

Swedish-Danish word intelligibility

Sebastian Kürschner, Charlotte Gooskens & Renée van Bezooijen
University of Groningen



university of
 groningen
 faculty of arts

The Thirteenth International Conference on Methods in Dialectology, Leeds, 4-8 August 2008
Workshop "Measuring linguistic relations between closely related varieties"

Danish and Swedish:

- Closely related, neighboring standard varieties
- In principle mutually intelligible

Research questions:

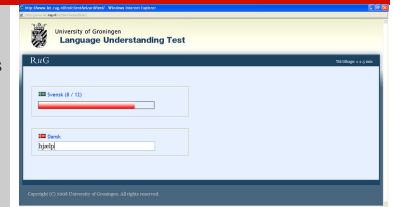
1. How well do Danes understand isolated Swedish words?
2. Which factors can explain the intelligibility of single words?

Material

345 cognate nouns chosen randomly from a list of 2500 highly frequent words

1. Intelligibility test

- Internet-based experiment on **word intelligibility**
- **Danish subjects** are auditorily presented with 384 nouns in Swedish (96 words per subject)
- **Task:** Write the Danish translation into a text field
- **Subjects:**
 - N = 29 (24 from Jutland, 5 from Funen)
 - Danish pupils, age 16-19 (mean 17)



Results

58.5 % correct answers

Test yourself!

<http://www.let.rug.nl/lrs/>
Login: **germanic**, password: **guest**

2. Explaining factors

Levenshtein-distance

- String edit distance used in Dialectometry
- Calculates the number of least necessary operations for mapping two sound strings
 - Insertions
 - Deletions
 - Substitutions

Version used here:

- Calculates a value of 0.5 if a difference is solely based on length, tenseness, or diphthongization sharing the same onset
- Correlation with correct answers: $r = -.32^{**}$

Example for calculating Levenshtein distance

Danish *guld* [gu1] vs. Swedish *guld* [gʉ1d] ,Gold^s

g	u	l	
g	ʉ	l	d
0	1	0	1

= 2/4 = 50 %

Words with „rare“ sounds

- Swedish sounds not existing in the Danish phoneme system
- Cf. retroflexes, [ʃ]
- Correlation with correct answers: $r = -.15^{**}$

Word-length (Swedish words)

- Correlation with correct answers: $r = .17^{**}$

Different word accent

- Cf. Da. *al'ternativ* vs. Sw. *alterna'tiv*
- Correlation with correct answers: $r = .02$

Different syllable number

- Cf. Da. *tegn* vs. Sw. *tecken*
- Correlation with correct answers: $r = .26^{**}$

Prosodic features

- Swedish tone accents ($r = .02$)
- Danish ,stod' ($r = .02$)

Neighborhood density

- Number of existing Danish words which deviate from the Swedish stimulus in only one sound
- Correlation with correct answers: $r = .14^{**}$

Etymology

- Native words vs. loanwords
- Correlation with correct answers: $r = -.10$

Orthography

- Danish test persons may have help from the conservative Danish orthography in identifying Swedish words

Swedish	l	a	n	d
Danish	l	æ	n	?
Orthography	l	a	n	d (2)

- Correlation with correct answers: $r = .15^{**}$

Word frequency in Danish

- Relative frequency in a written corpus
- Correlation with correct answers: $r = .08$

3. Multiple regression analysis

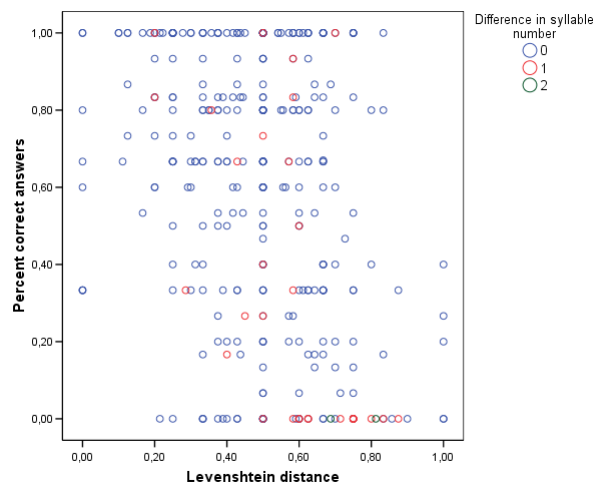
Enter mode

- $R = .49$, $df = 280$, $p < .001$

Stepwise mode

5 models:

1. Different syllable number
 $R = .31$, $R^2 = .09$, $p < .001$
2. + Levenshtein distance
 $R = .40$, $R^2 = .16$, $p < .001$
3. + Orthography
 $R = .43$, $R^2 = .19$, $p < .001$
4. + Frequency
 $R = .45$, $R^2 = .20$, $p < .001$
5. + Neighborhood density
 $R = .46$, $R^2 = .22$, $p < .001$



4. Conclusions

- Only 22 % of the variation in intelligibility scores can be explained by the factors considered
- Other linguistic factors to be considered
 - Help or confusion through language skills (cf. Sw. *brött* ,crime', translated as Da. *brød* ,bread' through German *Brot*)
 - Levenshtein distance which weights factors according to the assumed importance of their values for intelligibility, cf.
 - consonants > vowels,
 - stressed syllables > unstressed syllables,
 - etc.
- In the aggregate, nevertheless, Levenshtein distance correlates strongly with intelligibility: $r = .80$ (Gooskens 2006) for Scandinavian varieties translated by pupils from Copenhagen