

## D\_Tools v2.0: The Manual

*Paul Meara and Imma Miralpeix*

1. D\_Tools computes a statistic that reflects the lexical richness of a text. Specifically, D\_Tools computes the value of Malvern and Richards' **vocd** statistic.
2. The D\_Tools **workspace** looks like this:

**D\_Tools**

D\_Tools reports a D score for short texts

paste your text here:

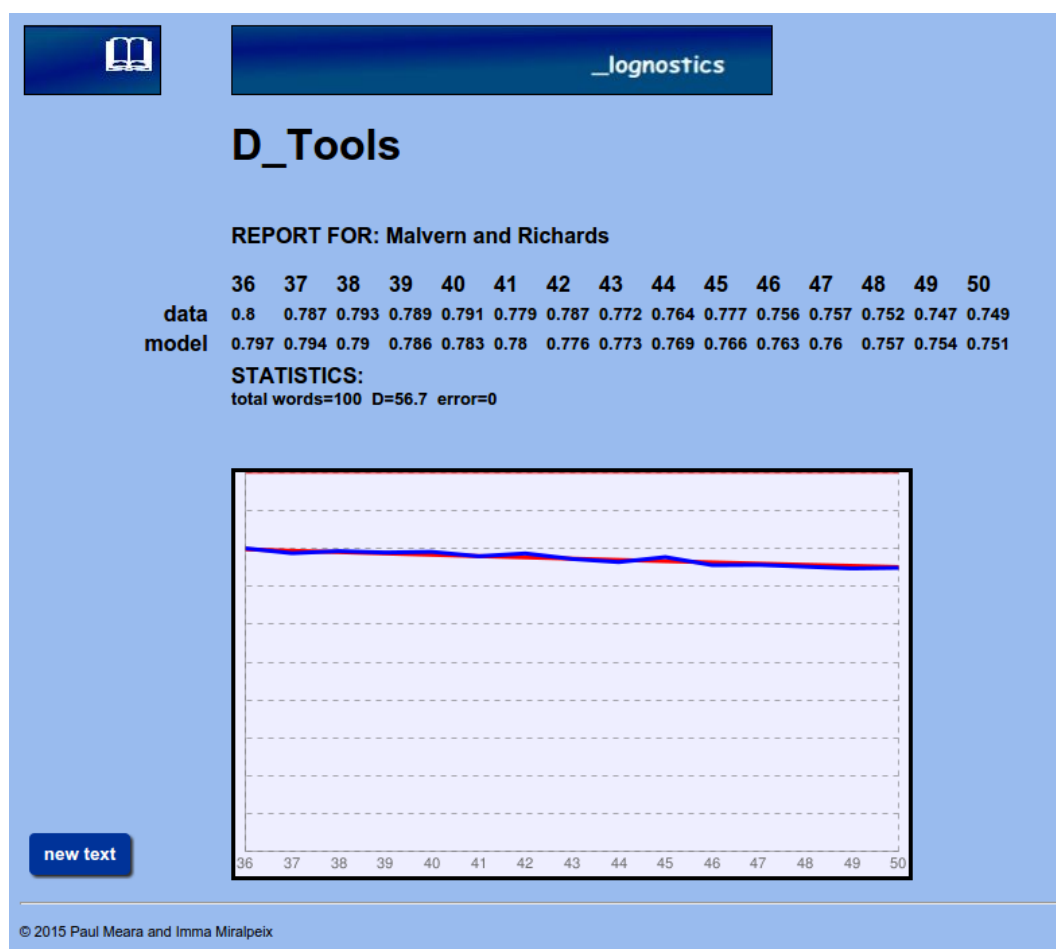
text name

ignore these characters

© 2015 Paul Meara and Imma Miralpeix

3. **Enter the text** you want to analyse into the text box. **Enter a name for your text** in the box immediately below the main text box.
4. You may need to **edit your text** for D\_Tools to function properly. Proper names like Abraham Lincoln should be entered as Abraham\_Lincoln if you want them to be analysed as single word units. If you are dealing with L2 texts, with a high level of errors, then you will need to decide how you should treat them. The two main options are correction and deletion.
5. You can add characters to the punctuation list at the bottom of the screen if you want D\_Tools to ignore them.
6. Click the **submit** button when you are ready to analyse your text.
7. **D\_Tools makes a report** on your text which looks like the figure on the next page. D\_tools works by taking a series of samples from the text and computing a Type-Token value for each of these samples. The program takes 100 samples of 35 words and computes a mean TTR for these samples. Then it takes 100 samples of 36 words and computes a mean TTR for those samples. This process is repeated with 100 samples of 37 words, 38 words, 39 words, and so on up to 50 words. D\_Tools then uses Malvern and Richards' formula to find a value of D which best matches this data set.

8. The D\_Tools report page looks like this:



The report shows the mean TTR score for each of the 15 sample sets (data), and the values that Malvern and Richards' formula generates for the best value of D (model). The graph shows the data values (blue) and the model values (red). Normally these two sets of values will be almost identical.

The report also details the number of words counted, the best estimate of D and an error score. The error score indicates how closely the model data matches the actual data. If this figure exceeds 0.1, then the model is NOT a good match for your data.

9. Note that D\_Tools takes random samples of words from your text, and this means that you will get small variations in the value of D if you run the program several times with the same text.

#### Background reading:

Malvern, D., Richards, B., Chipere, N. and Duran, P. (2004) *Lexical Diversity and Language Development: Quantification and Assessment*. Basingstoke: Palgrave Macmillan.

Read, J. (2005) Applying lexical statistics to the IELTS listening test. *Research Notes* 20, 12–16.

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