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
Title:

*Linguistic determinants of mutual intelligibility in Scandinavia*

http://www.let.rug.nl/micrela/
Inter-Scandinavian intelligibility
Research questions

› How well can Scandinavians understand each other’s vocabulary?

› What is the role of different linguistic factors for word intelligibility?
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
1. Frequency lists based on Danish, Norwegian and Swedish corpora

The three corpora must be constructed in the same way:

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

university of groningen
## Spontaneous speech

<table>
<thead>
<tr>
<th></th>
<th>Danish BySoc</th>
<th>Norwegian NoTa</th>
<th>Swedish GSLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>words</td>
<td>1.3 mio</td>
<td>900,000</td>
<td>1.3 mio</td>
</tr>
<tr>
<td>informants</td>
<td>80</td>
<td>165</td>
<td>371</td>
</tr>
<tr>
<td>activity</td>
<td>interviews and conversations</td>
<td>interviews and conversations</td>
<td>20 different activities</td>
</tr>
<tr>
<td>speaker variation</td>
<td>controlled</td>
<td>controlled</td>
<td>less controlled</td>
</tr>
</tbody>
</table>
Annotation

- 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?
**lemma ’room’**

<table>
<thead>
<tr>
<th>word</th>
<th>translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Da. værelse</td>
<td>Sw. rum</td>
</tr>
<tr>
<td>Sw. rum</td>
<td>Da. rum</td>
</tr>
</tbody>
</table>
Words can be understood via related words in other languages:

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

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

Sw. *fönster* – Da. *vindue* – Ge. *Fenster*
Cognates:
Sw. ägg – Da. æg ‘egg’

Noncognates:
Da. følelse – Sw. känsla ‘feeling’