Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues

Ewa Gumul
University of Silesia, Poland
ewa.gumul@us.edu.pl
https://orcid.org/0000-0003-0859-6067

Abstract

Retrospection in simultaneous interpreting research often uses either (1) transcripts of the source text or (2) recordings of the target texts as retrieval cues. This study tested their influence on the informativeness and the accuracy of retrospective reports in addition to the verbosity of the interpreters. The study also set out to examine the participants' perception of the cueing stimuli. The participants in the study were 36 trainee interpreters, who took part in an experiment consisting of interpreting a speech simultaneously and performing self-retrospection immediately after the interpreting task. They were divided into two groups: group A, which was exposed to a source-text transcript as the retrieval cue during retrospection, and group B, which relied on target-text recordings. The results suggest that the differences between the two retrieval cues may be less marked than is generally assumed: the three parameters of verbosity, informativeness and accuracy do not display statistically significant differences between the two experimental conditions. However, some discrepancies can be observed as far as the participants' perception of the cueing stimuli is concerned. The results also suggest that idiosyncratic reporting styles might have more impact on the retrospective reports than the type of cueing administered.

Keywords: simultaneous interpreting; trainee interpreters; process research; retrospective protocols; retrieval cues

1. Introduction

Retrospection is a process-oriented method that may be used to investigate the underlying cognitive processes involved in the interpreting task. It is hoped that retrospection provides some access – unfortunately, mostly indirect – to the mental processes of interpreters. Referred to in the Translation and Interpreting (T&I) literature as *retrospection* (e.g., Englund Dimitrova & Tiselius, 2009; Ivanova, 2000), *backward* or *delayed introspection* (Seeber, 2015) or *stimulated recall* (Leanza, 2005; Russell & Winston, 2014), this method consists of eliciting verbal reports from the participants after they complete interpreting tasks. A clear advantage over concurrent verbalizations (think-aloud techniques) is that retrospection is less invasive and does not interfere with the translation task, as there is no interaction between the research process and the interpreting process being analysed. As Buchweitz and Alves (2006)

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

observe, retrospective protocols "provide a window into human cognitive processes without interfering with the ongoing tasks" (p. 243).

Adapting this research method from psychology (Ericsson & Simon, 1980, 1993) to T&I research is not free of methodological challenges, though. This is because of the nature of the interpreting task, which imposes considerable cognitive load and also tends to be much longer than the tasks envisioned by Ericsson and Simon (1993) as being suitable for retrospection. Both of these factors may impede the efficient recall of underlying cognitive processes once the task is completed. The validity of retrospective research in this field has been questioned, and many researchers (e.g., Bartłomiejczyk, 2007; Englund Dimitrova & Tiselius, 2009, 2014; Hansen, 2005; Hild, 2015) point to the potential confounding factors and the inherent weaknesses of this method, such as

- the incompleteness of reports;
- the limited representational validity and non-veridicality of such verbal data;
- the non-feasibility of tapping into the short-term memory due to the posteriority of the retrospective task; and
- contamination with other contents of long-term memory (LTM), which may give rise to false memories being installed or the fabrication of mental events.[1]

Obtaining inferences and speculations about cognitive processes, rather than recall of the actual thought episodes (Ericsson, 2003; Ericsson & Simon, 1980, 1993), is also likely.

Nevertheless, retrospective protocols remain one of the few process methods available for simultaneous interpreting (SI) research. Notwithstanding its inherent limitations, retrospection can yield interesting and valuable data about the cognitive processes underlying interpreting, as construed by the participants, provided that the above limitations are addressed and reflected accordingly in the research designs. There is a clear need to work out optimal procedures for the use of this method in interpreting in order to ensure greater validity. Currently, the major obstacle to using retrospection is the limited comparability of individual studies, owing to their divergent research designs. Certain research design features can potentially either enhance or undermine the validity of this method, such as the choice of immediacy condition (immediate vs postponed retrospection), cueing methods (e.g., source-text (ST) transcript or target-text (TT) recording), the extent of the researcher's intervention (from self-retrospection to retrospective interviews), prior training (or the lack of it) and the instructions to the participants, explicitness and transparency of data-handling procedures, and any combination with other research methods (e.g., product-oriented research).

Owing to its posteriority to the interpreting task, retrospection entails tapping into the LTM store of interpreters or, more precisely, accessing what remains in LTM, as the process of memory decay is bound to have begun once the task is over. Therefore, eliciting retrospective data requires the use of effective retrieval cues that stimulate memory. These can be transcripts of the original speech (Englund Dimitrova & Tiselius, 2009, 2014; Ivanova, 1999, 2000; Tiselius & Jenset, 2011), a recording of the source speech (Shamy & de Pedro Ricoy, 2017), recordings of the interpreting outputs, usually with the ST audible in the background (Bartłomiejczyk, 2006; Gumul, 2006, 2017; Mead, 2002; Napier, 2004; Russell & Winston,

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

2014) or double-cueing that consists of providing both ST transcripts and TT recordings (Chang & Schallert, 2007; Tang, 2018; Vik-Tuovinen, 2002). The participants could be asked to perform self-retrospection, with no intervention on the part of the researcher (e.g., Bartłomiejczyk, 2006; Gumul, 2006, 2017). Conversely, the retrospective session may take the form of an interview (e.g., Leanza, 2005; Napier, 2004; Russell & Winston, 2014). In some cases, self-retrospection is either supplemented by additional questions asked by the researchers or followed by a retrospective interview (e.g., Ivanova, 2000; Tang, 2018; Vik-Tuovinen, 2002). Such questions are also intended to stimulate the memory of the participants.

The choice of divergent cueing methods in these studies is neither accidental nor random. Each researcher provides a rationale for adopting a given cue to elicit the interpreters' verbal protocols in their studies and points to the weaknesses of the other options. This is in fact probably the most contentious issue in designing a study that relies on retrospection, because each type of cueing entails the risk of distorting the data obtained and there is no consensus on which is a better solution.

Some researchers deliberately opt for cueing via an ST (Englund Dimitrova & Tiselius, 2009; Ivanova, 2000), arguing that cueing via the participants' own production (i.e., the TT recording) is not methodologically sound because it is more likely to create new cognitive processes rather than stimulate the recall of those experienced during the task of interpreting. Confronted by their own performance – which is not always satisfactory – interpreters inevitably tend to explain, justify their decisions and evaluate, which does not necessarily reflect what they were thinking about while interpreting (Englund Dimitrova & Tiselius, 2009; Hansen, 2005). This problem could be dealt with to some extent at different stages in the research process: first, by instructing the participants to refrain from making such comments; second, by effective coding; and, finally, by simply not taking such reports into account. However, it is not always feasible to separate the verbalizations that probe the original processes from those made *post factum*. Inevitably, we will always end up with some mixture of these in the data collected retrospectively, and cueing via product might aggravate that predicament.

The conviction of the superiority of cueing via ST is also rooted in Ericsson and Simon's (1993), idea that "recall is more likely to be successful if the cueing stimulus is encoded in the same way at recall as it was at the original presentation"(p. 117). In accordance with that, retrospection after interpreting is considered to be best cued by an ST, as was the interpreting itself. However, Ericsson and Simon were proposing a methodology for tasks not exceeding 15 seconds, which cannot be compared with the cognitive complexity of SI. Therefore, cueing via an ST transcript in SI is likely to stimulate memory to a much lesser extent than a TT recording, so using an ST will probably limit the completeness of the recall and result in inaccuracies.

Cueing via an ST entails another risk of distorting the data: not confronted with the evidence of their performance, the participants might possibly be more likely to succumb to the temptation of colouring their actual performance and report the process the way they would like it to have been. The issues of social desirability and impression management – tailoring one's behaviour in an attempt to influence what others think of them – highlighted in the

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

works on the use of verbal reports in consumer behaviour research (see, e.g., Büttner & Silberer, 2008) may also be applicable to T&I to a certain extent. Fully aware of the quality criteria and the clients' expectations, the interpreters might also verbalize in their protocols what they want the process to appear to have been rather than what it really was.

In brief, none of cueing methods is free from distortion, and most of the inherent constraints acknowledged for one retrieval cue would certainly also affect the other type of cueing to some extent. Cueing via an ST transcript would never be entirely free from inferences and *post factum* conclusions. Nor would cueing via a TT recording invariably render complete and accurate reports that bear no sign of impression management. The difference might not be clear-cut, as there is no ideal solution as far as the choice of cueing stimuli is concerned, and so this is probably one of the most contentious issues in retrospective research designs. Therefore, this article aims to contribute to improving methodological accuracy and transparency by testing the influence of two different ways of retrieval cueing (ST transcript and TT recording) on some aspects of verbal reporting empirically, as these are the two most frequently employed types of cueing. These are the research questions:

1. Does the type of cueing have any influence on the verbosity of the participants?
2. Does the type of cueing have any influence on the informativity of the participants' reports?
3. Is the discrepancy between product and process data regarding the use of strategies significantly lower when cueing is done via TT recording?
4. Is the discrepancy between product and process data regarding processing problems significantly lower when cueing is done via TT recording?
5. How do trainee interpreters perceive the efficiency of a given cueing stimulus?
6. To what extent do trainee interpreters perceive a given cueing stimulus as distorting their actual memories?

2. Research methods

This study applies a between-group design, where the performance of two groups is compared according to their reaction to different cueing stimuli. The study can be described as an independent measures experiment, as each of the participants is asked to complete only the task specific to the group to which they were assigned. The rationale behind choosing this type of experimental design rather than a repeated measures design was to avoid such confounding variables as carry-over effect or order effect. In this experiment, Group A was exposed to the ST transcript during the retrospective session, whereas members of Group B relied on the TT recordings of their own outputs.

Type of cueing serves as an independent variable in the experiment, whereas the informativeness and accuracy of reports and the verbosity of the participants are used as dependent variables.[2] The verbosity of the participants (the amount of talk) is measured as:

- the number of comments verbalized by each participant;
- the number of words per protocol; and
- the duration of the entire protocol.

In turn, the informativeness of the reports is measured in a concentration rate of four aspects of SI processing verbalized by the participants: (1) the problem, (2) the reported source of the problem, (3) the strategy and (4) the reason for adopting that strategy.

Process (retrospective protocols) and product data (TT recordings and transcripts) were contrasted with a view to testing the accuracy of the reports. The parameter of accuracy was measured using two types of indicator present in the product: (a) evidence of the use of strategies and (b) problem indicators (that are potentially indicative of increased cognitive effort) in the form of three types of disfluencies: hesitation markers, unfilled anomalous pauses exceeding 2 s and false starts.[3] In the first case, the strategies verbalized by the participants in the retrospective protocols are compared with the actual textual solutions in order to verify whether cueing via a TT (i.e., the participants' own production) – where they have direct access to the adopted strategies – guarantees significantly less discrepancy between their reports and product evidence than in the case of cueing via an ST transcript. The retrospective comments reporting processing problems are also compared with problem indicators in the product, that is, the TTs, in order to establish whether there is significantly less disparity between retrospective reports and product evidence when cueing is done using TT recording.

The test in this study was supplemented by survey research: a questionnaire was used to test the participants' perception of a given cueing stimulus and its influence on the retrospection task. The questionnaire contained seven items: three open questions and four multiple-choice questions, which also allowed the participants to provide comments. There were two versions of the questionnaire, depending on the experimental group (see Table 2). Therefore, for instance, group A, which used an ST transcript as the cue, was asked to what extent they found reading the ST transcript useful in stimulating memory during the retrospection and about the potential usefulness of another type of cue – the TT recording. In the case of group B these particular questions were reversed.

The participants in the study were 36 advanced interpreting students from the University of Silesia in Poland. Out of 41 advanced students taking a course of English–Polish simultaneous interpreting within the Translation and Interpreting programme at the University of Silesia, 36 volunteered to participate in the experiment and were willing to sign the informed consent. In order to avoid power relationships, none of the participants was a student of the researcher at the time of conducting the experiment. The participants formed a homogenous group of native Polish speakers with English as their B language. At the time of the experiment, each of them had completed at least 60 hours of prior training in simultaneous interpreting and the same amount of time training in other modes: consecutive and sight translation. There were 14 male and 22 female participants, with ages ranging from 21 to 25 ($M = 22.88$, $SD = 1.06$).

Owing to the limited availability of the participants and a need for a sample exceeding 16 in each of the experimental groups in order to make inferential statistical analysis possible,

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

random sampling was not possible. However, the participants were randomly assigned to the two experimental groups using a random number table. The one used in the study comes from Brzeziński (2008, p. 247). The researcher selected from a randomly chosen page the numbers equal to or smaller than $N - 1$, that is, 35. The participants who had been assigned the numbers appearing as the first 18 on this page were allocated to group A; the remaining participants formed group B.

The ST used in this study is Pamela Meyer's talk *How to spot a liar*.[4] In an effort to comply with the requirement of short task duration advocated for research adopting retrospective methods, only the fragment of the first 8 m and 43 s was used in the experiment. This segment of the text is 1,442 words long. The average speed of delivery of the ST is 170 wpm, which, by professional standards, is far above what is considered comfortable for simultaneous interpreters (see Barghout et al., 2015; Li, 2010). This text was selected for its high density of information, high speed of delivery and the presence of other numerous potential problem triggers, such as numbers, enumerations and cultural references.

All of these features were expected to induce increased cognitive load and prompt the participants to use coping or preventive tactics (strategies) when faced with processing problems, on which they might later possibly report during the retrospective session. Therefore, the difficulties envisaged in the source speech suggested that the potential lack of verbosity of the participants or the low informativeness in their reports would rather be due to reasons other than a lack of processing problems. As emphasized by Englund Dimitrova and Tiselius (2009), "a large number of problems in the process will presumably lead to a larger number of reports" (p. 115), which it was hoped would make the sample of data obtained in this study more representative.

This is also one of the reasons behind the choice of trainee interpreters as participants in this study – apart from their availability and the criterion of homogeneity of the group. Novices and trainee interpreters are believed to provide more extensive reports of the processing problems (see Ivanova, 2000) because the strategies have not undergone automatization to the same extent as in the case of professional interpreters with considerable experience.

The experiment was recorded in a standard teaching laboratory for simultaneous interpreting in the Faculty of Humanities at the University of Silesia in 2019. Prior to the interpreting task, the participants received a briefing on the subject-matter and the pragmatic setting of the speech. In order to avoid unnatural performance and white-coat effects, the actual aim of the study was not released. The participants were simply briefed that the material would be used in a research project on simultaneous interpreting of trainee interpreters. They were also informed of ethical issues: their anonymity and the possibility of withdrawing from the experiment at any stage of the procedure. All of the participants signed an informed consent form. Depending on the availability of the participants, the experiment was performed either individually or in groups ranging from two to four interpreters, each working on their own in a separate booth. The researcher was present in the laboratory, but did not interfere in any way in the retrospection of the participants.

Since the recency of a task is crucial in the retrospective procedure, the task was undertaken immediately following the interpreting task. There was only a short pause between the end

of the interpreting and the retrospective session, during which the participants received instructions on how to retrospect. The participants were asked to perform autonomous retrospection, which means that the only type of cueing was either the ST transcript or the TT recording, depending on the experimental group. Apart from being given initial instructions, during the retrospection the participants were neither asked any additional questions nor encouraged to make more comments. Additional questions might have drawn the participants' attention to unconscious processes and induced them to comment on them even in the absence of actual memories. There was no time limit for retrospection and the participants were free to spend as much time on it as they considered necessary.

Following the interpreting task, the participants in experimental group A were asked to read the ST transcript and verbalize any comments regarding their experienced and conscious decisions, that is, those they were aware of, taken at the time of interpreting. In turn, each participant in group B was asked to listen to the TT recording of their own output and, as with group A, report on processing problems they experienced during interpreting and their decisions. The procedure adopted in group B assumed they would stop the TT recording each time they wanted to make a retrospective comment. Their verbalizations were recorded on separate equipment (a portable recording device placed in each booth). Both groups were instructed to refrain from making remarks which occurred to them *post factum* only while reading the ST transcript or listening to the TT recording.

Both their interpreting outputs and the retrospective protocols were subsequently transcribed and coded. However, the analysis was not done by relying solely on the transcripts. Transcribed versions were regarded as essential for coding purposes, but the data were analysed using both the recordings and the transcripts. For the retrospective protocols the edited transcription was used with normalized orthography and punctuation. The interpreting outputs were transcribed verbatim, including all types of disfluency: pauses, false starts, hesitation markers, repetitions and unintelligible words.

The participants were free to choose the language of their comments or even to express themselves in a mixture of two languages, and some of them did. The vast majority of the retrospective comments were made in Polish, so the versions in this article are my own translations of the original verbalizations. In all the back-translations provided, care has been taken to follow closely the wording, register and form of the original comment.

The retrospective protocols were coded using the following coding scheme:

[**IP**]   to mark the identified problem
[**SP**]   whenever the participant reported on the source of the problem
[**AS**]   to designate the adopted strategy[5]
[**RS**]   for cases when the participant specified the reason for adopting a given strategy.

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

The transcripts of their interpreting outputs were coded in the following way:

[**HM**]   for hesitation markers
[**P**]   for anomalous pauses exceeding two seconds
[**FS**]   for false starts
[**S**]   for detectable strategies in the product.

All the stages of the study, from the phase of conducting it to transcription, coding and analysis of data, were performed by the author, the sole researcher in this study. The intention behind conducting the experiment in person and also performing the transcription personally was to facilitate the analysis of data and to gain more insights into the material obtained.

3. Analysis and discussion of results

3.1 Verbosity of the participants

The first research question concerned the extent of the influence of the type of cueing on the verbosity of the participants (i.e., the amount of talk they produced). It was hypothesized that because of the inherent limitations of cueing via an ST transcript, it might not be as effective in stimulating the memory and the interpreters would be less verbose in their reports. The parameter of verbosity was measured by participant in the number of comments, the number of words per comment and the duration of the entire protocol. Table 1 shows the descriptive statistics for the entire group, including the frequencies, means and standard deviation.

Next, the raw numerical data calculated for each of the participants were used to perform inferential statistics. Measured by the number of comments, the descriptive statistics suggest that the difference is marginal, and the inferential statistical analysis confirms that. As the results of the Shapiro Wilk test showed a normal distribution only in group A, the assumptions of a parametric one-way ANOVA test were not satisfied. Therefore, the between-group comparison had to be performed with the aid of a non-parametric test of Kruskal-Wallis, which showed that the difference was not statistically significant ($H = 0.703$, $p = 0.4$). The same test was applied to verify whether the difference in word count for each of the retrospective protocols between group A and B was statistically significant. Although the difference in terms of the mean value (μ) appears to be more noticeable than in the case of the number of comments – there are almost 20 per cent more words in the reports of group B – the results of the statistical analysis show that the difference is in fact not statistically significant ($H = 0.38$, $p = 0.53$). Therefore, the answer to the first research question is negative, since the type of cueing does not seem to exert a significant influence on the verbosity of the participants. The results imply that even though cueing via an ST transcript might not trigger the memory to the same extent as cueing via TT recordings, apparently it does not translate into the brevity of the reports as compared with the other cueing conditions. Table 1 shows the mean verbosity of the participants measured in number of comments, number of words per protocol and its duration.

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

Table 1. The mean verbosity of the participants measured in number of comments, number of words per protocol and its duration

	Cueing							
	ST transcript (A)				TT recording (B)			
	Total	Range	M (μ)	SD (σ)	Total	Range	M (μ)	SD (σ)
# comments	195	1–32	10.83	7.61	175	4–30	9.72	5.66
# words	5,890	7–583	327.22	179.12	7163	73–981	397.94	230.42
Length	5:57:08	7:39–37:07	19:50	–	8:01:23	15:15–39:28	26:45	–

The last parameter of the length of protocols was not subjected to statistical analysis since the clear differences in the duration of the retrospective sessions between group A and group B are simply due to the fact that reading an ST transcript usually takes less time than listening to the TT recording and having to stop it in order to make verbalizations. Judging by the length of some of the retrospective sessions in group A, some of the participants apparently scanned the text in search of the segment they remembered as having posed some difficulty and/or required some strategy. Moreover, the overall length of the retrospective session specified in Table 1 also includes periods of silence between comments. Therefore, the difference in duration should not be taken as a sign of different levels of verbosity.

One more conclusion on the level of verbosity can be drawn from the data. The large standard deviations for both the number of words and for the comments per participant point to substantial variations between participants. A qualitative analysis also shows that scarce reports of some participants contrast with abundant verbalizations of others. This observation is consistent with the results obtained by Englund Dimitrova and Tiselius (2009), who also noted considerable differences in the amount of talking between their participants. Trainee interpreters therefore appear to exhibit divergent reporting styles in verbosity.

The differences might in fact be due to individual retrospective styles rather than the cueing stimulus they are exposed to during retrospection. Such retrospective styles may result from personality traits, as some people are naturally more willing to share their thoughts and admit failure than others. Possibly, such differences might also stem from the novelty of the task and pre-training. However, in the present study, these two confounding factors were eliminated by ensuring that none of the participants had been subjected to the retrospective procedure before the experiment.

3.2 Informativeness of the reports

The second research question concerned the impact of the type of cueing on the informativeness of the retrospective protocols. It was hypothesized that cueing via a TT recording would induce richer reports, especially on strategies and problems experienced, since the participants were directly exposed to their outputs and heard the actual choices they had made. This parameter was measured in a concentration rate of four aspects of strategic processing in simultaneous interpreting verbalized by the participants: the identified

problem [IP], the reported source of the problem [SP], the adopted strategy [AS] and the reason for adopting a given strategy [RS].

All the types of problem affecting either the TT or its delivery fell into the first category [IP]: incoherent discourse, unfinished sentences, non-strategic omissions, mistranslations, mispronunciation and pauses. As sources of problems [SP], problem triggers were classified following Gile (2009) that are associated with ST features: speed of delivery, dense information content, enumerations, numbers, unknown words, cultural references and proper names which cannot be immediately recognized, and syntactic structures requiring reordering. The category of sources of problems also includes problems related to text processing under simultaneous interpreting constraints: coordination of simultaneous tasks (the efforts of listening and analysis, production and memory) and failure sequences (as described by Gile, 2009).

The broad category of strategies [AS] encompasses procedures which Gile (2009) refers to as tactics: comprehension, preventive, reformulation and coping tactics. These include, for instance, shortening or lengthening the EVS, segmentation, generalization, approximation, condensation, strategic omission, explaining or paraphrasing, compensation, transcoding. The last category [RS] includes the motivations of interpreters for adopting a given strategy, such as:

- avoiding being too far behind the speaker;
- wanting to concentrate on a subsequent ST segment;
- helping the receiver understand a foreign cultural reference or a metaphor;
- improving the text or its delivery, redundancy of an ST item and visual compensation in a speaker's presentation.

The first stage of the analysis involved a between-group comparison of each of the four aspects. Figure 1 shows that the difference between cueing via an ST transcript vs a TT recording according to reporting on the problems and their source is negligible. There appears, however, to be a visible disparity in reporting on the use of strategies. The difference is noticeable both when naming the strategy (79 in group A and 120 in group B) and when explaining the reason for using it (22 in group A and 41 in group B). This seems to support the hypothesis for research question 2. Influenced by the product itself, either the participants' memory is more efficiently stimulated or they tend to comment on what they observe in their outputs. However, this finding concerning a visible difference in strategy reporting was not corroborated by statistical analysis. In order to verify the significance of the disparity between groups A and B, I applied the ANOVA parametric test because the distribution proved to be normal in both participant groups when the Shapiro Wilk test was applied. The results of the one-way ANOVA reveal that the difference between the two groups in naming adopted strategies [AS] is not significant (F = 2.636, p = 0.11). Nor is the difference in reporting the reason for adopting a given strategy [RS], which also proves to be not significant (F = 1.633, p = 0.2).

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.
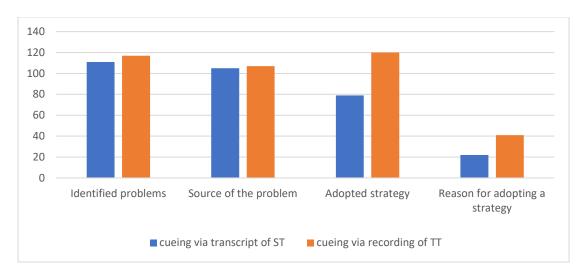
Figure 1. Between-group comparison concerning the four aspects of informativeness of the reports

In order to answer the second research question, further analysis was necessary. The concentration rate of four aspects of strategic processing was measured: the problem, the reported source of the problem, the strategy and the reason for adopting that strategy. The maximum concentration rate was assumed to be achieved when all four of these aspects were mentioned in one retrospective comment. Four remaining categories cater respectively for mentioning three aspects, two aspects, one aspect and none of them.[6] Example 1 below is an instance of verbalizing all four aspects of strategic processing in simultaneous interpreting. This trainee interpreter reports on his inability to produce the target-language (TL) version owing to the pace set by the speaker and the information density of the ST. As a consequence, he opted strategically to omit one segment of the text with a view to shortening the EVS and maximizing his performance in the subsequent segment.

(1) P26/B/RC5: *In the fragment **the speaker speaks quite fast and there was a lot of information** [SP] and that's why **I didn't manage to keep up with the text** [IP]. So **I decided to omit one bit of the text** [AS] **to save the one that followed and to be closer behind the speaker and to be able to translate it better** [RS].*

In descriptive statistics, the concentration rate shows some disparity between the two types of cueing (see Figure 2). All four aspects are mentioned more frequently by those participants who were exposed to the TT recording. The category encompassing three aspects also shows some prevalence among group B. In contrast, the remaining categories are dominated by experimental group A (ST transcript retrospection). Therefore the concentration rate of the four aspects appears to be higher in group B (TT recording as cue).

The between-group difference was further verified using the Kruskal Wallis test, as the data proved to be not normally distributed in the Shapiro Wilk test. Figure 2 presents the *H* and *p* values calculated for each category separately. Although the raw numerical data suggest higher concentration rates for group B, no difference between the two experimental groups was statistically significant. Therefore, the type of cueing does not appear to exert any decisive influence on informativeness or, at least, it would do so to a much lesser extent than might be inferred from the raw numerical data.
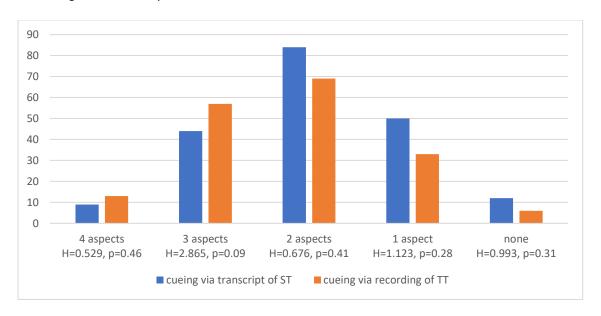
Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

Figure 2. Informativeness of retrospective comments

3.3 Accuracy of the reports

Analysis of the accuracy of the reports involved contrasting the data with the product – the TTs – following the study by Englund Dimitrova and Tiselius (2014). The first aspect examined in this part of the study is the use of strategies. Those verbalized by the participants in their retrospective protocols were compared with the actual textual solutions. Cueing via TTs guarantees direct access to the adopted solutions and is expected to result in more accurate reporting (Englund Dimitrova & Tiselius, 2009). Therefore, the research question posed here is not which type of cueing leads to more accurate reporting of strategies but whether the discrepancy between product and process data regarding the use of strategies is significantly lower when cueing is done via a TT recording.

To compare the reported strategies with the actual text solutions, three categories of correspondence were established. In the first case, reports matched the adopted strategies; in the second, a strategy was reported but there was no evidence of its use in the product. The last category comprised cases in which the strategy could be detected in the product but it was not reported in the retrospective protocol. Figure 3 displays the results for both experimental groups.
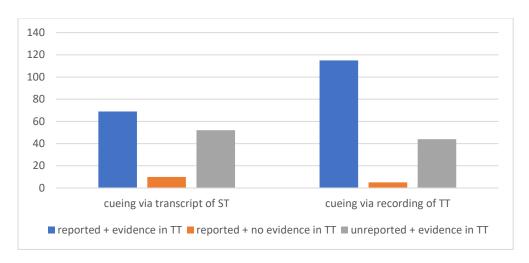
Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

Figure 3. Correspondence between reported strategies and strategies detected in the product

The results reveal that 87.34% of the strategies reported by group A (cueing via an ST transcript) coincide with the evidence found in the product and only for 12.66% of the reported shifts could no corresponding strategy be detected in the product. In the TT recording group, the ratio between evidence found in the TTs and the lack of evidence for reported strategies is 95.84% to 4.16%, which implies an only moderately higher accuracy for cueing via the TT recording. The difference in the number of unreported strategies is even less pronounced: 52 cases in group A and 44 in B. The results of the Kruskal Wallis test confirm a lack of statistical significance ($H = 0.757$, $p = 0.38$). Therefore, using the TT recording as a cueing stimulus appears not to guarantee a significantly higher accuracy of the reports in comparison with the other type of cueing as far as reporting conscious strategic processing is concerned.

The accuracy of retrospective protocols was also measured using disfluencies as problem indicators. Reports of problems and their sources were compared with problem indicators detected in the TTs. Since in cueing via a TT recording the participants are exposed to their own outputs while performing retrospection and processing problems can be inferred from the product, this type of cueing is expected to provide a more accurate picture of the problems experienced during interpreting. The aim of this part of the study was therefore to determine whether the discrepancies between product and process data regarding processing problems were significantly lower when cueing was done via the TT recording.

Accuracy was measured using three types of indicator of processing problem: hesitation markers, unfilled pauses exceeding two seconds and false starts. As in the case of the strategies, three categories of correspondence were established: processing problem was reported and its indicator was present in the product, processing problem was reported but there was no indicator in the product, or the processing problem was apparent from the product but remained unreported during retrospection.

The results show 76.66% of full correspondence between retrospective reports and problem indicators in the product for group A (cueing via the ST transcript), which means that in the case of 23.34% of the reports no corresponding indicator was found. The ratio for group B is 83.6% to 16.4%. These results are consistent with the difference in the number of unreported

problems – 46 cases in group A and 32 in group B, which also implies lower accuracy when cueing is done with the ST transcripts. However, the conclusions drawn from the raw numerical data were not confirmed when statistical tools were applied. The result of the Kruskal Wallis test showed no statistical significance of the difference ($H = 2.306$, $p = 0.12$). Therefore, as with the strategies, cueing via TT recording appears not to guarantee significantly higher accuracy of the reports when compared to using the ST transcript as retrieval cue, as far as reporting processing problems is concerned.
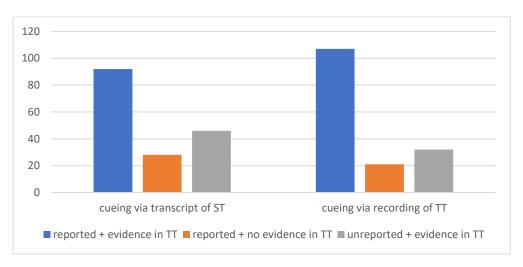

Figure 4. Correspondence between reported problems and problem indicators in the product

A similar analysis of accuracy was performed by Englund Dimitrova and Tiselius (2014) on retrospective reports cued by the ST transcript. The main difference between their results and those in this study is the proportion between reported and unreported problems. In their case, less than 20% of the problems detected in the product were reported by their participants in the retrospective protocols. In this study, unreported problems are much fewer. This difference is probably due to divergent operationalizations of processing problem indicators in the two studies. Apart from the disfluencies detectable in the product, Englund Dimitrova and Tiselius (2014) took into account changes of EVS, which could not be measured in this study because the TT recording featured only the TT (because of the research design adopted). Also, their threshold for unfilled pauses was much lower, beginning at only 0.5 s, whereas in this study only anomalous pauses lasting for more than 2.0 s were taken into account.

3.4 Participants' perceptions of the cueing stimuli

The participants' perceptions of the cueing they were exposed to was investigated with the aid of the questionnaires, which the participants were asked to fill in after the retrospective session. There were two versions of the questionnaire, depending on the type of cueing stimulus a given group was exposed to. The questionnaires contained seven questions each, with the questions 1–4 common to both versions. The first one was a multiple-choice question in which the participants were asked about the level of difficulty of recalling during retrospection, while the second one was a follow-up open question for those respondents who marked the options of either *difficult* or *very difficult* in question 1. They were asked to specify what made it difficult or very difficult for them to remember the decisions taken while

interpreting. In question 3 they were asked what facilitated recall and in question 4 what else could facilitate recall. Both 3 and 4 were open questions. Questions 5, 6 and 7 differed, depending on the type of cueing stimulus a given group relied on during retrospection. All three were multiple-choice questions with additional space for comments (see Table 2).

The answers for questions 1–6 provided data to answer the following research question addressed in this study: How do trainee interpreters perceive the efficiency of a given cueing stimulus? Let me stress that the participants had never performed a retrospective task before. This means that, when responding to the questionnaire, they had experienced only one type of cueing – the one assigned to their group. Therefore, their answers did not compare the two types of cueing, but were based only on their perception of the one they experienced during the study. It would be interesting to test the participants' perception after having experienced both of them, but, as I mentioned, this procedure was discarded in this study to avoid carry-over or/and order effects.

The ST-transcript participants perceived it as facilitating recall to a lesser extent than the TT-recording participants: ten out of 16 ST-transcript interpreters considered it either difficult or very difficult to remember the decisions taken while interpreting, whereas only five out of 16 of TT-recording participants rated remembering their decisions either difficult or very difficult. ST-transcript participants who considered cueing difficult tended to blame the inherent constraints of the simultaneous interpreting process as blocking recall during retrospection (four participants), as in the example below, rather than blaming it on a lack of access to their own outputs.

> (2) P02/A/Q2: *I think that simultaneous interpreting is such a stressful situation that my brain just "switches off" and I find it very difficult to remember my decisions afterwards.*

In a way this is consistent with the reason most often reiterated by participants in experimental group B (five participants), who claimed that it was the automaticity of text processing in SI and the intuitive, automated use of strategies which made it difficult for them to remember what had happened during interpreting:

> (3) P35/B/Q2: *To be honest, when interpreting simultaneously I don't think which strategy to use. I do it instinctively. I also couldn't compare my translation with the original.*

What is interesting, as can be seen in the example above, is that the participants exposed to the TT recording complained about the inability to consult the source text to a greater extent than those exposed to transcript of ST about the inability to listen to their outputs (seven in B and two in A).

Among the factors that facilitated recall during retrospection (question 3) for group A are, unsurprisingly, access to the ST transcript (ten participants), the characteristic features of the ST, such as humour, cultural references, numbers (3), the immediacy of retrospection after the interpreting task (2), the brevity of the text (2), and the problems experienced during interpreting (2), as in the example below:

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

(4) P01/A/Q2: *I could remember really well the strategies I used in those fragments which were especially difficult for me*.

The last reason was also mentioned by five participants in group B. Surprisingly, they mentioned the cue as facilitating retrospection with lower frequency than participants in the other group (only four participants). The stimulation of memory was also attributed to the immediacy of retrospection (4). Table 2 presents the results of the questionnaire (answers to multiple-choice questions).

Table 2. Results of the questionnaire (answers to multiple-choice questions)

		Cueing	
Multiple-choice questions	**Options**	**ST transcript (A)**	**TT recording (B)**
1. Was it difficult to remember during retrospection the decisions taken while you were interpreting?	It was easy	1	3
	Quite easy	2	7
	Neither easy nor difficult	5	3
	Difficult	6	4
	Very difficult	4	1
5. How useful was transcript of the ST *(version A)*/recording of the TT *(version B)*? Did it help you to remember the decisions taken while you were interpreting?	Very helpful	12	8
	Helpful	4	6
	Quite helpful	2	2
	Didn't matter	0	2
	Useless	0	0
6. Would listening to the recording of you interpreting *(version A)*/reading transcript of the ST *(version B)* be more useful?	Very helpful	8	3
	Helpful	6	2
	Quite helpful	3	6
	Wouldn't matter	1	3
	Useless	0	4
7. Do you think reading the transcript *(version A)*/listening to the recording of your interpreting *(version B)* distorts your memories of what you were thinking about while interpreting?	Doesn't distort them at all	10	8
	Slightly distorts	4	6
	Distorts	4	2
	Distorts them to a great extent	0	2

Among the factors that could possibly facilitate the retrospective task (question 4), members of the experimental group A (cued via the ST transcript) listed: listening to the TT recording instead of reading the ST transcript (4 participants), listening to the TT recording in addition to reading the ST transcript (3), better concentration during the interpreting task (2), access to the transcript of one's own interpreting output (1), and being informed prior to the interpreting task that they were supposed to report on their decisions after the task (1). For their part, the participants in group B (cued via the TT recording) mentioned as factors that could potentially facilitate retrospection: reading the ST transcript (3 participants), simultaneous listening to the TT recording and reading the ST transcript (2), slower pace of the speaker (2), being informed prior to the interpreting task that they were supposed to report on their decisions after the task (2), and the possibility of listening again to their outputs (2).

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

As far as the usefulness of the cueing stimuli is concerned, both groups perceived the cue they were exposed to as helpful. This finding was to be expected, but it is more interesting to see how it combines with the answers obtained to question 6. The majority of the participants from group A considered listening to their own TT recording as helpful to varying degrees, whereas in the case of group B there was no such unanimity. For three participants it would not matter if they could read the ST transcript, and four considered this type of cueing as of no use.

The respondents' answers to questions 1–6 reveal that the subjective perception of the participants may be the aspect that exhibits slightly more disparity between the two types of retrieval cue than the objective parameters of informativity, accuracy or verbosity. The difference can be noted in the perception of the usefulness of the cueing stimuli and the difficulty of recall.

The last research question in this study concerns the extent to which the trainee interpreters perceive a given cueing stimulus as distorting their actual memories. Based on the previous research (Englund Dimitrova & Tiselius, 2009, 2014), it was hypothesized that cueing via a TT recording would be perceived as entailing the risk of more distortions. However, the answers obtained for question 7 (see Table 2) showed no significant difference in that respect. A substantial proportion of the respondents (eight in group A and seven in group B) emphasized in their additional comments to that question that they believed it had not distorted them in any way and that, on the contrary, it helped them to stimulate their memory. Therefore, with such balanced results the hypothesis was not confirmed. Moreover, the additional comments made by some participants provide interesting evidence on how cueing via an ST transcript is perceived as installing false memories, as in the example below:

(5) P17/A/Q7: *Reading the transcript and analysing it made me think of better solutions and better strategies. It is difficult to describe what happens to me, but I feel that these new ideas were pushing out of my brain my memories and what really happened when I was interpreting*.

The results obtained for question 7 are consistent with the results on the accuracy of the retrospective reports, which also showed no significant difference between the two types of cueing.

4. Conclusions

The results of this study imply that the influence a retrieval cue exerts on the measurable parameters of retrospective reports might be less significant than it has so far been assumed. This conclusion should certainly be treated as tentative. First, the quantifiable parameters of verbosity, informativeness and accuracy account only for some aspects of the retrospective reports. Others, such as veridicality, for instance, cannot be measured or verified. Moreover, the large standard variation and the non-normal distribution of the data suggest that these parameters are highly dependent on the individual differences in reporting styles between the participants. Therefore, a larger sample is necessary to confirm the results obtained in this study before they can be generalized and extrapolated into a population of trainee interpreters. This last issue would also require researchers to ensure the greater external validity of the experiment by collecting data from more than one site. All the participants in this study come from the same interpreter training programme, and although this factor

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

should not constitute a confounding variable in this study, a multi-site design could improve its external validity and provide richer data.

The additional comments provided by two participants from experimental group B draw attention to something which might be a crucial aspect in determining the efficacy of a retrieval cue. They point out that they would have preferred to read the ST transcript because they find it easier to remember things if they see the text rather than listen to it. Therefore, visual or auditory dominance may prove to be an important factor in administering a retrieval cue as a prompt to the interpreters' memory. Possibly an experimental study pre-testing participants for their visual or auditory dominance could shed more light on the method of cueing used in retrospection.

References

Barghout, A., Ruiz R. L., & Varela. M. (2015). The influence of speed on omissions in simultaneous interpretation. An experimental study. *Babel*, 61(3), 305–334. https://doi.org/10.1075/babel.61.3.01bar

Bartłomiejczyk, M. (2006). Strategies of simultaneous interpreting and directionality. *Interpreting*, 8(2), 149–174. https://doi.org/10.1075/intp.8.2.03bar

Bartłomiejczyk, M. (2007). Introspective methods in conference interpreting research. In J. Arabski, (Ed.), *Challenging tasks for psycholinguistics in the new century: Proceedings of the 7th Congress of International Society of Applied Psycholinguistics.* (CD edition). Katowice.

Brzeziński, J. (2008). *Badania eksperymentalne w psychologii i pedagogice.* Wydawnictwo Naukowe Scholar.

Buchweitz, A., & Alves, F. (2006). Cognitive adaptation in translation: An interface between language direction, time, and recursiveness in target text production. *Letras de Hoje*, 41, 241–272.

Büttner, O. (2008). Tracing cognitive processes at the point of purchase: The validity of concurrent and retrospective verbal reports, In K. J. Perks & P. Shukla (Eds.), *Marketing landscapes: A pause for thought: Proceedings of the 37th EMAC conference 2008 [CD ROM]* (pp. 1–8). European Marketing Academy.

Büttner, O., & Silberer, G. (2008). Video-cued thought protocols: A method for tracing cognitive processes at the point of purchase, In T. Lowry (Ed.), *Brick and mortar shopping in the 21st century* (pp. 221–241). Lawrence Erlbaum.

Chang, C., & Schallert, D. (2007). The impact of directionality on Chinese/English simultaneous interpreting. *Interpreting*, 9(2), 137–176. https://doi.org/10.1075/intp.9.2.02cha

Englund D. B., & Tiselius, E. (2009). Exploring retrospection as a research method for studying the translation process and the interpreting process. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research* (pp. 109–134). Samfundslitteratur.

Englund, D. B., & Tiselius, E. (2014). Retrospection in interpreting and translation: Explaining the process?, *Monti. Special Issue – Minding Translation*, *1*, 177–200. https://doi.org/10.6035/MonTI.2014.ne1.5

Ericsson, K. A. (2003). Valid and non-reactive verbalization of thoughts during performance of tasks: Towards a solution to the central problems of introspection as a source of scientific data. *Journal of Consciousness Studies*, *10*(9–10), 1–18.

Ericsson, K. A., & Simon, H. A. (1980), Verbal reports as data. *Psychological Review*, *87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Rev. ed.). John Benjamins. https://doi.org/10.1075/btl.8

Gumul, E. (2006). Explicitation in simultaneous interpreting: A strategy or a by-product of language mediation?, *Across Languages and Cultures*, *7*(2), 171–190. https://doi.org/10.1556/Acr.7.2006.2.2

Gumul, E. (2017). *Explicitation in simultaneous interpreting: A study into explicitating behaviour of trainee interpreters.* Wydawnictwo Uniwersytetu Śląskiego.

Hansen, G. (2005). Experience and emotion in empirical translation research with think-aloud and retrospection. *Meta*, *50*(2), 511–521. https://doi.org/10.7202/010997ar

Hild, A. (2015). Retrospective protocols. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 351–353). Routledge.

Ivanova, A. (1999). *Discourse processing during simultaneous interpreting: An expertise approach* [Unpublished doctoral dissertation]. University of Cambridge.

Ivanova, A. (2000). The use of retrospection in research on simultaneous interpreting. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research* (pp. 27–52). John Benjamins. https://doi.org/10.1075/btl.37.05iva

Leanza, Y. (2005). Role of community interpreters in pediatrics as seen by interpreters, physicians and researchers. *Interpreting*, *7*(2), 167–192. https://doi.org/10.1075/intp.7.2.03lea

Li, C. (2010). Coping strategies for fast delivery in simultaneous interpretation. *JoSTrans: The Journal of Specialised Translation*, *13*, 19–25.

Mead, P. (2002). Exploring hesitation in consecutive interpreting: An empirical study. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 73–82). John Benjamins. https://doi.org/10.1075/btl.43.08mea

Napier, J. (2004). Interpreting omissions: A new perspective. *InterpretingI*, 6(2), 117–142. https://doi.org/10.1075/intp.6.2.02nap

Russell, D., & Winston, B. (2014). Tapping into the interpreting process: Using participant reports to inform the interpreting process in educational settings. *Translation & Interpreting*, *6*(1), 102–127. https://doi.org/10.12807/ti.106201.2014.a06

Seeber, K. (2015). Cognitive approaches. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 56–60). Routledge.

Shamy, M., & de Pedro Ricoy, R. (2017). Retrospective protocols: Tapping into the minds of interpreting trainees. *Translation & Interpreting*, *9*(1), 51–71. https://doi.org/10.12807/ti.109201.2017.a05

Tang, F. (2018). *Explicitation in consecutive interpreting.* John Benjamins. https://doi.org/10.1075/btl.135

Tiselius, E., & Jenset, G. B. (2011). Process and product in simultaneous interpreting: What they tell us about experience and expertise. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 269–300). John Benjamins. https://doi.org/10.1075/btl.94.20tis

Vik-Tuovinen, G.-V. (2002). Retrospection as a method of studying the process of simultaneous interpreting. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 63–71). John Benjamins. https://doi.org/10.1075/btl.43.07vik

---

[1] The term *veridicality* is used here, as in psychology, to mean simply the degree to which the retrospective reports accurately represent reality, as in the study of Büttner (2008).

Gumul, E. (2020). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 152–171.

---

[2] The term *verbosity* is used here in the neutral sense of the amount of talk, as it is used by Shamy and de Pedro Ricoy (2017, p. 55). It does not have any negative connotation and is not meant to imply excessive verbalizations or a superfluity of words.

[3] *Hesitation markers* are understood here as fillers in the form of non-lexical utterances, that is, meaningless strings of prolonged sounds (in Polish typically assuming the form of prolonged vowels: *yyy* or *eee*). By *false starts* we understand here words that are cut off mid-utterance and restarted or repeated words or syllables at the beginning of a turn.

[4] Source text: Meyer, P. (2011). *How to spot a liar*. Retrieved from
https://www.ted.com/talks/pamela_meyer_how_to_spot_a_liar/transcript

[5] By *designating* a strategy we mean identifying and reporting strategic behaviour. The "correct" name of the strategy was not a decisive criterion, given that the knowledge of terminology does not necessarily say much about either the process or the retrieval.

[6] The following report is an example of a verbalization in which none of the four aspects analysed in this study was mentioned: P03/A/RC6: *I remember I translated well this fragment about Henry Oberlander. In general this fragment was easy to understand because the speaker presented him first and then said what he did, so it was clear how it was related to lying*.