Are interpreters better respeakers?
An exploratory study on respeaking competences

Agnieszka Szarkowska & Łukasz Dutka
Agnieszka Chmiel, Agnieszka Lijewska, Krzysztof Krejtz, Krzysztof Marasek & Łukasz Brocki
Respeaking in Poland

- **Television**
  - No live subtitling on TV
  - Semi-live subtitling with QWERTY

- **Live events**
  - Conferences
  - Online interlingual respeaking for canonisation

- **SR software**
  - Newton Dictate (Newton Technologies)
  - Magic Scribe (Unikkon Integral)
Respeaking project (2014-2017)

- "Respeaking – process, competences, quality"
- Goal: to examine the competences of respeakers
- Research question: Are interpreters better respeakers?
- Three groups of participants
  - Interpreters and interpreting trainees
  - Translators and translation trainees
  - Control group (no interpreting/translation background)
Respeaking training

- No professional respeakers to test
- Respeaking training
  - Four two-day workshops (February - May 2015)
  - Pablo Romero Fresco, Carlo Eugeni, Juan Martinez
Tests in the project

- Working memory capacity tests
  - Digit Span
  - Listening Span
  - Reading Span
- Paraphrasing tests
- Proof-reading test
- Respeaking tests + semi-structured interview
Respeaking test

- **Intralingual respeaking** (Polish to Polish)
  - Four 5-minute clips in randomised order
  - Variables: slow/fast, one speaker/many speakers
  - Genres: speech, chat shows, news

- **Interlingual respeaking** (English to Polish)
  - Slow one-speaker speech
    (President Obama in Warsaw)
Participants

- 58 participants
  - 23 translators
  - 23 interpreters
  - 12 controls

Demographic data of the sample
- Mean age 27.48 (SD 5.71)
- 51 women, 7 men
Measurements

- Screenrecording
- Eyetracking
- EEG (Emotiv EPOC)
- Self-report on cognitive load
- NER
- WER
Preliminary results

- Participant sample: $N=42$
- Working memory capacity
- Self-reported cognitive load
- EEG cognitive load
### Working memory capacity

<table>
<thead>
<tr>
<th>Group</th>
<th>Listening Span</th>
<th>Reading Span</th>
<th>Digit Span (forward)</th>
<th>Digit Span (backward)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreters</td>
<td>56.55</td>
<td>36.77</td>
<td>6.22</td>
<td>5.50</td>
</tr>
<tr>
<td>Translators</td>
<td>54.35</td>
<td>36.35</td>
<td>6.17</td>
<td>5.58</td>
</tr>
<tr>
<td>Controls</td>
<td>49.57</td>
<td>35.71</td>
<td>6.28</td>
<td>5.42</td>
</tr>
</tbody>
</table>

N=42, ANOVA, p<.05
# Self-reported cognitive load

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Demand</td>
<td>How mentally demanding was the task?</td>
<td>Extremely undemanding to extremely demanding (7-point scale)</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>How hurried or rushed was the task?</td>
<td>Extremely unrushed to extremely rushed (7-point scale)</td>
</tr>
<tr>
<td>Frustration</td>
<td>Describe your level of irritation, stress or annoyance while respeaking this clip.</td>
<td>Extremely low to extremely high (7-point scale)</td>
</tr>
<tr>
<td>Difficulty Level</td>
<td>How easy or difficult was the task?</td>
<td>Extremely easy to extremely difficult to understand (7-point scale)</td>
</tr>
<tr>
<td>Concentration / Engagement</td>
<td>To what extent could you concentrate on the task (i.e. without thinking about other things)?</td>
<td>Not at all to all of the time (7-point scale)</td>
</tr>
</tbody>
</table>

*adapted from Kruger et al. (2014)*
Difficulty level

7-point scale:
1 – very difficult
7 – very easy
Mental demand

7-point scale:
1 – very undemanding
7 – extremely demanding
Frustration

7-point scale:
1 – very low
7 – very high

Graph showing the level of frustration for interpreters, translators, and controls for different speaking rates: slow one speaker, slow many speakers, fast one speaker, fast many speakers.
Concentration – self-report vs. EEG

Self-report (1-7 scale)

EEG
Preliminary respeaking data sample

- Participants
  - 5 translators & 5 interpreters
- Video
  - One speaker
  - Slow speech
- NER
- WER
- BLEU
# Accuracy rate in NERstar

<table>
<thead>
<tr>
<th></th>
<th>ACCURACY RATE (NER value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translators</td>
<td>93.29</td>
</tr>
<tr>
<td>Interpreters</td>
<td>94.39</td>
</tr>
</tbody>
</table>
## Edition vs. recognition errors

<table>
<thead>
<tr>
<th></th>
<th>Edition errors</th>
<th>Recognition errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translators</td>
<td>27.66</td>
<td>72.34</td>
</tr>
<tr>
<td>Interpreters</td>
<td>28.75</td>
<td>69.33</td>
</tr>
</tbody>
</table>
## Reduction

<table>
<thead>
<tr>
<th></th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translators</td>
<td>12.13</td>
</tr>
<tr>
<td>Interpreters</td>
<td>10.49</td>
</tr>
</tbody>
</table>
Word error rate (WER)

- **Reference text**: transcription from respeakers
- **Analysed text**: SR output

<table>
<thead>
<tr>
<th>Group</th>
<th>WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translators</td>
<td>18.93</td>
</tr>
<tr>
<td>Interpreters</td>
<td>14.43</td>
</tr>
</tbody>
</table>
**BLEU (Bilingual Evaluation Understudy)**

- Machine translation measure
- **Reference** text: transcription of the original speaker
- **Analysed** text: transcription from respeakers

<table>
<thead>
<tr>
<th>Group</th>
<th>BLEU RATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translators</td>
<td>56.11</td>
</tr>
<tr>
<td>Interpreters</td>
<td>55.15</td>
</tr>
</tbody>
</table>
So...?

- So far... no conclusive evidence that interpreters are better respeakers
- Ongoing study – more data to come
- What next?
  - Correlate quality in respeaking with other measures: what makes a good respeaker?
  - Implications for respeaker training
Acknowledgements

This study was supported with the grant no. 2013/11/B/HS2/02762 from the National Science Centre for the years 2014-2017

Agnieszka Lijewska  Agnieszka Chmiel  Krzysztof Marasek  Łukasz Brocki  Krzysztof Krejtz
www.facebook.com/RespeakingProject
www.avt.ils.uw.edu.pl
a.szarkowska@uw.edu.pl
lukasz.stanislaw.dutka@gmail.com