**Inputs:**

* List of German Words (almost 2 Mio with diff gram.forms)
* Duden (library in python):
  + Meaning
  + Synonyms
  + Antonyms
  + Typical coocurance
  + Meanings, Examples, Idioms
* Child Corpora (with CRAN)
* German Corpora (maybe)

**To look at:**

* Frequencies of every prefix, how many words exist with a certain prefix
* (prob. With the nouns or verbs)
* Coocurancies with prepositions
* The most similar prefixes (in synonyms)
* The least similar prefixes (in antonyms)
* Sentiment Analysis (positive/ negative/ neutral meaning)
* Term frequency – inverse document frequency (TF–IDF looks at how frequently a word appears in a document and its importance relative to the whole set of documents) – Corpus analysis
* How many variants does usually one stem has (with how many prefixes and which ones) (and if there is one, what is the probability that one certain prefix comes after it is known that another one exists, Bayes, conditional probability)
* How different is the meaning of the same stem with different prefixes
* Topic modelling (probabilistic latent semantic indexing) ??