Measuring mutual intelligibility in Scandinavia

Sebastian Küürschner, s.kurschner@rug.nl
Rijksuniversiteit Groningen
Albert-Ludwigs-Universität Freiburg im Breisgau

Outline

1. Background: Mutual intelligibility in Scandinavia
2. Research project „Linguistic determinants of mutual intelligibility in Scandinavia”
3. Measuring linguistic distances
Background: Mutual intelligibility in Scandinavia

- The Scandinavian languages:
  - Mainland Scandinavian languages: Danish, Norwegian (Bokmål and Nynorsk), Swedish
  - Island Scandinavian languages: Faroese, Icelandic
Mutual intelligibility

- For the Mainland Scandinavian languages, mutual intelligibility is in principle possible.
- Semi-communication (Haugen 1966):
  - communication in closely related languages
  - each involved person uses her/his mothertongue
Danish, Norwegian, and Swedish are historically closely related
- same root: North-Germanic languages
- intense language contact in the Middle Ages with Low German Hanse-tradesmen
  - high number of similar loanwords in all three languages
  - grammatical simplification
- common language policy
  - will to semi-communicate
Research on mutual intelligibility in Scandinavia

- **Competence**
  - How good are Scandinavians in understanding one another? (Delsing/Lundin-Åkesson 2005)

- **Discourse:**
  - How do participants of multilingual conversation act to make communication possible / more convenient?
  - Accommodation studies (Börestam Uhlmann 1994)
  - Gaining linguistic competence and negotiating discourse (Zeevaert 2004)

- **Attitudes:**
  - How do linguistic/political attitudes influence the success of the conversation?

- **Linguistic prerequisites of semi-communication mostly unexamined**
Faculty of Arts
University of Groningen

Project „Linguistic determinants of mutual intelligibility in Scandinavia“

• Ongoing research project (started in 2006) at the University of Groningen
• Project leader: Charlotte Gooskens
• Main research questions:
  – What linguistic distances can be found between the spoken Scandinavian languages at different linguistic levels (vocabulary, phonetics, morphology, syntax)?
  – To what extent are the linguistic distances on the different linguistic levels predictors of inter-Scandinavian intelligibility?

• Aim:
  – Measure of linguistic differences by use of dialectometric methods
  – Development of a model predicting and explaining mutual intelligibility in Scandinavia
Method

• finding objective criteria to measure linguistic distances
  – validating the results of objective methods with results from perception experiments

• isolating other factors which influence the possibilities of semi-communication
  – e.g., experience and attitudes are non-linguistic factors influencing the possibility of mutual intelligibility
How to measure linguistic distances?

- **linguistic distances can be measured on different linguistic levels**
  - **lexicon:**
    - how many words are cognate?
  - **phonetics/phonology:**
    - how much phonological distance is there between cognates?
  - **morphology:**
    - what is expressed where and how?
  - **syntax:**
    - to what extent are the syntactic systems similar?
Measuring methods

• Dialectometry
  – Methods for the measuring of distances between dialects
  – Levenshtein distance
  – Conditional entropy
Levenshtein distance

- computational method for comparison of related language varieties
- mostly used for measuring phonetic differences (Heeringa 2004, Nerbonne/Heeringa 2000)
- string mapping: comparing two strings
  - the costs of operations necessary for mapping are calculated
  - operations are insertions, deletions, and substitutions
  - word length is accounted for
Levenshtein distance: example

Danish *hjemme* – Swedish *hemma*, 'at home'

<table>
<thead>
<tr>
<th>j</th>
<th>e</th>
<th>m</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>e</td>
<td>m:</td>
<td>a</td>
</tr>
</tbody>
</table>

\[
\begin{array}{cccc}
1 & 0 & 0.5 & 1 \\
\end{array}
\]

\[= \frac{2.5}{4} = 62.5\%\]

Danish *guld* – Swedish *guld*, 'gold'

<table>
<thead>
<tr>
<th>g</th>
<th>u</th>
<th>l</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>u</td>
<td>l</td>
</tr>
</tbody>
</table>

\[
\begin{array}{cccc}
0 & 1 & 0 & 1 \\
\end{array}
\]

\[= \frac{2}{4} = 50\%\]
Levenshtein distance: Hypothesis and problems

- Hypothesis: The phonetic distance of two languages cannot exceed a certain degree for mutual intelligibility to be possible.
- The distance between the Scandinavian languages is not symmetrical:
  - Perception tests show that Danish is more difficult to understand for Swedes than Swedish is for Danes.
- Levenshtein distance cannot capture this kind of asymmetry.
Conditional entropy

• measure of linguistic distance which is sensible to asymmetry
• basis: probability of a sound to match the same sound in a corresponding word from another language
• frequency is accounted

Da.  kommer /kəmər/  savner /sɑːvər/
Sw.  kommer /kɔmːər/  saknar /sɑːknaːr/

,comes‘  ,misses‘

Da.  ø
     ε
     a

Sw.
Moberg/Gooskens (in press) examined phonetic entropy in Swedish-Danish contrast in formal and informal speech.

For most parts, the entropy of Danish given Swedish is higher than the other way round.

I.e., there is evidence for asymmetrical distances between Danish and Swedish.

These results match with results from perception experiments.

Linguistic factors can be taken as part of the explanation for (asymmetric) mutual intelligibility.

Question arising:
- How far is written language involved?
Summary

• the possibility of mutual intelligibility is partly determined by linguistic factors, which for this reason should be considered in research
• the Scandinavian languages prove a good test situation to measure these linguistic factors, for symmetrical as well as asymmetrical measures
• methods from dialectometry can lead to objective linguistic results, which can be compared to the results of actual intelligibility tests for validation
• methodically, non-linguistic factors need to be isolated as far as possible, but in an overall view, the integration of both linguistic and non-linguistic factors can help to achieve realistic models of mutual intelligibility
References


