Determinants of mutual intelligibility between closely related languages in Scandinavia

Charlotte Gooskens

Learner Language, Learner Corpora
University of Oulu, 5-6 October 2012
Overview

› Background
› Measuring intelligibility
› Determinants of intelligibility
› Present project
Background

Linguistic determinants of mutual intelligibility in Scandinavia

› 1 January 2006 – 1 January 2011
› Financed by NWO (The Netherlands Organisation for Scientific Research)
› Project members: Sebastian Kürschner, Nanna H. Hilton, Anja Schüppert, Renée van Bezooijen, student assistants
› [http://www.let.rug.nl/~gooskens/project/](http://www.let.rug.nl/~gooskens/project/)
Background

Semicommunication

› Haugen (1966)

› Speakers of different but related languages communicate with each other each speaking their own language and still comprehend each others

› = receptive multilingualism
Inter-Scandinavian intelligibility

Diagram showing the inter-Scandinavian intelligibility between Norwegian, Swedish, and Danish listeners.
Inter-Scandinavian intelligibility

Norwegian listener

Swedish

Danish

Norwegian

Swedish

Danish listener

Norwegian

Swedish listener

Danish
Inter-Scandinavian intelligibility

Norwegian listener

Swedish

Danish

Norwegian

Swedish

Danish listener

Norwegian

Swedish listener

Danish
Inter-Scandinavian intelligibility

- Mutual intelligibility is often imperfect
- Alternative for a lingua franca or active command of L2
- Scientific knowledge is lacking about determinants of mutual intelligibility
Central question

› To what extent are (extra-)linguistic factors determinants of intelligibility?
Measuring intelligibility
Measuring intelligibility

› **Opinion testing**
  How well do people *think* they understand the other language variety?

› **Observations**
  How well do people understand each other in *real* language situations?

› **Functional testing**
  How well do people *actually* understand the other language variety?
Measuring intelligibility

Methods for functional testing

› Open questions
› Multiple choice
› Translations
› ...

 rijksuniversiteit
groningen
Types of speech

› Spontaneous speech
› Read texts
› Single words
› ...

Measuring intelligibility
Determinants of intelligibility
Determinants of intelligibility

› Non-linguistic
  • Attitude
  • Contact

› Linguistic
  • Lexicon
  • Sounds
  • Orthography
  • Prosody
  • Syntax
Determinants of intelligibility

› Non-linguistic
  • Attitude
  • Contact

› Linguistic
  • Lexicon
  • Sounds
  • Orthography
  • Prosody
  • Syntax
Determinant of intelligibility: lexicon
What is the relationship between lexical distances and intelligibility?
Lexicon

Project *Internordisk sprogforståelse*, INS (*Inter-Nordic comprehension*), (Delsing & Lundin Åkesson 2005)

› 690 secondary school pupils

› Intelligibility:  % of correct answers to 5 open questions about a news item
Measurement of lexical distance: percentage of non-cognates

Example of a non-cognate:
› Danish *dreng* vs. Swedish *pojke* ‘boy’
Correlation between intelligibility scores and lexical distances is not significant ($r = -.42, p = 0.11$)

Lexical distances play a minor role in the predictability of intelligibility in the Scandinavian language area because the differences are small (values between 0 and 3.6%)
Determinant of intelligibility: sounds
What is the relationship between phonetic distances and intelligibility?
Measurement of phonetic distances: 
Levenshtein algorithm

› Heeringa (2004)
› Measures the phonetic distance between related language varieties
› Compares the sounds of cognate word pairs
› Counts how many sounds minimally must be substituted, added or removed in order to change the sounds of one word into the sounds of another word
› Total distance is obtained by summing word distances
Example Levenshtein distance

Danish *ligne* vs. Swedish *likna* ‘be like’

\[
\begin{array}{cccc}
| & l & i: & n & \varepsilon \\ \\
| & l & i: & k & n & \alpha \\
\end{array}
\]

\[
1 & 1
\]

\[
(1+1=2)/5 = 40\% \text{ distance}
\]
Correlaties verstaanbaarheid met fonetische afstand

Correlation between intelligibility scores and phonetic distances

\[ r = .81^{**}, \ p = .00 \]
› At the text level phonetic distances explain intelligibility to a high extent...

› ...but to be able to draw conclusions about the role of specific linguistic characteristics we need to look at the word level
Intelligibility of isolated words

Kürschner, Gooskens & van Bezooijen (2008)

- Internet experiment testing Swedish-Danish mutual intelligibility at the word level

- [http://www.let.rug.nl/lrs](http://www.let.rug.nl/lrs)
  
  login: germanic
  password: guest
Intelligibility of isolated words

Test words

› 384 frequent Swedish and Danish nouns
› Recordings of Standard Swedish and Danish
Intelligibility of isolated words

Listeners

› Danish and Swedish high school pupils
› 15-19 years
› Native speakers
Intelligibility of isolated words

Procedure

› Subjects listened to words from the neighboring language
› Translations into mother tongue
› Intelligibility = percentage of correctly translated words
Intelligibility of isolated words

Percentage of correctly translated words

<table>
<thead>
<tr>
<th></th>
<th>% correct translations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedes</td>
<td>50%</td>
</tr>
<tr>
<td>Danish</td>
<td>60%</td>
</tr>
<tr>
<td>Danes</td>
<td>70%</td>
</tr>
<tr>
<td>Swedish</td>
<td>80%</td>
</tr>
</tbody>
</table>

listeners → test language

Swedes Danish

Danes Swedish
Intelligibility of isolated words

› Correlation between phonetic distance and word intelligibility low ($r = -0.27$, $p < 0.001$)

› Many idiosyncrasies in single words
Intelligibility of isolated words

Analysis of errors can give information about listener strategies

› Phonetic confusions of sounds
  e.g. Sw. /k/ is often perceived as /g/ by Danes
  Sw. *klass*, Da. *klasse* ‘class’ is translated into Da. *glas* ‘glass’

› Influence of neighbor words
  e.g. Sw. *kør*, Da. *kor* ‘choir’ is often translated into Da. *kør* ‘drive’

› Interference from foreign languages
  e.g. Sw. *hot*, Da. *trussel* ‘threat’ is often translated into Da. *varm* ‘hot’
Intelligibility of isolated words

› Unknown sounds
  e.g. Sw. retroflexes [dʈʂn] and postalveolar-velar fricative [ʰ], Da. *stød* and [ˈð̝̃]

› Word length
  longer words are better recognized than shorter words because they have fewer neighbors

› Differences in number of syllables
  Da. *mængde* vs. Sw. *mängd* ‘quantity’

› Word frequency
  frequent words are more likely to come to the subjects’ minds immediately than infrequent words

› Orthography
Determinant of intelligibility: orthography
Danish pronunciation has changed more than Swedish pronunciation

Both orthographies are conservative

Therefore Danes can more often use their orthography when listening to Swedes than vice versa

Example

<table>
<thead>
<tr>
<th>Danish</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>hund [hun]</td>
<td>hund [hund] ‘dog’</td>
</tr>
</tbody>
</table>
Orthography

- **Consistent** sound-grapheme correspondence:
  L2 sound = L1 orthography

- **Inconsistent** sound-grapheme correspondence:
  L2 sound ≠ L1 orthography

<table>
<thead>
<tr>
<th>Danish</th>
<th>gift</th>
<th>/gift/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td>gift</td>
<td>/jift/</td>
</tr>
</tbody>
</table>
Orthography

![Bar chart showing the mean number of consistent sounds for Danish words (Swedish listeners) and Swedish words (Danish listeners). The Swedish words have a significantly higher mean number of consistent sounds, indicated by the ** symbol.]
Orthography

Experiment Schüppert & Gooskens (2012)

› Does orthography play a role in the intelligibility of spoken language?
Stimuli

› 112 Danish-Swedish cognates
› Pronunciation differs by one sound
› This sound is consistent or inconsistent with L1 orthography
Orthography

Subjects
› 29 Danish students
› 25 Swedish students

Task
› Translation into L1
Percentages of correctly translated words

Orthography
Orthography plays a role in the intelligibility of spoken language
Determinant of intelligibility: prosody
Hilton, Schüppert & Gooskens (2011)

› Could speech rate determine intelligibility results?

› Hypothesis: Danes speak faster than Norwegians and Swedes, and are therefore more difficult to understand
Recordings

News items from national broadcasting companies:

› Norwegian: 18 speakers (9 men, 9 women)
› Swedish: 18 speakers (9 men, 9 women)
› Danish: 19 speakers (9 men, 10 women)

› ±26 minutes (55 speakers), mean 28 seconds per speaker
Measurements of speech rate

Two kinds of measurements:

1. Number of *phonetic* syllables per second
   - Example: Danish *skatteydere der ikke*: how many syllables?
   - Are counted automatically my means of a PRAAT script
Measurements of speech rate

Two kinds of measurements:

1. Number of phonetic syllables per second
   - Example: Danish skatteydere der ikke: how many syllables?
   - Are counted automatically by means of a PRAAT script

2. Number of canonical syllables per sekund
   - Example: Danish skatteydere der ikke has 8 syllables: skat-te-y-de-re-der-ik-ke
   - Manually counted
1. Danes, Norwegians and Swedes produce the same number of phonetic syllables per second
2. Danes produce more *canonical* syllables per second than Swedes and Norwegians
Speech rate

Danes reduce more syllables than Swedes and Norwegians
Speech rate

Conclusion

› Danes produce more canonical syllables per second than Norwegians and Swedes

› Measurements of the number of phonetic syllables per second show that the differences are due to a larger degree of syllable reduction in Danish
Implications for intelligibility

› A large degree of reductions means:
  • More phonological syllables disappear when Danes speak than when Norwegians and Swedes speak
  • Danish articulation is less precise

› In a follow-up intelligibility experiment we found that this is indeed problematic for Swedes and Norwegians
Determinant of intelligibility: syntax
Hilton, Gooskens & Schüppert (accepted)

› Do syntactic differences play a role in the intelligibility of a closely related language?

› What is the relative influence of syntactic versus phonetic differences on intelligibility?
Research design

› Investigation tests the effect of idiosyncratic Norwegian syntactic constructions on Danes’ comprehension of Norwegian.

› e.g. difference in particle placement

<table>
<thead>
<tr>
<th>Norwegian:</th>
<th>Han</th>
<th>tok</th>
<th>av</th>
<th>brillene</th>
</tr>
</thead>
<tbody>
<tr>
<td>subj</td>
<td>verb</td>
<td>part</td>
<td></td>
<td>obj</td>
</tr>
<tr>
<td>Danish:</td>
<td>Han</td>
<td>tog</td>
<td>brillerne</td>
<td>af</td>
</tr>
<tr>
<td>subj</td>
<td>verb</td>
<td>obj</td>
<td></td>
<td>part</td>
</tr>
</tbody>
</table>
Research design

- Danish listeners are asked to decide whether the content of a sentence is plausible or not (binary choice)

  e.g. implausible: *elefanten slog ordet op*  
  ‘the elephant looked up the word’

  e.g. plausible:  *journalisten skrev en artikel ud*  
  ‘the journalist printed the article’

- Response time and number of correct answers are measured
Research design
› Crossed design
› 5 types of syntactic constructions
› 8 sentences per construction
› 2 plausability conditions
› 4 different linguistic conditions:

<table>
<thead>
<tr>
<th></th>
<th>Danish syntax</th>
<th>Norwegian syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danish sounds</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Norwegian sounds</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Results

› Syntactic differences between L1 and L2 impede intelligibility...

› ... but sound differences play a larger role for intelligibility than syntactic differences
Present project
Present project

**Mutual intelligibility of closely related languages in Europe: linguistic and non-linguistic determinants**

- 1 September 2011 – 1 September 2015
- Financed by NWO (The Netherlands Organisation for Scientific Research)
- Project members: Vincent van Heuven, Anja Schüppert, Wilbert Heeringa, Renée van Bezooijen, Jelena Golubovic, Femke Swarte, Stefanie Voigt
- [http://www.let.rug.nl/~gooskens/project/](http://www.let.rug.nl/~gooskens/project/)
Research questions

1. What is the level of mutual intelligibility of closely related languages in Europe?
2. What factors play a role in mutual intelligibility?
3. How well do speakers of closely related European languages understand each other in non-native English compared to semi-communication?
## Present project

### Languages to be tested

<table>
<thead>
<tr>
<th>Germanic</th>
<th>Romance</th>
<th>Slavic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danish</td>
<td>French</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>Dutch</td>
<td>Italian</td>
<td>Croatian</td>
</tr>
<tr>
<td>English</td>
<td>Portuguese</td>
<td>Czech</td>
</tr>
<tr>
<td>German</td>
<td>Romanian</td>
<td>Polish</td>
</tr>
<tr>
<td>Swedish</td>
<td>Spanish</td>
<td>Slovak</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Slovene</td>
</tr>
</tbody>
</table>