Linguistic Inquiry and Word Count: *LIWC2015*

Operator's Manual

James W. Pennebaker, Roger J. Booth, Ryan L. Boyd,
and Martha E. Francis

Getting Started

The LIWC2015 program comes with a single application file for either Windows or Macintosh. The dictionaries that come with the program include three internal default dictionaries for LIWC2015, LIWC2007, and LIWC2001.

A variety of free dictionaries for use in basic research are available to LIWC users from the LIWC website, www.LIWC.net/dictionaries. These include LIWC dictionaries in Spanish, German, Russian, and many other languages as well as specialty dictionaries that other researchers have built. It is recommended that you download whatever dictionaries you need to your computer so that you will have ready access to them.

Running LIWC2015 for the First Time

There are three versions of LIWC2015:

- Macintosh
- Windows PC (for 64-bit systems)
- Windows PC (for 32-bit systems)

Most PC users with a relatively recent computer have a 64-bit machine (you can find which PC version by clicking on your computer's **Start** button, then click **Computer**, and then **System**). Alternatively, you can search for the **System Information** program in your **Start** menu -- your system type will be listed under **System Type** in the main **System Summary** menu.

Once you have downloaded the LIWC installation program, follow the on-screen instructions to install it. Once installed, it should be available in your Applications folder (for Mac) or Start/Programs menu (for PC). The first time that you run LIWC2015, you will need to provide the activation serial number that was e-mailed to you at the time of purchase. The serial number looks something like this:

LIWC2015-VM3A-TY2Q-RAQL-CNJ3-4H9U
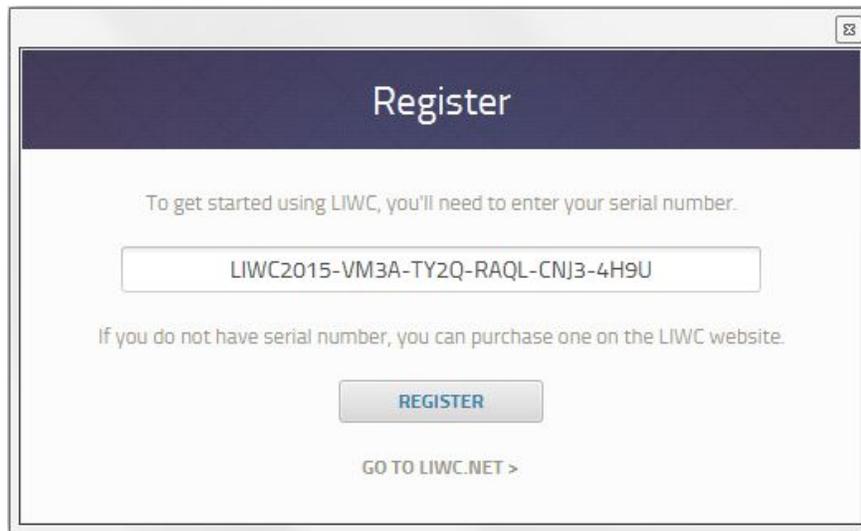
Input your serial number into the *Registration* screen, then click the **Register** button to begin validation. Make sure that the serial number is formatted <u>exactly</u> how you see it in the e-mail that you receive. Following registration of your serial number, you will be prompted to activate your copy of LIWC2015.
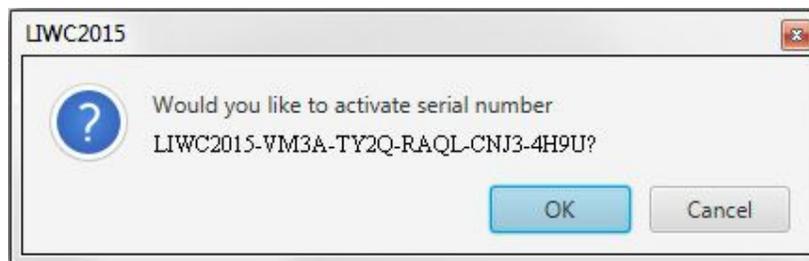
Once you select **OK**, the application will validate your copy of LIWC2015. Upon successful activation, you will receive the message "Product Activation Succeeded." LIWC2015 is now ready to use.

> *Important:* Save your serial number in case you need to reinstall LIWC at a later date. Depending on the license that you purchased, the serial number will remain active for either one month or indefinitely. If you purchased the one month rental license, you will need to renew your LIWC license after one month (from date of purchase).
>
> On purchase, you will be able to activate LIWC2015 on only two computers. If you plan to transfer your copy of LIWC2015 to a third computer, you will need to first **Deactivate** your serial number. You are allowed to transfer your serial number a total of three times. For more information, see the **Serial Number Issues** later in this manual.

Analyzing Text in LIWC2015

In LIWC2015, there are three primary ways to analyze text:

- Choose one or more text files for analysis
- Choose entire directories and subdirectories containing text files for analysis
- Choose text columns within an Excel file (or other delimited spreadsheet file, such as CSV).

The table below describes all of the file types that LIWC2015 can analyze. If all of your files are standard English, all of the following formats will work equally well. However, if your files or your custom dictionary are in a non-English language with non-English characters, Word document files are recommended.

Text Files	
.docx	Standard MS Word 2007 (and later) format. Also used by popular word processing software including LibreOffice and OpenOffice
.doc	Standard MS Word 97-2007 file format for text documents. *Note*: LIWC2015 cannot analyze MS Word files that were saved in Word-6 or earlier (pre-1997). If you want to analyze such files, you will need to open them in a more recent MS Word version and resave them.
.txt	Plain text file format -- the most common type of text file, sometimes called an "ASCII" file. See the **Settings for Opening and Saving Files** section of this manual for more details.
.rtf	Rich Text Format files. An older text file format that was commonly used between 1987 and 2008.
.pdf	Portable Document Format, or PDF -- common with Adobe's PDF-reader software. Your PDF file must contain computer-readable text in order to be analyzed by LIWC2015. *Note:* LIWC2015 is not able to process scanned pictures of text (for example, scanned book pages) or other text embedded within images.
Spreadsheet Files	
.xlsx	Standard spreadsheet file for MS Excel from 2007 up to the present version.
.xls	Standard MS Excel spreadsheet file format. Originally used in versions

	of MS Excel prior to 2007.
.csv	Comma Separated Value format. A common format for data files. Note that different CSV files may have different formats depending on their country of origin. Please refer to the **Settings for Opening and Saving Files** section of this manual for more details.

Analyzing Text Files in LIWC2015

The two primary ways of analyzing text files are by using the **Analyze Text** feature and the **Analyze Text in Folder** feature. Both features can be found under the **File** menu within the main LIWC2015 application.

If you select **Analyze Text**, you will be asked to find the specific file(s) that you would like to analyze. If you choose to **Analyze Text in Folder**, you will be asked to select a folder containing text files, then whether you would like to include subdirectories. If you would like to analyze all text files that are contained in all subdirectories (including subdirectories within subdirectories), then you should select **Yes**. Otherwise, select **No** to analyze only the files contained in the folder that you chose.

> *Tip:* If you have a large number of text files to process, it is most efficient to put them all into one directory (or one directory with multiple subdirectories).

Analyzing Spreadsheet Files in LIWC2015

You may already have all of your data in an Excel spreadsheet file. For example, if you had conducted an open-ended survey with 5,000 participants, you might enter variables such Participant ID, age, and sex in columns A, B, and C, respectively. In column D, you could have participants' text-based responses to an open-ended question. In these cases, it can be tedious to copy all of your text from a spreadsheet into separate files, analyze the text, then merge the results back into your original dataset. We've fixed this problem. Presto! Problem solved! In LIWC2015, it's easy and fun to directly analyze text already inside of a spreadsheet file.

To analyze text within a spreadsheet, select the **Analyze Excel/CSV file** option from the **File** menu, or select **Analyze Text** from the *Welcome* screen. Once you have selected your spreadsheet file, you will be prompted to identify which column(s), row(s), or cell(s) you wish to analyze. For example, if your text responses are in column D, you would highlight this entire column, then choose **Analyze**. Once LIWC has analyzed the text in this column, The LIWC output would appear in the first empty column of your spreadsheet -- in this example, column E.

> *Note:* In some cases, your spreadsheet might have multiple cells per row that contain text. If you select multiple columns for each row, the text in these columns will be combined for text analysis.

> ***Important Note about Excel Files:*** MS Excel spreadsheet files have a limit on how much text can be stored in a single cell -- any text that exceeds this limit will be forever lost by MS Excel. If your dataset contains large amounts of text in each cell, we recommend

saving your text files in a format that does not have these limitations (e.g., plain text or MS Word files).

Viewing and Analyzing LIWC2015 Output

After processing your text files, LIWC2015 displays the results on screen. For every new analysis that you run on a PC, LIWC2015 will create a new tab for displaying results -- this allows you to have the results from multiple analyses open at the same time (see below). For Mac users, each analysis is in a separate window.

After your files have been analyzed, you can save the results by going to the **File** menu and selecting **Save Results**. Results can be saved in multiple formats, including tab-delimited text files, CSV files, and Excel files. Most statistical packages (e.g., SPSS, R, SAS, SciPy, Weka) are able to read data files in one or all of these formats.

> *Tip:* If you have analyzed a large number of files, or if your dictionary has a large number of categories, it may take some extra time for LIWC2015 to prepare the results for display within the application.

Most LIWC2015 output variables are expressed as percentage of total words. There are six exceptions: word count (WC; raw word count), (WPS; mean words per sentence), and four summary variables: Analytic, Clout, Authentic, and Tone. Each of the summary variables are

standardized composites based on previously published research. The composites have been converted to percentiles based on large corpora of texts described in the LIWC2015 Language Manual.

Settings for Opening and Saving Files

Region and Encoding Settings

Within the **Options** menu, clicking on the **Settings** option will allow you to choose the ways in which your files are read, formatted, and saved. The default settings are set to handle most English text using North American conventions for decimals. For most files, UTF-8 encoding (a version of Unicode) should be sufficient. However, in some cases, your text files may have a specific encoding that contains characters not handled by UTF-8. In these cases, you should select the correct encoding for your files within the **Settings** menu. File encodings may vary by country or language -- you should typically select the standard coding for your region if you are experiencing issues.

> ***Note:*** Region and encoding settings apply to plain text (.txt) and CSV files. For other file types, such as MS Word and MS Excel files, LIWC2015 will automatically determine the correct encoding.

Country Settings

Countries and regions also differ in how numbers are represented (including decimal notation). In North America, *pi* is expressed as 3.14, in other parts of the world it is 3,14. To change LIWC2015 to your region's standard notation (for both input and output files), go to the **Settings** menu.

CSV Format Settings

CSV files are a standard format for holding tables of data, usually using a comma (,) to denote new columns and quotation marks (") to denote cells containing text. However, different regions of the world and different academic fields can have different formatting practices for CSV files. The CSV options in the **Settings** menu can be changed to the appropriate formatting of your files. Note that the settings that you select will apply to both input files that you choose to analyze, as well as output files (when saving your results in CSV format).
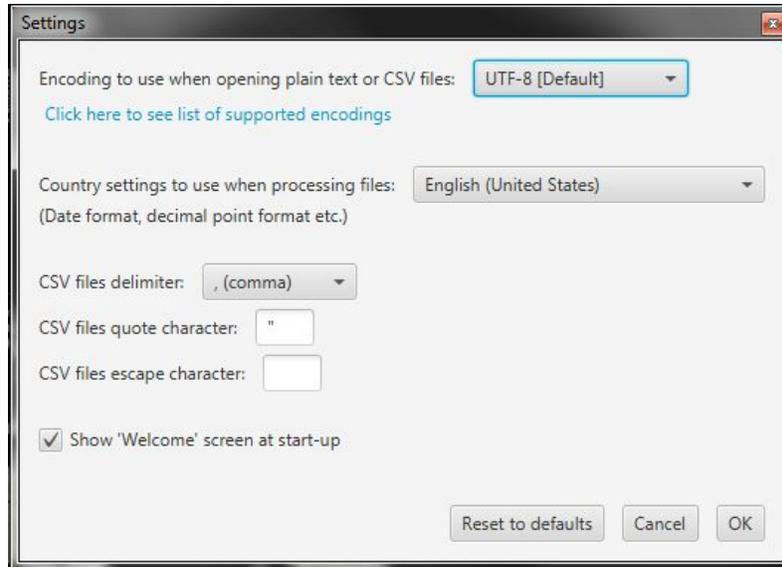
Customizing LIWC2015 Output

In some cases, you may prefer to analyze only a subset of language dimensions rather than the full set of LIWC variables. For example, you might want to only analyze positive and negative emotion words. To do this, select the **Categories** menu from within the **Options** menu. Within each broad category (e.g., standard information, linguistic dimensions etc.), checkboxes are available for each LIWC2015 variable. By checking the box next to different variables, you are selecting to include them in your LIWC2015 results and output. Note that the category preferences will remain in effect until they are re-checked and will be saved when you close the application. To use all dimensions, click on the "Select All" button.

> *Tip:* The following variables can only be turned on when using the LIWC2015 internal dictionary: Analytic, Clout, Authentic, Tone.
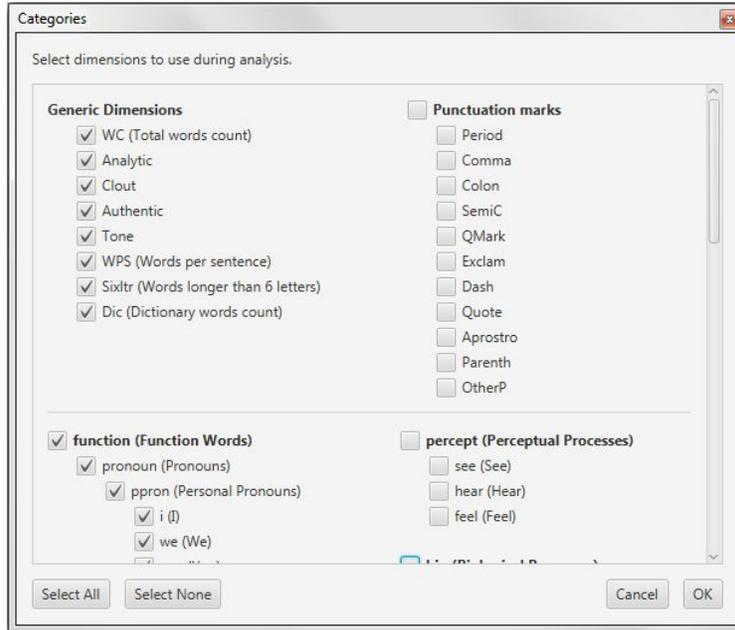
Analyzing Text in Segments

Each text file analyzed by LIWC2015 can be treated as a whole or broken into segments. This is controlled by the **Segments** option under the **Options** menu. You have five choices:

- *No text segmentation*
- *Define number of segments*, where you can choose how many equally-sized segments you wish to divide your text files into
- *Define words per segment*, wherein you choose how many words will be in each segment
- *Define segments by special characters*, where you can segment a file after the occurrence of a character such as "#" or even a word like "[break]"
- *Define segments based on the number of carriage returns* (or hard line breaks) in your text.

Note that in the LIWC results output file, the second column refers to the actual segment sequence
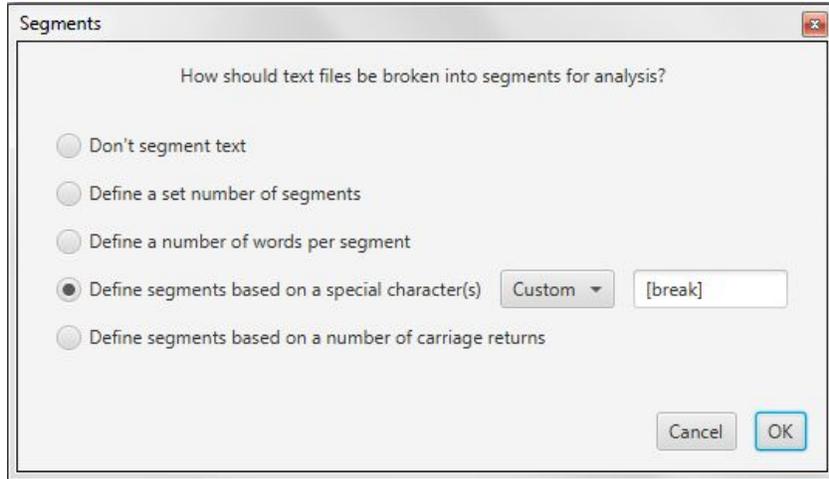
Tip: Your segmentation options are saved each time you exit the LIWC2015 application. When you open the application, check your *Welcome* screen to ensure that your segmentation options are set up to your liking.
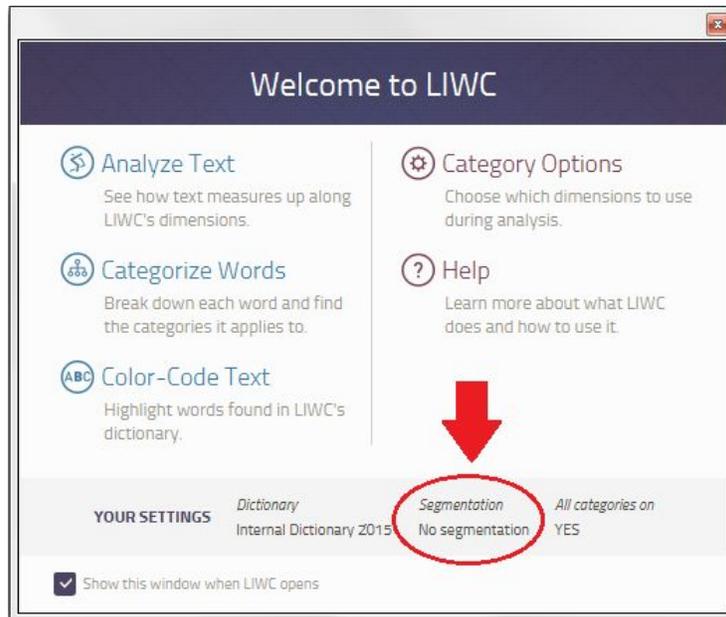
Color Coding Text

It is sometimes helpful to see which words within a text are being counted by LIWC. To do this, simply click on the "Color-Code" option and select the file you would like to analyze. The application will display the analyzed text, and dictionary-recognized words will be in red.
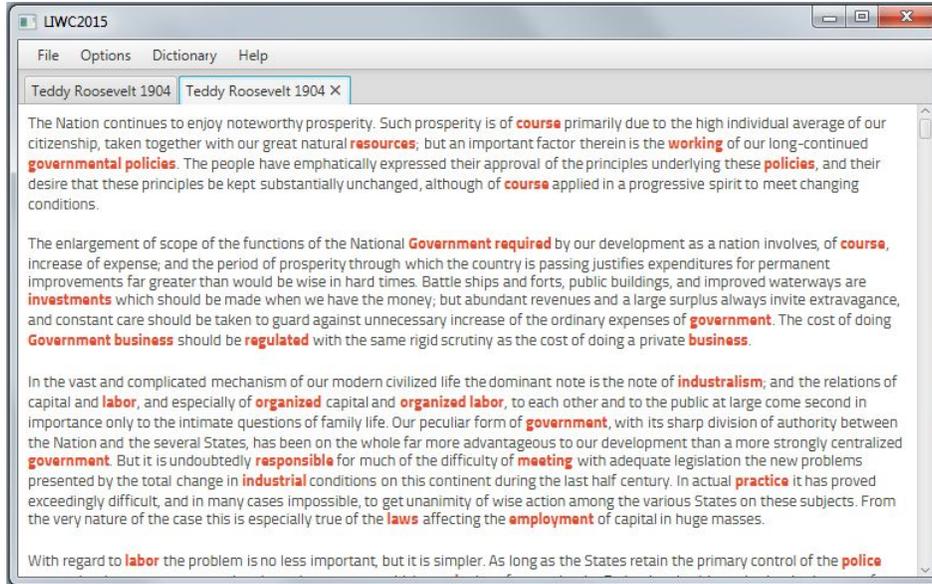
By default, LIWC2015 will color-code <u>all</u> words contained in your dictionary. If you would only like to see highlighted words from a single category, go to the **Categories** menu turn off all of the other variables except for the one that you are interested in. For example, in the above image, the LIWC2015 "work" category was selected from the **Categories** menu.

> *Note:* LIWC2015 will only color-code one text file at a time.

Word Counting Conventions

With the exception of the summary variables and words per sentence, all LIWC2015 output variables are expressed as percentage of total words. One would think that calculating word count would be straightforward. One would be wrong. For example, how many words are in the following sentence?

> *The red-shirted lawyer gave her ex-boyfriend $300 out of pity :(.*

There are either 9, 10, 11, or 12 words depending on how you count. MS Word's internal counter sees 11 words, LIWC counts 12, wordcounter.net computes 10, and wordcounter.com says 9. Herein lies the problem. LIWC2015 counts red-shirted as two words (since it is not in the LIWC dictionary), and ex-boyfriend as a single word since it is in the dictionary. The MS Word counter assumes all hyphenated words are single words. Both LIWC and MS Word count the emoticon as a word whereas the two online counters do not. Finally, wordcounter.com does not count $300 as a word. LIWC2015 relies on a slightly different computational method than the previous versions of LIWC. Consequently, the word counts might differ slightly.

The "Punctuation…" item of the **Categories** menu allows you to determine how LIWC2015 counts and reports punctuation characters. After selecting various punctuation characters from this menu, LIWC2015 will count them and report them as a percentage of total words. For parentheses, LIWC2015 counts <u>pairs</u> of parentheses. Also, the "Other Punctuation" category

includes all punctuation marks not counted in the other punctuation categories as well as ASCII characters from 33-47, 58-64, 91-96, 123-126.

> *Note*: All Punctuation output is expressed as a percentage of total words.

Creating and Using Custom Dictionaries

LIWC2015 permits use of dictionaries and categories created by the user. This is done by selecting the "Load new dictionary…" option from the **Dictionary** menu. External dictionaries must be contained inside of plain text saved with a ".dic" extension. LIWC2015 will also read dictionary files from older versions of LIWC that have a ".dic" file type as well. If you create your dictionary file in a program like MS Excel or MS Word, you will need to save your file as a plain text (.txt) file, then rename it to a ".dic" file. Your dictionary files must be set out in the following format:

%			
1	firstperson		
2	secondperson		
3	thirdperson		
4	singular		
5	plural		
6	possessive		
%			
I	1	4	
I'*	1	4	
my	1	4	6
our	1	5	6
you	2		
their	3	5	6
us	1	5	
we'*	1	5	

This rather incomplete pronoun dictionary is divided into two parts. The first part, between the two "%" characters, identifies which categories will be created. The second half, following the second "%" includes which words to count and which categories they should be added to.

The category definitions in the top half always start with a number followed by the name of the category. The category name can only be a single word (NOT two or more words). The category name follows the category number by a tab. It makes no difference if the names or words are capitalized or not. LIWC is not case sensitive.

The words below that second "%" are the ones that will be counted and are each assigned to one or more of the pre-assigned categories. Hence, the word "I" is associated with category 1 (the

word dictionary of "firstperson") and category 4 (a singular rather than plural pronoun), the word "our" is associated with categories 1 (firstperson), 5 (plural), and 6 (possessive).

All word categories to be counted <u>must</u> be separated by a tab (old dictionary files that use spaces will need to be updated). In addition to counting whole words, LIWC2015 also counts word stems. Word stems are partial words terminated by an asterisk (*). Use of an asterisk at the end of the word or word stem signals LIWC2015 to ignore all subsequent letters. Consequently, "I'*" will count the words I'm, I'll, I'd, I've and so forth in categories 1 and 7.

Multi-Word Phrases

Although LIWC was originally created to examine words or word stems, it can now count phrases. Note, however, that word phrases are counted as a single word. For example, the phrase "kind of" will be treated as a single word, rather than each separate word ("kind" and "of") being counted individually. This feature is primarily intended to be used for brief utterances and very specific concepts that you wish to classify in your texts.

Helpful tips

- The first line of the dictionary must be the % symbol, followed by the Category numbers and names. After the last Category name, a second % symbol must be inserted to signal the beginning of the Word entries and numbers.

- It is more efficient to create your own dictionary than to try to integrate your dictionary into the LIWC2015 default dictionary.

- Category names should be ONE word only with no punctuation. Category names do not need to be in alphabetical order.

· Words should be listed in alphabetical order. Word entries should only appear once.

OK: Love   500   504

NOT OK: Love   500

                      Love   504

- Be careful in the use of asterisks to avoid misclassifying words. For example, the following case is problematic:

NOT OK: Thank*   27

                      Thanksgiving 94

In this case, the word "Thanksgiving" would only be counted in category 27 since LIWC2015 does not double-count words. The asterisk at the end of "Thank*" will capture <u>all</u> words that begin with *Thank*.

Preparing Written Text For LIWC2015 Analysis

LIWC2015 can analyze text within separate text files or within a single Excel-compatible spreadsheet file. In both cases, the accuracy of LIWC2015 output data is determined by the quality of the text samples themselves. In order to ensure the best results, it is necessary to properly prepare text samples for LIWC2015 analysis. The essential steps for essay text organization, entry, and editing are as follows:

1.   **Text file organization.**

Each language sample should be put in its own file and named in a systematic and meaningful way.

2.   **Text file computer entry.**

Essays should be saved on your computer as MS Word documents or as standard plain text (e.g., ASCII) files. Files written and saved by other word processing programs (e.g., Word Perfect) may not work. Also, files in formats such as html, jpg, etc. will not provide valid output. A good rule of thumb is that files ending in .doc, .docx, .txt, .rtf, or .pdf will probably provide accurate results.

> *Note:* PDF files that contain computer-readable text can be processed by LIWC2015. Images within PDF files, including scanned images of text (but not computer-readable text) <u>cannot</u> be analyzed by LIWC2015.

3.   **Cleaning the text files.**

As a general rule, LIWC2015 can work with most naturally-produced files. Occasional misspellings are rarely a problem. Many people like to run all files through standard spell-check programs. Because LIWC2015 converts all text files to lower case before processing them, grammar, capitalization, and sentence structure do not need to be corrected. Be sure to check if your files may be corrupted or contain large portions of unwanted text.

Naming Your Files

Because the file names are part of the output file, certain conventions should be adopted in the preparation of the files and file names:

- **Separate files for separate text samples**. LIWC2015 analyzes data one file at a time. If participants write responses to two questions or perhaps write on two separate days, each question or day should be a separate file. If responses to both questions (or both days' writing) are within the same file, LIWC2015 will analyze them as a single writing sample. If you are analyzing a data table (e.g., Excel files, CSV files), each language sample should be located in separate rows or columns.

- **File names should be descriptive,** including ID number, condition, and question or day number. For example, data from a study with two conditions and three days of writing might be saved in files using this naming strategy:

  [PARTICIPANT#][DAY#].[CONDITION] -- 4568day1E.txt, 4568day2E.txt, and 4568day3E.txt

Typing Conventions: Writing and Interview Samples

In making corrections or cleaning text files, keep in mind what your goals are in analyzing the data. LIWC2015 does not discriminate between upper- and lower-case letters. It can only count words that are in its dictionaries. Misspellings, colloquialisms, foreign words, and abbreviations are usually not in the internal dictionaries. The following items should be checked before any files are analyzed:

1.   **Spelling, abbreviations, contractions.**

Remember that LIWC2015 typically analyzes common words. Proper names, scientific or technical terms, or rarely-used words are not parts of the internal dictionaries -- which means if they are misspelled, it doesn't matter. For more precise results, be sure and correct spelling errors for common words. Both United States and British spellings are included in the internal dictionaries.

Meaningful abbreviations should be spelled out. "Jan" should be January. More obscure abbreviations or acronyms, such as "AT&T", can remain as such unless you have reason to want the term to be expanded and counted as four separate words: "American Telephone and Telegraph".

Common verb contractions are in the dictionary and do not need to be changed. These include: don't, won't, isn't, shouldn't, can't, couldn't, I'm, I'll, I'd, we're, we'd, you're, he's, it's, etc. Most others will be simply counted as possessive nouns: "Sally's shoes" will be counted the same way as "Sally's going to the store." In the second case, change "Sally's" to "Sally is."

> *Helpful tip*: Don't lose a lot of sleep about spelling. The more files and words you have, the less it matters. If your study has 100 people who write 1,000 words each, even if 1 percent of your words are misspelled, it probably won't matter. However, consistent misspellings should be corrected.

2.   **End of sentence markers and hyphens.**

The Words per sentence (WPS) category is based on the number of times that end-of-sentence markers are detected. These include all periods (.), question marks, and exclamation points. One potential problem is that common abbreviations (such as "Dr.", "Ms.", "U.S.A.", "D.O.A.") will be counted as multiple sentences unless the periods are

removed. Be careful that the removal of the periods doesn't make a new word. For example, the United States, or "U.S.", becomes "US" (1st person plural pronoun) when the periods are removed. In this case, change it to "USA".

Time markers (e.g., 6 a.m. or 7:30 p.m.) can also be a problem. Because "a.m." without the periods is a verb, "am", change time to 6am or 7:30pm.

When words start or end with hyphens, they are read by LIWC2015 as part of the word. LIWC2015, for example, lists "chit-chat" as a meaningful word in one of its dictionaries. In cases of hyphenated phrases such as "this-or-that" LIWC2015 will count it as three separate words since "this-or-that" is not in the dictionary.

3.   **Common Internet Notations**

Several types of internet notations can lead to less accurate results. Examples of this include e-mail addresses, hashtags, and web URLs and other links. Like misspellings, the occasional internet notation here or there will not likely impact your results in any significant way. However, you should consider altering them or removing them from you sample prior to analysis if your text data has a large amount of these types of notation. A few examples are shown below.

Entry Type	*Entry Example*	*Recommended Replacement:*
E-mail Address	your.name@example.com	subEmailaddress
URL address	http://www.LIWC.net	subURLaddress
Hashtag	#LIWCisawesome	subHashtag
Twitter Handle	@jwpennebaker	subTwittername

Note that many types of "netspeak" that is used as shorthand interpersonal communication (e.g., "lol", "4ever") are captured by the LIWC2015 dictionary and do not need to be altered in your text files.

4.   **Other common problems:**

Typed entry	*Change to:*
w/	with
b/	between
&	and
'cause	because
and/or	and - or
'an or 'n	and
mos	months
sec	second
@	at

Transcribing Oral Exchanges: Special Problems

Although not designed for spoken language, we have found LIWC2015 to be useful in analyzing conversations and interviews. To accommodate certain dimensions of spoken language, we have adopted the following conventions:

1. **Nonfluencies.**

    Hm, hmm, uh, uhh, uhm, um, umm, and er are part of the nonfluency dictionary. Other forms will generally not be caught (e.g., ooooh should be changed to um if used as a nonfluency).

    Stuttering can be accommodated by altering the stuttering part of a phrase to a nonfluency marker. For example, "The, the bo-, the boat went into the water" could be changed to "Uh, the boat went into the water." The transcriber will have to decide how many uh's would be appropriate.

    Uh-uh and uh-huh should be changed to "no" and "yes". Huh? should be changed to "what?" Or, if you are very, very proper, to "Excuse me madam, I didn't quite catch what you said."

2. **Fillers.**

    Everyday speech is littered with "meaningless" fillers. Unfortunately, these fillers use some of the most important words in our dictionaries. Watch out for the following:

    You know. As in, "we went, you know, to the store and, you know, bought gum." Change to one word: youknow. "We went, youknow, to the store..."

    I mean. As in, "we went, I mean, to the store..." Change to one word: Imean.

    I don't know. As in, "we went, I don't know, to the store..." Change to: Idontknow.

    Like. "We went, like, to like the store and like we like bought like gum." Be careful with like because sometimes it is used appropriately. As a nonfluency, change it to: rrlike. Note that all words starting with "rr" will be coded as a nonfluency. Hence, if you are transcribing audiotapes made in the 1950's, the word "well" would likely be used the way "like" is today. Hence, you would enter it as "rrwell."

3. **Transcribers' comments.**

    LIWC2015 is designed only for expressions in the form of language. Transcribers often insert remarks, such as [subject laughs], [shaky voice], [whispers]. We recommend removing these.

    Occasionally, the transcriber cannot understand a word or passage. Rather than writing [can't understand word] or [?], it is recommended that you just leave the unheard word out. For entire passages, don't insert anything.

Troubleshooting and Technical Support

Problems with Opening LIWC2015

Problems with Opening LIWC2015 on a Mac

After downloading and installing LIWC2015 on Mac computers, some users receive an error message that LIWC cannot be opened because it is not an approved application. To open it the first (and only) time, go to the **Finder** tab, then click on **Go,** then click on **Applications**. Once in the **Applications** menu, hold down the "control" button (next to your space bar) and click on the LIWC2015 icon.

Problems Opening LIWC 2015 on a PC

If you receive an error message when first opening LIWC2015 after installation, it is possible that the wrong version has been installed. For example, Windows 10 will display a message such as:

> *This app can't run on your PC*
> *To find a version for your PC, check with the software publisher.*

Note this message will be different depending on your version of Windows (e.g., Windows 7, Windows XP).

An error message like this typically means that you are trying to run the 64-bit version of LIWC2015 on a 32-bit machine. In order to fix this, first uninstall the 64-bit version of LIWC2015, then download and install the 32-bit version of LIWC2015.

Serial Number Issues

A 20-30 character alphanumeric serial number is required to activate your LIWC program. When you purchased LIWC2015, your serial number was sent to you by email. If you did not receive the email, check your Spam folder. If it does not arrive within an hour, contact technical support (see email address below).

Invalid or expired serial number

You may receive this error due to one of several reasons:

- You copied the number incorrectly.

- You are only allowed to install LIWC2015 on two of your computers at time. You must **deactivate** your copy of LIWC2015 prior to installation on a new computer. This can be done within the LIWC2015 application under the **Help** menu.

- You have purchased a lease license of LIWC2015 (rather than purchased a full copy) and your time limit has expired. You will need to reactivate your program by going to

http://www.LIWC.net and purchasing a new serial number. Your lease contract begins on the day of purchase.

Problems with External Dictionaries

One of the most common causes of problems when loading an external dictionary is improper formatting of the dictionary file. The most common problems in building external dictionaries include:

- Forgetting to put "%" at the top of the file and another "%" after the language categories (and before the actual words you want the program to count). Read the earlier section on **Creating Custom Dictionaries**.

- In the top part of the dictionary, having language categories with two or more words. A category can only be a single word.

- In the bottom half of the dictionary, having blank lines or lines with only a search word with no corresponding language category number.

- Having stray tabs, punctuation, or letters on one or more of the lines in either the top or bottom of the dictionary.

MS Word Dictionary Files

In the earlier versions of LIWC, all dictionary files were required to end in ".dic". No longer. You can now write your dictionary in MS Word and save it as a ".docx" or ".doc" file. You can also add the extension .txt (if it is a text file). Or .wbl (if it is a werble file -- warning, there is no such thing as a werble file. This is LIWC2015 dictionary-making humor).

> **Important warning**: It is strongly advised that any dictionary files you make are so labelled so that they are not confused with other document or text files. If you have made a dictionary to capture feelings of guilt, you might name it "Guilt-Dictionary.docx." It is also recommended that you save all of your dictionaries in a single directory on your computer.

Plain Text Dictionary Files

It is also fine to save your dictionary files in text format using either a standard word processing program such as MS Word or in a readily-available text editor such as Notepad or Notepad++ (https://notepad-plus-plus.org/). When finished, you can save your dictionary with the extension ".dic" or ".txt". To make your life simpler, re-read the **Important Warning** above.

Problems with Displaying LIWC Results inside of Application

The results show normal word counts, but results are mostly zeroes

If your results file shows that LIWC2015 is counting the number of words but not recognizing the individual words in your text files, your LIWC2015 encoding setting might need to be changed. Ensure that LIWC2015 is set to the correct encoding under the **Settings** menu. If issues continue to persist, try saving the files with a different encoding (e.g., UTF-8), then set the LIWC2015 encoding accordingly prior to reanalyzing your files.

It is also possible that you are analyzing files that were saved in a very old version of Word (Word 6 or earlier). If this is the case, simply open each in a newer version of Word and re-save them.

Results do not display, or take a long time to display

LIWC2015 is relatively quick. The entirety of all 37 plays by Shakespeare take about 22 seconds to analyze, depending on your computer. If your processor is slow and you have very little memory, very large files may take several seconds or up to a minute to analyze. Processing time also depends on the number of files you are analyzing. Unless you are analyzing many hundreds of thousands of files at a time, this is unlikely to be an issue.

Similarly, spreadsheet files that are too large may cause LIWC2015 to become slow when presenting results. Spreadsheet files that have many hundreds of thousands of rows may need to be divided into two or more files in some cases.

If the program freezes, it is possible that you are trying to process a file that has been corrupted in some way. Try to isolate the file that LIWC is having trouble with and determine if its contents are problematic.

One or more of the rows in my results are blank

If your output contains blank rows, the most common cause is that you have one or more text files that are in an unsupported format. To see additional error details, double-click on the left-most cell in the blank row (i.e., cell under the "Filename" column). This most commonly occurs when trying to analyze older word processing documents, such as files saved in versions of MS Word prior to 2003.

Viewing the Internal LIWC dictionaries

You can view and download the words associated with each of the language categories associated with the LIWC2015, LIWC2007, and LIWC2001 internal dictionaries. All three can be exported by going to the **Dictionary** menu and clicking on **Export Internal Dictionaries**. When you export the internal dictionaries, copies of the internal default dictionaries are made in your specified location. These exported dictionaries are not the actual dictionaries that the LIWC program runs.

Important reminder: LIWC dictionaries are for research purposes only and cannot be shared with other labs. Any commercial use of the LIWC dictionaries is forbidden. If you would like to use

part or all of any LIWC dictionary, contact LIWCpermissions@receptiviti.com. Note that Receptiviti, Inc. oversees all commercial use and licensing of LIWC.

New Help for Old Friends

If you are familiar with LIWC2007, this section is for you. There are some changes that will delight you and a few that might annoy you at first. Let's get to it.

Change in some variable names. All variable names are now singular. Instead of the old "preps" and "articles", they are now called "prep" and "article." The old cognitive mechanisms (cogmech) dictionary has been renamed the cognitive processes (cogproc) dictionary.

Word counting algorithm update. The ways words are counted are improved and are slightly different from LIWC2007.

New language formats can be processed. LIWC can now run non-image-based pdf and rtf files. The biggest advance is the ability to analyze text within Excel spreadsheets in .xls, .xlsx, and .csv formats. See section earlier in this document for more information.

The output has a different appearance and saves differently. You can figure this out on your own.

Make your own dictionaries using Word. It's easy. Read the instructions on how to do it.

The Mac and PC versions are virtually identical. Look at word categories on a PC the way you could on the Mac in LIWC2007.

Big Data friendly. LIWC2015 can analyze giant data sets and can probably handle as many files or lines of text as you could dream of. (This last sentence does not apply to computer scientists or engineers).

Changes in our dictionaries and categories:

Small changes in most of the LIWC dictionaries. Using more current statistical methods, we have been able to omit words that were not being classified correctly and add words that previous versions missed. Virtually no LIWC2015 dimension is completely identical to its LIWC2007 counterpart.

Forced retirement of some language categories. Sadly, the following variables have been removed from LIWC2015:

- inclusive -- weak psychometrics, most are now in Conjunction category
- exclusive -- terrible psychometrics; a new, better version is Differentiation
- inhibition -- very low base rate category
- past tense verb -- updated to a more inclusive Past Focus category
- present tense verb -- updated to a more inclusive Present Focus category
- future tense verb -- updated to a more inclusive Future Focus category
- humans -- broken down to far more interesting Female and Male categories

Welcoming a new set of language dimensions. The new LIWC2015 dimensions better capture important linguistic, psychological, and social processes. These include:

- Four summary variables based on algorithms that the Pennebaker Lab has published over the last several years. References to the relevant papers are available in the Pennebaker, Boyd, Jordan, Blackburn, and Chung (2015) language manual. Due to prior commercial agreements, the precise algorithms are not available. Note that the four summary variables have been re-scaled so that they reflect a 100-point scale ranging from 0 to 100. The variables include:
  - Analytical thinking -- a high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking.
  - Clout -- a high number suggests that the author is speaking from the perspective of high expertise and is confident; low Clout numbers suggest a more tentative, humble, even anxious style.
  - Authentic -- higher numbers are associated with a more honest, personal, and disclosing text; lower numbers suggest a more guarded, distanced form of discourse.
  - Emotional tone -- a high number is associated with a more positive, upbeat style; a low number reveals greater anxiety, sadness, or hostility. A number around 50 suggests either a lack of emotionality or different levels of ambivalence.
- Adjectives -- based on the most commonly-used adjectives across multiple formal and informal text samples.
- Comparison words -- where the author is comparing one entity with another, includes words like as, bigger, best, smaller, etc.
- Interrogatives -- who, what, where, how, etc.
- Female references -- pronouns (she, her, hers) and nouns (female, woman, girl, etc.)
- Males references -- pronouns (he, his, him) and nouns (male, man, boy, etc.)
- Differentiation -- a class of words where people are distinguishing between entities, people, or ideas, including words such as "but", "without", and "hasn't".
- Drives -- an overarching dimension that captures the following needs, motives, drives:
  - Affiliation -- McClelland-like dimensions including reference to others
  - Achievement -- references to success and failure, achievement striving
  - Power -- references relevant to status, dominance, social hierarchies
  - Reward focus -- references to rewards, incentives, positive goals, approach
  - Risk focus -- references to dangers, concerns, things to avoid
- Time orientation variables:
  - Past orientation -- a mix of past-tense verbs and references to past events/times
  - Present orientation -- present-tense verbs and references to present events/times
  - Future orientation -- future-tense verbs and references to future events/times
- Netspeak -- Inclusion of words frequently uses in social media and text messaging (bae, lol, 4ever) and very basic punctuation-based emoticons, such as :) and ;). Emojis and html or image-based emoticons are not included. These words and symbols are included in the appropriate dictionaries. For example, "b4" is coded in both the preposition and netspeak categories.

Technical Support

Technical support for set-up and hardware/software compatibility can be obtained by sending an email to info@liwc.net. Further assistance is available from the first author, James W. Pennebaker, Department of Psychology, The University of Texas, Austin, Texas 78712 (Pennebaker@psy.utexas.edu), Roger Booth, School of Medical Science, The University of Auckland, Auckland, New Zealand (rj.booth@auckland.ac.nz), or Ryan L. Boyd (ryanboyd@utexas.edu). More extended consultation is available on a fee basis.