

V_LexSig v1.0: The Manual

Paul Meara

1. V_LexSig (Lexical Signatures) is a small utility program that takes as input two data sources. The first is a set of target words (this list will normally be a short list of 8-12 target words which are of interest to the researcher). The second data source is a text which may contain some of the target words or not. The program returns a binary description of the text in terms of the target word list (e.g. 0 1 0 0 1 0 1 1, where 1 means the word is present in the text and 0 that it is not).

V_LexSig works with a single text. The companion program, **SigSorter**, works with collections of Lexical Signature descriptions generated by V_LexSig and allows you to perform some simple counts and sorting operations on sets of texts described in this way.

2. The V_LexSig workspace screen is shown below. V_LexSig requires two pieces of data: a short list of words and a text. V_LexSig will work comfortably with short texts generated by L2 speakers, up to about 500 words in length. It will also work in languages other than English.

3. You may need to do some editorial tidying of the text that you want to work with. For example, you may want to mark some words as lexical phrases, rather than separate words. Use an underscore to do this: e.g. *planning_permission*.

The screenshot shows the V_LexSig workspace interface. At the top left is a book icon, and at the top right is the text "_lognostics". The main title is "Lexical_Signatures". Below the title is a description: "Lexical_Signatures checks a text for the occurrence of a small set of target words and returns a binary string that describes the text in terms of these target words." The form contains the following fields and buttons:

- "paste your target words here:" followed by a text input field.
- "name of the target list:" followed by a text input field.
- "paste your text here:" followed by a large text area.
- "name of the text:" followed by a text input field.
- "ignore these characters" followed by a text input field containing "...!?"
- "START" and "CLEAR" buttons.
- Copyright notice: "© 2015 Paul Meara" in the bottom left corner.

4. You can type the target word set directly into the program or copy and paste it from another source. The short text for analysis should be in .txt format. Copy and paste the text you want to analyse into the larger text box. Next, provide a short identification code for this text in the box marked **name of the text**.

Your text may include some characters that you want to ignore. A list of common punctuation marks is included in the box labelled **ignore these characters**. If your text contains other characters that you want to ignore you can add them to this list.

5. Click on the **Start** button to process the text.

6. Click on the **Clear** button to reset the workspace.

7. An example of V_LexSig report page is shown below. This report shows that *ThreeBears.txt* was tested against Target Word List 'TList12'. This list was made up of 10 words, and *ThreeBears.txt* included word1, word2, word5, word6, word7 and word8. Word3, word4, word9 and word10 did not appear in this text.



_lognostics

Lexical Signatures

Report for text: ThreeBears.txt
Target word list: TList12

1100111100/TList12/ThreeBears.txt

Copy this report string into a text file for later use

NEW TEXT

© 2015 Paul Meara

If you are examining a large number of texts, then you should open a text file and keep all the report strings for your texts in this file. Files of this sort can be processed using the **SigSorter** program (also on Lognostics website).

8. Click on the **New text** button to analyse another text.

Background reading:

Meara, P.M., Jacobs, G. and Rodgers, C. (2002) Lexical Signatures in foreign language free-form texts. *ITL Review of Applied Linguistics* 135-136, 85-96.

Meara, P.M., Rodgers, C. and Jacobs, G. (2000) Computational assessment of texts written by L2 speakers. *System* 28 (3), 345-354.

For a more detailed discussion of *V_LexSig*, see **PM Meara and I Miralpeix (2017)** *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.