



# Measuring linguistic distances between the Scandinavian languages

Charlotte Gooskens  
Scandinavian department  
University of Groningen

Workshop on 'The Nordic Corpus Bank', 1 Nov 2007  
Session I: The genres of the corpus canon

# Research project

26-11-2007 | 2

Title:

*Linguistic determinants of mutual intelligibility in  
Scandinavia*

<http://www.let.rug.nl/micrela/>

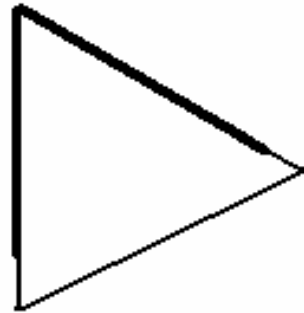


university of  
groningen

# Inter-Scandinavian intelligibility

26-11-2007 | 3

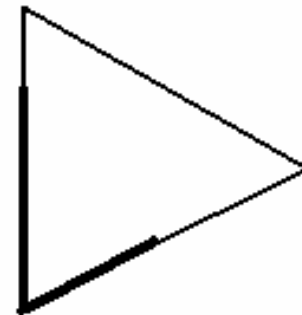
Norwegian listener



Danish

Swedish

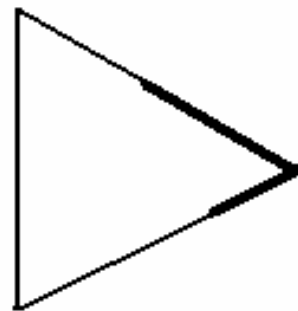
Norwegian



Swedish

Danish listener

Norwegian



Swedish listener

Danish



university of  
groningen

# Research questions

26-11-2007 | 4

- › How well can Scandinavians understand each other's vocabulary?
- › What is the role of different linguistic factors for word intelligibility?



university of  
groningen

# Research plan

26-11-2007 | 5

1. Frequency lists based on Danish, Swedish and Norwegian corpora
2. Annotation
3. Linguistic distance measurements
4. Word intelligibility tests
5. Relationship between linguistic distances and word intelligibility



university of  
groningen

# 1. Frequency lists based on Danish, Norwegian and Swedish corpora

26-11-2007 | 6

The three corpora must be constructed in the same way:

- › written & spoken
- › formal & informal
- › genres
- › size
- › lemmatized
- › ...



university of  
 groningen

# Spontaneous speech

26-11-2007 | 7

	Danish BySoc	Norwegian NoTa	Swedish GSLC
words	1.3 mio	900.000	1.3 mio
informants	80	165	371
years	1986-88	2004-2006	1978-2000
activity	interviews and conversations	interviews and conversations	20 different activities
speaker variation	controlled	controlled	less controlled



university of  
groningen

# Annotation

26-11-2007 | 8

- › frequency
- › word class
- › phonetic transcriptions
- › etymological information (native word, loan word)
- › [translations into the other two neighbouring languages](#)
- › [translations into other modern languages](#) (Engels, German, French...)
- › [cognates – non-cognates](#)
- › ...

Automatic or manual?



university of  
groningen



# Annotation

26-11-2007 | 9

lemma 'room'

word	translation
Da. <i>værelse</i>	Sw. <i>rum</i>
Sw. <i>rum</i>	Da. <i>rum</i>



university of  
groningen

# Annotation

26-11-2007 | 10

Words can be understood via related words in other languages:

Sw. *potatis* – Da. *kartoffel*– Eng. *potatoes*

Sw. *fåtöjl* – Da. *lænestol* –Fr. *fauteul*

Sw. *fönster* – Da. *vindue* – Ge. *Fenster*



university of  
groningen

# Annotation

26-11-2007 | 11

Cognates:

Sw. *ägg* – Da. *æg* ‘egg’

Noncognates:

Da. *følelse* – Sw. *känsla* ‘feeling’



university of  
groningen