Documentation

Introduction

The TTR-Analyzer is a Tool to calculate, present and compare the relationship between *Wordtypes* and *Wordtokens* in given Texts. This "*Type-Token-Relation"* is commonly used in different academic fields, but its calculation requires either competence in programming or tedious (and error-prone) manual computing. The TTR-Analyzer aims to aid academics interested in working with TTRs.

Installation

Windows

Download the folder "TTR-analyzer-windows". It contains the file "TTR-analyzer.exe". To run the program, simply execute this file. No further installation necessary.

Linux/Mac

To use the tool on other platforms than windows, it has to be compiled manually. Download and install some version of Allegro Common Lisp. The "Free Express Edition" should suffice and can be downloaded from franz.com (http://franz.com/downloads/clp/survey).

After installation, run the IDE and open the TTR-Analyzer-Project, which you find in the folder "Source Code". You then can either compile the program or run it directly out of the IDE.

Functions

Type-Token-Relations

Type-Token-Ratio

A simple to understand an interpret index that compares the quantity of Types and Tokens. It's dependent on the text length. The longer text, the less influence has a new Type on the TTR. So it gets smaller and smaller, the longer the text goes on. Which means, that from two text, the significant larger one will have automatically a lower TTR (in most cases), regardless of quality.

The TTR-Analyzer can show the total TTR, or plot a graph displaying the full distribution of TTR-values over each single Text.

It is computed whit following formula: $Types \div Token$

KGM

The "Köhler-Galle-Method" is a more stable kind of measure that takes length and dynamic of a text into account. The final value doesn't differ from the TTR, but it graph does. So it's more useful if you want to look at the flow of information of chosen text.

It too can be drawn as a graph plot.

The formula is: $\frac{CountOfTypesAtPosition+CountOfTypesInText-\frac{Position*CountOfTypesInText}{text\ length}}{text\ length}$

MLTD

The “Measure of textual Diversity” is not used to measure any distribution over a text. Instead it tries to give a measure for vocabulary richness. It considers the text length. A higher MLTD stands for a higher the variance in the vocabulary. Other than the simple TTR, the shorter text won’t have a “better” MTLD.

With the MTLD being a measure, that calculates one single value over the whole text (and not parts of it); it is not useful to plot it as graph.

Stemming

When looking at the word in our text, there is more than one way to do identify the types. In some cases, we might simply want to define each different *word form* as a new type. In those cases *house* and *houses* would be seen as two types. This may be interesting when we for one want to know something purely about the variance of the text vocabulary. But in cases, where we (for example) want to take a look into the flow of information of a given text, the *word form* doesn’t suffice. Is it *house* or *houses*, it doesn’t matter, the author is still seemingly talking about the same thing, the same lexical meaning.
So we want to look at types of *lexemes*, rather than *word forms*.

While the TTR-Analyzer doesn’t feature a full-fledged lemmatizer, it is capable to find out the *stem* of each word via *stemming*(which is usually accurate enough).
It uses the well-known “Porter Stemmer” (http://tartarus.org/~martin/PorterStemmer/). The current version of TTR-Analyzer is able to stemm English, German and Russian texts.
It can be chosen individually for any one text if and how it should be handled.

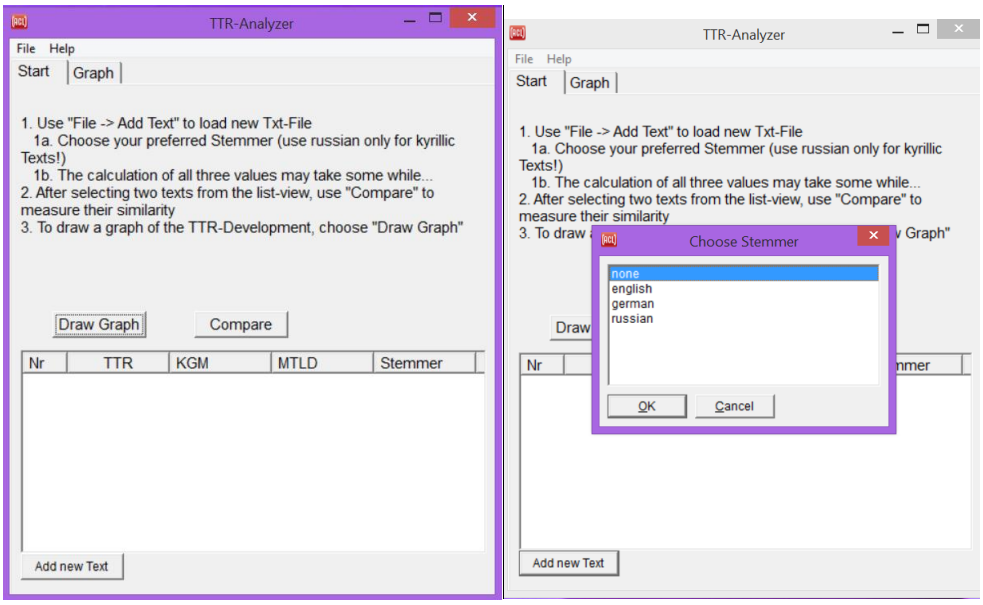
Correlation

Just looking at the graph of two TTR-Distributions is rarely enough to clearly see to what degree their values differ. That’s why the TTR-Analyzer comes with a build in function, which does just that.

Spearman's rho (http://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient) measures the correlation of two variables. It yields a value between 0 and 1, with 1 implying, that both data sets are identical. So the higher Spearman's rho, the more similar are the two text (regarding TTR/KGM).

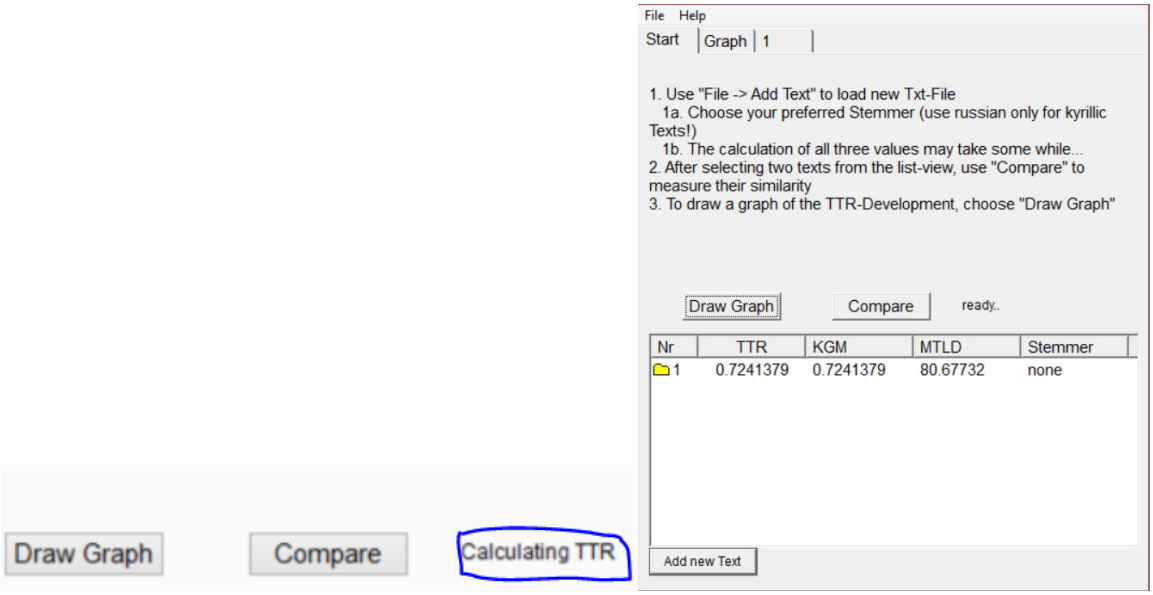
Tutorial

Although the TTR-Analyzer is quite easy to use, a quick Tutorial might help to get you familiar with the Tool.

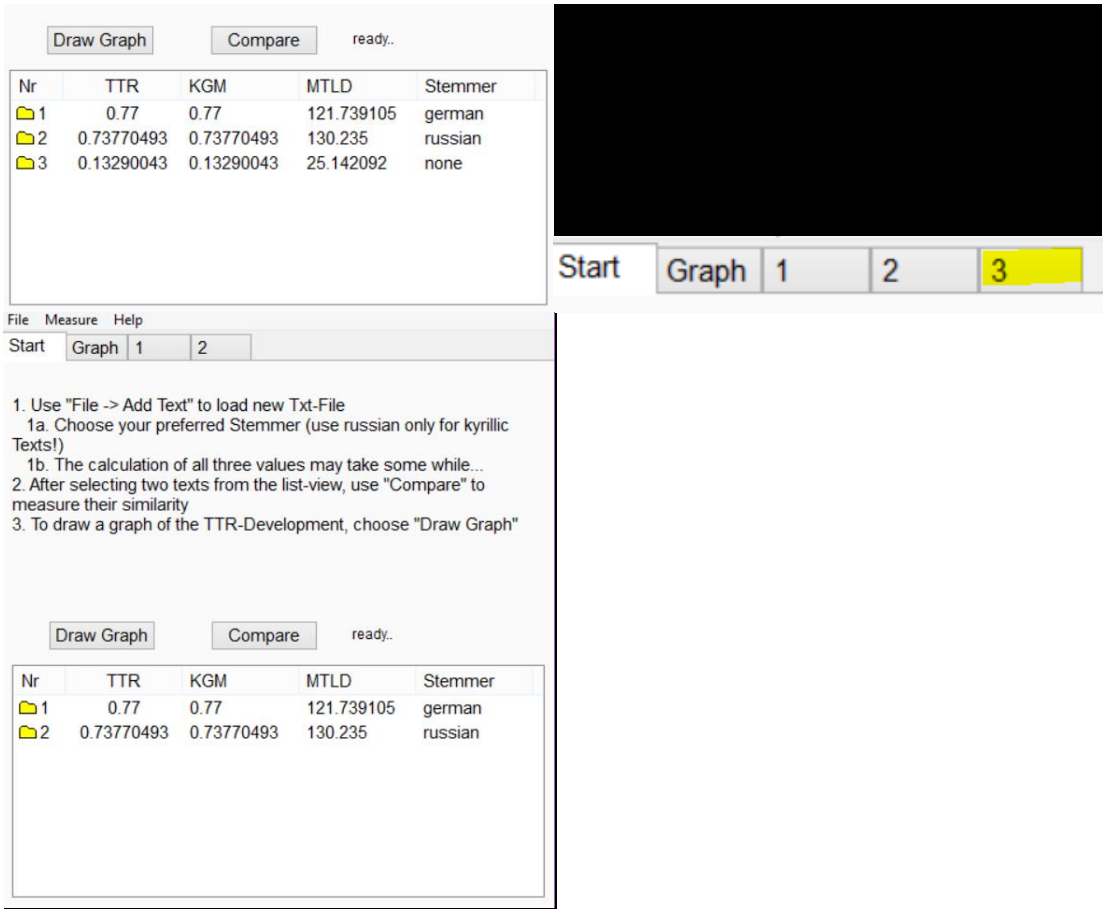


Clicking on “File-> Add Text” in the menu or the button “add new Text”, will first let you choose the text you want to work with and asks you afterwards, what method of stemming you want apply.

Depending on the size of the chosen text, it might take a while till the text is fully loaded into the tool. While loading, the current Calculation is shown right above the list-view. When ready, the values are shown in the list-view.



Additional Texts will be shown there too. The content of your uploaded files can be accessed via the Tab with the Texts-Number on it. To delete a loaded text, just double click on its tab.



The Button „Draw Graph“ draws a graph of the type, that’s selected via a popup-dialog
The Button “Compare” only works, when exactly two text are selected in the list-view.