

tlCorpus Concordance Software

2021

User Guide

Professional Corpus Query Software

Version 12.0.15

Copyright © 2004-2021 TshwaneDJe Software

<https://tshwanedje.com/>

Table of Contents

General Notes to the User Guide	1
Frequently Asked Questions (FAQ).....	1
ChangeLog (Detailed List of New Features).....	2
Additional Resources.....	2
Terminology Used	1
Keeping Your Software Up To Date	2
Getting Started with tlCorpus: QuickStart Guide	3
Setting Up a New Corpus.....	3
Adding Texts to the Corpus.....	3
Adding Files Via Drag-And-Drop.....	3
Saving Your Corpus.....	3
Unsaved Changes Indicator *.....	4
Working With Multiple Corpora.....	4
Running Searches.....	4
Garbled Text?.....	4
PDF File Support	5
“Configure Corpus” Commands	6
Right-click Commands.....	6
Selecting Multiple Texts At Once.....	6
List of Right-Click Commands in Text List.....	6
Important Notes on the Text File “Encoding” [Intermediate to Advanced].....	7
Search	8
Tabbed Interface.....	8
Search Status Indicators (Tab Heading).....	8
Tab Selection Shortcut Keys.....	8
Closing the Current Tab.....	8
Closing All Open Searches.....	8
Search Right-click Commands.....	8
Selecting Multiple Lines At Once.....	8
List of Search Right-click Commands.....	9
Sort Order.....	10
Launching Multiple Search Queries At Once.....	10
“Sample” Option.....	10
Advanced Searches with Regular Expressions.....	10
Collocation Summary	12
Generating Wordlists	13
Wordlist Tool.....	13
Exporting The Wordlist To a File.....	13
Wordlist Right-click Commands.....	13
Overall Corpus Statistics.....	14
Generating a List of Occurring Characters	15
Calculating/Generating N-Grams	16
Character Bigrams.....	16
Word N-Grams.....	16
Web Crawler	17
Options.....	17
Output Folder.....	17

Monitor Clipboard.....	17
Don't Download URLs Already In Corpus.....	17
Automatically Add Downloaded Files To Corpus.....	17
Adding Multiple URLs at Once.....	17
Adding Multiple URLs at Once With Address Ranges.....	17
Right-to-left Languages.....	19
Font Problems.....	20

General Notes to the User Guide

tlCorpus is a software application for managing, querying and working with corpora (singular “corpus”). A corpus is an electronic database of actual texts, such as book, novels, poems, or news articles, in one or more languages. Concordance software represents a powerful tool to assist anyone who works with language regularly to study actual language usage in detail. It allows you to search the text (i.e. displaying 'concordance lines') to show actual usage of a word, it allows you to generate and save frequency-based word-lists (e.g. dictionary compilers can obtain a list of the top-most-frequently used words in a language to select for inclusion in their dictionaries), find frequent collocations (i.e. multi-word compounds), more easily generate a breakdown of word senses for particular words, and more.

tlCorpus is fully internationalized, that is, it supports all world languages. (Technically, this means all fields will accept **Unicode** characters.) If you have problems displaying characters for a particular language, this is usually either a font problem (e.g. you need to choose a font that supports the given characters), or the problem may be encoding-related. Please see the documentation for assistance if you encounter any problems working with data in your language.

tlCorpus is created by the makers of TLex, the industry-leading software for compiling dictionaries, used by many major publishers and individuals worldwide.

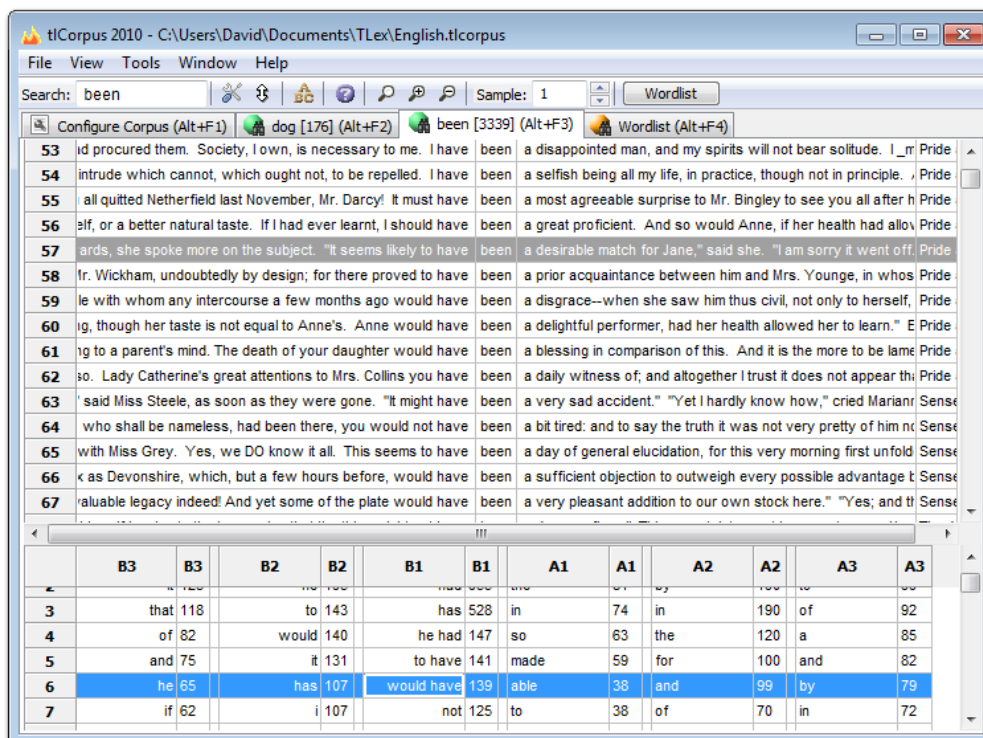


Figure 1: Screenshot of tlCorpus showing a search for "been" on a small sample English corpus. The section at the bottom is a "collocation summary"; see the relevant chapter for more details.

Frequently Asked Questions (FAQ)

For additional documentation, tips and tricks, answers to common queries, information on undocumented features, and other information for getting the most out of your software, it is

suggested that you regularly check our FAQ (“Frequently Asked Questions”) document, available online at: <http://tshwanedje.com/faq.html>. The FAQ also contains supplementary documentation (e.g. on new features) that has not yet made it into the User Guide.

ChangeLog (Detailed List of New Features)

The most up-to-date information on the newest features and fixes in the software is always published in the official ChangeLog at: <http://tshwanedje.com/tshwanelex/changelog.html>

Additional Resources

Free tutorial videos are available online via: <http://tshwanedje.com/videos/>

Also subscribe to our YouTube channel, where we periodically release new ‘tips and tricks’ videos, how-to videos and demo videos:

 https://www.youtube.com/channel/UCD_mfcdc234peP84Ooahgsg

Terminology Used

Corpus: A collection of texts, in one or more languages, in electronic format.

Concordance Line: A single search result of a search query on the corpus; these are typically displayed in the format of “lines”.

Frequency: How often a particular word appears in the corpus.

Collocation: A multi-word compound (e.g. phrases or idioms, such as “draw the line”).

Lexicography: The theory and practice of making dictionaries.

Word Sense: A single meaning of a word (a single word may have multiple meanings, i.e. 'senses').

Tokens: Total number of words in a text or texts. (Note that depending on the language, the term “words” may not be strictly linguistically correct; this is why the term “token” is used instead of “word”.) Example: The number of tokens in “It is a dog eat dog world” is seven.

Types: The number of uniquely occurring words in a text or texts (compare 'Tokens'). Words that appear more than once are counted once. This will always be less than or equal to the number of tokens. Example: The number of types in “It is a dog eat dog world” is six (as “dog” is counted only once).

TTR: Type-to-Token Ratio, typically expressed as a percentage. This is the number of types, divided by the number of tokens ('word count'). Example: The TTR of “It is a dog eat dog world” is 85.7% ($6/7 * 100$). The TTR, roughly, expresses a measure of the lexical density of a text or corpus.

A1, A2, A3, ...: “First word after”, “Second word after”, “Third word after”, etc. This denotes the word in the given position relative to the word matching a search term. (Note that some sources use the abbreviations **R1, R2, R3, ...** (for “right”); we prefer “A” as the term “right” would be incorrect for right-to-left languages.)

B1, B2, B3, ...: “First word before”, “Second word before”, “Third word before”, etc. This denotes the word in the given position relative to the word matching a search term. (Note that some sources use the abbreviations **L1, L2, L3, ...** (for “left”); we prefer “B” as the term “left” would be incorrect for right-to-left languages.)

Right-to-left: Indicates a language written in the direction of right to left, such as Hebrew or Arabic.

Keeping Your Software Up To Date

TshwaneDJe continually releases new “maintenance updates” for tlCorpus. These updates contain important and useful improvements and bugfixes, and it is strongly recommended that you keep your software up to date. You can check if there is a newer release available for your version of the software by using the “Help/Check for updates” menu option. This will open your web browser to a site that tells you if an update is available, and if so, tells you where and how to download it.

Maintenance updates (not to be confused with actual version *upgrades*) are free. When you install a maintenance update, re-activation is typically not required.

Getting Started with tlCorpus: QuickStart Guide

Setting Up a New Corpus

The first step in creating a corpus is to *obtain* some texts. There are many possible sources, e.g. gathering documents from the Internet, or scanning books. Once you have some text files, you are ready to begin. A “document” in tlCorpus (e.g. that you 'save' or 'open' or 'close') *is* the collection of texts. You can add or remove texts at any time.

Adding Texts to the Corpus

To add a single text file to your corpus, select “File/Add text”.

To add multiple files, use “File/Add texts ...”. Here you can select a folder containing multiple text files to add. If you enable the 'recursive' option, then text files in sub-folders of the selected folder will also be added.

You can also add a new text file based on text on the system clipboard using the “Add text from clipboard” command.

Adding Files Via Drag-And-Drop

You can also add files to your corpus by *dragging and dropping* either files or folders from your system file manager (e.g. Windows Explorer), onto the “Configure Corpus” window. If you drop folders, the contents of the folder will be added recursively.

Saving Your Corpus

Once you have added some texts, you probably want to save the configured corpus, so that you can quickly load it and continue the next time you use tlCorpus. You can do this by selecting “File/Save” (or press Ctrl+S).

To save the corpus under a different name or to a different location on the hard disk, use “File/Save As ...”.

<p>Important Note: The tlCorpus corpus file does not contain the actual content of the text files themselves – it only contains the names of the files on the hard disk, and some additional details, and loads the files from their original location when needed. If you need to copy or move a corpus to a different computer, you must also separately copy or move all the text files.</p>
--

Note that the last-opened corpus will be re-opened automatically whenever you run tlCorpus.

Unsaved Changes Indicator *

When you make changes to the corpus, e.g. add or remove text files, the changes will not be saved immediately by default; rather, the changes will reflect as “unsaved changes”, and a “*” character will appear in the title bar to indicate this. To save the changes, select “File/Save” (or press Ctrl+S).

Working With Multiple Corpora

You can easily work with multiple different corpora with tlCorpus. You can do this by saving each corpus under its own file name. To close the current corpus, use “File/Close”. You can then either set up a new one, or open a different one by using “File/Open Corpus” (Ctrl+O).

If you want to work with two different corpora *simultaneously*, you can also launch two copies of tlCorpus simultaneously, and load a different corpus in each one. **Tip:** You can rapidly switch between these (and other applications) in Windows using the “Alt+Tab” shortcut key (hold in Alt, and press Tab one or more times to select the other copy of tlCorpus that is running).

Running Searches

As soon as you have added texts to your corpus, you can start running searches by simply typing your search word in the “Search” box in the toolbar (near the top left of the tlCorpus window), and hitting “Enter”. See the “Search” chapter for more details.

Garbled Text?

IMPORTANT: If the text comes out looking like garbage, or some characters display incorrectly, please review the section titled “Important Notes on the Text File “Encoding”! Also, the problem may be caused by incorrect font settings.

PDF File Support

For t!Corpus on Microsoft Windows, PDF file support should work “out the box”.

For t!Corpus on Mac OS X, you must first install the “pdftotext” application available here:

<http://www.bluem.net/en/mac/packages/>

“Configure Corpus” Commands

Right-click Commands

If you right-click on one or more texts, you will be presented with a number of commands, explained below.

Selecting Multiple Texts At Once

You can select (and apply commands to) multiple texts at once. You can do this by holding in “Ctrl” while clicking on additional texts that you wish to select. You can also select a range of texts by holding in “Shift” while clicking with the mouse.

List of Right-Click Commands in Text List

Select all	Select all texts in the corpus
Select all in same folder	Select all texts in the same folder exactly (i.e., not its subfolders) as the text(s) selected in the list. (Please note this <i>adds</i> to any existing selection.)
Select all in same folder tree	Select all texts in the same folder (or its subfolders) as the text(s) selected in the list. (Please note this <i>adds</i> to any existing selection.)
Open (text file)	Open the selected text file(s) in the system’s default text editor. (If the original file is a type that requires internal pre-conversion, e.g. PDF or Microsoft Word, this opens the final generated text file, if it has been converted already.)
Open (original file)	For file types requiring internal pre-conversion (e.g. PDF or Microsoft Word or HTML), this opens the <i>original</i> PDF/Word/HTML file.
Open location	Open the folder(s) in which the selected text file(s) reside, in the system’s default file manager (e.g. Windows Explorer)
Open URL	Opens the website(s) in your default Web browser from which the selected text(s) was/were originally downloaded, if the address is available (e.g. if they were downloaded using the integrated Web crawler)
Copy path	Copy the full path of the selected text file(s) to the clipboard, e.g. “c:\MyCorpus\MyFile.txt”
Copy location path	Copy the folder name(s) in which the selected text file(s) reside, to the clipboard, e.g. “c:\MyCorpus”
Copy URL	Copies the website address(es) of the selected text(s) if available (e.g. if they were downloaded using the integrated Web crawler)
Add text	Select a single text to be added to the current corpus
Add texts	Select one or more texts to be added to the current corpus
Add text from clipboard	Create a new text file using the current contents of the system clipboard, and add it to the corpus. This allows an additional method to quickly and easily paste text from other applications (e.g. Web browser) “into” your corpus.
Remove selected texts	Remove the selected text(s) from the current corpus
Auto-detect language/encoding	Run the built-in automatic detection of language and encoding on the selected text(s). NOTE: This may detect information incorrectly, or be unable to detect. In that case, you can manually override the detected defaults by using “Select language/encoding”.

Select language/encoding	Manually select the language and/or encoding of the selected text(s). NB: See notes below.
Select language	Manually select the language for the selected text(s) (same as above option, but without encoding selector)
Wordlist	Generates wordlist for selected text(s) only
Character Info	Builds a list of all occurring characters in the selected text(s), and then displays those characters immediately in the Character Tool

Important Notes on the Text File “Encoding” [Intermediate to Advanced]

The text file format on computers (e.g. “.txt” file) was unfortunately never “standardized” in terms of how information is actually stored within the file itself. Many different standards were developed, by different organizations and at different times, for different languages and purposes that served their needs at the time. These different standards are known as “encodings”, and are specified as a list of “mappings” from the internal “numbers” in the text file data, to actual characters (e.g. in most encodings, the letter “A” is represented by the number “65”, but in some encodings it may be represented by a different number, or even be absent entirely). (The proliferation of different encodings was particularly bad before the world started moving to Unicode, which was intended to solve some of these problems.) Thus a given (say) Chinese text file, may have the characters stored as either the so-called “Big5 encoding” (an older, pre-Unicode standard), or in a Unicode-based encoding (though note, to confound the issue further, there are multiple possible *Unicode encodings* too, e.g. “utf8”, “utf8 with signature”, “utf16” etc.).

Unfortunately, it is often not possible to automatically detect the encoding of a text file. But a software application, such as tlCorpus, *needs* to know the encoding in order to open the text file correctly. If the software does not know the encoding, the text file will most likely either come out as entirely garbage, or some characters may become garbled.

You can tell tlCorpus what ‘encoding’ your text file is, by right-clicking on the text and selecting “Select language/encoding”.

IMPORTANT: tlCorpus does not support all possible encodings (there are thousands). We are steadily expanding the number of supported encodings. **Please contact us** if you run into an encoding that you need supported (or converted), and run into trouble doing so.

CONVERTING FILE ENCODINGS: If tlCorpus does not immediately support the encoding your file is in, the most simple solution is usually to convert the text file encoding outside of tlCorpus, using software such as a text editor, and re-save the file in a different encoding that tlCorpus supports. **The recommended best-supported encodings are generally “UTF-8 (with signature)” or any of the two “Unicode” (UTF-16) encodings.**

Search

You can run searches by simply typing your search word in the “Search” box in the toolbar (near the top left of the tlCorpus window), and hitting “Enter”.

Tabbed Interface

Each search that you run will open in a new “tab”. You can start a new search, even while the current one is busy, by simply typing a new search and hitting “Enter” again. Multiple searches remain open in parallel by default. You can switch between tabs by clicking on the tab heading, or using the tab selection shortcut keys (Alt+F1 to Alt+F9).

Important Note: The *first tab* will always be the corpus configuration window itself, where you can view or make changes to the list of files in your corpus.

Search Status Indicators (Tab Heading)

The currently selected search will be prioritized, but when it completes, other open unfinished searches will automatically continue to run in the background. The tab header icon will remain **orange** while a search is still busy, and will turn **green** when the search completes.

The number of results (corpus lines) is shown in square brackets in the tab heading.

Tab Selection Shortcut Keys

You can use the shortcut keys “Alt+F1” to “Alt+F9” to jump to tabs 1 through 9.

Closing the Current Tab

The current tab can be closed using “Window/Close”, or pressing Ctrl+W.

Closing All Open Searches

You can close all currently open searches at once using “Window/Close All Searches”.

Search Right-click Commands

If you right-click on one or more search lines, you will be presented with a number of commands, explained below.

Selecting Multiple Lines At Once

You can select (and apply right-click commands to) multiple search result lines. You can do this by holding in “Ctrl” while clicking on additional lines that you wish to select. You can also select a range of lines by holding in “Shift” while clicking with the mouse.

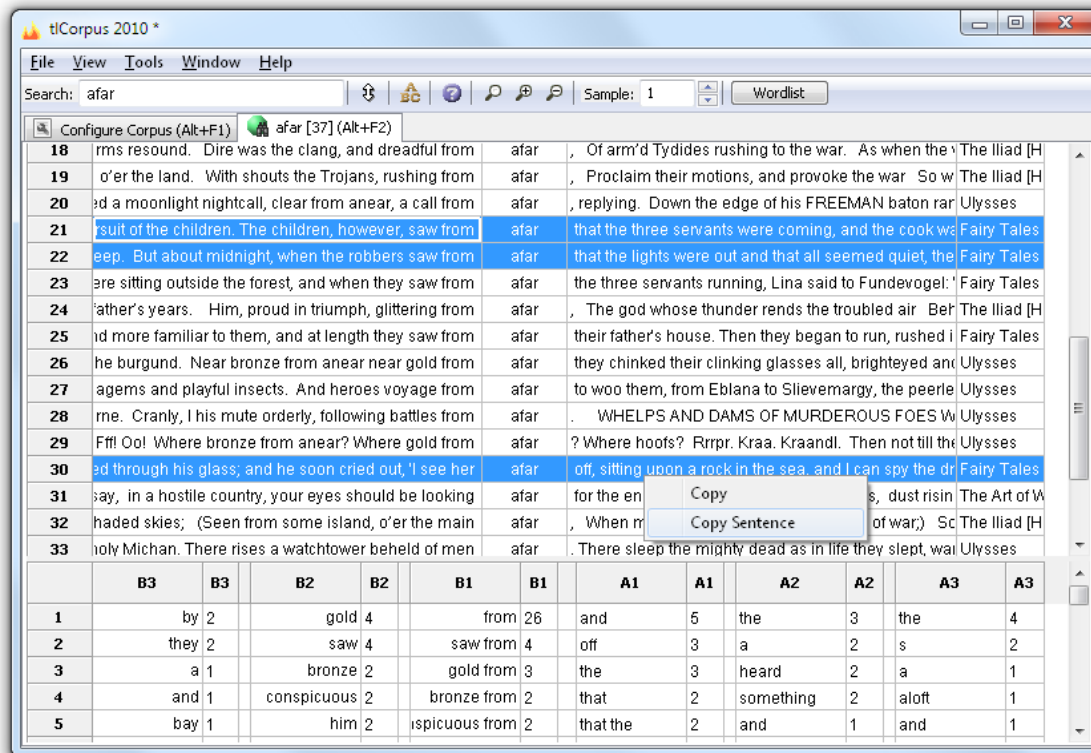


Figure 2: Selecting and copying multiple concordance lines at once.

List of Search Right-click Commands

Copy	Copy entire line(s) to clipboard
Copy Sentence	Copy sentence(s) to clipboard
Copy Example Sentence for TLex	Copy the sentence(s) to the clipboard in a special format that allows you to paste it directly into TLex as a ‘usage example’. <i>[Not relevant in tICorpus Standalone]</i>
Copy with Tabs (for Excel) (Ctrl+C)	Copy line(s) to clipboard in a format that allows it to be pasted into Excel with correct multiple column layout.
Copy Match(es)	Build a list of all unique matching words/phrases in the center column, and copy it to the clipboard. This is particularly useful when using wildcards (i.e. regular expressions), as this allows you to quickly do things like “build a list of all words that start with ‘g’ in this language”, or “build a list of all words containing the vowel ‘u’”.
Select All	Select all lines (search results) in this results list
Sorting	Invoke the configuration window for changing the sort order of results.
Open	Open the text file(s) for the current line(s) (in your default text editor)
Open location	Open the folder in which the text file(s) for the current line(s) resides (in your default system file manager, e.g. Windows Explorer)
Copy path	Copy the full path for the text file(s) to the clipboard, e.g. “c:\MyCorpus\MyFile.txt”

Copy location path

Copy the folder name(s) in which the selected text file(s) reside, to the clipboard, e.g. "c:\MyCorpus"

Sort Order

You can change the order of search results by the "Tools/Sorting" dialog. Select the desired sort criterion, and use "Move up" and "Move down" to change its priority. For example, moving "Word Before Search Term" to the top will sort the corpus lines (results) according to the word *immediately to the right* of the search word itself, allowing you to quickly spot various common collocates (multi-word compounds) to the right of the search word.

Launching Multiple Search Queries At Once

You can launch multiple search queries simultaneously by separating them with "#" in the search box, e.g. "fat#overweight#obese" will instantly launch three separate search query tabs.

"Sample" Option

The toolbar along the top of the tlCorpus window contains a box labelled "Sample". This allows the search to "skip" result lines, and only return every "N'th" result. For example, if you enter "10", only every 10th result will be returned. This can be useful for searches for common words, that may return a lot of results from the first few files in the corpus. By "sampling" every N'th result across the corpus, the results will be better spread out over multiple source text files in the corpus.

Advanced Searches with Regular Expressions

The search tool accepts **regular expressions**. This allows you to construct more advanced searches, such as wildcard searches.

Broadly speaking, regular expressions are an industry standard, and many detailed references are available on the World Wide Web, such as the following recommended Wikipedia article: http://en.wikipedia.org/wiki/Regular_expression. Thus, only a very brief overview will be provided here.

Following are a few examples of regular expression constructs:

Character wildcard: "." can be used to indicate "one of any character", thus the search string "li.e", will match "like", "line", "life", "lime", "lice", etc. (but not "lie").

One or more: "+" denotes that the preceding character may be repeated one or more times, e.g. "lo+se" will find occurrences of both "lose" and "loose".

Zero or more: "*" denotes that the preceding character may be repeated zero or more times, e.g. "sail.*" will match "sail" and its variants, e.g. "sailing", "sailed", "sails".

Square brackets allow you to specify a specific range of possible characters to search for at a given spot in the text string, e.g. an alternate search string for searching for occurrences of both "gray"

and “grey” would be “gr[ae]y”. This will match “gr”, followed by either an “a” or an “e”, followed by “y”.

Ranges of characters may also be specified within the square brackets, e.g. “[a-z]” will match any single alphabetic character between “a” and “z” inclusive. The regular expression search string “^[a-z][a-z][a-z]\$” would thus, for example, match all three-letter words.

If you wish to actually search for one of the characters that are used as special regular expression query characters, you can do so by “**escaping**” the character – this is done by preceding it with a “\” character, e.g. “\.” to search for a full stop (period).

Collocation Summary

When a search completes, the search window will display a summary of the most frequent collocates of the matching search word(s). This is displayed for the various positions, e.g. A1, A2, A3, B1, B2, B3 etc.

Generating Wordlists

Wordlist Tool

To generate a list of all words that appear in your corpus, click on the “Wordlist” button in the toolbar.

The list includes the number of occurrences of each word, and is sorted is sorted by descending frequency (i.e. most common words first).

Following is an example of the output for English:

```
the,81909
of,45889
and,42561
to,34553
a,25782
in,24753
he,23482
his,14708
was,14625
that,12351
it,11301
with,11299
is,11165
...
```

Exporting The Wordlist To a File

To generate and save a wordlist from your corpus to a text file, there are two options:

1. “File/Export wordlist”
2. Use the “Wordlist” button in the toolbar, and then click the “Save ...” button.

Wordlist Right-click Commands

If you right-click on a word in the wordlist window, you can access the following commands:

Search	Launch a new tlCorpus search window for the selected word
Web Search	Launch a web search (in your browser) for the selected word
Web Image Search	Launch a web images search (in your browser) for the selected word
Character Info	Displays the selected string in the Character Tool
Copy	Copy the selected word to the system clipboard

Overall Corpus Statistics

The wordlist tool also calculates and displays the number of types, tokens and the type-to-token ratio (TTR) for the entire corpus (see the chapter “Terminology Used” for an explanation of these terms).

Generating a List of Occurring Characters

You can generate a list (in comma-separated text file format) of all occurring characters in your corpus using the “**File/Export character list**” command. The list is sorted by occurrence count, in the format “Character,Count”, e.g.:

```
e,22424  
a,18053  
t,17933  
...
```

Note also that if you right-click on one or more texts in the “Configure Corpus” window, you can also select the “Character Info” command, which generates a list of occurring unique characters in the selected text(s), and displays them in the “Character Info” window.

Calculating/Generating N-Grams

Broadly speaking, there are two primary different types of “n-grams” - ‘character n-grams’, and ‘word n-grams’. ‘Character n-grams’ represent adjacent *alphabetic characters*, e.g. “th” and “he” are two bigrams of the word “the”. ‘Word n-grams’ represent adjacent *orthographic words*, e.g. given the sentence “the cat sat”, we would have the following word bigrams: “the cat”, and “cat sat”.

The variable “n” represents the number of adjacent characters or words to include, e.g. when n is 2, then we generate bigrams, as in the examples above. For n=3, we generate trigrams (e.g. for the sentence “the cat sat down”, there are two word trigrams, “the cat sat” and “cat sat down”), and so on.

Character Bigrams

You can generate a list (in comma-separated text file format) of all occurring pairs of characters in your corpus using the “**File/Export character bigrams**” command. The list is sorted by occurrence count, in the format “NGram,Count”, e.g.:

```
th,71  
he,58  
in,58  
...
```

Word N-Grams

You can generate a list (in comma-separated text file format) of all occurring *word* N-grams in your corpus using the “**File/Export word N-grams**” commands. The list is sorted by occurrence count, in the format “NGram,Count”, e.g.:

```
the cat,22  
cat sat,18  
...
```

Web Crawler

The integrated “Web Crawler” tool allows you to quickly download website addresses (URLs) and automatically add the downloaded file(s) to the present corpus.

Options

Output Folder

This option allows you to select the folder in which to save downloaded files.

Monitor Clipboard

If the checkbox option “Monitor clipboard and automatically download URLs” is ticked, tlCorpus will monitor the contents of the clipboard for website addresses (e.g. text starting with “http:”), and automatically queue any new addresses for download by the Web Crawler. This allows you to, for example, sit in your Web browser (while tlCorpus is open in the background), and simply right-click on desired links and use the 'copy link to clipboard' option (in your Web browser) on multiple addresses (e.g. newspaper articles or documents or even image files etc.) that you wish to download.

Don't Download URLs Already In Corpus

When the Web Crawler adds a downloaded text or HTML file to the corpus, it records the original address (URL) as a property of the new text added. When this option is ticked, the Web Crawler will not download any URLs that appear to have already been downloaded and added to the corpus previously.

Automatically Add Downloaded Files To Corpus

If this is enabled, then any downloaded files (presuming they are applicable file types that tlCorpus can handle), will automatically be added to the current corpus. This is enabled by default, but can be disabled with this option. (This also allows the tlCorpus Web Crawler to be used as a more general-purpose downloader/spider.)

Adding Multiple URLs at Once

Multiple URLs or URL patterns can be specified at once (when adding from the clipboard), using multiple lines, one URL or URL pattern per line.

Adding Multiple URLs at Once With Address Ranges

It is possible to automatically add an entire numerically-incremented *range* of addresses to be downloaded, e.g.:

<http://www.newspaper.com/index.php?article&id=300>
<http://www.newspaper.com/index.php?article&id=301>
...
<http://www.newspaper.com/index.php?article&id=400>

This can be done using the syntax '|startnum,endnum|', e.g.:

<http://www.newspaper.com/index.php?article&id=|301,400|>

This would retrieve 100 articles.

The numbers may optionally also be front-zero-padded, if specified in the start number, e.g.:

<http://foo.com/IMG|0001,0100|.jpg>

This would download all of:

<http://foo.com/IMG0001.jpg>
<http://foo.com/IMG0002.jpg>
...
<http://foo.com/IMG0100.jpg>

Note that the address range syntax also works for non-text file-types, such as images.

Right-to-left Languages

Languages written right-to-left, such as Arabic and Hebrew, are supported by tlCorpus. In the corpus configuration tab, tick the “Right-to-left” checkbox for files containing primarily right-to-left text.

Font Problems

It is recommended that you obtain a copy of “Arial Unicode MS”; this contains support for most Unicode characters.