Evidence of cognitive effort in simultaneous interpreting: Process versus product data

EWA GUMUL

Received 26.10.2018, received in revised form 03.07.2019, accepted 04.09.2019.

Abstract

The present study aims at measuring the level of correspondence between the problems related to increased cognitive effort reported by simultaneous interpreting trainees in their retrospective verbal protocols and problems indicators identified in the target texts. The study triangulates process analysis (retrospective protocols) with product analysis (manual comparison of source and target texts). The corpus of the study consists of about 75 hours of recordings of 240 interpreting outputs and the recordings of 5,005 retrospective comments. The results imply that increased cognitive load involved in simultaneous interpreting and the resulting cognitive effort experienced by trainee interpreters is not always manifested in the product.

Keywords

retrospective protocols, process research, simultaneous interpreting, Gile's Effort Models, cognitive effort, cognitive load
Wysiłek kognitywny w tłumaczeniu symultanicznym. Analiza procesu i produktu przekładu

Abstrakt

Celem niniejszej pracy jest zbadanie zależności pomiędzy wysiłkiem kognitywnym zwerbalizowanym przez tłumaczy symultanicznych w protokołach retrospektywnych a oznakami wysiłku kognitywnego widocznymi w tekście docelowym. Badanie przeprowadzono za pomocą metody retrospekcji oraz analizy produktu przekładu. Korpus badawczy stanowi 240 tłumaczeń oraz 5005 komentarzy retrospektywnych. Wyniki pokazują, że zwiększone obciążenie kognitywne charakterystyczne dla przekładu symultanicznego i będącego efektem wysiłek kognitywny, którego doświadczają tłumacze, nie zawsze uwidacznia się w tekście przekładu.

Słowa kluczowe

protokoły retrospektywne, badania and procesem przekładu, tłumaczenie symultaniczne, Modele Wysiłkowe Gile’a, wysiłek kognitywny, obciążenie kognitywne

1. Introduction

The aim of this study is to test the level of coincidence between the problems related to increased cognitive effort reported in retrospective protocols of simultaneous interpreting trainees and the problem indicators encountered in the target texts. The research conducted by Englund Dimitrova and Tiselius (2014) shows that the majority of reported problems coincide with the problem indicators identified in the target texts whereas not all problems encountered in the target texts are reported. Given that the latter might be due to a number of reasons related to the limitations of the method, such as memory capacity, potential unwillingness to report problems, and also automaticity in using coping tactics or strategies when dealing with recurrent problems, in this study I will focus only on processing capacity
problems reported during retrospection. My aim is firstly to see how they are manifested in the SI renditions and secondly to ascertain whether any of the reported problems are effectively masked by the efficient use of coping tactics and strategies and thus are not detectable in the target-language versions. Unlike in the studies of Englund Dimitrova and Tiselius (2009, 2014), in which they examined a broad spectrum of problems, I will limit my analysis to those problems related to increased cognitive load and processing capacity overload or saturation following the premises of Gile’s Effort Models (1985, 1995, 2009, 2016, 2017). Thus, for the purpose of process analysis, cognitive effort has been operationalized as: failure sequences, evidence of competing efforts (competition hypothesis), evidence of working close to cognitive overload (tightrope hypothesis), and negative effect of problem triggers. Previous research (Gumul 2018) revealed that 531 comments out of the total of 5,005 obtained in the study referred to the aspects related to Gile’s Effort Models and 108 out of 120 interpreters participating in the experiment made at least one remark reporting them. In this study I will reanalyse these 531 comments comparing each of them with the corresponding target-text segment in search of the problem indicator in the product. The ones analysed in this study are: anomalous pauses, omissions in the target text, repairs, grammatical errors, mispronunciations, and disfluencies in the form of hesitation markers and false starts. These problems indicators are assumed to be indicative of cognitive effort. Such claim might seem precipitated as there is no scientific evidence unequivocally linking such infelicities to cognitive effort. However, since cognitive effort was already reported by the subjects in their retrospective reports, the assumption has been made that these disfluencies, errors and omissions in the product are likely to reflect cognitive effort in these particular cases.

The paper begins with the presentation of the construct of cognitive load and a brief overview of literature regarding the existing models, relying in particular on Gile’s Effort Models, as this framework is the basis of analysis conducted for the
The high cognitive load involved in the task of simultaneous interpreting is what distinguishes this mode of interpreting from other forms of translation. As indicated by Seeber, it can either refer to “the processing load imposed on the performer by a particular task (...), or the perceived mental effort the performer invests in the task” (2015: 60). This multidimensional construct originating from the field of psychology was applied extensively for the first time to interpreting research by Gile (1985, 2008) to account for linguistic infelicities, distortions and loss of information observed in simultaneous interpreting, although the broader idea of cognitive processing has been present in interpreting studies since the early 1970s (e.g., Gerver 1971,1 quoted in Pöchhacker 2016; Kirchhoff 1976/2002).

The cognitive load involved in simultaneous interpreting is central to many approaches and models accounting for the mental complexity of this task, especially those aiming at exploring the process of interpreting at the innermost, essentially micro-process-oriented, levels of analysis identified by Pöchhacker (2004, 2016), i.e. cognitive and neural (see Figure 1).

---

Of all the potential levels of analysis indicated by Pöchhacker (2016), it is the cognitive sphere which has attracted the highest degree of analytical interest, giving rise to a substantially higher number of models than any other level, which is evidenced by the comprehensive overviews of the existing cognitive processing models in interpreting presented by Moser-Mercer (1997/2002), Gile (2003), Setton (2003, 2015), Pöchhacker (2016), and Korpal (2017). In the subsequent sections I will present the selected models describing processing in SI from the cognitive perspective focusing mostly on Gile’s Effort Models, for which empirical evidence is sought in the present study.

3. Cognitive processing models

Early attempts at modeling interpreting within the cognitive perspective began in the 1970s in the experimental psychology period of interpreting research with the contributions of Gerver (1976) and Moser (1978). Both of their models were built upon
the information processing approach from the field of psychology. Gerver’s (1976) flowchart-type model of the mental structures and procedures involved in input processing and output production in simultaneous interpreting assumes the existence of control processes that play a vital role in “the distribution of attention to the different components of the task” (Gerver 1976: 193). Whereas Gerver’s (1976) model is essentially psychological in nature, Moser’s (1978) processing model of simultaneous interpreting is an attempt to seek interdisciplinary cooperation between psychologists and interpreters. She elaborated her model of memory structures and processing operations in the form of a flow diagram presenting the temporal course of simultaneous interpreting on the basis of Massaro’s speech-comprehension model. Like Gerver’s (1976) contribution, Moser’s (1978) model also accounts for the simultaneity of processing stages and the necessity of cultivating divided attention between input comprehension and output production. She also indicates that the operations involved in SI compete for available processing capacity. At the same time another model of simultaneous interpreting appeared, elaborated by Kirchhoff (1976/2002), however it went largely unnoticed outside of German-speaking countries until the publication of its translated version in Pöchhacker and Shlesinger’s (2002) Interpreting Studies Reader. Kirchhoff (1976/2002) models simultaneous interpreting from a psycholinguistic perspective and her multi-tasking model not only explains the mental processes involved in simultaneous interpreting, but also accounts for psycholinguistic processing difficulties resulting both from reaching or exceeding processing capacity limits and divergent linguistic structure of the language pair involved.

Unlike the approaches discussed so far, which were firmly rooted within psychology, Chernov’s (1979/2002, 2004) probability-prediction model is based on discourse-oriented linguistics. The model assumes that processing in simultaneous interpreting is to a large extent expectation-based and the key factors which facilitate anticipation are the inherent redundancy of
linguistic expression and the probabilistic nature of discourse. While acknowledging the importance of surface structure in forming predictions about the incoming discourse, Chernov also emphasizes the role of background knowledge stating that “[t]he semantic component of the suggested model interacts most closely with the individual’s store of knowledge in general [...], and with the situational context of communication in particular” (1979/2002: 106). The aspect of knowledge-based processing in SI is what links Chernov’s (1979/2002, 2004) model to the Paris School model which also emphasizes the role of the interpreter’s external knowledge. In the triangular process model developed by Seleskovitch (1962), in which she shows the relationship between the sense of the message and the two languages involved, interpreting is viewed as the act of deverbali-
sation, i.e. reducing words to nonverbal sense (Seleskovitch and Lederer 1984). Whereas the triangular model has a more general implementation applying to both consecutive and simultaneous interpretation as well as to translation, Lederer’s (1981) model of eight mental operations, also elaborated within the théorie du sens, accounts for the specificity of SI, modeling this mode of interpreting in terms of continuous interplay between multiple concurrent tasks. In line with the general tenet of the interpre-
tive theory, this model also acknowledges that conceptualization occurs by integrating linguistic input with prior knowledge.

Another paradigm which accounts for the role of extratextual knowledge is the cognitive-pragmatic model developed by Setton (1999). He models simultaneous interpreting beyond strictly cognitive mental operations and aims at incorporating communicative aspects in the analysis of mental processes underlying the interpreting task within the framework of the Relevance The-
ory. The pragmatic context in which the communication takes place is considered to be of vital importance at all stages of cognitive processing. The features of discourse processing which are believed to ease the effort expended in simultaneous interpreting are the construction of mental models and the activation of knowledge schemas. Setton also emphasizes the role of the
The construct of processing capacity is also central to Seeber’s (2011, 2013) Cognitive Load Model, in which he attempts to describe the amount of cognitive load generated during the task of simultaneous interpreting. With this objective in mind, he designed a conflict matrix for simultaneous interpreting that accounts for interference and overlapping of concurrent tasks. This model subsumes the separate cognitive load model for four SI strategies: waiting, stalling, chunking and anticipation, which are commonly adopted in the case of structurally divergent languages. In his later contribution Seeber (2013, see also Seeber and Kerzel 2012) set out to measure cognitive load involved in simultaneous interpreting by means of pupillometry. As Seeber (2011) indicates, his model is meant as a competing account to Gile’s Effort Models, which we shall present in the subsequent section as the conceptual framework for the analysis conducted in this study.

4. Gile’s Effort Models

Gile (1995, 1997, 2009) models simultaneous interpreting as a process consisting of three concurrent operations requiring processing capacity: the Listening and Analysis Effort (L), the Production Effort (P), and the Memory Effort (M). The first part involves all of the comprehension-oriented operations from receiving the sound waves to the final decisions about the meaning of the text. The second part, the Production Effort, extends beyond the mere delivery of the target text, as it also includes the formation of a mental representation of the message and speech planning operations. Finally, the Memory Effort refers to all of the mental operations related to the storage of information in memory. This effort assumes the increased demand on short-term memory during the task of simultaneous interpreting (Gile 1997: 198).
Gile (1995, 1997, 1999) complements the equation with the Coordination Effort (C) stressing that coordination function (or “executive” function) also consumes attentional resources. The model assumes that two more conditions must be satisfied for successful performance in simultaneous interpreting. First, L + P + M + C must be less than the Total Available Processing Capacity (TAPC). Second, the processing capacity management condition must be satisfied, which means that “the capacity available for each effort (LA, MA, PA, and CA) must be equal to or larger than its requirements for the task at hand” (Gile 1997: 199).

Gile (1995, 2009) regards all of the operations constituting the simultaneous interpreting process – that is listening and analysing, producing the target text, and storing it in memory – as non-automatic. He relies on evidence from cognitive psychology and psycholinguistic research which confirms that each of these processes has controlled components, given that they all include deliberate action requiring decisions and resources.

In fact, as Gile (1997: 197) emphasizes, the name of the model was chosen to accentuate the non-automatic nature of these operations. It is the very nature of simultaneous interpreting which makes the three processes of comprehension, speech production and memorizing far less automatic than it is the case in any other communication situation.

Firstly, the comprehension effort is more active in SI, given that the interpreters cannot afford to practice selective listening. Unlike conference delegates who can just focus on what they are interested in, interpreters have to concentrate on every aspect of the source speech to be able to render the text. The effort expounded on listening and analysis is also believed to be more intense in interpreting due to the unshared knowledge constraint, which means that the interpreters are bound to be less familiar with the subject matter and terminology than the delegates to whom the source text is addressed (Gile 2009: 16).

Also speech production entails more effort under simultaneous interpreting conditions. Given that interpreters do not
speak on their own behalf, but have to render somebody else’s speech, they cannot resort to conventional strategies for improving fluency like opting for habitual combinations of words or rearrangement of sequences of words and ideas. Gile observes that “interpreting constraints force interpreters to deviate from their habitual speech production patterns” (2009: 165), which obviously lowers the automaticity of this mental operation.

Finally, the same applies to the third constituent of this model – the memory effort, which is considered to be an essentially non-automatic operation since it involves the storage of information for later use (Gile 2009: 166).

The non-automatic nature of the three efforts is one of the operational assumptions underlying Gile’s Effort Models. The others are the competition hypothesis and the tightrope hypothesis.

The competition hypothesis postulates the competitive allocation of effort between the three non-automatic processes. The model assumes that even though the three efforts may be to some extent cooperative, their coexistence will invariably increase processing capacity requirements (Gile 1999: 156). In mathematical terms, Gile (1999) represents his competition hypothesis with one equation (1) and three inequalities (2–4):

1. \[ \text{TotC} = \text{C(L)} + \text{C(M)} + \text{C(C)} \]
2. \[ \text{C(i)} \geq 0 \quad i = \text{L, M, P} \]
3. \[ \text{TotC} \geq \text{C(i)} \quad i = \text{L, M, P} \]
4. \[ \text{TotC} \geq \text{C(i)} + \text{C(j)} \quad i, j = \text{L, M, P} \text{ and } i \text{ different from } j \]

which are explained as follows:

- equation (1) represents the total processing capacity consumption
- inequality (2) means that each of the three Efforts requires some processing capacity
- inequality (3) means that the total capacity consumption is at least equal to that of any single Effort
- inequality (4) means that the total capacity consumption is at least equal to that of any two Efforts performed in conjunction (in other
words, adding a third Effort means adding further capacity consumption). (Gile 1999: 156-157)

The third operational assumption of the Effort Models – the tightrope hypothesis – presupposes that most of the time interpreters work close to cognitive saturation. Gile emphasizes that in simultaneous interpreting we have to take into account that interpreting failures are also caused by “chronic cognitive tension between processing capacity supply and demand” (2009: 182) apart from insufficient linguistic or extralinguistic knowledge or mistakes.

Unlike most of the existing cognitive models of simultaneous interpreting, which do not take into account the external factors that further hinder the process of the cognitively demanding interpreting task, Gile identifies several sources of SI processing failure (2009: 192ff) referring to them as problem triggers. He defines them as “factors and conditions which increase processing capacity requirements or make the interpreter more vulnerable to attention lapses and attention management errors” (Gile 2015: 136). Problem triggers are grouped according to processing capacity requirements imposed by individual efforts. It is worth noting that the very presence of problem triggers does not necessarily engender problems with processing capacity. Problem triggers can only be treated as potential sources of errors or omissions, but whether they occur or not depends on the context. For instance, an informationally dense segment may come at the end of the sentence and additionally be followed by a pause. Then, the Listening and Analysis Effort is no longer active, and the whole processing capacity can be directed to the Memory and Production components (Gile 1995: 174).

Within his Effort Models, Gile (1995) also introduces the idea of failure sequences, which assumes that a given problem trigger might potentially give rise to EOIs (errors, omissions, or infelicities – see Gile 2015) as a result of processing capacity mismanagement rather than only because of its inherent difficulty. One possible scenario of a potential failure sequence is when
the interpreter is faced with an incoming speech segment requiring additional capacity resources for production (e.g., a speech segment of high density in terms of either fast delivery rate or dense information content). As a result the interpreter may be forced to delay producing the target-language version until more processing capacity is available for the Production Effort. Obtaining extra processing capacity is possible after the interpreter has been freed from the Listening Effort, that is, working on the incoming speech segment. This, however, may impose excessive strain on the Short-Term Memory Effort because of the backlog of incoming input segments that has accumulated in the meantime. If the interpreter tries to deal with the problem by directing more processing capacity to the Short-Term Memory Effort, this may lead to losses in the capacity aimed for the Listening and Analysis Effort, putting the comprehension of another incoming segment at risk (Gile 1997: 200). Many other types of failure sequences are possible (see Gile 2009: 173). The above simulation clearly shows that failure sequences do not necessarily affect the problematic segment that triggered them, but may occur at a distance, influencing the rendition of segments that pose no particular difficulty (Gile 1995: 175).

Gile’s conceptual framework of the Effort Models has been criticised for not taking into account the socio-pragmatic perspective of the interpreting process (see, e.g., Pym 2008). Indeed, considering the contextualists versus cognitivists divide, Gile’s (1995, 1997, 2009) model is clearly on the cognitive side. This, however, only reflects the predominant orientation of this paradigm, and by no means does it mean that Gile denies any influence of the socio-pragmatic factors on simultaneous interpreting. In fact, in the most extensive description of the model, which appeared in both editions of his “Basic Concepts and Models for Interpreter and Translator Training” (Gile 1995, 2009), we can find references to the pragmatics of the communicative situation. It is evident in the discussion of laws underlying the choice of preventive and coping tactics. As Gile points
out, interpreters are guided by self-protection (risk-aversion) and seek maximum information recovery and maximum communication impact of speech (Gile 1995: 201-204; Gile 2009: 211-214; see also Gile 2016, 2017).²

These two communication-oriented aspects of Gile’s Effort Models, as well as those essentially related to cognitive processing, we shall see reflected in retrospective protocols of trainee interpreters analysed in the present study.

5. Research design

As indicated in the title of this paper, the present study relies on two types of data: the product data and the process data. The former consists of recordings and transcriptions of simultaneous interpreting outputs,³ whereas the latter includes recordings and transcriptions of retrospective protocols of the interpreters participating in the experiment. The target-texts amount to approximately 75 hours of recordings and the accompanying retrospective material encompasses 5,005 comments. The material was recorded for the purpose of a larger project aiming at investigating explicitation in simultaneous interpreting (Gumul 2017). Nevertheless, the research design of the former study makes it possible to use the gathered corpus of target texts and verbal reports for the analysis of other aspects of simultaneous interpreters performance.

The subjects in the study form a fairly homogenous group of advanced interpreting trainees with a comparable level of

---

² Gile’s Effort Models are described in more detail in the article which searched for evidence of Gile’s Effort Models in retrospective protocols of simultaneous interpreters (Gumul 2018).

³ Note that Englund Dimitrova and Tiselius (2009, 2014) refer to the interpreting outputs as process data, whereas in this study process data refers to retrospective verbal reports, as they are assumed to provide more direct information about the process of interpreting and are one of the tools in interpreting process research (cf. e.g., Hild 2015). Studies relying on verbal reports (either think-aloud protocols or retrospective verbal reports) are considered as process-oriented studies, in contrast to product-oriented studies, which rely only on the target texts.
previous training in conference interpreting and the same language combination. All the participants were native speakers of Polish with English as language B in their combination of working languages. They were students from three Polish higher education institutions: the University of Silesia in Katowice, the University of Gdańsk, and the University College of Social Sciences in Częstochowa (formerly the College of Foreign Languages at the time the material was recorded). Prior to the experiment, the students received between 120 and 150 hours of conference interpreting training.

The source texts used in the study belong to three different genres. They comprise typical text types, routinely interpreted using simultaneous mode: conference presentations, commencement addresses, and political speeches (see Sources). The interpreters were asked to interpret in both directions – from English into Polish and from Polish into English. In order to ensure a relative comparability of source texts in each direction, the selected texts are to a certain extent comparable in terms of the topic, the context in which they were delivered, and most importantly the texts constituting each pair share some features of orality, which is directly related to information density. The conference presentations were delivered during a seminar on medical ethics and concern both the moral and legal aspects of refusing to help a patient. These two speeches are apparently written texts delivered orally. In turn, the commencement addresses were both presented at art schools and are fairly general speeches spontaneously delivered, with the lowest degree of prior preparation in terms the exact wording and form. Finally, the political speeches both concern the Iraqi conflict and are

\[4\] The study cannot be described as truly experimental, as it does not comply with the requirements of experimental research both in terms of the distribution of subjects between experimental and control groups, and the random assignments of subjects to such groups. The term experimental is employed here in its broadest sense, which is often used in the translation studies to refer to the studies which are carried out in a controlled laboratory setting and involve the manipulation of certain variables.
pre-prepared speeches designed to closely resemble oral discourse.5

Each source text was interpreted by 40 interpreters working in both directions of interpreting, for a total of 240 target texts. The source texts used in the experiment are authentic speeches. However, in order to control the variable of the rate of delivery, which has a profound impact on various aspects of text processing in simultaneous interpreting, the texts were recorded in laboratory conditions by native speakers of the respective languages. For English source texts, the rate was kept at an average of 110–120 words per minute,6 whereas in the case of source texts delivered in Polish, it was 80–90 words per minute. The difference in word count stems from the systemic differences between Polish and English, since Polish words tend to be longer. These values might be considered to be roughly equivalent on the basis on the syllable count per minute (Gumul 2017).

The experimental procedure (see footnote 5 above) involved recording the interpreting outputs (dual-track recording) in standard laboratories used for teaching simultaneous interpreting. Prior to the interpretation, the subjects received a thorough briefing concerning the pragmatic setting of each speech, that is, the details concerning the identity of the speaker, the profile of the target audience, the time, the venue, and the subject matter of each text. The cue used to stimulate recall during retrospection was a dual-track recording of the target texts (with the source text in the background). The subjects performed self-retrospection,7 meaning there was no prompting on the part of the

---

5 A detailed description of the corpus and the rationale for the selection of the texts are available in Gumul (2017).

6 The average of 120 words per minute is the rate of delivery of the source text in English which is generally believed to be the most comfortable and optimal for simultaneous interpreters to ensure the best quality of interpreting (e.g., Gerver 1975, Pöchhacker 2004, Bartłomiejezyk 2016).

7 Each participant was given control of the in-built recorder in the booth and was asked to stop the recording each time he or she remembered a consciously taken decision and to comment on it aloud. The retrospective
researcher. All of the instructions were given before they started listening to their outputs. The subjects were asked to report all decisions taken consciously during the interpreting task. The objective of the study was not disclosed as to avoid influencing the subjects’ verbalisations. This aspect of the research design made it possible to use the data obtained to investigate other aspects of the simultaneous interpreting process, apart from the original aim of analysing explicitation (the results of which were reported in Gumul 2017).

The 5,005 retrospective comments obtained in the study were analysed in search of reports of experienced cognitive effort operationalised here, following the framework of Gile’s (1995, 1997, 2009) Effort Models, as failure sequences, evidence of competing efforts, evidence of working close to cognitive overload, and negative effect of problem triggers. As indicated in the introduction to the present paper, 531 verbalisations report on the above-mentioned aspects related to Gile’s Effort Models (2009). A detailed account of the results of this stage of the analysis was described in the previous article (Gumul 2018). In this study, the target text segments for which the interpreters reported increased cognitive load are analysed for problem indicators. In other words, the analysis aims at establishing the level of correspondence between process reports and the indicators present in the product.

In this study the problem indicators that can potentially reflect cognitive effort experienced by interpreters and resulting processing problems are: anomalous pauses, omissions in the target text, repairs, grammatical errors, mispronunciations, and disfluencies in the form of hesitation markers and false starts. Some of the above categories need to be clarified further. As far as pauses are concerned, I have taken into account only

comments were recorded on the external source (portable dictation devices placed in the booths) activated by the researcher before the retrospective session.
anomalous unfilled pauses exceeding two seconds. Filled pauses are accounted for within the category of hesitation markers. As for omissions, the only type relevant for the present study are those leading to the loss of information content, which means I excluded surface-form omissions resulting in implicitation or leading to an acceptable level of condensation. Omissions and unfilled pauses are two separate categories in this study, assuming that not every omission might lead to a pause exceeding two seconds. The category of repair also requires some clarification. For the purpose of this study I excluded those self-corrections which apparently stem from lexical search under time pressure and those in which both lexical items are acceptable equivalents in a given context. The results of my previous research reveal that such reiterations, that on the surface look like repairs, are in some cases intended to clarify the message, i.e. are performed either to help the receiver to decode the message, to avoid ambiguity, or to improve the text by adding emphasis (see Gumul 2017). As for grammatical mistakes, judging only the product it might be difficult to distinguish them from grammatical errors which obviously are not in most cases indicators of increased cognitive effort, but simply signs of linguistic incompetence, especially when interpreting into a language B. However, given that the present study only focuses on the analysis of target text segments for which the subjects reported increased cognitive effort, and in the absence of other recurrent grammatical incorrectness of the same type in other segments of a given target text, the assumption has been made that in such text segments they are mistakes that result from processing problems.

---

8 Two seconds is the threshold value for anomalous pauses in many studies investigating fluency in simultaneous interpreting (e.g., Pradas Macías 2006). In the study of Pradas Macías (2006) pauses exceeding two seconds in simultaneous interpreting are considered to have a negative impact on the evaluation of fluency.
For the purpose of analysis, four categories of possible correspondence (or lack thereof) between the process data and the product have been established:

1. **RC and I** – cognitive effort reported in retrospective protocols corresponds to a problem indicator in the target text (in the reported segment or the following, which points to a failure sequence). This category also includes the cases of cognitive effort reported at the global level in general comments.

2. **RC and no I (strategy)** – cognitive effort is reported in retrospective protocols, but there is no apparent corresponding problem indicator in the target text, which attests to an efficient use of preventive or coping tactic. The strategy is either reported or visible in the product.

3. **RC and no I (no strategy)** – cognitive effort is reported in retrospective protocols, but there is no corresponding problem indicator in the target text and no preventive or coping strategy has been reported or could be identified in the target text.

4. **RC post and I** – the last category subsumes those cases in which the cognitive effort is reported in the retrospective protocols but the interpreters admit it is *a posteriori* remark triggered by listening to their outputs and it was not a conscious decision taken during the interpreting task. Such verbalisations usually take the form of a supposition (e.g., “probably I didn’t hear this part of the sentence”, “perhaps I was already tired” etc.). These verbal reports were not taken into account in the previous study (not included in the total count of 531 analysed verbalisations) as they are not retrospective and were triggered by the cueing stimuli. For that reason, in this study, the results for this category are presented separately (see the last column of Table 1) and are not included in Figure 2.

### 6. Results and discussion

The results reveal that in the majority of cases the reported increased cognitive effort corresponds to problem indicators in the
product (see Figure 2 and Table 1). As many as 80% of retrospective comments reporting cognitive effort coincide with one or more indicators identified in the corresponding target texts’ segments or those immediately following them (carry-over effect, which is characteristic for SI).

The number of verbal reports which do not coincide with any increased cognitive effort indicator in the product is relatively high (19%, and the total of 99 cases) compared with the results obtained by Englund Dimitrova and Tiselius (2014). In their study there were no such examples in the performance of student interpreters and 15% in the case of professionals, and still for the latter group the researchers were able to find some justification suggesting that in their corpus the examples illustrating this tendency are not fully representative (Englund Dimitrova and Tiselius 2014: 193). The difference between the two studies might be due to a number of reasons. First, we need to consider the sample size. The smaller the size of the corpus, the less likely we are to find examples for less frequent phenomena. Thus, it is natural that in a study analysing the performance of six interpreters certain phenomena might be underrepresented. Another potential reason for the difference is the operationalisation of the analysed problems. As indicated in the introduction, the two studies take into account a different scope of processing problems. The disparity between the results might also be due to a different level of student training in interpreting. The subjects in the study of Englund Dimitrova and Tiselius (2014) were taking the introductory course to translation and interpreting, whereas all the participants in the present study received a complete training in all modes of conference interpreting. Therefore, it might be the case that the subjects in the present study, who are at the final stages of their training, have already mastered more efficient techniques of masking processing problems. As could be expected, the results for the third category show that it is a negligible phenomenon most probably due to the inherent weaknesses of the method of stimulated re-
call, like memory limitations, a wish to demonstrate cooperation with the researcher, and influence of the cue.

This last factor is also at play in the last identified category of verbal reports which are not truly retrospective. These *a posteriori* remarks report on experienced cognitive effort after having noticed the indicators in their own outputs. They do not reflect a conscious train of thought experienced during the interpreting process, and as such are not taken into account in the total count of analysed reports (the identified 531 comments). Nevertheless they were included in the table below, as these remarks can be considered as an interesting source of information on how trainee interpreters perceive indicators of processing problems in their own outputs (see the last column in Table 1).

![Figure 2](image.png)

**Figure 2**
The coincidence between reports of cognitive effort and their indicators in the target texts.
Table 1
The coincidence between reports of cognitive effort and problem indicators in the target texts

<table>
<thead>
<tr>
<th></th>
<th>NoRC</th>
<th>RC and I</th>
<th>RC and no I (strategy)</th>
<th>RC and no I (no strategy)</th>
<th>RC post and I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem trigger</td>
<td>157</td>
<td>96</td>
<td>58</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Failure sequence</td>
<td>52</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Competition hypothesis</td>
<td>231</td>
<td>187</td>
<td>41</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Tightrope hypothesis</td>
<td>91</td>
<td>91</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>531</td>
<td>426</td>
<td>99</td>
<td>6</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 1 shows how these four categories of correspondence between process and product data are represented in the four aspects representing the premises of Gile’s Effort Models. The first column (marked NoRC) presents the total number of verbal reports for each of them, while in the subsequent three columns the numbers are split according to the level of correspondence between reported problems and their indicators visible in the interpreting outputs. It is interesting to note that in the case of the reports testifying to failure sequences and the tightrope hypothesis, the subjects’ verbalisations always coincide with the problem indicators. The result is hardly surprising given the fact that the scale of experienced processing problems is greater in the case of the saturation or overload of the processing capacity, which is associated to a much greater extent with failure sequences and the tightrope hypothesis than competing efforts or the occurrence of a problem trigger.

The first analysed example represents the first category of correspondence between the process and the product data (RC
and I), in which the problem verbalised during retrospection coincides with problem indicators in the target text. In order to free the production effort and avoid taxing the Short-term Memory Effort with the proper name that has been identified by the interpreter as a potential problem trigger, this participant opts for omitting the title. The cognitive effort he experiences is visible in the surface structure indicators of processing problems: the pause (marked in the transcript by a double slash “//”) and the hesitation marker:

(1)

ST (T2):⁹ szanowni państwo / przypadl mi wielki zaszczyt wygłoszenia inauguracyjnego wykładu dla was studentów i absolwentów szkół artystycznych / był to pomysł Rektora Akademii Sztuk Pięknych we Wrocławiu Jego Magnificencji Profesora Jacka Szewczyka / bardzo dziękuję za to zaszczytne zaproszenie

TT (P22): ladies and gentlemen eee I have the great honour to present this inaugural lecture for you / the students and graduates of Artistic Schools // it was the idea of the vice-chancellor of the Academy of Art in Wrocław eee* professor Jacek Szewczyk / I would like to thank him very much for this kind invitation

RC1: Translating the first fragment I omitted the words ‘His Magnificence’ because I wanted to focus all my attention on a correct rendering of his name and surname. I assumed that this kind of titles are not that important for the Americans.

In the above example, the analysis of the recording and the measurement of the length of the EVS reveal that the anomalous pause of 2.5 seconds coincides with the occurrence of the item “Jego Magnificencji” (“of His Magnificence”). There is also a hesitation marker “eee” immediately before the title and the

⁹ The examples from the corpus of the study (source texts, target texts and retrospective comments) are coded in the following manner: the number of a source text (T2 in this case), the coding number of a given participant (P22 here), and the number of a successive retrospective comment within each output (RC1 in this example).

¹⁰ The asterisk always marks the segment of the target text to which the retrospective comment following it refers.
name of the rector which can also be attributed to the increased cognitive effort when focusing on the correct rendition of non-contextual information (the proper name in this case).

In this study, the assumption has been made that the indicator might not necessarily appear in the segment that proves to be problematic, but in line with Gile’s (1995, 2009) idea of failure sequences it might occur at a distance, for instance in the subsequent segment. The example (2) below shows problem indicators (hesitation markers and false starts) not only in the very segment in which the interpreter experienced problems with coordinating the efforts of listening, memorising, and production involved in SI, but also in the subsequent segment, a considerable part of which was omitted as a result of this increased cognitive effort.

(2)

ST (T2): aby sztuka była sztuką przez wielkie S musi się stać częścią rozwoju naukowego i technologicznego / albo nawet ten rozwój wyprzedzać / tak jak było w renesansie / włoscy artyści byli wtedy jedynymi naukowcami / dlaczego nie może tak być dzisiaj? / to tylko zależy od świadomości i postawy artystów / macie łatwy dostęp do każdej dziedziny wiedzy i codziennie około szesnastu wolnych godzin

TT (P25) to make your art the art written with capital A it it has to be a part of / of development or even to be faster than the technological development /* I want you to be also scientists as artists were in the past why cannot it be like this today? it only belyyy / the artist has the access to yyy you’ve got the access to all realms of knowledge

RC20/P25/T2/A-B: In the fragment about scientists I missed the very beginning and I omitted it. I had no opportunity to go back to that, so I only mentioned the scientists. Simplifying it was my way of saving the fragment. Also the next sentence wasn’t translated because I didn’t manage to reenter the text after this terrible failure at the beginning. This omission wasn’t a deliberate, conscious decision.
In this particular case (example 2 above), the failure sequence is explicitly reported by the interpreter, but in fact the corpus abounds with examples in which the product analysis of the transcripts and recordings suggests that the indicator might be a direct result of a problem reported in the preceding part of the text.

As shown in Figure 2, 19% of the retrospective comments reporting on the cognitive effort do not coincide with problem indicators in the corresponding target-text segments. The use of strategy is either reported explicitly during retrospection or it is clearly visible in the product (the interpreting output). It is the case with example (3), in which there is no indicator of processing problems in the target text, but the retrospective protocol reports on the use of the preventive strategy of reordering information:

(3)

ST (T2): w czasie mojego życia byłem świadkiem powstania kolorowej / a później HD / telewizji / elektrycznej gitary / komputera stacjonarnego / laptopa / DVD / generalnie elektroniki i technologii cyfrowej / pomijam niezwykłe osiągnięcia w wielu dziedzinach nauki

TT (P24): throughout my life I have seen / I have seen a television in colour / I have seen HD television / electric guitar computer laptop / DVD / I have seen all these things being created / invented* / I do not mention many other great achievements

RC19: Here the speaker started enumerating a lot of things and I couldn't remember all of them. That's why I decided to omit the part in which he says they were invented. Instead, to catch up with the speaker, I quickly enumerated all these things saying it practically at the same time as the speaker. And only at the end of the list I added that he witnessed them being invented. That's why I added it at the end. I suppose it sounds a bit awkward but thanks to this decision I managed to enumerate all these things although it was very difficult for me. If I hadn't done that, I guess I would have forgotten at least some of them.
In the above example, the interpreter decided to shift part of the information to the end of the segment. In this way she avoided an excessive load on short-term memory when dealing with this fragment of the text. Enumerations are considered to be one of the major problem triggers in interpreting since they are vulnerable to lapses of attention due to their high density of information and inherent signal vulnerability (given that the individual items are difficult to recover from the context). Changing the order of elements in an enumeration is one of the preventive tactics described by Gile (2009), who attributes its efficiency to saving the processing capacity. He points out that “by reformulating the last elements first, it is possible to pick them up before they have been processed in depth and integrated fully into the semantic network” (Gile 2009: 205–206).

Reported processing problems are often masked by the use of explicitation in the form of physical addition of extra elements (e.g., reiteration of lexical items, addition of modifiers and qualifiers, meaning specification) (see Gumul 2017). In the subsequent example, the interpreter resorts to adding a proposition recoverable from the context which acts in this case as a filler and allows her to allocate more time to the listening and analysis effort:

(4)

ST (T4): niektórzy prawnicy interpretują to nawet tak daleko że w zasadzie każdy lekarz ze względu tylko na to że wykonuje ten szczególny zawód zaufania publicznego / jest w pewnym sensie gwarantem bezpieczeństwa zdrowotnego każdego obywatela będącego w potrzebie / nie wszyscy prawnicy z tym się zgadzają

TT (P71): some lawyers interpret it that in fact each doctor only because he is a doctor* he fulfils this specific public trust functions in some sense a guarantor of health for each citizen in need / but not all lawyers agree with this

RC15: I needed a while to think about how to translate the expression ‘zawód zaufania publicznego’. That’s why I added ‘only because he is a doctor’. It doesn’t change the meaning of the text, but it gave me some time to think.
The retrospective protocols analysed in the study differ substantially as to the reporting of an indicator. More than half of the reports verbalising problems with increased cognitive load (223 comments) do not mention any indicator. These interpreters describe experienced difficulties, but do not comment on how these problems affected the target text and do not refer explicitly to the resulting disfluencies, omissions or errors. As we can see in Figure 3, 47.7% of retrospective comments mention problem indicators. Omissions are the most often mentioned indicator (72 comments – see Table 2). Omissions are also often mentioned along with another indicator – 19 out of 23 comments describing multiple indicators list omission as one of them. The second most frequently reported indicator is the category of unfilled anomalous pauses, which amount to 57 instances.

Figure 3
Indicators reported in retrospection vs. indicators not mentioned
Table 2
Indicators found in text segments referred to in retrospection

<table>
<thead>
<tr>
<th>Types of indicators</th>
<th>Reported indicators</th>
<th>Indicators identified in TTs (reported and unreported)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omissions</td>
<td>72</td>
<td>131</td>
</tr>
<tr>
<td>Pauses</td>
<td>57</td>
<td>97</td>
</tr>
<tr>
<td>Hesitation markers and false starts</td>
<td>19</td>
<td>71</td>
</tr>
<tr>
<td>Repairs</td>
<td>21</td>
<td>27</td>
</tr>
<tr>
<td>Grammatical errors</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Mispronunciations</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Multiple indicators</td>
<td>23</td>
<td>79</td>
</tr>
<tr>
<td>TOTAL</td>
<td>203</td>
<td>426</td>
</tr>
</tbody>
</table>

The following example is the case of reporting an unfilled pause, which the interpreter attributes to her fatigue. This retrospective comment, which might attest to the tightrope hypothesis, reports explicitly on the indicator of increased cognitive load, which in this case is not just a matter of prolonged ear-voice span, but involves a considerable omission:

(5)
ST (T4): lekarz powinien odmówić niektórych działań i niektórych czynności właśnie ku pożytkowi chorych / medycyna partnerska stwarza pewien problem i ja przedstawie Państwu moją interpretację i proszę żebyście się Państwo ze mną nie zgodzali / bo ja się mogę mylić / medycyna partnerska wymaga uzgodnienia sprawy między lekarzem i pacjentem

TT (P49): doctors should sometimes refuse to help people there are some there are some situations / in which they should do that / partnership medicine causes certain problems /* partnership medicine requires an agreement between a patient and a doctor

RC9: I omitted the fragment in which the speaker explains the
concept of partnership medicine. I decided it wasn’t very important and preferred to wait for the next fragment which is visible in the long pause I made. I did it because I was very tired.

Another example, in which the cognitive effort indicator is explicitly reported, shows how the competing efforts of listening and analysis on the one hand, and production on the other lead to a grammatical mistake, which is apparently more of a slip of the tongue. In this text segment there are multiple processing problems indicators. Apart from the incorrect grammatical form, there are five hesitation markers which, judging from the EVS measurement, can be directly attributable to mismanagement of processing capacity while searching for an equivalent of a problematic lexical item:

(6)

ST (T4): no i druga sytuacja w której jest zerwana ta nić porozumienia zaufania szacunku wzajemnego / przede wszystkim wtedy kiedy pacjent po prostu nie akceptuje lekarza / wówczas / moim zdaniem / lekarz ma pełne prawo do tego żeby nie podejmować się leczenia takiego pacjenta

TT (P46): and the other situation is eee when the bond between doctor and a patient is yyy closed the situation yy in which the patient don’t respect trust* a doctor / and a doctor can refuse eee curing aaa patient

RC10: Here I was desperately trying to find an equivalent for ‘zerwana nić’ [lit. ‘broken thread’ here in the figurative meaning of a mutual understanding] and it took me so much time that I lost the description of this ‘nić’. I tried to compensate for that in the second part of the sentence, in which, instead of acceptance for a doctor, I talk about respect and trust, but unfortunately completely ungrammatically because I was still thinking about how to say ‘zerwana nić’ in English.

The problems experienced by this interpreter can be explained in terms of Gile’s (1995, 2009) Gravitational Model of linguistic availability, which assumes that lexis, apart from the most basic high frequency vocabulary, belongs to the variable part of the
language resources stored in the memory and is, therefore, less accessible under the time constraint of SI and therefore requires more processing capacity. Thus, the interpreter might need more time to access certain lexical items, as was the case in the example above.

As indicated in the section specifying the aims of the study, I have not made any quantitative analysis of occurrences of problem indicators which were not mentioned in the retrospective protocols. This is obviously partly due to the size of the corpus. Moreover, such cases are not easily quantifiable, as a given instance of experiencing an increased cognitive effort might give rise to multiple problem indicators. I believe that counting these individually might blur the proportion between reported and unreported cognitive load. In fact, such an attempt has been made by Englund Dimitrova and Tiselius (2014). Their research design facilitated the task as they analysed a small corpus. Moreover, the unit of analysis in their study was a processing problem, which is more easily quantifiable and identifiable at the micro level, relying only on the product rather than the constructs of failure sequences or competing efforts used for the purpose of analysis in the present study. It has to be stressed that the finding of Englund Dimitrova and Tiselius (2014), namely that there is a substantial amount of processing problems that remain unreported, has been confirmed in this study. Although no attempt has been made to obtain the exact figures, it is evident that the corpus abounds with such examples. There are numerous cases of unreported problems possibly related to cognitive effort, which appears to confirm the results obtained by Englund Dimitrova and Tiselius (2014).

Finally, it is worth mentioning that we cannot rule out that at least part of the indicators are most probably mentioned in the retrospective protocols due to the influence of the cue. Since the cueing was done via the product, i.e. the interpreters were exposed to the recordings of their own outputs in order to stimulate memory during the retrospective session, there is a likelihood that not all of them reflect a conscious train of thought
during the interpreting process, but have been spotted by the interpreters while listening to their outputs. Some interpreters openly admit it and such cases were not taken into account (the category of RC post and I), but there might be more cases of a posteriori comments, which is attributable to the inherent weakness of the method and cannot be avoided.

7. Conclusions

The results of the study imply that increased cognitive load involved in simultaneous interpreting and the resulting cognitive effort experienced by the interpreters might not necessarily be manifested in the product, i.e. the target text. The fact that the majority of the reports of cognitive effort correspond to processing problems indicators is a fairly predictable result. However, the absence of indicators in the outputs of the advanced interpreting students constitutes an interesting finding. The difference between the results of this study and the one by Englund Dimitrova and Tiselius (2014) implies that the higher proportion of effectively masked cognitive effort may be due to the mastery of the technique of interpreting, which entails the use of efficient coping and preventive tactics. In this way it could possibly be seen as directly attributable to the level of expertise and experience. Nevertheless, in order to corroborate such claims we need further large-scale research on professional interpreters.

References


Sources


Ewa Gumul
ORCID iD: 0000-0003-0859-6067
University of Silesia in Katowice
Institute of Linguistics
Gen. Stefana Grota-Roweckiego 5
41-205 Sosnowiec
Poland
ewa.gumul@us.edu.pl