

Paraphrasing in respeaking – comparing linguistic competence of interpreters, translators and bilinguals

Agnieszka Chmiel (corresponding author)

Department of Translation Studies, Adam Mickiewicz University, Poznań, Poland

al. Niepodległości 4, 61-674 Poznań, Poland

achmiel@amu.edu.pl

+48 61 8293517

Agnieszka Lijewska

Department of Psycholinguistic Studies, Adam Mickiewicz University, Poznań, Poland

al. Niepodległości 4, 61-674 Poznań, Poland

alijewska@wa.amu.edu.pl

+48 61 8293525

Agnieszka Szarkowska

Institute of Applied Linguistics, University of Warsaw, Warsaw, Poland

ul. Dobra 55, 00-312 Warsaw, Poland

Centre for Translation Studies, University College London, London, UK

GORDON SQUARE, 16-18, LONDON, WC1H 0AG

a.szarkowska@uw.edu.pl

+48 22 5526033

Łukasz Dutka

Institute of Applied Linguistics, University of Warsaw, Warsaw, Poland

ul. Dobra 55, 00-312 Warsaw, Poland

lukasz.dutka@uw.edu.pl

+48 22 5526033

Paraphrasing in respeaking – comparing linguistic competence of interpreters, translators and bilinguals

Respeaking is a method of producing subtitling for live events and TV programmes. Respeakers repeat speakers' utterances so that they may be changed by speech recognition software into subtitles for the deaf and hard of hearing. Respeakers need to paraphrase the text so that it conforms with temporal and spatial constraints of subtitling. Due to the similarities between respeaking, interpreting and translation, we tested interpreters, translators and bilingual controls on a paraphrasing task to see if interpreters or translators would manifest any advantage thanks to experience. Following a respeaking training, the participants were asked to paraphrase sentences with semantic redundancies, oral discourse markers and false starts in a simultaneous and delayed condition. Contrary to our predictions, we found that experience did not modulate paraphrasing quality or speed in general, but interpreters did outperform other groups when eliminating semantic redundancies, which were also the most difficult reformulations to tackle for all participants. The data suggest that while interpreters and translators are not better predisposed to become respeakers than regular bilinguals, at least as regards the paraphrasing performance, certain aspects of the interpreting experience (the need to express meaning concisely within time constraints) may offer slight advantage in producing well-formed respoken subtitles.

Keywords: respeaking; paraphrasing; interpreting; translation; subtitling for the deaf and hard-of-hearing

1. Introduction

Accessibility is one of eight priority areas within the European Commission's European Disability Strategy 2010-2020, which entails inclusion of people not only with physical but also sensory disabilities. When it comes to individuals with hearing disabilities, their access to live programmes or events can be increased through *respeaking* – a method of producing subtitles to live programmes in real time using speech recognition technology. As defined by Romero-Fresco (2011, p. 1), 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 speech recognition software, which turns the recognised utterances into subtitles displayed on the screen with the shortest possible delay”. The target group of respoken subtitles are mainly deaf and hard of hearing viewers (Eugeni, 2007, 2008b), as well as language learners and foreigners (Díaz-Cintas & Remael, 2007; Vanderplank, 1988). As the demand for respeaking is expected to grow in the light of accessibility policies, training programmes for qualified respeakers are being launched both in the academia and audiovisual translation industry (Romero Fresco, in press; Romero-Fresco, 2012). One of the questions which is immediately raised when candidates for respeakers are sought is that of the necessary competences. The aim of our study was to check whether interpreters or translators, as opposed to regular bilinguals, already possess the skills and competences which make them better candidates for respeakers. To this end, unlike Szarkowska et al. (under review), we did not look at the general respeaking performance across the three groups. Instead, we devised a sentence paraphrasing task to investigate paraphrasing as a fundamental skill in respeaking administered in a highly controlled environment. This enabled us to look for advantages manifested by particular participants with respect to a number of crucial factors which jointly determine the quality of respeaking performance.

1.1. Respeaking

Respeaking, also known as ‘voice writing’, has been used since 2001 (Lambourne, 2006; Romero-Fresco, 2011). The vast majority of respeaking is done intralingually, that is within the same language, but it can also be employed interlingually, i.e. to translate between different languages (de Korte, 2006; Robert & Remael, 2017; Romero Fresco & Pöchhacker, 2017).

The process of respeaking resembles that of simultaneous interpreting (Eugeni, 2008a; Marsh, 2006; Romero-Fresco, 2011). Prior to commencing the work, both respeakers and interpreters have to prepare for the task by familiarising themselves with expected terminology. During the task, they need to simultaneously listen to the original text, produce their linguistic output and constantly monitor its quality. While the interpreter usually renders the original speech into another language and this rendition is immediately available to the audience, the respeaker's spoken words are turned into text using speech recognition software and are later displayed as subtitles on viewers' TV screens with a delay usually lasting several seconds (Ofcom, 2015). Apart from transferring the verbal content of the original speaker's utterance, respeakers also need to insert punctuation marks. This is usually done by voicing appropriate commands, for instance the respeaker says "comma" and the speech recognition software turns this command into the punctuation mark ",". In some countries, like the UK, respeakers also need to monitor the respoken output and correct errors whenever possible, which puts even more strain on their cognitive resources; in other countries, the editing is done by another person. Finally, both interpreters and respeakers usually work in teams, changing every 30 minutes or so, in order to maintain good quality of the service. Otherwise, the strain on working memory may become too extensive, resulting in decreased performance quality.

Respeaking is rarely done verbatim: owing to fast speech rates, overlapping speech and limitations of human working memory, respeakers need to condense the original text (Luyckx, Delbeke, Van Waes, Leijten, & Remael, 2010; Romero-Fresco, 2009). They rephrase the original, removing the unnecessary elements of spoken language, such as false starts, or repetitions. Research has shown that condensation and reduction in respeaking mostly occur in the case of information overload, delay, fast speech rates and overlapping speech (Luyckx et al., 2010). In order to cope with these issues, respeakers resort to different

strategies, which according to Luyckx et al. (2010, p. 23) include “using a shorter near-synonym or equivalent expression” (p. 23), omitting entire units, replacing questions with affirmative sentences and simplifying indicators of modality. Although deaf communities normally expect verbatim subtitles (Neves, 2008; Robson, 2004; Szarkowska & Laskowska, 2015), in practice such subtitles would have to be displayed with very high presentation times, requiring extremely fast reading speeds (Ofcom 2015). Thus, certain reformulations are part and parcel of respeaking.

Live subtitling with respeaking always involves a degree of delay – also referred to as time lag or latency (Mikul, 2014) – between the original speech and the corresponding subtitle (Romero-Fresco, 2011). The delay usually varies from a few to several seconds; in a recent Ofcom report (2015) the average delay in live subtitling in the UK was found to be 5.6 seconds. It is caused by the nature of the live subtitling production process: it simply takes some time for respeakers to listen to the original utterance, produce its respoken version and later for the software to turn it into text and then broadcast as subtitles. The delay has been found to be the most disturbing factor for deaf and hard of hearing viewers (Mikul, 2014).

Respeakers are trained to speak in what is known as ‘respeaking units’, which ideally should comprise from five to seven words (Romero-Fresco, 2011). This is because some speech recognition software, such as the frequently used Dragon Naturally Speaking manufactured by Nuance, requires respeakers to make a brief pause so that the spoken text can be turned into written output. Unlike in interpreting, the respoken text needs to fit limited subtitle space on the screen. As long sentences extending over many lines are difficult to read in subtitling (Díaz-Cintas & Remael, 2007), respeakers should aim to contain their sentences in two lines (Romero-Fresco, 2011). Using what is known as ‘the salami technique’ (Jones, 2002), i.e. cutting the original spoken text into appropriate units, respeakers are required to produce sentences up to 12 words and not longer than 24 words (Romero-Fresco, 2011). This

approach implies the necessity to paraphrase the original text. In the present study we used a paraphrasing task to examine advantages among candidates for respeakers. We decided to specifically look for those advantages among translators, interpreters as compared to regular bilingual speakers given that previous research, reviewed extensively below, demonstrated paraphrasing to be closely linked to translation and interpreting training and performance.

1.2. Paraphrasing

According to Paradis (2007, p. 17), “[p]araphrasing is a phenomenon which does not differ in kind from translation, that is, saying more or less the same thing with different words.”

Paraphrasing can be seen as intralingual translation or ‘rewording’ (Jakobson, 1959), where the same language is used to convey the same message using different phrasing. Paraphrasing has been frequently used as a task to be compared with translation based on the assumption that it involves similar linguistic processing without the interlingual component (Moser, 1983; Russo & Salvador, 2004; Whyatt, Stachowiak, & Kajzer-Wietrzny, 2016).

A number of previous studies showed important parallels between paraphrasing in intra- and interlingual translation. Zethsen (2009) compared interlingual and intralingual translations and found little difference between the two activities. Whyatt et al. (2016) compared paraphrasing with interlingual translation performed by professional translators by means of eye-tracking and key-logging. They found that interlingual translation was more cognitively demanding than paraphrasing. The two tasks displayed similar patterns of text production and pausing, which Whyatt et al. (2016) interpreted as evidence that the participants transferred their translation skills into paraphrasing.

Paraphrasing is part and parcel of subtitling. Owing to the limited time and space available for subtitles on screen, the original utterance often undergoes paraphrasing (Ivarsson & Carroll, 1998). This can be achieved through simplifying verbal periphrases, turning

complex sentences into simple ones, generalising enumerations, simplifying indicators of modality, changing negations or questions into affirmative clauses, turning indirect speech into direct and passive into active, omitting question tags, manipulating theme and rheme, etc. (Díaz-Cintas & Remael, 2007; Gambier, 2006). Gottlieb (1998) argues that “the full transcription/translation of the spoken discourse in films and television is seldom desirable” (p. 247) because subtitlers – and we believe this also includes respeakers – can take advantage of and reduce redundancy.

Apart from written and audiovisual translation, paraphrasing is an important activity in conference interpreting. It has been used in aptitude tests for candidates to conference interpreting programmes (Moser, 1983; Russo & Salvador, 2004). It is also applicable as an exercise in introductory and more advanced interpreter training courses (Bartłomiejczyk, 2015; Gillies, 2001; Setton & Dawrant, 2016). Professional interpreters use it as an interpreting strategy (Liontou, 2012) or a reformulation tactic (Gile, 2009) for example to help the interpreter express the source text meaning when lacking the proper term, to express an idea more succinctly or to interpret idioms (Chmiel, 2015; Tzou, Vaid, & Chen, 2016).

Anderson (1994) compared the performance of conference interpreters in three tasks: interpreting, paraphrasing and shadowing (which involves simultaneous verbatim repetition of the source text). She measured ear-voice span (EVS), i.e. the delay between the source text item uttered by the speaker and the corresponding target text equivalent uttered by the interpreter. EVS is considered an important manifestation of processing in simultaneous interpreting (Lee, 2002; Timarová, Dragsted, & Hansen, 2011). It can vary depending on the source text speed (Barik, 1973; Lee, 2002), text type (Lamberger-Felber, 2001) and sentence length (Lee, 2002). In general, longer EVS might reflect processing difficulty. Anderson (1994) found no difference between EVS in interpreting and paraphrasing, suggesting a similar processing load in both tasks. In terms of quality, interpreting and paraphrasing did

not differ on information content, but interpreting resulted in lower text cohesion and more grammatical and style-related infelicities.

Christoffels and de Groot (2004) also compared paraphrasing, interpreting and shadowing. They found that the quality of interpreting and paraphrasing was equally high but lower than that of shadowing. The temporal delay was the smallest for shadowing, followed by interpreting and the longest latencies were obtained for paraphrasing. Christoffels and de Groot interpreted this finding in combination with the output performance data as evidence against similarity of paraphrasing and interpretation.

Even though the above cited research is not completely unequivocal with respect to the similarity of paraphrasing and translation or interpreting, it still indicates that some of the processing might be similar across these activities. The present study compares the performance of interpreters, translators and regular bilinguals in a paraphrasing task in order to identify the group from which would-be respeakers should preferably be recruited.

2. The present study

In the present study we asked interpreters, translators and bilingual controls who have completed initial training in respeaking to paraphrase sentences. Thereby we aimed to identify the group which is better equipped to coping with paraphrasing. As mentioned above, respeaking performance can be assessed qualitatively and quantitatively. Respeaking is of higher quality when the highest possible semantic equivalence between the source text and the respoken output is achieved. Additionally, the deletion of redundancies is required and the reformulation through synonyms is expected so that the respoken output can be easily transformed into well-formed subtitles. As regards the quantitative assessment, the smaller the delay between the source text and the respoken output, the better.

The sentences used in the study contained redundant material as well as phrases which required synonymic substitutions. The participants were asked to paraphrase the sentences to form respoken subtitles, i.e. they were required to omit redundancies and replace phrases with more succinct synonyms. The paraphrases were later assessed both qualitatively (in terms of semantic equivalence and accuracy of the expected changes) and quantitatively (in terms of the delay between the source text and the output).

By the very nature of their professions and as a result of their professional training, interpreters and translators have had more experience with reformulation (indispensable during translating) than bilinguals. Therefore, we expected interpreters and translators to outperform regular bilinguals in paraphrasing. This would be manifested in more accurate¹ and more semantically equivalent paraphrases obtained from interpreters and translators than from bilinguals.

We used bilinguals, rather than monolinguals, as controls because numerous linguistic studies have found co-activation of both languages and language non-selective access, which directly influences performance (Spalek, Hoshino, Wu, Damian, & Thierry, 2014; Van Assche, Duyck, & Hartsuiker, 2012)). If we had used monolinguals to be compared with interpreters and translators, these groups would differ in two respects: the translation/interpreting skill and the number of known languages. Thus, we would not be able to tell which could explain any differences revealed by our study. Additionally, both bilinguals and interpreters/translators enjoy the bilingual advantage (Poarch & Bialystok, 2015), which may entail improved cognitive processing in comparison with monolinguals.

By manipulating the experimental stimuli (introducing semantic redundancies, oral discourse markers and false starts), we wanted to see if certain types of stimuli are more difficult for the participants to paraphrase than others. This could help adjust future

¹ We understand accuracy as performing the required deletions of the unnecessary material (e.g. oral discourse markers, hesitations, etc.).

respeaking training to the specific needs of trainees. We predicted that semantic redundancies would be more difficult to paraphrase than oral discourse markers and false starts, which would be reflected in the drop of paraphrasing accuracy and semantic equivalence in the sentences with semantic redundancies. Furthermore, since interpreters and translators are experienced in reformulation, we expected them to outperform bilinguals particularly when coping with semantic redundancies. We expected to see smaller differences between groups when paraphrasing sentences with oral discourse markers and false starts.

By introducing two conditions, i.e. paraphrasing simultaneously and with a delay, we wanted to see if the simultaneous nature of the task was particularly problematic to those participants who have no experience with simultaneous interpreting (translators and bilinguals). We expected to see the main effect of condition with simultaneous paraphrasing being less accurate than delayed paraphrasing (similarly to the findings obtained by Christoffels & de Groot, 2004). We also predicted that interpreters would cope with the temporal constraints of the simultaneous condition better than translators and bilinguals. This would manifest in the quantitative analysis as shorter EVS (faster delivery of the paraphrases) in the simultaneous condition displayed by interpreters as compared to translators and bilinguals. Actually, we predicted that other groups, untrained in simultaneous interpreting, would wait longer and engage in quasi-consecutive interpreting (i.e. they would wait until the end of the sentence to start paraphrasing even if allowed to paraphrase simultaneously). Such a result would show that simultaneity is a real problem to participants who have not been trained in interpreting. When it comes to the delayed condition (where paraphrasing can start only after the stimulus sentence has been presented in full), we expected bilinguals to wait longer (to manifest longer latencies) before starting to paraphrase a given sentence than interpreters and translators. In this case, the disadvantage of bilinguals in the delayed condition would again stem from their smaller experience with reformulation.

2.1. Method

We asked the participants to perform intralingual paraphrasing of sentences in their mother tongue (Polish). We applied a mixed factorial design in the study: 3 (group: interpreters, translators, bilingual controls) x 3 (sentence type: semantic redundancy, oral discourse marker, false start) x 2 (condition: simultaneous, delayed). Group was a between-group factor and sentence type and condition were within-subject factors.

2.2. Participants

57 participants including 49 females and 8 males took part in the study. Their mean age was 27.44 (SD 5.75). 49 participants studied or graduated from linguistics/philology departments, 7 majored in other areas and one person did not have a university degree. The participants were divided into three groups based on their self-reported interpreting/translation experience: there were 22 interpreters and interpreting trainees with at least two years of interpreting experience, 23 translators and translation trainees with at least two years of translation experience and 12 bilingual controls with no experience in either interpreting or translation. Interpreting or translation experience is understood here as any exposure to interpreting or translation, including market or classroom experience. For the sake of simplicity, the first two groups will be shortly called interpreters and translators in this paper.

All participants underwent an introductory respeaking training in the form of a two-day (16-hour) workshop, conducted by experienced respeaker trainers. During the workshop, they were introduced to the fundamentals of respeaking, which included exercises on shadowing and multi-tasking, paraphrasing the text, dictating punctuation marks, working with their voice and intonation for respeaking. During the workshop, participants respoke several clips representing different TV genres, including speeches, chat shows and news. They

also tried interlingual respeaking from English into Polish. The participants were familiarised with the FAB Subtitler Live programme. They also created their own voice profiles on the Newton Dictate software for the Polish language manufactured by Newton Technologies. Finally, they also learnt about quality assessment in respeaking using the NER model (Romero Fresco & Pöchhacker, 2017; Romero-Fresco & Martínez Pérez, 2015).

2.3. Materials

We used 60 Polish sentences as materials in the present study. They were matched for length measured by the number of characters ($M=108$, $SD=6.87$) and the number of syllables ($M=40$, $SD=1.31$). The sentences were created on the basis of potential subtitles that would fit regular screen constraints (Díaz-Cintas & Remael, 2007), i.e. they were approximately 80 characters long to fit into two lines on the screen ($M=81.4$, $SD=2.48$). The potential subtitles, in turn, were created by manipulating sentences taken from Polish press articles on everyday topics. We turned the potential subtitles into longer experimental sentences by neutral padding or stalling, i.e. adding words that did not carry much meaning. Eventually, the stimuli sentences were approximately 30 percent longer than the potential subtitles. We created the stimuli sentences on the basis of potential subtitles because we wanted it to be possible to create professional two-line subtitles from these experimental sentences.

We included three types of manipulations in the sentences to create 20 sentences with semantic redundancies, 20 sentences with redundant oral discourse markers and 20 sentences with false starts, which were supposed to be deleted. In order to create the stimulus sentences with semantic redundancies one word from the potential subtitle was replaced in the stimulus sentence with a phrase carrying the same meaning, e.g. “alcohol-free” was replaced with “not containing any alcohol” or “medieval” with “originating from the medieval times”. The rationale behind this manipulation was that condensation of content is a frequently required

operation in various types of intra- or interlingual transfer (including interpreting, subtitling and respeaking). Stimulus sentences with oral discourse markers were created by including two stalling devices into each sentence. Stalling devices (Majewska-Tworek, 2014) are words or phrases used frequently in oral discourse which do not carry any meaning but prolong the utterance and give the speaker some time to find words to continue. They obviously disrupt the fluency of oral discourse. Some examples include: *właściwie* (actually), *znaczy* (I mean), *można powiedzieć* (one can say). False starts were created by including the following disfluencies into the sentences: adjectives were used with a wrong gender (Polish adjectives are gender-marked) and were then repeated with the right gender, or nouns were used in the wrong number or case (Polish has seven cases) and were then repeated with the right number or case. Table 1 presents sample stimulus sentences together with the potential subtitles they originated from.

[Insert Table 1 here]

All 60 sentences were recorded by a female native speaker of Polish at the approximate speed of 110 words per minute. Two sets of stimuli were created so that each participant listened to half of the sentences of each type in the simultaneous condition and half in the delayed condition. The order of the conditions and the sentences in the sets were counterbalanced across all participants.

2.4. Procedure

The participants were asked to perform intralingual paraphrasing of Polish sentences heard over headphones. As these were isolated sentences, there was no accompanying visual material. They first received instructions in Polish. We asked them to repeat the sentence they could hear so that it can potentially become a good respooken subtitle. We told them that in

order to achieve that goal they would have to paraphrase the sentence or shorten it by approximately one third. We asked them to introduce any necessary changes in the paraphrase but to strive at reproducing as much content as possible. We then presented the participants with an example stimulus sentence and a target respoken paraphrase. In the simultaneous block, they were asked to start respeaking as soon as they wanted and in the delayed block they were asked to wait until the end of the sentence to start respeaking. Each experimental block was preceded by a practice session with five sentences. The practice session could be repeated if needed. The sequence of the experimental blocks was counterbalanced across the participants. The whole experimental procedure took approximately 30 minutes. We recorded the participants' output in Audacity in order to obtain two synchronized audio tracks for further analysis.

2.5. Analysis

In order to analyse the results, we used linear mixed effects (LME) models via the lme4 package (Bates, 2013) within R (version 2.13.1) (Baayen, 2008; R Development Core Team, 2013). This type of analysis treats participants and items as random effects and replaces the traditional ANOVA F1 and F2 analyses. The participants' spoken output was manually transcribed and further analysed qualitatively and quantitatively. The qualitative analysis was based on the rating of semantic equivalence between the stimuli and the output as well as on the scoring of the paraphrasing accuracy. The quantitative analysis was based on the assessment of the temporal aspect of the output where EVS was calculated in the simultaneous condition and latency was analysed in the delayed condition.

2.5.1. Semantic equivalence

The quality of the paraphrased sentences understood as semantic equivalence between the source sentence and the paraphrased sentence was rated by a group of independent raters with the use of a specially designed rating tool (see Figure 1). This was an application displaying a source sentence (top window) and the transcribed paraphrased sentence (bottom window) along with the rating scale (as vertically aligned radio buttons) where the judge was able to perform the rating. The semantic equivalence of each sentence was rated by two raters on a scale from 0 to 6 with the following meaning: 0 – not equivalent at all, 1 – equivalent to a very small extent, 2 – equivalent to a small extent, 3 – equivalent to a medium extent, 4 – equivalent to a great extent, 5 – equivalent to a very great extent, 6 – identical). The raters were translation trainees so that they were familiar with various linguistic operations and restructuring entailed by the translation process. The inter-rater reliability was high (Kendall's coefficient of concordance $W=.676$, $p<.001$). The average rating was treated as a continuous variable indicating the quality of each sentence paraphrased by a given participant.

[Insert Figure 1 here]

2.5.2. Accuracy

Scoring was used to assess the accuracy of the paraphrased sentences with respect to changes expected of the participants. The paraphrases were given 1 point if redundancies (both oral discourse markers or both false starts) were omitted or 0 points if they were not removed or if they were removed partially. In sentences with semantic redundancies, 1 point was given if these semantic redundancies were replaced with shorter synonyms (e.g. to use the example presented above, when “not containing alcohol” was replaced with “alcohol-free”). If no such replacement was performed, the sentence was given 0 points.

2.5.3. EVS and latency

In order to compare groups on how they coped with the difficulty of paraphrasing in the simultaneous vs. delayed condition, we adopted two temporal indicators: EVS and latency. EVS was calculated as the time between the moment a given word was heard in the source sentence to the moment it was uttered by the participant in the paraphrased sentence. In line with Christoffels and de Groot (2004), we conducted these analyses by selecting every third paraphrased sentence for a given participant. We then time-aligned each of these selected paraphrased sentences with its source sentence. Next, we chose three content words in each of the source sentences and calculated the EVS for the corresponding three words in the paraphrased sentence. The averaged EVS (in ms) for each paraphrased sentence was next treated as a continuous variable in the analysis. Since the nature of the task in the delayed condition excluded the use of EVS, we opted for latency instead. Latency was measured (in ms) as the time between the end of the source sentence and the onset of the paraphrased sentence. The rationale was that the more difficult it was to paraphrase a given source sentence, the longer it took for the participants to start paraphrasing this sentence.

2.6. Results

2.6.1. Semantic equivalence

The LME model included participants as random factors and the following predictors: group, condition, sentence type and a two-way interaction between group and condition. Table 2 presents descriptive statistics for the three groups. There was no difference between bilinguals and interpreters ($b=-.301$, $t=-1.244$, $p=.21$), between interpreters and translators ($b=.057$, $t=0.29$, $p=.76$) and between translators and bilinguals ($b=.244$, $t=1.03$, $p=.3$).

[Insert Table 2 here]

There was a significant effect of condition ($b=.457$, $t=10.81$, $p<.001$). Contrary to our predictions, sentences paraphrased in the simultaneous condition ($M=4.24$, $SD=1.31$) received higher semantic equivalence rating than sentences paraphrased in the delayed condition ($M=3.79$, $SD=1.34$). Table 3 presents descriptive statistics for sentence types. There was a marginally significant difference between the rating for sentences with false starts and oral discourse markers ($b=-.083$, $t=-1.71$, $p=.08$) and between sentences with false starts and semantic redundancies ($b=.091$, $t=1.86$, $p=.06$), but there was no difference between sentences with discourse markers and semantic redundancies ($b=-.007$, $t=-.14$, $p>.05$).

[Insert Table 3 here]

There was a significant two-way interaction between bilinguals and interpreters and condition ($b=-.338$, $t=-3.04$, $p=.002$), between interpreters and translators and condition ($b=.385$, $t=4.31$, $p<.001$), but not between translators and bilinguals and condition ($b=-.046$, $t=-.43$, $p=.66$).

This suggests that the quality of the paraphrases was more affected by the simultaneity of the task in interpreters than in translators and bilinguals. As shown in Figure 2, the interpreters' performance in the simultaneous condition was rated significantly lower (less semantically equivalent) than the performance of translators and bilinguals.

[Insert Figure 2 here]

2.6.2. Accuracy

In order to analyse accuracy (1 or 0 points in scoring) we fitted a logistic regression model with participants as a random effect and the following predictors: group, condition, sentence type, a two-way interaction between group and sentence type and between condition and sentence type. Figure 3 presents mean accuracy for all conditions.

[Insert Figure 3 here]

Table 4 presents mean accuracy scores for the three experimental groups. We found no difference between bilinguals and interpreters ($b=.06$, $z=.16$, $p=.86$), between interpreters and translators ($b=-.03$, $z=.09$, $p=.92$) or between translators and bilinguals ($b=-.02$, $z=-.08$, $p=.93$).

[Insert Table 4 here]

Contrary to our predictions, we found no effect of condition ($b=.14$, $z=.89$, $p=.36$). The average accuracy score was .64 ($SD=.47$) for the simultaneous condition and .62 ($SD=.48$) for the delayed condition. The sentence type effect was reliable. We found a statistically significant difference between sentences with false starts and discourse markers ($b=.66$, $z=3.17$, $p=.001$), between sentences with discourse markers and semantic redundancies ($b=-2.16$, $z=-10.55$, $p=.001$) and between sentences with semantic redundancies and false starts ($b=1.503$, $z=7.97$, $p<.001$). As expected, the mean accuracy was the lowest for semantic redundancies ($M=.37$, $SD=.48$). It was higher for false starts ($M=.71$, $SD=.45$) and the highest for oral discourse markers ($M=.82$, $SD=.38$).

Additionally, there were a few reliable two-way interactions. Bilinguals and interpreters differed in the way how accurately they paraphrased sentences with false starts

and discourse markers ($b=1.21, z=2.26, p=.023$). This interaction was driven mainly by the fact that interpreters received low ratings for paraphrasing sentences with false starts. We also found a reliable interaction between bilinguals and interpreters in paraphrasing semantic redundancies and false starts ($b=-1.92, z=-4.008, p<.001$). The difference in difficulty between paraphrasing these two types of redundancies was greater for bilinguals than for interpreters. A similar interaction was found between interpreters and translators paraphrasing semantic redundancies and false starts ($b=1.06, z=2.42, p=.015$). We also found a marginally significant interaction between condition and discourse markers and semantic redundancies ($b=-.66, z=-1.68, p=.092$) and a reliable interaction between condition and semantic redundancies and false starts ($b=1.34, z=3.69, p<.001$). It seems that paraphrases of semantic redundancies gain accuracy when the condition changes from simultaneous to delayed while paraphrases of discourse markers and false starts lose accuracy when they have to be performed with a delay. No other two-way interactions were found significant.

2.6.3. EVS and latency

We trimmed the EVS data by deleting outlying observations above and below three standard deviations in each group (outliers constituted 1% of all data). The LME model included participants as random factors and group as a fixed predictor. We found no significant differences between the groups although interpreters ($M=3173$ ms, $SD=1464$) did manifest numerically shorter EVS than bilingual controls ($M=3247$ ms, $SD=1652$) and translators ($M=3260$ ms, $SD=1603$).

We also trimmed the latency data by deleting outlying observations above and below three standard deviations in each group (outliers constituted 3% of all data). We fitted a similar LME model with participants as random factors and group as a fixed predictor. We found reliable differences between interpreters and bilinguals ($b=-3662.30, t=-3.03, p<.01$)

and between translators and bilinguals ($b=-2522.67$, $t=-2.099$, $p=.04$). Interpreters did not significantly differ from translators. The mean latencies were as follows: 2140 ms (SD=2246) for interpreters, 3170 ms (SD=4073) for translators and 5763 ms (SD=6293) for bilinguals.

2.7. Discussion

In the present study we looked for advantages during paraphrasing for respeaking which might stem from the participants' previous experience in interpreting and translation. Contrary to our predictions, in the qualitative analysis of the output we found no reliable group effects. This means that none of the groups produced significantly more equivalent paraphrases across conditions. In fact, if looking just at the numerical differences, it was the semantic equivalence of the bilinguals' paraphrases that was rated the highest. Similarly, we found no group differences in the accuracy scores, which means that all groups performed the desired types of paraphrases equally efficiently. It seems that the interpreting and translation experience does not give interpreters or translators any straightforward advantage that would transfer directly into a qualitatively better performance in paraphrasing. This finding could be interpreted as evidence for the perception of paraphrasing and translation as two different tasks (Christoffels & de Groot, 2004) and this is why experienced translators and interpreters did not outperform bilinguals, who lacked translation/interpreting experience. However, this conclusion has to be treated with caution. Even though the results did not provide evidence for a straightforward advantage for any of the participant groups (and the similarity of translation and paraphrasing), the data obtained revealed several interesting interactions, indicating that the groups indeed differed in the way they handled the task across time constraints and sentence types. This, in turn, suggests that the participants' previous experience and training did surface in the task.

As far as semantic equivalence of the stimuli and the output is concerned, we found reliable interactions between bilinguals and interpreters and condition as well as between interpreters and translators and condition, but not between translators and bilinguals and condition. This pattern of results suggests that interpreters were more affected by the simultaneity of the task than translators and bilinguals and that this effect was negative. In the simultaneous condition interpreters' output was rated as significantly less semantically equivalent than the output of translators and bilinguals. This could be taken as evidence that, contrary to our predictions, interpreters coped with the simultaneity of the task worse than the other groups. However, such a conclusion is justified only to a certain extent.

The interpretation of this pattern becomes clearer when combined with accuracy data (where participants were scored for eliminating redundancies as instructed during the initial respeaking training). The analysis of the accuracy revealed that interpreters were better than translators and bilinguals at eliminating semantic redundancies. It seems that the interpreting experience makes it easier for interpreters to deal with semantic paraphrasing, which is in line with previous research pointing to “expressive flexibility” (Russo & Salvador, 2004, p. 421) as an important quality in interpreting trainees. Further, it can be hypothesized that more pronounced reformulation of semantic redundancies performed by interpreters might have influenced the raters when judging semantic equivalence of the sentences. Even though the raters had been instructed to reward high semantic equivalence (between the stimuli and the paraphrases) rather than the verbatim rendition of the stimulus sentences, it is possible that verbatim renditions received higher rating in the end, especially as the highest mark (6) was described as the one to be given to identical sentences. In this case, the word “identical” could have been understood as identical in form and not in meaning. That would automatically lead to a disadvantage of those sentences that included greater reformulation of semantic redundancies (hence their verbatim similarity was smaller), which is what interpreters were

best at. If this explanation is legitimate, it could be taken as evidence that the interpreters are better than other groups at eliminating redundant information in order to form short, concise respoken output which could potentially be easily turned into well-formed subtitles.

As far as group differences are concerned, we analysed the participants' output not only qualitatively but also quantitatively, in order to check which participant groups can best cope with time constraints of paraphrasing for respeaking. In real-life respeaking, the delay with which subtitles appear on the screen, which is directly influenced by how fast the paraphrases are delivered, is an important factor in quality assessment (Ofcom 2015, Robert & Remael 2017). Contrary to our predictions, we found no reliable differences between interpreters, translators and bilinguals on the EVS measure. This result can be indirectly explained by the results obtained by Anderson (1994) and Christoffels and de Groot (2004). It was expected in both studies that EVS would be longer in interpreting as compared to paraphrasing because the former is a more complex and difficult task. However, Anderson (1994) found no difference in EVS between interpreting and paraphrasing performed by conference interpreters, while Christoffels and de Groot (2004) found larger EVS for paraphrasing than for interpreting. The authors of the last study concluded that paraphrasing was in fact more difficult than interpreting and that "the paraphrasing task should not be regarded as the monolingual equivalent of simultaneous interpreting" (Christoffels & de Groot, 2004, p. 235). Therefore, the interpreter advantage might not be directly transferrable to the paraphrasing task, which could explain why interpreters did not outperform the non-interpreting groups on the EVS measure.

Furthermore, we predicted that interpreters would manifest shorter latencies in the delayed condition than translators and bilinguals, which would reflect their faster processing and smaller cognitive load. As predicted, interpreters started their paraphrases the fastest, while bilinguals waited the longest. We found that both interpreters and translators differed

from bilinguals, but interpreters did not significantly differ from translators. Since interpreters have to regularly deal with time constraints in their interpreting practice, it is understandable that they would be ready to paraphrase sooner after hearing the whole sentence than bilinguals. But why did translators outperform the bilingual controls on that measure? In fact, many experienced translators subject themselves to time constraints in their work to gain efficiency. Translators might engage in subvocal sight translation when writing their target text, especially when they are efficient touch-typists, and only then do they come back to the text to revise their output. Although it might serve as a tentative explanation, this suggestion is obviously rather speculative and requires empirical investigation.

Because interpreters and translators regularly engage in restructuring and rephrasing the source target content when producing the target text (Chmiel, 2015; Gile, 2009; Liontou, 2012), we expected some differences in the way the experimental groups would handle different types of paraphrases. In general, we found only marginal differences between false starts and the other two types of sentence types and no difference between discourse markers and semantic redundancies. Thus, false starts turned out to be the easiest to paraphrase (i.e. eliminate). We found no reliable group by sentence type interactions, suggesting no difference between interpreters, translators and bilinguals in the way they paraphrase various types of sentences. The findings for accuracy scores (where elimination of redundancies was awarded) were more consistent with our predictions. All sentence types differed significantly from one another. Semantic redundancies were the most difficult to eliminate, while oral discourse markers were the easiest.

We predicted that simultaneous paraphrasing will pose more difficulty to the participants than delayed paraphrasing since the latter does not entail so much multitasking. We found the main effect of condition on the semantic equivalence measure. However, it was inconsistent with our predictions. Simultaneously performed paraphrases received higher

average ratings than delayed paraphrases. It seems that the concurrent processing facilitated paraphrasing, probably due to lower memory load. These results are at a variance with the findings by Christoffels and de Groot (2004), who found a significant effect of condition in paraphrasing: the quality of performance was better in the delayed condition than in the simultaneous one.

Contrary to our predictions, the analysis of the accuracy scores brought no reliable effect of condition. Paraphrases were equally accurate in the simultaneous and delayed condition. We found no interactions between group and condition, but we did find two interactions between condition and sentence type. Semantic redundancies were more accurately paraphrased in the delayed condition while discourse markers and false starts were more accurately paraphrased in the simultaneous condition. The last two paraphrases involve omissions and thus can be easily tackled when processing simultaneously. On the other hand, paraphrasing of semantic redundancies requires lexical retrieval and can benefit from lesser time constraints.

Admittedly, the tested sample size is a limitation of our study and it might have influenced the results to some extent. These limitations stem from the very nature of the study. For the purposes of this research we could only recruit participants among those individuals who completed an introductory respeaking training. Given that there was an uneven number of interpreters, translators and bilinguals among the participants, these three groups were disproportionately represented in our study. This, along with the fact that interpreters and translators are not numerous anyway, has posed a direct limitation on the sample sizes within our study.

3. Conclusions

The aim of the present study was to look for advantages stemming from previous training and experience which would surface during paraphrasing for respeaking. Thereby we wanted to check whether interpreters, translators or bilinguals would make better candidates for respeakers. We assessed their paraphrasing performance qualitatively and quantitatively. Additionally, we tested which types of the required manipulations are most challenging during paraphrasing. To this end, we asked the participants to paraphrase different types of sentences in a simultaneous and in a delayed condition. The results did not reveal any straightforward group advantages but showed that interpreters are better than others at eliminating redundant semantic material during paraphrasing (as indicated by accuracy scores). This can be seen as interpreters' advantage leading them to produce more concise output, which can be easily transformed into well-formed respoken subtitles.

Our analysis also showed that during paraphrasing the elimination of semantic redundancies is most challenging to all participants and requires more time than omitting false starts or oral discourse markers. These results can be taken as evidence that during respeaking training particular emphasis should be placed on practicing semantic reformulations under time constraints.

Interestingly enough, since our data did not reveal any clear and straightforward advantages of any of the participant groups in paraphrasing for respeaking, these results could serve as an encouragement for anybody willing to become a respeaker. Simultaneously, some of the differences we did find can help tailor future respeaking training to the specific needs of the participants. Once this is the case and once respeaking training attracts many more participants in the future, this study could be replicated with greater sample sizes and in other languages in order to look for more subtle differences. However, more importantly, when respeaking training becomes attractive to a greater number of people, that will mean that the

demand for respeakers has increased. This, in turn, will be vital, as it will mean that the needs of all members of the society are more likely to be catered for.

Acknowledgements

This work was funded by the National Science Centre Poland based on decision no. DEC 2013/11/B/HS2/02762. We would like to thank Łukasz Brocki for programming the rating tool.

References

- Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap: Empirical research in simultaneous interpretation* (pp. 101-120). Amsterdam: Benjamins.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*: Cambridge University Press.
- Barik, H. C. (1973). Simultaneous interpretation: Temporal and quantitative data. *Language and Speech*, 16(3), 237-270.
- Bartłomiejczyk, M. (2015). Wprowadzenie tłumaczenia symultanicznego. In A. Chmiel & P. Janikowski (Eds.), *Dydaktyka tłumaczenia ustnego* (pp. 207-226). Katowice: Stowarzyszenie Inicjatyw Wydawniczych.
- Bates, D. (2013, 2013). *Linear mixed model implementation in lme4*. University of Wisconsin.
- Chmiel, A. (2015). Przetwarzanie w tłumaczeniu symultanicznym. In A. Chmiel & P. Janikowski (Eds.), *Dydaktyka tłumaczenia ustnego* (pp. 227-247). Katowice: Stowarzyszenie Inicjatyw Wydawniczych.

- Christoffels, I. K., & de Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition*, 7(3), 227-240. doi:10.1017/s1366728904001609
- de Korte, T. (2006). Live inter-lingual subtitling in the Netherlands: historical background and current practice. *InTRAlinea*.
- Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. Manchester: St. Jerome.
- Eugeni, C. (2007). Il rispeakeraggio televisivo per sordi: per una sottotitolazione mirata del TG. *InTRAlinea*, 8.
- Eugeni, C. (2008a). Respeaking the TV for the deaf: for a real special needs-oriented subtitling. *Studies in English Language and Literature*(21), 37-48.
- Eugeni, C. (2008b). A sociolinguistic approach to real-time subtitling: Respeaking vs. shadowing and simultaneous interpreting. *English in International Deaf Communication*, 72, 357-382.
- Gambier, Y. (2006). Subtitling. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 258-263): Elsevier.
- Gile, D. (2009). *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia: Benjamins.
- Gillies, A. (2001). *Tłumaczenie ustne. Poradnik dla studentów*. Kraków: Tertium.
- Gottlieb, H. (1998). Subtitling. In M. Baker (Ed.), *The Routledge encyclopedia of translation studies* (pp. 244-248). Manchester: St. Jerome.
- Ivarsson, J., & Carroll, M. (1998). *Subtitling*. Simrishamn: TransEdit HB.
- Jakobson, R. (1959). On linguistic aspects of translation. In R. A. Brower (Ed.), *On translation* (pp. 232-239). Cambridge, MA: Harvard University Press.
- Jones, R. (2002). *Conference interpreting explained*. Manchester: St. Jerome.

- Lamberger-Felber, H. (2001). Text-oriented research into interpreting: Examples from a case-study. *Hermes*, 26, 39-63.
- Lambourne, A. (2006). Subtitle respeaking: A new skill for a new age. *InTRALinea*.
- Lee, T.-H. (2002). Ear voice span in English into Korean simultaneous interpretation. *Meta*, 47(4), 596-606.
- Liontou, K. (2012). *Anticipation in German to Greek simultaneous interpreting*. (Unpublished doctoral thesis), University of Vienna.
- Luyckx, B., Delbeke, T., Van Waes, L., Leijten, M., & Remael, A. (2010). Live subtitling with speech recognition causes and consequences of text reduction. *Across Languages and Cultures*, 14, 15-46.
- Majewska-Tworek, A. (2014). *Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny*. Wrocław: Quaestio.
- Marsh, A. (2006). Respeaking for the BBC. *InTRALinea*.
- Mikul, C. (2014). *Caption Quality: Approaches to standards and measurement*. Sidney: Media Access Australia.
- Moser, B. (1983). Testing interpreting aptitude. In W. Wilss & G. Thome (Eds.), *Translation theory and its implemetation in the teaching of translating and interpreting* (pp. 318-325). Saarbrucken: Association Internationale de Linguistique Appliquée.
- Neves, J. (2008). Ten fallacies about subtitling for the d/Deaf and the hard of hearing. *Journal of Specialised Translation*, 10, 128–143.
- Ofcom. (2015). Measuring live subtitling quality. Results from the fourth sampling exercise. Retrieved from https://www.ofcom.org.uk/_data/assets/pdf_file/0011/41114/qos_4th_report.pdf

- Paradis, M. (2007). The neurofunctional components of the bilingual cognitive system. In I. Kecskes & L. Albertazzi (Eds.), *Cognitive aspects of bilingualism* (pp. 3-28): Springer.
- Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a Model for Multitasking. *Developmental Review, 35*, 113-124. doi:10.1016/j.dr.2014.12.003
- R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Robert, I., & Remael, A. (2017). Assessing quality in live interlingual subtitling: a new challenge. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 14*.
- Robson, G. D. (2004). *The closed-captioning handbook*. Amsterdam: Elsevier.
- Romero Fresco, P. (in press). Respeaking: subtitling through speech recognition. In L. Pérez-González (Ed.), *The Routledge Handbook of Audiovisual Translation Studies*. London/New York: Routledge.
- Romero Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR model. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 14*.
- Romero-Fresco, P. (2009). More haste less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics, 6*(1), 109-133.
- Romero-Fresco, P. (2011). *Subtitling through speech recognition: Respeaking*. Manchester: St. Jerome.
- Romero-Fresco, P. (2012). Respeaking in translator training curricula: Present and future prospects. *The Interpreter and Translator Trainer, 6*(1), 91-112.
- Romero-Fresco, P., & Martínez Pérez, J. (2015). Accuracy rate in live subtitling: The NER model. In J. Díaz Cintas & R. Baños Pinero (Eds.), *Audiovisual translation in a global*

- context. Mapping an ever-changing landscape* (pp. 28-50). London: Palgrave Macmillan.
- Russo, M., & Salvador, P. (2004). Aptitude to interpreting: Preliminary results of a testing methodology based on paraphrase. *Meta: Journal des traducteurs*, 49(2), 409–432.
- Setton, R., & Dawrant, A. (2016). *Conference Interpreting - A Trainer's Guide*. Amsterdam: Benjamins.
- Spalek, K., Hoshino, N., Wu, Y. J., Damian, M., & Thierry, G. (2014). Speaking two languages at once: unconscious native word form access in second language production. *Cognition*, 133(1), 226-231. doi:10.1016/j.cognition.2014.06.016
- Szarkowska, A., Krejtz, K., Dutka, Ł., & Pilipczuk, O. (under review). Are interpreters better respeakers?
- Szarkowska, A., & Laskowska, M. (2015). Poland – a voice-over country no more? A report on an online survey on subtitling preferences among Polish hearing and hearing-impaired viewers. In Ł. Bogucki & M. Deckert (Eds.), *Assessing audiovisual translation* (pp. 179–197). Bern: Peter Lang.
- Timarová, S., Dragsted, B., & Hansen, I. G. (2011). Time lag in translation and interpreting: A methodological exploration. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies*. (pp. 121-146). Amsterdam: John Benjamins.
- Tzou, Y.-Z., Vaid, J., & Chen, H.-C. (2016). Does formal training in translation/interpreting affect translation strategy? Evidence from idiom translation. *Bilingualism: Language and Cognition*, 1-10. doi:10.1017/s1366728915000929
- Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers of Psychology*, 3, 174. doi:10.3389/fpsyg.2012.00174

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT journal*, 42(4), 272-281.

Whyatt, B., Stachowiak, K., & Kajzer-Wietrzny, M. (2016). Similar and different: cognitive rhythm and effort in translation and paraphrasing. *Poznań Studies in Contemporary Linguistics*, 52(2), 175-208.

Zethsen, K. K. (2009). Intralingual translation: An attempt at description. *Meta: Journal des traducteurs*, 54(4), 795-812. doi:10.7202/038904ar