Investigating the competences of interlingual respeakers – a preliminary study

Łukasz Dutka, Agnieszka Szarkowska, Krzysztof Krejtz, Olga Pilipczuk
What is respeaking?

A speaker speaks a text → Respeaker respeaks/ translates the text → Speech recognition software transforms the utterance from speech to text → Respeaker/ Editor corrects any errors → Respeaker/ Editor releases a subtitle to be displayed on screen
Measurements

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

del respeaking interlinguistico. Avevamo 58 partecipanti. La maggior parte di loro erano femmine. L'età media era 27 anni. Abbiamo
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
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
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)
Respeaking test

- Screen recording
- Eyetracking
- EEG (Emotiv EPOC)
- Self-report on cognitive load
- Semi-structured interview
What is cognitive load?

- **Cognitive load** can be understood as the amount of cognitive resources (mental effort) used in the working memory during task completion (Sweller, 1988).
  - *Limited cognitive resources* (individual differences in working memory capacity) (Miller, 1956)
  - Each task people perform requires some cognitive resources

- Tasks may exhaust individuals’ working memory capacity leading to **cognitive overload** (e.g. Chandler & Sweller, 1991).

- Working memory capacity may be enhanced by e.g. training or professional experience.
Cognitive load and respeaking

- Multidimensional construct representing the load that performing a particular task imposes on the respeaker’s cognitive system (adapted from Paas & van Merriëboer, 1994)

- Cognitive overload more of a problem in interlingual respeaking
Types of cognitive load (Mayer 2002)

- **Intrinsic**
  - Inherent quality of the material to the participant based on its difficulty

- **Extraneous**
  - The way the information is presented

- **Germane**
  - The remaining available cognitive resources - CL that people use to process and comprehend material
  - Related to processes that contribute to the construction and automation of schemas

- The higher the intrinsic and extraneous load, the less capacity remains in working memory for germane CL, which may result in cognitive overload
Preliminary results on interlingual respeaking

1. Working memory capacity
2. EEG cognitive load
3. Self-reported cognitive load
Interlingual vs. intra lingual respeaking

- Polish speech by prime minister Kopacz
- English speech by president Obama
  - one speaker, slow
## Working memory capacity

<table>
<thead>
<tr>
<th>Group</th>
<th>Listening Span (LSPAN value)</th>
<th>Reading Span (RSPAN value)</th>
<th>Digit Span (forward, fTE_ML)</th>
<th>Digit Span (backward, bTE_ML)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreters</td>
<td>57.6</td>
<td>39.34</td>
<td>6.4</td>
<td>5.5</td>
</tr>
<tr>
<td>Translators</td>
<td>55.35</td>
<td>35</td>
<td>6.17</td>
<td>5.65</td>
</tr>
<tr>
<td>Controls</td>
<td>50.16</td>
<td>35.16</td>
<td>6.25</td>
<td>5.5</td>
</tr>
</tbody>
</table>
LSPAN – total no. of sentence errors

F(2,55) = 2.4, p = 0.09 (marginally significant)
LSPAN - sentence speed errors

$F(2,55) = 2.93, \ p = 0.06$ (marginally significant)
Erickson Flanker Task with Fish (Christ et al. 2011)

- Erickson Flanker task is one of the cognitive inhibition tasks
- It measures the ability to suppress responses that are inappropriate.
  - The faster (lower response latency) accurate subjects’ responses the higher suppressing ability
- Often used for measuring selective attention.
- Information selection working memory function may be crucial for managing cognitive load.
The same response as the target (similar symbols).

The opposite response of the target (different symbols).
Cognitive load

- Can be measured
  - Directly (EEG)
  - Indirectly (self-report)
EEG results (Emotiv EPOC)

- Based on Emotiv Affectiv algorithms
  - Engagement
  - Frustration
EEG engagement (concentration)

<table>
<thead>
<tr>
<th>Group</th>
<th>Interpreters</th>
<th>Translators</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intralingual</td>
<td>.57 (SD=.04)</td>
<td>.58 (.05)</td>
<td>.58 (.03)</td>
<td>.58 (.04)</td>
</tr>
<tr>
<td>Interlingual</td>
<td>.57 (.05)</td>
<td>.53 (.11)</td>
<td>.59 (.06)</td>
<td>.55 (.09)</td>
</tr>
<tr>
<td>Total</td>
<td>.57 (.04)</td>
<td>.55 (.09)</td>
<td>.58 (.04)</td>
<td>.57 (.07)</td>
</tr>
</tbody>
</table>

ANOVA (language & group)
Language: F(1,75)=1.68, p = 0.1988
Group: F(2,75)=1.18, p = 0.3125
### Self-reported cognitive load

<table>
<thead>
<tr>
<th>MENTAL DEMAND</th>
<th>How mentally demanding was the task?</th>
<th>Extremely undemanding to extremely demanding (7-point scale)</th>
</tr>
</thead>
<tbody>
<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)
Self-reported difficulty level
intra-lingual vs. inter-lingual respeaking

7-point scale:
1 - very difficult
7 - very easy

ANOVA (language & group)
Language: F(1,98)=12.30, p = 0.0007
Group: F(2,98)=3.38, p = 0.0380
Self-reported mental demand

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

ANOVA (language & group)
Language: F(1,99) = 18.99, p = 0.0000
Group: F(2,99) = 3.75, p = 0.0269
Self-reported frustration

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

ANOVA (language & group)
Language: F(1,99)=17.00, p = 0.0001
Group: F(2,99)=3.61, p = 0.0305
Self-reported engagement/concentration

7-point scale:
1 - no concentration
7 - full concentration

ANOVA (language & group)
Language: $F(1,99)=0.52$, $p = 0.4723$
Group: $F(2,99)=0.82$, $p = 0.4409$
Interlingual respeaking – concentration

Self-report (1-7 scale)  EEG

no statistically significant correlation
Searching for a competence model in respeaking

- Expertise in simultaneous interpreting is not a function of discrete processing abilities but of task-specific skills (Liu 2001)
  - These may include selective processing, efficient output monitoring and allocation of working memory resources
- Competence acquired through extensive time-on-task
Conclusions

- Cognitive measures do not show differences in competence.
- Working memory tests - Some indications of better working memory (might suggest transfer of interpreting skills into respeaking).
- In general, interpreters are not significantly better when it comes to working memory.
What next?

- Quality in respeaking (finding a measure)
  - NERstar for intralingual (nor particularly suited for interlingual)
  - Adapting BLEU and other machine-translation measures

- Relate cognitive load to performance
  - Cognitive overload will result in lower performance
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
www.facebook.com/RespeakingProject
www.avt.ils.uw.edu.pl
lukasz.stanislaw.dutka@gmail.com
a.szarkowska@uw.edu.pl