rijksuniversiteit groningen

Linguistic determinants of the intelligibility of Swedish words among Danes

› Sebastian Kürschner
› Charlotte Gooskens
› Renée van Bezooijen

Idea

- Linguistic factors have been identified to explain intelligibility between closely related languages/varieties in prior research
- Prior research based on aggregate differences of whole text understanding between varieties
- Assumption: To understand texts in a foreign variety, you need to be able to identify a certain amount of words
- Aim: Identify the linguistic factors which determine the intelligibility of single words, presented in isolation

rijksuniversiteit groningen

Languages

- Danish listeners confronted with Swedish words
- Danish and Swedish are well-known to be mutually intelligible to some degree

rijksuniversiteit groningen

Explaining factors

- We included only **linguistic factors**
  - **Attitudes** were not tested, but usually do not correlate strongly with intelligibility scores
  - We excluded the results of subjects who probably had much **prior experience** with Swedish
- Linguistic factors based on
  - general results on word intelligibility from psycho-phonetic literature
  - language-specific factors identified in literature on Swedish and Danish

rijksuniversiteit groningen

Intelligibility experiment

Test words:

- 347 cognate nouns
- randomly selected from a database with parallel lists of 2575 frequent spoken words
- concepts known to secondary school pupils between 15 and 19 years of age
- read aloud by a male speaker of Standard Swedish

rijksuniversiteit groningen

Intelligibility experiment

Subjects:

- 36 native speakers of Danish
- pre-university education
- aged 16 to 19

Task:

- translation of Swedish test words into Danish
- via internet

rijksuniversiteit groningen

Intelligibility experiment

Design:

- 4 groups of subjects
- 5 to 19 subjects per group
- 96 test words per group

Factors considered for explanation

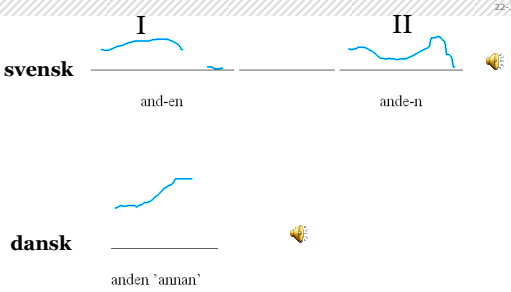
- Levenshtein distance
  - normalized
  - 0.5 for
    - segment length: [u] vs. [uː]
    - [+/-tense]: [u] vs. [ʊ]
    - [R] vs. [r]
- Foreign sounds

ɧ fx sv. *sjukdom*[ɧʉːkdɔmː] da. *sygdom* [syːdɔm]
retroflekser fx sv. *jord* [juːɖ] da. *jord* [joːˈɒ]

Factors considered for explanation

- Word length
- Word accent differences
  - Danish '*kontekst* vs. Swedish *konˈtext* 'context'
- Difference in Syllable number
  - Sw. *choklad* vs. Danish *chokolade* 'chocolate'

Factors: Word accent differences



Factors: Stød

fx	+ stød	- stød
	hvalen	*valen*
	køber	*køber*
	musen	*musen*
	skal	*skal*
	tør	*tør*

Factors

- Neighborhood density
  - Sw. *säng*
  - Da. *seng* + *syng* 'sing', *senge* 'beds', *hæng* 'hang', *and stæng* 'close'
- Etymology
  - native vs. loan word
- Orthography
  - Danish *hånd* [hɔnˀ] vs. Swedish *hand* [hand]
- Danish word frequency

Results

› Intelligibility: 57.7 % correct identifications of the cognates

› Explaining factors:

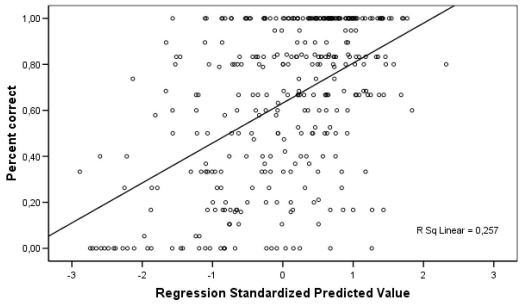
Table 1: Correlation of the intelligibility scores with linguistic factors

Factor	Correlation (*r*)	Significance (*p*)
Levenshtein distance	-.33	< .001
Foreign sounds	-.13	< .05
Word-length of Swedish words	.25	< .001
Word accent difference	.01	
Difference in syllable number	-.25	< .001
Lexical tones	.04	
Stød	-.02	
Neighbourhood density	-.17	< .01
Etymology	-.1	
Orthography	.21	< .001
Word frequency	.04	

Results: Multiple regression Dependent: Intelligibility scores

method	entered variables	model	result
enter	all linguistic factors	all linguistic factors	$R = .51$, $R^2 = .26$ $p < .001$
Stepwise 1st step	all linguistic factors	Levenshtein distance	$R = .34$, $R^2 = .11$ $p < .001$
Stepwise 2nd step	all linguistic factors	Levenshtein distance, orthography	$R = .44$, $R^2 = .19$ $p < .001$
stepwise 3rd step	all linguistic factors	Levenshtein distance, orthography, different syllable number	$R = .47$, $R^2 = .22$ $p < .001$
stepwise 4th step	all linguistic factors	Levenshtein distance, orthography, different syllable number, neighbourhood density	$R = .49$, $R^2 = .23$ $p < .001$
Stepwise 5th step	all linguistic factors	Levenshtein distance, orthography, different syllable number, neighbourhood density, word frequency	$R = .50$, $R^2 = .24$ $p < .001$

Scatter plot



Discussion

› Why can the variation not be explained to a higher degree?

- unpredictable factors in single words (voice quality, speech rate, etc.) which are levelled in aggregate scores
- idiosyncratic characteristics of single words

Some characteristics

› Confusion with other languages learned

- Swedish *art* 'sort' - Danish *kunst* 'art'

› neighbourhood closeness

- Swedish *fel* [fe:l] 'mistake' translated as Danish *fæl* [fæ'l] 'foul' instead of correct *fejl* [faj'l]

› non-corresponding phonemes

- Swedish *stat* [sta:t] 'state' translated as Danish *start* [sda:'d] 'start' instead of correct *stat* [sdɛ:'d]
- Swedish *frukt* 'fruit' translated as Danish *frygt* 'fear' instead of correct *frugt*