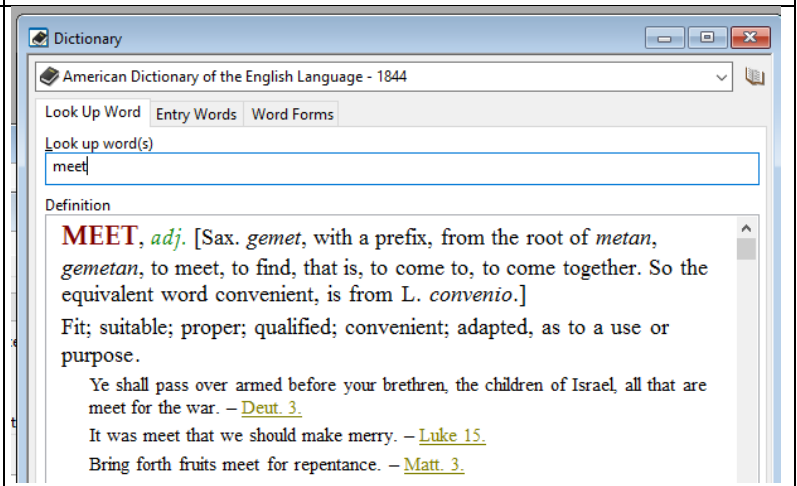


The screenshot shows two windows side-by-side. The left window is titled "Book of Mormon | Alma 17:31" and shows the English text: "31 And it came to pass that he flattered them by his words, saying: My brethren, be of good cheer and let us go in search of the flocks, and we will gather them together and bring them back unto the place of water; and thus we will preserve the flocks unto the king; and he will not slay us." The right window is titled "Libro de Mormón | Alma 17:31" and shows the Spanish text: "31 Y ocurrió que los alentó con sus palabras, diciendo: Hermanos míos, sed de buen ánimo, y vayamos a buscar los rebaños, y los recogeremos y los traeremos otra vez al abrevadero; y así preservaremos los rebaños del rey, y no nos matará."

## Webster's 1844 English Dictionary

If you download Webster's dictionary from the Bookstore, you can click on any English word in any text and see the definition in a Dictionary window.

- Other dictionaries in English or other languages could be converted to a WordCruncher format.



The screenshot shows a window titled "Dictionary" with the subtitle "American Dictionary of the English Language - 1844". The "Look Up Word" field contains "meet". The definition is: "**MEET**, *adj.* [Sax. *gemet*, with a prefix, from the root of *metan*, *gemetan*, to meet, to find, that is, to come to, to come together. So the equivalent word convenient, is from L. *convenio*.] Fit; suitable; proper; qualified; convenient; adapted, as to a use or purpose. Ye shall pass over armed before your brethren, the children of Israel, all that are meet for the war. — [Deut. 3](#). It was meet that we should make merry. — [Luke 15](#). Bring forth fruits meet for repentance. — [Matt. 3](#)."

## English Parallel Bible

The LDS Bible has HEB or GR footnotes that give alternate translations for about 1% of Bible words. This Parallel Bible lets you see the KJV translation and a definition for all words.

- See Hebrew and Greek words (אלהים, Χριστός) or transliterated words ('ēlōhîm, Christós).
- Click on a Hebrew or Greek word to see a brief or longer definition.
- Double-click on a word to see every place the word occurs.
- See the LDS View “Getting Started Tutorial” in the Help menu for more information.

## Hebrew and Greek dictionaries

Strong's Hebrew and Greek Lexicons are included with the *English Parallel Bible*.

- Click on a Hebrew or Greek word (e.g., 'ēlōhîm), to see a brief definition and how the word was translated in the Bible.
- You will also see links to a pdf of *Brown Driver Briggs Hebrew and English Lexicon* (BDB), or *Thayer's Greek Lexicon* where you will see longer definitions.

## Hidden Tags

The Hebrew words in the Dead Sea Scrolls collection have hidden tags, i.e., lemma or dictionary form of the word, part of speech information, and a brief English definition.

- As your mouse hovers over a word, the hidden tags appear in the Tagword window (at the bottom).
- You can search lemmas, definitions, or part of speech tags.



## Corpus Linguistics

A corpus may include the collected writings of a single author organized by genre.

The 0.9 million-word Shakespeare corpus is organized by genre. The plays are referenced by act, scene, and line.

- In the example, *merry* is the search word.
- The green line represents the relative size of the text in each genre.
- When the red line (actual %) is longer than the green line (expected %) as in *Comedies*, the search word occurs more often than expected.

[3] Com - Err. II-i:88

87 *Adr.* His company must do his minions grace,  
 88 Whilst I at home starve for a merry look:  
 89 Hath homely age th' alluring beauty took

▲ Com ...

Table of contents	Actual	≈Rating	≈Actual/Expected
Comedies	79	3.5	
Histories	48	-0.1	
Tragedies	28	-2.7	
Romances	18	-0.9	
Poems	7	-0.7	

A corpus may include the writings of many authors organized by year.

The 19 million-word LDS Conference corpus (i.e., Journal of Discourses and Conference Reports) is organized by decade, year, talk, and paragraph.

- In the example, *atone* is the search word.
- The green line represents the relative size of the text in each decade. Before 1970, conference was usually three days instead of two.

Each WordCruncher corpus may read, annotated, studied, or analyzed with any WordCruncher tool.

[112] 1860s1860, ES Submission 1120 • JD 7:354

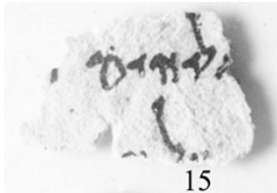
will also lead us in the same direction; and no atonement of our Lord and Savior Jesus Christ can alter that eternal law, any more than he can

▲ 1860s ... • JD 11:97

Table of contents	Actual	≈Rating	≈Actual/Expected
1830s	2	0.0	
1850s	107	-7.5	
1860s	45	-9.6	
1870s	77	-10.0	
1880s	70	-7.5	
1890s	17	-3.6	
1940s	36	-10.0	
1950s	181	-2.1	
1960s	152	-4.9	
1970s	180	-2.8	
1980s	309	10.0	
1990s	500	10.0	
2000s	692	10.0	
2010s	809	10.0	

The Dead Sea Scroll corpus is organized by the caves (e.g., 1Q, 4Q) where the scrolls were found, and by scroll, fragment, and line.

- In the example, אלהים (*Elohim*) is the search word.
- The green line represents the relative size of the text from each cave. Cave 4Q was the largest collection of texts.
- Click on a link to see and enlarge an [image](#) of a scroll.
- Hold the mouse over a word to see the [lemma](#), [part of speech](#), and [definition](#) in the tag window.



[3] 1Q1 (1QGen) Frg. 15:1 • 1Q1 f15:1

אלהים

Table of contents

Table of contents	Actual	≈Rating	≈Actual/Expected
1Q	211	4.4	
2Q	8	-2.1	
4Q	474	-3.1	
5Q	3	-1.4	
6Q	5	-0.9	
8Q	30	5.1	

Hebrew-Aramaic Texts (Biblical)

אלהים אלהים Heb NOUN comm masc plur God, gods

## Publisher Tools

### Publisher Toolkit

The WordCruncher Publisher Toolkit for Windows enables you to create your own WordCruncher eBooks and corpora. The toolkit includes:

- **Indexer** converts txt, Word docx, and other file formats into a WordCruncher eBook.
  - Check for duplicate reference levels
- **Image Library Manager** organizes images for a book into one file with labels for viewing, linking, and distribution.

Email [wordcruncher@byu.edu](mailto:wordcruncher@byu.edu) to request the *free* Toolkit and more information. Include your name, city, state, a short project description, and the current format of your documents, books, or collections.

### Proofing Tools

Documents often have typos or scanning errors. Spell check is a useful tool in word processors. You can use several WordCruncher tools described above to find and fix errors. Some tools below are only available to those with publisher permissions. Church translation supervisors have used these tools to check prepublication translations. Editors can use these tools to check prepublication books or documents.

- **Search Result Publisher Reports**
  - Check hyperlinks
  - Check synchronization
- **Text Window**
  - Check styles, formatting, etc.
- **Character Usage Report**
  - Check for spurious characters as well as mismatched parentheses, brackets, or quote marks.
  - Find and see any character in context.
- **WordWheel** filters (\*, \* ...)
  - Find common OCR or typing errors.
- **Word list**
  - Find words assigned to an incorrect style or language.
- Text > **Vocabulary Frequency Distribution Report**
  - Compare two sections of a book. Words that occur rarely in one, but never in the other may be typos.
- Search > **Phrase List Matrix Report**
  - Check translation of key words or phrases.

## Project Requests and Support

The WordCruncher team is now part of the Office of Digital Humanities. We provide the following services for BYU faculty and staff:

- Help install WordCruncher on Windows and Mac computers, and on iPhones and iPads.
- Help faculty and students use existing eBooks and corpora in offices, classrooms, and computer labs.
- Help create WordCruncher books and corpora to support faculty research and teaching.
- Train and help faculty, research assistants, and students use WordCruncher and the Publisher Toolkit. We provide this service in offices, labs, and classrooms.

### Availability

#### University Labs

WordCruncher is available in the following BYU computer labs:

- JFSB computer labs
- JSB computer lab for religion faculty and TA's.
- HBLL computer labs

---

<sup>1</sup> AntConc lets users load a lemma file when needed. WordCruncher ignores case and uses an Expansion Lexicon of lemmas or a stemmer.

<sup>2</sup> AntConc allows regex text searches (e.g., A[<sup>^</sup>p]+pass) within a search window of n characters. WordCruncher uses wildcard word searches that are not limited by a search window size.

<sup>3</sup> AntConc shows a 'barcode' plot. Frequency Distribution shows distribution by book, chapter, etc., and the Vocabulary Dispersion Report shows a 'barcode' plot for each word in a book.

<sup>4</sup> Clusters show hitword(s) followed by range of words. Neighborhood report shows two-sided neighbors like KWIC.

<sup>5</sup> AntConc shows 'barcode' plot for search words. Vocabulary Dispersion Report shows a 'barcode' plot for all words.