Inter-coder reliability and the effects of coder background and linguistic structure.

Thøgersen, Jacob; Heegård, Jan

Publication date: 2013

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-ND

Citation for published version (APA):
1. Research question

- Phonetic coding and transcriptions are not objective.
- Acoustic structure plays a significant role for the reliability of a transcription: some structures are easier to differentiate than others.
- Coders’ disciplinary backgrounds affect what they hear.

- We propose:
  1. Also semantics, i.e. phonological contrast, must be considered.
  2. Listener (dis)agreement correlates with level of difficulty in perceptability.
  3. Reduction phenomena should be viewed as graded: listener disagreement reveals intermediate stages in word forms’ pronunciation.

2. Variables

3. Method

- 9 listeners (4 trained linguists, 4 students of other subjects, 1 experimenter with different background).
- 2303 tokens (1289 for /-də/, 872 for /-æðə/, 149 for /-ŋæ/).
- Listened to tokens in isolation (through Praat script).
- Task for /-də/ and /-æðə/: Which word is heard (semantic contrast)?
- Task for /-ŋæ/: Is the schwa assimilated?
- Perceived short forms = The number of listeners who heard the token in its short form.

4. Results, intercoder agreement

/-/də/ (+Semantic, +Easy): Highest agreement
/-/æðə/ (+Semantic, -Easy): Medium agreement
/-/ŋæ/ (-Semantic): Least agreement

5. Results, listener background

<table>
<thead>
<tr>
<th></th>
<th>Linguists</th>
<th>Non-linguists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intergroup agreement</td>
<td>0.940</td>
<td>0.932</td>
</tr>
<tr>
<td>Chi-squared</td>
<td>123, df = 1</td>
<td>p&lt;0.001</td>
</tr>
</tbody>
</table>

6. Results, coder confidence

- Coders could choose to listen again to tokens if they were in doubt.
- The graphs show the relationship between such stimulus repeats and listener (dis)agreement.

7. Results, tokens with high and low agreement

<table>
<thead>
<tr>
<th></th>
<th>High agreement, unreduced</th>
<th>High agreement, reduced</th>
<th>Low agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeats by perceived short forms /-də/</td>
<td>6.4%</td>
<td>6.2%</td>
<td>14.1%</td>
</tr>
<tr>
<td>Repeats by perceived short forms /-æðə/</td>
<td>10.6%</td>
<td>10.4%</td>
<td>22.1%</td>
</tr>
<tr>
<td>Repeats by perceived short forms /-ŋæ/</td>
<td>12.2%</td>
<td>12.0%</td>
<td>24.5%</td>
</tr>
</tbody>
</table>

8. Conclusions

- Coding reliability depends on semantic functional load of target. Phonetic contrasts which are distinctive are inherently easier to perceive.
- Some phonetic contrasts are inherently more difficult to perceive.
- Trained linguists have an advantage over lay listeners on some but not all variables.
- In our study, linguist showed higher agreement only on the more difficult variables.
- Intercoder agreement correlates highly with intercoder confidence as measured by repeats of stimulus.
- Intercoder (dis)agreement appears to correlate with segmental ambiguity.
- Intercoder disagreement is not noise in our investigation, but evidence against a polar interpretation of reduction phenomena.