

COGNITIVE LOAD IN INTRALINGUAL AND INTERLINGUAL RESPEAKING – A PRELIMINARY STUDY

AGNIESZKA SZARKOWSKA
University of Warsaw
a.szarkowska@uw.edu.pl

KRZYSZTOF KREJTZ
University of Social Sciences and Humanities,
Warsaw

ŁUKASZ DUTKA
University of Warsaw

OLGA PILIPCZUK
University of Warsaw

ABSTRACT

In this paper we present preliminary results of the study on the cognitive load in intralingual and interlingual respeaking. We tested 57 subjects from three groups: interpreters, translators and controls while respeaking 5-minute videos in two language combinations: Polish to Polish (intralingual) and English to Polish (interlingual). Using two measures of cognitive load: self-report and EEG (Emotiv), we found that in most cases cognitive load was higher in interlingual respeaking. Self-reported mental effort that the participants had to expend to complete the respeaking tasks was lower in the group of interpreters, suggesting some parallels between interpreting and respeaking competences. EEG measures showed significant differences between respeaking tasks and experimental groups in cognitive load over time.

KEYWORDS: Respeaking; interpreting; cognitive load; live subtitling; audiovisual translation.

1. Introduction¹

Respeaking is a complex cognitive activity, requiring the coordination of multiple cognitive resources. When producing live subtitles for an audiovisual pro-

¹ This work was funded by the National Science Centre Poland based on decision no. DEC-2013/11/B/HS2/02762. Many thanks to all our participants for their time and patience during the experiments.

gramme, a respeaker must simultaneously listen to the original speech and translate the text – intra- or interlingually – as well as add appropriate punctuation and edit the text if necessary.

In this paper we examine the cognitive load in a group of subjects while they were respeaking two clips: intralingually (from Polish to Polish) and interlingually (from English to Polish). We hypothesise that because interlingual language transfer is a more complex activity, the cognitive load will be higher in respeaking between different languages than within the same language. To test this hypothesis, we compare two measurements of cognitive load: indirect (self-report) and direct (EEG) as well as discuss the findings from the semi-structured interview which followed the respeaking test.

We begin with a brief introduction to respeaking and respeaking landscape in Poland, followed by a short description of the respeaking project, which this study is part of, and then finally, we present the results on the cognitive load in intra- and interlingual respeaking in the three groups of subjects.

2. Respeaking

Respeaking is a method of creating live subtitling with the use of speech recognition (SR) technology. It is “a technique in which a respeaker listens to the original sound of a live programme or event and respeaks it, including punctuation marks and some specific features for the deaf and hard of hearing audience, to a SR software, which turns the recognised utterances into subtitles displayed on the screen with the shortest possible delay” (Romero Fresco 2011: 1). Respeaking has been used since 2001 (Lambourne 2006; Romero Fresco 2011) and it is mainly employed as an intralingual tool, but it can also be used to translate between languages (den Boer 2001).

Respeakers need to master both linguistic and technical competences (see Arumí Ribas and Romero Fresco 2008; Eugeni 2008; Romero Fresco 2011). The former include the ability to paraphrase and condense the text in one language or to translate the words spoken in a live programme into another language. Among technical competences are the knowledge of SR software and subtitling skills. Other necessary skills are the ability to listen and speak at the same time, as well as the ability to simultaneously listen to other speakers and to one’s own voice in order to control the intonation and correctness of the respo-ken text.

2.1. Respeaking in Poland

Slavic languages, as highly-inflected and non-positional, pose many challenges to speech recognition (Ziółko and Ziółko 2011). This, along with commercial considerations, influenced the fact that SR for Polish was developed much later than for English, with two leading desktop systems Magic Scribe (offered by a Polish company Unikkon Integral) and Newton Dictate (from Newton Technologies, a company based in the Czech Republic) becoming available in 2009 and 2014, respectively. As a result, respeaking made a delayed debut in Poland in June 2013 at an accessibility conference at the Polish parliament. The canonization mass of popes John Paul II and John XXIII in Rome (April 2014) was the first event with remote respeaking in Polish and it was also a turning point with live subtitling in Polish made available online to a mass audience. In June 2015, the conference presenting the EU Single Digital Market initiative in Warsaw marked the first use of interlingual (English-Polish) respeaking.

At the end of 2015, as we write this paper, live subtitling for unscripted broadcasts has not been implemented on Polish television, and respeaking is mainly used as one part of all-in-one access service for live events, including sign language interpreting, induction loop and live audio description, all available on the spot and online. Since the number of events with respeaking is still relatively small, there are just a few respeakers working on the Polish market and their work experience is rather limited.

2.2. Respeaking project

This study is part of a larger project on respeaker competences, whose goal is to examine the process of intra- and interlingual respeaking as well as skills and competences required in respeaking.

Due to the fact that there was no respeaking on Polish television and also no professional respeakers at the time of this study, we could not test differences between respeaking experts and novices. Instead, we tested three groups of participants:

- (1) Interpreters and interpreting trainees, i.e. people with well-trained working memory (see Chincotta and Underwood 1998, Yudes et al. 2012, Köpke and Nespoulous 2006), having good linguistic and interpreting skills;
- (2) Translators and translation trainees, i.e. people with good linguistic and translation skills;

- (3) Control group, i.e. people with good English skills, but no translation/interpreting background.

All participants underwent two days of respeaking training, conducted by professional respeaker trainers from the United Kingdom, Italy and Switzerland.

Apart from the main respeaking test, participants were also subjected to working memory, paraphrasing, and proofreading tests. In this paper, we only report the results from part of the main respeaking test related to cognitive load.

3. Cognitive load

Cognitive load is a “multidimensional construct representing the load that performing a particular task” imposes on a person’s cognitive system (Paas et al. 2003a: 64). The amount of load depends on the one hand on the person’s individual characteristics, such as expertise level, age, prior knowledge related to the task, and on the other hand, on task characteristics, including the pace and difficulty of the task, task novelty, etc.

3.1. Cognitive load in novices and experts

Cognitive load theory posits that the human cognitive architecture consists of long-term memory, which is assumed to have “a virtually unlimited capacity” (Brünken et al. 2003: 54), and working memory, which is limited both in capacity (Baddeley and Hitch 1974) and in duration (Peterson and Peterson 1959; Just and Carpenter 1992; Šárka et al. 2015). Long-term memory contains hierarchically organised cognitive schemas “that are used to store and organize knowledge by incorporating multiple elements of information into a single element” (Kirschner et al. 2011: 1). Working memory is responsible for the “mechanisms and processes that control, regulate and actively maintain task-relevant information” (Miyake and Shah 1999 cited after Brünken et al. 2003: 54). If the processing capacity of the working memory is low or approaches zero, then it is likely that the person will experience high cognitive load or overload. The load on working memory can be reduced by using automated schemas in long-term memory (Kalyuga et al. 1999: 351; Paas et al. 2003a: 2). Schemas can “bypass working memory during mental processing thereby circumventing the limitations of working memory” (Pass et al. 2003: 64).

According to Kalyuga et al. (2003, p. 24), the concept of schemas can be used to explain differences between novices and experts. Because experts have many hierarchically organised domain-specific schemas, they can more easily than novices recognise patterns and categorise different elements into a single unit. Unlike in novices, experts' schemas are highly automated. This, in turns, reduces their load on working memory, gives more capacity for processing the task and may positively impact their performance on a task.

While it is generally believed that expertise is domain-specific (see Ericsson and Charness 1994), there is some evidence that experts may transfer their expertise to other tasks, that is apply the previously acquired knowledge to the new task (see Kimball and Holyaok 2000). In other words, as noted by Muñoz Martin (2014: 7), “expert skills may be successfully applied to non-identical tasks”.

3.2. Cognitive load in intra- and interlingual respeaking

Thanks to a number of overlapping competences between interpreters and respeakers – including the ability to listen to the original speech and translate it into another language (or repeat it in the same language) while simultaneously monitoring their linguistic output – in this study we wanted to find whether interpreters will be better predisposed to respeaking, as they had already developed some cognitive schemas necessary to perform such tasks. We set out with an assumption that, having developed and automated some of the schemas useful in respeaking in their long term memory, interpreters will experience lower cognitive load than the other two groups in both respeaking tasks (intra- and interlingual).

Another assumption we wanted to test was whether – given that the act of transferring a text from one language into another is a more cognitively complex activity – the interlingual respeaking task will trigger higher cognitive load than the intralingual one.

3.3. Measuring cognitive load

Paas and van Merriënboer (1994) distinguish three measurable dimensions – or “assessment factors” – of the cognitive load: mental load, mental effort and performance. *Mental load* can be thought of as “an indication of the expected cognitive capacity demands and can be considered an a priori estimate of the cogni-

tive load” (cited after Paas et al. 2003b: 64). *Mental effort*, in turn, is “an aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task” (Paas et al. 2003b: 64). It can therefore reflect the actual cognitive load. Finally, *performance* is an aspect of the cognitive load that shows the person’s achievements in carrying out the task. If the complexity of the task increases, people can compensate for an increase in mental load by expending more mental effort to execute the task, keeping performance at the same level (Paas et al. 2003b: 67).

According to Choi and van Merriënboer (2014: 227), assessing the cognitive load based on mental load is task-centred, focussing on the characteristics of the task, whereas the assessment based on mental effort is more human-centred, focussing on the mental resources expended by the person to deal with the task. The assessment of the cognitive load based on performance relies on the assumption that if people carry out the task faster and produce fewer errors, then their cognitive load is smaller compared to those who make more errors and take more time to complete the task.

In our respeaking study, given the higher complexity involved in translating or interpreting a text from one language into another, we hypothesised that the interlingual respeaking task will induce more *mental load* than the intralingual one. In turn, the *mental effort* expended by respeakers will be larger in the case of interlingual respeaking. We also thought that the *mental effort* will be higher in the case of translators and controls, as they did not yet develop cognitive schemas in their long term memory which would help them deal with such tasks and which would reduce the potential strain on their working memory when respeaking. Because of their expertise, it is possible that interpreters already acquired schemas in their long-term memory which they can use in respeaking to avoid overburdening their working memory, which in turn may result in a better respeaking performance.² This should potentially translate into lower cognitive load.

Cognitive processes can be measured using subjective techniques, such as rating scales, and objective physiological techniques, such as measures of brain or heart activity (see also Brünken et. al. 2003). Rating scale techniques “are based on the assumption that people are able to introspect on their cognitive processes and to report the amount of mental effort expended” (Paas et al. 2003b: 66). As demonstrated by Paas on cognitive load (1992) and earlier by Gopher and Braune (1984), people “are quite capable of giving a numerical in-

² In this paper we do not yet look into performance measures, which will be reported elsewhere.

dication of their perceived mental burden” (Paas et al. 2003b: 66). In this study we used two types of measurement techniques: subjective rating scales and objective physiological technique (EEG), which we explain below.

3.3.1. Self-report

In the study reported here, participants had to assess their cognitive load after each respeaking task using self-report rating scales. We used an adapted version of the self-report questionnaire by Kruger et al. (2014), which was in turn derived from the NASA task load index (Hart and Lowell 1988). This scale for measuring mental workload had also been successfully used to assess the level of translation difficulty by Sun and Shreve (2014).

In our study, after each respeaking task, participants were asked to self-report their cognitive load on 1–7 scale in five categories: difficulty level, temporal demand, mental demand, frustration, and engagement (see Table 1).

Table 1. Self-report questionnaire on cognitive load.

| Parameter | Question after the respeaking task | Rating scale |
|------------------|--|--|
| Difficulty level | How do you assess the general difficulty of the clip you just respoke? | Very easy to very difficult (7-point scale) |
| Temporal demand | How do you assess the pace of the clip you just respoke? | Very slow to very fast (7-point scale) |
| Mental demand | How mentally demanding was the task? | Very undemanding to very demanding (7-point scale) |
| Frustration | Describe your level of irritation, stress or frustration while respeaking this clip. | Very low to very high (7-point scale) |
| Engagement | To what extent could you concentrate on the respeaking task (without thinking about other things)? | I couldn't concentrate at all to I was fully concentrated all the time (7-point scale) |

3.3.2. Electroencephalography (EEG)

Another measure of cognitive load we used in this study was an objective physiological measure: EEG, a neuroimaging technique which detects electrical activity of the brain by using electrodes placed on the scalp. EEG makes it possible to “noninvasively measure brain activity in authentic, real-world settings” (Antonenko et al. 2010: 428).

In the present study we used a low-cost EPOC Emotiv EEG headset which detects brain activity. Based on special algorithms, the activity is “interpreted” as different types of emotions: short term excitement, long term excitement, engagement/boredom, meditation and frustration. Each indicator is scored from 0 to 1; higher scores correspond to greater emotion intensity. The emotion indicators are defined as follows:

- (1) Short term and long term excitement – a “feeling or awareness of physiological arousal in a positive sense” (Harrison 2013).
- (2) Engagement/boredom – “the alertness experienced by a person and the conscious direction of attention towards a task-relevant stimulus” (Harrison 2013). Lower scores may indicate boredom.
- (3) Meditation – a feeling of calmness and relaxation. “Meditation represents how calm a person is: the higher the score, the calmer they are” (Harrison 2013).
- (4) Frustration – a negative feeling of frustration. The higher the score, the more frustrated the study participant is.

4. Method

In this paper, we report on the differences in cognitive load between the three groups of participants in the two respeaking tasks: intra- and interlingual, using two measures of cognitive load: self-report and EEG. We expect these two measures to correlate.

We hypothesise that: (1) interlingual respeaking is more cognitively demanding than intralingual respeaking, and (2) interpreters will experience lower cognitive load when respeaking both tasks owing to a number of similar competences required in interpreting and respeaking.

4.1. Participants

Fifty-seven participants (50 women, 7 men) took part in the respeaking tests, out of whom 23 were translators, 22 interpreters, and 12 controls with no interpreting/translation experience. The interpreters and translators had English as either B or C language. The control group did not have any experience in either translation or interpreting, but spoke fluent English.

The mean age of the participants was 27.48 (SD 5.71), ranging from 19 to 51 years.

47 participants were graduates of linguistics/philology departments, 9 participants studied at other faculties, one did not have a university degree.

Based on the number of years of experience in interpreting and translation (see Table 2), we divided interpreters and translators into two subgroups: experts (3 or more years of experience) and novices (less than 3 years of experience). Given that the development of expertise requires years of “deliberate practice” (Ericsson et al. 1993) and “individual differences in attained levels of performance are not fully understood” (Ericsson 2000: 187), we treat the notions of ‘experts’ and ‘novices’, following Shreve (2002: 161), as two extreme poles of a continuum rather than two separate clear-cut groups.

Table 2. Work experience in interpreting and translation by group.

| Experience | Interpreters and interp. trainees | Translators and translation trainees |
|------------|-----------------------------------|--------------------------------------|
| None | 1 (9.09%) | 1 (4.35%) |
| 1–2 years | 11 (50.00%) | 8 (34.78%) |
| 3–4 years | 4 (18.18%) | 9 (39.13%) |
| 4 + years | 6 (27.27%) | 5 (21.74%) |

4.2. Procedure

Participants were tested individually. They were first given a study information sheet to read and were asked to sign an informed consent to take part in the study. During the respeaking test, the participants’ cognitive activity was monitored with Emotiv EPOC headset. After fitting and calibrating the EPOC head-

set, participants had a short 3-minute training session with a mock intralingual respoking task (slow one-speaker speech) to familiarise themselves with the experimental set-up. The test proper consisted of four randomised intralingual tasks and one interlingual task. After respoking each video, the participant had to self-report the cognitive load for this task, by answering the five questions listed in Table 1.

Each participant used their own voice profile, which was created during the respoking workshops. We used speech recognition software for Polish manufactured by Newton Technologies. While respoking, participants' words were transformed from speech to text by Newton, and they were displayed in FAB Subtiter LIVE, professional software used in live subtitling on television. This means that while respoking, participants could simultaneously monitor their output which was being displayed on the screen.

Participants were instructed to provide television respoking intra- and interlingually, i.e. to repeat or translate the original speech, adding the necessary punctuation marks. They were not asked to use colours for different speakers or to mark speaker changes.

The participants' output was audio recorded and later transcribed for further analysis of respoking performance (which will be reported elsewhere).

After the test, the participants underwent semi-structured interviews, in which they were prompted to assess the difficulty of particular videos, to comment on the test, etc.

4.3. Materials

The respoking test consisted of four 5-minute videos in intralingual Polish-to-Polish respoking and one 5-minute video with interlingual English-to-Polish respoking. The main focus was placed on intralingual respoking, as it is this type of respoking that is most representative of market practices. There were two variables in the intralingual respoking task: speech rate (fast/slow) and number of speakers (one/many). In this paper we only report the results from the comparison of the slow, one-speaker task in intralingual respoking and compare it to the slow, one-speaker interlingual respoking task. We believe they were relatively comparable in terms of the number of speakers, duration and the number of words in the original text to be respoken (see Table 3).

Table 3. Characteristics of the respeaking tasks discussed in this paper.

| | Intralingual task | Interlingual task |
|-----------------|--|--|
| Speech | Prime Minister Ewa Kopacz's New Year's address (2014/2015) | Beginning of President Barack Obama's speech delivered on 4 June 2014 in Warsaw at the 25th Anniversary of Freedom Day |
| Speech rate | 108 words per minute | 94 words per minute |
| Duration | 4 minutes 47 seconds | 4 minutes 55 seconds |
| Number of words | 519 | 463 |

4.4. Study design. Dependent and independent variables

The final dataset for the statistical analyses consists of recordings measured during two experimental tasks: intralingual respeaking and interlingual respeaking.

The main dependent variables from self-report measures of cognitive load were difficulty level, temporal demand, mental demand, frustration and engagement. All the self-report measures were evaluated on 7-points Likert-type scale with high values indicating high level of frustration, effort, difficulty, etc. For analyses of self-report data, 3×2 mixed design ANOVA was used, where the experimental group was a between-subject factor and the task was a within-subject factor.

The EEG variables were frustration, meditation, engagement, short and long-term excitement (see section 3.3.2). All data from Emotiv EEG recordings were carefully checked in terms of variance and missing data. EEG recordings showing no variance or including missing values, which indicate measurement problems or errors, were excluded from the analyses. For the sake of accurate comparisons, only those participants were chosen who completed both experimental tasks (interlingual and intralingual). We need to note here that nine participants (one interpreter, four translators and four controls) refrained from taking part in the interlingual respeaking task. These steps reduced the initial sample, resulting in the final sample consisting of 3,175,229 data records from 30 participants ($N = 13$ in the translators group, $N = 12$ in the interpreters group, and $N = 5$ in the control group). All analyses presented in the paper were based on that sample.

The analyses were conducted with R programming language for statistical computations (R Core Team, 2015). The basic analytical procedure for EEG da-

ta was mixed-design $3 \times 2 \times 10$ Analyses of Variance (ANOVA) with the lme4 package, followed by pairwise comparisons with Tukey's correction when needed. In all ANOVAs, the experimental group (translators vs. interpreters vs. controls) was treated as a between-subject factor, whereas both tasks (interlingual vs. intralingual) and time intervals (10 levels) as within-subjects factors. The time interval factor was calculated by dividing the task completion time for each participant into 10 equal intervals.

5. Results

We first present the results of the self-reported cognitive load, then the EEG data and, finally, semi-structured interview held at the end of the respeaking test between the researcher and the participant.

5.1. Self-reported measures of cognitive load

Difficulty. The analysis of variance on task difficulty revealed a significant main effect of task, $F(1,27) = 9.02$, $p < 0.01$, eta-squared = 0.13. The interlingual task was evaluated as more difficult ($M = 3.43$, $SE = 0.29$) than the intralingual one ($M = 2.2$, $SE = 0.29$), see Figure 2. However, neither the main effect of group, $F(1,27) = 2.07$, $p > 0.1$ nor the interaction effect, $F(1,27) < 1$, were statistically significant. This indicates that none of the groups differed in how they perceived the task difficulty.

Temporal demand. The two-way ANOVA with task temporal demand evaluation showed a statistically significant main effect of the experimental task, $F(1,27) = 11.05$, $p < 0.01$, eta-squared = 0.16. The intralingual task was evaluated as faster ($M = 3.80$, $SE = 0.27$) compared to interlingual ($M = 2.53$, $SE = 0.27$), see Figure 2. Neither the main effect of experimental group, $F(1,27) < 1$, nor the interaction effect, $F(1,27) < 1$, were statistically significant.

Mental demand. More interesting results were shown by ANOVA on the evaluation of mental demand. Similarly to the above statistical tests, we found the main effect of task, $F(1,27) = 15.04$, $p < 0.001$, eta-squared = 0.21. Again, the interlingual task was evaluated as more cognitively demanding ($M = 5.76$, $SE = 0.26$) compared to the intralingual task ($M = 4.30$, $SE = 0.26$), see Figure 2. The analysis showed that the main effect of the experimental group was marginally

significant, $F(1,27) = 2.59$, $p = 0.09$, eta-squared = 0.09. Pairwise comparisons with Tukey's correction showed that interpreters reported lower cognitive effort ($M = 4.6$, $SE = 0.31$) than translators ($M = 5.48$, $SE = 0.30$), $t(27) = 2.27$, $p = 0.07$ (see Figure 1). The mental effort reported by control group participants ($M = 5.01$, $SE = 0.36$) was not significantly different from the effort reported by translators and interpreters. ANOVA also showed that the interaction effect of the experimental group and tasks was not significant, $F(1,27) < 1$.

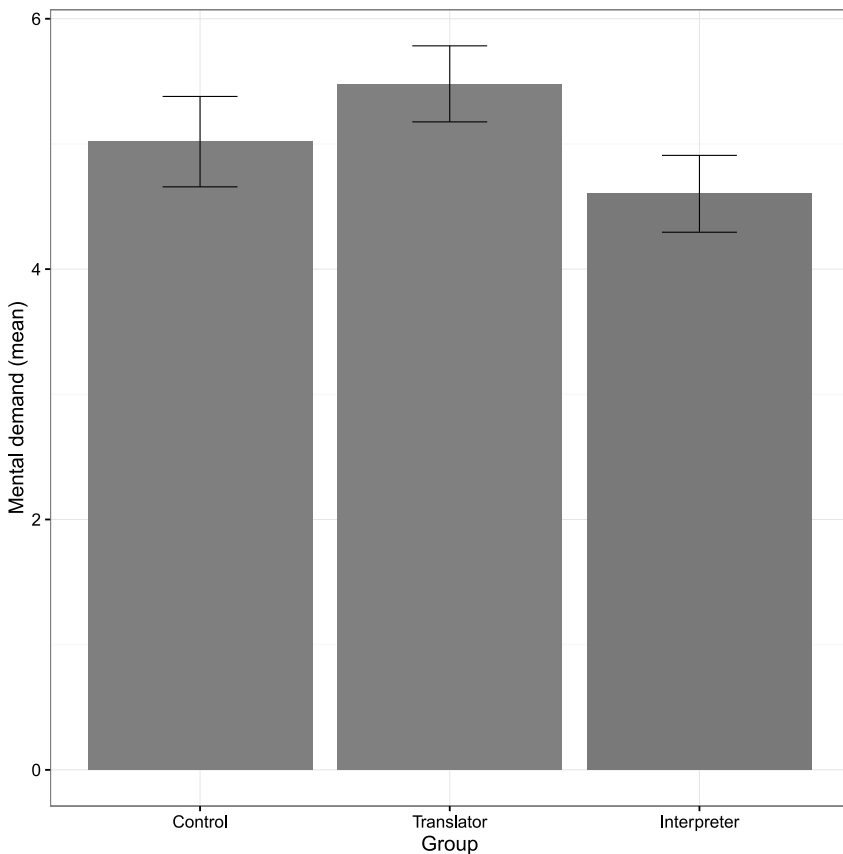


Figure 1. Differences in self-reported mental demand in experimental groups (error bars represent ± 1 standard error).

Frustration. In line with our predictions, the analysis of variance for self-reported frustration revealed a significant main effect of task, $F(1,27) = 13.14$, $p < 0.001$, eta-squared = 0.15. The interlingual task triggered higher frustration among study participants ($M = 4.64$, $SE = 0.33$) than the intralingual task ($M = 3.08$, $SE = 0.33$), see Figure 2. The main effect of the experimental group, $F(2,27) < 1$, as well as the interaction effect, $F(2,27) < 1$, were statistically insignificant, which means that frustration was not higher or lower in any of the groups.

Engagement. Contrary to predictions, the analysis of variance with self-reported engagement as a dependent variable, showed no statistical differences between experimental groups, $F(2, 27) < 1$, experimental tasks, $F(1, 27) < 1$, or the interaction between these two factors, $F(2, 27) = 1.52$, $p > 0.1$ (see also Figure 2).

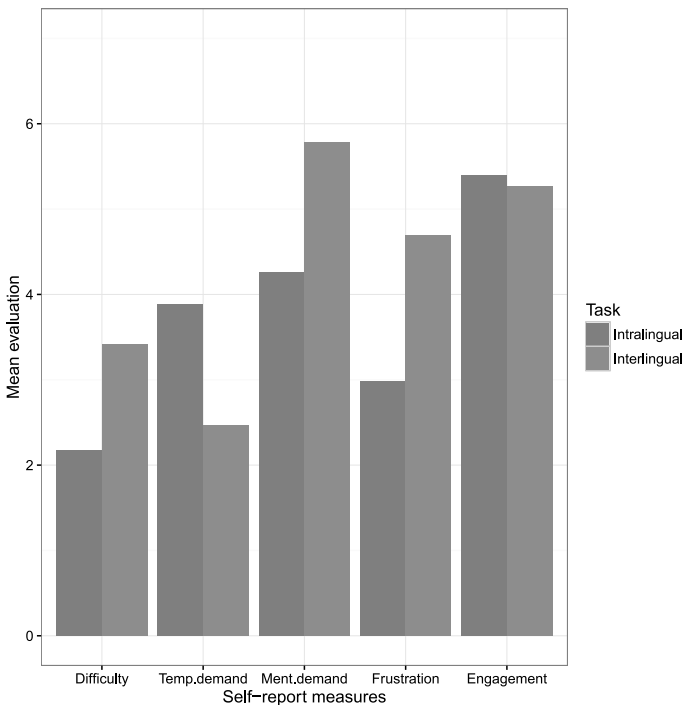


Figure 2. Differences between experimental tasks on all self-reported measures of cognitive load.

5.2. EEG measurements of cognitive load

We start with reporting the EEG results where we found statistically significant differences.

Frustration. The three-way ANOVA, $3(\text{group}) \times 2(\text{task}) \times 10(\text{time})$, on frustration measured by Emotiv EEG showed a significant main effect of time, $F(4.49, 121.31) = 7.43$, $p < 0.001$, eta-squared = 0.09. The post-hoc analyses revealed that frustration was significantly higher from 3rd to 9th time period compared to the beginning and end of the task (see Figure 3).

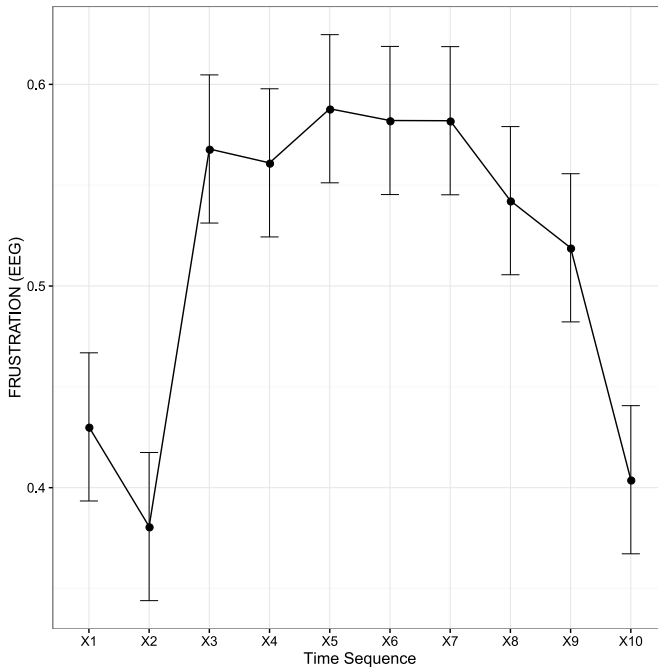


Figure 3. Dynamics of Frustration measured with Emotiv EEG in the respeaking tasks (error bars represent ± 1 standard error).

The main effect of task was marginally significant, $F(1, 27) = 3.36$, $p = 0.08$, eta-squared = 0.02. The intralingual task triggered higher frustration ($M = 0.55$,

SE = 0.03) than the interlingual one ($M = 0.48$, SE = 0.03). Interestingly, the interaction of the experimental group and time was marginally significant, $F(8.99, 121.31) = 1.70$, $p < 0.1$, eta-squared = 0.04. Figure 3 presents the dynamics of frustration during the respeaking tasks for the three experimental groups. The highest frustration in the middle of tasks was measured in the control group. The frustration among interpreters increased sharply from the beginning to the 6th time period to similarly sharply drop after that time. Frustration among translators remained relatively stable during the whole task duration.

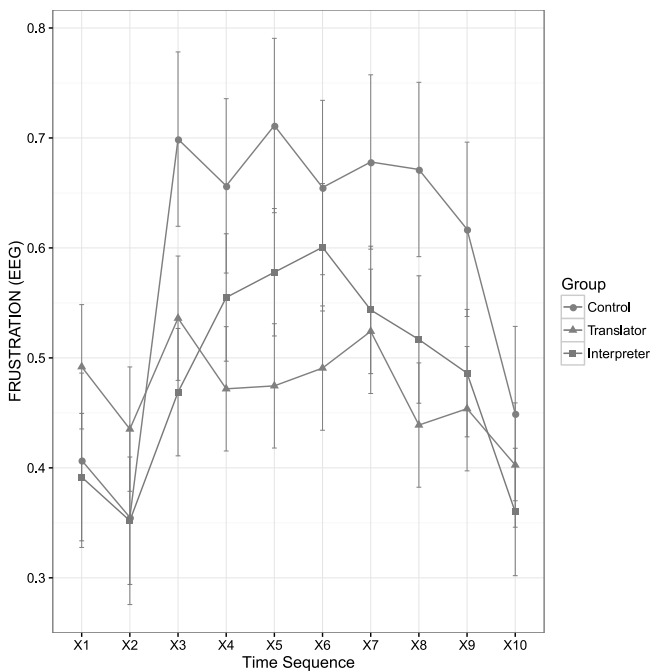


Figure 4. Dynamics of frustration measured with Emotiv EEG during both respeaking tasks by experimental group (error bars represent ± 1 standard error).

All the other main and interaction effects in this analysis of variance were not significant.

Meditation. The second analyses of variance on Emotiv EEG data used Meditation as the dependent variable. The analysis also showed statistically significant main effect of time, $F(3.34, 90.2) = 2.93$, $p < 0.05$, eta-squared = 0.02. The pairwise comparisons revealed that during 6th, 7th, and 8th time periods participants were calmer than in the 2nd one.

Most interestingly the interaction of time, task and group was marginally significant, $F(9, 121.52) = 1.89$, $p = 0.06$, eta-squared = 0.02. To understand the three-way interaction, it was broken into two separate 3(group)x10(time) ANOVAS for each task separately. For the interlingual task, the interaction between time and group was statistically significant, $F(7.05, 95.24) = 2.37$, $p < 0.05$, eta-squared = 0.03 (see Figure 5a), whereas for the intralingual task, it was not significant, $F(3.5, 94.39) = 1.33$, $p > 0.1$ (see Figure 4b). Figure 4a as well as pairwise comparisons clearly show that when respeaking interlingually, the group of translators experienced lowest calmness during the whole task duration. In contrast, during the intralingual respeaking task, experimental groups did not differ on dynamics of Emotiv EEG Meditation indicator (see Figure 5b).

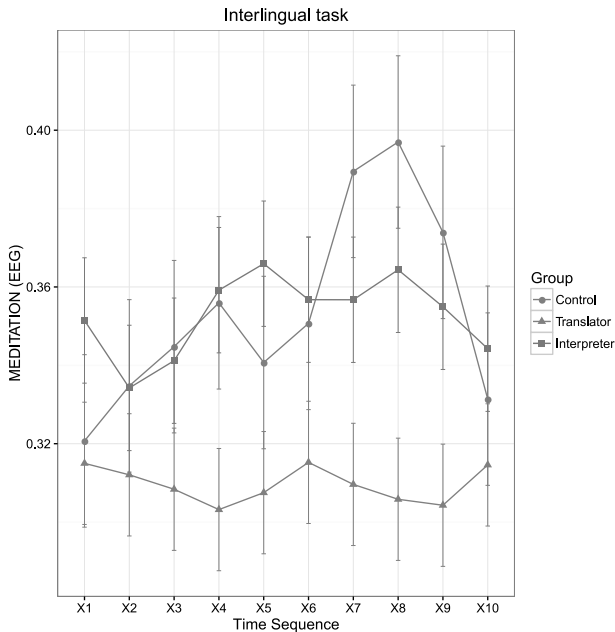


Figure 5a. Interaction of time and experimental group on the Meditation parameter in the interlingual task (error bars represent +/-1 standard error).

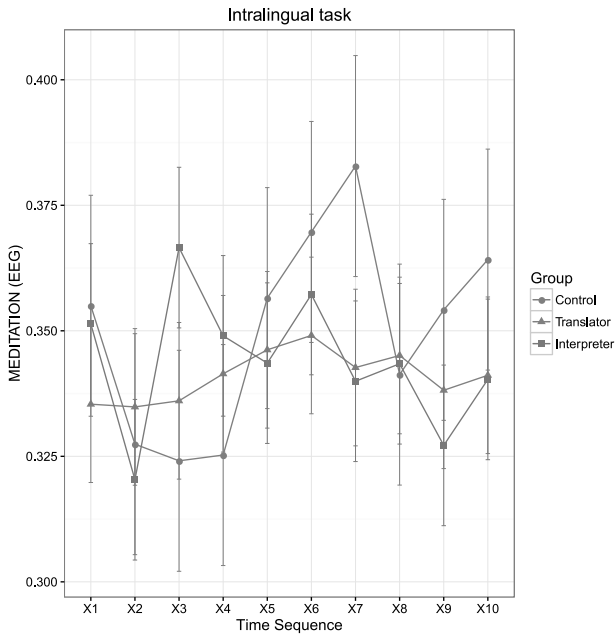


Figure 5b. Interaction of time and experimental group on the Meditation parameter in the intralingual task (error bars represent +/- 1 standard error).

Further analyses of variance with Emotiv EEG measures were conducted on Emotiv engagement. In line with the analysis of self-report engagement, we did not obtain any – even marginally – statistically significant effect. The same lack of significant results was obtained for analysis of Emotiv EEG long-term excitement as a dependent variable. For short-term excitement, the ANOVA revealed only main effect of time being marginally significant, $F(3.85, 103.96) = 2.15$, $p = 0.08$, eta-squared = 0.02. The effect can be evaluated as very weak since it explains only 2% of the entire variance of short-term excitement.

Summing up the EEG results, higher frustration was experienced by all participants in the interlingual re-speaking task compared to the intralingual task. Frustration was highest in the middle part of the tasks. Participants from the translators group experienced less calmness during the whole interlingual re-speaking task, at the same time being steadily frustrated. Interpreters experienced high frustration in the middle of the tasks, however the high level of frustration lasted relatively briefly compared to the translators and the control

group. On the other hand, they also were relatively and steadily calm during the whole duration of tasks. Lastly, no statistically significant effects were found for short-term excitement and engagement measured with Emotiv EEG.

We did not find any statistically significant difference between experts and novices in either self-report or EEG measures.

5.3. Semi-structured interview

When asked which clip they considered to be the most difficult, 37.5% participants stated during the semi-structured interview after the respeaking test that it was the interlingual respeaking task. Only 7% said it was the intralingual clip. Among the difficulties enumerated by the participants related to the interlingual task were: finding the right Polish equivalent for some of the terms used by President Obama in his speech (such as *the Righteous Among the Nations*, *Velvet Revolution*, *the Home Army*) and comprehension problems (e.g. related to the way President Obama pronounced some Polish terms, like *Lech Wałęsa*, *Stanisław Kostka*, *kielbasa* or *Jan Karski*). Some participants stated that although they considered the clip to be easy (slow pace of speech, familiar topic), they still had problems with rendering it into another language and that it was interpreting/translation that made it much more difficult to respeak, as they were not used to such type of task.

As regards the intralingual respeaking task, some participants reported problems with respeaking long sentences and with following the pace of speech delivered by Prime Minister Kopacz. This may have stemmed from the fact that although the speech was delivered orally, it was pre-scripted and read out from a prompter, rather than delivered naturally.

6. Discussion

In this study, we wanted to find whether interlingual respeaking induces more cognitive load than intralingual respeaking and whether the cognitive load is lower in the interpreting group.

In the self-report questionnaire, we tested five parameters of cognitive load (see Table 1). In three of them, we found significant differences between intra- and interlingual respeaking: difficulty level, mental demand and frustration. The difficulty level parameter is related to the notion of *mental load* (see Paas et al. 2003b) and may be regarded as an estimation of cognitive load required during

a task, resulting from the characteristics of the task. The interlingual task in our study was considered by all participants to be more difficult than the intralingual task (3.43 and 2.2, respectively, on the 1–7 scale). The mental demand parameter used in the self-report questionnaire can be taken to represent the *mental effort* required by participants to complete the task (see Paas et al. 2003b). It reflects the cognitive load participants had to expend when performing the tasks. Similarly to the difficulty parameter, participants reported significantly higher mental demand in the interlingual task. Not only did the participants think that the interlingual task was more difficult than the intralingual task, but they also declared to have had to invest more effort into respeaking interlingually, suggesting both higher mental load and mental effort in the interlingual respeaking task.

Analyses of the self-report measures of the cognitive load showed that the interlingual task was evaluated as more difficult, mentally demanding, and frustrating compared to the intralingual one across the three groups of participants. These findings clearly indicate the potentially larger cognitive load during the interlingual respeaking task which may be due to the fact that comprehending the source text was more difficult when it was in a foreign language.

The interlingual task was evaluated as less demanding in terms of time (speech rate) than the intralingual task. This difference may be related to the higher speech rate in the intralingual clip (see Table 2). It points to a potentially interesting finding that the pace of the task may be as important indicator of cognitive load as the type of translation task (intralingual vs. interlingual).

One of the most important findings of this study is that although interpreters did not differ from translators and controls in how they perceived the difficulty of both tasks, they reported lowest mental demand. We believe this finding shows that interpreters may have developed schemas in their long-term memory relevant to this type of task, which helped them reduce the strain on their working memory involved in respeaking. It also points to the possibility that there was a transfer of skills from interpreting to respeaking, particularly as regards the ability to listen to the source text while simultaneously producing the target text and monitoring their output. These findings now need to be correlated with respeaking performance measures.

The self-report measures did not fully correspond with the EEG measures when it comes to the cognitive load involved in intra- vs. interlingual respeaking. Despite higher self-report values for the interlingual task, the EEG data showed higher frustration in the intralingual task. While we are not certain how to interpret these results, we believe this may reflect the fact that while some participants were trained in interpreting and/or translation between two different

languages, they were not accustomed to respeaking within the same language. Being trained in linguistics and translation studies, participants may have also intuitively allocated more difficulty to linguistic transfer between languages rather than within the same language. Another reason to explain this finding may be related to the speech rate and the number of pauses in both clips. While in the interlingual clip President Obama was delivering his speech at a very slow rate (92 wpm) and was making frequent pauses, Prime Minister Kopacz in the intralingual clip was reading from a prompter and did not make many pauses, which resulted in a slightly higher speech rate (107 wpm). The differences in speech rates in both tasks were also noticed by participants in the self-report questionnaire, as the intralingual task was rated higher in terms of temporal demand, although the difference between the two tasks did not reach statistical significance (see Figure 2). All this may indicate that frustration levels correlate with the pace of dialogue: the higher the speech rate, the higher the frustration experienced by the participants.

As rightly pointed out by Antonenko et al. (2010: 427), “subjective rating scales do not give insight into fluctuations in instantaneous load over time”. The EEG data we collected allowed us to examine the dependent variables such as Frustration and Meditation in the respeaking task in relation to time. The highest Frustration values were detected in the middle of the tasks (periods 3 to 9 out of 10). This might suggest that the participants found it hard to follow the pace of the original speaker and over time their ear-voice span became longer,³ which resulted in an increased burden on their working memory and eventually cognitive overload. The sense of not being able to cope with the task triggered frustration.

Frustration levels, as measured by EEG, differed depending on the participant group. The highest levels were recorded in the control group, which may be an indication of the highest cognitive load experienced by this group. This is not surprising, as they did not have any background in interpreting/translation. Importantly, interpreters’ Frustration dramatically dropped towards the end of respeaking tasks, which can be an effect of their cognitive resources release. We may expect that the interpreters developed cognitive strategies to deal with highly cognitively demanding tasks like respeaking.

Frustration measures corresponded with the measure of calmness (Emotiv EEG Meditation). The results show that translators experienced lowest calmness while respeaking the interlingual clip, which may show that the interlingual task was more frustrating for them as it required using other set of cognitive re-

³ We will report on the ear-voice span elsewhere.

sources than they were accustomed to. Interpreters experienced significantly calmer emotional state, probably being more confident of their abilities to complete the interlingual respeaking task accurately. Being more trained in performing parallel tasks such as simultaneous listening and transferring text between languages, at the same time monitoring their linguistic output, interpreters may have developed the schemas useful in respeaking, which may explain higher calmness indicators in the case of this group and the interlingual task.

6.1. Limitations of the study

An important limitation of this study is that we did not examine whether the length of the task affects the cognitive load and performance: since participants had five clips to respeak, it would be interesting to see whether the measures differed depending on fatigue, which possibly increased along with the duration of the experiment.

Because of the limited number of participants in the study, we could not rely on inferential statistics to examine detailed differences between expert interpreters and translators on the one hand, and novice interpreters and translators on the other.

7. Conclusions

To the best of our knowledge, this is the first study comparing cognitive load in intra- and interlingual respeaking. Our findings show that interlingual respeaking was perceived by participants as more cognitively demanding than intralingual respeaking, as it is an activity more complex, possibly requiring more cognitive effort and expertise. Although we expected to find more prominent differences between interpreters and the other two groups across all categories of cognitive load, the interpreters were indeed found to have experienced lower cognitive load in some categories, particularly in self-reported mental demand.

The results also point to the importance of speech rate of the original clip - fast pace of speech increases the cognitive effort necessary in respeaking, as shown by our EEG results, where the faster intralingual clip was found to trigger more frustration than the slower interlingual clip.

Further studies on cognitive load in respeaking could look into differences between professional respeakers and novices in other languages, countries, and

language combinations. It would also be interesting to test other types of clips and genres.

We intend to further analyse the time intervals to examine what we term “respeaking crisis points”, i.e. moments when respeakers’ performance was at its lowest and where frustration was the highest. We are also planning to correlate cognitive load with respeaking performance quality.

REFERENCES

- Antonenko, P., F. Paas, R. Grabner and T. van Gog. 2010. “Using electroencephalography to measure cognitive load”. *Educational Psychology Review* 22. 425–438.
- Arumí Ribas, M. and P. Romero Fresco. 2008. “A practical proposal for the training of respeakers”. *The Journal of Specialised Translation* 10. 106–127.
- Baddeley, A.D. and G. Hitch. 1974. “Working memory”. In: Bower, G.H. (ed.), *The psychology of learning and motivation: Advances in research and theory* (vol. 8). New York: Academic Press. 47–89.
- Brünken, R., J. Plass and D. Leutner. 2003. “Direct measurement of cognitive load”. *Educational Psychologist* 38(1). 53–61.
- Boer, C. den. 2001. “Live interlingual subtitling”. In: Gambier, Y. and H. Gottlieb (eds.), *(Multi)media translation: Concepts, practices and research*. Amsterdam/Philadelphia, Benjamins. 167–172.
- Chincotta, D. and G. Underwood. 1998. “Simultaneous interpreters and the effect of concurrent articulation on immediate memory. A bilingual digit span study”. *Interpreting* 3. 1–20.
- Choi, H.-H., J.J.G Merriënboer and G.W.C. Paas. 2014. “Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load”. *Educational Psychology Review* 26(2). 225–244.
- Ericsson, K.A. 2000. “Expertise in interpreting: An expert-performance perspective.” *Interpreting* 5(2). 187–220.
- Ericsson, K.A., R.T. Krampe and C. Tesch-Römer. 1993. “The role of deliberate practice in the acquisition of expert performance”. *Psychological Review* 100. 363–406.
- Eugeni, C. 2008. “A Sociolinguistic approach to real-time subtitling: Respeaking vs. shadowing and simultaneous interpreting”. In: Cynthia, J., K. Bidoli and E. Ochse (eds.), *English in international deaf communication* 72. Bern: Peter Lang. 357–382.
- Gopher, D. and R. Braune. 1984. “On the psychophysics of workload: Why bother with subjective measures?” *Human Factors* 26. 519–532.
- Harrison, T. 2013. “The Emotiv mind: Investigating the accuracy of the Emotiv EPOC in identifying emotions and its use in an Intelligent Tutoring System”. Technical report, Department of Computer Science and Software Engineering University of Canterbury.

- Hart, S.G. and L.E. Staveland. 1988. "Development of Nasa-Tlx (Task Load Index): Results of empirical and theoretical research." In: Hancock P.A. and N. Meshkati (eds.), *Human mental workload*. Amsterdam: North-Holland. 139–183.
- Just, M.A. and P.A. Carpenter. 1992. "A capacity theory of comprehension: Individual differences in working memory". *Psychological Review* 99(1). 122–149.
- Kalyuga, S., P. Chandler and J. Sweller. 1999. "Managing split attention and redundancy in multimedia instruction". *Applied Cognitive Psychology* 13. 351–371.
- Kalyuga, S., P. Ayres, P. Chandler and J. Sweller. 2003. "The expertise reversal effect". *Educational Psychologist* 38(1). 23–31.
- Kimball, D.R. and K.J. Holyoak. 2000. "Transfer and expertise". In: Tulving, E. and F.I.M. Craik (eds.), *The Oxford handbook of memory*. New York: Oxford University Press. 109–122.
- Kirschner, F., L. Kesterand G. Corbalan. 2011. "Cognitive load theory and multimedia learning, task characteristics, and learner engagement: The current state of the art". *Computers in Human Behavior* 27(1). 1–4.
- Köpke, B. and J.-L. Nespoulous. 2006. "Working memory performance in expert and novice interpreters". *Interpreting* 8. 1–23.
- Kruger, J., E. Hefer and G. Matthew. 2014. "Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2". *Journal of Eye Movement Research* 7(5). 1–15.
- Lambourne, A. 2006. "Subtitle respeaking. A new skill for a new age". *Intralinea*. <http://www.intralinea.it/specials_old/respeaking/eng_more.php?id=447_0_41_0_M>
- Paas, F. and J.J.G. van Merriënboer. 1994. "Instructional control of cognitive load in the training of complex cognitive tasks". *Educational Psychology Review* 6. 51-71.
- Paas, F., A. Renkl and J. Sweller. 2003a. "Cognitive load theory and instructional design: recent developments". *Educational Psychologist* 38(1). 1–4.
- Paas, F., J.E. Tuovinen and H. van Gerven Tabbers. 2003b. "Cognitive load measurement as a means to advance cognitive load theory". *Educational Psychologist* 38(1). 63–71.
- Peterson, L.R. and M.J. Peterson. 1959. "Short-term retention of individual verbal items". *Journal of Experimental Psychology* 58. 193–198.
- R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. <<https://www.R-project.org/>>.
- Romero Fresco, P. 2011. *Subtitling through speech recognition: Respeaking*. Manchester: St Jerome.
- Shreve, Gregory M. 2002. "Knowing translation: cognitive and experiential aspects of translation expertise from the perspective of expertise studies". In: Riccardi, A. (ed.), *Translation studies: Perspectives on an emerging discipline*. Cambridge: Cambridge University Press. 150–171.
- Sun, S. and M.G. Shreve. 2014. "Measuring translation difficulty. An empirical study". *Target* 26(1). 98–127.
- Timarová, Š., I. Čeňková and R. Meylaerts. 2015. "Simultaneous interpreting and working memory capacity". In: Ferreira, A. and J.W. Schwieter (eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting*. Amsterdam: Benjamins. 101–126.

- Yudes, C., P. Macizo and T. Bajo. 2012. “Coordinating comprehension and production in simultaneous interpreters: evidence from the articulatory suppression effect”. *Bilingualism: Language and Cognition* 15. 329–339.
- Ziółko, M. and B. Ziółko (2011). *Przetwarzanie mowy* [Speech processing]. Kraków: Wydawnictwo AGH.

Address correspondence to:

Agnieszka Szarkowska
University of Warsaw
Institute of Applied Linguistics
Dobra 55
00-312 Warszawa
Poland
a.szarkowska@uw.edu.pl