Otto-Friedrich Universität Bamberg
Lehrstuhl für englische Sprachwissenschaft einschließlich Sprachgeschichte
Workshop Methods and Theories
Dozenten: Julia Schlüter, Fabian Vetter
Nazanin Zadegan Hassani, Jasmin Roth

Miniproject A17

1. Topic

=> examination of a possible correlation between the length of an English word and its frequency

2. Steps in detail

1. Click on project A 17 in the VC
2. Click on the link provided there, then choose the link "General lists..."
3. Again, choose the link "General lists..." and tick the boxes of **length, Freq_HAL** and **Log_Freq_HAL**
4. Scroll down, choose **E-Mail**, enter your E-Mail adress and click "Execute Query"
5. Open your E-Mails and download the file
6. Open the file with **Notepad++**
7. Press Ctrl + F, choose "Ersetzen", enter in *suchen nach* "," and in *ersetzen durch* ";", then click "Alle ersetzen"
8. Repeat step 7 and enter in *suchen nach* "." and in *ersetzen durch* ",", then click **speichern**
9. Open the file now with **Excel**
10. Choose the empty cell F, click **Formen => mehr Funktionen => Statistisch => KORREL => ** choose column B and C
11. Repeat step 10 and choose column B and D
12. Mark the columns A, B, C and D, click **Daten => Sortieren** (nach Freq_HAL absteigend)
13. We face two results

3. Interpretation of the results

- 0,080220165: This number indicates that there is nearly no correlation between word length and frequency.
- 0,35259482: This number indicates that there is kind of a correlation between word length and the logarithm of the frequency

WHY? => The distribution of word frequency in English is very uneven; therefore, no correlation comes up which is why we need the logarithm of the frequency in order to get a result.

This result demonstrates that **there is a tendency for shorter words to occur more often in the English language.**