


 rijksuniversiteit
 groningen

27-1-2009 | 1

Other linguistic distances

Other linguistic distances

- > Lexical
- > Syntactic
- > Morphological


 rijksuniversiteit
 groningen

27-1-2009 | 2


 rijksuniversiteit
 groningen

27-1-2009 | 3

Lexical distances

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*


 rijksuniversiteit
 groningen

27-1-2009 | 4

Lexical distances

27-1-2009 | 5

To be considered:

- > Number of words
- > Speaking style
- > Lexical domain
- > Word class
- > Content word vs. function word
- > Frequency
- > ...


 rijksuniversiteit
 groningen

Lexical distances

27-1-2009 | 6

Can be **asymmetric**:

Example 'room'

<u>Swedish</u>	<u>Danish</u>
<i>rum</i>	<i>rum</i>
	<i>værelse</i>

Lexical distances should be measured from
 A to B and from B to A


 rijksuniversiteit
 groningen


 rijksuniversiteit
 groningen

27-1-2009 | 7

Relating lexical distances to intelligibility scores

Relating lexical distances to intelligibility scores

27-1-2009 | 8

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)


 rijksuniversiteit
 groningen

Relating lexical distances to intelligibility scores

27-1-2009 | 9




 rijksuniversiteit
 groningen

Relating lexical distances to intelligibility scores

27-1-2009 | 10




 rijksuni
 groningen

Relating lexical distances to intelligibility scores

27-1-2009 | 11

Task:

- > answer 5 open questions

Scores:

- > completely correct: 2 points
- > partly correct: 1 point
- > incorrect: 0 points
- > maximum score: 10 points per listener (= 100%)


 rijksuniversiteit
 groningen

Correlations between linguistic distances and intelligibility

27-1-2009 | 12

Dependent variable:

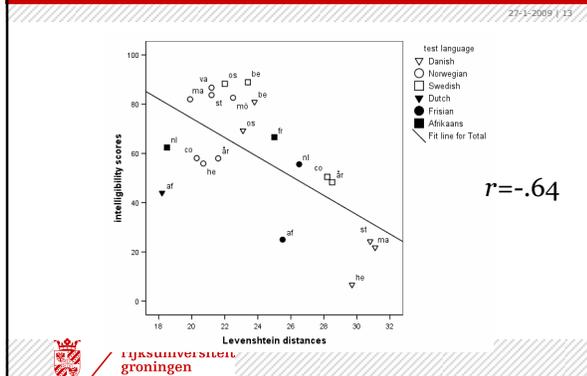
- > % correct answers per listener group and language

Independent variables:

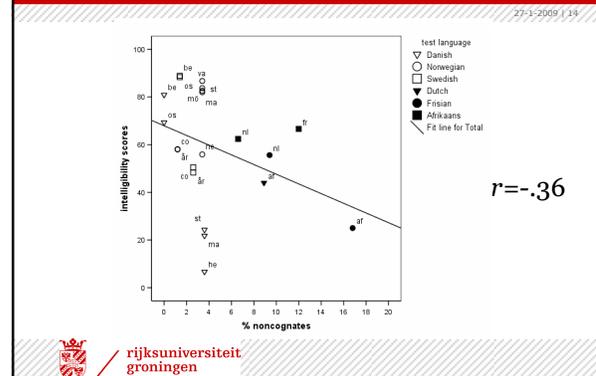
- > phonetic distances (Levenshtein) between test languages and native language of the test persons
- > lexical distances (% non-cognates)


 rijksuniversiteit
 groningen

Relation between intelligibility and Levenshtein distances



Relation between intelligibility and lexical distances



Relation between intelligibility and lexical distances

27-1-2009 | 15

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

$R = -.71$

Lexical distances do not contribute significantly

rijksuniversiteit groningen

Relation between intelligibility and lexical distances

- 27-1-2009 | 16
1. Intelligibility can to a high extent be predicted by phonetic distances
 2. Lexical distances play a smaller role in the intelligibility of Scandinavian languages because there are so few differences
 3. There are larger lexical differences between Dutch, Frisian and Afrikaans and they seem to have a larger effect on intelligibility
- rijksuniversiteit groningen

Relative importance of phonetic and lexical distances

27-1-2009 | 17

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?

rijksuniversiteit groningen