Mutual intelligibility between closely related languages

Charlotte Gooskens
Renée van Bezooijen

Charles University in Prague, 7 October 2009
Overview

› background
› intelligibility testing
› factors determining intelligibility
› four investigations
› future research
Background
Background

› *Linguistic determinants of mutual intelligibility in Scandinavia*

› Financed by NWO (The Netherlands Organisation for Scientific Research)

› 1 January 2006 – 1 January 2011

› Members of the project group:
  - Nanna Haug Hilton
  - Anja Schüppert
  - Renée van Bezooijen
  - Charlotte Gooskens
  - student assistants

› [http://www.let.rug.nl/~gooskens/project/](http://www.let.rug.nl/~gooskens/project/)
Semicommunication

› Haugen (1966)

› \(\approx\) nonconvergent/asymmetric/bilingual discourse, receptive bilingualism

› Speakers of different but related languages each speak their own language and still comprehend each others’ languages

› Mutual intelligibility is sometimes imperfect and asymmetric
Background

Observed semicommunication

› Danish - Norwegian - Swedish (Haugen 1966, Maurud 1976....)
› Czech - Slovakian (Budovičá 1987)
› Czech - Polish (Hansen 1987)
› Spanish - Portuguese (Coseriu 1988, Jensen 1989, Zeevaert 2002)
› Italian - Spanish (Hansen 1987)
› German - Dutch (Haz 2002)
› Frisian - Dutch (Feitsma 1986)
› Croatian - Serbian (Haugen 1990)
› Hindi - Urdu (Haugen 1990)
› Icelandic - Faeroese (Braunmuller & Zeevaert 2001)
› Macedonian - Bulgarian (Haugen 1990)
› Russian - Bulgarian (Braunmuller & Zeevaert 2001)
Central questions

› How can the mutual intelligibility between closely related languages be measured?

› How can the relevant (extra-)linguistic factors be measured?

› To what extent are the (extra-)linguistic factors predictors of intelligibility?
Intelligibility
Measuring intelligibility

› **Opinion testing:**
  How well does the listener **think** he understands the other language variety?

› **Functional testing:**
  How well does the listener **actually** understand the other language variety?

› **Observations:**
  How well do people understand each other in **real** language situations?
# Measuring intelligibility

<table>
<thead>
<tr>
<th>Type of research</th>
<th>Type of speech</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>opinion test</td>
<td>• spontaneous speech</td>
<td>• open questions</td>
</tr>
<tr>
<td></td>
<td>• read texts</td>
<td>• multiple choice</td>
</tr>
<tr>
<td></td>
<td>• isolated words</td>
<td>• translations</td>
</tr>
<tr>
<td></td>
<td>• ...</td>
<td>• reaction times</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• ...</td>
</tr>
<tr>
<td>functional test</td>
<td>• spontaneous speech</td>
<td>• opinion scales</td>
</tr>
<tr>
<td></td>
<td>• read texts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• isolated words</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ...</td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>• spontaneous speech</td>
<td>• counting phenomena</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Factors explaining intelligibility
Factors explaining intelligibility

› Non-linguistic

› Linguistic (distances)
Factors explaining intelligibility

Non-linguistic

› attitude
› contact
› orthography
Explaining factor: attitude

What do you think of the Danish language?
beautiful □ □ □ □ □ □ ugly

What do you think of the Danes?
kind □ □ □ □ □ □ unkind

Would you like to live in Denmark?
yes □ maybe □ no □
Explaining factor: attitude

› It is plausible that in real life attitudes play a role in intelligibility
› Still statistic relationships have hardly been found
› The test situation is likely to block attitudes
Explaining factor: contact

I watch Danish television...

- once a week
- once a month
- once a year
- less often

I meet Danes...

I am in Denmark...
Explaining factor: contact

- It is certain that contact plays a role in intelligibility
- Still statistic relationships have hardly been found
- Scandinavians have little contact
Explaining factor: orthography

Danish

Swedish

\textit{hund} [hun] \hspace{1cm} \textit{hund} [hund] \text{ ‘dog’}
Factors explaining intelligibility

Linguistic distances

› pronunciation
› lexicon
› morphology
› syntax
Explaining factor: phonetic distance

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
Explaining factor: phonetic distance

Example Levenshtein distance:

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

```
  l    i:    n    η
  l    i:    k    n    a
  1 1
```

\[(1+1=2)/5 = 40\%\] difference
Explaining factor: lexical distance

Percentage of non-cognates

Example of a non-cognate:
> Danish *pige* vs. Swedish *flicka* ‘girl’
Investigations
Investigations

› Text comprehension with open questions
› Intelligibility of isolated words
› Intelligibility among children
› The role of syntax in intelligibility
Text comprehension with open questions
Text comprehension with open questions

- The project *Internordisk sprogforståelse*, INS (*Inter-Nordic comprehension*)
- Supported by the Nordic Cultural Fund
- Test groups from all Nordic countries
- Intelligibility of the three mainland Scandinavian languages (Danish, Norwegian and Swedish) is tested
- Questionnaire about attitudes and contact
Our contribution to INS:

› To investigate to what extent differences in inter-Scandinavian intelligibility can be explained by linguistic distances

› To compare results to intelligibility between speakers of Dutch, Frisian and Afrikaans
Text comprehension with open questions

Scandinavian subjects

› 690 secondary school pupils
› between 16 and 19 years
### Design of Scandinavian study

<table>
<thead>
<tr>
<th>Subjects from</th>
<th>Test language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Danish</td>
</tr>
<tr>
<td>Denmark:</td>
<td></td>
</tr>
<tr>
<td>Århus</td>
<td>-</td>
</tr>
<tr>
<td>Copenhagen</td>
<td>-</td>
</tr>
<tr>
<td>Norway:</td>
<td></td>
</tr>
<tr>
<td>Bergen</td>
<td>47</td>
</tr>
<tr>
<td>Oslo</td>
<td>57</td>
</tr>
<tr>
<td>Sweden:</td>
<td></td>
</tr>
<tr>
<td>Malmö</td>
<td>44</td>
</tr>
<tr>
<td>Stockholm</td>
<td>41</td>
</tr>
<tr>
<td>Finland:</td>
<td></td>
</tr>
<tr>
<td>Mariehamn</td>
<td>22</td>
</tr>
<tr>
<td>Vasa</td>
<td>-</td>
</tr>
<tr>
<td>Helsinki</td>
<td>9</td>
</tr>
</tbody>
</table>
West-Germanic subjects

› 81 secondary school pupils
› between 16 and 17 years
**Design of West-Germanic study**

<table>
<thead>
<tr>
<th>Subjects:</th>
<th>Test language</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dutch</td>
<td>Frisian</td>
<td>Afrikaans</td>
</tr>
<tr>
<td>Dutch</td>
<td>-</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Frisian</td>
<td>-</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>Afrikaans</td>
<td>15</td>
<td>17</td>
<td>-</td>
</tr>
</tbody>
</table>
Method

› News item (250 words)
› Read aloud in test language
› 5 open questions
› Intelligibility was expressed as percentage of correctly answered questions per test group
Text comprehension with open questions

Linguistic distances

› Extra recordings from each of the nine towns in Scandinavia

› All Scandinavian and West-Germanic recordings were transcribed phonetically

› Phonetic and lexical distances were calculated
Text comprehension with open questions

Correlation between intelligibility and lexical distances

$r = -.36$
Text comprehension with open questions

Correlation between intelligibility and phonetic distances

\[ r = -0.64^* \]
Text comprehension with open questions

Large advantage:

› Realistic (ecologically valid)

Disadvantages:

› Difficult to assess representativity of stimulus text
› Single unintelligible words may have a large effect on intelligibility
› Difficult to construct open questions
› Difficult to decide when a question is correctly answered
› Many varieties needed in order to be able to correlate intelligibility with linguistic distances
› Provides only an overall impression of the role of phonetic and lexical distances for the intelligibility
Intelligibility of isolated words
Intelligibility of isolated words

- Internet experiment involving seven languages in the Germanic language area
- Word comprehension
Intelligibility of isolated words

Research questions

› How well do speakers of various Germanic languages understand each other’s vocabularies?

› Which linguistic factors play a role in the intelligibility?
Test words

› 384 nouns
› Randomly selected from a list of 2575 highly frequent spoken words
› Translated into seven languages
› Recordings of standard language speakers
Intelligibility of isolated words

Test words are non-cognates or cognates

Example non-cognate:

Du. *lichaam*, De. *krop*, ‘body’
Fr. *fyts*, Zw. *cykel*, De. *Fahrrad* ‘bicycle’

Examples cognates:

Fr. *strategy*, Du. *strategie*, De. *strategi* ‘strategy’
Intelligibility of isolated words

Recordings

- Swedish (Sw)
- Norwegian (No)
- Danish (Da)
- German (Ge)
- Low German (LG)
- Frisian (Fr)
- Dutch (Du)
# Intelligibility of isolated words

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Test language</th>
<th>Da</th>
<th>Sw</th>
<th>No</th>
<th>LG</th>
<th>Ge</th>
<th>Fr</th>
<th>Du</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sw</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fr</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Du</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
### Intelligibility of isolated words

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Da</th>
<th>Sw</th>
<th>No</th>
<th>LG</th>
<th>Ge</th>
<th>Fr</th>
<th>Du</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sw</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Fr</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Du</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Intelligibility of isolated words

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Test language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Da</td>
</tr>
<tr>
<td>Sw</td>
<td>x</td>
</tr>
<tr>
<td>No</td>
<td>x</td>
</tr>
<tr>
<td>Da</td>
<td></td>
</tr>
<tr>
<td>Fr</td>
<td>x</td>
</tr>
<tr>
<td>Du</td>
<td>x</td>
</tr>
</tbody>
</table>
## Intelligibility of isolated words

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Da</th>
<th>Sw</th>
<th>No</th>
<th>LG</th>
<th>Ge</th>
<th>Fr</th>
<th>Du</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sw</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Du</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>Translation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td><em>Ik weet dat hij naar huis komt</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frisian</td>
<td><em>Ik wit dat er thús komt</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td><em>Ich weiß dass er nach Hause kommt</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish</td>
<td><em>Jeg ved at han kommer hjem</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td><em>Jag vet att han kommer hem</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norwegian</td>
<td><em>Jeg veit at han kommer hjem</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

‘I know that he comes home’
Intelligibility of isolated words

Listeners

- 1400 high school pupils
- 15-19 years
- Speaking test language at home
Intelligibility of isolated words

Procedure

- Internet-based
- Subjects listened to 96 words via head phones
- Translations into mother tongue

- [http://www.let.rug.nl/lrs](http://www.let.rug.nl/lrs)
  login: germanic
  password: guest
The result of the test person is shown:
The results of all test persons are shown:
Intelligibility of isolated words

Calculations

Intelligibility = percentage of correctly translated words

› Ignoring spelling mistakes
  ex. Dutch *kultuur* instead of *cultuur* for Danish *kultur* ‘culture’

› Allowing alternative translations
  ex. Dutch *winkel* or *boetiek* for Low German *Laden* ‘shop’
Intelligibility of isolated words

% cognates per language pair

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>% Cognates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Da No</td>
<td>100</td>
</tr>
<tr>
<td>No Sw</td>
<td>90</td>
</tr>
<tr>
<td>Da Sw</td>
<td>80</td>
</tr>
<tr>
<td>Du Ge</td>
<td>70</td>
</tr>
<tr>
<td>Du LG</td>
<td>60</td>
</tr>
<tr>
<td>Da Ge</td>
<td>50</td>
</tr>
<tr>
<td>Da Du</td>
<td>40</td>
</tr>
<tr>
<td>Da LG</td>
<td>30</td>
</tr>
<tr>
<td>Da Fr</td>
<td>20</td>
</tr>
</tbody>
</table>
Intelligibility of isolated words

% correctly translated cognates:
Intelligibility of isolated words

% correctly translated inherited words, from high to low

<table>
<thead>
<tr>
<th>Test Language</th>
<th>Correct Translations</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>80</td>
</tr>
<tr>
<td>Da</td>
<td>70</td>
</tr>
<tr>
<td>Sw</td>
<td>60</td>
</tr>
<tr>
<td>No</td>
<td>50</td>
</tr>
<tr>
<td>Da</td>
<td>40</td>
</tr>
<tr>
<td>Ge</td>
<td>30</td>
</tr>
<tr>
<td>Da</td>
<td>20</td>
</tr>
<tr>
<td>Sw</td>
<td>10</td>
</tr>
<tr>
<td>Da</td>
<td>0</td>
</tr>
<tr>
<td>Fr</td>
<td>0</td>
</tr>
<tr>
<td>Da</td>
<td>0</td>
</tr>
<tr>
<td>Du</td>
<td>0</td>
</tr>
</tbody>
</table>

rijksuniversiteit Groningen
Intelligibility of isolated words

% correctly translated loan words, from high to low

Loan words

% correct translations

listeners →
test language →

No Sw No No Du Du Da Da Da Sw Da Da Da Fr Da Fr

rijksuniversiteit
groningen
Intelligibility of isolated words

From small to large difference between inherited and loan words

<table>
<thead>
<tr>
<th>Listeners</th>
<th>Test Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Da</td>
</tr>
<tr>
<td>Sw</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Sw</td>
</tr>
<tr>
<td>Da</td>
<td>Du</td>
</tr>
<tr>
<td>No</td>
<td>Du</td>
</tr>
<tr>
<td>Da</td>
<td>Ge</td>
</tr>
<tr>
<td>Sw</td>
<td>Da</td>
</tr>
<tr>
<td>Da</td>
<td>Fr</td>
</tr>
<tr>
<td>Ge</td>
<td>Da</td>
</tr>
<tr>
<td>Du</td>
<td>Sw</td>
</tr>
<tr>
<td>Da</td>
<td>Da</td>
</tr>
<tr>
<td>Sw</td>
<td>Da</td>
</tr>
<tr>
<td>Da</td>
<td>Du</td>
</tr>
<tr>
<td>Sw</td>
<td>Da</td>
</tr>
<tr>
<td>Da</td>
<td>Du</td>
</tr>
<tr>
<td>Da</td>
<td>Du</td>
</tr>
<tr>
<td>Sw</td>
<td>Du</td>
</tr>
<tr>
<td>Da</td>
<td>Da</td>
</tr>
<tr>
<td>Sw</td>
<td>Du</td>
</tr>
<tr>
<td>Da</td>
<td>Da</td>
</tr>
<tr>
<td>Sw</td>
<td>Du</td>
</tr>
<tr>
<td>Da</td>
<td>Da</td>
</tr>
<tr>
<td>Sw</td>
<td>Du</td>
</tr>
<tr>
<td>Da</td>
<td>Da</td>
</tr>
</tbody>
</table>

% correct translations

- Inherited words
- Loan words
Intelligibility of isolated words

11 factors considered for prediction of intelligibility of Swedish words by Danish listeners

- Levenshtein distance
- Foreign sounds
- Word length
- Word stress differences
- Differences in number of syllables
- Lexical tones
- Stød
- Neighbourhood density
- Etymology (native words versus loan words)
- Orthography
- Word frequency
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 neighbour words
  e.g. Sw. *körr*, 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 among children
Intelligibility among children

Ph.D.-project Anja Schüppert:

Nonlinguistic factors can be neutralised by testing young children:

• Cannot read
• Have no knowledge of foreign languages
• Have had little contact with neighbouring countries
• Have less strong attitudes towards neighbouring languages
Intelligibility among children

Subjects

• 19 Danish children from Odense, 4-6 years old
• 26 Swedish children from Vaxjö, 4-6 years old
• 20 Danish adults from Odense
• 19 Swedish adults from Vaxjö
Intelligibility among children

Stimulus material

- 50 nouns (cognates) that are frequent in child language and early acquired
- Read aloud in Danish and Swedish
Intelligibility among children

Test

• Danes listen to Swedish words and Swedes listen to Danish words
• Per word four pictures are presented
• Subject points to the picture corresponding to the test word on a touch screen
• Response time is measured
Example
Swedish subject hears Danish æble ‘apple’ and sees the following pictures

The subject chooses rightmost picture
Intelligibility among children

Intelligibility results

![Bar chart showing mean response time for children and adults speaking Danish and Swedish.](chart.png)
Intelligibility among children

Attitude elicitation

› Children and adults were asked if they thought the language sounded...
› ...less nice than their native language (-1)
› ...as nice as their native language (0)
› ...nicer than their native language (1)
Intelligibility among children

Attitude results

-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

mean attitude

Reaction Regarding:

- children
- adults

Danish
Swedish
Intelligibility among children

Plans

Measure intelligibility and attitudes among children of various ages
The role of syntax in intelligibility
The role of syntax in intelligibility

Research questions

› 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?
The role of syntax in intelligibility

Research design

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

e.g. difference in particle placement

Norwegian: Han tok av brillene
            subj verb part obj

Danish: Han tog af brillerne
            subj verb obj part

‘he took off the glasses’
The role of syntax in intelligibility

Research design

› 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 is measured in addition to number of correct answers
One informant hears:

<table>
<thead>
<tr>
<th></th>
<th>Without Noise</th>
<th></th>
<th>With Noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Norwegian</td>
<td>Danish</td>
<td>Norwegian</td>
<td>Danish</td>
</tr>
<tr>
<td></td>
<td>syntax</td>
<td>syntax</td>
<td>syntax</td>
<td>syntax</td>
</tr>
<tr>
<td>Danish phonology</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Norwegian phonology</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
The role of syntax in intelligibility

Research design

›6 types of syntactic constructions
›8 sentences per construction
›2 plausability conditions
›4 different linguistic conditions
›2 noise conditions

768 total sentence count
The role of syntax in intelligibility

Plans

› Experiment still in design stage, but will be completed and conducted during the coming year with informants in Denmark

› If results show that syntactic differences do indeed impede intelligibility the project can be extended to listener groups in Norway and Sweden.
Future research
Future research

› Develop a general model of intelligibility among closely related languages