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

Jacob Thøgersen, University of Copenhagen
Jan Heegård, Copenhagen Business School

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

![Polar interpretation of reduction phenomena. Intercoder (dis)agreement appears to correlate with segmental ambiguity. The graphs show the relationship between stimulus repeats and listener (dis)agreement.](image)

![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. Inter-coder agreement correlates highly with inter-coder confidence as measured by repeats of stimulus. Inter-coder (dis)agreement appears to correlate with segmental ambiguity.](image)

![Inter-coder reliability and the effects of coder background and linguistic structure.](image)

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

4. Results, intercoder agreement

<table>
<thead>
<tr>
<th>Semantic contrast</th>
<th>Long form</th>
<th>Short form</th>
</tr>
</thead>
<tbody>
<tr>
<td>/-də/</td>
<td>brugt ‘used’ (past., long., indef.)</td>
<td>brugt ‘used’ (past., long., indef.)</td>
</tr>
<tr>
<td>/-ə/</td>
<td>blanded ‘mixed’ (past., long., indef.)</td>
<td>blanded ‘mixed’ (past., long., indef.)</td>
</tr>
<tr>
<td>/-ŋ/</td>
<td>mange ‘many’ (past., indef.)</td>
<td>/-ŋ/ (past., indef.)</td>
</tr>
</tbody>
</table>

‘Easiness’

<table>
<thead>
<tr>
<th>Unreduced</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>/-də/</td>
<td>d</td>
</tr>
<tr>
<td>/-ə/</td>
<td>δ</td>
</tr>
<tr>
<td>/-ŋ/</td>
<td>η</td>
</tr>
</tbody>
</table>

5. Results, listener background

<table>
<thead>
<tr>
<th></th>
<th>Linguists</th>
<th>Non-linguists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-group agreement</td>
<td>0.940</td>
<td>0.952</td>
</tr>
<tr>
<td>‘Short form’ codes</td>
<td>65.4%</td>
<td>52.4%</td>
</tr>
<tr>
<td>Between group differences</td>
<td>Chi squared = 123, df = 1, p&lt;0.001</td>
<td>Chi squared = 16.7, df = 1, p=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 stimuli repeats and listener (dis)agreement.

7. Results, tokens with high and low agreement

![High agreement, unreduced.](image)

![Low agreement.](image)

![High agreement, reduced.](image)

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, linguists showed higher agreement only on the more difficult variables.
- Inter-coder agreement correlates highly with inter-coder confidence as measured by repeats of stimulus.
- Inter-coder (dis)agreement appears to correlate with segmental ambiguity.
- Inter-coder disagreement is not noise in our investigation, but evidence against a polar interpretation of reduction phenomena.