Linguistic determinants of the intelligibility of Swedish words among Danes

- Sebastian Kürschner
- Charlotte Gooskens
- Renée van Bezooijen

Idea

- Linguistic factors have been identified to explain intelligibility between closely related languages/varieties in prior research
- Prior research based on aggregate differences of whole text understanding between varieties
- Assumption: To understand texts in a foreign variety, you need to be able to identify a certain amount of words
- Aim: Identify the linguistic factors which determine the intelligibility of single words, presented in isolation

Languages

- Danish listeners confronted with Swedish words
- Danish and Swedish are well-known to be mutually intelligible to some degree

Explaining factors

- We included only linguistic factors
  - Attitudes were not tested, but usually do not correlate strongly with intelligibility scores
  - We excluded the results of subjects who probably had much prior experience with Swedish
- Linguistic factors based on
  - general results on word intelligibility from psycho-phonetic literature
  - language-specific factors identified in literature on Swedish and Danish

Intelligibility experiment

Test words:

- 347 cognate nouns
- randomly selected from a database with parallel lists of 2575 frequent spoken words
- concepts known to secondary school pupils between 15 and 19 years of age
- read aloud by a male speaker of Standard Swedish

Subjects:

- 36 native speakers of Danish
- pre-university education
- aged 16 to 19

Task:

- translation of Swedish test words into Danish
- via internet
Intelligibility experiment

Design:

- 4 groups of subjects
- 5 to 19 subjects per group
- 96 test words per group

Factors considered for explanation

- Levenshtein distance
  - normalized
  - 0.5 for
    - segment length: [u] vs. [u]
    - [e] vs. [e]
    - [R] vs. [r]
- Foreign sounds

\[\text{fx sv. gjukdom [gjukdom:] da. sygdom [sygdom]}
\text{retrofleksjer fx sv. jord [jord] da. jord [jor'd]}\]

Factors: Word accent differences

- Word length
- Word accent differences
  - Danish ‘kontext’ vs. Swedish ‘kon’text’ ‘context’
- Difference in Syllable number
  - Sw. choklad vs. Danish chokolade ‘chocolate’

Factors: Stød

\[\begin{align*}
\text{fx + stød} & \quad \text{- stød} \\
\text{hvalen} & \quad \text{valen} \\
\text{køber} & \quad \text{køber} \\
\text{musen} & \quad \text{musen} \\
\text{skal} & \quad \text{skal} \\
\text{tør} & \quad \text{tør}
\end{align*}\]

Factors

- Neighborhood density
  - Sw. säng
    and staäng ‘close’
- Etymology
  - native vs. loan word
- Orthography
  - Danish hånd [hænd] vs. Swedish hand [hand]
- Danish word frequency
Results

› Intelligibility: 57.7 % correct identifications of the cognates

› Explaining factors:

Table 1: Correlation of the intelligibility scores with linguistic factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Correlation (r)</th>
<th>Significance (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levenshtein distance</td>
<td>-.33</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Foreign sounds</td>
<td>-.13</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Word-length of Swedish words</td>
<td>.25</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Word accent difference</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Difference in syllable number</td>
<td>-.25</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Lexical tones</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Stød</td>
<td>-.02</td>
<td></td>
</tr>
<tr>
<td>Neighbourhood density</td>
<td>-.17</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Etymology</td>
<td>.1</td>
<td></td>
</tr>
<tr>
<td>Orthography</td>
<td>.21</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Results: Multiple regression

Dependent: Intelligibility scores

<table>
<thead>
<tr>
<th>Method</th>
<th>Entered variables</th>
<th>Model</th>
<th>t-stat</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>All linguistic factors</td>
<td>All</td>
<td>-10.24</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>Levenshtein distance</td>
<td>Levenshtein distance</td>
<td>-9.72</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>Orthography</td>
<td>Orthography</td>
<td>8.07</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Step 4</td>
<td>Different syllable number</td>
<td>Different syllable number</td>
<td>7.19</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Step 5</td>
<td>Neighbourhood density</td>
<td>Neighbourhood density</td>
<td>5.97</td>
<td>&lt; .001</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

› Why can the variation not be explained to a higher degree?
  • unpredictable factors in single words (voice quality, speech rate, etc.) which are levelled in aggregate scores
  • idiosyncratic characteristics of single words

Some characteristics

› Confusion with other languages learned
  • Swedish art ‘sort’ - Danish kunst ‘art’
  • neighbourhood closeness
  • non-corresponding phonemes
    • Swedish frukt ‘fruit’ translated as Danish frugt ‘fear’ instead of correct frugt