


 rijksuniversiteit
 groningen

27-1-2009 | 1

Relating Levenshtein distances to intelligibility scores

Research question

27-1-2009 | 2

- How well can Levenshtein distances predict intelligibility?


 rijksuniversiteit
 groningen

Investigations

27-1-2009 | 3

Beijering, Gooskens & Heeringa (2009):
 The intelligibility of 18 Nordic language varieties among Danes

Kürschner, Gooskens & Van Bezooijen (2009):
 The intelligibility of 384 Swedish words among Danes


 rijksuniversiteit
 groningen


 rijksuniversiteit
 groningen

27-1-2009 | 4

The intelligibility of 18 Nordic language varieties among Danes

Material

27-1-2009 | 5

- recordings of "The North Wind and the Sun"
- 18 Nordic language varieties
- mean 98 words
- phonetic transcriptions of the cognates (historically related words)


 rijksuniversiteit
 groningen

Material

27-1-2009 | 6

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.





 rijksuniversiteit
 groningen

Material

27-1-2009 | 7

8 Norwegian varieties

rijksuniversiteit
 groningen

Material

27-1-2009 | 8

Standard Swedish (Stockholm)
+
3 Swedish varieties

rijksuniversiteit
 groningen

Material

27-1-2009 | 9

Standard Danish (Lyngby)
+
3 Danish varieties

rijksuniversiteit
 groningen

Material

27-1-2009 | 10

Finland Swedish (Helsinki)

rijksuniversiteit
 groningen

Material

27-1-2009 | 11

Faeroese (Torshavn)

rijksuniversiteit
 groningen

Intelligibility

27-1-2009 | 12

Stimulus material:

- > 6 sentences in 6 varieties (latin square design)

Test persons:

- > 18 groups of high school pupils from Copenhagen, aged between 15 and 20 (average 17,6)

Task:

- > translate word for word into Standard Danish

rijksuniversiteit
 groningen

Correlation

27-1-2009 | 13

Dependent variable :

- › % correctly translated words per variety

Independent variable:

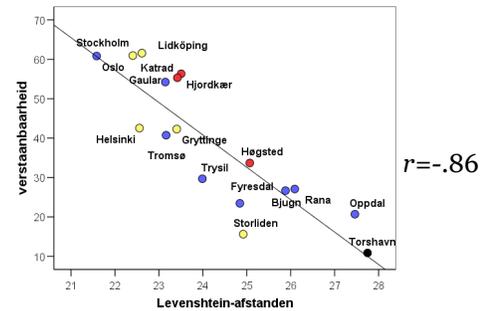
- › Levenshtein distances between Standard Danish and the 18 varieties



rijksuniversiteit
groningen

Relation between intelligibility and Levenshtein distances

27-1-2009 | 14



rijksuniversiteit
groningen

Relation between intelligibility and Levenshtein distances

27-1-2009 | 15

- › It is possible to predict intelligibility of language varieties to a high extent by means of Levenshtein distances



rijksuniversiteit
groningen

Consonants vs vowels

27-1-2009 | 16

- › Are consonants or vowels more important for the intelligibility?



rijksuniversiteit
groningen

Consonants vs vowels

27-1-2009 | 17

- › Word recognition in English depends more on correct consonant identification than on the correct identification of vowels (Van Ooijen 1994 and references).
- › **Consonants** are more important for the semantic identity of a word than **vowels**; they function as reference points in words.



rijksuniversiteit
groningen

Consonants vs vowels

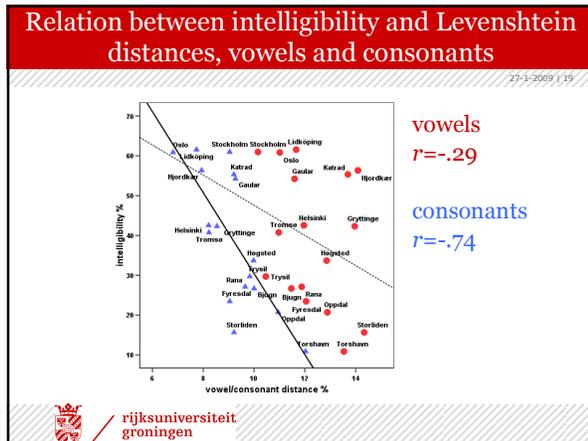
27-1-2009 | 18

Hypothesis:

- › Deviations in vowels are less damaging for the intelligibility of a closely related language than deviations in the consonants.



rijksuniversiteit
groningen



The intelligibility of 384 Swedish words among Danes

rijksuniversiteit groningen

27-1-2009 | 20

- ### LRS experiment
- 27-1-2009 | 21
- Material:**
- > Recordings of 347 frequent Swedish cognates
 - > Phonetic transcriptions
- Test persons:**
- > 38 Danish high school pupils aged 16-19
- Task:**
- > Translate words into Standard Danish
- rijksuniversiteit groningen

- ### Correlation
- 27-1-2009 | 22
- Independent variable:**
- > Levenshtein distances between Swedish and Danish words
- Dependent variable :**
- > % correct translations per words
- rijksuniversiteit groningen

Relation between intelligibility and Levenshtein distances

27-1-2009 | 23

- > $r = -.33$
- > Other factors must be considered in order to predict intelligibility at the word level

rijksuniversiteit groningen