Other linguistic distances

- Lexical
- Syntactic
- Morphological

Lexical distances

Simple measure:
% cognates in a representative language sample (text, corpus, frequency word list etc.)

Cognate:
e.g. English butter Dutch boter

Non-cognates:
e.g. English butter Danish smør

Lexical distances

To be considered:
- Number of words
- Speaking style
- Lexical domain
- Word class
- Content word vs. function word
- Frequency
- ...

Lexical distances

Can be asymmetric:
Example 'room'

Swedish    Danish
rum         rum
varsel
Relating lexical distances to intelligibility scores

Stimulus material:
› news item (256 to 262 words)
› Danish, Norwegian and Swedish; Dutch, Frisian and Afrikaans

Test persons:
› 16 groups (in total 488) high school pupils from 9 places in Scandinavia, aged between 16 and 19 (average 17.0)
› 5 groups (in total 81) Dutch, Frisian or Afrikaans high school pupils, aged between 16 and 17 (average 16.6)

Task:
› answer 5 open questions

Scores:
› completely correct: 2 points
› partly correct: 1 point
› incorrect: 0 points
› maximum score: 10 points per listener (= 100%)
Relation between intelligibility and Levenshtein distances

\[ r = -0.64 \]

Relation between intelligibility and lexical distances

\[ r = -0.36 \]

Relation between intelligibility and lexical distances

Linear regression analysis with the independent variables phonetic distance and lexical distance:

\[ R = -0.71 \]

Lexical distances do not contribute significantly

Relative importance of phonetic and lexical distances

Experimental testing:

Four versions of an utterance:

1. Lexicon of language A pronounced by a speaker of A
2. Lexicon of language B pronounced by a speaker of A
3. Lexicon of language A pronounced by a speaker of B
4. Lexicon of language A pronounced by a speaker of A

Syntax, morphology?