How much variance in writing competence can be explained by morphology?

Identification of indicators for competence assessment of students’ essays and development of a prototype for computer-assisted text analysis

Jasmine Bennöhr

Humboldt-Universität zu Berlin & Landesinstitut für Lehrerbildung und Schulentwicklung, Hamburg

Aims of the project
- identify indicators or subdomains for competence to assess students’ strengths and weaknesses
- report results of individual students to teachers who can then give feedback or plan lessons and cater to special needs accordingly
- initial investigations of indicators on the morphological level

Why start with morphology?
- categorizable building blocks of language which feature in every text
- less susceptible to spelling errors than entire words

Data
- students’ essays in the German language from the city of Hamburg’s longitudinal KESS (Kompetenzen und Einstellungen von Schülerinnen und Schülern) study
- supplemented with language test results for validation
- core dataset of 9000 texts (paper and pencil) which were rated along several dimensions
- there are approximately 400 texts of grade 4 in digital format whose language competence ratings were used for this analysis in combination with a language test result of grade 7

Methodology
- normalization of variables: count of morphemes per word
  The following statistics were used on the morpheme counts:
  1. correlations
  2. regression analysis
    - human assigned ratings as dependent variable
  3. factor analysis
    - identify morphological dimensions by qualitative analysis of factors
    - 10%-significance level achieved in each test

Preliminary Results
- use of prefix ‘zer’, suffixes ‘lich/e’, ‘ion’ and ending ‘rt’ as in ‘funktioniert’ are correlated with high ratings of language competence
- use of prefix ‘ver’ and endings ‘rt/ren’, ‘nen’, ‘ten’ correlate with high test results in grade 7
- two even more robust morphemes are identified; robust in the sense that they correlate with both the rating of language competence in grade 4 and additionally the language test results in grade 7
- use of prefix ‘ent’ is a robust indicator of high competence
- use of prefix ‘ge’ is a robust indicator of low competence
- regression analysis: a model with six variables explains 35.8 % of the total variance of essay ratings
- results of factor analysis are thus far unsatisfactory: components are not easily interpretable. E.g. desirable dimensions such as ‘suffixes which indicate past tense’ have not yet been observed.

Limitations
- spelling errors are currently not taken into account (e.g. versuchen vs. fersuchen)
- searching for morphemes may yield strings that can be mistaken for morphemes (e.g. 1. Ver-se vs. Versehen; 2. Ver-d instead of Pferd)

Further work
- continue digitalization of texts to increase sample size
- deal with spelling errors: correct manually or use of prefix ‘ver’ and endings ‘rt/ren’, ‘nen’, ‘ten’ correlate with high test results in grade 7
- two even more robust morphemes are identified; robust in the sense that they correlate with both the rating of language competence in grade 4 and additionally the language test results in grade 7
- use of prefix ‘ent’ is a robust indicator of high competence
- use of prefix ‘ge’ is a robust indicator of low competence
- regression analysis: a model with six variables explains 35.8 % of the total variance of essay ratings
- results of factor analysis are thus far unsatisfactory: components are not easily interpretable. E.g. desirable dimensions such as ‘suffixes which indicate past tense’ have not yet been observed.

References