UNIWERSYTET ŚLĄSKI
W KATOWICACH

WYDZIAŁ HUMANISTYCZNY

INSTYTUT JĘZYKOZNAWSTWA

MARCELINA PIETRYGA

NUMER ALBUMU: 8919

THE INFLUENCE OF TRANSLATION DIRECTIONALITY ON THE COGNITIVE EFFORT OF TRAINEE TRANSLATORS

ROZPRAWA DOKTORSKA

Rozprawa doktorska napisana pod kierunkiem

dr hab. Ewy Gumul, prof. UŚ

KATOWICE 2024

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my supervisor, Dr Hab. Ewa Gumul, Prof. UŚ, without whose invaluable help this PhD thesis would have never come to the final shape. Thank you for your support, guidance, time, patience, advice, inspiring discussions, and always finding time to answer my endless questions. I have learned a lot from you during all these years.

I am also very grateful to my two tutors, who helped me with experimental design and data analysis. I would like to thank Dr Hab. Agnieszka Chmiel, prof. UAM, for guiding me into the eye-tracking world, and Dr Hab. Agnieszka Turska-Kawa, prof. UŚ, for all the help with the statistical issues.

I would also like to express my gratitude to Dr Maciej Bożek, the Head of the Research Laboratory of the Institute of Psychology at the University of Silesia in Katowice. Thank you for your consent to conduct the study in the Laboratory and your supervision and help during the experiment.

I am very grateful to Prof. Dr Alexis Hervais-Adelman for proofreading Section 2.6. Your insightful comments helped me to improve its content.

A huge thank you to all the participants and experts who agreed to take part in this project. Without your participation, I would not have been able to conduct this research.

I would also like to express my gratitude to Dr David Schauffler for proofreading this thesis. Your comments allowed me not only to improve my work but also to look at the text from a different linguistic angle.

Last but not least, I am grateful to my closest ones, my parents, grandparents, Michał and my friends. Thank you for always believing in me and in this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Directionality denotes translating either into someone's native language (L1) or foreign language (L2) (Whyatt 2019: 79). In 1988, a well-known translation researcher, Peter Newmark, stated that translators should work solely into their L1. There was also a belief that L1-L2 translation evokes significantly higher cognitive effort (Fonseca 2015; Whyatt 2018). It can be observed that the Golden Rule of Translation (Newmark 1988) is usually followed in countries where high diffusion languages, like English or Spanish, are the official languages (Ferreira et al. 2021). However, translators working in countries where languages of low diffusion are spoken, like Poland, the Czech Republic and Croatia, frequently admit working into their L2 on a daily basis (Pavlovič 2007; Whyatt & Kościuczuk 2013; Mraček 2018). Furthermore, some even claim that it is their preferred translation direction. This practice is reflected in translation courses, as L1-L2 translation direction is an important part of curricula in many Central Europe universities (Pavlovič 2010; Gumul 2017a). In fact, translation trainees also frequently admit that they prefer to work into their L2 (Pietryga 2022). These observations suggest that L1-L2 translation does not always evoke higher cognitive effort, and even translation trainees are able to overcome the L2-L1 translation advantage. Therefore, various researchers have attempted to verify whether working into L2 is more cognitively demanding. However, it seems that the question remains open. Even though some studies unequivocally concluded that the L1-L2 direction evokes higher cognitive effort (e.g., Buchweitz & Alves 2006), frequently, some of the analysed variables did not reach the level of statistical significance (e.g., in works by Pavlovič & Jensen 2009; Whyatt 2019; Hunziker Heeb 2020), or the results indicated L2-L1 as a more effortful direction of translation (e.g., in work by Ferreira et al. 2021). These inconclusive results lead to a significant research gap in research on cognitive effort and directionality.

The aim of this study is two-fold. The primary objective is to investigate the influence of directionality on cognitive effort. The secondary objective is an attempt to add new variables to the research on cognitive effort and directionality. To the best of my knowledge, variables like eye-key span and the weighted rating of the NASA-TLX have been proven as valuable indicators of cognitive effort. However, they have not been analysed in the context of directionality before. I also decided to analyse directionality based on translating collocations. The preliminary study (Pietryga 2022) revealed that collocations are believed to be crucial problem triggers regardless of the translation

direction. Moreover, advanced knowledge of collocations is a crucial skill for translators (Pellicer-Sánchez et al., 2022; Sonbul et al., 2022). Since translating collocations has not been analysed in the context of directionality and cognitive effort before, there is a significant research gap that I would make an attempt to fill.

I adopted a mixed-methods approach using both subjective and objective methods. The data were collected by eye-tracking, keylogging, retrospective verbal reports and two questionnaires: NASA-TLX and a self-designed questionnaire. Such triangulation of methods allowed me to study the influence of directionality on cognitive effort both globally and locally. The data gathered in this way made it possible to answer research questions I formulated in this study. These were:

- How does directionality influence cognitive effort?
- How do the three stages of the translation process (orientation, drafting, revision) differ in each direction?
- How do participants describe cognitive effort related to translating collocations in both directions?
- How does directionality influence translation accuracy?

This thesis is structured into two parts. Chapters 1-3 provide a review of the literature on directionality, translation process research and cognitive effort. The empirical part consists of Chapters 4-7. Chapter 1 opens with various definitions of directionality. Then, I provide an outline of attitudes towards directionality and focus on the assumptions of the Golden Rule of Translation (Newmark 1988). Subsequent sections raise the issue of languages of low diffusion and low resources. The chapter closes with the phenomenon of translation asymmetry and the Revised Hierarchical Model (Kroll & Steward 1994).

Chapter 2 covers the topic of translation process research. The presentation of the most popular process methods and the types of data collected using these methods allows me to refer to them in the next chapters when discussing the issue of directionality and cognitive effort and elaborating on the study design. I begin the discussion of process methods with verbal reports. Then, I move on to keylogging and eye-tracking. I also focus on the newest neuroimaging methods, like EGG, PET and fMRI. The chapter closes with a short discussion of questionnaire studies. Additionally, each section raises the issue of the advantages and drawbacks of each method.

Chapter 3 discusses the issue of cognitive effort. It opens with definitions of two phenomena that are the subject of high-tension debate: cognitive effort and cognitive load (Gieshoff & Hunziker Heeb 2023). I also refer to the three most popular models of cognitive effort and cognitive load: the Effort Model (Gile 1995), the Cognitive Load Model (Seeber 2011), and the Cognitive Load Theory (Sweller et al. 1998). The next section presents the indicators and measures of cognitive effort developed by Chen et al. (2012) and studied by Ehrenberger-Dow et al. (2020). The main part of the chapter is devoted to the outline of the current state of research on cognitive effort and directionality. I also draw attention to the decreased level of cognitive effort resulting from the default translation (Halverson 2019).

The main objective of Chapter 4 is to discuss the research design of the experimental study. It opens with presenting the main aims and motives that led me to conduct the study. I also like to emphasise the differences between the experimental and quasi-experimental research designs. However, following the common practice (e.g., in Korpal 2016b), I will refer to my study as an experiment. In the next sections, I elaborate on the formulated research questions and adopted hypotheses, grounding them in the literature on directionality and cognitive effort. Independent and dependent variables are the subject of the subsequent sections. I also describe the materials and methods used in the study and provide details about the participants. Since the opinion of the Ethics Committee was required, I devoted Section 4.3.7 to a summary of ethical issues. The chapter closes with the presentation of the experimental procedure.

Chapter 5 presents the analysis of the results. It is structured into six sections. Five of them correspond to five adopted hypotheses. Each section opens with descriptive statistics results. I use histograms to present the distribution of the variables, and when possible, I discuss kurtosis and skewness. Then, I move on to inferential statistics. I conduct either paired t-tests or Wilcoxon tests. Each section closes with a summary of the results and confirmation or rejection of the adopted hypothesis. The last section of Chapter 5 is devoted to the linear regression results.

The aim of Chapter 6 is to discuss the results and provide answers to the research questions. It is structured into four sections that correspond to the four research questions. Each section includes a summary of the results as well as their interpretation in the context of assumptions on cognitive effort and directionality.

The empirical part closes with Chapter 7, which presents the concluding remarks. I also briefly discuss the limitations of the study and potential avenues for further research.

Chapter 1. Directionality

The aim of Chapter 1 is to review the literature on the notion of directionality. It begins with an overview of various definitions. Next, I discuss the notions of native tongue, mother tongue, and foreign language. Section 1.2 presents an overview of attitudes toward directionality over the years. It starts with the perception of directionality in ancient translations and the Middle Ages, then moves on to the approach of Martin Luter. Subsequently, the Golden Rule of Translation proposed by Newmark (1988) is discussed. Section 2 ends with the contemporary approach to the notion of directionality. In Section 3, I focus on the phenomena of languages of low diffusion and low resources. Particular attention is also devoted to translation practices among languages of low diffusion. The chapter closes with a discussion of the Revised Hierarchical Model (Kroll & Steward 1994), which provides the grounds for linguistic asymmetry, which is frequently discussed, referring to directionality.

1.1 Definition of directionality

To gain a full picture of issues related to directionality, it is necessary to define it first. There have been many attempts to define directionality, for example, by Beeby Lonsdale (2001), Pavlovič (2007a), Chen (2020), and Rodríguez-Inés (2022). All the definitions appear very similar, as they draw attention to translating in two directions, into L1 and L2[1]. In my study, I will follow the definition developed by Whyatt (2019, p. 79), who claims that directionality embraces "work[ing] into their [translators] first or "native" (L1) language or out of their L1 and produc[ing] translations into their "first foreign" language (L2)." The quoted definition indicates that there are several ways to describe translating into L1 and L2. Translating into L1 is usually referred to as direct translation or translating into the native language[2]. L2 translation functions also under the terms inverse translation, or translating into a foreign language (e.g., Pokorn 2011; Fonseca 2015; Ferreira & Schwieter 2017; Ferreira & Schwieter 2017; Mraček 2018; Whyatt 2019).

[1] It should be noted that the terms L1 and L2 did not originate in the field of translation process research. As pointed out by Hunziker Heeb (2020, p. 33), they are "adopted from the field of second language acquisition and commonly reflect the chronology of acquisition." Hammarberg (2014, p. 5) explains that "L1 and L2 are defined through the priority/posteriority distinction. L1, being encountered first, develops as the original system, and L2 is subsequently added to the already established L1."

[2] There are also names like the A language (L1) and the B language (L2) that are used in interpreting studies (e.g. Bartłomiejczyk 2004; Gumul 2017a, Hunziker Heeb 2020, Bartłomiejczyk & Gumul 2024).

Terms like mother tongue, foreign language, and native speaker are also worth discussing since they may cause misunderstandings (Pokorn 2005; Rodríguez-Inés 2022). Pokorn (2005: 1) suggests that often, not enough attention is paid to the scope of their meaning. In fact, they are getting vague in times of globalisation when it is relatively easy for people to travel, learn new languages, and change their place of living (Pokorn 2005; Ferreira et al. 2021; Rodríguez-Inés 2022). It is even more difficult to precisely define mother tongue, foreign language and native speaker when referring to people who have immigrated to foreign countries, especially in their early years. One should also not ignore the situation of linguistic minorities, who are frequently bilingual. Usually, they are able to speak fluently both the official language of the country where they live and their minority language (Pokorn 2005: 4, 14-16). These situations may have various results, like replacing one's first language with the official language of the new place of living, being bilingual in both languages or even completely forgetting the language of the country where a person was born (Pokorn 2005: 6-7). In the study by Ferreira et al. (2021: 127), 19 out of 20 participants assigned to the Spanish-dominant group were born outside the USA. However, at the time of the study, not only did they live in the USA, but also they were active translators in the English-Spanish language pair.

To provide some kind of solution to various approaches to the discussed terms, Pokorn (2005) presents the following definitions. A mother tongue "is not only the first language according to the time of acquisition but the first with regard to its importance and the speaker's ability to master its linguistics and communicative aspects" (Pokorn 2005, p. 3). The researcher also attempts to explain the fundamental differences between native and non-native speakers. Referring to previous studies, for example, by Coppieters (1987) and Long (1990), Pokorn (2005: 13-14) observes that although some foreign language speakers manage to get linguistic abilities almost identical to those demonstrated by native speakers, in most of the cases, non-native speakers may struggle with collocations and idiomatic expressions, and they may be prone to express their thoughts in L2 more explicitly than in L1.

Therefore, some researchers believe that descriptions like direct, inverse, foreign, or native should be avoided. They claim that, first of all, these names are not representative enough and, therefore, do not embrace the correct scope of each category. Secondly, they make the user subconsciously divide the directions into a better or more favourable direction and the more problematic one (Pavlovič 2007a; Whyatt 2019; Hunziker Heeb 2020; Rodríguez-Inés 2022). That is why, following researchers like

Pavlovič (2007a), Ferreira (2014), and Hunziker Heeb (2020), I will use the neutral terms: L1 and L2 to discuss translation directions.

1.2 An overview of attitudes towards directionality throughout history

Throughout history, the attitude towards directionality has undergone many changes. Currently, this topic is being studied by many researchers. Articles and entries related to directionality can be found, for example, in publications like the Routledge Encyclopaedia of Translation Studies (Baker 1998; 2001), the Handbook of Translation Studies (Gambier & van Doorslaer 2011), and the Handbook of Translation and Cognition (Schwieter & Ferreira 2017). Some grants have also been awarded to study this phenomenon. For example, the Poland National Science Centre founded the *EDiT* project (*Effects of Directionality in the Translation Process and Product)* conducted by Bogusława Whyatt in 2016-2019[3]. The project entitled *Directionality in translation qualitative and sociological aspects*, conducted by David Mraček, Tomáš Duběda, and Vanda Obdržálková received funding from the Czech Science Foundation (Mraček 2018: 220). Thus, it can be observed that the major interest in this topic appeared in the 21st century. Nevertheless, the practice of translating not only into L1 but also into L2 can be actually traced back to ancient times (Beeby Lonsdale 2001; Pokorn 2005; Duběda & Obdržálková 2021).

1.2.1 Brief history of attitudes towards directionality

As pointed out by Beeby Lonsdale (2001: 64-65) and Pokorn (2005: 34), L1-L2 translation did not raise any concerns and was widely practised until the end of the Middle Ages. The researchers give examples of the first Bible translations from Greek into Latin, done by speakers whose L1 was Greek rather than Latin. They also enumerate L1-L2 translations of Buddhist sacred texts. According to the researchers, L1-L2 translation was a common practice also in the Middle Ages. Beeby Lonsdale (2001: 65) draws attention to translations of texts from Arabic and Jewish, frequently translated via a third language, done in collaboration by translators with various L1s.

However, the attitude towards directionality changed after Martin Luther announced his 95 theses (Beeby Lonsdale 2001; Pokorn 2005; Whyatt & Kościuczuk 2013; Kościuczuk 2016). Probably, it was the first moment in the history of directionality

[3] source: https://projekty.ncn.gov.pl/index.php?projekt_id=293638, visited 26 January 2022

when doubts were raised about the accuracy and validity of L1-L2 translation. Pokorn (2005, p. 25) describes the impact of Luther's statements as follows:

> it seems to have been Martin Luther who, in defence of his translation for the first time, explicitly considered his knowledge of the TL [target language] as a decisive advantage over his critics (Luther 1993:18-22), which led many of his readers to the conclusion that one can translate satisfactorily only into one's own language. Luther's conviction was taken over and strengthened by the first and second, nationalist generation of Romantic authors, who also made a great contribution to the rise of national philologies.

Nevertheless, this view did not prevent translators and interpreters from working into their L2. According to Beeby Lonsdale (2001: 65), the practice continued for the following centuries, for example, in the case of Latin, the language frequently used in science. Later, L1-L2 translation and interpreting were frequently practised in countries that either wanted to have some control over their translations, like the former USRR or wanted to expand their relationships with other countries, as in the case of the Middle Eastern countries (Gumul 2017a; Hunziker Heeb 2020; Bartłomiejczyk & Gumul 2024). However, the belief that translators and interpreters should work only into their L1 to produce a stylistically correct text managed to take root in translation theory. It was favoured by the Paris School (Seleskovitch and Ledered 1989) and appeared systematically in many writings, such as Delille, Humbolt, and Hugo (Pokorn 2005: 30-31).

1.2.2 The Golden Rule of Translation

While Martin Luther's opinion had laid the foundations for favouring L2-L1 translation, it is probably the statement of the well-known, contemporary opposer of L1-L2 translation, Peter Newmark, the formulator of the Golden Rule of Translation, which had the most significant influence on this aspect of translation practice and research. Writing in 1988, in his Textbook of Translation, Newmark states that "I shall assume that you, the reader, are learning to translate into your language of habitual use since that is the only way you can translate naturally, accurately and with maximum effectiveness" (Newmark 1988, p. 3). Moreover, in an earlier publication, Newmark assessed the L2 abilities of translators even more radically. He claims that the translator "usually knows that he cannot write more than a few complex sentences in a foreign language without writing something unnatural and non-native" (Newmark 1981, p. 180). He continues by

saying that the aspects that will definitely be spotted as unnatural in L1-L2 translation are usually collocations. Although this problem has already been mentioned as one of the features distinguishing native speakers from non-native ones, it should be remembered that bilingualism is enumerated as only one of the translation competencies required from translators (e.g., PACTE 2000, 2003). Hunziker Heeb (2020: 34) observes that although bilingualism is included in many models of translators' competencies, in fact, none of the models emphasises in which direction translations should be performed. What is more, most translators have received more linguistic training than average foreign language speakers. However, according to Kościuczuk (2016: 5), opinions supporting the supremacy of L2-L1 translations can also be found elsewhere. He claims that even works published by Chomsky, in which he studies the features of the native language, may influence the belief in the superiority of L2-L1 translation.

As observed by Pokorn (2005: 26), the Golden Rule of Translation by Newmark had such a strong impact on translation literature discussing directionality and translation practice that it has subsequently been repeated in many other textbooks that are used in translation courses. In his article, Kościuczuk (2016: 5-6) enumerates many contemporary textbooks, all of them published after the year 2000, the authors of which favour L2-L1 translation. In fact, even Venuti (1995) points out that according to the previous traditional axiom, translations should not indicate that the text has a foreign origin, let alone being the outcome of the L1-L2 translation process (Pokorn 2005: 33). Thus, many researchers (e.g., Bartłomiejczyk 2004; Pokorn 2004, Pavlovič 2007a; Fonseca 2015; Kościuczuk 2016) observe that there is still a conviction that both translators and interpreters should work only into their L1. Kościuczuk (2016, p. 4) refers to an argument that "[i]t is often believed that native speakers know their L1 perfectly well and will always outperform non-native users of this language." For example, L2-L1 translation and interpreting are practised in international organisations like the European Union and the United Nations (Biel 2007; Chmiel 2016; Buchowska 2017; Gumul 2017a, Graves et al. 2022; Bartłomiejczyk & Gumul 2024).

A study whose objective was to verify if native speakers are able to distinguish translations performed by natives (L2-L1 translation) from translations performed by non-natives (L1-L2 translation) was conducted by Pokorn (2004). The study involves three types of translations into English. Two of them were performed by translators whose L1 was Slovene; thus, they performed L1-L2 translation. Subsequently, two translations were performed by translators whose L1 was English, so they performed L2-L1

translation. Finally, three pairs of translators produced three other target texts (TTs). Each pair consisted of one person whose L1 was English and one person with high knowledge of Slovene (these were a native speaker, an immigrant, and a philologist). The author of the study observed that in the case of the first two types of translations (i.e., texts translated by one person), only 56% of respondents managed to correctly distinguish whether it was L1-L2 or L2-L1 translation. Moreover, the participants also had some problems estimating the number of people involved in the translation process. In the case of translations performed as a collaboration of two translators with different L1s, only 10% of participants correctly decided that more than one person translated the text. Thus, the results demonstrate that even if translators work into their L2, the results of their work are often indistinguishable from L2-L1 translation. This study by Pokorn (2004) provides firm evidence that L1-L2 translation should not be undermined or rejected and ensures its high quality and accuracy.

1.2.3 Current perception of directionality

Currently, the attitude towards directionality is two-fold. On the one hand, both Pokorn (2004) and Kościuczuk (2016) observe that some translation agencies and some translators themselves still claim that translations should be done only into L1. As has already been mentioned, in many well-known organisations, translators usually work into their L1 (Bartłomiejczyk 2004; Whyatt & Kościuczuk 2013; Gumul 2017a; Bartłomiejczyk & Gumul 2024). However, as observed by Kościuczuk (2016, p. 7)

> [w]ith such a strong bias against [L1-]L2 translation in the EU, it comes as a surprise that neither EN 15038 (European Quality Standard for Translation Service Providers) nor ISO 17100: 2015 (Translation services – Requirements for translation services) requires that the translator's target language is their native language.

Nevertheless, Rodríguez-Inés (2022) observes that some organisations associating professional translators, like ATRAE (Asociación de Traducción y Adaptación Audiovisual de España), completely preclude the possibility of translating into translators' L2.

Ferreira et al. (2021: 119) point out that the stigmatisation of L1-L2 translation also influenced the field of translation process research. According to them, despite the growing interest in the topic of directionality, which can be observed based on the number of works devoted to L1-L2 translation, this phenomenon has still not been thoroughly

analysed. There are many aspects that have not been answered and discussed yet. It is worth mentioning that nowadays, translation scholars do not approach the topic of directionality with any bias, but rather, they make attempts to examine how the translations in the two directions differ based on different variables. As pointed out by Apfelthaler (2019, p. 156),

> [t]oday, driven by empiricism, it [research on directionality] has adopted a much more descriptive discourse, albeit often with a somewhat emancipatory message, to ensure that non-native translation and interpreting – and their protagonists receive the objective, the fair treatment they deserve. The general goals, topics, methods and theoretical frameworks of directionality have converged with those of other areas of translation studies to the benefit of the discipline as a whole as well as non-academic stakeholders, given the potential impact of its findings and its responsibility to address the needs of the wider society.

Thus, Ferreira and Schwieter (2017, p.102) believe that "[i]t seems that we have come to the point in the development of translation studies in which old beliefs on directionality can no longer be accepted." Therefore, directionality is studied in various contexts. For example, Korpal and Jankowiak (2021) analysed the influence that each direction may have on emotions during interpreting. Whyatt et al. (2021) focused on potential differences in L1-L2 and L2-L1 translation when using online resources. Tomczak and Whyatt (2022) studied directionality in the context of lexical selection in professional translators. Bartłomiejczyk and Gumul (2024) analysed disfluencies in interpretations into A and B languages performed by professional interpreters in the European Parliament.

Currently, the Golden Rule of Translation is being challenged by the needs of the translation market (Whyatt & Kościuczuk 2013: 66). Scholars unanimously agree that there is a demand for L1-L2 translation, which can no longer be ignored (Ferreira 2014; Ferreira et al. 2016). This tendency can be observed based on various studies in which professional translators admit that they regularly receive commissions for L1-L2 translation (e.g., Pavlovič 2007a; Mraček 2018; Rodríguez-Inés 2022). Detailed numbers will be presented in the next section.

The importance of L1-L2 translation practice is emphasised by both translation scholars and translation students. Therefore, attention to L1-L2 translation is paid at the university level. In some countries, like Poland, translation course curricula offer both L2-L1 and L1-L2 translation exercises (Pavlovič 2010; Gumul 2017a; Rodríguez-Inés

2022). In the study by Pietryga (2022), one of the questions in the questionnaire filled in by students was: *Do you believe it is important to practice L1-L2 translation?* 90 out of 93 students answered affirmatively to this question. Some of them also pointed out the market's needs. Pavlovič (2007a), Whyatt and Kościuczuk (2013) and Pietryga (2022) also asked about participants' opinions on the Newmark's Golden Rule of Translation. In the case of students participating in the study by Pietryga (2022), 23.1% of respondents (21 students) firmly disagreed with this rule. Interestingly, the longer the span of translation training the students receive, the higher the number of those who disagree with Newmark's rule. Answers to questionnaires submitted by professionals (Pavlovič 2007a; Whyatt & Kościuczuk 2013) also show strong disagreement with the so-called traditional axiom. However, there were many opinions, especially in the study by Pavlovič (2007a), which supported it. Interestingly, the study by Bartłomiejczyk (2004) devoted to interpreting revealed that students who thought that one of their interpreting directions yielded better quality, in almost 50% of cases, named the L1-L2 direction. However, the answers submitted in the questionnaire were not compared to the results of students' work at that point.

1.3 Languages of low diffusion

The belief that one should avoid translating into L2 prevailed among the researchers whose L1 was English and who analysed translation markets dealing with major languages (Kościuczuk 2016: 6). As discussed by Ferreira et al. (2021: 127), in the case of the USA, L2-L1 translations constitute about 60% of commission for both English and Spanish. Nevertheless, translation trends look quite different in countries like Poland and Croatia, where L1-L2 translation constitutes a large part of the translation market. In the case of Croatia, almost all translators who admitted working into L2 claimed that this direction constituted a majority or near majority of their commissions (Pavlovič 2007a: 86). On the Polish market, the amount of commissions into L2 varied from 5% to 100%, depending on the translator (Whyatt & Kościuczuk 2013).[4] However, unlike English, Polish and Croatian are languages of low diffusion.

An attempt to define the concept of a language of low diffusion was made by Pavlovič (2007b p. 7), who claimed that it is "a language not widely used outside its primary linguistic community or frequently acquired as a second language." Whyatt and

[4] A detailed analysis of translation practices on various European translation markets is presented in Pietryga (2022).

Pavlovič (2021: 2-3) suggest that although such language is characterised as having a limited or low diffusion, the designation is not restricted to languages of minorities or threatened with the risk of extinction. As the researchers emphasise, for more than half of all languages used worldwide, the number of speakers does not exceed 1% of all people. For example, Polish, also classified as a language of low diffusion, is spoken by many users (Whyatt & Kościuczuk 2013; Ferreira & Schwieter 2017; Whyatt 2018). According to the EU data, there are almost 37.0 million inhabitants of Poland[5], and it can be assumed that the great majority of them use Polish daily. Apart from Polish, scholars also enumerate Hungarian, Slovene, Croatian, Czech, Danish, and many others as languages of low diffusion (Ferreira and Schwieter 2017; Whyatt 2018; Mraček 2019; Ferreira et al. 2021).

It has been observed that the criteria based on which languages are assigned to the low and high diffusion categories are partly connected with relations between languages (Kościuczuk 2016; Whyatt & Pavlovič 2021; Mikolič Južnič et al. 2021). This observation is explained in detail by Mikolič Južnič et al. (2021: 243) based on the literary translation market. The researchers discovered that the more frequently a given language is translated into other languages, the higher its status. In the case of literary translation, but very likely also in other genres, the first place is unquestionably occupied by English. According to Mikolič Južnič et al. (2021, p. 243), English "accounts for at least 40% of the world market in translated books (up to 70% of the European market)." French and German can be found in the second place. However, their translations constitute only about 10%. The rest of the languages get either up to 3% of the translation market or less than 1%. These data show the crucial position English has in the translation market. That is why Whyatt and Pavlovič (2021: 3) refer to it not only as the language of high diffusion but also as the language of unlimited diffusion. Sometimes researchers also refer to it as the contemporary lingua franca (e.g., Pavlovič 200a; Whyatt & Kościuczuk 2013; Mraček 2019; Rodríguez-Inés 2022). Interestingly, nowadays, the number of English foreign language speakers is several-fold higher than that of English native speakers (Whyatt & Pavlovič 2021: 2).

[5] source: https://european-union.europa.eu/principles-countries-history/key-facts-and-figures/life-eu_en, visited 30 April 2024.

1.3.1 Translation practices among languages of low diffusion

As already discussed, L1-L2 translation holds a special place among languages of low diffusion. Researchers frequently point out that, in this case, the practice of L1-L2 translation is not only frequent but also crucial or even inevitable (Pokorn 2005; Pavlovič 2010; Ferreira 2014; Duběda & Obdržálková 2021; Ferreira et al. 2021; Whyatt & Pavlovič 2021). There is a common belief that communities using languages of low diffusion would be unable to exchange information at the international level without L1-L2 translation (Pokorn 2005; Whyatt & Pavlovič 2021). Whyatt and Pavlovič (2021: 5) observe contemporarily high growth in the demand for translations into and from languages of low diffusion. The possibilities offered to the members of the European Union may influence this trend. For example, based on the no preference rule the documentation has to be translated into all EU languages (Biel 2007, 2016; Buchowska 2017).[6] Moreover, Bartłomiejczyk and Gumul (2024: 39) observed that native Polish speakers working in the European Parliament frequently work both in and out of their L1, as opposed to native English speakers who work solely into their L1. However, the demand for translating into and from languages of low diffusion also fluctuates due to migration waves. The growing interest in the languages of low diffusion also extends into the translation process research field. This can be illustrated, for example, by the fact that the Interpreter and Translator Trainer journal devoted a whole issue to the topic of languages of low diffusion.

Researchers observe that translations from languages of low diffusion into major languages are usually performed by translators whose L1 is the language of low diffusion and who translate into L2 (Whyatt & Kościuczuk 2013; Kościuczuk 2016; Mraček 2018; Whyatt at al. 2021). First of all, this results from the fact that there are few native speakers of major languages, like English, who not only live in countries like Poland, Croatia or Hungary but also work as translators in a given language pair. There are also other circumstances leading to this practice. Based on the observation of the Polish market, Whyatt and Kościuczuk (2013, p. 73) enumerate the following elements:

- Polish translators working into English are easier to recruit than the few English translators working out of Polish, and, as can be expected, more competitive rates are offered by non-native translators. In an age of austerity, such economic parameters gain significance.
- Translation clients believe that professional translators, due to their competence, are able to provide translations in either direction.
- Clients prefer to entrust work to their regular translators for reasons of confidentiality and their previous experience.

[6] The details of translating the EU law and documents are discussed, for example, in Biel (2016).

There is also a perception that when working into their L2, translators can be assisted by native speakers of this language (Pokorn 2005, 2011; Mraček 2019). A study which aimed to analyse the scope of such assistance based on the Czech translation market was conducted by Mraček (2019). Data regarding the cooperation between native and non-native speakers during and after the translation process were collected in the form of a pre-experiment questionnaire. The language pair was Czech, as L1, and English and French as L2s. The results indicated that only 8 out of 40 translators always or most of the time consult native speakers while translating. More than half of respondents (21 out of 40) claimed that they are sometimes assisted with such help. However, the number of translators who stated that L2 native speakers never assist them while translating is far from insignificant. As many as 11 out of 40 translators gave this answer. These translators who are assisted by L2 native speakers usually claimed to need help with issues like vocabulary and style.

However, the majority of respondents (35 out of 40) claimed that their TTs undergo a proofreading process. Twenty-one of them reported that the texts are proofread either always or most of the time, and 12 reported that their TTs are sometimes revised. Only two translators stated that their texts are not proofread. Nevertheless, the respondents revealed that the proofreaders are not always L2 native speakers. Sometimes, the TT is checked by a Czech native speaker with professional knowledge of a given L2 (English or French).

Questions related to the topic of proofreading were also asked in questionnaires by Pavlovič (2007a) and Whyatt and Kościuczuk (2013) in their studies analysing respectively Croatian and Polish translation markets. In Pavlovič's study, more than half of the respondents (31 out of 61) claimed that an L2 proofreader sometimes checks their texts. In contrast, 14 out of 61 respondents stated that their texts never undergo any revision process. The number of translators who state that their texts are revised either always (2) or, for most of the time (10) is relatively small. In the study by Whyatt and Kościuczuk (2013), who focused on the Polish translation market, almost half of the translators stated that their translations are sometimes revised by L2 native speakers and only 11 out of 55 claim that their TTs are always proofread. However, the respondents additionally suggested that their clients usually do not require such services.

1.3.2 Languages of low resources

Apart from languages of low diffusion, Whyatt and Pavlovič (2021: 4) also distinguish languages of low resources. As they continue, "the scarcity of resources, including human and digital resources, is a serious challenge to developing translation and other language technologies" (Whyatt & Pavlovič 2021, p. 4). Although sometimes a language can belong to both named categories, there are cases when languages of low diffusion cannot also be categorised as languages of low resources. An example discussed by the researchers is Czech. It is definitely a language of low diffusion; however, its aggregate of various online data is rather substantial.

It has also been observed that the number of online resources for major languages is larger than for languages of low diffusion, even if they do not belong to the category of low-resource languages (Kościuczuk 2016; Whyatt 2019; Whyatt et al. 2021). In this case, Mraček (2019: 211) draws particular attention to English, which has many online resources such as parallel texts.

An analysis of the influence of directionality on using online resources was conducted by Whyatt et al. (2021). The researchers observed that "the positive correlation between the time spent in the Internet browser and the total time needed to translate the experimental texts was slightly stronger for L2 than for L1 translation" (Whyatt et al. 2021, p. 12). It was also observed that the characteristics of searches differ in each direction. It appears that translators type into the browser more detailed queries when working into their L2. However, the researchers found that the direction of translation does not significantly influence the amount of time spent in online resources.

1.4 Revised Hierarchical Model

The differences between L1-L2 and L2-L1 translations are usually traced back to the idea of the bilingual lexicon and the manner in which words in both languages are stored and accessed in the bilingual mind. Researchers were particularly interested in whether there are two separate lexicons for each language or a common one for both of them (discussed, e.g., in Kroll & Steward 1994; Whyatt 2012; Chmiel 2016). It was found that the meaning is hierarchically stored in the bilingual mind. According to de Groot (2002, p. 33), "it consists of at least two layers of memory representations (or 'nodes'). One of the layers of nodes stores the meanings of words, and the second stores their forms." Thus, another question appeared: "whether the two languages are segregated or integrated at both levels or at one of the levels" (Whyatt 2012, p. 84). The Revised

Hierarchical Model developed by Kroll and Steward (1994) provided some answers to this question. However, before the Revised Hierarchical Model appeared, there were two main concepts for accessing L2 words. These were word association and concept mediation models by Potter et al. (1984).

1.4.1 Word association model and concept mediation model

In their study from 1984, Potter et al. formulated two hypotheses. The first one, the word association hypothesis, predicted that "new words in the second language (e.g., French) are directly associated with the words in the first language (e.g., English). As long as the second language remains weaker than the first one, this association is used in understanding and speaking the second language" (Potter et al. 1984, p. 23). The second, concept mediation hypothesis predicted that "second language words are not directly associated with first-language words, but instead are associated with the nonlinguistic concept common to two words" (Potter et al. 1984, p. 23). To verify these hypotheses, they conducted two experiments.

Participants of the first experiment were fluent bilinguals with Chinese as L1 and English as L2. They were presented with 96 words in pictures or written words in Chinese or English. Participants were assigned to two groups. The task of the first group was either to name or to translate the word appearing on the screen[7]. At the same time, the second group paired the same words with their superordinate categories. According to Kroll and Steward (1994, pp. 150-151), the word association and concept mediation models have the following prediction-related word translation and picture naming tasks:

> the word association model [hypothesis] predicts that translation from the first language (L1) to the second (L2) should be faster than picture naming in L2. The model assumes that translation from L1 to L2 relies on lexical links and can thus bypass conceptual access. However, picture naming, which requires conceptual access, should first be mediated through conceptual memory and through the first language; only then can the link from L1 to L2 be retrieved. The concept mediation model [hypothesis] predicts that translation into L2 and picture naming in L2 should be similar because both require conceptual access prior to retrieval of the L2 word.

The results confirmed the predictions of the concept mediation hypothesis for fluent bilinguals. Naming words in L2 was not significantly longer than the L1-L2 translation task. It means that words in both languages are directly associated with the corresponding

[7] The task of naming or translating single words is a common practice in studies discussing the Revised Hierarchical Model. According to Potter et al. (1984, p. 23), it stems from "the well-known difference in naming time for pictures and written words. Words can be named (…) faster than the pictures of the same items."

concept. This observation was also confirmed by the task performed by the second group as they assigned words to their categories with a comparable speed for both words and pictures (Potter et al. 1984: 27-31).

Novice learners of French, with English as L1, participated in the second experiment. The stimuli were similar to those used in experiment one. This time, the stimuli were presented in English or French. Participants performed the naming and translation tasks. In the next step, some of them were asked to recall the words used in the experiment. The results revealed a significant latency between naming pictures in L2 and translating into L2. It appeared that the translation task took longer. According to the authors of the study, the results constitute proof confirming the word association hypothesis (Potter et al. 1984: 31-34).

Potter et al. (1984, p. 24) developed two models based on this experiment. The first is the word association model, according to which "access to and from the second-language word (L2) is exclusively via the first language word (L1)." In the concept mediation model, "the bilingual second language is associated directly with the relevant underlying concept and thus is only indirectly associated with the first-language word."

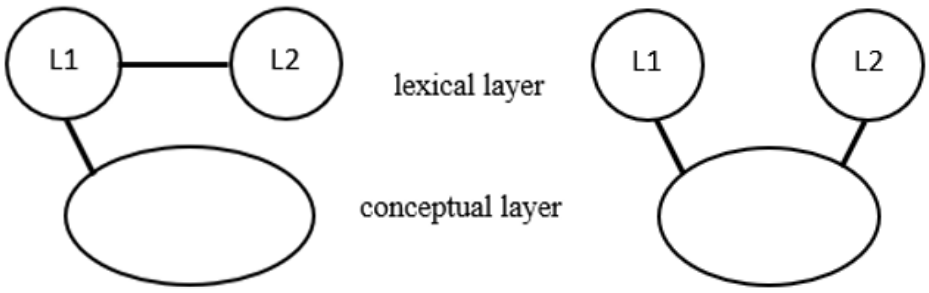


Figure 1. Word association model (on the left) and Concept mediation model (on the right) adapted from de Groot (2002, p. 37).

1.4.2 Development of the Revised Hierarchical Model

Models developed by Potter et al. (1984) seem to be two distinct approaches to the access of meaning in L2. However, the researchers also pointed out the third option, which is "an intermediate model in which second language learners start only with lexical associations, but gradually develop direct links between the second language lexicon and concepts, as in the concept mediation model" (Potter et al. 1984, p. 24). Although Potter et al. (1984) did not further develop this idea, it was thoroughly studied by Kroll and Steward ten years later. At that time, there started to appear ideas according to which word association and concept mediation models do not have to be mutually exclusive but

can just refer to various levels of L2 proficiency (Kroll & Steward 1994: 149-151). Kroll and Steward (1994) were interested in how the two languages are interconnected in bilinguals' minds. They conducted three experiments in which participants, as in the previous experiments, were supposed to translate words and name pictures in L1 and L2. The first two experiments were related to the effect of category interference and focused only on L1. However, the last experiment included both L1 and L2 and focused on bilinguals (who were neither translators nor translation trainees) and their translation abilities.

As pointed out by Kroll and Steward (1994, p. 152), "the goal of the first experiment was to see whether the category interference effect in picture naming observed by Kroll and Curley (1988) under between-subject condition could be replicated within-subjects when subjects use only their first language to respond." The task of the students with English as their L1 was to name pictures and words appearing on the screen in English. The stimuli were divided into two groups. Words and pictures were either semantically related and belonged to twelve different categories, or they were randomly mixed. Next, students were asked to recall as many words and names of pictures as possible. The results of this experiment reveal that, in line with previous studies, participants named words faster than the pictures presenting the same items. Moreover, the two analysed groups reached a statistically significant difference in the picture naming task. Pictures presented in random order appeared to be named faster than those from categorised lists. Such a relationship did not occur in the case of word naming. Like the previous studies, the recall task results show that pictures are recalled better than words. A statistically significant interference among semantically related categories and picture naming was observed. It led to the belief that concept mediation plays a crucial role in picture naming tasks - the phenomenon known as category interference. Thus, the results of an earlier study by Kroll and Curley (1988) were confirmed.

The second experiment aimed to check whether "the source of semantic interference in picture naming is in the mapping between semantic representations and lexical entries" (Kroll & Steward 1994, p. 155). Students with English as their L1 participating in the experiment were supposed to name words and pictures appearing on the screen. The stimuli and the tasks were the same as in the first experiment. This time, trials were not limited to either pictures or words but contained both words and pictures. A recall task was excluded. The results resembled the ones from the first experiment. Participants named words faster than pictures. Nevertheless, there was no category

interference as they had to name both words and pictures within one trial. It means that trials in which items belong to the same category were not named faster than trials with randomised lists of items. According to the researchers, this may suggest that the lack of category interference in picture naming during the second experiment may be the result of a "continuous access to related concepts [that] produces increased activation at the conceptual level, which makes it more difficult to select than the single lexical entry that best names the pictures" (Kroll & Steward 1994, pp. 156-157).

Finally, the third experiment, which led to the development of the Revised Hierarchical Model, was conducted on a group of bilinguals whose L1 was Dutch and L2 was English. Participants were supposed to translate or name words in their L1 and L2. As in previous experiments, words were assigned to categorised and random lists. The experiment ended with a recall task. The objective of the third experiment was to analyse whether category interference is also present among bilinguals during the translation task. Moreover, Kroll and Steward observed in their earlier studies that when translating in both directions, L1-L2 translation is always slower than L2-L1 – the phenomenon known as translation asymmetry. In line with previous studies, the results of the third experiment showed that participants name words faster than they translate them. Moreover, it took them longer to name words in their L2 than in their L1. In the translation task participants needed more time to work into their L2 than into their L1. Category interference occurred only when bilingual participants translated into their L2. It suggests that L2 words may be accessed through concept mediation in this translation direction. Since these observations confirm the predictions of translation asymmetry, Kroll and Steward believed that the assumptions of the word association and concept mediation model needed to be adjusted to the new findings. According to the Revised Hierarchical Model:

> L1 is represented as larger than L2 because, for most bilinguals … more words are known in the native than in the second language. Lexical associations from L2 and L1 are assumed to be stronger than those from L1 to L2 because L2 to L1 is the direction in which second language learners first acquire the translations of new L2 words. The links between words and concepts, however, are assumed to be stronger for L1 than for L2. According to this asymmetric strength model, when a person acquires a second language beyond the stage of very early childhood, there is already a very strong link between the first language lexicon and conceptual memory. During the early stages of second language learning, second language words are attached to this system by lexical links with the first language. As an individual becomes more proficient in the second language, direct conceptual links are established.
>
> (Kroll and Steward 1994, pp. 157-158)

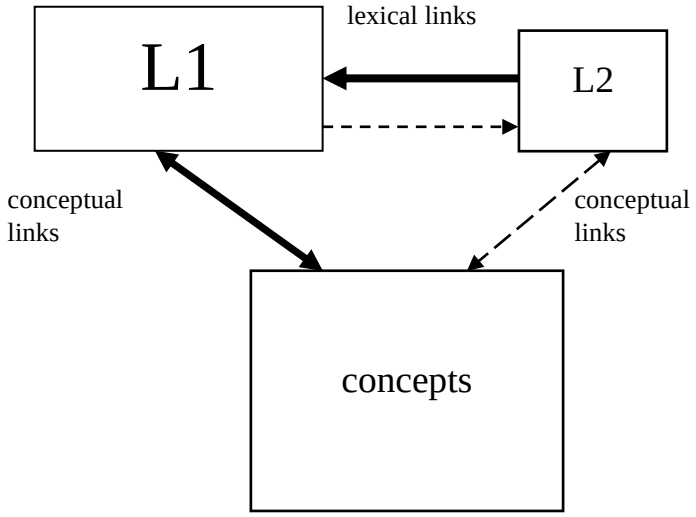


Figure 2. Revised Hierarchical Model adapted from Kroll and Steward (1994, p. 158).

As observed by some researchers (e.g., de Groot 2002; Kroll et al. 2010; Whyatt 2012), the Revised Hierarchical Model is, in fact, composed of the two previous models: word association and concept mediation. It is worth mentioning that in the revised version of the model, its authors pointed out the changes in the strength of links between the components as bilinguals become more fluent in their L2 (Kroll & Steward 1994: 158). Thus, translation asymmetry visible at the early stages of bilingualism diminishes or disappears in fluent bilinguals (de Groot 2002; Kroll et al. 2002; Chmiel 2016; Whyatt 2018; Ferreira et al. 2021). According to Brysbaert and Duyck (2010, pp. 360-361), the Revised Hierarchical Model has four main contributions to studies focusing on bilingualism. These are:

> *the separation of lexical and conceptual representations* … it was the first to clearly outline the implications of the hierarchical model for the issue of shared vs. separate representations in bilingual language processing. *Separate lexicons and selective access* … it made a distinction between the L1 and L2 lexicons … *Asymmetries between L1 and L2 processing* … *the developmental aspect of bilingualism*. Finally, the RHM [Revised Hierarchical Model] appealed to many researchers in language acquisition because it included an interesting view of how the organisation of bilingual memory changed as a function of increasing proficiency.

The assumptions of the Revised Hierarchical Model were verified by several other studies. For example, the study by Garcia et al. (2014) was divided into two experiments. The participants of the first one were bilinguals with English as L1 and Spanish as L2. They were divided into two groups: having either low or high expertise in translation. However, none of the participants received formal translation training. Their task was to read words in L1 and L2 and to perform L2-L1 and L1-L2 translation of words. According to the results, the difference in reading times for L1 and L2 words occurred only in the

group with low expertise in translation. The fact that they read words faster in their L1 may prove that at the early stage of bilingualism, links between L1 and concepts are stronger than the ones between L2 and concepts. In line with the Revised Hierarchical Model, translation asymmetry disappeared among highly proficient bilinguals. Not only latencies in reading times but also directionality effect during the translation task were not found among the group with high expertise in translation.

The same study design was replicated in the second experiment by Garcia et al. (2014). This time, the participants were bilinguals with formal translation training, with Spanish as L1 and English as L2. They were divided into three groups: beginning students of the translation programme, advanced students of the translation programme, and professional translators. Most of the results followed the findings from the first experiment. It appeared that the higher the level of translation expertise the participants had, the faster they performed the tasks. Moreover, participants usually translated concrete words and cognates more quickly. This effect was not observed in the reading tasks. However, no statistically significant difference was found between advanced students and professional translators. Since the results from both experiments seem similar, researchers summed up that formal translation training may not influence the bilingual processing of words but rather impact other competencies possessed by translators.

The work by Ferreira and Schwieter (2014) is another example of a study based on the Revised Hierarchical Model. Researchers use the Stroop word-translation task to "investigate the asymmetrical nature of L1 and L3 mappings of words to concepts. Furthermore … [they] explore whether or not the semantic relatedness effect is modulated by semantic restrictedness of the words-to-be-translated" (Ferreira & Schwieter 2014, p. 91). According to the researchers, the main idea of the Stroop word-translation task is that besides the words that are supposed to be translated, participants also see some distracting words or pictures. The distracting items either are or are not related to the word that is going to be translated. The participants' L1 was English, and L3 was Spanish. However, they were not fluent bilinguals. Participants translated in the L1-L3 an L3-L1 directions. The results confirmed the assumptions of the Revised Hierarchical Model. The fact that the semantic relatedness effect was relatively low may suggest that beginning L3 speakers rather use lexical links while accessing L3 meaning.

The study by Chmiel (2016) exemplifies the application of the Revised Hierarchical Model to interpreting studies. The researcher analysed translation

asymmetry and the influence of the context on interpreting behaviour among professional unidirectional and bidirectional interpreters. Although both groups consisted of highly fluent and experienced bilinguals, they may experience translation asymmetry due to the specific characteristics of their translation practice. Unlike in the previously discussed studies, participants of the study by Chmiel translated words that appear either in full sentences or on their own. Additionally, the context of sentences was manipulated. Words appeared in high context, low context, or without context. It is believed that semantically constrained context facilitates word recognition. Following this assumption, all participants interpreted words in high context faster. However, contrary to the adopted hypotheses, translation asymmetry was revealed only among bidirectional interpreters. According to Chmiel (2016), translation asymmetry may not reappear due to the characteristics of translation practice. As the study shows, the assumptions of the Revised Hierarchical Model are not always confirmed in other studies. Thus, the model faces some criticism as well.

1.4.3 Critique of the Revised Hierarchical Model

Although the Revised Hierarchical Model had a significant impact on the fields of bilingualism and directionality, it did not avoid criticism. The main criticism of the model was expressed in an article titled "Is it the time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?" by Brysbaert and Duyck (2010). Although the researchers begin their article with the enumeration of contributions the model has brought to the discipline, they question some of its assumptions in the next paragraphs. Brysbaert and Duyck (2010) present five main objections to the Revised Hierarchical Model in their study.

The first one is related to the assumption that L1 and L2 words are stored in separate lexicons and thus differently accessed in the bilingual mind. Since there is evidence that L2 words significantly impact L1 words, they should not be stored separately. In a number of studies described by Brysbaert and Duyck (2010: 363-364), L2 words influenced the processing of L1 words. For example, participants were distracted by L2 words, so it took them longer to choose the correct L1 word. The second objection refers to lexical connections between words in both languages. Brysbaert and Duyck (2010) object to translation priming while translating into L1. They claim that studies that revealed such an effect might have an incorrect research design. Moreover, they enumerate some studies where priming either did not occur or was not significant.

The third objection is related to the simplicity of the Revised Hierarchical Model, which is not reflected in real language processing. According to Brysbaert and Duyck (2010: 365-366), a significant problem lies in the fact that words do not have ideal equivalents, and there may be many translation options. This problem is also noticed by other researchers (e.g., de Groot 2002; Whyatt 2012). Moreover, Whyatt (2012: 89) claims that this issue is not restricted solely to the Revised Hierarchical Model but appeared as well in the two previous versions, namely in word association and concept mediation models. De Groot (2002: 45) suggests that the model in its current form is not only too simple but may also be incomplete. She specifically points out various forms of words, like written and phonological, that are not included in the Revised Hierarchical Model. The next objection raised by Brysbaert and Duyck (2010: 366) refers to the type of links between L2 words and concepts and during the L1-L2 translation process. They suggest that concept mediation plays an important role as well. This argument is also raised by de Groot (2002: 41). And finally, Brysbaert and Duyck (2010, p. 367) question "to what extent all semantic information is language-independent." They claim that some kind of semantic information may be restricted to a particular language. Based on this analysis, Brysbaert and Duyck (2010) believe that some Revised Hierarchical Model assumptions should be modified and adjusted to the new findings from recent studies.

As indicated, Brysbaert and Duyck (2010) are not the only researchers questioning the assumptions of the Revised Hierarchical Model. Some doubts can also be found in the studies by de Groot (2002) and Whyatt (2012). Besides the aspects that overlap with the findings by Brysbaert and Duyck (2010), de Groot (2002: 41) observes that sometimes translation asymmetry predicted in the Revised Hierarchical Model does not appear. Frequently, it is translation into L2, which is faster. Similar results can also be found among studies analysing the issue of directionality based on the translation of longer passages of texts. For example, the study by Whyatt (2019) shows that the influence of the direction of translation on time to translate the texts did not reach statistical significance. Moreover, de Groot (2002) and Whyatt (2012) find that the organisation of bilingual memory may differ among bilinguals and has a more complex structure. De Groot (2002: 45) observes that people do not acquire all words in the same manner. Usually, it is easier and faster to learn specific words. Thus, they may be directly linked to a corresponding meaning. The researcher also finds out that not all L2 words are stored and accessed in the same way. For example, "relatively many high-frequency words are stored in concept-mediation representations, whereas relatively many low-frequency

words are stored in word-association representations" (de Groot 2002, p. 43). The L2 learning process also has a significant impact on the organisation of L2 words in memory.

Kroll et al. (2010) responded to the objections in their article "The Revised Hierarchical Model: A critical review and assessment." It appears that the authors of the Revised Hierarchical Model are aware of some of its misconceptions. They agree that the model they presented lacks non-selectivity while activating the languages. Nevertheless, they strongly disagree with that words from both languages may be stored in one common lexicon. Secondly, Kroll et al. (2010) agree that L2 words may be accessed directly. Thus, L1 mediation is not an obligatory path in this case. They decided that some changes in the scheme of the Revised Hierarchical Model, presented in section 1.4.2, may also be needed. Kroll et al. (2010) opt for changing the weak bidirectional link between L2 words and concepts into a unidirectional line. Finally, they agree with the objections related to the preciseness of the model, especially referring to the semantic representations. However, Kroll et al. (2010) also criticise the study by Brysbaert and Duyck (2010). Their main objection is related to distinguishing between word production and word recognition, as the Revised Hierarchical Model was supposed to refer only to the first issue. They do not support the claim that L2 is always conceptually mediated like L1. Thus, the lack of translation asymmetry in many studies is explained by differences in study designs, like the level of word frequency. To sum up, they emphasise that the need to implement some model adjustments results from field development. However, they stand up for and support the importance of their model and are against rejecting its assumptions.

Chapter 2: Translation process research

This chapter discusses the most popular methods used in the translation process research. It begins with the basic categorisation of process methods and a discussion of triangulation, an approach prevailing in translation process research. The discussion of the process methods starts with verbal reports, including retrospective reports and think-aloud protocols. Then, I move on to keylogging. I devote particular attention to the stages of the translation process and Translog, one of the most popular software for keylogging. In the next section I focus on eye-tracking. I discuss the advantages and drawbacks of this method as well as basic eye-tracking variables. Next, I move on to neuroimaging techniques, including EEG, PET and fMRI. The last section is devoted to questionnaires in translation process research.

2.1 Methods of translation process research: preliminary observations

Interest in the translation process appeared relatively recently, in the second half of the 20th century, and has subsequently increased in popularity into the 21st century (Göpferlich & Jääskeläinen 2009; Płużyczka 2011; Gumul 2019b; Apfelthaler 2019). According to researchers (e.g., Kussmaul & Trikkonen-Condit 1995; Jääskeläinen 2010; Płużyczka 2011, 2013; Sun 2011; Gumul 2019b), the first studies in the translation process, conducted in the 80s and 90s utilised various forms of introspection. However, the new century abounds in new methods using modern computer technologies like keylogging, eye-tracking, and scanning the human brain. As a result, introspection has become less common (Sun 2011; Płużyczka 2011, 2013; Gumul 2019b).

Methods of translation process research can be classified in many ways. Alves (2015) assigns the enumerated methods to the generations of studies. The researcher observes that the oldest method – think-aloud protocols – was applied frequently in the first generation when scholars "aimed at identifying what happens in the translation process" (Alves 2015: 21). Keylogging, classified as the second generation, appeared in the following years and has been frequently triangulated with other measures. The eye-tracking method is assigned to the third generation (Alves 2015: 22-23). Mellinger and Hanson (2020: 173) also add a fourth generation, referred to as the phase of translation studies in their work. They describe it as "ongoing methodological innovation and triangulation." They advocate implementing methods strongly associated with medicine, like fMRI, measuring blood pressure or stress hormones. It is worth noting that the

methodological development strongly influenced the main research interests, shifting them to psycholinguistics (Alves 2015: 21-23).

Frequently, translation process research methods are divided into two categories: subjective and objective methods (e.g., in Gumul 2019). This classification refers to the forms of data elicitation. Subjective methods embrace various types of self-reports, like think-aloud protocols or retrospective verbal reports. In this case, only participants themselves can verbalise the information about the translation process and decision-making. Unfortunately, this kind of data is prone to the participants' interference. For example, they may intentionally hide information. The objective category includes methods like keylogging, eye-tracking, and neuroimaging techniques. In this case, the data are elicited more objectively, utilising special equipment, for example, keyloggers or eye-trackers. As a result, participants are not able to influence the data (Gumul 2019: 172).

Krings (2005), cited in Płużyczka (2011), proposes quite a different type of organising these methods. The researcher draws attention to the exact time of their application. He discusses two main groups, real-time and post-hoc measures, and then two subgroups (Płużyczka 2011: 182). The category of real-time measures embraces methods using electronic equipment either in the form of various computer programmes, like keylogging or technologies known from the field of medicine, like EEG (Krings 2005, in Płużyczka 2011: 182). They do not interfere with the translation process and do not involve a translator's active response. Thus, it can be stated that they gather data in the background of the translation process (Whyatt 2012; Seeber 2013; Alves 2015; Kajzer-Wietrzny et al. 2016). Płużyczka (2011: 182), referring to Krings (2005), also assigns think-aloud protocols to real-time measures since they have to be conducted simultaneously with the translation task. However, in this case, the level of the translator's active involvement is much higher. The post-hoc measures group includes various types of target text analyses, retrospective reports, and questionnaires. To implement such methods, the translation task must first be finished (Krings 2005, in Płużyczka 2011: 182).

2.2 Triangulation

Although scholars discussing the topic of the translation process organise the methods in a variety of ways, they are consistent in that none of the presented measures could precisely reflect mental processes. Each of the methods has some drawbacks and limitations (Göpferlich & Jääskeläinen 2009; Płużyczka 2011; da Silva 2015;

Kumpulainen 2015; Gumul 2019a; Hunziker Heeb 2020). Therefore, triangulation, defined by Sun (2011, p. 935) as "the use of multiple methods to examine a research problem so that biases can be eliminated," seems to be a response to these problems (Gumul 2019b,2020; Alves 2015). According to researchers, triangulation was first developed by Jakobsen (1999) (Sun 2011; Alves 2015).

As has been observed, the multiplication of methods could positively impact the study design. According to Jakobsen (2003: 70), the greatest advantage of triangulation lies in the fact that the limitations of one method, which disenable obtaining some kind of data, could be compensated for by the possibilities offered by other measures. Sun (2011: 936) emphasises that "[t]he more research methods one adopts in one's research, the more complex research questions one might be able to answer." Triangulation is also the most appropriate approach to translation process research for da Silva (2015: 176-177), who favours using more than one method.

Despite the researchers' optimism, triangulation has not escaped some criticism. For example, Sun (2011: 936) recalls the work by Fielding and Fielding (1986), who claimed that "using multiple methods or data sources does not necessarily increase validity, reduce bias or bring objectivity to research, as different methods often measure different aspects of a phenomenon." In fact, the problem of inadequate methods choice cannot be ignored (Sun 2011: 936-937). Such methods may exclude each other and preclude data gathering and analysis. Therefore, the researcher emphasises the importance of an accurate method choice. Sun (2011: 936-937) also draws attention to the fact that the number of methods may impact the length of the research process.

The popularity of triangulation has increased in recent years. For example, five methods were utilised in the Transcomp project, the objective of which was to study translation competence (Göpferlich 2009). These were keylogging, screen recording, think-aloud protocols, retrospective reports and a questionnaire. Five methods were also applied to the study by da Silva (2015), who enumerates the following methods: "[d]ata elicitation techniques were questionnaires, direct observation, keylogging, screen recording, and free retrospective protocols" (da Silva 2015, p. 186). Researchers frequently decide to use two or three methods. For example, Gumul and Pietryga (manuscript under preparation) decided to use eye-tracking, two types of retrospective verbal reports and product analysis to study the explicitating behaviour of interpreting trainees during a sight translation task. Gumul (2017b), in her research discussing the issue of explicitation in simultaneous interpreting, decided to use retrospective verbal

reports, product analysis, and a questionnaire. Quite an uncommon connection of methods was chosen by Korpal (2016a), who not only applied semi-structured interviews but also measured heart rate and blood pressure in his study on stress during interpreting. The research by da Silva et al. (2017) is an example of a study implementing two methods. The researchers utilised the Translog programme, representing the method of keylogging and eye-tracking, to study the effort in translation and post-editing in reference to directionality. These examples indicate the impressive research possibilities that triangulation could offer the researcher. It can be applied to study various topics within the field of translation.

2.3 Verbal reports

As already discussed, verbal reports were classified within the first generation of translation process research (Alves 2015: 21). In fact, applying various forms of verbalisations was the first attempt to analyse the translation process (Jääskeläinen 2001, Göpferlich & Jääskeläinen 2009). Works by Lörscher (1992) and Jääskeläinen (1993) are examples of implementing verbal reports in translation and interpreting studies. However, the method itself did not originate in the field of translation but has been adopted from cognitive psychology (Ericsson & Simon 1984). It was observed that psychologists frequently and successfully ask their participants to verbalise their mental processes. Ericsson and Simon emphasise that:

> [t]he thought process can thus be described as a sequence of states, each state containing the end products of cognitive processes, such as information retrieved from long-term memory, information perceived and recognised, and information generated by interference. The information in a state is relatively stable and can thus be input to a verbalisation process and reported orally.
> (Ericsson and Simon 1993, p. 13)

Therefore, the researchers assumed that verbal reports could also be a source of valuable information about the cognitive processes in the case of the translation process. However, as pointed out by Jääskeläinen (2001, p. 266), "[v]erbal reports have had a highly controversial history in psychology, ranging from unconditional acceptance by structuralists to total rejection by behaviourists. Contemporary assessments tend to be less extreme."

Verbal reports can be divided into two broad categories based on the time when the participants are supposed to verbalise their thoughts (Saldanha & O'Brien 2013; Russell & Winston 2014). The method known as think-aloud protocols (TAPs), which is

a concurrent verbalisation, takes place while performing a translation task (Jääskeläinen 2010; Sun 2011; Russell & Winston 2014). When the participants are asked to verbalise their thoughts and conscious decisions after the task has already been finished, we talk about retrospective reports (Ericsson & Simon 1993; Englund Dimitrova & Tiselius 2014; Gumul 2019b, 2020b). Temporal constraints of the types of verbal reports significantly influence the contexts in which each method can be applied. Simultaneous production of the TT and verbalisation of thoughts is possible in the case of translation. Therefore, TAPs are frequently applied to study the translation process (Pavlovic 2007, 2010). It can be observed that the very first studies applying verbal reports used TAPs rather than retrospective reports, which appeared later (Piotrowska 2000; Hansen 2005; Alves et al. 2009). The nature of interpreting disenables verbalisation during the TT production. That is why retrospective reports, performed after the task has been finished, are frequently chosen to study the process of interpreting (Englund Dimitrova & Tiselius 2009; 2014; Gumul 2017b, 2021a,b). Retrospective reports are also implemented in TPR (translation process research). For example, Ferreira et al. (2018) used retrospective reports to analyse directionality and the decision-making process.

2.3.1 Retrospective reports

Retrospection is described by Englund Dimitrova and Tiselius (2009, p. 101) as "an introspective method, tapping subjects' cognitive processes via their own reports." Retrospective reports enable getting at least some insight into the process data which cannot be captured using any other quantitative methods (Herring & Tiselius 2020: 58). As already mentioned, the task of participants is to verbalise all the conscious processes occurring in their minds during the already finished process of translation or interpreting (Gumul 2019b, 2020a, 2020b; Englund Dimitrova & Tiselius 2014). As a result, the researcher obtains a set of comments that should refer to the translation or interpreting product. According to Saldanha and O'Brien (2013: 122-123), reports may also have a written form. They enumerate various options like translation diaries or integrated problem and decision report. However, Herring and Tiselius (2020: 56) found some inconsistencies in the nomenclature referring to retrospection. They emphasise that there is a wide range of different names used in various studies that also denote different types of retrospection, beginning from retrospective interviews used by Mead (2002) to stimulated recall, which could be found in the study by Russell and Winston (2014), and retrospection' and retrospective protocols/reports (Englund Dimitrova & Tiselius 2009;

Gumul 2020b), as well as retrospective process tracing (Herring 2018, Gumul and Herring 2022). It should also be noted that usually, researchers deliberately use various names to distinguish between many types of retrospection.

Ericsson and Simon (1993, p. 16) characterise retrospection in the following way:

> a subset of the sequence of thoughts occurring during performance of a task is stored in a long-term memory. Immediately after the task is completed, there remain retrieval cues in short-term memory that allow effective retrieval of the sequence of thoughts.

Performed after the task has already been finished, retrospective session does not require attention divisibility and does not hinder or disrupt the task (Ferreira 2014; Gumul 2019b, 2020b). It also does not impose any additional cognitive load (Jakobsen 2011; Gumul 2019b).

All types of verbal reports enable verbalisations of only conscious processes. Therefore, any automatised processes would not be included in the reports, which brings the issues of their validity and completeness (Ivanova 2000; Hansen 2005; Bartłomiejczyk 2007; Saldanha & O'Brien 2013). Since retrospection is a purely subjective method, there is also a risk that participants will consciously avoid verbalising some of their thoughts (Vik-Touvinen 2002; Gumul 2019b, 2020a).

To somehow solve these problems, scholars recommend triangulating retrospective reports with objective quantitative methods, like keylogging or eye-tracking (Vik-Tuovinen 2002; Jääskeläinen 2017, Shamy & de Pedro Ricoy 2017; Gumul 2020a). Retrospective reports may be triangulated with several other methods to gather data about the translation process. Probably one of the most frequent combinations consists of retrospective reports and keylogging. For example, Ferreira (2014) applied retrospective reports and keylogging using the Translog programme to analyse recursive movements and translation problems in the context of directionality. In the case of simultaneous interpreting process research, retrospective reports are frequently combined with product analysis (Gumul 2017b). For example, Gumul (2021) used retrospective reports and the analysis of the interpreting product to study indicators of stress as well as strategies implemented to overcome it. In an earlier study by the same researcher, this combination of methods served to analyse explicitation in simultaneous interpreting (Gumul 2017b). The characteristics of sight translation give the possibility to triangulate retrospective verbal reports and eye-tracking. Such a combination of methods was used, for example, in the study by Gumul and Pietryga (manuscript under preparation). In this case, the

researchers were able to compare verbalisation related to explicitation with gazing patterns identified while performing explicitation during the sight translation task.

2.3.1.1 Immediacy condition in retrospection

Since verbalisation takes place after the translation task is finished, there is a considerable risk that such information will be naturally forgotten (Hansen 2005; Jääskeläinen 2017). This phenomenon is known as a 'recency effect' or 'immediacy condition' (Ericsson & Simon 1984; Ivanova 2002; Gumul 2017b). It includes two aspects that may influence the verbalisations: the length of the task after which the retrospective session takes place and the time that passes from the end of the task to the beginning of the retrospective session (Ericsson & Simon 1984; Gumul 2019b). Retrospection can be divided into two types. Immediate retrospection is conducted right after a very short task. Ericsson and Simon (1993, p. 16), when describing retrospection conducted as a part of psychological research, point out that "for tasks that can be completed in 0.5 – 10 seconds, we would expect the subject to be able to recall the actual sequence of their thoughts with high accuracy and completeness. With longer durations, recall will be increasingly difficult and incomplete." Such a time constraint is also related to the capacity of short-term memory that is used by participants of psychological studies when they verbalise conscious processes (Gumul 2021c: 80-81). Delayed retrospection takes place when the pause between the task and verbalisation is extended, reaching even a few days, or when the task preceding retrospection is longer than 15 seconds (Cohen & Honsenfield 1981; Hansen 2005; Gumul 2019b). Translation scholars recommend immediate retrospection to get as much data as possible (Hansen 2005; Jääskeläinen 2017; Gumul 2019b). However, the characteristics of the translation and interpreting processes exclude the possibility of conducting a proper immediate retrospection in the case of task length and the pause preceding the retrospective session (Gumul 2019b, 2020b).

First of all, translation process is always much longer than the recommended few seconds (Gumul 2019b, 2020a). Translators work at their own pace and manner of translating, e.g., frequently moving back and forth within the text (Englund Dimitrova & Tiselius 2009: 111). A proper immediate retrospection is also not possible in the case of interpreting. Herring and Tiselius (2020, p. 60) emphasise that "the interpreting task preceding a retrospective process tracing must be long enough to accurately simulate a real-life interpreting task." There were attempts to divide the interpreting process into

smaller parts to meet the needs of the fMRI study, as described in the study by Hervais-Adelman and Babcock (2019: 742). However, scholars unequivocally emphasise that it does not reflect the actual process of simultaneous interpreting.

Secondly, Gumul (2019: 177) points out that in the interpreting process research retrospective session is conducted as soon as possible, after the interpreting task is finished. In fact, immediate retrospection can occur only in artificial laboratory conditions, for example, during experimental research. As the researcher continues, for interpreters in their genuine work environment, retrospection is frequently possible only after long hours or even the next day. For example, in the study by Vik-Touvinen (2002: 64), participants had the possibility to provide their verbal reports only later the same day. Moreover, the researcher states that for many participants, the retrospective session took place a few days after the interpreting task.

The enumerated examples show that the characteristics of translation and interpreting tasks forced researchers to redefine immediate retrospection. Both the analysed task and the pause preceding retrospection are significantly longer than in the recommendations made by Ericsson and Simon (1993). Thus, it can be concluded that the method of retrospection is adopted from psychology rather than borrowed in an unchanged form (Gumul 2019b, 2021; Herring & Tiselius 2020).

2.3.1.2 Cueing

The aim of cueing is to trigger participants' memory and to minimise the consequences of the recency effect (Saldanha & O'Brien 2013; Gumul 2020b, 2021c). Therefore, researchers frequently emphasise the importance of an adequate cueing choice (Englund Dimitrova & Tiselius 2009; Shamy & Pedro Ricoy 2017; Gumul 2020b). Herring and Tiselius (2020, p. 61) claim that "[t]he ideal cue thus triggers the participant's memory without skewing memory by adding information." In the case of triangulation, possible cueing depends on the task type and methods used. All the types of cueing are usually limited to various forms of the source text (ST) or the target text (Shamy & Pedro Ricoy 2017; Herring 2018; Gumul 2019b, 2020a, 2020b; Herring & Tiselius 2020). There is also a possibility to trigger memory by asking questions, the so-called verbal probes (Hering 2018; Gumul & Herring 2022).

Cueing in the form of the ST is probably the most frequently chosen option (Herring 2018; Gumul 2020b; Herring & Tiselius 2020). Researchers usually emphasise its ecological validity. It is believed that reading a ST does not trigger new cognitive

processes, which may happen in the case of TT reading (Gumul 2019b, 2020b). Therefore, ST cueing should facilitate verbalisations related solely to cognitive processes rather than trigger performance assessment (Jääskeläinen 2017: 219). In her study, Gumul (2020b: 160) finds that in the case of ST cueing, some participants remembered parts of the text that were especially difficult. They provided verbalisation after finding them in the transcript. Nevertheless, Gumul (2020b, p. 154) suggests that ST cueing is also not devoid of problems; "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." Moreover, participants will probably remember and verbalise less without seeing their own performance (Gumul 2020b: 154).

Another type of cueing is using a target text. In this case, translators produce verbalisations while seeing their own performance either in a written or a recorded form (Herring 2018). According to Shamy and de Pedro Ricoyy (2017: 53), this type of cueing positively influences the number of verbalisations. Nevertheless, there is one important drawback of TT cueing emphasised by many researchers. While seeing their own performance after it has already been finished, participants may find their mistakes and try to justify their decisions. As a result, they may try to assess their TT, which can lead to the occurrence of new cognitive processes. Gumul (2020b, p. 154) also mentions that "it is not always feasible to separate the verbalisations that probe the original processes from those made *post factum*." Verbalisations in which participants assess their TT frequently lead to a number of comments not related to the translation process (Ivanova 2000; Englund Dimitrova & Tiselius 2009; Jääskeläinen 2017; Gumul 2019b, 2020b; Herring & Tiselius 2020). It is also worth noting that seeing their own mistakes may be uncomfortable for both experienced translators and novices.

The effect of both types of cueing on the informativeness, accuracy and verbosity of verbalisations was analysed by Gumul (2020b). The participants were students who, after performing L2-L1 interpreting, were asked to perform retrospection cued either by the ST transcript or their performance in the form of a TT recording. Surprisingly, no statistically significant difference was found between ST and TT cueing and their influence on the three aspects enumerated above. Students were also asked to fill in a questionnaire in which they could express their opinions about both types of cueing. ST cueing was often chosen as the kind of memory trigger enabling verbalisation. However, there were suggestions in both groups that having the other type of cueing may appear

helpful. To sum up, probably each type of cueing, especially if inadequately chosen, may distort the verbalisation process (Saldanha & O'Brien 2013; Englund Dimitrova & Tiselius 2014; Gumul 2019b, 2020b).

In the third type of cueing, a retrospective interview, participants are asked direct questions (Gumul 2020). The researcher asks questions intending to help subjects to verbalise their thoughts or to gain information about a specific behaviour (Gumul 2020a: 44). This type of cueing also functions under the names verbal probes (Herring 2018) and a question-probed retrospection (Herring & Tiselius 2020). Researchers emphasise that in order to use verbal probes properly, the experimental design should be carefully planned and identical for all the participants. Each participant should be asked an identical set of questions related solely to their thought processes. The questions should not imply any answers. Therefore, lack of validity is enumerated as the main drawback of such cueing (Gumul 2020a: 44). Participants should not have the impression that the researcher suggested a correct answer. They also should not artificially create a comment when they would not naturally elaborate on such aspects (Vik-Touvinen 2020; Gumul 2020a). Question-probed retrospection was used, for example, in the studies by Herring (2018), Gumul and Herring (2022) and Gumul and Pietryga (manuscript under preparation).

It has also been observed that it is not necessary to provide any type of cueing (Herring & Tiselius 2020: 61). As Vottonen and Kujamäki (2021: 6) point out, cueing always makes the process of retrospection and the whole experiment longer. Moreover, Herring (2018, p. 129) observes that "[u]ncued retrospection was thus identified as the best option for gaining more direct, unfiltered insight into the interpreter's thought process." However, in this case, the probability of producing sufficient and valuable comments decreases drastically (Herring & Tiselius 2020: 61).

Source and target text cueing may be implemented in a variety of ways to trigger memory. Transcripts of the interpreted speech were used in the studies by Tiselius and Jenset (2011) and by Englund Dimitrova and Tiselius (2009, 2014). The researchers state that "[t]he cue was a transcript of the ST, and the subjects were asked to go through it, sentence by sentence, trying to recall their thoughts and actions while they interpreted and translated it and described it in their own words" (Englund Dimitrova & Tiselius 2009, p. 118). Bartłomiejczyk (2007), in her study analysing the influence of directionality on the strategies used during conference interpreting, provided her participants with recordings of both the source speech and their performance. She points out that "[t]hey [participants] were asked to listen to the recording of the source text and

their own interpreting and try to remember what they had thought when interpreting the text" (Bartłomiejczyk, 2007, p. 6).

Technological advancement, especially keylogging and eye-tracking, positively impacts the scope of cueing. Besides their target texts, participants can watch the recording of their translation process, including implementing modifications, as well as the points that their eyes focus on (Englund Dimitrova & Tiselius 2009, 2014; Saldanha & O'Brien 2013). The replay function of the Translog programme was used for cueing, for example, in the studies by Ferreira (2014) and Ferreira et al. (2018). The researcher describes the procedure of her study in the following way: "[t]he translation processes were recorded with Translog. After completing each task, the participants were shown a replay of the Translog recording. They were asked to comment on possible difficulties and solutions to the problems they encountered" (Ferreira 2014, p. 116).

Another option is to use screen recording. This method could be found in the study by Vottonen and Kujamäki (2021). The researchers implemented the Camtasia Studio 8 screen recording programme to trigger verbalisations. Nevertheless, the idea of showing participants detailed and often unconscious information about their translating and interpreting behaviour is subject to some criticism. Saldanha and O'Brien (2013: 141) point out that instead of eliciting verbalisations, it may inhibit them. The risk is especially high for subjects unfamiliar with such programmes and replay functions. The result of focusing on unknown pictures and interesting data on the screen may be a lack of any verbalisation.

Some researchers combine various forms of cueing, implementing a different one in a multi-stage retrospection. Herring's (2018) retrospective process tracing (RTP) is an example of that (also used by Gumul & Herring 2022, 2023). RTP includes three stages of retrospection: "[t]he first stage was uncued retrospection, followed by minimally-cued retrospection, and then verbal probes" (Herring 2018, p. 129). The author of the method points out that the rationale for providing diverse cueing was also the task itself. There is no source text in the case of dialogue interpreting. Herring (2018: 129-130) also suggests that this method allows the researcher to get pure, uninterrupted information first and then move to more specific aspects. In the second stage, the memory was minimally triggered to remind the participants of the general topics of the ST. In the case of the study by Herring (2018), the cueing was presented in the form of a list, including the main parts of the interpreted dialogue. Gumul and Herring (2022, 2023) presented participants with a ST transcript. In the last stage of the RPT, participants were asked precise questions about

their interpreting process. According to Herring (2018: 130), the asked questions allow the participants to focus on those aspects of the interpreting process that are of particular interest to the researcher.

2.3.1.3 The langue of retrospection

There is also the issue of the language in which verbalisations should be made. There is a variety of approaches to this topic. For example, Saldanha and O'Brien (2013: 127) emphasise that the researcher should decide on just one language in which the subjects are expected to verbalise their thoughts. This point of view is backed up by two main arguments:

> [f]irst, the choice of language might impact the ability of the participants to verbalise, especially if they are using a second or third language. Second, in the case of translation, the participant is already engaged in a bilingual processing task, so there may be interference between the source and target languages and the language used for verbalisation - participants are very likely to produce bilingual protocols.
>
> (Saldanha and O'Brien 2013, p. 127)

However, it is also a common practice to allow the participants to use whichever language they wish and mix L1 and L2. This can be found in the studies by Bartłomiejczyk (2007), Gumul (2021a) and Gumul and Pietryga (manuscript under preparation). Bartłomiejczyk (2007: 6-7) points out that such freedom of language choice facilitates verbalisation as participants focus less on the language's grammatical and stylistic correctness.

2.3.1.4 The role of the researcher during retrospection

It is emphasised that researchers' involvement in the process of retrospection should be as minimal as possible unless they are asking questions as a form of a memory trigger (Shamy & de Pedro Ricoy 2017; Gumul 2019b, 2020a; Herring & Tiselius 2020). When interrupting too often, they increase the possibility of obtaining unreliable data. Participants may either want to help the researcher get interesting reports or feel constantly observed and, therefore, hide some information (Hansen 2005; Saldanha & O'Brien 2013). This phenomenon is known as the white-coat effect (Gumul 2021b: 30).

According to Gumul (2019, 2020a), the researcher's active involvement in the retrospective session should end after they have provided understandable instructions. However, there are also opinions that the researcher should be present in the room during the session in case of any technical issues or doubts related to the correctness of the

performance (Hansen 2005; Herring & Tiselius 2020). Herring and Tiselius (2020, p. 63) give precise instructions for the researcher's presence in the room:

> [a]s the participant starts the retrospection, it is advisable that the researcher sit slightly behind the participant so as not to be in their line of sight. Our experience is that it is important for the researcher to be physically present in case of technical difficulties and also to add the human dimension, including encouragement. The participant is likely to ask for confirmation that they are on the right track or have understood the instructions. In such cases, it is important for the researcher to be encouraging but without interfering.

The exact procedures can be found in another study by Tiselius. Writing in 2011, Tiselius and Jenset point out that during their study, the researcher did not leave the room where the retrospective session was conducted in case of any problems. However, they were out of the participant's sight. In the study by Vottonen and Kujamäki (2021), the researcher was present during the session and intervened only when the pauses between the comments were too long.

2.3.1.5 Transcription and coding

After the retrospective reports have been recorded, the subsequent phase is data transcription and coding (Ferreira 2014; Gumul 2020b). Transcribing the recording is often described as the longest part of the study. The recording may also have poor quality or many disfluencies (Saldanha & O'Brien 2013; da Silva 2015). There is also an option to use some programmes facilitating the process or to commit this task to an outside company. Transcription is also considered one of the most important stages of a study that uses any kind of verbal reports (Saldanha & O'Brien 2013; Herring & Tiselius 2018; Gumul 2020a). According to Ferreira et al. (2018, p. 106), it "allow[s] for quantitative and qualitative analysis of the retrospective protocols." Thus, transcription is believed to be the beginning of data analysis and interpretation.

Saldanha and O'Brien (2013) and Gumul (2020a) discuss two main types of transcription: naturalised and denaturalised. In the case of naturalised transcription, the written text is not entirely faithful to what has been recorded. Elements like long pauses, hesitation markers, false starts, repetitions, and other characteristic features of the spoken language are not written down. Nevertheless, Saldanha and O'Brien (2013: 129) warn that it may not give the complete picture of the thoughts and verbalisations. Denaturalised transcription contains all those elements that are omitted in the case of the first option. This time, the text is an exact, complete, and faithful version of the recording (Saldanha

& O'Brien 2013; Gumul 2020a). According to Gumul (2020a: 45), the choice of the type of transcription should be thoroughly thought about and influenced by the type of information required by the research questions.

Interestingly, Shamy and de Pedro Ricoy (2017: 59) suggest that some scholars do not support the practice of transcribing the recordings. The most frequent argument used is the fact that even the most faithful transcription may falsify the data to some extent. The recording always gives a complete picture of the verbalisations. It is especially important for languages used in the oral rather than written form, for example, regional varieties of Arabic.

The next phase is to code the transcribed data (Saldanha & O'Brien 2013; Gumul 2020a; Herring & Tiselius 2020). The codes should correspond to the issues discussed in the study (Gumul 2020a: 47). Saldanha and O'Brien (2013: 130) suggest that the coding system may be developed either beforehand or after the transcription is finished based on the gathered information. Gumul (2020a: 47-48) emphasises that it is of high importance to firstly mark the verbalisations for the actual part of the ST they are referring to. The researcher should also separate the comments useful for their study from those that do not refer to the translation process or are not related to the objective of the study. For example, Gumul (2017) classified the obtained comments into two categories: relevant and irrelevant. Relevant comments were characterised as "all verbalisation complying with the instructions […] reporting on decisions consciously taken during an interpreting task" (Gumul 2017, pp. 147-148). Irrelevant comments included "observations about the quality of the output, observations clearly made *a posteriori*, comments explicitly referring to solutions adopted automatically, and also protocols reporting explicitation without giving any reason" (Gumul 2017b p. 148). Interestingly, around one/fourth of all comments were classified as irrelevant. According to Saldanha and O'Brien (2013, p. 130), "[v]erbal protocols could be coded at macro, micro and both levels." The first type of coding refers to general observations. The second one embraces very specific, precisely localised phenomena. Herring and Tiselius (2020: 65) point out that researchers may create their own system of codes or reuse the ones already implemented either in their own studies or those used by other researchers.

For example, Gumul (2021), in her study discussing stress in interpreting, states that "the transcripts were coded for the three types of disfluencies (…) hesitation markers, false starts and anomalous pauses. In turn, retrospective protocols were coded for reports of stress and/or anxiety" (Gumul 2021b, p. 31). Ferreira et al. (2018: 107), in their study

analysing the decision-making process in the context of directionality, divided verbalisations into two broad categories*:* problem identification and potential solutions, which were subsequently divided into many specific subcategories like postponing the final decision, personal preferences, fluency or consulting external resources.

2.3.1.6 Main drawbacks

Apart from the many advantages of retrospective reports, there are also some drawbacks researchers should be aware of when designing the study. Some of them have already been enumerated in the previous sections. First and foremost, there is a considerable risk of incompleteness (Bartłomiejczyk 2007; Saldanha & O'Brien 2013; Englund Dimitrova & Tiselius 2014; Shamy & de Pedro Ricoy 2017; Ferreira et al. 2018; Gumul 2020a,b, 2021). Some information may be missing as the result either of as a conscious choice by the participants or of memory decay. The problem of automaticity cannot be ignored. Gumul (2019: 182) points out that professional translators and interpreters with long years of experience who have already worked out their own system of translating and interpreting and in the case of whom the automatisation process has occurred may not be able to exactly depict their thought processes. For example, the differences between professionals and students have been discussed by Ivanova (2000). The author observes that "[i]n the course of retrospection, the trainees identified more processing problems compared with the reports of the experts" (Ivanova 2020, p. 42). Experts were focused more on the adequacy of the message.

Another problem is related to personal features (Ivanova 2000; Sun 2011; Shamy & de Pedro Ricoy 2017). Some people naturally verbalise a lot and can easily convey their thoughts. For others, such a situation may be unnatural or uncomfortable. They may also be shy and taciturn (Gumul 2019b: 181-182). Both Jääskeläinen (2017: 22) and Saldanha and O'Brien (2013: 126) emphasise the influence of language and culture on verbalisations. In some cultures, people are used to easily express their feelings and thoughts. In others, they are more reserved.

There is also the issue of the amount and quality of comments the researcher obtains. Both Bartłomiejczyk (2007) and Gumul (2017b) point out that in their studies, a substantial number of comments do not refer to the translation process. Likewise, Englund Dimitrova and Tiselius (2014: 195) find that subjects usually verbalise less than half of the actual problems they encounter while translating or interpreting. Moreover, they observed that participants may describe processes or problems that could not be

found in the translation product. They suggest that sometimes there is only a minor indicator of a given problem in the output; therefore, it is not analysed, though the participant managed to point it out in their comments (Englund Dimitrova & Tiselius 2014: 193). Nevertheless, the number of comments should not be conflated with their accuracy (Englund Dimitrova & Tiselius 2009: 126). Many comments may be created post-factum or do not refer to the translation process. In contrast, the reports, including small amounts of comments, may cover only the most crucial factors.

2.3.2 Think-aloud protocols

Think-aloud protocols, also known as concurrent verbal reports, are another type of verbalisations used frequently in translation process research (Jääskeläinen 2001, 2010; Pavlovič 2007b; Piotrowska 2002, 2007; Russel & Winston 2014). TAPS are defined by Jääskeläinen (2001, p. 266) as follows:

> [it] involves asking a translator to translate a text and, at the same time, to verbalise as much of his or her thoughts as possible (…) think aloud is concurrent (takes place simultaneously with the task performance) and undirected (subjects are not asked to verbalise specific information).

There are many similarities between TAPs and retrospective protocols. For example, the role of the researcher, transcription, and coding usually look the same. As already mentioned, the key difference is the time of verbalisation. Both translation and verbalisation take place at the same time. Therefore, when verbalising their conscious decisions, subjects usually rely on short-term memory (STM) or working memory (WM) (Ericsson & Simon 1993; Englund Dimitrova & Tiselius 2014; Gumul 2019b). There are important advantages to think-aloud protocols. First of all, in contrast to retrospective reports, there is no risk of memory decay. Secondly, when subjects are engaged in two tasks (translating and verbalising), they are more likely to focus solely on the details of the translation process when thinking aloud (Ivanova 2000; Russel & Winston 2014).

However, multitasking is frequently enumerated as the main drawback of TAPs. Many participants are not able to perform two tasks simultaneously (Hansen 2005; Russell & Winston 2014). This results in a lower quality of either the TT or the verbalisations. There is a debate about whether verbalisations performed at the same time as translation impose an additional, exhaustive load on the cognitive system (Hansen 2005; Jääskeläainen 2010; Sun 2011; Gumul 2019b, Sun et al. 2020).

One of the first studies to analyse this problem was conducted by Jakobsen (2003). The researcher triangulated TAPs with keylogging to analyse the influence of verbalisations on translation performance. He focused on three aspects: speed, pauses as the indicators of segments and revisions. Both professional translators and graduating students participated in this study. Their task was to translate in and out of their native language, and the analysed language pair was Danish-English. Jakobsen (2003, p. 79) found that "TA [think-aloud] slowed down target text (TT) production; both L2 to L1 and L1 to L2 translation, whether by semi-professionals or by experts, were slower when done with think-aloud than when done without think-aloud." It also significantly influenced the number and nature of the segments. There were usually more segments that were shorter compared to translating without TAPs. Nevertheless, it was observed that verbalisations co-occurring with translation did not influence the revision phase. Therefore, it is suggested that "think-aloud does not change the course or structure of thought process, except for a slight slowing down of the process" (Jääskeläinen 2010, p. 371). Other scholars also elaborate on this topic. They suggest that delays resulting from the design of TAPs may seriously hinder data gathering in studies focusing on translation speed (Saldanha & O'Brien 2013: 124).

Hansen (2005: 513) draws attention to pauses occurring during concurrent verbalisations. Participants are expected to constantly verbalise their thoughts; however, there are points when they stop verbalising. This may be either a result of the exhaustion of the cognitive system or a sign of a very complex problem-solving process. In the first case, Jääskeläinen (2000, p. 75) emphasises that "a very demanding (…) task tends to use up all the available processing resources and none are left for producing verbalisations." Referring to the second case, Bernardini (2001: 257) points out that the processes occurring in subjects' minds are so intense that they are not able to verbally convey them. Therefore, pauses should never be treated as indicators of gaps in thought processes. There are also suggestions that pauses may result from the inability to constantly comment on the currently performed work. In fact, describing processes undergoing in one's mind is already a problem in the case of retrospective reports. Thus, performing TAPs may be uncomfortable for many people. That is why a short warm-up exercise is recommended to overcome the feeling of unnaturalness (Kussmaul & Trikkonen-Condit 1995; Jääskeläinen 2000; Russel and Winston 2014).

2.3.2.1 Monologue and dialogue think-aloud protocols

Think-aloud protocols are usually applied in two forms: monologues or dialogues (Kussmaul & Trikkonen-Condit 1995; Piotrowska 2000, 2007; Sun 2011; Saldanha & O'Brien 2013). In the case of the first option, verbalisations are performed by just one person, working on their own (Piotrowska 2007: 81). Kussmaul and Trikkonen-Condit (1995: 179-180) observe that researchers usually favour the monologue version. However, in this case, subjects are also most prone to be unable to verbalise.

The second option is known under the names dialogue protocols (Kussmaul & Tirkkonen-Condit 1995; Piotrowska 2007; Saldanha & O'Brien 2013), joint translation (Jääskeläinen 2000; Bartłomiejczyk 2007), and collaborative protocols (Pavlovič 2010). According to Göpferlich and Jääskeläinen (2009: 171), it was introduced in the translation process research after the monologue version had already been used. Bartłomiejczyk (2007, p. 2) explains that it "involves asking two or more subjects to translate a text together and analysing their discussion." Dialogue protocols were used in the study by Pavlovič (2010, p. 66), who describes the method as "verbal reports in which subjects perform translation tasks in groups (..) working on the same source text (ST) and making translation decisions based on consensus." The main aim of dialogue protocols is to induce participants to verbalise naturally without the feeling of artificiality or of constantly being observed. Since participants frequently disagree with the choices offered by their partners, they have to elaborate on their opinions and translation decisions, which positively impacts the number and content of verbalisations (Kussmaul & Trikkonen-Condit, 1995; Jääskeläinen 2000; 2009, Tomaszkiewicz 2013). Jääskeläinen (2017, p. 220) observes that this "open[s] a window to the ways in which meanings and solutions to problems are negotiated in social interaction". To sum up, participants are supposed to discuss and cooperate to achieve their goals (Piotrowska 2007: 81).

However, dialogue protocols did not avoid criticism. The main argument against using this type of TAPs is that instead of analysing the processes occurring in the translators' black box, researchers have to deal with the outcome of processes generated by achieving a compromise between two people (Kussmaul & Trikkonen-Condit 1995: 181). There is also a considerable risk that participants will be unable to work together, for example, due to differences in their personalities. In the most extreme case, one of the participants takes a superordinate role and does not agree to any suggestions (Kussmaul & Trikkonen-Condit 1995: 182). An example of a researcher who does not support dialogue protocols is Bernardini (2001, p. 243). She claims that "conversation involves

reworking thoughts to make them conform to socially established norms, a process which might sensibly alter the information attended to." A similar point is made by Ericsson and Simon (1998: 181). They claim that dialogues may negatively influence verbalisations and recommend using the monologue version.

2.3.2.2 Pedagogical advantages of think-aloud protocols

Researchers frequently emphasise the pedagogical advantages of TAPs (Piotrowska 2000, 2002, Bernardini 2001). Concurrent verbalisations may positively impact students' learning process. They learn how to find and describe problems encountered during the translation process. TAPs also enhance their abilities to apply and explain translation solutions. (Kussmaul & Trikkonen-Condit 1995; Piotrowska 2007; Russell & Winston 2014).

Pavlovič (2010, p. 83) enumerates the general pedagogical functions of TAPs: they may serve "as a diagnostic tool (…), as a monitoring tool (…), to learn about the students' translation styles and thus help them to get to know their strengths and weaknesses." An example of using TAPs for pedagogical aims is discussed in the study by Piotrowska (2002). The researcher investigated the influence of instructions on trainees' awareness of the translation process. Bawej (2015) used written TAPs as a foreign language learning exercise. The participants' task was to translate from German into Polish (L2-L1) and simultaneously write down all their thoughts. As a result, the researcher (teacher) gets feedback related to problems encountered by the students and any gaps in their knowledge. Pavlovič (2010) used collaborative TAPs to study the influence of directionality on the students' decision-making process and description of translation choices while translating a text. This time, the language pair was Croatian-English.

2.4 Keystroke logging

The method of keystroke logging or keylogging was classified as the second generation of translation process research when researchers started applying triangulation on a larger scale (Alves 2015: 22). The method is described by Saldanha and O'Brien (2013, p. 132) as:

> [t]he use of a special software programme that records all keys pressed on the keyboard, as well as mouse movements and pauses between key presses during a text production task (…) Keystroke logging tools produce a log file containing a record of all the keys pressed during a task, such as text production characters (up and down arrows, right and left arrows, etc.), as well as pauses and their duration, and text and revision indicators such as deletions, and cut and paste operations.

As pointed out by Jakobsen (2003, 2011), the originator of the first keylogging programme designed especially to study the translation process, it provides the researcher with the exact data related to all characters that have been inserted and deleted. It also provides information on modifications implemented into the target text and navigation within the text. As a result, it is possible to analyse the final translation product and its development during the different stages of the translation process (Jakobsen 2006, Kajzer-Wietrzny et al. 2016; Muñoz & Apfelthaler 2022). Since keylogging programmes also present numerical data, it is possible to conduct quantitative analysis (Whyatt 2012: 330).

Miller and Sullivan (2006: 2) observe that keylogging originates from writing research, where scholars were interested in how the written text is developed by its author. According to Jakobsen (2011: 38), translation process researchers were looking for a programme which could be triangulated with the already widely used verbal reports. He states that "[there] was a conviction that if it were possible to record converging qualitative think-aloud and quantitative behavioural data, our hypotheses about the translation process would gain important support" (Jakobsen 2011, p. 38). The need to develop a keylogging programme designed solely for translation research is understandable since, after the year 2000, most translations were already being done on computers (Jakobsen 2003: 69-70). Due to computer use, translators were able to accelerate their work, immediately make any corrections and direct their cognitive resources solely to the translation process. Therefore, Jakobsen (2003: 70) named this method 'type-along protocols.' Kajzer-Wietrzny et al. (2016, p. 6) describe it as "the cornerstone of the modern translation process research." There was a growing interest in introducing keylogging to translation studies at the end of the 20th and the beginning of the 21st century. Lauffer (2002: 62) explains that, initially, researchers used to put cameras in front of computer monitors to record the translation process.

The implementation of keylogging programmes has opened up new possibilities for researchers analysing the translation process, especially its cognitive aspects. For example, Serbina et al. (2015, p. 12) suggest that "[b]ased on this behavioural data and the intermediate version of translations, assumptions about cognitive processing during translation can be made." Keylogging is usually triangulated with eye-tracking or verbal reports. Since the works utilising eye trackers will be discussed in Section 2.5, here I would like to focus on the studies where keylogging was triangulated with retrospective reports.

One of the topics frequently analysed using keylogging is the characteristics of the writing process in translation and its differences in comparison to monolingual writing. This phenomenon was studied, for example, by Immonen (2006). The researcher used retrospective reports and keylogging to analyse pauses in writing and translating. The first task of the participants was to write a short text in their native language. In the second task, they performed an L2-L1 translation (from English into Finnish). It appeared that the location of pauses tends to be similar in writing and translation. Immonen (2006, p. 329) points out that "pause duration was the greatest between paragraphs and diminished down towards the smaller linguistic units." However, some differences in their length were observed (Immonen 2006: 329). In the case of larger segments, like paragraphs, pauses produced during writing are longer than the ones produced during the translation task. However, pauses appearing in shorter passages of text tend to be longer during the transition task. The data obtained through retrospective reports were mainly related to the decision-making process (Immonen 2006: 319).

Keylogging was also triangulated with verbal reports in the study by Buchweitz and Alves (2006). This time, the researchers used TAPs rather than retrospective reports. The topic of their study was very broad as it encompasses directionality, recursiveness and two groups of participants: professionals and students. The language pair was English-Portuguese, with Portuguese as the L1. As the researchers describe, they focused on the following types of data generated with the help of a keylogging programme, "the total number of pauses, the total number of keystrokes (text production and revision), and a total number of what is called translation units" (Buchweitz & Alves 2006, p. 248). Both groups of participants appeared to work longer when translating into their L2 and the L1-L2 direction also resulted in fewer revisions (Buchweitz & Alves 2006: 251, 254). An innovative project utilising the Task Segmented Framework to analyse keylogging data in terms of translation fluency was conducted by Muñoz and his colleagues (e.g., Muñoz & Caradona 2019; Muñoz & Apfelthaler 2022). The tool they developed allows for a very detailed analysis of typing behaviour during the translation process. They divided the translation process into segments and focused on a very advanced distinction between pauses and other units during which translators do not show any typing activities like delays and respites (Muñoz & Apfelthaler 2022).

2.4.1 Translog

It is impossible to discuss the method of keylogging without mentioning one of the most popular keylogging programmes, Translog (Carl 2012). It is the first keylogging programme designed with the aim of studying the translation process. It was developed by Jakobsen and Schou (1999) (Schou et al. 2009; Whyatt 2012; Saldanha & O'Brien 2013; Carl et al. 2016). This first version appeared in 1995 and since then has undergone many modifications (Muñoz 2009; Schou et al. 2009; Jakobsen 2006, 2011; Carl 2012). As pointed out by Muñoz (2009, p. 87), "[e]ver since Translog made it possible to record and measure typists' keyboard activities, it has been one of the favourite data collection tools." Writing fifteen years after releasing Translog, Schou et al. (2009: 40) found that it had been installed almost 100 times worldwide. By the time this work is being written, the numbers are substantially higher. The author of the programme, Jakobsen (2006, p. 96), describes its development in the following way: "[t]he original purpose for which Translog was created was to be an automatic, subject-independent tool for collecting hard, supplementary process data to the softer data collected by means of introspection, retrospection and think-aloud."

Probably the most characteristic feature of Translog is its division into two components. They are currently named Translog-Supervisor and Translog-User (Buchweitz & Alves 2006; Immonen 2006; Jakobsen 2006; Carl 2012; Whyatt 2012). As the name suggests, the first component is intended for researchers (Jakobsen 2006; Carl 2012; Whyatt 2012). They first design the experiment, adjusting features like the font size, and colour of both the ST and the TT, and insert the ST. The researcher can also decide on localising the ST and TT windows. After the data were recorded, the Supervisor component serves to analyse the data. Whyatt (2012, pp. 329-330) enumerates the following types of statistical data: "the duration of the TL production, the total number of user events (TUE), which include a keystroke or a cursor/ mouse movement, number of characters noted as text production (TP), all the text that was eliminated or erased by the translator (TE)." It is also possible to replay the translation process in real-time, which may serve as a cueing during a retrospective session (Carl 2012; Whyatt 2012). There are three options for data presentation: linear view, user view, and pause plots (Carl 2012: 4110-4111). The second component, Translog-User, has fewer functions and is intended for the participants. Here, their only task is to start the recording session and type the translation (Jakobsen 2006; Carl 2012; Whyatt 2012).

As pointed out by Carl (2012: 4109), the programme has three main functions: creating a project file, running and recording a Translog-II session, and replaying and analysing a recorded log file. As in the case of any keylogging programme, Translog records all keystrokes and mouse movements. In the latest version, they are saved as "1) insertion, 2) deletion (deletion and backspace), 3) navigation (cursor movements), 4) copy/cut-and-paste, 5) return key, or 6) mouse operations" (Carl 2012, p. 4108). Researchers often draw attention to the high precision of the recorded data (Jakobsen 2006; Carl 2012; Ferreira 2014; da Silva 2015; Kumpulainen 2015).

There are also other programmes utilised for keylogging. Nevertheless, none of them was designed particularly for translation process research (Carl 2012, Saldanha & O'Brien 2013). Researchers frequently use Inputlog (Leijten & Van Waes 2013). Likewise Translog, it records all the data and provides the researcher with detailed statistics. The main feature that distinguishes it from Translog is its environment. Inputlog operates within the Microsoft Word environment (Leijten & Van Waes 2006: 76). Therefore, it is usually used by researchers whose studies are conducted outside the laboratory conditions, on the translator's private computers, or in translation agencies. For example, it was implemented by Teixeira and O'Brien (2017) in their study related to workplace ergonomics concerning different translation tools that improve translators' work. They gathered the data in a genuine work environment, including real commissions, as well as computers that the translators usually work on. Therefore, the data-gathering process could not in any way disturb the natural work conditions.

2.4.2 Translation phases

Based on the functions of the keylogging programme, Jakobsen (2002) managed to study and divide the translation process into three broad phases: orientation, drafting and revision. The first phase, orientation, is described by Jakobsen (2002, p. 192) as

> the time delay between the appearance of the source text on the screen and the typing of the first text production key. The only kinds of keystrokes that can occur here are mouse clicks to scroll the source text and to resize to the source text window.

The aim of this phase is to familiarise participants of the study with the text that is supposed to be translated (Jakobsen 2002; Englund Dimitrova 2010; Dragsted & Carl 2013). During the analysis of translation styles, Dragsted and Carl (2013: 140) observed

that the orientation phase may differ across various translators and translation trainees. They found out that not all of them read the whole ST beforehand. Some start translating right after reading only part of the text and getting the main idea about its characteristics. As Dragsted and Carl (2013, p. 141) point out, this phase "includes the planning required for translating the first segment." The authors divide translators and translation trainees into three groups based on their planning behaviour (Dragsted & Carl 2013: 142). The first group is described as head-starters – people who do not read the whole ST beforehand but rather immediately start translating. The second category is known as quick planners, who also do not read the whole ST but at least devote some time to getting the main idea. The last group, scanners, "scan the text rapidly" (Dragsted & Carl 2013: 142).

Right after the orientation phase is finished, the drafting phase occurs. According to Jakobsen (2002, pp. 192-193), it "runs from the first text production keystroke until the first typing of the final punctuation mark (or equivalents keystroke)." This is the part of the translation process when the text in a foreign language appears on the computer screen. That is why Jakobsen (2002: 193) observes all kinds of keystrokes used in this phase. As the researcher continues, translators continuously fix their mistakes or adjust the text during the drafting phase. Therefore, they frequently go back to the ST window. Researchers triangulated keylogging and eye-tracking to analyse the so-called ST online orientation (Dragsted & Carl 2013: 142). They observed two tendencies: looking ahead to the next phrases that will be translated soon or looking back to the phrases that have just appeared in the TT. Moreover, various researchers also observe two types of revisions: online and end revisions (Jakobsen 2002, 2003; Malkiel 2009; Dragsted & Carl 2013). Online revisions denote modifying the TT text during the drafting phase, whereas end revision constitutes a separate translation phase (Jakobsen 2002; Bauchwietz & Alves 2006; Dragsted & Carl 2013).

Finally, there is the last phase, revision, during which "the text that was drafted in the middle phase [drafting] is reviewed. Typically, text production speed is considerably lower than in the middle phase because the main activity is monitoring existing text" (Jakobsen 2002, p. 193). These revisions are also known as end revisions (Jakobsen 2002, 2003; Dragsted & Carl 2013). In this case, translators usually want to check some of their translation solutions, move back to the ST, or improve punctuation (Jakobsen 2002: 193). Malkiel (2009: 150) draws attention to the fact that revisions do not always lead to corrections. Sometimes, translators present their ideas in a different, improved way. This

phase occurred only in part of translations analysed in the study by Dragsted and Carl (2013: 147). Often, lack of time forces the participants to skip it.

The length of orientation, drafting and revision phases, and the group differences between students and professionals were analysed, for example, in the study by Jakobsen (2002: 193-194). The researcher observed that drafting is the lengthiest of the three phases. It consumes about 70-80% of the whole translation time. Revision usually takes about 20%, whereas orientation takes only 1-2%. Moreover, there are substantial differences between students and professionals. It appears that, unlike students, professionals devote more time to the first and the last phase. The same phases, however, under different names also appear in other studies. For example, Carl and Kay (2011, p. 960) discuss:

> *skimming*, in which the translator acquires a preliminary notion of the content of the source text, *drafting* in which the actual transcription is typed; and revision, in which some or all of the translated text is reread, typos are corrected, and sentences possibly reformulated.

The three phases of translation are discussed in the context of various translation phenomena like directionality, cognitive effort and translation styles. For example, Whyatt et al. (2021) focused on using internet resources and their influence on the translation process. The study was conducted in the Polish-English language pair, based on two different text types: a product description and a film review. In the first phase, the number of online searches was significantly higher in the case of translating into the L1 for both text types. However, an opposite situation can be observed during the second phase. This time, the number of searchers in the L1-L2 translation outnumbered those in the L2-L1 direction. No significant difference was found in the last revision phase.

Another example comes from a study conducted by Alves et al. (2009), who, as one of the aims of their study, discussed the influence of directionality on the three translation phases. This time, keylogging was triangulated with eye-tracking and retrospective reports. The variable analysed with the keylogging programme was the duration of translation. It appeared that directionality might influence the time spent on each translation phase. The only statistically significant difference was found in the revision phase, which is longer than the L1-L2 translation (Alves et al. 2009: 278).

2.4.3 Advantages and drawbacks of keylogging

There are many advantages to using keylogging in translation process research. First of all, it is a non-invasive method. The only task the participant is obliged to perform is the translation itself (Jakobsen 2006; Whyatt 2012; Kajzer-Wietrzny et al. 2016). Keylogging programmes record the translation process in the background without interrupting it. Therefore, it does not require translators to split their attention between two separate tasks and imposes no additional cognitive load (Saldanha & O'Brien 2013: 133).

The second advantage is related to the type of data gathered through keylogging. These are very precise data that allow the researcher to analyse phenomena like recursive movements or pauses. It is impossible to gather this kind of data using any other method (Jakobsen 2006; Miller & Sullivan 2006; Sun 2011; Saldanha & O'Brien 2013; da Silva 2015; Kajzer-Wietrzny et al. 2016). Whyatt (2012, p. 328) points out that it "can be applied to study the interaction of cognitive processes by analysing the temporal patterns of text production and pauses in the flow of typing which reveal the dynamic processing involved in text comprehension and text production."

Another aspect is accessibility. The two most popular keystroke loggers, Translog and Inputlog, can be downloaded and used without fees. Moreover, the research centre in charge of Translog II provides a database that allows researchers to reuse the data gathered by other scholars (Muñoz & Apfelthaler 2022: 9-10). An important feature of keylogging is the possibility of its triangulation with other methods (Whyatt 2012; Saldanha & O'Brien 2013). By joining keylogging with eye-tracking, the researchers get the information that is not restricted to writing behaviour (Jakobsen 2011: 46-47).

It is recommended to triangulate keylogging with other methods. Although keylogging offers detailed numerical data, it cannot fully enter into the translator's black box. Namely, the analysis of the keylogging data does not provide the reasons behind the observable processes (Jakobsen 2006, 2011; Sun 2011; da Silva 2015). Therefore, Saldanha and O'Brien (2011: 133) emphasise the need to triangulate keylogging with more subjective methods like TAPs or retrospective reports. This can result in detailed numerical data and verbalisations describing mental processes going on in the translator's mind. However, it should always be remembered that neither verbal report provides complete data. There is no guarantee that what the translator says precisely depicts their thoughts.

The second drawback results from translators' characteristics and preferences. Although working on a computer is nowadays a common practice (Jakobsen 2003; Saldanha & O'Brien 2013), not all translators will type at an identical speed. There is a risk that some of them may fall behind the rest of the group. According to Saldanha and O'Brien (2013: 134), typing speed may be a serious confounding variable. As the researchers continue, personal preferences, such as computer or keyboard type, may also have an impact on the recorded data. This aspect should not be ignored, especially when the experiment is conducted in laboratory conditions since all participants should work on the same equipment.

2.5 Eye-tracking

Eye-tracking is classified as the third generation of TPR (Alves 2015:22-23). It was developed at a time when translation scholars not only based their works on triangulation but were also searching for more and more accurate data-gathering methods. Eye-tracking is broadly described by Saldanha and O'Brien (2013, p. 136) as "the process of recording the point of gaze of a person and the movement of the eyes from one point to another." To record the behaviour of human eyes, researchers use equipment called the eye tracker. Like most translation process research methods, eye-tracking has not been developed by translation scholars but borrowed from other fields, mainly cognitive science and psychology (Hvelplund 2017a; Doherty 2018). For example, in the past, it "was most popular in aviation medicine, when researchers examine the response of people in extreme conditions triggered in flight simulations" (Płużyczka 2018, p. 113). Although eye-tracking has been applied relatively recently in translation studies, researchers became interested in the behaviour of human eyes in the 19th century (Grucza 2011; Płużyczka 2018).

A detailed description of the development of the eye-tracking method and the innovations applied to the eye trackers lies beyond the scope of this study. However, it has been thoroughly described, for example, in works by Płużyczka (2018) and Stachowiak-Szymczak (2019). To present a brief picture of the eye-tracking history, the first non-invasive eye tracker was used in the first half of the 20th century (Grucza 2011; Płużyczka 2018). Nowadays, eye trackers use infrared light to register eye movements (Saldanha & O'Brien 2013; Korpal 2015; Płużyczka 2019a). Its functioning is described, for example, in the study by Teixeira and O'Brien (2018, p. 35), who state that:

> most existing eye tracker models use a combination of small infrared cameras to capture the position of the eyes in relation to the screen at very short intervals, defined by the eye tracker's operating frequency. […] The position is a coordinate on a xyz axis having the screen as a reference (x and y are the horizontal and vertical positions of the gaze on the screen, while z is the difference between the eyes and the screen surface).

The first stage of an eye-tracking study is the calibration of the equipment. According to Teixeira and O'Brien (2018: 42), during this part of the study, the equipment "learn[s] the characteristics of each participant's eyes, such as eye size and distance between eyes, as well as to know how the eyes are positioned when fixating specific points on the screen." The participant's task is to focus their eyes on a small point that appears in different places on the screen. After that, the researcher obtains results that provide information on how accurate the data are (how close the eye was to the given point) and can make some corrections, for example, in participants' distance from the computer. Currently, there are different types of eye trackers (O'Brien 2009; Grucza 2011; Korpal 2015; Saldanha & O'Brien 2013; Duchowski 2017; Hvelplund 2017a; Teixeira & O'Brien 2018). Researchers have at their disposal eye trackers which can be placed on the desk, attached to the computer monitors, the ones which are easily portable, or even glasses that the participant puts on.

According to Płużyczka (2013, 2019a), the first works applying eye-tracking in translation process research appeared after the year 2000. With time, the method gathered greater popularity. Hvelplund (2017a: 251) even suggests that the growing popularity of eye-tracking leads to the point where it is frequently used either as the main research method or triangulated with other methods. Several factors contribute to its successful adaptation in the translation process. First of all, it makes it possible to gather an entirely new type of data. While keylogging explores the writing process, eye-tracking shows how translators read the texts and what they look at (Płużyczka 2018, 2019a). The significance of this kind of data in the translation process research can be justified by the observation that "sight is for humans one of the most important senses. Nearly 80% of all sensory impressions are delivered to the brain via the visual channel. Vision provides also information at the highest level of speed" (Płużyczka 2018, p. 101). It is, therefore, a crucial factor in the case of translation.

Secondly, like keylogging, eye-tracking is a non-invasive method (Zwierzchoń-Grabowska 2011; Korpal 2015). The human eye is not able to detect infrared light. Because of that, the eye is not distracted and can work normally. Calibration is the only phase that draws the participant's attention to the equipment. In the following steps, they

are not required to perform any additional task or to divide their attention (Zwierzchoń-Grabowska 2011: 204). What is more, eye-tracking provides objective data. It registers the behaviour the participant is usually unable to influence (Zwierzchoń-Grabowska 2011; Płużyczka 2019b). The registered data is also characterised by very high precision. Grucza (2011: 157) points out that the eye-tracker can detect even the smallest and fastest eye movements. Finally, the gathered data have a quantitative form that enables statistical analysis (Zwierzchoń-Grabowska 2011; Saldanha & O'Brien 2013; Doherty 2018; Płużyczka 2019a). As a result, the hypotheses can be accepted or rejected based on solid numerical data. Apart from columns and tables, data can also be presented in various graphic forms, like heat maps, which show the points where participants most frequently looked (Saldanha & O'Brien 2013; Płużyczka 2019a). Therefore, Chmiel and Mazur (2013, p. 192) call it a "new window to the translator's mind." Based on these features, it could be stated that eye-tracking has brought many new possibilities to translation process research.

2.5.1 Eye-mind hypothesis

It is impossible to discuss the eye-tracking method without mentioning the eye-mind hypothesis. As mentioned by Hvelplund (2017a, p. 250), "[t]he analysis of eye movement data in any kind of research rests on the notion that the focus of visual attention can tell us something about the focus of cognitive attention." In this case, most authors refer to the eye-mind hypothesis developed by Just and Carpenter (1980). Based on their study of the reading process, the researchers observed that participants fixated their eyes on every word in the text, except for some short words, especially functional words. This observation led them to formulate the eye-mind hypothesis, which states that "the eye remains fixated on a word as long as the word is being processed. So the time it takes to process a newly fixated word is directly indicated by the gaze duration", and as a result, "there is no appreciable lag between what is being fixated and what is being processed" (Just & Carpenter 1980, pp. 330, 331). Therefore, these words that are considered difficult or less understandable tend to be fixated on longer. It is also possible that the eyes move back (the so-called refixation) to read a problematic word one more time (Just & Carpenter 1980: 331). These observations allow the researchers to study the reading cycle. Following this assumption, translation process researchers frequently analyse fixations as indicators of cognitive effort (e.g., Jensen et al. 2009; Hvelplund 2017b). For example, when describing their variables, Pavlovič and Jensen (2009, p. 98) state that "an

increase in average fixation duration is considered synonymous with increased cognitive effort."

Nowadays, the eye-mind assumption did not avoid criticism. The main objection raised by some translation researchers is briefly summarised in the work by Hvelplund (2017b, p. 73), who states that

> [t]he basic assumptions entail that *only* ST reading related activities take place during ST reading and *only* TT reading related activities take place during TT reading. While this is certainly a reasonable assumption for the vast majority of the time, there are instances during which the translator works with tasks other than those indicated by the type of reading.

Evidence confirming this observation can be found in the study by Balling et al. (2014), who suggest that translating is a parallel process. It means translators simultaneously perform two tasks, the ST and the TT processing. The researchers analysed the translation of segments with different word order in the English-Danish language pair. They observed that:

> segments that are incongruent between the source text and the target text are looked at longer than congruent segments, indicating that processing is parallel since the necessity for transposition of word order in the TT seems to be anticipated during reading of the ST.
>
> (Balling et al. 2014, p. 246)

Evidence of parallel processing is also found in the study by Hvelplund (2017b: 70) when translators fixated their eyes on the ST, but at the same time, they were already writing the TT.

Since the translation process requires concurrent target text typing, looking at the computer screen and processing the ST and TT, such an attention split may be challenging for some translators. The issue of attention split was analysed in the study by Sharmin et al. (2008), who focused on changes in visual attention under different conditions. Participants' typing skills may also influence the obtained data. The results revealed that people with better typing skills are able to devote more visual attention to the screen, which can be observed based on the higher number of fixations and longer total gaze time (Sharmin et al. 2008: 37-39). However, the results of the spatial saccadic movements study by Płużyczka (2016, 2019) may provide evidence that translators constantly process the text even if they look away from the computer screen.

2.5.2 Basic variables registered by eye trackers

Researchers usually enumerate three types of eye movements registered by eye trackers. These are fixations, saccades, and pupil dilation (Chmiel & Mazur 2013; Korpal 2015; Hvelplund 2017a; Stachowiak-Szymczak 2019). Fixations are "eye movements which stabilise the retina over a stationary object of interest" (Duchowski 2017, p. 44). It means that during a fixation, the eye is not moving but rather keeps still and focuses on an object or word (Stachowiak-Szymczak 2019: 66). According to researchers, the average time when the eye is not moving is around 200-250 ms (Jakobsen & Jensen 2008; Korpal 2015). However, this observation also has its opponents. It is suggested that every eye behaves differently, and the fixation time may be longer (Just & Carpenter 1980: 330). It is widely believed that fixation is the time when the eye registers information (Korpal 2015; Stachowiak-Szymczak 2019). Fixations can be analysed based on their average or total duration (e.g. in, Jakobsen & Jensen 2008; Jensen et al. 2009; Pavlovič & Jensen 2009; Hvelplund 2017b; Whyatt et al. 2021), and their number in a particular part of a text or the whole task (e.g. in, Jakobsen & Jensen 2008; Chmiel & Mazur 2013; Hvelplund 2017; Płużyczka 2019a).

Saccades are "rapid eye movements used in repositioning the fovea to a new location in the visual environment" (Duchowski 2017, p.40). In a more graphic description, Stachowiak-Szymczak (2019, p. 66) names them " 'jumps' from one fixation to another." The eye does not register any information during saccades (Balling et al. 2014; Płużyczka 2016; Stachowiak-Szymczak 2019). The analysis of saccadic movements is thoroughly described in the works by Płużyczka (2016, 2019b), who focuses on spatial saccadic movements during sight translation. The researcher observes that when encountering translation problems, translators tend to move their eyes away from the computer screen. This can easily be spotted on the gaze maps produced by the eye tracker. Płużyczka (2019b: 51-52) continues that such behaviour may be an observable result of referring to long-term memory, where the translator attempts to find some solution.

Pupil dilation refers to pupils' "size or diameter" (Stachowiak-Szymczak 2019, p. 67). It denotes one of the eye's behaviours that a person is not able to control or influence (Balling et al. 2014: 238). Therefore, it is frequently measured as an indicator of cognitive load (e.g., in Jensen et al. 2009; Pavlovič & Jensen 2009; Hvelplund 2017). Nevertheless, many scholars draw attention to its susceptibility to environmental conditions like light

or emotions, which may make the data unreliable (O'Brien 2009, Saldanha & O'Brien 2013, Płużyczka 2019b).

2.5.3. Eye-tracking in translation and interpreting process research

It can be observed that eye-tracking is widely used in translation process research, especially when triangulated with other methods. It makes it possible to study additional aspects of the translation process, like reading. It was studied, for example, by Jakobsen and Jensen (2008), who analysed eye movements in four different types of reading: reading for comprehension, reading in preparation for translating, reading during sight translation, and finally, reading while typing a written translation. Their research project included both students and professional translators. The researchers focused on four variables: "1) task time, 2) the total number of fixations ('fixation count'), 3) the total duration of all fixations during the execution of the task ('total gaze time), and 4) the average duration of individual fixations" (Jakobsen & Jensen 2008, p. 108). The results show that various reading purposes significantly influence the observed eye behaviour in each of the four variables. The task characteristics, especially the possibility of moving back to the text or remembering as much necessary information as possible, increase or decrease the obtained results (Jakobsen & Jensen 2008: 120-121).

The topic of reading and translating was further developed almost ten years later by Hvelplund (2017b). He focused mainly on reading during the translation process and enumerated its four types. These are "ST reading, ST reading while typing a text, TT reading (reading of existing TT text), and TT reading while typing a text (reading of emerging TT)" (Hvelplund 2017b, p. 60). This time, the researcher triangulated eye-tracking with keylogging and based his research on pupillometry and fixation analysis. He observed significant differences resulting from the task the translator is engaged in. It appears that translators have to invest more effort in TT processing (Hvelplund 2017b: 69). The highest cognitive effort is evoked by concurrent TT typing and reading (Hvelplund 2017b: 72).

Eye-tracking is also implemented to investigate various aspects of directionality in translation, especially in the context of cognitive effort. One of the first directionality and cognitive effort studies that based the analysis solely on eye-tracking was conducted by Pavlovič and Jensen (2009). This topic was also studied, for example, by Whyatt (2018, 2019), who triangulates eye-tracking with keylogging. In a subsequent study, eye-tracking served to analyse participants' eye behaviour while using external resources in

both directions of translation (Whyatt et al. 2021). Another example of implementing eye-tracking in studies on directionality is the work by da Silva et al. (2017). He analysed the influence of directionality on post-editing based on variables like fixation count and total reading time.

There have also been attempts to use eye-tracking in studies conducted at the translators' workplace, for example, by Teixeira and O'Brien (2018). As they observe, this practice involves additional obstacles. There is another problem besides all the necessary agreements that researchers must get to conduct their experiment and recruit participants. Most professional translators usually utilise two screens in their work. Still, the eye tracker can be connected to only one of them (Teixeira & O'Brien 2018: 39). The constant switching between the computer screens results in considerable data loss. Nevertheless, in the case of Teixeira and O'Brien's study (2018, p. 46), eye-tracking was not the main research method as "[t]he eye-tracking data were used mainly to obtain an initial understanding of how translators interact with the different elements of a standard CAT tool" and, thus the encountered obstacles did not influence the results negatively.

Eye-tracking is also used to study various processes in interpreting. Sight translation was investigated, among others, by Chmiel and Mazur (2013). The researchers implement eye-tracking to track the progress made by students and examine whether there are any behavioural differences among the two groups of students. Gumul and Pietryga (manuscript under preparation) used eye-tracking to study explicitating shifts in sight translation among translation trainees. An example of eye-tracking used to analyse dialogue interpreting can be found in the study by Tiselius and Sneed (2020). They focused on cognitive load and directionality issues, which were analysed based on the gaze patterns. Stachowiak-Szymczak (2019) used eye tracking to study the changes in eye behaviour during simultaneous and consecutive interpreting as a reaction to various types of cognitive load. This can be related to some specific aspects of speech that the interpreter is interpreting or to the visual stimuli that come to the interpreter's eyes.

2.5.4 Drawbacks of eye-tracking

Although eye-tracking has yielded many new possibilities for the translation process research, it also has some drawbacks. One of them is data loss, which may result from the human body's natural behaviour, especially its susceptibility to environmental conditions (Doherty 2018: 80). One of the most frequently found problems is the influence of light (O'Brien 2009; Saldanha & O'Brien 2013; Korpal 2015; Płużyczka

2019). Saldanha and O'Brien (2013, p. 138) comment that "[i]t is well known that light affects pupil size, with the pupil constricting in high-intensity light environments and dilating in low light." Because of that, lighting conditions could easily disrupt the results, especially if one of the variables is pupil dilation. Another problem is related to different lighting conditions among participants. This can make the obtained data unreliable and impossible to compare. Therefore, it is highly recommended to keep as identical lighting conditions among all participants as possible (Saldanha & O'Brien 2013: 138). Light is not the only factor that affects pupil dilation. An effect of larger pupils may also be the result of various stimulants, such as coffee or stress (O'Brien 2009; Saldanha & O'Brien 2013; Korpal 2015). Likewise, sound may have a negative influence on eye behaviour (O'Brien 2009; Saldanha & O'Brien 2013).

Another frequent problem during eye-tracking studies is the participant's movements (O'Brien 2009; Saldanha & O'Brien 2013). Participants should sit as still as possible, and high emphasis is put on not moving their heads. It is possible to use chin rest to prevent participants from making unexpected, rapid movements (Saldanha & O'Brien 2013: 139). However, such conditions are highly uncomfortable and, in the case of interpreting studies, impossible to implement. A less invasive method of preventing participants from significant movements is proposed by Saldanha and O'Brien (2013, p. 139), who suggest that "[a] practical way from preventing substantial changes in head position (…) is to provide a non-swinging chair, but the one adjustable for height." Apart from the physical body or head movements, participants should be able to work without devoting most of their work time to looking at the computer keyboard (O'Brien 2009: 257). The eye tracker can gather the data only when participants look at the screen. Otherwise, the gathered data are incomplete and very poor.

As Saldanha and O'Brien (2013, p. 139) pointed out, "some participants may just not be suited to the collection of gaze data." Some obstacles in data collection may also be caused by eye makeup or various eye conditions (O'Brien 2009; Saldanha & O'Brien 2013; Korpal 2015). The first problem may be minimised by asking participants beforehand not to wear makeup on the day of the eye-tracking experiment. However, eye conditions may result in more significant issues. Serious problems may also be related to eye colour. O'Brien (2009: 257) observes that the equipment does not always track subjects with dark eyes correctly.

2.6 Neuroimaging techniques

As previous sections have suggested, translation and interpreting process research often benefits from methods developed in other disciplines. Increasingly rapid scientific development and the emergence of new equipment allow researchers to study behavioural measures and brain activity (Saldanha & O'Brien 2013; Kemmerer 2015; Garcia et al. 2016; Garcia 2019). Such an observation is possible, among others, by non-invasive methods such as electroencephalography (EEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI), which are assigned by the scholars to the current fourth generation of translation studies (Mellinger & Hanson 2020: 173). These methods started to be applied to translation studies in the 21st century. They were borrowed from the field of medicine, where they are widely used, for example, to study various brain diseases or damages (Hansen-Schirra 2017; Garcia 2019; Płużyczka 2019a; Borghardt et al. 2021). As has been observed, neuroimaging techniques have opened a new spectrum for interpreting and translation studies (Zwierzchoń-Grabowska 2011: 199). For example, Garcia et al. (2016, p. 22) state that "the inclusion of neuroscience methods in translation studies could be critical to understanding how translation and interpreting mechanisms are embedded into other neurocognitive domains and, more generally, within the human organism."

Like the previously discussed process methods, EEG, fMRI, and PET have some weaker points. One of them is that they provide high-quality data solely in spatial or temporal resolution. In the case of high-spatial-resolution, the researcher obtains "accurate information about which brain region is implicated in a given process" (Garcia et al. 2016, p. 23). Both PET and fMRI possess this feature (Garcia et al. 2016; Garcia 2019). The high-temporal resolution "reveal[s] precisely when that process took place" (Garcia et al. 2016. 23). EEG is classified within this group (Garcia et al. 2016; Garcia 2019; Hervais-Adelman & Babcoc 2020; Borghardt et al. 2021).

Hervais-Adelman and Babcock (2020: 741) describe two additional categories of neuroimaging methods classification: functional and structural. Functional studies aim to "reveal which brain areas are implicated in the execution of SI [simultaneous interpreting]" (Hervais-Adelman & Babcock 2019, p. 741). They usually use methods like PET and fMRI. Structural study design "aim[s] to examine the relationship between brain structure and SI" and implements methods like EEG but also fMRI (Hervais-Adelman and Babcock 2019, p. 741).

Researchers also advise triangulating neuroimaging techniques with other methods (Garcia et al. 2016: 40). This practice can be observed, for example, in the study by Borghardt et al. (2021), who used three methods: eye-tracking, EEG, and fMRI, to study the processing of Easy Language. Moreover, neuroimaging studies, especially those using PET and fMRI, should always have some control conditions (Garcia et al. 2016; Hervais-Adelman & Babcock 2019). Only then researchers are able to distinguish the brain regions that are activated during translation or interpreting. That is why, besides interpreting, one of the most frequently analysed phenomena is probably bilingualism (Hervais-Adelman et al. 2011, 2015).

2.6.1 EEG

Electroencephalography (EEG) is a method that serves to "measure electric activity on the human scalp by using (…) electrodes (…). When a participant perceives auditory or visual stimuli, the neurotransmissions move to the scalp surface" (Borghardt et al., 2021 p. 4). The number of electrodes placed on the scalp may differ in each study. A higher number of electrodes creates the possibility of obtaining better spatial resolution of the sources of the recorded electrical activity (Kemmerer 2015: 62). It has been observed that the electric activity of the brain can be divided into four categories, the so-called frequency bands (Garcia 2019: 66). These are

> - Alpha waves (α): Alpha waves are neural oscillations occurring at the frequency of 8-13 Hz. They represent a relaxed waking state with closed eyes.
> - Betha waves (β): Beta waves occur at a frequency of 13-30 Hz. These physiological waves represent both relaxed states as well as mental activity.
> - Theta waves (θ): Theta waves occur at a frequency of 4-8Hz. Depending on the test subject, they can be either physiological (infants or babies in waking state or adults in sleep stages II and III) or pathological waves (adults in waking state).
> - Delta waves (δ): Delta waves occur at a frequency of 0.5-4 Hz and are characterised by bow- to trapezoid-shaped form. For infants as well as adults sleep stages I and II, they are regarded as physiological, whereas they are considered pathological for the waking EEG of adults.
>
> (Hansen-Schirra 2017, p. 239)

EEG was first used in the 20th century. Like all neuroimaging techniques, it has been employed in various studies related to brain diagnostics. However, it has also provided many interesting data in linguistic research (Kemmerer 2015; Hansen-Schirra 2017; Garcia 2019; Borghardt et al. 2021).

One of the most frequently used forms of the EEG study is the analysis of so-called event-related potentials (ERPs). ERPs are defined by Borghardt et al. (2021, p. 4)

as "[t]he potentials connected to a specific event (…) [they] appear as soon as the participant perceives a stimulus." Garcia et al. (2016, p. 30) provide further details of ERPs analysis, stating that their study is based on "averaging evoked responses (…) The result is a waveform featuring negative and positive peaks, whose amplitudes and latencies can be compared across conditions." ERPs are examined, for example, in the study by Janyan et al. (2009), who analysed cognate processing based on interpreting. The ERPs results of participants whose task was to interpret single words revealed that the human brain processes cognates and non-cognates differently, depending on their concreteness level.

However, it should be noted that EEG studies are not free from problems. First and foremost, EEG is known to be a very sensitive method. Signals can be easily distorted by human behaviour, like the movements of the eyes or muscles (Saldanha & O'Brien 2013; Hervais-Adelman & Babcock 2019; Garcia 2019; Borghardt et al. 2021). Borghardt et al. (2021: 6) even suggest conducting EEG experiments in a soundproof booth. That is why it is not recommended for participants to read long passages of texts (Hansen-Schirra 2017; Garcia 2019; Borghardt et al. 2021). Rapid eye movements, as well as additional processing which occurs during this task, may negatively influence the data. To overcome such distortion, stimuli can be presented at a slow rate. Scholars also tend to show their participants single words. This constitutes a severe obstacle in the case of translation and interpreting studies, as the translation and interpreting processes never implement single words or sentences.

2.6.2 PET and fMRI

Positron emission tomography (PET) can be defined as a method that "measures regional cerebral blood flow – and, by inference regional neural activity – by tracking the distribution of a radioactive isotope throughout the brain" (Kemmerer 2015, p. 44). Developed in the second half of the 20th century, PET was used mainly until the beginning of the next century (Garcia et al. 2016; Garcia 2019). Because of its good spatial resolution, it can provide information about the exact localisation of active brain regions during particular processes (Kemmerer 2015; Garcia et al. 2016). For example, Rinne et al. (2000) studied simultaneous interpreting and directionality using PET. They found that the direction of interpreting influences cerebral activation patterns.

Nevertheless, it has some serious drawbacks. First of all, the isotope has to be injected directly into the bloodstream to conduct the study (Kemmerer 2015; Garcia et

al. 2016; Garcia 2019). Although the amount is not great and it leaves the bloodstream relatively rapidly, some ethical issues arise and precautions must be taken (Kemmerer 2015: 44). Worth mentioning are also high costs, as well as the need for a highly specialised group of experts who are authorised to work with the equipment and maintain all the safety standards (Garcia et al. 2016: 25). Currently, PET's place has been overtaken by fMRI (Kemmerer 2015: 46).

Functional magnetic resonance imaging (fMRI) is defined by Borghardt (2021, p. 6), who explains that it

> is a neuroimaging technique that uses MRI scanners to investigate changes in brain function. MRI is based on the use of a very strong magnetic field, which energetically excites hydrogen atoms in the body. The energy emitted later can be measured and localised.

As pointed out by Kemmerer (2015: 35), the strength of this magnetic field has a particular influence on the quality of the obtained data. A crucial observation for any fMRI study is that blood flow increases together with neural activity (Kemmerer 2015; Borghardt et al. 2021). Garcia et al. (2016, p. 27) emphasise that "[d]uring cognitive tasks blood flow increases in relevant areas (…) generating a slight shift in the MRI signal known as blood-oxygen-level-dependent (BOLD) effect." The history of fMRI studies is relatively short. The method itself only appeared at the end of the 20th century. However, its frequent application since the year 2000 has resulted in many studies being published during this time (Kemmerer 2015; Garcia 2019). Besides an excellent spatial resolution, fMRI does not require an injection of any substance. Nevertheless, its costs are still high but below those produced by PET (Kemmerer 2015: 45).

fMRI is widely implemented in the interpreting process research by Hervais-Adelman (e.g., 2015, 2020) and his colleagues. For example, the study from 2015 aimed to analyse brain activity and language control during simultaneous interpreting (Hervais-Adelman et al. 2015). As in the case of the remaining methods, it is impossible to interpret long text passages while being placed in the MRI scanner. Therefore, participants interpreted only four sentences long stories. Shadowing and passive listening served as control conditions, and their results were compared to those from the interpreting task. The results have shown that regions activated during interpreting also activate during other tasks, like shadowing. Although the number of such areas observed by Hervais-Adelman et al. (2015: 4733) is greater than in other works, their results comply with the previous studies. It can be concluded that there is no single region activated only during interpreting (Garcia et al. 2016).

Nevertheless, scholars using fMRI have to struggle with some difficulties. One of them is the extreme noise caused by the scanner (Kemmerer 2015; Hervais-Adelman & Babcock 2019). Hervais-Adelman et al. (2015) used the sparse imagining protocol to overcome this problem. They were able to record the utterances during the silent pauses between the scans. The spatial conditions are also rather uncomfortable, as the participants are required to lie motionless (Hervais & Babcock 2019; Borghardt et al. 2021).

2.7 Questionnaires

In this short section, I would like to briefly discuss the implementation of questionnaires in translation and interpreting process research. As Mellinger and Hanson (2020:174) observe, in contrast to other fields like social science, questionnaires play a relatively minor role in the case of translation and interpreting studies. However, they "can help with understanding the various traits and characteristics of translators and interpreters, on the one hand, and the users of language services, on the other" (Mellinger & Hanson 2020, p. 174). In contrast to methods like keylogging, eye-tracking, and neuroimaging, surveys are usually conducted after the translation process has already been finished. Thus, they are an offline method or a post-hoc measure (Borghardt et al. 2021:7).

The popular belief that not much effort is required to construct, conduct, and then analyse questionnaires, especially compared to other methods, is quite misleading (Saldanha & O'Brien 2013; Mellinger & Hanson 2020). An ill-conceived design may easily lead to a low response rate and unreliable data. Describing the survey methodology, Saldanha and O'Brien (2013, p. 153) enumerate four main mistakes:

> the first is a coverage error, which refers to when some part of the population is not included in the study. The second is a sampling error when some parts of the population have a higher probability of being included in a survey than other parts of the population. The third error type is known as a nonresponse error, which arises when members of the sample do not answer the questionnaire at all or answer only some questions. The fourth error type is a measurement error, which occurs when the actual response (i.e. the answers given) differs from the 'true' response (i.e. the facts or beliefs of the participants).

One of the most important features of a proper survey is the questionnaire design. Saldanha and O'Brien (2013: 153-159) emphasise the desirability of questions. They suggest that the questions presented in the questionnaire should strictly correspond to the research questions. Researchers should pay attention to their structure, avoiding the ones

which are too long, ambiguous, or difficult to follow. One can choose between open and closed questions or the Likert scale (Saldanha & O'Brien 2013; Mellinger & Hanson 2020).

Moreover, Saldanha and O'Brien (2013) discuss some disadvantages the researchers should be prepared to face when deciding to conduct a survey. First of all, there is an issue related to personal and sensitive data and the anonymity of the participants. It may raise various ethical questions. Both the sample size and the type of questions should not impair participants' anonymity (Saldanha & O'Brien 2013: 154; 161-163). Secondly, questionnaires may also be affected by the so-called white-coat effect, for example, when the respondents modify their answers either due to the fear of the researchers or a desire to help them in their study (Saldanha & O'Brien 2013: 153).

There are three main aims of questionnaires in the case of translation process research: "to collect background information on research participants; to collect data on facts, opinions, attitudes, behaviour, etc. or to combine the collection of both" (Saldanha & O'Brien 2013, p. 152). The first type can easily be found in studies on directionality, where researchers frequently ask about the years of translation and interpreting experience, types of commissions, or the direction in which translators and interpreters perform their services. For example, such questions were asked in studies by Pavlovič (2007a) and Whyatt and Kościuczuk (2013), who analysed the situation of directionality on the translation markets in their countries. It is worth mentioning that the studies by Bartłomiejczyk (2004) and Pietryga (2022) utilise questionnaires as the only research method. Questionnaires providing information about participants' observations or feelings can also be found outside the topic of directionality. For example, in the study by Gumul (2020b), the survey was an additional research method through which students described their opinions about the types of cueing used during the retrospective session. Researchers can either construct their own questionnaires or apply a ready-to-use template of scales. Such a practice was adopted, for example, by Chen (2017), who employed the NASA Task Load Index in her preliminary study focusing on cognitive load.

Chapter 3. Cognitive effort in the translation process

The aim of this chapter is to discuss the phenomena of cognitive effort and cognitive load. The chapter opens with definitions of both terms. In the next section, I discuss the Effort Model by Gile (1995), as well as two models focusing on cognitive load. These are the Cognitive Load Model by Seeber (2011) and the Cognitive Load Theory by Sweller et al. (1998). The last model refers to pedagogy. Section 3 is devoted to indicators of cognitive effort. They are organised into four measures: subjective, physiological, behavioural, and performance. A subsequent section presents an overview of studies on directionality and cognitive effort both in translation and interpreting. The lack of conclusive results from these studies constitutes a point of departure for my PhD study. Finally, the last section is devoted to default translation, a new perspective focusing on structures that decrease cognitive effort.

3.1. Definitions of cognitive effort and cognitive load

Translation and interpreting process researchers agree that translating and interpreting are very demanding and effortful activities (Ehrensberger-Dow et al. 2020; Gile & Lei 2020; Hunziker Heeb 2020, Hunziker Heeb et al. 2021). For example, Ehrensberger-Dow et al. (2020: 220) observe that substantial cognitive resources are required to understand the source text and produce the target text. Thus, such multitasking may be highly problematic for beginning translators and interpreters as well as translation and interpreting trainees. Gile and Lei (2020, p. 263) discuss this process in detail and point out the following elements:

> the search for ad hoc information, both linguistic and extra-linguistic, to identify target-appropriate lexical and phraseological usage, as well as reflection and decisions on how to address the tension arising from differences in information-explicitness requirements between the source and target language.

However, activities like translating and interpreting established phrases that are well-known are considered to be relatively effortless (Gile & Lei 2020: 263). Moreover, manual aspects of the translation process, like typing, are considered to be effortless and automatised (Jakobsen 2003; Hvelplund 2016; Hunziker Heeb 2020).

Researchers frequently use names like cognitive load (e.g., in Seeber 2013; Gumul 2021a) and cognitive effort (e.g., in Ferreira et al. 2016; Whyatt 2018) to describe difficulties resulting from the translation and interpreting process. Nevertheless, as has

been observed, it is highly difficult to precisely define both terms (Seeber 2013; Hunziker Heeb 2020; Ehrensberger-Dow et al. 2020; Gieshoff &Hunziker Heeb 2023). Gieshoff and Hunziker Heeb (2023, p. 4) present a very graphic description according to which "load can be understood as corresponding to the price tag of a task whereas effort is the amount the person performing is willing to pay." Many researchers have attempted to define both phenomena. For example, Seeber (2013, p. 19) claims that "on the one hand, this construct represents the load imposed on the performer by a particular task. On the other hand, it represents the perceived effort invested by a performer during the execution of that task." Although he uses both terms in one definition, it can be seen that they have different functions. Cognitive load is an external phenomenon, while effort is an internal one. A similar relationship can be concluded from the definition presented by Ehrensberger-Dow et al. (2020, p. 221), who "associate cognitive load with the complexity of the stimuli and task (i.e., source text, situation, commission and so on), and cognitive effort with an actual response by the task performer." Nevertheless, in my study, I will follow the definition of translator effort developed by Hunziker Heeb (2020, p. 48), who perceives it as "the total effort the translator expends during the translation task … The target text is then the product of this translator effort." Not only does it precisely refer to the translation situation, but also its author claims that the terms translator effort and cognitive effort are synonymous (Hunziker Heeb 2020: 66).

It should be noted that cognitive effort management is considered to be an individual matter. Translators and interpreters not only decide how much attention and effort is needed to translate or interpret a text, but also they should be able to adjust their effort to the changing circumstances (Ehrensberger-Dow et al. 2020; Hunziker Heeb 2020; Hunziker Heeb et al. 2021). It is also worth mentioning that humans possess only a limited amount of cognitive resources (Hunziker Heeb et al. 2021: 52). Thus, more difficult tasks will require translators and translation trainees to invest more cognitive resources to successfully accomplish the task (Gieshoff & Hunziker Heeb 2022). That is why Hvelplund (2016: 149) draws attention to the significance of cognitive efficiency, according to which translators are able to effectively manage their cognitive resources, investing them only in those processes that require such effort.

Although cognitive effort is a frequently discussed topic, it only appeared in the field of translation and interpreting research in the second half of the 20th century (Seeber 2013; Gile & Lei 2020). Initially, it was discussed mainly in the context of simultaneous interpreting, as it was believed that interpreters constantly work on the verge of the

saturation level of their cognitive resources. This phenomenon is known as the tightrope hypothesis (Gile & Lei 2020: 263, 274). Its popularity in interpreting studies is reflected in the two most common models of cognitive effort and cognitive load, which I will discuss in the next section. Nevertheless, Gile and Lei (2020: 265) also observe a growing interest in cognitive effort and cognitive load in the field of translation. For example, Michaela Albl-Mikasa and her team received funding from the Swiss National Science Foundation to conduct a project entitled Cognitive Load in Interpreting and Translation (CLINT).[8] Translation and interpreting researchers and neuropsychologists cooperated to study processing English as a lingua franca among bilinguals, interpreters, and translators. According to Gile and Lei (2020: 265), cognitive effort studies can result in important advantages and new findings. First of all, cognitive effort can be investigated in the context of its influence on the target text quality, as "[t]he relation [between cognitive effort and quality] is not straightforward" (Gile & Lei 2020, p. 265). Secondly, it can help to analyse other phenomena, like directionality (analysed in this study), explicitation (e.g., Gumul 2021a and the CLESI project[9]), emotions (e.g., Hunziker Heeb et al. 2021), and post-editing (e.g., Lacruz 2017).

3.2 Models of cognitive effort and cognitive load

As has already been pointed out, two crucial models of cognitive processing in the field of translation and interpreting stem from interpreting studies. These are the Effort Model by Gile (1995) and the Cognitive Load Model by Seeber (2011). In fact, there were some attempts to adjust Gile's Effort Model to translation. However, the Cognitive Load Model by Seeber is associated solely with interpreting.

3.2.1 Effort Model

This model was developed by Gile in 1995. According to its author, there are two reasons which stand behind the development of the Effort Model, and these are:

> - Interpretation requires some sort of mental "energy" that is available only in limited supply.
> - Interpretation takes up almost all of this mental energy and sometimes requires more than is available, at which times performance deteriorates.
>
> (Gile 1995, p. 161)

[8] Source: https://www.zhaw.ch/en/linguistics/institutes-centres/iued-institute-of-translation-and-interpreting/research/clint/ visited 19 July 2022.
[9] Source: https://projekty.ncn.gov.pl/index.php?projekt_id=558652 visited 15 December 2023.

The Effort Model consists of three non-automatic components that simultaneously compete for the limited, available resources. Interpreters struggle with The Listening and Analysis or Comprehension Effort, The Production Effort, and The Memory Effort (Gile 1995). The Listening and Analysis Effort refers to the stage in which the interpreter hears the source text, starts to analyse it, and decides how to interpret it. In this case, Gile (1995, p. 162) describes "all comprehension oriented operations." The interpreter produces orally the target text during The Production Effort. As Gile (1995, p. 165) points out, "it is defined as the set of operations extending from the mental representation of the message to be delivered to speech planning and the performance of the speech plan." Finally, The Mental Effort appears when the interpreter has to keep the already heard information in their short-term memory, simultaneously listening to the speaker and waiting for upcoming information to build a cohesive target text. Based on these components, Gile (1995, p. 169) built an equation according to which:

> [the] process consist[s] of the three Efforts described above, namely the Listening and Analysis Effort L, the Short term memory Effort M, and the Speech production Effort P, plus a Coordination Effort C, which is required to coordinate the three other Efforts …
>
> SI= L+P+M+C

Gile (1995) also attempted to adapt his Effort Model to other types of interpreting and translation. As consecutive interpreting and sight translation lie beyond the scope of this study, I will discuss only the Effort Model in translation. According to Gile (1995: 185-16), translation consists of only two components: The Reading and Analysis and The Writing components. Nevertheless, the researcher claims that "processing capacity requirements are much lower in written translation than in either mode of interpretation, precisely because in translation there are no competing Efforts" (Gile 1995, p. 185). Thus, he believes that since translation does not impose any time constraints, the two components do not compete for the translator's limited cognitive resources. Contrary to this view, Hvelplund (2016: 152), in his study on the efficient allocation of cognitive resources in translation, claims that the components enumerated in Gile's Effort Model are also crucial for written translation.

Seeber (2013: 20-21) emphasises that through his Effort Model, Gile managed to present the complex issue of simultaneous interpreting and various modes of effort in a manner that is easy to understand. Nevertheless, as the researcher points out, the

simplicity may also turn out to be its drawback. It may be difficult to empirically verify such a simple model.

3.2.2 Cognitive Load Model

Another model of simultaneous interpreting, the Cognitive Load Model, was developed by Seeber in 2011. According to its author, the Cognitive Load Model "takes into account the amount of load generated by individual concurrent tasks" and is related to Wicken's (1984) Multiple Resource Model (Seeber 2011, p. 187). As opposed to the already-mentioned Gile's Effort Model, it does not assume that there is just one common cognitive resource for all activities. Seeber (2011, pp. 187-188) divides simultaneous interpreting into two tasks: language comprehension and language production, which are further divided into so-called vectors, like "perceptual auditory verbal processing of input and output 'C,' and verbal-response processing of output 'R' … interference 'I' … is calculated whenever two or more tasks overlap." Like Gile (1995), Seeber (2011) also includes the memory component. In this case, it is working memory, where information is stored while the interpreter is waiting for other components to create the output.

According to Seeber (2011, p. 189), his Cognitive Load Model has several advantages that may capture the cognitive demands of simultaneous interpreting better than the Effort Model (1995). The first one is related to the strength of vectors that present the load in a more detailed manner. Thus, the Model is able "to reflect local cognitive load as a function of both input and output features." Moreover, the division of the model into subtasks helps analyse how interpreters deal with cognitive load in various conditions, especially considering syntactic asymmetry. Finally, Seeber (2011: 189) suggests that his model enables the quantitative analysis of cognitive load.

3.2.3 Cognitive Load Theory

Hunziker Heeb et al. (2021: 52-53) attempted to adjust the assumptions of the Cognitive Load Theory to translation process research. Nevertheless, the Cognitive Load Theory developed by Sweller et al. (1998) refers to cognitive load in the teaching and learning environment. It is not adopted in translation process research. According to Sweller et al. (1998), there are three types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load "refers to the complexity of the information being processed and was related to the concept of element interactivity" (Sweller et al. 2019, pp. 263-264). Extraneous cognitive load is related to "how the information is presented

and what the learner is required to do by the instructional procedure" (Sweller et al., 2019, p. 264). This type of load may inhibit the learning process. Thus, the aim of a teacher is to design a task in such a way that its instructions do not evoke any additional, unnecessary cognitive load (Sweller et al. 1998: 262). Discussing the third type of cognitive load, Sweller et al. (2019, p. 264) point out that "[g]ermane cognitive load was defined as the cognitive load required to learn, which refers to the working memory resources that are devoted to dealing with an intrinsic cognitive load rather than extraneous cognitive load." It is worth mentioning that the three components are additive, meaning that if one of them is too high, it takes the resources available for the remaining two (Sweller et al. 1998: 264).

In the translation context, Hunziker Heeb et al. (2021, p. 53) believe that the first type of load "is inherent in translating as such, but it amount depends on the level of expertise and background knowledge of each translator." Extraneous cognitive load may be associated with various problems resulting from the translation assignment like "a lack of documentary resources, problems with tools that cause cognitive friction in human-computer interaction … or poor physical or organisational conditions" (Hunziker Heeb et al. 2021, pp. 52-53). Germane cognitive load is related to parts of the translation process that does not cause major problems, for example, because of an understandable source text.

3.3 Indicators and measures of cognitive effort

Various indicators like fixations, pauses, task length, etc., are used to measure cognitive effort in the translation process (Ferreira et al. 2016; Shreve & Lacruz 2017; Ehrensberger-Dow et al. 2020; Gile & Lei 2020). Hunziker Heeb (2020, p. 49) even points out that "[t]he performed cognitive processes manifest themselves in process activities and, conversely, the analysis of these activities allow us to make inferences about the cognitive effort that the translator has expended." Nevertheless, researchers emphasise that such indicators are just indirect measurements; in fact, it is impossible to measure cognitive effort directly (Chen 2017; Gile & Lei 2020). According to Gile and Lei (2020: 269), quantitative methods of cognitive effort measurement allow solely for detecting differences in cognitive effort, whether it increases or decreases. However, they are not able to pinpoint the reason behind such observations.

There are various methods of measuring cognitive effort. For example, Hunziker Heeb (2020: 52-56) discusses a classification based on the units of analysis. She proposes

five units of analysis. The first unit uses micro-units. It embraces the analysis of single words. According to Hunziker Heeb (2020: 52), micro-units analysis originated in the field of psycholinguistics and, currently, is used mainly in studies focusing on bilinguals. For example, it was used in the studies leading to the development of the Revised Hierarchical Model, discussed in section 1.4. The second unit of analysis is devoted to phenomena appearing in the source text and target text. These are usually various problem triggers. Hunziker Heeb (2020: 52-53) enumerates examples like metaphors (analysed, for example, in Fröster Hegrenæs 2018) or rich points (analysed, for example, in Dragsted 2012). Collocations analysed in this PhD project also fall into this category. The third unit is devoted to the analysis of complete sentences. According to Hunziker Heeb (2020: 53-54), it is most frequently used in studies focusing on the usage of CAT tools. The analysis of the translation process phases, developed by Jakobsen (2003), is classified as the fourth unit of analysis. Finally, the fifth unit embraces the macro-level analysis, the analysis of the whole text. According to Hunziker Heeb (2020: 55), in this case, researchers frequently devote their attention to pauses.

Chen (2017, p. 648) observes that "[b]ecause cognitive load is a multi-dimensional construct, a single measure cannot provide a comprehensive picture." She identifies four measurable aspects of cognitive load in interpreting that translate into four cognitive load measures (Chen 2017: 647-648). In the case of subjective measures, researchers use psychometric rating scales to measure the feeling of cognitive effort. Performance measures embrace the quality of interpreting. The researchers can use either primary or secondary task measures. Physiological measures are related to the activation of various body parts; this includes, for example, pupillometry. According to Chen (2017, p. 648), the last category, analytical measures, is "usually provided by experts or derived from models or task analysis." It is based on the characteristics of the interpreting task, for example, its complexity.

However, I would like to draw particular attention to the division discussed by Ehrensberger-Dow et al. (2020). Referring to the work by Chen et al. (2012), they claim that "cognitive effort can be assessed through (i) subjective measures, (ii) physiological measures, (iii) behavioural measures, (iv) performance measures" (Ehrensberger-Dow et al. 2020, p. 223). As can be observed, there is a visible similarity between the categories described by Chen et al. (2012) and Chen (2017). Interestingly, while describing these concepts, both research groups originally referred to the phenomenon of cognitive load

rather than cognitive effort. Nevertheless, I am going to follow the description of these methods provided by Ehrensberger-Dow et al. (2020).

3.3.1 Subjective measures

According to Ehrensberger-Dow et al. (2020, p. 223), "[s]ubjective measures provide insight into an individual's perception and experience of cognitive effort during a language processing task." Various types of questionnaires, as well as think-aloud protocols and retrospective reports, belong to this category. Retrospective reports were used to analyse cognitive effort, for example, in the study by Gumul (2019a). The aim of the study was to verify whether the cognitive effort reported by participants in their retrospective reports can also be found in the interpreting product. The researcher focused on the following indicators of cognitive effort that can be detected in retrospective reports: "failure sequences, evidence of competing efforts, evidence of working close to cognitive overload, and negative effect of problem triggers" (Gumul 2019a, p. 26) The following indicators of cognitive effort were searched for in the interpreting product: "anomalous pauses, omissions in the target text, repairs, grammatical errors, mispronunciations, and disfluencies in the form of hesitation markers and false starts" (Gumul 2019a, p. 26). Although in most cases, cognitive effort reported in the retrospective reports has an identifiable reflection also in the product of interpreting, there is a considerable number of retrospective comments for which no corresponding indicator of cognitive effort was found. These results reveal that the method allows for tracing the processes that occur in the translator's black box but are not visible in the product. Another advantage of subjective measures is mentioned by Ehrensberger-Dow et al. (2020: 224), who draw attention to the simplicity of their use. Subjective measures do not require the use of any special equipment. Nevertheless, one of the main drawbacks of subjective measures cannot be ignored. As in any type of verbal reports and questionnaires, participants are able to verbalise only conscious efforts. Moreover, they can easily manipulate the data, for example, due to the white-coat effect.

Probably the most popular questionnaire, designed particularly to measure cognitive workload, is the NASA Task Load Index [NASA-TLX] (Hart & Staveland 1988), which is also used in this study. It was employed, for example, in the study by Hunziker-Heeb and Gieshof (2022, 2023). The researchers investigated whether translators perceive any differences between cognitive load and cognitive effort. The study was based on two out of six ratings from the NASA Task Load Index, Mental

Demand and Effort. Forty-eight participants, both translators and interpreters, whose L1 is German and L2 is English, took part in the study. Participants first performed the translating or interpreting tasks and were subsequently asked to fill out the NASA-TLX questionnaire. The results revealed an interesting trend. While the cognitive effort was rated slightly higher than the cognitive load among the group of interpreters, an opposite pattern was found among the group of translators. Nevertheless, cognitive effort and cognitive load were generally assessed as being quite similar. Sometimes, participants pointed out a rather comfortable level of both measurements.

3.3.2 Physiological measures

Physiological measures are described by Ehrensberger-Dow et al. (2020, p. 224) as "complementary to subjective measures of cognitive effort in that they can provide a more objective indication of the amount of cognitive capacity devoted to a particular task." These measures are strictly related to the human nervous system. The first type of variable that belongs to the category of physiological measures is pupil dilation. Pupil dilation was analysed in the context of cognitive effort, for example, in the study by O'Brien (2008). Eye-tracking was also triangulated with a subjective measure in the form of a questionnaire study. The researcher devoted her study to CAT tools and focused on the influence of fuzzy matches on the level of cognitive effort. Fuzzy matches are defined by O'Brien (2008, p. 79) as "source-text segments contained in a translation memory that are similar to a source-text segment requiring translation in a new text." Five translation trainees working in a German-English language pair participated in this study. Nevertheless, the pupil dilation results did not unequivocally point out how fuzzy matches influence cognitive effort during translating. It appeared that when there is a high level of match between the source text segment that is supposed to be translated and the segment saved in the translation memory, the level of median pupil dilation is relatively high. Then, pupil dilation increases until the match rates fall to 70%. From that point, not only the match level but also the medial pupil dilation decreases. (O'Brien 2008: 86-87).

Heart rate is another variable classified as a physiological measure. Ehrensberger-Dow et al. (2020: 225) point out that heart rate is frequently analysed as a reaction to stress during interpreting tasks, which may be directly linked to an increased cognitive effort. According to Rojo López et al. (2021: 593), there are also additional measures of stress in translation and interpreting studies, like galvanic skin response or cortisol level. However, a detailed analysis of these lies beyond the scope of this study. The work by

Rojo López et al. (2021) may serve as an example of a study using heart rate as one of the variables. They decided to conduct an analysis of heart rate as an indicator of stress among 23 interpreting trainees during their final exam. Students' L1 was Spanish, and L2 was English, and they worked in both directions. Besides mean heart rate, the researchers also focused on speech rhythm. The results revealed that students' mean heart rate increased as they proceeded to interpret the utterings, while it was lower during the listening phase. Interestingly, the data from the mean heart rate did not directly translate into the results of this exam (Rojo López et al. 2021: 602).

The main advantage of physiological measures is associated with their high objectivity as well as the individuality of measurement. Nevertheless, both pupil dilation (as discussed in Section 2.5.4) and heart rate are not devoid of obstacles, as they may be easily influenced by other factors that may result in obtaining unreliable data (Ehrensberger-Dow et al. 2020: 225). Moreover, Rojo López et al. (2021: 593) suggest that translation and interpreting researchers are unsure whether such measures could be directly related to the analysed behaviour, like stress. Thus, as in any case, the triangulation of methods is highly recommended.

3.3.3 Behavioural measures

Ehrensberger-Dow et al. (2020, p. 225) claim that "behavioural measures have been used to derive the effort associated with multimodal communication, text comprehension, and text production." Process methods like eye-tracking and keylogging, as well as studies analysing gestures and facial expressions, definitely belong to this category.

For example, both eye-tracking and the analysis of gestures can be found in the study by Stachowiak-Szymczak (2019). The subject of the study was beat gestures, defined as "gestures aimed at stressing parts of speech or the content of what is being said" (Stachowiak-Szymczak 2019, p. 76). In the case of gaze patterns, Stachowiak-Szymczak (2019: 86) focused on mean fixation duration. The primary objective was to verify whether changes in cognitive load are reflected in eye movements and gestures. Both professional interpreters and interpreting trainees participated in her study. The participant's task was to interpret six speeches from Polish (L1) into English (L2). The results revealed some characteristics of the analysed variables. For example, increased mean fixation duration was registered when participants interpreted numbers. Moreover, interpreters tend to gesticulate more while interpreting lists of items. However, as stated

by Stachowiak-Szymczak (2019: 131), these data allow for further examination of the applied measures and their relation to cognitive issues.

Sjørup (2013) employed eye-tracking and keylogging to analyse cognitive effort while translating metaphors. The main aim was to test whether translating metaphors requires more cognitive effort than translating literal expressions. She focused on two stages: comprehension and production, which could be compared to Jakobsen's (2003) orientation and drafting phases (Sjørup 2013: 13). The study was conducted on a group of 17 professional translators with Danish as L1 and English as L2. In the case of eye-tracking, Sjørup (2013) analysed variables like fixation time (fixation duration), number of fixations, and first pass duration. Following Hyönä et al. (2003), Sjørup (2013, p. 84) defines first pass duration as "the summed duration of all fixations on a target region before exiting it." For keylogging data, particular attention was drawn to production time. However, the researcher was not able to confirm whether it is always more effortful to translate metaphors than literal expressions. For example, various translation strategies applied to translate metaphors require various levels of cognitive effort. Sjørup (2013, p. 205) concluded that "metaphors are perhaps more unexpected, but not necessarily more cognitively effortful to comprehend than literal expressions." It was also found that the actual production of metaphor in the target language may require more cognitive effort (Sjørup 2013: 208).

3.3.4 Performance measures

The analysis of performance measures is related to the observation made by Gile (1995) in his Effort Model, assuming that the performance of translators and interpreters may deteriorate as a response to the increased cognitive effort. As pointed out by Ehrensberger-Dow et al. (2020, p. 227), "[p]erformance measures are … a subgroup of behavioural measures, as they consider behaviour with regard to particular performance criteria." This category includes various indicators like accuracy, fluency, complexity, and disfluencies, such as pauses during target text production. Ehrensberger-Dow et al. (2020, p. 227) also mention the analysis of secondary tasks, defined as "responding quickly to a visual or auditory signal while performing a primary task such as interpreting or translation."

For example, the phenomenon of pausing behaviour in the case of translation is discussed in the study by Kruger (2016). The researcher explicitly says that "pauses are seen as indexical of cognitive effort" and devoted her study to the analysis of cognitive

processes occurring during pauses (Kruger 2016, p. 27). Translation trainees, with Afrikaans as their L1 and English as their L2, participated in this study using eye-tracking and keylogging. According to Kruger (2016, p. 48), "[t]he findings of this study point out to a complex relationship between pause duration, the syntactic position of the pause, reading behaviour and cognitive effort." For example, it appeared that during pauses, translators usually focus their attention either on the source text or both the source text and the target text (Kruger 2016: 34). This means that when a pause occurs, translators tend to read further parts of the source text to proceed with the translation process, or they often perform some online revisions and introduce minor changes; thus they have to read both the source text and target text. That is why the longest pauses were found when participants switched their attention from the source text to the target text reading. The results revealed no significant differences between the location of the pauses and their duration. The only statistically significant result was found at the word boundary level (Kruger 2016: 34-47).

In the case of interpreting studies, both fluency and pauses are discussed, for example, by Chmiel et al. (2022). They based their study on speeches delivered in the European Parliament and their interpretations. The interpretations were performed both into English and Polish. This time, both silent and filled pauses that influence interpreting fluency were analysed. The speed of the source text delivery appeared to influence the number of both types of pauses. More silent pauses appeared during the interpretation of slower source texts. The number of filled pauses is significantly higher in the case of more compressed source texts and shorter interpretations, which may indicate some cognitive effort.

3.4 Cognitive effort and directionality

Research in the translation process has produced a sizeable body of findings on directionality. Unfortunately, researchers did not manage to obtain conclusive results. Frequently, they either did not reach the level of statistical significance or only one of the analysed variables was statistically significant. There are also studies in which the results suggest that translating into L1 may require more cognitive effort.

The work by Buchweitz and Alves (2006) is one of the few studies confirming that translating into L2 is more effortful than translating into L1. It aimed to analyse how the language of the previously translated text influences the second translation task. The study was conducted on the English-Portuguese language pair, with Portuguese as L1.

The participants were translation graduates with some translation experience and translation trainees. They were supposed to translate two texts, one in each direction. The researchers adopted total task time, the number of segments, and the number of keystrokes as indicators of cognitive effort. The results from the total task time confirmed the hypothesis, which assumed that L1-L2 translation requires more effort. Interestingly, not only the whole task but also the stage of writing, or following Jakobsen's (2003) terminology, the drafting phase took longer while translating into L2. L2-L1 orientation took 8% of the total task time, writing 49%, and revision 43%. L1-L2 writing took 70% of the total task time, revision only 22%. In contrast, the amount of time devoted to the orientation phase was identical as in the case of L2-L1 translation. According to the researchers, the reason for the shorter revision time in L1-L2 translation may lie in "an exhaustion of the translator at the end of the process" (Buchweitz & Alves 2006, p. 254). Two remaining variables, the number of segments and the number of keystrokes, also indicated that L1-L2 is a more effortful direction. Thus, all variables in this study showed that translating into L2 requires more effort.

The same language pair was analysed in the study by Fonseca (2015), who focused on translators' writing profiles. Professional translators, whose task was to translate four texts (two in L2-L1 and two in L1-L2 translation directions), participated in the study. Also in this case the participants' L1 was Portuguese. Nevertheless, this study did not show major differences in translators' work styles in the context of directionality. It appeared that they adopted the Drafter/Reviser profile and recursive sub-profile in both directions. It means that "professional translators tend to improve their text by monitoring their translation process during both the drafting and revision phases, regardless of the direction in which they are working" (Fonseca 2015, p. 123). Thus, the researcher concludes that the results did not indicate that L1-L2 is a more effortful translation direction.

Several studies partially confirmed that L1-L2 translation requires more effort. Usually, either only one hypothesis about directionality was confirmed, or just one variable indicated L1-L2 as a more cognitively demanding translation direction. For example, da Silva et al. (2017) analysed directionality in terms of translating and post-editing. One out of four research questions referred to the influence of translation direction on cognitive effort. They wondered whether "[i]t is cognitively more effortful to translate and post-edit into the L2 than into the L1" (da Silva et at. 2017, p. 108). As the remaining questions referred to the relationship between translating and post-editing,

they lie beyond the scope of this study. The study by da Silva et al. (2017) was conducted on the Chinese-Portuguese language pair, with Chinese as L1. The task of 18 professional translators was to translate and post-edit texts in two directions. Since the researchers used eye-tracking and keylogging, they adopted "the number of fixations on ST AOI [area of interest], [t]otal reading time (sum of all fixations) on ST AOI, [n]umber of fixations on TT AOI, [and] [t]otal reading time (sum of all fixations) on TT AOI" as indicators of cognitive effort (da Silva et al. 2017, p. 119). One out of four enumerated variables, the total reading time of the source text, did not reach statistical significance. The results indicated that translating into L2 requires more effort solely in the writing (drafting) phase. As pointed out by da Silva et al. (2017, p. 121), "[t]he participants tended to gaze longer and fixate more times on the TT while performing the inverse tasks [L1-L2 translation]; conversely, they tended to fixate almost twice as much on the ST while performing the direct tasks [L2-L1 translation] as compared to the inverse tasks." Interestingly, the results indicated that translators always need more time to process the text in their L2, whether it is the target text, as in the case of L1-L2 translation, or the source, while translating into their L1.

The hypothesis that L1-L2 translation requires more effort was also not unequivocally confirmed in the research by Pavlovič and Jensen (2009). This was one of the first studies triangulating eye-tracking and keylogging to investigate directionality. The researchers analysed four indicators of cognitive effort: gaze time, average fixation duration, total task length, and pupil dilation. They aimed to verify whether target text processing always requires more effort, regardless of the translation direction (hypothesis 1). They also wondered whether L1-L2 is a more effortful direction (hypothesis 2). The two subsequent hypotheses referred to the phases of the translation process. Processing of the source text (orientation) was supposed to evoke more effort in L2-L1 translation, and drafting could require more effort in L1-L2 translation. The study focused on both professional translators and translation trainees. Thus, the last hypothesis was related to the differences between professionals and trainees. However, this issue lies beyond the scope of this study.

The results revealed that only the first hypothesis was fully confirmed by all the adopted variables. Hypothesis 2 was partially confirmed. Half of the analysed variables (task length and pupil dilation) showed that L1-L2 requires more effort. The difference between L2-L1 and L1-L2 translation was statistically significant only for the measure of pupil dilation (Pavlovič & Jensen 2009: 101). The third hypothesis revealed quite

interesting results. Gaze time was the only variable confirming that orientation requires more effort in L2-L1 translation. However, the difference between the two directions did not reach the level of statistical significance. The results from the average fixation duration confirmed this hypothesis only in the case of students, while only professional translators had increased pupil dilation during the L1-L2 translation (Pavlovič & Jensen 2009: 102-103). Likewise, an unexpected trend appeared in the case of the fourth hypothesis. The assumption that drafting requires more effort in L1-L2 translation was confirmed solely by pupil dilation. Both average fixation duration and total gaze time showed the opposite pattern, indicating that the L2-L1 drafting phase is more effortful (Pavlovič & Jensen 2009: 104). Thus, none of the hypotheses related to differences in effort evoked by L2-L1 and L1-L2 translation was fully confirmed.

Another researcher who analysed various aspects of directionality is Ferreira. In her study from 2014, Ferreira focused on the issue of recursive movements in the context of directionality. Following Buchweitz and Alves (2006), Ferreira (2014, p. 111) defines recursiveness as "the online revisions of the target text observed in keylogging data recorded in real-time as the translation process unfolds." Thus, the indicators of cognitive effort that she adopted in her study were recursive movements, recorded by the keylogging programme, and the number of verbalisations obtained during the retrospective session. Eight professional translators, with Portuguese as L1 and English as L2, who participated in her study were supposed to translate texts ascribed to two groups. The texts were either related by their topic or completely unrelated. Two out of four hypotheses referred directly to cognitive effort. First of all, Ferreira (2014: 115) aimed to verify the number of recursive movements in L1-L2 translation when translators are translating thematically unrelated texts. She was also interested in the number of comments appearing in retrospective reports describing the L1-L2 translation of unrelated texts.

The results turned out to be quite surprising. It appeared that the number and the character of recursive movements vary depending on the relatedness of the topic of translated texts. Translators produce more recursive movements during L1-L2 translation of related texts, but only when they make changes in vocabulary and syntax. They more frequently return to the text during L2-L1 translation to implement modifications to typing mistakes while translating unrelated texts (Ferreira 2014: 119). Therefore, Ferreira (2014, p. 125) draws the following conclusions: "in the context of related texts, L1[L2-L1] translation was more demanding than L2 [L1-L2] translation. However, in the context

of texts on different topics, results show that L2 translation required more effort than L1." Moreover, the researcher observed some differences in the issues described in the L1-L2 and L2-L1 retrospective comments. For example, translators were more critical in their comments while discussing their L2-L1 translation process. They produced, in general, more comments during retrospective sessions after translating unrelated texts from L1 to L2. It confirmed hypothesis four (Ferreira 2014: 120; 123). Nevertheless, it can be concluded that these results did not provide an unambiguous answer as to which direction of translation requires more cognitive effort.

In 2016, Ferreira analysed the issue of cognitive effort and directionality together with her colleagues. This time, they focused on data obtained through an eye tracker. They adopted total task length, fixation count, average fixation duration, and gaze time as indicators of cognitive effort. A relatively small group of four professional Spanish-English translators, with either Spanish or English as their L1, translated two texts, one in each direction. The results indicate that all participants need more time translating into their L2. Total task time and average fixation duration are the variables that unequivocally confirmed that L1-L2 translation requires more cognitive effort. In the case of fixation count, the hypothesis was true for 3 out of 4 participants (Ferreira et al. 2016: 71-72). Ferreira et al. (2016: 73-74) also discussed cognitive effort in terms of areas of interest. The results partially demonstrated that source text processing requires more effort while translating into L1. Only one participant showed the reverse pattern. However, the reverse hypothesis that target text processing requires more effort during L1-L2 translation was rejected. The results did not prove that L1-L2 translation requires more attention devoted to searching for information in Internet resources. Although this time, the group of participants was considerably smaller, and it was not possible to conduct inferential statistics, the results still do not show which direction of translation requires more cognitive effort.

A lack of statistically significant differences between the two directions of translation was reported in the study by Ferreira et al. (2021). The researchers implemented eye-tracking, keylogging, and retrospective reports. They analysed mouse and keypress events as well as fixations, saccades, gaze duration, and gaze index as indicators of cognitive effort. The study focused not only on cognitive effort and directionality but also on the issue of language dominance. As in the case of all previously discussed studies conducted by Ferreira, professional translators (32 participants), having either English or Spanish as their L1, participated in this study (Ferreira et al. 2021: 127).

They translated one text in each translation direction. Worth noting is the fact that this time, the number of participants allowed the researchers to verify the results through inferential statistics. Unfortunately, in the case of most of the variables, statistical tests did not indicate either of the directions as requiring higher cognitive effort.

The phenomenon of directionality and cognitive effort was analysed in detail in the EDiT grant project conducted by Whyatt. She analysed the influence of directionality on cognitive effort using eye-tracking and keylogging. In the study from 2018, Whyatt presented the preliminary results of her project based on ten professional translators. The analysed language pair is Polish-English, with Polish as the participants' L1. The task of the participants was to translate two texts of different genres in both directions. Whyatt (2018, p. 101) adopted the following indicators of cognitive effort: "(1) average fixation duration on the source text during the first reading, prior to translation (2) typing speed during target text production (3) number of pauses longer than 5 seconds during target text production and (4) the percentage of the entire task time devoted to the final revision of the target text." The results were divided into three translation phases: orientation, drafting, and revision (Jakobsen 2003).

While discussing cognitive effort during the orientation phase, Whyatt (2018) explains that participants tended to fixate longer when translating into L1. This is in line with the results obtained by da Silva et al. (2017), which suggested that translators need more time to process the source text in their L2. The results also revealed that participants spent a similar percentage of time reading the source text in both directions (Whyatt 2018: 107). In the case of drafting, participants tend to type faster when translating into their L1. However, the difference in their pausing behaviour was marginal. There were, on average, 34.1 pauses during L2-L1 translation and 32.1 pauses during L1-L2 translation. The analysis of the revision phase revealed that this process lasted longer in L1-L2 translation. However, the difference in the duration of L1-L2 and L2-L1 revisions and the percentage of time devoted to this stage in both directions was very small (Whyatt 2018: 104). These results appeared to be contrary to the results obtained by Buchweitz and Alves (2006). Whyatt (2018, p. 107) summarises her study's results, concluding that "for the 10 translators whose data have been analysed, the direction made hardly any difference in the selected measures which operationalise their cognitive effort. In other words, L2 translation does not seem more effortful than L1 translation."

The EDiT project was continued, and the subsequent results appeared in the study from 2019. The design of the study was identical to the one from 2018. However, this

time, 30 Polish-English translators participated in the study, which resulted in 26 complete data sets (Whyatt 2019: 85). This time, the researcher focused on the total task time and quality of the target text assessed by proofreaders. Since the indicator of translators' cognitive effort referring to directionality was total task time, I would like to focus solely on this variable. The total task time in this study did not reach the level of statistical significance. It means that L1-L2 translation did not appear to be a more effortful direction of translation (Whyatt 2018: 88). Nevertheless, the study by Tomczak and Whyatt (2022), based on the same data set and experimental procedure, revealed interesting results. The objective of the study was to analyse the influence of directionality on the process of lexical selection, defined as "a process of mapping meaning, or concepts into words" (Tomczak & Whyatt 2022, p. 120). The authors of the study analysed variables like verbal fluency, online resource use, deletion and changes introduced into words, pauses, and penalty points given by proofreaders. The results from the verbal fluency task confirmed the translation asymmetry discussed in section 1.4.1, as translators gained more points for enumerating words in their L1. Moreover, they used online resources more frequently while translating into L2. They also made more changes to the target text during the L1-L2 translation and received more penalty points in this translation direction. However, a closer look at corrections made by proofreaders shows that these were usually minor mistakes.

Hunziker Heeb (2020) devoted her doctoral thesis to differences between L2-L1 and L1-L2 translation, with a particular emphasis placed on cognitive effort. Hunziker Heeb (2020: 68) searched for differences between both directions of translation in the context of the translation process, product, and cognitive effort. As this chapter is devoted to the phenomenon of cognitive effort, I will only refer to the latter aspect. Three groups of professional translators participated in this study: bidirectional translators and unidirectional translators translating into their L1s, either English or German. Hunziker Heeb (2020: 99-101) divided indicators of cognitive effort into four categories: character production effort analysed based on the character count, revision effort based on the number of performed revisions, information search effort in which she counted the number of searchers performed by the translators, coordination effort based on the number of pauses and some eye-tracking data, and mean fixation duration.

Discussing the results, Hunziker Heeb (2020: 156-166) compared three groups, bidirectional translators translating in two directions and each of the unidirectional group with bidirectional translators translating either into their L1 or L2. It appeared that for the

bidirectional group, only one of the variables, character production, reached a statistically significant difference. It was higher while translating into L2. Just one of the variables reached a statistically significant difference in the comparative analysis of bidirectional translators working into their L2 and unidirectional translators whose L1 is English. It appeared that the information search effort was higher when translating into L2. Interestingly, none of the variables managed to reach a statistically significant difference among the third group, where bidirectional translators working into their L1 were compared with unidirectional translators working into their L1 (German). Thus according to Hunziker Heeb (2020: 180), hardly any differences can be found in terms of cognitive effort expended in the translation process by bidirectional and unidirectional translators.

The presented overview of studies analysing the influence of directionality on cognitive effort draws on various methods. Researchers usually employ retrospective reports, keylogging, and eye-tracking. Nevertheless, Rodríguez-Inés (2022) emphasises that studies on directionality do not provide repetitive results, which can be observed based on the examples of studies described hereinabove. However, it is also worth mentioning that studies on directionality focus on various language pairs. Sometimes, both languages have a relatively high diffusion (e.g., Ferreira et al. 2021), whereas, in the case of studies conducted by Whyatt (2018, 2019), Polish is a language of low diffusion, while English is a contemporary lingua franca. It is worth noting that the characteristics of participants vary among studies. On the one hand, translation trainees and graduates participated in the study by Buchweitz and Alves (2006), and on the other hand, Ferreira et al. (2016, 2021) focused solely on the group of professional translators.

3.5 Cognitive effort and default translation

Previous sections have shown that translation process researchers frequently focus on increased cognitive effort resulting from various difficulties of the translation process. Although translation is an effortful activity, not all phenomena appearing in the translation process require the same level of cognitive effort. While some choices seem more effortful and, frequently, more time-consuming, others proceed relatively smoothly. As suggested by Hegernæs and Halverson (2022), it is important to analyse not only those phenomena that increase cognitive effort but also the ones in the case of which cognitive effort is at a minimum level, with particular emphasis put on the reasons leading to such a low level. In this case, researchers point out two phenomena: literal translation and

default translation, both of which are associated with an easiness present in the translation process.

Although the two concepts function separately (e.g., in studies by Halverson (2015) or Hegernæs and Halverson (2022), sometimes, both terms are used interchangeably. For example, Schaeffer and Carl (2014, p. 29) present the following definition in their study:

> the *formal correspondence hypothesis*, the *literal translation default rendering procedure*, the *law of interference*, and the *monitor model* are all related concepts which seem to assume that one-to-one literal translation correspondences are easier to produce than translations that formally deviate from the source text.

Schaeffer et al. (2016, p. 189) define literal translation as "the first or default solution a translator applies to the source text." A similar definition, including both terms, can be found in work by Trikkonen-Condit (2005, p. 408), who claims that "literal translation is a default rendering procedure." Following works by Halverson (2015, 2019), I would like to focus on each term separately, firstly discussing the phenomenon of literal translation and then turning to default translation with particular attention to the analysis of N-grams.

3.5.1 Literal translation

As some scholars (such as Halverson 2015) point out, literal translation is a well-established phenomenon in translation studies. It was discussed as early as 1958 in work by Vinay and Darbelnet, who classified it as one of the translation techniques. Later, it can also be found in work by Newmark (1988, p. 70), who states, "I believe literal translation to be basic translation procedure, both in communicative and semantic translation, in that translation starts from." Nevertheless, it is treated not only as one of the techniques or procedures but also as a target text feature that often struggles with some criticism (Halverson 2015, 2017, 2019). While describing the contemporary approach to literal translation, scholars usually refer to the study by Scheaffer and Carl (2014, pp. 29-30), who distinguished three features of literal translation that can be found during comparative analysis of the source text and target text. These are "a) Word order is identical in the source and target languages. b) Source and target text items correspond one-to-one. c) Each source word has only one possible translated form in a given context." It is believed that the already-mentioned correspondence and similarity between the source text and target text is a crucial feature leading to decreased cognitive effort, making the translation process faster and devoid of difficulties (Schaeffer & Carl 2014; Halverson

2015). Translators do not have to devote much cognitive effort to form the corresponding target text since the source text words have only one possible equivalent. Therefore, any additional translation alternatives appearing in the translation process prevent literal translation and increase cognitive effort.

This view is supported by the results from Schaeffer's and Carl's (2014) study. The first objective of their study was to develop a repetitive metric that will enable literality of translation to be measured. The second one was to measure the increasing cognitive effort that appears when literal translation is not possible. The scholars utilised in their study both keylogging and eye-tracking. The results indicated that:

> - More translation choices lead to longer reading and processing time
> - Longer relative source-target distortions increase gaze activity
> - Regressions are more effortful than progressions
> - Translators and post-editors map not only the source text against the target but also the target against the source text.
>
> (Schaeffer & Carl 2014, p. 36)

Nevertheless, it is believed that it is impossible to obtain a perfect literal translation as the majority of words have more than one acceptable translation. Translators should also bear in mind aspects like cultural elements and semantics, grammar and word order of the target language, which frequently do not allow for a literal translation of a given phase (Schaeffer et al. 2016; Schaeffer & Carl 2017).

3.5.2 Default translation

As opposed to literal translation, default translation is a relatively new concept in translation and interpreting studies. According to Hegrenaes and Halverson (2022), it was first presented in front of a conference audience in 2019. The first works devoted to default translation were published in 2015 and 2019 by Halverson, who leads the research on this topic (Hegrenaes & Halverson 2022). Since the concept of default translation does not have a long history within translation and interpreting studies, some of its aspects are still under-researched.

Default translation is defined as "a particular phase of translation production. In this phase, translators demonstrate stretches of uninterrupted production. The text produced is assumed to draw on easily accessible, routinised knowledge" (Halverson 2019, p. 190). Due to the perceived easiness of translation during this phase, it is also directly linked to a decreased cognitive effort. As the author of the definition pointed out,

default translation can be described as both a phase and a mode of translation (Hegrenaes & Halverson, 2022). Muñoz (2021) classifies default translation as the fourth wave of translation and interpreting studies. Its assumptions are in line with the findings of the school of cognitive translatology and 4EA cognition (embodied, embedded, enacted, extended, and affective) (Muñoz 2017). Therefore, as Halverson (2019) points out, it deviates from the previous perception of translating as a problem-solving procedure. Due to its characterisation as uninterrupted production, default translation can also be compared to the unchallenged translation concept developed by Carl and Dragsted in 2012 (Halverson 2019; Hegrenaes & Halverson 2022). Unchallenged translation occurs when "translation production proceeds smoothly in a parallel or in a sequential mode with only some words look-ahead" (Carl & Dragsted 2012, p. 138). Similarities between default and unchallenged translation can easily be seen in the quoted definition, as both concepts point out decreased cognitive effort. Unlike unchallenged translation, the challenged one "is characterised by delayed text production and associated with extended reading activities into the ST or TT context" (Carl & Dragsted 2012, p. 138). In this case, the cognitive effort may be substantially increased.

The types of knowledge that are crucial for default translation to occur are also worth mentioning. Halverson (2019, p. 190) enumerates three elements: "bilingual linguistic knowledge, metalinguistic knowledge (including knowledge of communication norms), and knowledge of the specific task." It is also believed that the ability to perform default translation develops due to routine and translation practice (Hegrenaes & Halverson 2022).

Some differences between literal and default translation can be observed at this point. While the first concept refers to either the procedure or quality of a text, default translation denotes the translation phase. Moreover, literal translation is not perceived as a translation solution acquired with practice. The main difference between literal and default translations lies in the structural similarity between the source text and the target text. It is a fundamental condition in the case of literal translation. However, it is not necessary for default translation to occur. Moreover, defaults can be found and analysed only in the process data, both in writing and reading (Halverson 2019). Therefore, methods like eye-tracking and keylogging are recommended to analyse it.

3.5.2.1 N-grams

A study investigating whether defaults are structurally similar in source and target language was conducted by Hegernæs and Halverson (2022). They also attempted to find out what structures may be translated as defaults. As the researchers point out, there are no limits on the type of structures that may appear in default translation. Nevertheless, they draw attention to the fact that constructs known as N-grams should include at least two words. According to Hegernæs and Halverson (2022), the N-gram refers to a structure made of words that are highly likely to appear together in translation as an outcome of an uninterrupted translation string. These sequences are known as defaults. As the researchers continue, there are various numbers of words that appear together. In their study, the smallest number is two, which forms bi-grams, and the highest number is six, forming six-grams. The data were obtained from eight translations by four students whose L1 was Norwegian and L2 was English. The translation process was recorded by a keylogging programme. The researchers utilised the Inputlog software. Later, the data were analysed using Sketchengine, which helped to select those segments that may appear to be a default, according to criteria adopted by the researchers. These include, for example, structurally complete phrases. The phrase was supposed to appear no less frequently than three times in all analysed translations. Only then could it be further analysed as an N-gram.

The results of the study by Hegernæs and Halverson (2022) did not confirm the findings from the study by Schaeffer and Carl (2013, 2014, 2017). It provides evidence that literal translation and default translation are distinct phenomena. First of all, the results from Hegernæs and Halverson (2022) reveal an interesting trend regarding structural similarity. It appears that there is usually a lack of structural similarity between defaults and the source text. Interestingly, such a similarity can often be found between segments that are not classified as defaults. Therefore, default translation may also appear when the source and target languages are not similar (Halverson 2019). Secondly, defaults in the form of N-grams of various lengths were found in more than 20% of the analysed segments. This illustrates the large scale of this phenomenon. Hegernæs and Halverson (2022) state that, on average, almost 100 defaults can be found in translations by one participant of their study. Therefore, as Halverson (2019) discussed, default translation may raise new possibilities within cognitive translatology. It may provide useful information about activities assumed as effortless or when no cognitive effort is reported in processual data.

Chapter 4. Methodology

The aim of this chapter is to present the PhD project study design. In the first sections, I discuss the aim of the study and the reasons that led me to conduct it. Since I denote this study as an experiment, I draw a distinction between the notions of experiment and quasi-experiment in the next section. Then, I move on to research questions and hypotheses. I also discuss dependent and independent variables, with particular attention to the state of research and possible research gaps. The last sections are devoted to the details of the study designs of the pilot and the main studies.

4.1 The aim of the study

The aim of this study is two-fold. First and foremost, it aims to analyse the influence of directionality on cognitive effort among translation trainees. The second aim is to add new variables, namely eye-key span (EKS) and the weighted rating of the NASA-TLX, to studies on directionality and cognitive effort. I would also like to verify their application in this scope.

A mixed-methods approach was adopted to analyse the influence of directionality on cognitive effort. The following methods were used in the study: eye-tracking, keylogging, retrospective verbal reports, NASA Task Load Index and a self-designed questionnaire. The application of both objective and subjective methods enabled me to obtain both qualitative and quantitative data. Since the experiment was conducted using five different methods, it was possible to gain more detailed insight into the translator's black box during the translation process.

The independent variable analysed in this study is directionality, defined as translating into L1 or L2 (Whyatt 2019: 79). I am going to analyse the influence of directionality on the following dependent variables denoting cognitive effort: average fixation duration, total gaze time, eye-key span, total task time, duration of the orientation, drafting and revision phases, number of pauses longer than 5s, mean length of pauses longer than 5s, self-reports of cognitive effort, weighted rating of the NASA Task Load Index and accuracy. The application of variables that have already been used in the previous studies analysing directionality, like average fixation duration, total gaze time, and total task time, allows me to compare the results obtained in this experiment with the existing studies. Variables like EKS and the weighted rating of the NASA Task Load Index constitute the novelty of this study. They have been identified as measurable

indicators of cognitive effort, but to the best of my knowledge, they have not been used to investigate directionality yet.

4.2 Rationale behind the study

There are several reasons that led me to conduct this study. First and foremost, the question of which direction of translation evokes higher cognitive effort remains unanswered. In many cases, the results obtained in previous studies are inconclusive since they did not reach the level of statistical significance. Some of them suggest that contrary to popular belief, it is the L2-L1 translation direction that requires higher cognitive effort. Because the issue of directionality and cognitive effort is discussed in detail in Section 3.4, I would like to remind here just a few examples. L1-L2 translation was unequivocally proved to be a more effortful direction in the study by Buchweitz and Alves (2006). This was confirmed by all the analysed variables. In the study by Pavlović and Jensen (2009), only some variables indicated the L1-L2 direction as a more effortful one. However, most of them did not reach the level of statistical significance. Interestingly, two eye-tracking variables, average fixation duration and gaze time during the drafting phase, indicated that L2-L1 translation requires higher cognitive effort. Likewise, no statistically significant difference between L1-L2 and L2-L1 translation processes was found in the study by Ferreira et al. (2021). The influence of directionality on cognitive effort was also investigated in the Polish-English language pair. The studies by Whyatt (2018; 2019), in which directionality was analysed using eye-tracking and keylogging, did not confirm that the L1-L2 direction evokes higher cognitive effort. The differences between the directions of translation were statistically insignificant. Therefore, in this study, I aim to address this research gap created by inconclusive results.

Another rationale behind the analysis of the influence of directionality on cognitive effort was the results from my preliminary study (Pietryga, 2022). I conducted a questionnaire study among students from two Polish universities to investigate their experience with directions of translation and their attitudes towards them. The results indicated some discrepancies among the students depending on the amount of training they had received. According to the general results, L1-L2 is perceived as more cognitively demanding. However, students who received more than 120 hours of translation training claimed that L2-L1 translation is more effortful. A similar division can be observed in the case of the question of the preferred translation direction. The general results indicated that students prefer to translate into their L1. However, those

who received more than 120 hours of translation training would rather work into their L2. It is worth noting that as many as 98.9% of participants agreed that L1-L2 translation should be practised during university courses. These results suggest that the influence of directionality on cognitive effort still requires further analysis.

Finally, since the participants of this study are translation trainees, the results may have didactic implications, influencing the curriculum of translation courses. The results will demonstrate the L1-L2 and L2-L1 translation competencies students possess upon entering the translation market. The study will also give the possibility to confront the subjective perception of directionality from the preliminary results with the results of this study obtained through the combination of subjective and objective process methods.

4.2.1 Collocations as problem triggers

The preliminary results from Pietryga (2022) revealed an interesting link between translation direction and collocations as problem triggers. It appeared that regardless of the translation direction and the amount of translation training, students perceived vocabulary like collocations and idioms as being the most problematic aspect of translation. The notion of collocation appeared in linguistics in the 20th century. It was introduced by John Rupert Firth. However, it is claimed that he did not manage to define it precisely (Bernardini 2007; Religa 2009; Przybylska 2020). For the purpose of this study, I adopt the definition developed by Teubert (2004, p. 174), who defines collocation as a unit of meaning with "the co-occurrence of two or more words." It can be observed that co-occurrence is frequently enumerated as the most characteristic feature of collocations (e.g., in Carter 1994; Lewis 2000).

Many researchers emphasise that collocations frequently cause serious problems during the translation process. On the one hand, they are believed to be an essential part of fluent language speaking. On the other hand, the collocations learning process seems to be one of the most complex and time-consuming parts of foreign language acquisition (Newmark, 1981; Pokorn, 2005; Wolter & Yamashita, 2015; Pellicer-Sánchez et al., 2022; Sonbul et al., 2022). There are also some differences in the use of collocations by native and non-native speakers. Non-native speakers use fewer collocations than native speakers. There may also exist the problem of L1 interference (Pellicer- Sánchez et al., 2022; Sonbul et al., 2022). Moreover, collocation learning and translating are affected by factors like congruency, type and frequency of collocations and L1 and L2 fluency (Wolter and Yamashita 2015; Sonbul et al. 2022).

A study analysing the influence of collocation congruency, type and knowledge on the accuracy of translation was conducted by Sonbul et al. (2022). The results indicated that translation trainees tend to translate congruent collocations more correctly than incongruent ones. However, no similar effect was found for parts of speech, collocations length and frequency. The type of collocation knowledge may also influence the translation process. The researchers analysed two types of knowledge, form recall and recognition, but only the first one reached the level of statistical significance.

Moreover, Sonbul et al. (2022: 411) draw attention to a significant research gap in studies analysing translating collocations. To the best of my knowledge, there is no study in which translating collocations as problem triggers was analysed in the context of cognitive effort and directionality. Therefore, I would like to utilise the measures of eye-key span, self-reports of cognitive effort and translation accuracy to analyse translating collocations. It gives the possibility to get insight into the translation process and product, as well as students' perspective of the collocations translation process, including possible problems and tactics used to overcome them.

4.3 Study design

4.3.1 Experiment vs. quasi-experiment

Following the study by Korpal (2016b), I intend to categorise this study as experimental, although it has a more quasi-experimental design. Brzeziński (2008: 13-22) enumerates four rules of a well-designed experiment. First, all hypotheses and the analysed variables should be related to the same paradigm. Second, there should be compliance between the methods used in the experiment and theoretical justifications for adopted hypotheses. The experimental design should eliminate all confounding variables that may affect the results. Third, the researcher should be able to generalise the results to the whole population. Fourth, the results should be interpreted according to the theoretical background as well as the model tested in the experiment. Brzeziński (2008: 51) also points out that a proper experimental study can be manipulated, controlled and measured. For the experiment to be manipulated, it is crucial to randomly assign the participants to the study and control groups (Liu 2010: 100). Such randomisation has several positive effects enumerated by Brzeziński (2008: 48). First, the possible confounding variables have an identical impact on both groups. Second, the groups are more homogenous. Therefore, the individual differences between participants should not influence the results. There are various randomisation methods. Brzeziński (2008: 48)

discusses a popular option of using the random number table. There are also various platforms available online, for example, Randomiser.org.

Nevertheless, Gumul (2020a: 37) states that in the case of translation and interpreting process research, the study design rarely has the form of a proper experiment. Frequently, the characteristics of the study make it impossible to randomly assign participants to control and study groups. Moreover, the research design of directionality studies usually employs only one group of participants who translate the texts in two directions, L1-L2 and L2-L1. These studies are also usually categorised as experiments, although it is impossible to create a control group. Thus, the name experiment is commonly used in translation and interpreting process research (Liu 2010; Gumul 2020a). Examples can be found in Ferreira et al. (2016), Whyatt (2019), and Tomczak and Whyatt (2022).

This thesis will also follow this trend, although it was impossible to randomly assign the participants to the study and control groups. However, to preserve some form of randomisation, both in the case of the pilot study and in the main study, the order of directions of translation was counterbalanced and randomly assigned to the participants. For this purpose, I used the online platform Randomiser.org.[10] The rationale behind this was to avoid the confounding variable of fatigue resulting from the length of the experiment rather than from a particular translation direction.

4.3.2 Research questions and hypotheses

In this study, I would like to answer four research questions. These are:

Research question 1: How does directionality influence cognitive effort?

Research question 2: How do the three stages of the translation process (orientation, drafting, revision) differ in each direction?

Research question 3: How do participants describe cognitive effort related to translating collocations in both directions?

Research question 4: How does directionality influence translation accuracy?

Based on the previous studies, four hypotheses were formulated to be verified in this study. Following the assumptions stemming from the Golden Rule of Translation (Newmark 1988) and translation asymmetry (Kroll and Steward 1994), all the adopted

[10] Source: https://www.random.org/lists/, visited at 27th July 2023.

hypotheses follow the general assumption that L1-L2 translation is more cognitively effortful than L2-L1 translation. Although this assumption is not always true, research questions and hypotheses frequently refer to it, for example, in studies by Pavlović and Jensen (2009), Ferreira et al. (2016), and Whyatt (2019).

Hypothesis 1: Cognitive effort operationalised by total gaze time, average fixation duration, total task time, the number of pauses longer than 5s, mean pause length of pauses longer than 5s, EKS, and weighted rating of the NASA-TLX is higher in L1-L2 direction.

Some of these variables have already been analysed in the context of directionality and cognitive effort. The results of these studies show possible differences, indicating that L1-L2 translation may result in a higher cognitive effort. Although some of them did not reach the level of statistical significance, the general results imply the need for further analysis. For example, in the study by Pavlović and Jensen (2009), the results of two out of four indicators revealed higher cognitive effort in the L1-L2 translation. These were task length and pupil dilation. The second measurement reached the level of statistical significance. The average number of pauses longer than 5s. was analysed in the study by Whyatt (2018). She observed that it is slightly higher in the L1-L2 translation direction. Ferreira et al. (2016) found that average fixation duration and total task time were significantly higher in the L1-L2 translation. Total task time was longer in the L1-L2 direction in the study by Hunziker Heeb (2020).

Hypothesis 2: The orientation phase is shorter in the L1-L2 direction.

This hypothesis is based on the assumption that L2 word processing may evoke higher cognitive effort. Such an observation was made in the study by da Silva et al. (2017). The researchers based their opinion on the number of fixations and gaze time that were higher during processing the source text in L2. The hypothesis adopted in this thesis is also supported by the results from previous studies, like Ferreira et al. (2016) and Whyatt (2018). Ferreira et al. (2016) observed that in the case of L2-L1 translation, source text processing evokes higher cognitive effort than target text processing. Based on the results from mean fixation duration, Whyatt (2018) concluded that the orientation phase is more cognitively demanding in the L2-L1 translation direction.

Hypothesis 3: The drafting phase and revision phase are longer in the L1-L2 direction.

The explanation of this hypothesis stems directly from the phenomenon of translation asymmetry (Kroll & Steward 1984). Due to the existing asymmetry, translation trainees may need more time to translate and then revise the text in their L2. This observation was confirmed based on eye-tracking measures by da Silva et al. (2017). As already mentioned, they found that processing a text written in a foreign language is more cognitively demanding. Cognitive effort during the drafting phase was also analysed by Pavlović and Jensen (2009). The results showed that one out of three indicators of cognitive effort, pupil dilation, was statistically higher while translating the text from L1 to L2.

Hypothesis 4: The content and number of retrospective reports will indicate higher cognitive effort in translating collocations in the L1-L2 direction.

This hypothesis serves as the extension of Hypothesis 1 adopted in this study. I assume that higher cognitive effort in the L1-L2 translation direction may lead to greater informativeness of retrospective reports regarding translating collocations. As pointed out by Englund Dimitrova and Tiselius (2009: 115), translation problem triggers positively impact the number of retrospective self-reports. Therefore, higher cognitive effort may also lead to a greater number of verbalisations since students may face more translation problems. As a consequence, they will have to make more conscious translation decisions. The relation between the retrospective reports and directionality was already analysed by Ferreira (2014). She observed that translators verbalise more after translating text in their L2. Therefore, she concludes that retrospective reports provide a valuable measure of lexical problems.

Hypothesis 5: Translation accuracy will be greater in the L2-L1 direction.

This hypothesis is derived directly from Newmark's Golden Rule of Translation (1988, p. 3), which claims that L2-L1 translation "is the only way you can translate naturally, accurately and with maximum effectiveness." The Effort Model by Gile (1995) also assumes that increased cognitive effort has a negative influence on translators' and interpreters' performance. Moreover, the study by Tomczak and Whyatt (2022) shows

that translators make more lexical mistakes when translating into their L2, even though the difference between the two directions did not reach statistical significance. It should be emphasised that, to the best of my knowledge, translation accuracy has not been studied yet in terms of directionality. However, Whyatt (2019) focused on a related topic – the influence of directionality on the quality of translation. The results show that there may be some differences in the quality of translation depending on the translation direction. However, they did not reach the level of statistical significance. Pavlović (2007b) also investigated the quality of translation. The L1-L2 translation appears to be of inferior quality, confirming the statement of Newmark. These results may lead to the assumption that translation direction may significantly influence translation accuracy.

4.3.3 Independent variable

Directionality (see Section 1.1) serves as the independent variable analysed in this study. I follow the definition developed by Whyatt (2019, p. 79), who explains that in the case of directionality, "translators work into their first or "native" language (L1) or out of their L1 and produce translations in their "first foreign" language (L2)." Although directionality has been the subject of research, neither direction of translation has been unequivocally reported to evoke significantly higher cognitive effort, as researchers did not manage to obtain conclusive results.

In this work, directionality is studied in the Polish-English language pair. Polish is the L1 of the participants, and English is their L2. A difference in the distribution of these languages should be emphasised. English is described as a contemporary lingua franca (e.g., Pavlovič 2007a; Rodríguez-Inés 2022). It also disposes of many resources like dictionaries or parallel texts that can support translators and translation trainees in their daily work. Polish belongs to the group of languages of low diffusion (see Section 1.3), which are "not widely used outside its primary linguistic community or frequently acquired as a second language" (Pavlovič 2007b p. 7).

4.3.4 Dependent variables

In this study, I focus on the following dependent variables that can serve as measurable indicators of cognitive effort: average fixation duration, total gaze time, eye-key span, total task time, the length of the translation process phases: orientation, drafting and revision, number of pauses longer than 5s, mean length of pauses longer than 5s, weighted rating of the NASA Task Load Index, self-reports of cognitive effort, and

accuracy. Following the categories of cognitive effort measures discussed by Ehrensberger-Dow et al. (2020), the enumerated variables embrace subjective, behavioural, and performance measures (see Section 3.3).

Average fixation duration is applied as a global variable and is analysed quantitatively. In this study, I would like to adopt the definition of average fixation duration established by Pavlovič and Jensen (2009, p. 98), according to whom "[t]he average fixation duration indicator (…) is based on total gaze time and the absolute number of fixations." There are several translation process studies which applied average fixation duration as a measure of cognitive effort, also in the context of directionality. As pointed out by Whyatt (2018, p. 99), it can be assumed that "longer average fixation duration (…) [is an] indicative of more effortful processing needed to solve a translation problem." Probably one of the first studies in which cognitive effort was operationalised by average fixation duration was the study by Pavlovič and Jensen (2009). They examined whether the results from average fixation duration and other eye-tracking variables are higher in the L1-L2 translation direction. However, they managed to fully confirm just one of their hypotheses. The average fixation duration appeared to be longer in the L1-L2 translation, but only for students. Average fixation duration analysis can also be found in some more recent studies exploring the issue of directionality and cognitive effort, such as Ferreira et al. (2016), Whyatt (2018), and Whyatt et al. (2021). For example, the study by Ferreira is one of the few in which translators produced significantly longer average fixation duration in the L1-L2 translation. In the case of the study by Whyatt (2018), the average fixation duration was analysed during the orientation phase. Professional translators appeared to produce longer average fixation duration during L2-L1 translation while reading the text in a foreign language.

Total gaze time serves as a global variable and is analysed quantitatively. Various definitions of total gaze time can be found in the literature on translation and interpreting process. I would like to follow the one proposed by Jensen et al. (2009, p. 322) since the authors provided a detailed description of the variable. According to them, total gaze time is defined as "the combined duration, in milliseconds, of all fixations on a given area of the screen during a given task." Total gaze time has already been studied in the context of directionality and cognitive effort. Many studies indicated that increased total gaze time can be directly linked to increased cognitive effort (e.g., Jensen et al. 2009; Pavlovič & Jensen 2009; Sjørup 2013; da Silva et al. 2017; Whyatt 2018). As pointed out by Jensen et al. (2009: 322), the relation between total gaze time and cognitive effort is deeply

embedded in literature and derives from the eye-mind hypothesis by Just and Carpenter (1980). One of the first modern studies analysing total gaze time in the translation context was probably the work by Jakobsen and Jensen (2008), in which the researchers focused on various types of reading in translation. In the following years, total gaze time was a subject of analysis, for example, in works by Pavlovič and Jensen (2009). The results of their study did not confirm the hypothesis that L1-L2 translation results in longer total gaze time. Contrary to researchers' expectations, both students and professionals produced longer gaze time during L2-L1 translation.

Eye-key span (EKS) denotes "the time lag between a fixation on a ST word and the first keystroke related to producing its TT equivalent" (Dragsted 2010, p. 50). The concept of time lag is defined by Timarová et al. (2011, p. 121), who describe it as "the temporal delay between source text (ST) input and target text (TT) output." EKS evolved from other time lag measures, namely, an ear-voice span (EVS). EVS is analysed in the case of interpreting studies where measuring time lag has a longstanding tradition (Dragsted & Hansen 2008). It was recently analysed, for example, in studies by Gumul (2006), Lijewska et al. (2017), and Collard and Defrancq (2019). It is also possible to analyse the time lag in sight translation. In this case, an eye-voice span (IVS) is measured. IVS is the subject of analysis, for example, in studies by Chmiel et al. (2020), Chmiel and Lijewska (2022), Wenchao (2023), and Gumul and Pietryga (manuscript under preparation). The concept of time lag in interpreting studies is still evolving, and its new variations appear. For example, the pen-eye-voice span (2018) and the ear-pen span (2020) were introduced by Chen when investigating consecutive interpreting.

There are two manners of measuring EKS in translation, either from the first or from the last fixation before the typing activity occurs (Dragsted 2010: 51). As pointed out by Dragsted (2010, p. 51), EKS from the first fixation "seems to span the entire preparation or planning phase preceding the production of a word." However, it is prone to many distortions resulting from refixations on the same word or fixations on other parts of the text. Therefore, I decided to analyse EKS from the last fixation because it "indicates the immediate effort of switching from the reading mode to the writing mode" (Dragsted 2010, p. 51). As Dragsted (2010, p. 51) continues, "the EKS from the last fixation invariably involves a coordination/transformation effort, because, during this time span, a fixation on a ST word is actively transformed into a TT equivalent which is typed in the TT window of the screen." As a result, EKS from the last fixation indicates cognitive effort appearing directly before the typing activity. EKS is also a very objective type of

variable in that, like other eye-tracking and keylogging variables, it is resistant to participants' conscious behaviour. (Timarová et al. 2011: 122).

EKS, although utilised in translation studies since the beginning of the 21st century, is highly under-researched. Writing in 2011, Timarová et al. even pointed out that "EKS measurement in translation process studies using eye-tracking and keystroke logging technologies is still in its infancy" (2011, p. 134). Although the present study was conducted more than ten years later, not many works covering the topic of the EKS had appeared. Two main studies discussing EKS are the works by Dragsted (2010) and Timarová et al. (2011). In her study, Dragsted (2010) focused on the EKS in coordinating reading and writing processes among students and professionals. The results indicated that students produce longer EKS than professionals. It was observed that students are less able to simultaneously read the ST and produce the TT. Timarová et al. (2011) also compared EKS between students and professionals. Their study revealed that professionals produce shorter EKS that are usually more stable and not prone to many fluctuations. Therefore, the authors of the study suggest that "[t]he higher EKS values among students may be taken as an indicator of the time it takes for the student to construct meaning based on the ST and to switch mode to commence production" (Timarová et al. 2011, p. 132). That is why both Dragsted (2010) and Timarová (et al. 2011) agree that EKS constitutes a reliable measurement of cognitive effort experienced during the translation process.

It shall be emphasised that until now, EKS has not been studied in the context of directionality, which is the subject of this study. Moreover, this is the first study applying EKS as a local variable measuring cognitive effort in translating collocations, and it is analysed quantitatively. Both studies, by Dragsted (2010) and Timarová et al. (2011), focused on different units of the text, like the beginning of the sentence or random word pairs. They also compared EKS produced by students to those produced by professionals. Therefore, my study is the first one focusing particularly on a group of advanced students.

Total task time is applied as a global variable, referring to the whole translation process and is analysed quantitatively. Although the variable of total task time has been present in the translation process studies since the early 2000s, I would like to follow one of the most recent definitions applied in the study of cognitive effort and directionality. It was developed by Hunziker Heeb (2020, p. 82), who uses the term 'process duration' and defines it as "the time it took the translator to accomplish the task, i.e. to produce a target text she thought fit for its purpose." Although the author, Hunziker Heeb, modified

the name of the variable, it still follows one of the most popular definitions of total task time by Pavlovič and Jensen (2009, p. 94). They defined it as "the total time it took the subject to complete the given translation task." Many studies (e.g., Pavlovič & Jensen 2009; da Silva et al. 2017; Whyatt 2019; Whyatt et al. 2021) indicate that total task time may be successfully analysed as an indicator of cognitive effort in the context of directionality. For example, Pavlovič and Jensen (2009, p. 98) emphasise that total task time "is considered synonymous with increased cognitive effort in that we equate processing time with cognitive effort." However, it is worth noting that the results of the studies did not confirm translation asymmetry. The L1-L2 translation was statistically longer than the L2-L1 translation only in the case of one study by Pavlovič and Jensen (2009). In the case of the remaining studies, the results did not reach the level of statistical significance.

The analysis of the length of the translation process phases: orientation, drafting and revision (see Section 2.4.2) serves as the extension of the previous variable, total task time and is analysed quantitatively. It enables a more detailed analysis of possible differences in the duration of translation in two directions. The length of the translation process phases has been pointed out as a reliable indicator of cognitive effort. It was studied, for example, in works by Whyatt (2018) and Whyatt et al. (2021). In the case of the first work, the researcher analysed the amount of time invested in each of the phases during the L1-L2 and L2-L1 translation process. It appeared that both the drafting and revision phases were slightly longer during L1-L2 translation, which can be associated with higher cognitive effort. In their second work, Whyatt and her colleagues (2021) analysed the three stages of the translation process in relation to the time spent in online resources. The results indicated that directionality had a statistically significant influence on the orientation and drafting phases but did not influence the revision phase. In the case of orientation, translators devoted more time to online resources while translating into their L1. In the case of drafting, L1-L2 required more Internet searches.

The number of pauses longer than 5s serves as a global variable analysed during the drafting phase. It is also analysed quantitatively in this study. Pauses are defined by Muñoz and Apfelthaler (2022, p. 23) as "time spans of no-recorded activities assumed to be mainly task-related, which tend to happen between words and higher language units." As the researchers continue, pauses: "are conscious, intentional and part of the keylogged task flow – which they break down into task segments – but not of typing" (Muñoz & Apfelthaler, 2022, p. 23). They were analysed as indicators of cognitive effort in many

studies (e.g., Alves 2006; Immonen 2006; Dragsted 2010; Whyatt 2018; Whyatt 2019; Muñoz & Apfelthaler 2022). Timarová et al. (2011, p. 123) suggest that "[p]auses have been used as the primary indicator of cognitive processing, interpreted as indicators of hesitation and boundaries between text production units or, in translation, translation units and segments." However, a threshold of the minimum pause duration should be established to distinguish pauses from the usual typing pace or other similar phenomena (Dragsted 2010; Immonen & Mäkisalo 2010; Kruger 2016; Muñoz & Apfelthaler 2022).

Various thresholds of pause duration can be found in the literature. Kruger (2016) identified pauses of a minimum length of 3s as cognitive effort indicators. While Hunziker Heeb (2020) opted for the minimum length of 5s to study cognitive effort in the context of directionality. Probably one of the lowest thresholds was applied in work by Muñoz and de León (2018), who defined it as 200ms. What is more, the researchers introduced the categorisation of pauses based on their length; "short pauses, possibly associated to typing micro strategies; long pauses, customarily linked to problem-solving activities; and pauses between thresholds, or mid pauses" (Muñoz & de León 2018, p. 44). Following the studies by Whyatt (2018, 2019) and Hunizker Heeb (2020), who analysed pause duration in the context of directionality, I decided to focus on the number of pauses longer than 5 seconds as indicators of cognitive effort. Interestingly, in the case of the study by Whyatt (2018), just a slight difference in the number of pauses longer than 5s was observed between the directions. There were only two pauses more in the L1-L2 translation direction. In the later study, higher cognitive effort was identified in the reverse translation direction. Such results indicate the need for further research on this variable regarding the influence of directionality on cognitive effort.

The mean length of pauses longer than 5s serves as a global variable and is analysed quantitatively. As pointed out by Immonen and Mäkisalo (2010, p. 45), "the pause length reflect[s] the demands of the cognitive processing." Mean pause length was analysed, for example, in works by Immonen (2006) and Immonen and Mäkisalo (2010). Immonen and Mäkisalo (2010) compared translation and writing processes. They analysed pause length before three different units: clauses, phrases and words. The results indicated that the segment following the pause has a significant influence on its length. It was observed that in the case of translation, the longest pauses appear before shorter units. These results confirmed the findings from the previous study by Immonen (2006). The researcher observed some differences in the mean pause length in translation and writing.

Pauses preceding longer units like sentences and paragraphs are longer during writing, while pauses preceding shorter units like words are longer during the translation task.

I decided to apply the variable of the mean length of pauses longer than 5s to my study to further develop the analysis of the number of pauses. The number of pauses has already been discussed in relation to directionality and, in this respect, led to inconclusive results. However, the influence of directionality on the mean pause length is still unexplored. To the best of my knowledge, the only study that analysed mean pause duration in the context of directionality was conducted by Hunziker Heeb (2020); however, the results did not reach the level of statistical significance. In fact, the analysis of the mean pause duration is derived from interpreting studies. It is frequently applied in the interpreting process research, where particular attention is placed on the duration of so-called silent pauses, for example, in studies by Gumul (2021a; 2021b) and Chmiel et al. (2022).

Weighted rating of the NASA Task Load Index is applied as a global variable and is analysed quantitatively. As pointed out by Chen (2017, p. 650), "[t]he NASA Task Load Index (Hart & Staveland 1988) is one of the most widely used scales for measuring mental load." The NASA Task Load Index was applied in many translation and interpreting studies, for example, by Sun and Shreve (2014), Zhou et al. (2022), and Gieshoff and Hunziker Heeb (2022, 2023). One of the aims of the study by Sun and Shreve (2014) was to verify whether the NASA-TLX is a reliable research tool that can be utilised to measure translation difficulty. Even though the researchers focused on only four out of six factors of the NASA-TLX, the study confirmed the reliability of the scale. Gieshoff and Hunziker Heeb (2022, 2023) and Zhou et al. (2022) applied the NASA-TLX to measure cognitive effort in translation and interpreting tasks. Gieshoff and Hunziker Heeb (2022, 2023) found that cognitive effort and cognitive load ratings refer to closely related concepts. It appeared that translators and interpreters may consider both elements to be somewhat similar. Zhou et al. (2022) focused on the impact of task complexity on self-reported cognitive effort. The results indicate that the complexity of the translation task significantly influences the perceived cognitive effort measured by the weighted rating of the NASA-TLX. To the best of my knowledge, this thesis is the first study that uses the weighted rating of the NASA Task Load Index to study the influence of directionality on cognitive effort. Moreover, the studies mentioned above applied only selected factors of the NASA-TLX weighted rating and in the case of this thesis, the complete questionnaire is used.

Self-reports of cognitive effort serve as local and global variables and are analysed quantitatively and qualitatively. Such a combination of quantitative and qualitative analysis of the reports can be found in Pavlovič (2010). It is believed that "verbal protocols offer an invaluable insight into translation process in general, and the decision-making aspect of those processes in particular" (Pavlovič 2010 p. 83). That is why retrospective protocols were also used to analyse cognitive effort, for example, in studies by Gumul (2018, 2019a), where the researcher compared self-reports of cognitive effort with problems found in the TT.

Verbalisations in the form of retrospective reports (Ferreira 2014) and collaborative think-aloud protocols (Pavlovič 2007b; 2010) were also analysed in the context of directionality. Pavlovič (2010) looked for self-reports referring to tentative solutions applied while translating in both directions. The researcher focused both on the number and the content of self-reports. The results indicated that although similar categories of comments can be detected in both directions, participants tend to produce more verbalisations about tentative solutions to problems during L2-L1 translation. A similar study was conducted by Ferreira (2014: 121), who also focused on the number of self-reports referring to categories introduced by Pavlovič (2010), like actions, problems, solutions and verbalisations. According to the results, L1-L2 translation caused more problems, and translators were not able to solve them as effectively as in the case of the L2-L1 direction. To the best of my knowledge, retrospective reports have not yet been analysed in relation to collocations. Therefore, I decided to analyse the content and the overall number of self-reports of cognitive effort that are related to the process of translating collocations.

Accuracy is applied as a local variable and is measured in relation to the translation of collocations. It is analysed quantitatively. Accuracy is believed to be one of the two measures that allow for tracing cognitive effort during the translation process (Timarová et al. 2011; Sun 2015). The relationship between accuracy and directionality can be traced back to the work by Newmark (1988), who, in his Golden Rule of Translation, suggests that only translation into L1 can be described as accurate. Accuracy serving as an indicator of cognitive effort in the context of directionality is under-researched. However, a similar phenomenon, the influence of directionality on the quality of translation, was studied by Whyatt (2019). The results indicated that directionality may have an impact on the quality of translation. Grammar tends to be of poorer quality in L1-L2 translation. However, more punctuation and sense problems were found in the L2-L1

direction. Accuracy is also frequently analysed in the context of interpreting. For example, de Groot (1992) studied the influence of the frequency of words on accuracy, Stachowiak-Szymczak and Korpal (2019) focused on the influence of numbers on accuracy, Gieshoff and Albl-Mikasa (2022) conducted a unit-based accuracy analysis, and Nicodemus and Emmeray (2015) focused on signed language interpreting.

Following the study by Whyatt (2018, 2019) and Whyatt et al. (2021), I asked persons not related to this study to act as proofreaders. Since their task was not to proofread the whole text but to verify the translation accuracy of chosen interest areas, I refer to them as experts. In this thesis, accuracy is assessed by four experts (three for each direction). Two of them are native speakers of Polish, who are also English scholars. The third expert is a native speaker of Polish and a Polish scholar, while the fourth expert is a native speaker of English and an English scholar. As a result, the accuracy is assessed, by the competencies of native speakers and philologists. The experts were instructed to assign 0, 1 or 2 points depending on how accurate the translation of a given collocation is. The scale was based on the work by Andermann and Rogers (1997), as they raised the subject of translation assessment in the didactic context. They suggest that "[t]he assessment may take the form of a judgement such as 'task completed', 'task partially completed' or 'task not completed'" (Andermann & Rogers 1997, p. 61). I decided to expand these categories by assigning points to them. 0 points were assigned to the category 'task not completed', 1 point to the category 'task partially completed', and 3 points to the category 'task completed'.

4.3.5 Methods

A mixed-methods approach was adopted in this study. The following qualitative and quantitative methods for data collection were used both in the pilot and in the main study.

Eye-tracking was applied as an objective method. The eye-tracker used in the study was Eyelink Portable Duo, recording monocularly in the remote mode. It means that the participants were able to freely move their heads and use both the keyboard and the screen. However, they were asked to sit as still as possible, avoid any sudden movements, and reduce their head movements to the minimum to avoid losing the data. Due to the remote recording mode, participants were obliged to wear a sticker on their forehead that allowed the tracker to properly trace the participant's head position. For calibration, the

setting of 13 points was applied. The sampling rate at which the tracker was recording the data was 1000 Hz.

Keylogging was applied as an objective method. Following the studies by, for example, da Silva et al. (2017) and Kajzer-Wietrzny et al. (2016), the keylogging software used in the study was Translog II (Carl 2012). Texts used in the study were presented on a grey background to avoid unnecessary pupil dilation after the calibration phase. The screen was divided into two windows. The source text was presented in the upper window, and the target text was typed in the lower window. Participants could work at their normal pace and manner of translation and freely edit the target text. However, they were not allowed to consult any external resources such as online sources or paper dictionaries. Such a possibility constitutes a risk of a significant reduction of the cognitive effort that the participants may experience while facing problems in the translation process.

Retrospective verbal reports were applied in the form of self-retrospection. In this subjective method, participants were asked to comment freely on their cognitive processes appearing during the translation task. Following the studies by Englund Dimitrova and Tiselius (2009; 2014), source texts were used as the retrieval cues. Another rationale behind using source texts rather than target texts was the study by Gumul (2020b). Its results revealed that the type of cueing influences neither the informativeness nor the accuracy of retrospective reports. Moreover, extracting the target texts from the keylogging programme to use them as the stimulus would significantly lengthen the time that would have to pass between the translation task and the retrospective session. That could negatively influence the self-reports, as the retrospective session should be performed as soon as possible after the translation task (Ericsson & Simon 1993, Gumul 2021c). No other form of cueing was provided. Retrospective protocols were transcribed and coded for the purpose of the analysis.

The NASA Task Load Index (Hart & Staveland 1988) is a questionnaire designed to study workload in various settings. It was applied in this study as a subjective method. The NASA-TLX was already used in translation and interpreting process research, for example, by Chen (2017) in her study of cognitive load in the interpreting process. The workload is measured based on a weighted average of six elements: mental demand, physical demand, temporal demand, performance, effort and frustration. The questionnaire consists of two measurements: rating scales and sources of workload. First, participants decide how demanding each of the six elements is by marking an appropriate

point on a scale. The scale and the definitions of the elements are presented in Appendix 9. In the next step, participants obtain 15 pairs of the already mentioned elements, and their task is to decide which element from the pair contributed more to the perceived workload. The questionnaire was distributed to participants in paper form after the retrospective session was finished.

Another subjective method applied in this study is a self-designed questionnaire.[11] This questionnaire consisted of two questions and was distributed to the participants in a paper form. It was the very last task of the experimental procedure. The aim of the questionnaire was to determine whether any of the translation directions was perceived by the participants as more cognitively demanding. They also had the chance to explain their decision. The questionnaire is included in Appendix 8.

4.3.6 Materials

In order to analyse the influence of directionality on cognitive effort, all participants translated two texts, one from Polish into English (L1-L2) and one from English into Polish (L2-L1). The texts used in the study were retrieved from the National Geographic website. They were modified to ensure matching lengths and difficulty levels, aligning with the study's objectives. The texts can be characterised as popular science texts covering the topic of animals. However, they did not require any specialist knowledge or prior preparation. The Polish text discusses the topic of the oldest tortoise in the world. The English text is related to dogs and their manner of processing praise. These texts were used both in the pilot and main studies. Both texts can be found in Appendices 5 and 6.

The texts were balanced in terms of their readability features according to the Fog index. The readability of the Polish text was 11, which was verified through the Logios website.[12] It was also analysed through Jasnopis.pl[13], which serves to verify the readability of Polish texts. It assessed its readability as 4/7, which means that the text is difficult and should be understood by readers who have secondary education. The readability of the English text was 11.3, and it was verified through the Readability Formulas website[14]. According to this website, such a level of readability can be

[11] The questionnaire was added to the research design after the pilot study to complement the remaining methods and give the participants the possibility to present their position on directionality.
[12] Source: https://dozabawy.logios.dev/, visited 5th February 2022
[13] Source: https://www.jasnopis.pl/about-us/, visited 5th February 2022
[14] Source: https://readabilityformulas.com/ visited 6th February 2022

characterised as difficult and should be understood by readers with a high-school education.[15] To the best of my knowledge, there is no tool that can compare readability formulas across multiple languages. Therefore, I had to utilise different websites to verify each text. The texts were also similar in their length. The Polish text had 168 words, and the English text had 171 words. It is worth mentioning that the length of the texts was influenced by the constraints of eye-tracking and keylogging programmes. Longer texts would have required scrolling, and the eye-tracking software used in this study does not implement scroll compensation in the screen recording mode. Therefore, there was a high risk of data loss in the case of longer texts requiring scrolling. A detailed comparison of the two texts is presented in Table 1 below.

I also manipulated the texts to ensure an equal number of collocations, the translation of which was analysed in terms of their accuracy, eye-key span, and retrospective reports. Initially, each text contained 14 collocations, including seven noun collocations and seven verb collocations. However, as some of the collocations seemed to be too close together, there was a possibility that the participants might perceive them as one long phrase. Therefore, I decided to restrict the number of collocations to 12 in each text. The final sample of collocations consisted of six noun and six verb collocations in each text. The collocations in the Polish text were retrieved from the Narodowy Korpus Języka Polskiego (NKJP) (Pęzik 2012). All collocations used in the text belong to the 21 most popular collocations. The collocations in the English text were retrieved from the British National Corpus (BNC) (Davies 2004). All collocations used in the text belong to the 46 most frequent collocations, and only one of them was listed as the 46th most frequent. The remaining ones belong to the 25 most frequent collocations. The list of collocations used in the study is included in Appendix 7.

The source texts and their translations were recorded by Translog II, the programme used for keylogging. The texts were written in Times New Roman, black font, size 22, with double spacing. This setting allowed for proper eye-tracking data gathering. The texts were presented on a grey background to avoid unnecessary pupil dilation after the calibration phase was finished.

[15] Source: https://readabilityformulas.com/freetests/six-readability-formulas.php, visited 21st July 2023

	POLISH SOURCE TEXT	ENGLISH SOURCE TEXT
topic	The oldest turtle in the world	Processing praise by dogs
number of words	168	171
FOG index	11	11.03
number of collocations	12	12

Table 1. Details of the source texts.

4.3.7 Ethical issues

Since students of the University of Silesia in Katowice participated in this study, a positive opinion of the Ethics Committee was required to conduct the experiment. The Committee thoroughly assessed the study design and instructions for the participants before issuing an opinion. Firstly, they checked whether the participants were properly informed about the study. All the participants of the pilot study and the main study received detailed instructions about the course of the experiment, its predicted duration, and the used methods. It was emphasised that none of the methods are harmful to people. They received all the information again right before the experimental procedure.

Before the experimental procedure began, participants signed the consent to have their data recorded, analysed and published, for example, in the form of a PhD thesis and articles. The consent and the GDPR form are included in Appendices 1 and 2. Moreover, participants of the experiment remained anonymous, as their data were pseudonymised. Their surnames were replaced by numbers. I also removed from verbal reports information, which may lead to their identification. It should be emphasised that none of the participants were my students. In such a way, I wanted to minimise the white-coat effect resulting from the dependence relationships. The Committee also recommended that in line with good practices, the participants should obtain some benefits for their involvement in the study and the time devoted to it. Therefore, university gadgets were distributed as compensation to all participants of the pilot study and the main study.

The Ethics Committee also ensured that the experts responsible for assessing the accuracy of the translation of collocations were adequately informed about their role in the project. Each expert signed the consent to assess the accuracy of the translation of collocations. These documents contained detailed information about their task in the experiment.

Finally, a positive opinion no. KEUS192/12.2021 was issued by the Ethics Committee at the University of Silesia in Katowice. From this moment, I was allowed to invite participants to take part in the experiment and to conduct the pilot study and then the main study.

4.3.8 Pilot study research design

The pilot study was conducted between the 1st and the 4th of April 2022. Its aim was twofold. First of all, it served to verify the research design, adopted methods, and procedure and to implement all necessary modifications to the main study design. Secondly, it aimed to gain preliminary results on how directionality influences cognitive effort. Six translation trainees (five women and one man; age: 22-24, M=23.5, SD=0.8) who were 4th-year English philology students of the translation and interpreting programme at the University of Silesia in Katowice participated in the pilot study. All pilot study participants were recruited on a voluntary basis. They reported either normal or corrected-to-normal vision. At the moment of the pilot study, all the participants had completed at least 90 hours of translation classes.

The pilot study began with lexical and typing speed pre-tests conducted at the Faculty of Humanities at the University of Silesia in Sosnowiec. These tests aimed to prevent confounding variables such as excessively low linguistic abilities or excessively slow typing that would impede the translation process. Additionally, demographic data of the participants were gathered. Firstly, participants were asked to fill in the Lextale test (Lemhöfer & Broersma 2012), which served to verify their proficiency in English as their L2. During the test, conducted via an online platform, they were presented with 60 words, and their task was to decide whether a given word was an existing word in English[16]. The results of the Lextale test indicated that all the pilot study participants have a very proficient knowledge of English (M=92.5%, SD=3.29). In the next step, participants were asked to assess their six linguistic abilities: listening, reading, spoken interaction, spoken production and writing in their L1 (Polish) and L2 (English) by filling in the self-assessment grid table prepared by the Council of Europe (2001). The self-assessment grid tables are presented in both languages in Appendices 3 and 4. The results indicated that all participants assessed their abilities as being at least at the B2 level in English and Polish. It is worth mentioning that the majority of participants indicated that they have

[16] Source: https://www.lextale.com/whatislextale.html, visited 20th July 2023

the C1 or C2 level of English, and there were only several persons who indicated the B2 level. The next step consisted of typing speed tests in Polish and English conducted via an online platform, 10FastFingers.com[17]. Participants were asked to copy the texts in two languages that were presented to them on a computer screen. The results showed comparable typing skills in both languages. The mean typing speed was 40 words per minute (WPM) for Polish (SD: 7.8) and 39.5 WPM (SD: 6.5) for English. Information about the participants of the pilot study is summarised below.

age	M: 23.5
number of participants	6
year of studies	4
number of completed hours of translation classes	>90h
Lextale test	92.5%

Table 2. Summary of pilot study participants' demographic data.

	POLISH	**ENGLISH**
linguistic skills	>B2	>B2
typing speed	40 WPM	39.5 WPM

Table 3. Information about the pilot study participants' skills.

The experiment took place on a different day in the Research Laboratory of the Institute of Psychology at the University of Silesia in Katowice. The experiment began with the PEBLE typing speed test (Mueller & Piper 2014) which served as a practice of using the keyboard that the laboratory is equipped with. This time, participants copied non-existing words composed of random order of letters. After the PEBL typing speed test was completed, participants were asked to place an eye-tracking sticker on their foreheads, sat in front of the computer, and were reminded of the course of the experiment and the tasks they were being asked to perform. I began the preparation of the equipment, as well as the calibration of the eye-tracker.

[17] Source: https://10fastfingers.com/, visited 25th July 2023

The experiment consisted of three tasks in two directions, L1-L2 and L2-L1. The tasks are translation, retrospective session and filling in the NASA Task Load Index. The order of directions was counterbalanced to avoid a confounding variable in the form of fatigue resulting from the length of the experiment. Participants were randomly assigned to the order of directions of translation. Randomisation was conducted using an online platform, Randomiser.org. There were 5-minute breaks after the translation task and after the retrospective session. A 15-minute break took place after filling in the NASA-TLX in the first direction so that participants could rest before performing the same three tasks in the opposite direction.

During the translation task, participants were asked to translate the text presented to them on the screen in Translog II software. The materials used in the study have been discussed in section 4.3.6. The eye-tracker and the keylogging programme recorded the translation process. Participants were informed that they should work at their normal pace, as they usually do during classes or homework assignments. However, they were not allowed to consult any external resources either in online or in paper form. As mentioned in Section 4.3.5, any use of external resources could significantly influence the study results.

The second task was the retrospective session, during which the participants performed self-retrospection. They received written instructions on the retrospective session and were asked to comment freely on the conscious processes that they experienced during the translation task. Participants were informed that they could verbalise as many retrospective comments as they wished and were not restricted by any time limits. As mentioned in Section 4.3.5, source texts were used as the retrieval cues. Following the study by Tiselius and Jenset (2011), the researcher was present in the room during the retrospective session. However, I was outside participants' field of vision and had assured them earlier that I was not listening to or focusing on what they were saying. The only purpose of my presence was to help in case of any technical problems or arising questions. The last task was to fill in the NASA Task Load Index. The participants obtained a complete questionnaire together with instructions. After a 15-minute break, the three tasks were repeated in the reverse direction.

A short debriefing session took place after the whole experimental procedure was finished, which was aimed at gaining information about the retrospective session. The answers helped me implement the modifications in the main study design. Two questions were asked:

- Which type of cueing is better: source text, target text, target text replay in the Translog II, or target text replay in the DataViewer?
- Was the researcher's presence in the room during the retrospective session confusing or uncomfortable to you?

Some participants reported that they would prefer the cueing in the form of either the target text or the target text that is replayed in Translog II. However, there were also opinions indicating that the present cueing was adequate. In the case of the second question, some participants answered affirmatively, suggesting they would feel more comfortable on their own during the retrospective session. One participant stated that they would prefer to talk directly to me. The remaining participants did not object to my presence in the room during the retrospective session.

A summary of the steps of the pilot study is presented below:

1) Filling in the Lextale test;
2) Filling in the self-assessment grid table regarding linguistic abilities in English and in Polish;
3) Performing typing speed tests in English and in Polish;
4) Performing PEBL typing speed test;
5) Preparation for the experimental study and calibration of the eye-tracker;
6) Translation task;
7) Retrospective session;
8) Filling in the NASA-TLX questionnaire;
9) 15-minute break;
10) The steps 5-8 were repeated in the reverse direction of translation;
11) Short debriefing session.

Steps 1-3 were conducted at the Faculty of Humanities at the University of Silesia in Sosnowiec, and steps 4-11 were conducted another day in the Research Laboratory of the Institute of Psychology at the University of Silesia in Katowice.

The following modifications to the study design were made after the pilot study. First of all, based on the comments from the debriefing session, I decided that in the main study, I would not be present in the room during the retrospective session, and participants would perform self-retrospection alone. It should also help to minimise the white-coat effect. Another rationale can be found in studies by Bartłomiejczyk (2007) and Gumul (2006, 2017). In the case of both studies, participants always performed self-retrospections while being on their own in the interpreting booths. Secondly, following

the participants' comments, the 15-minute break separating experimental tasks in two directions was reduced to 5 minutes. Participants frequently stated that there was no need for such a long break. For the same reasons, 5-minute breaks between the tasks were cancelled. Such a decision allowed me to expedite the experiment and reduce the fatigue effect. Moreover, some of the participants had reported that they did not need any form of a break. Finally, I decided to add an extra questionnaire that would be filled in after the experiment, in which participants will have a possibility to indicate whether they perceive any direction of translation as more cognitively demanding. It will allow me to compare the results obtained through objective methods like eye-tracking and keylogging with participants' subject perception of directionality. The pilot study also allowed me to develop the categorisation of cognitive effort reported in retrospective reports that will be discussed thoroughly in the analysis.

The results of the pilot study indicated higher cognitive effort in the L1-L2 translation direction. However, due to the small number of participants, a calculation of inferential statistics was impossible, and the results were based solely on descriptive statistics. These results indicated that the topic is worth further analysis, including inferential statistics and linear regressions.

4.3.9 Main study research design

The main study was conducted between 29th April and 9th June 2022 at the University of Silesia. I decided to replicate the research design of the pilot study and include the modifications discussed in Section 4.3.8. The modified research design will be presented in detail in the following subsections.

4.3.9.1 Participants of the main study

Thirty-five translation trainees who were 5th-year English philology students of the translation and interpreting programme at the University of Silesia in Katowice participated in the main study. All participants were recruited on a voluntary basis. The group comprised 29 women and six men aged 23 to 26 (M= 24, SD= 0.78).

All participants reported normal or corrected-to-normal vision. However, due to some calibration issues resulting, for example, from the fact that some participants wore either glasses or lenses or were unable to prevent themselves from making too expressive movements, the eye-tracking data from 10 out of 35 participants had to be discarded. That is why, when analysing the eye-tracking variables (i.e., average fixation duration, total

gaze time, and eye-key span), I will refer to 25 participants who managed to produce valuable eye-tracking data. Nevertheless, the participants whose eye-tracking data were lost managed to produce valuable data gathered through other methods used in the study, keylogging, retrospective sessions, and questionnaires. That is why I decided to include the data from all 35 participants in the analysis of the remaining variables to prevent excessive data loss. Therefore, the final sample consists of 25 eye-tracking data sets and 35 data sets gathered through keylogging, retrospective sessions, and questionnaires.

All translation trainees were participating in the study a few weeks before their final examinations in translation and interpreting and the defence of their MA thesis. Thus, it can be assumed that they had high translation skills as they were on the verge of entering the translation market. They had completed at least 60h of translation classes. To be more precise, 2 out of 35 students indicated that they completed 60-120h of translation classes, which equals 3-4 semesters. Nineteen of the students said that they had completed 120-180h of translation classes, which equals 4-6 semesters, and 14 students stated that they had completed more than 180h of translation classes. A summary of the main study participants' demographic data is presented below.

age	M: 24
number of participants	35 (25 participants qualified for eye-tracking variables analysis)
year of study	5
number of finished hours of translation classes	>4 semesters
Lextale test	76.8

Table 4. Summary of main study participants' data.

4.3.9.2 Procedure of the main study

The main study procedure began with lexical and typing speed pre-tests conducted at the Faculty of Humanities at the University of Silesia in Sosnowiec. The tests served to verify the participants' language skills before they performed the experimental tasks. I

also collected demographic data at that point. The first of the tests was the Lextale test[18] (Lemhöfer & Broersma 2012) filled in via an online platform. The results of the Lextale test indicated that participants have a proficient knowledge of English (M= 76.8%; SD=10.9).

Participants were also asked to assess their linguistics abilities in English (L2) and Polish (L1) using the self-assessment grid table developed by the Council of Europe (2001). Their task was to assess five aspects of both languages, according to the scale from A1 to C2. These were listening, reading, spoken interaction, spoken production and writing (Appendices 3 and 4). In the majority of cases, participants reported a C2 level of their L1 (Polish), with only some of them indicating a C1 level. In the case of English (L2), the participants reported that their abilities reached at least the B2 level, although the overwhelming majority indicated either C1 or C2 level.

To assess their typing speed in both languages, participants were asked to perform a typing speed test via an online platform. It appeared that participants had similar typing skills in both languages. The mean speed was 43.3 WPM (SD=14.91) in the case of Polish and 45.7 WPM (SD=12.98) in the case of English. A summary of the main study lexical and typing speed pre-tests is presented in the table below.

	POLISH	**ENGLISH**
linguistic skills	>C1	>B2
typing speed	43.3 WPM	45.7 WPM

Table 5. Information about the main study participants' skills.

Like in the case of the pilot study, the experiment was conducted at the Research Laboratory of the Institute of Psychology at the University of Silesia in Katowice and began with the PEBLE typing speed test (Mueller & Piper 2014). It aimed to familiarise the participants with the computer keyboard they would use during the experiment. After the PEBLE typing speed test was completed, the participants were asked to place a sticker on their foreheads, and the preparation of the equipment and calibration proceeded. After the participants were reminded of the procedure and the order of the tasks, the experiment

[18] Source: https://www.lextale.com/whatislextale.html, visited 20th July 2023

started. Like the pilot study, the experiment in the main study consisted of three tasks performed in two directions, in and out of the participants' L1. To avoid a confounding variable in the form of fatigue, the order of directions of translation was counterbalanced and randomised through the web tool Randomiser.org.

The first task was to translate the text presented on the screen. During this task, the eye-tracking data were recorded by the Eyelink Portable Duo eye-tracker and the keylogging data were recorded by the Translog II programme. As in the case of the pilot study, participants of the main study were instructed to work at their normal pace. They were also reminded that using both online resources and paper dictionaries is prohibited.

The second task was a retrospective session. Following the modifications made after the pilot study, participants were placed in a small room, similar to an interpreting booth, where they were asked to comment freely on their conscious processes during the translation task. Participants were also informed that there are no time or length limits for their reports. Following the study by Gumul (2020a), they could verbalise their reports in Polish or English. Nevertheless, none of the participants decided to do this task in their L2 (English). What is most important, this time, I was not present in the room. The cueing in the form of STs remained unchanged. Following the comments obtained during the pilot study, there was no break after the first and second task. Moreover, none of the participants reported the need for such a break.

In the third task, participants were asked to fill in the NASA Task Load Index questionnaire, to which they obtained all the necessary instructions. Participants remained alone in the room while filling in the questionnaire; however, they could ask questions at any time. The lack of a researcher's presence should help them to feel more comfortable while performing the task. After filling in the NASA Task Load Index questionnaire, an obligatory 5-minute break took place. In the next step, the already discussed three tasks were repeated in the reverse direction. Contrary to the pilot study, participants were asked to fill in a short questionnaire related to directionality (Appendix 8) after completing all the tasks in both translation directions. They were asked if any of the directions of translation was more difficult and what is the reason behind their opinion.

A summary of the steps of the main study is presented below:

1) Filling in the Lextale test;

2) Filling in the self-assessment grid table related to linguistic abilities in English and in Polish;

3) Performing typing speed tests in English and in Polish;

4) Performing PEBL typing speed test;
5) Preparation for the experimental study and calibration of the eye-tracker;
6) Translation task;
7) Retrospective session;
8) Filling in the NASA-TLX questionnaire;
9) 5 minutes break;
10) The steps 5-8 were repeated in the reverse direction of translation;
11) Filling in the questionnaire on the directionality of translation.

As in the case of the pilot study, steps 1-3 were conducted at the Faculty of Humanities at the University of Silesia in Sosnowiec and steps 4-11 were conducted another day in the Research Laboratory of the Institute of Psychology at the University of Silesia in Katowice.

Chapter 5. Results and analysis

The aim of this chapter is to present the analysis of the results obtained in the main study.[19] The chapter is divided into six sections. Five of them correspond to the adopted hypotheses, and the last one is devoted to linear regression. I begin the analysis with descriptive statistics, providing information about the mean results. I will present histograms and analyse kurtosis and skewness for the results that do not have a normal distribution. Finally, I focus on inferential statistics, discussing the results of the statistical tests. It will allow me to confirm or reject the adopted hypotheses.

All the eye-tracking variables, total gaze time, average fixation duration and EKS, were extracted from the recordings and analysed using the DataViewer software. DataViewer also helped me to establish the duration of the orientation, drafting and revision phases. In the case of keylogging variables, the number of pauses longer than 5s and the mean length of pauses longer than 5s, I used the Translog Supervisor component. Retrospective reports were transcribed and then manually coded. The eye-tracking and keylogging data also helped me to establish codes for the retrospective reports. The statistical analysis was performed using the SPSS software. As already stated, in the case of eye-tracking variables, I analysed 25 participants who produced valuable eye-tracking data sets. In the case of the remaining variables, I refer to 35 participants.

5.1 Hypothesis 1

In Hypothesis 1, I assumed that cognitive effort operationalised by total gaze time, average fixation duration, total task time, the number of pauses longer than 5s, mean pause length of pauses longer than 5s, EKS, and the weighted rating of the NASA-TLX is higher in L1-L2 direction.

5.1.1 Total gaze time

The mean results of the total gaze time variable indicate some differences between the two directions of translation. The total gaze time in the L1-L2 direction was 872906 ms (SD= 465204.1), which is around 15 minutes. In the L2-L1 translation direction, it was 662403.43 ms (SD= 250377.4), which is around 11 minutes. It appears that participants gazed longer at the screen while translating from Polish into English. The

[19] Detailed results for each variable are included in the Appendix.

detailed differences between both directions of translation can be observed in histograms presented in Figure 3 below.

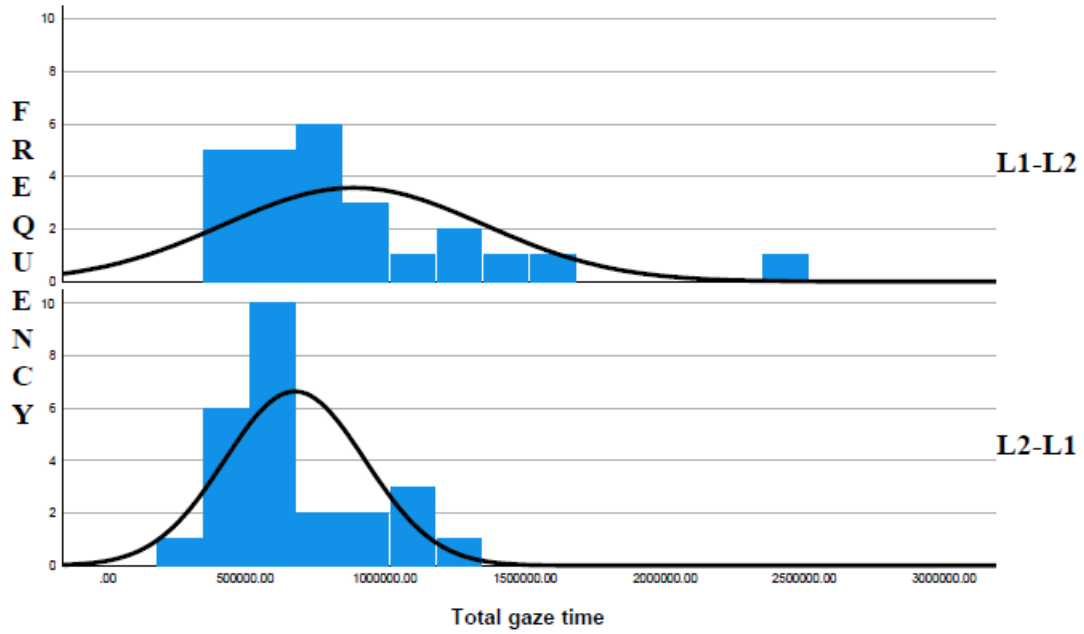


Figure 3. Total gaze time - histograms.

The histograms indicate some differences in the distribution of the total gaze time variable. In the case of the L2-L1 direction, it reaches an observable peak of total gaze time longer than 500000.00 ms. A classic shape of the normal distribution can be observed in the first part of the graph. The results higher than the mode are rather flattened, with a smaller peak of the results longer than 1000000.00 ms. No outliers appeared when participants translated from English into Polish. The results from the L1-L2 translation direction are more flattened. There are no extremely low results, and a high number of results is close to the mode. There is also a more extended distribution of the total gaze time, with an observable outlier of the results longer than 3000000.00 ms. Based on the kurtosis and skewness results (Table 6), leptokurtic distribution can be observed in the L1-L2 translation direction. The histogram is also right-skewed. In the case of the L2-L1 translation direction, the results are close to 0.

	L1-L2	**L2-L1**
kurtosis	4.8	-0.2
skewness	1.9	0.7

Table 6. Total gaze time - kurtosis and skewness.

To further verify the results, I decided to check the normal distribution of the variable and then conduct an appropriate statistical test. The Shapiro-Wilk test showed that there was a normal distribution in the case of the L2-L1 total gaze time (p=0.101). However, no normal distribution was found in the L1-L2 total gaze time (p<0.001). Therefore, I assumed there was no normal distribution, so I conducted the nonparametric Wilcoxon test. The test results indicated significant differences between the means (p=0.002) and reached the level of statistical significance[20]. This indicates that cognitive effort operationalised in total gaze time is significantly higher in the L1-L2 translation direction.

5.1.2 Average fixation duration

The mean results of the average fixation duration also indicate that the L1-L2 translation direction may impose higher cognitive effort. The average fixation duration in the L1-L2 translation was 295.05 ms (SD= 69.3). In the L2-L1 translation direction, it was slightly shorter, 282.5 ms (SD= 61.2). The results show that, on average, participants fixated longer while translating text from Polish into English.

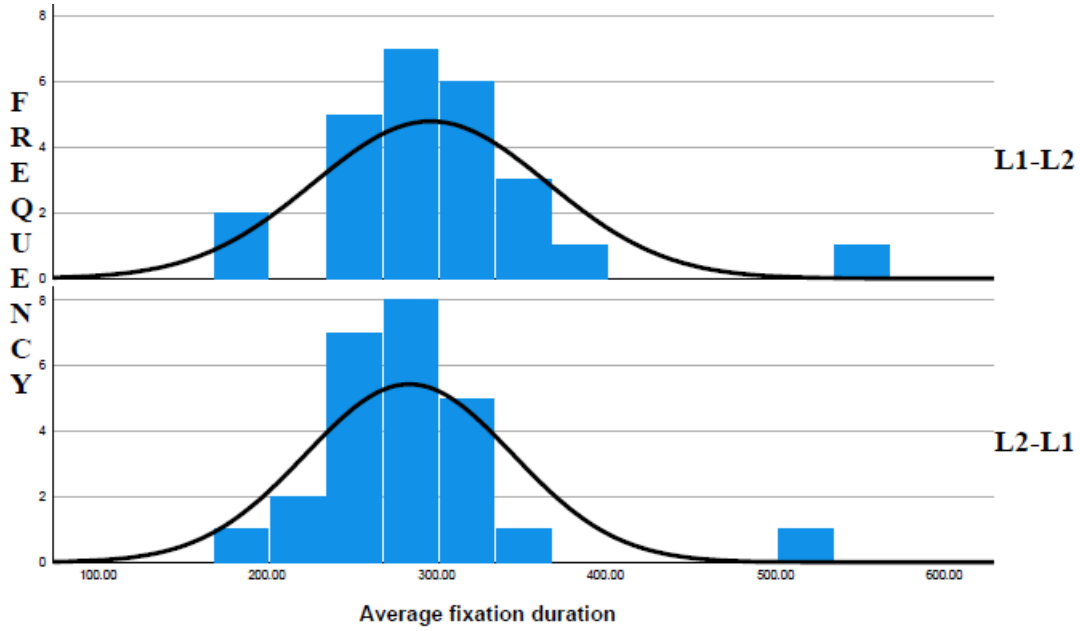


Figure 4. Average fixation duration - histograms.

[20] The summary of statistical tests results is presented in Appendix 11.

The histograms presented in Figure 4 reveal some differences and similarities between the average fixation duration distribution in the L1-L2 and L2-L1 translation directions. First of all, outliers, in the form of very long average fixation durations, can be observed in both directions. Moreover, in the case of the L1-L2 direction, outliers in the form of extremely short results appear as well. It suggests a greater variability of the average fixation duration while translating from Polish into English. Secondly, in the case of both histograms, a classic normal distribution shape can be observed in the second part of the graph that begins with the mode. Both histograms seem to have a similar peak of average fixation duration close to 300 ms. However, the peak frequency is higher in the L2-L1 translation direction. Beginning with the modes, the frequencies of the results start to decrease. The measures of shape (Table 7) indicated that a leptokurtic distribution can be observed in both cases.

	L1-L2	**L2-L1**
kurtosis	5.3	8.4
skewness	1.5	2.2

Table 7. Average fixation duration - kurtosis and skewness.

The results of the inferential statistics show no normal distribution of the average fixation duration in both directions (L1-L2, $p=0.006$; L2-L1 $p<0.001$). Therefore, I decided to conduct the nonparametric Wilcoxon text. The results reached the level of statistical significance ($p=0.021$). It can be concluded that there are significant differences in the analysed variable. The average fixation duration is significantly longer in the L1-L2 translation direction.

5.1.3 Eye-key span [21]

Since I was interested in the immediate cognitive effort leading to a correct translation, I decided to exclude EKS appearing before all inaccurate or incomplete translations. Therefore, when the whole experimental procedure was finished, I extracted translations of collocations from the target texts. In the next step, four experts assessed translations of collocations for their accuracy. Then, those EKS values before the translations of collocation, which received 0 points from at least one of the experts, were

[21] This section appears as the article (Pietryga, manuscript under review)

excluded from further analysis as being inaccurate. It can be assumed that these EKS values do not reflect the actual effort because translations resulting from such EKS time lag were recognised as mistakes. Therefore, it did not lead to a successful translation solution.

I also decided to exclude from the analysis the EKS preceding translations of collocations that were modified during the translation process, regardless of whether such modifications were made in the drafting or the revision phase (Jakobsen 2002). Such an EKS also does not indicate the immediate effort that is the object of analysis in this study. Some cognitive processes may still occur in participants' brains, leading to further modifications of translations of collocations. There were also cases where participants managed to translate just one component of the collocation while the second word was added later in the translation process. Such EKS was also not included in the analysis. To sum up, in the case of the L2-L1 direction of translation, 170 out of 350 EKS values preceding translations of collocations were excluded from the analysis. In the case of the L1-L2 direction of translation, 172 out of 350 EKS values preceding translations of collocations were excluded from further analysis.

The mean values of the EKS indicated some difference between L1-L2 (M= 6591.3 ms; SD= 4969.3) and L2-L1 translation directions (M= 6250.3 ms SD= 3238.3). It can be observed that participants produced longer EKS during L1-L2 translation. A detailed distribution of the mean values of the EKS variable is presented in the histograms in Figure 5.

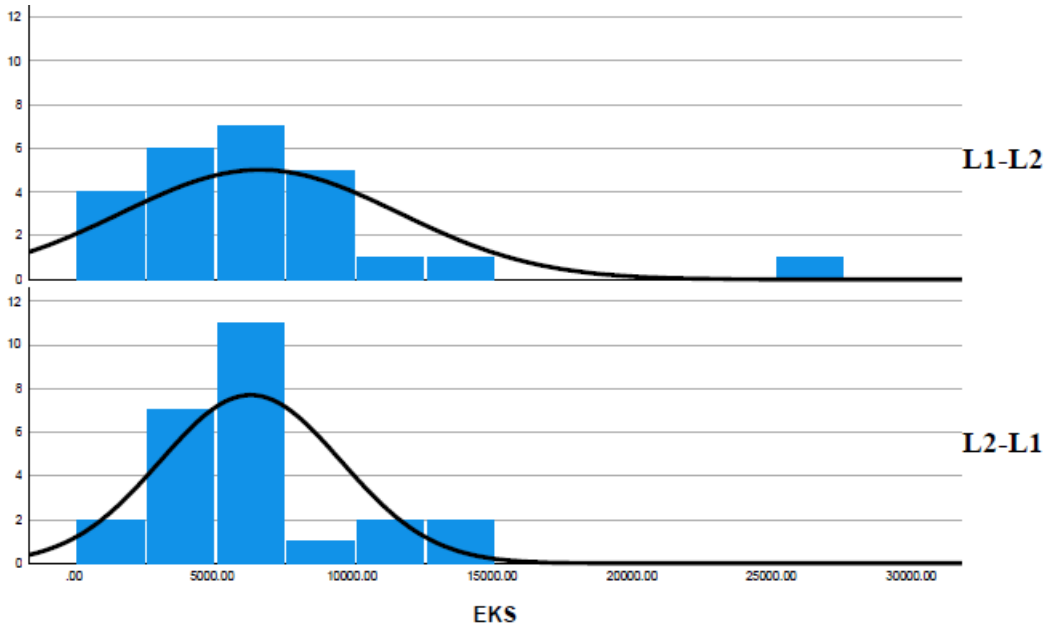


Figure 5. EKS – histograms.

In the case of the L2-L1 translation direction, quantitative differences exist among the results, with an observable peak of EKS longer than 5000 ms. There are relatively few extreme values and no outliers. The classic shape of a normal distribution can be observed in the first part of the graph. The extremely high results, higher than the mode, are flattened. In the case of the L1-L2 direction, the quantitative differences among the results are definitely smaller. The graph is more flattened. Another crucial difference is that in contrast to the L2-L1 direction, some outliers in the form of a very long EKS appeared in the L1-L2 translation direction. However, as Timarová et al. (2011: 129) pointed out, one should be cautious while analysing outliers, especially if they are substantially longer than the mean values. Taking into account memory constraints, such long EKS values may frequently mean data loss or refixations.

For further analysis, I decided to verify the distribution of the EKS in both directions. The results of the Shapiro-Wilk test revealed that no normal distribution of the EKS variable can be found both in the L2-L1 translation direction ($p < 0.001$) and in the reverse translation direction ($p = 0.014$). It allows me to compare kurtosis and skewness in both directions. The data summarised in Table 8 indicate a leptokurtic distribution in the L1-L2 direction, and the histogram is right skewed. In contrast, a distribution closer to the normal one can be observed in the L2-L1 translation direction.

	L1-L2	**L2-L1**
kurtosis	8.9	0.8
skewness	2.6	1.1

Table 8. EKS – kurtosis and skewness.

Inferential statistics was conducted to further verify the mean results. Since there was no normal distribution of the EKS variable, I decided to conduct the nonparametric Wilcoxon test. The test did not reach statistical significance ($p = 0.545$). It means that the difference between EKS in the L1-L2 and L2-L1 translation direction is statistically insignificant. Since no statistical significance was found, I was interested in whether the results of the statistical test are different when all the EKS values, taken from all 350 translations of 14 collocations in both directions of translation, are included, regardless of the points assigned by the experts, or the moment when the participants typed the translation of collocations. Therefore, I decided to include all EKS values obtained during the experiment to verify the results. Likewise, the mean values indicated higher cognitive

effort operationalised by longer EKS in the L1-L2 direction of translation (M=6408.4 ms; SD=3529.3) compared to the L2-L1 direction of translation (M= 6188.9 ms; SD= 3543.7). The results of the Wilcoxon test confirmed the previously obtained results. Once again, no statistical significance was found (p= 0.81).

5.1.4 Total task time

The mean length of the total task time variable indicated that participants' translation process lasted longer in the L1-L2 direction (M=1303357.5 ms; SD= 520075.8). It was around 22 minutes. In the reverse direction, L2-L1, the task took them, on average, 1041109 ms (SD= 299093.1), which is around 17 minutes. A detailed distribution of this variable is presented in Figure 6 below.

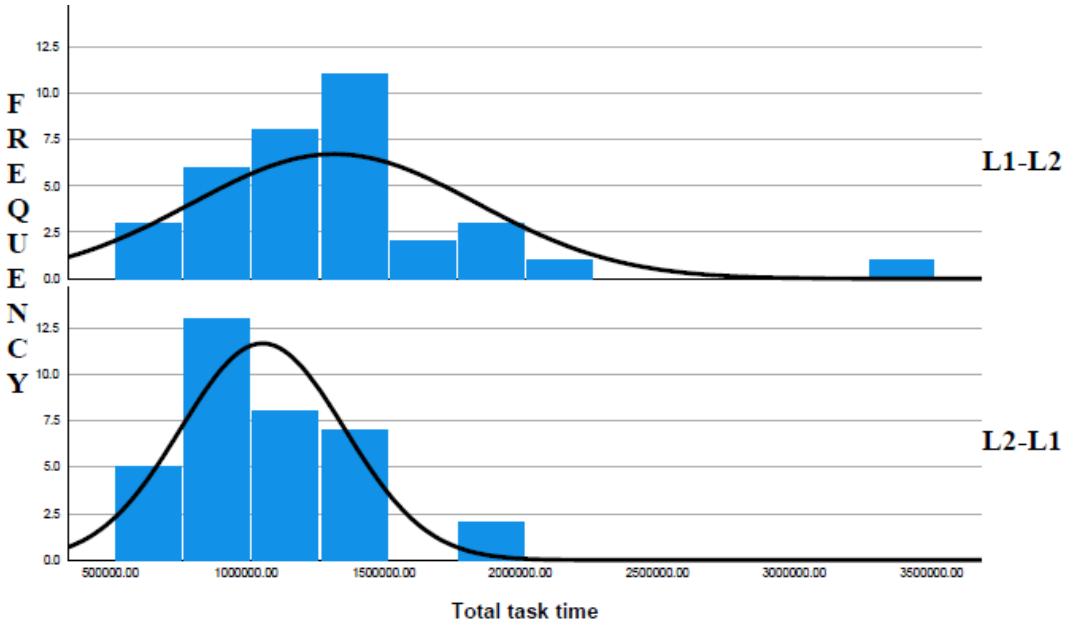


Figure 6. Total task time – histograms.

The L1-L2 histogram has the classic shape of normal distribution in its left part. However, results higher than the mode are less frequent. There are also some outliers in the form of an extremely long total task time that is close to 3500000 ms = 58 min. The results are less widely distributed in the L2-L1 translation direction, and there is one observable peak. Outliers can be found in this translation direction as well. Nevertheless, this time, they include results that are shorter than the outliers in the L1-L2 translation direction. The analysis of the kurtosis and skewness (Table 9) showed that the leptokurtic

distribution can be found in the L1-L2 translation direction. This histogram is also right-skewed. The results of the L2-L1 translation are close to 0.

	L1-L2	L2-L1
kurtosis	8.5	0.4
skewness	2.3	0.6

Table 9. Total task time - kurtosis and skewness.

The results of the Shapiro-Wilk test revealed that there is a normal distribution in the case of the L2-L1 total task time ($p=0.118$). However, no normal distribution is observed in the L1-L2 total task time ($p<0.001$). Thus, there is, in general, no normal distribution. Therefore, I decided to conduct a nonparametric Wilcoxon test. The results of the test indicated that there are significant differences between the two directions of translation ($p=0.003$). It can be concluded that participants spent significantly more time during L1-L2 translation, which means that this translation direction evokes significantly higher cognitive effort.

5.1.5 Number of pauses longer than 5 s

Descriptive statistics indicates differences in the total number of pauses longer than 5s. Participants produced in total 1064 pauses during their L1-L2 drafting phase and 848 pauses during the L2-L1 drafting phase. On average, participants paused 30.4 times during the L1-L2 translation (SD=16.4) and 24.2 times during the L2-L1 translation (SD=12.5). According to the total and mean results, more pauses occurred when participants translated the text from Polish to English.

The L1-L2 histogram has a more flattened shape (Figure 7 below). There are many average results that are close to the mode (Mo= 35) and mean (M= 24.4). There is also a more extended distribution of the variable. However, no single observable peak can be found. Two types of outliers expressing extremely low and extremely high results can be found in this translation direction. There is one observable peak (Mo= 16) in the case of the L2-L1 translation direction. This time, just one type of outliers can be observed. This is extremely high results that exceed 60 pauses longer than 5s.

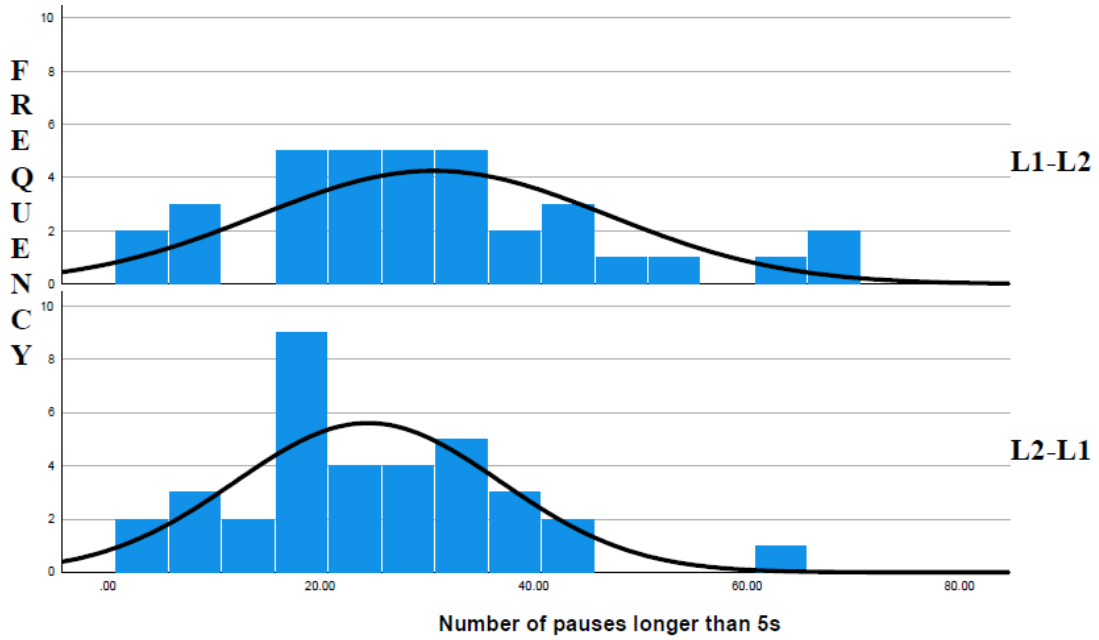


Figure 7. Number of pauses longer than 5s – histograms.

To further verify the results, I conducted the Shapiro-Wilk test, which indicated that this variable has a normal distribution in both directions (L1-L2, p=0.268 and L2-L1, p=0.128). Therefore, I was able to conduct the parametric paired t-test. The results of the tests revealed statistical differences between the mean values (p=0.009). It can be observed that cognitive effort operationalised in the number of pauses longer than 5s is significantly higher in the case of the L1-L2 translation. The results confirmed the assumptions of hypothesis one.

5.1.6 Mean length of pauses longer than 5s

Inferential statistics showed that the mean length of pauses longer than 5s was slightly longer during the L1-L2 translation. The mean pause length was around 10.4s (SD=2.1). In the case of the L2-L1 translation direction, participants, on average, produced pauses that lasted 9.7s (SD= 1.99). As can be observed, the differences between the mean values were marginal.

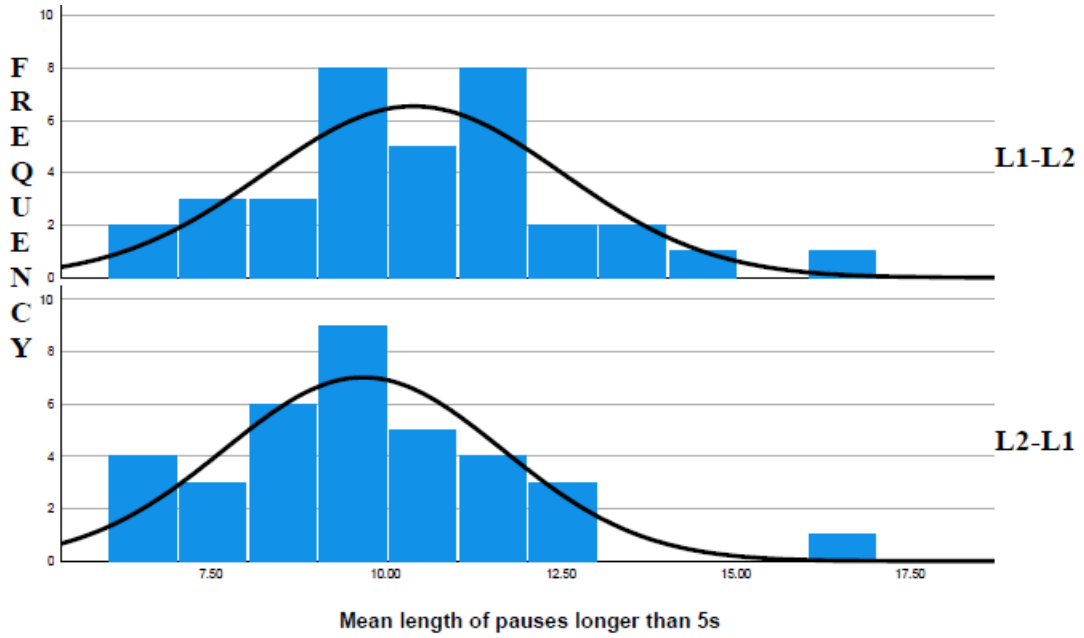


Figure 8. Mean length of pauses longer than 5s – histograms.

The histograms in Figure 8 showed differences in the distribution of the mean pause length in both translation directions. There is a more extended distribution of the results in the L1-L2 translation direction. Moreover, two identical peaks can be observed. The shape of the histogram is close to the classic one. In the L2-L1 translation direction, participants produced a higher number of very short and very long pauses. There is also an observable peak of the mean pause length. Interestingly, both histograms present similar outliers, including very long mean pause length.

The Shapiro-Wilk test showed a normal distribution of the variable in the L1-L2 translation direction (p= 0.415). No normal distribution was found in the case of the reverse translation direction. However, as p=0.045, the result is close to the threshold value of statistical significance. Therefore, I assumed that there is a normal distribution in the case of both directions. It allowed me to conduct a paired t-test. However, the results did not obtain the level of statistical significance (p=0.098). It can be concluded that there are no statistical differences between the mean pause length in the L1-L2 and L2-L1 translation direction. The hypothesis was not confirmed.

5.1.7 Weighted rating of the NASA-TLX

The results of the descriptive statistics showed marginal differences in the mean values of the NASA-TLX weighted rating. It was 50.7 (SD=12.3) in the L1-L2 translation

direction and 47 (SD=15.4) in the reverse translation direction. The mean values of the six rating scales (Figure 9) showed many similarities between the two translation directions. First of all, regardless of the translation direction, participants highly evaluated their performance. The evaluation of the L2-L1 translation is just slightly higher. Secondly, it can be observed that the ratings for mental demand and effort are similarly high in both translation directions. They significantly exceed 40 out of 100 points on a scale. The results were slightly higher in the L1-L2 translation direction in both cases. Participants decided that physical and temporal demands were rather low in both directions.

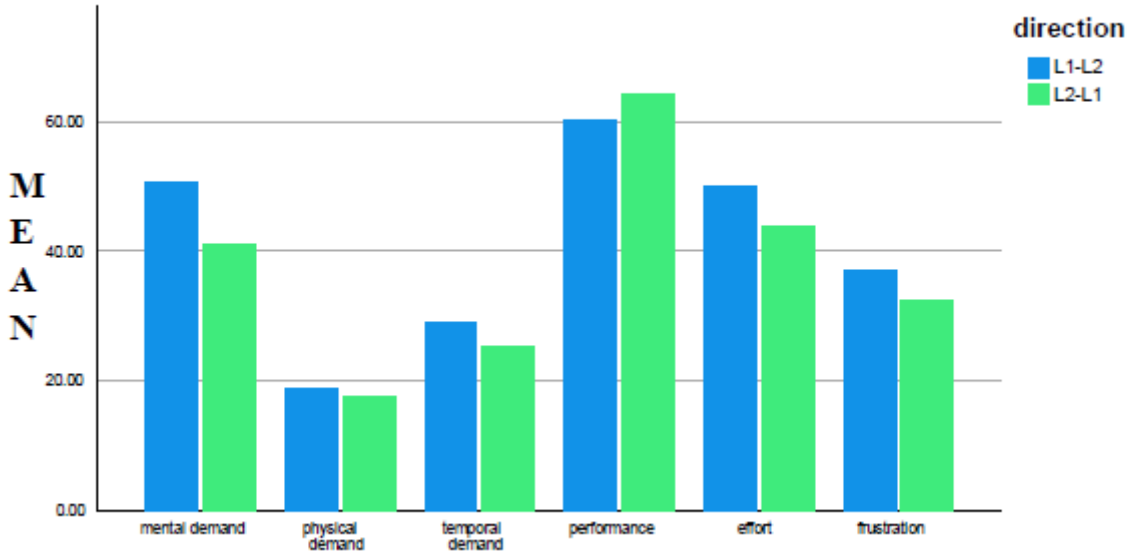


Figure 9. NASA-TLX - mean values of rating scales.

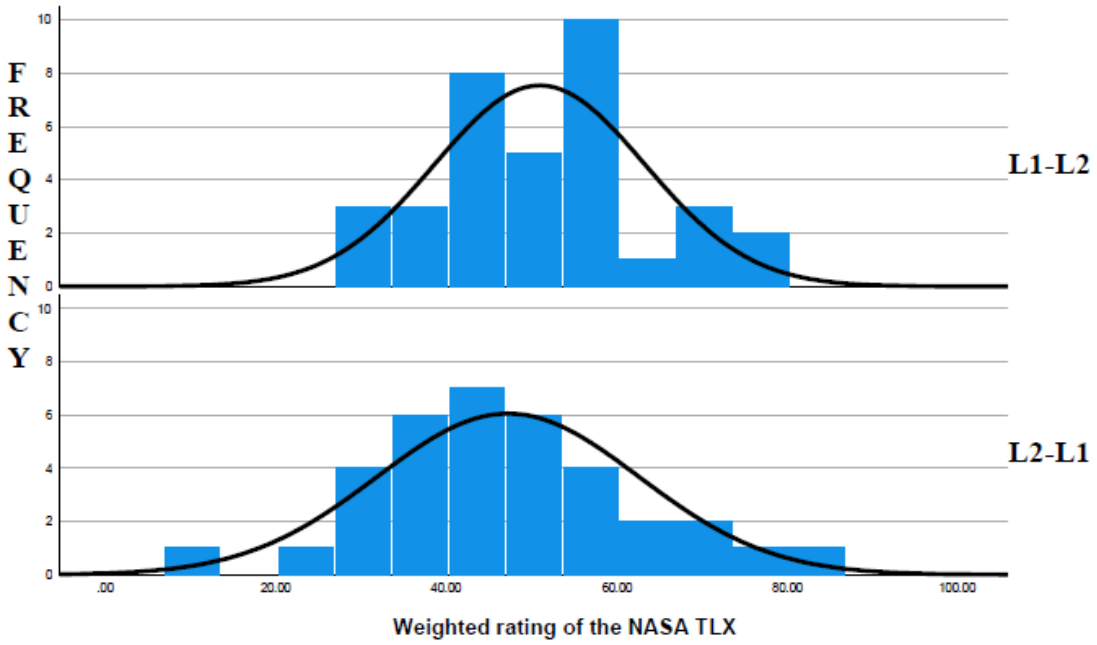


Figure 10. Weighted rating of the NASA-TLX – histograms.

The histograms presented in Figure 10 point out some differences in the distribution of the weighted rating of the NASA-TLX in the L1-L2 and L2-L1 translation directions. The lower histogram is more flattened, and there is a more extended distribution of the results. There is also a relatively small number of low and high results. Outliers, including very low results, can also be observed. In the case of the L1-L2 translation direction, there are no low results and no outliers. There is also a greater fluctuation among the results. Moreover, two peaks can be observed.

The results of the Shapiro-Wilk test showed that this variable has a normal distribution in both translation directions (L1-L2 p=0.07 and L2-L1 p=0.62). As a result, I was able to conduct a paired t-test. The results of the paired t-test indicated that there are significant differences between the weighted rating of the NASA-TLX (p=0.045). It can be concluded that cognitive effort operationalised in the NASA-TLX weighted rating is significantly higher in the L1-L2 translation direction, as predicted in the hypothesis. However, it is worth noting that the result of the paired t-test is close to the threshold value of p=0.05. Therefore, the results may be a trend rather than regularity.

5.1.8 Summary of Section 5.1

Based on the results obtained through inferential statistics, Wilcoxon and paired t-tests, it can be observed that Hypothesis 1, assuming that cognitive effort, operationalised by total gaze time, average fixation duration, total task time, the number

of pauses longer than 5s, mean pause length of pauses longer than 5s, EKS and the weighted rating of the NASA-TLX, is higher in L1-L2 direction, was not fully confirmed. It was corroborated for four out of seven variables. These were total gaze time, average fixation duration, total task time, and the number of pauses longer than 5s. In the case of two variables, EKS and the mean length of pauses longer than 5s, the results did not obtain statistical significance; in the case of the weighted rating of the NASA-TLX, the paired t-test obtained a level of statistical significance (p= 0.045). It means that the result is close to the threshold value of p= 0.05. Thus, it can be treated as a trend rather than a solid result.

5.2 Hypothesis 2

In Hypothesis 2, I assumed that the orientation phase is shorter in the L1-L2 direction. Descriptive statistics indicate that in L1-L2 as well as in L2-L1 translation directions orientation phases were rather short. Contrary to predictions, the orientation phase, on average, lasted longer in the L1-L2 translation (M=37899.6ms= 0.63min; SD=35035) than in L2-L1 (M= 29771.17ms= 0.5min; SD=30125.2). These results reveal that the hypothesis cannot be confirmed. In contrast to the assumptions, it is the L2-L1 orientation that was shorter. The histograms in Figure 11 present in detail differences in the distribution of the variable in both translation directions.

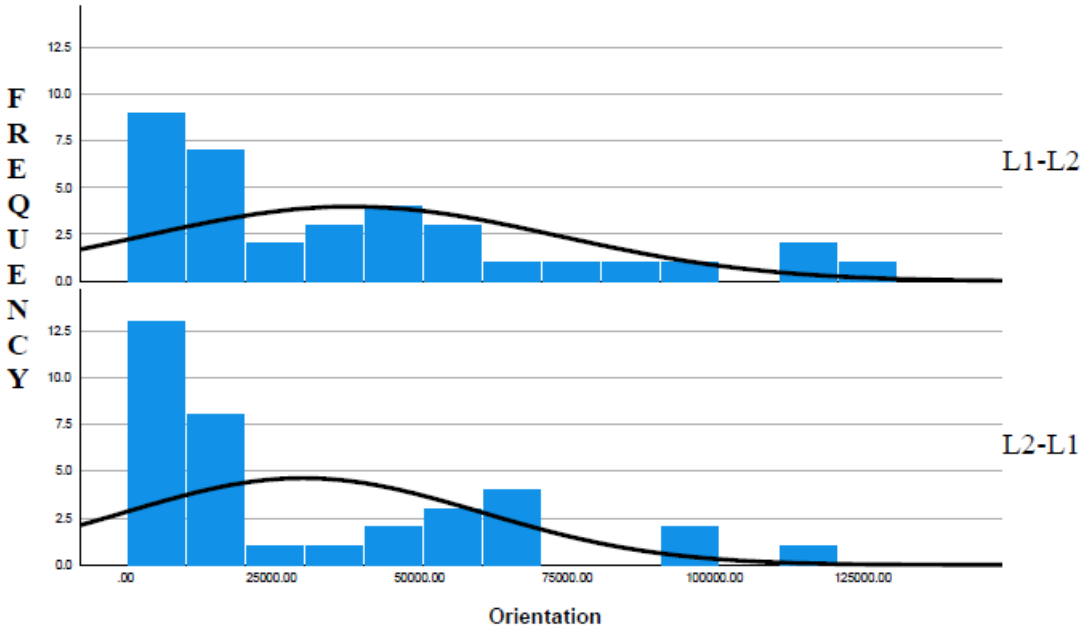


Figure 11. Orientation phase length – histograms.

Histograms in Figure 11. show some differences and similarities in the distribution of the variable. First of all, there is a high number of extremely low results in both directions. This means that both when performing L1-L2 and L2-L1 translations, many participants read only a few words of the source text, which resulted in a very short orientation phase. Secondly, in both cases, the histograms have a rather flattened shape. Finally, some outliers, including extremely long orientation phases, can be observed. However, it is worth noting that the outliers are longer in the L1-L2 translation direction. The main difference between the two histograms can be observed in the middle part, near the median. There is a greater frequency of results that are close to the median in the L1-L2 translation direction (Mdn.= 24074ms) than in the L2-L1 translation direction (Mdn. = 11856ms). A more extended distribution of the variable can also be observed in this direction. The analysis of kurtosis and skewness (Table 10) shows that both histograms are right-skewed. The kurtosis is close to 0 in both translation directions.

	L1-L2	**L2-L1**
kurtosis	0.8	0.5
skewness	1.2	1.2

Table 10. Orientation - kurtosis and skewness.

Although the results of the descriptive statistics excluded the possibility of confirming Hypothesis 2, I still wished to verify whether the difference between the duration of the orientation phases reaches the level of statistical significance. First, I analysed the distribution of the variable. The lack of normal distribution can be found both in the L1-L2 translation direction ($p<0.001$) and in the L2-L1 translation direction ($p<0.001$). As there is no normal distribution in both translation directions, I decided to conduct the non-parametric Wilcoxon test. The test results did not reach the level of statistical significance ($p= 0.171$). The difference in the duration of the orientation phase in the L1-L2 and L2-L1 translation directions is not statistically significant. The hypothesis was not confirmed.

5.3 Hypothesis 3

In Hypothesis 3, I assumed that the drafting phase and revision phase are longer in the L1-L2 direction.

5.3.1 Drafting phase

According to descriptive statistics, the drafting phase lasted, on average, longer in the L1-L2 translation direction than in the L2-L1 translation direction. The mean duration of the L1-L2 drafting phase was 952537.4 ms (SD= 310027.6), which is around 16 minutes. The L2-L1 average drafting phase was slightly shorter, as it lasted, on average, 811197.4 ms (SD= 234147.6), which is around 14 minutes. The differences in the distribution of the drafting phases are presented in histograms in Figure 12 below.

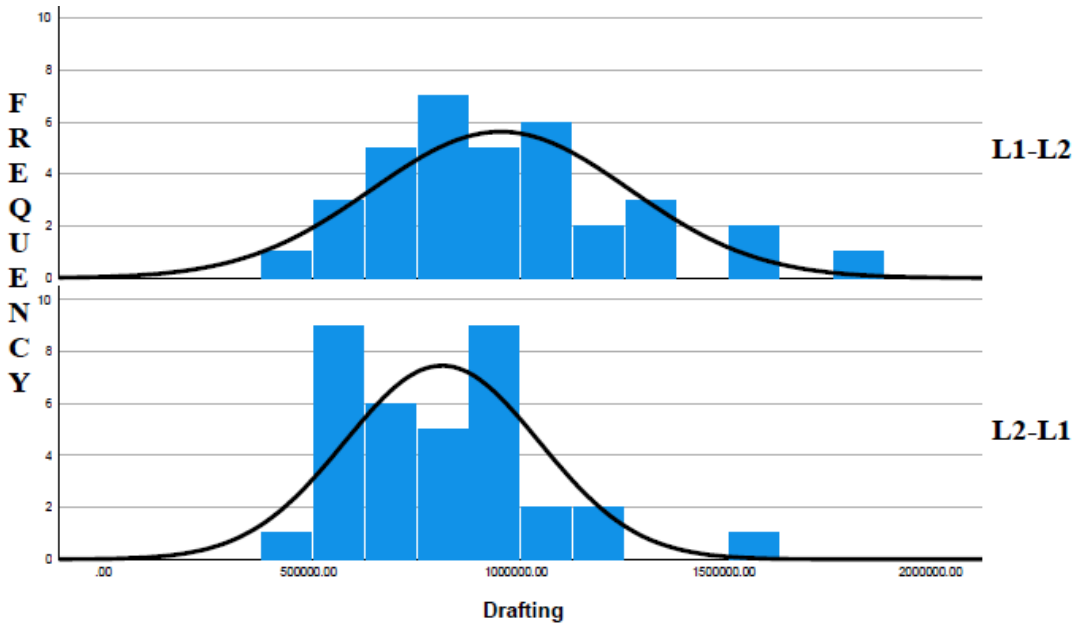


Figure 12. Drafting phase length – histograms.

The L1-L2 histogram shows a classic normal distribution shape in the graph's left part. Two groups of outliers, including extremely long drafting times, can be observed in this translation direction. There is also a more extended distribution of the variable in this translation direction. In the case of the L2-L1 translation direction, there are two peaks with a similar frequency. One group of outliers can be found in the L2-L1 translation direction. These are extremely long drafting phases, which, however, are shorter than L1-L2 outliers.

The results of the Shapiro-Wilk test revealed that there is a normal distribution of the variable in both translation directions (L1-L2 p= 0.5; L2-L1 p= 0.1). Therefore, I decided to conduct the paired t-test. The results of the t-test indicated that there is a statistically significant difference between the mean values (p= 0.01). It can be stated drafting phase is significantly longer in the L1-L2 translation direction.

5.3.2 Revision phase

The mean results of the descriptive statistics seem to confirm the adopted hypothesis, as the L1-L2 revision phase was, on average, longer than the L2-L1 revision phase. The mean duration of the revision phase in the L1-L2 translation direction was 312920.5 ms (SD= 315406.5), which is around 5 minutes. In the case of the L2-L1 revision, it was 200140.4 ms (SD= 154481.8), which is around 3 minutes.

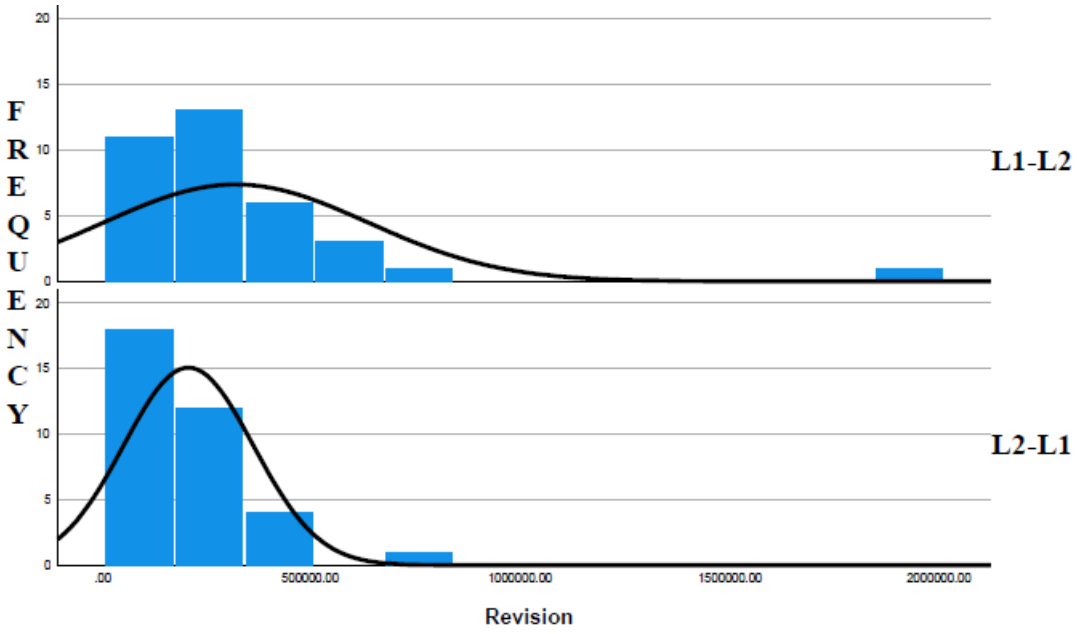


Figure 13. Revision phase length – histograms.

As shown in the histograms in Figure 13, there are many extremely short revision phases in both translation directions. This means that regardless of the translation direction, participants did not tend to devote much time to revise their translations. However, the L1-L2 histogram is more flattened. Moreover, participants produced some outliers in the form of extremely long revision phases in the L1-L2 translation direction. The mode in this direction is 23037ms (0.4 min), which is related to almost 15 observations. In the case of the L2-L1, the number of observations belonging to mode (Mo= 8873 ms= 0.15min) is higher, as there are almost 20 observations. It is worth noting that the outliers in this direction are definitely shorter than in the L1-L2 translation direction. Both histograms have the classic normal distribution shape in their second part, presenting longer results. The analysis of kurtosis and skewness (Table 11) indicated that the L2-L1 and L1-L2 revision phases are of leptokurtic and right-skewed distribution.

	L1-L2	L2-L1
kurtosis	18.5	7.2
skewness	3.8	2.1

Table 11. Revision - kurtosis and skewness.

The results of the Shapiro-Wilk test indicated that there is no normal distribution in the case of both translation directions, as $p<0.001$. Thus, I decided to conduct the nonparametric Wilcoxon test. According to the results of the Wilcoxon test, the difference between the mean values of the revision phase in both translation directions is statistically significant ($p=0.007$). The L1-L2 revision phase is significantly longer than the L2-L1 revision phase.

5.3.3 Summary of Section 5.3

The results of the inferential statistics, paired t-test and Wilcoxon test, confirmed the adopted Hypothesis 3, stating that the drafting and revision phases are longer in the L1-L2 direction. Both variables, the length of the drafting phase and the length of the revision phase, reached the level of statistical significance.

5.4 Hypothesis 4

In Hypothesis 4, I assumed that the content and number of retrospective reports would indicate higher cognitive effort in translating collocations in the L1-L2 direction. Firstly, I analysed the verbosity of all the retrospective reports. I employed three measures of verbosity introduced by Gumul (2020b) and used in Gumul and Herring (2022, 2023). These are: "the number of comments[22] verbalised by each participant, the number of words per protocol, the duration of the entire protocol" (Gumul 2020, p. 156). The retrospection after the L1-L2 translation lasted, on average, 5:28 min (SD=3:42). The average retrospection after the L2-L1 translation was slightly shorter, and it lasted 5:09 min (SD=3:35). The measurement of the number of words per protocol indicated that participants produced slightly more words during the retrospection after the L2-L1 translation (M= 299.3, SD= 237), compared to the retrospection after the L1-L2 translation (M =293, SD= 193.5). In total, participants verbalised 329 self-reports (M= 9.4, SD=4.3) related to the L1-L2 translation, of which there were 113 self-reports of

[22] In the later part of the thesis, I operationalise the number of comments as the number of self-reports.

cognitive effort related to translating collocations. During retrospection after the L2-L1 translation, they verbalised 261 comments, of which there were only 54 self-reports of cognitive effort related to translating collocations. These numbers suggest that participants may experience higher cognitive effort during the L1-L2 translation process.

In the next step, I categorised the self-reports of cognitive effort related to translating collocations into three levels[23]:

- Level 1- participants describe the process of translating collocations, for example, PE32/T/6: "I decided to literally translate '**skupia uwagę** mieszkańców wyspy' as 'gathers the attention', it seems to me that this idiom works in English as well;'"[24]
- Level 2 – participants describe the process of changing a decision or introducing modifications, for example, PE04/T/4: "at the beginning, I translated the fourth sentence in the following way: '**lost his smell** and he **is blind'**. I wanted it to look nicely, so I changed it into 'his smell and sight;'"
- Level 3 – participants explicitly state that there was a problem during the translation process, for example, PE32/T/11: "I had a problem with '**odcisk łapy'**, and I decided to use explicitation 'a mark in the sand.'"

Out of 113 L1-L2 self-reports of cognitive effort related to translating collocations, 22 self-reports (M= 0.6, SD=1) were classified as Level 1, 2 self-reports (M= 0.05, SD=0.3) were classified as Level 2, and 89 self-reports (M=2.5, SD=1.9) were classified as Level 3. The high number of self-reports belonging to Level 3 suggests that the participants not only experienced serious problems while translating collocations in this translation direction but were also able to explicitly report these issues. In the reverse translation direction, out of 54 self-reports of cognitive effort related to translating collocations, 30 self-reports (M=0.8, SD=1.1) were classified as Level 1, 7 self-reports (M= 0.2, SD=0.4) were classified as Level 2, and 17 self-reports (M=0.5, SD= 0.9) were classified as Level 3. This time, participants produced more self-reports describing the decision-making process rather than pointing out explicit cognitive effort and problem-solving strategies.

[23] The discussed levels of cognitive effort related to translating collocations were developed by me based on the data from the retrospective verbal reports.

[24] The quotations of self-reports of cognitive effort are derived from retrospective reports verbalised by participants of this study. Following the study by Gumul (2017), the quoted examples are marked with the number of the participant, the abbreviation of the source text, where 'D' stands for 'dog' and 'T' stands for 'tortoise' and the number of the self-report. I also marked in bold the analysed collocation.

Interestingly, in both cases, participants produced the smallest number of self-reports in the category Level 2. The results are summarised in Table 12 below.

Translation direction	Total number of comments referring to translation of collocations	Level 1 (number of comments)	Level 2 (number of comments)	Level 3 (number of comments)
L1-L2	113	22 (M= 0.6, SD=1)	2 (M= 0.05 SD=0.3)	89 (M=2.5 SD=1.9)
L2-L1	54	30 (M=0.8, SD=1.1	7 (M=0.2, SD=0.4)	13 (M=0.5, SD= 0.9)

Table 12. Cognitive effort related to translating collocations - number of self-reports.

There are some observable differences between the translation directions. On the one hand, most L1-L2 self-reports indicate explicit problems during translating collocations. On the other hand, in the L2-L1 direction, most of the self-reports were related to decision-making. Descriptive statistics, the content and the number of retrospective reports indicate that the L1-L2 translation evoked increased cognitive effort. To verify this hypothesis, I decided to conduct inferential statistics. I begin with the total number of self-reports of cognitive effort related to translating collocations, and then I move on to the Levels of cognitive effort.

Both translation directions are characterised by the lack of normal distribution of the total number of self-reports of cognitive effort related to translating collocations (L1-L2 $p=0.03$, L2-L1 $p<0.001$). Since there is no normal distribution, I decided to conduct a non-parametric Wilcoxon test. The results of the test reached the level of statistical significance ($p<0.001$). It means that participants produced significantly more self-reports of cognitive effort related to translating collocations in the L1-L2 translation direction.

Likewise, no normal distribution was found in the number the self-reports classified as Level 1 (L1-L2 $p<0.001$, L2-L1 $p<0.001$). Therefore, I decided to conduct a non-parametric Wilcoxon test. Unfortunately, the results did not reach the level of statistical significance ($p=0.138$). The difference in the number of self-reports describing

the effort during translating collocations that were classified as Level 1 is not statistically significant.

No normal distribution was also found in the case of the self-reports that were classified as Level 2 (L1-L2 $p<0.001$, L2-L1 $p<0.001$). I conducted a Wilcoxon test; however, no statistical significance was found ($p=0.1$). There is no significant difference in the number of self-reports of cognitive effort related to translating collocations classified as Level 2 in both translation directions.

A normal distribution was found in the case of self-reports that were classified as Level 3 in the L1-L2 translation direction ($p=0.026$). However, no normal distribution was observed in the same category in the reverse translation direction ($p<0.001$). Therefore, I assumed that there is, in general, no normal distribution, and I decided to conduct a non-parametric Wilcoxon test. The results of the test revealed statistical significance ($p<0.001$). It means that the number of self-reports of cognitive effort related to translating collocations classified as Level 3 is significantly higher in the L1-L2 translation direction than in the L2-L1 translation direction.

The participants also verbalised a cognitive effort that did not result from translating collocations. They frequently reported problems with translating particular words. For example,

> PE38/D/3: In sentence number 5, I was not sure whether ‘hemisphere’ means ‘półkule’, but it made sense to me that in Polish, we say ‘prawa półkula’, ‘lewa półkula’, so I left it like that.

Participants produced 75 such self-reports during the retrospection after the L1-L2 translation ($M= 2.1$, $SD=1.9$) and 65 self-reports during the retrospection after the L2-L1 translation ($M=1.9$, $SD=1.7$). No normal distribution was observed in both translation directions (L1-L2 $p<0.001$; L2-L1 $p<0.001$). Unfortunately, the results of the Wilcoxon test did not reach the level of statistical significance ($p= 0.1$).

Interestingly, in both translation directions, participants also verbalised a lack of cognitive effort during the translation process. For example, one participant stated,

> PE22/T/4: I did not have any problems with sentence number 4. Likewise, with sentence number three.

There were 12 self-reports verbalising a lack of cognitive effort related to the L1-L2 translation direction (M= 0.3, SD= 0.7) and 14 such self-reports related to the L2-L1 translation direction (M= 0.4, SD= 0.7). No normal distribution was found in both translation directions (L1-L2 $p<0.001$; L2-L1 $p<0.001$). The results of the Wilcoxon test did not reach the level of statistical significance ($p=0.527$). Therefore, it can be stated that there are no significant differences in the number of self-reports reporting a lack of cognitive effort.

5.4.1 Summary of Section 5.4

The descriptive statistics results indicate that participants generally verbalised more self-reports of cognitive effort related to the L1-L2 translation of collocations. Verbalisations belonging to all three Levels of cognitive effort can be found in both translation directions. In the case of L1-L2 translation direction, the majority of self-reports of cognitive effort related to translating collocations were classified as Level 3. In the reverse translation direction, the largest number of self-reports of cognitive effort related to translating collocations was classified as Level 1. Nevertheless, the inferential statistics did not fully confirm the adopted hypothesis. The results of the Wilcoxon tests reached the level of statistical significance only for the general number of self-reports of cognitive effort related to translating collocations and for Level 3 of self-reports. No statistical significance was found for the number of self-reports classified as Level 1 and 2. To sum up, Hypothesis 4, assuming that the content and number of retrospective reports would indicate higher cognitive effort in translating collocations in the L1-L2 direction, cannot be fully confirmed. Likewise, no statistical significance was found for the categories of cognitive effort not resulting from translating collocations and lack of cognitive effort.

5.5 Hypothesis 5

In Hypothesis 5, I assumed that translation accuracy will be higher in the L2-L1 direction. Each expert assessing translation accuracy could assign from 0 to 2 points to the translation of a given collocation. Collocations were assessed by three experts in each direction. It means that the highest number of points a participant could receive for the translation of a collocation was 6.

All L1-L2 translations of collocations received in total 1767 points (M=50.5, SD=7). Experts assigned 0 points 202 times, 1 point 349 times, and 2 points 709 times.

Based on the percentage of scores presented in Figure 14 below, it can be observed that experts assessed translations of collocations as highly accurate in more than half of the cases (56.27%). According to the experts, 16.03% of translations of collocations were completely inaccurate. In 27.7% of cases, experts decided that translations of collocations were partially accurate.

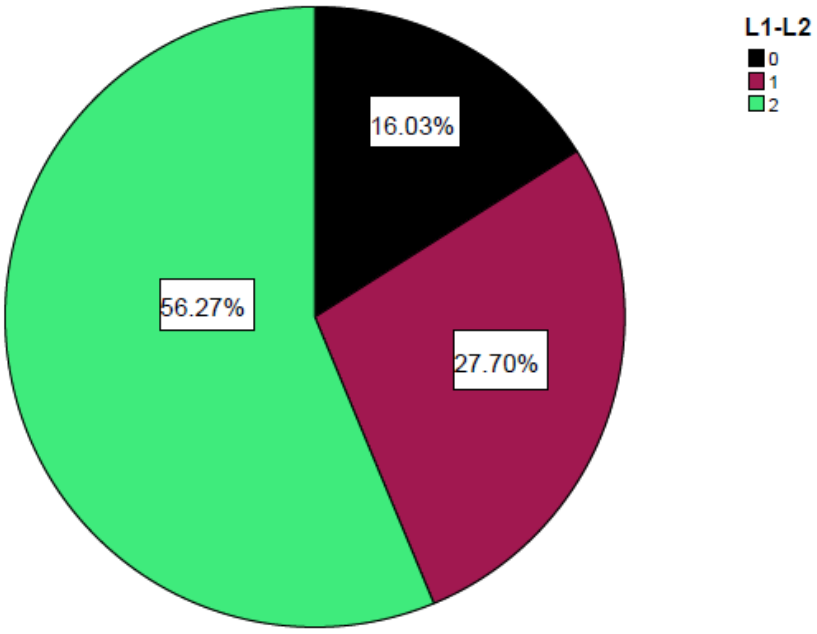


Figure 14. L1-L2 translation accuracy.

Participants had the least problem with the collocation “rozpiera go energia’, as it received a total number of 201 points. In fact, only 7 out of 35 participants did not receive 6 points for the translation of this collocation. However, none of the participants obtained 0 points from any of the experts. ‘Pamiątkowy certyfikat’ appeared to be the most problematic collocation in this direction. Participants received in total only 99 points for the translation of this collocation. Four participants received in total 0 points for their translation, and only 2 out of 35 participants obtained the maximum 6 points.

L2-L1 translations of collocations received in total 1975 points (M=56.4, SD=5), which is over 200 points more than in the reverse translation direction. 0 points were assigned 170 times, 1 point was assigned 195 times, and 2 points were assigned 890 times. This time, experts decided that translations of collocations have a very high accuracy in almost three/fourth of cases (70.63%), which is more than in the reverse translation direction. The number of collocations assessed as partially accurate (15.48%) and completely inaccurate (13.89%) has diminished compared to the L1-L2 translation direction.

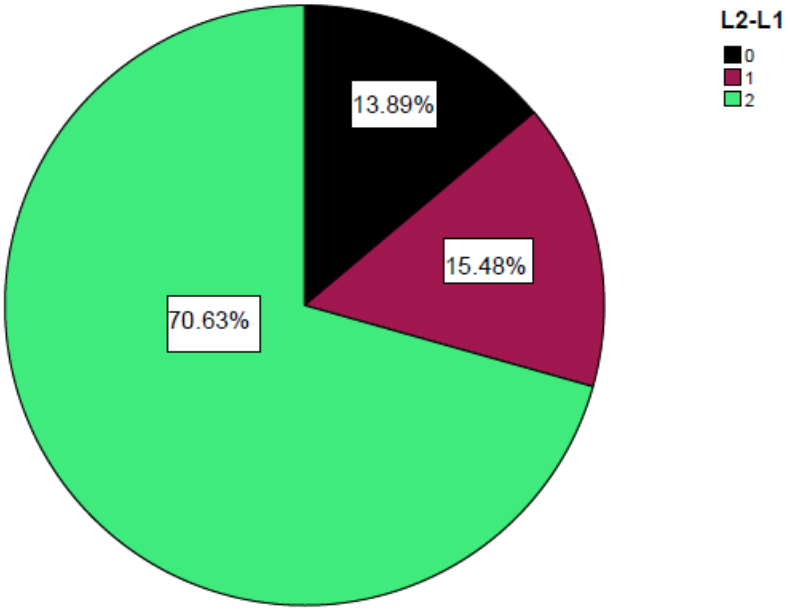


Figure 15. L2-L1 translation accuracy.

The collocation ‘common ancestor’ received the highest number of points (204), and only two out of 35 participants did not obtain the maximum number of points. The collocation that received the smallest number of points was ‘to analyse intonation.’ It received, in total, 140 points. Nevertheless, this time, none of the participants received in total 0 points for their translation of this collocation. Based on the descriptive statistics, it could be observed that the accuracy of the L2-L1 translations was higher. A detailed distribution of the mean translation accuracy is presented in Figure 16 below.

The L1-L2 histogram (Figure 16) seems to be more flattened, with an observable peak of 4.08 points. Interestingly, outliers, including a very small number of points, can be found solely in this translation direction. There is also a separate group of results between 3 and 4 points with an observable peak. The results from the L1-L2 translation direction are characterised by a more extended distribution and greater variability than the ones from the L2-L1 translation direction. Following Hypothesis 5, the results of the L1-L2 translation accuracy are relatively high. In fact, mean translation accuracy begins with more than 3 points, which is an outlier. The majority of the results are gathered between 4 and 6 points.

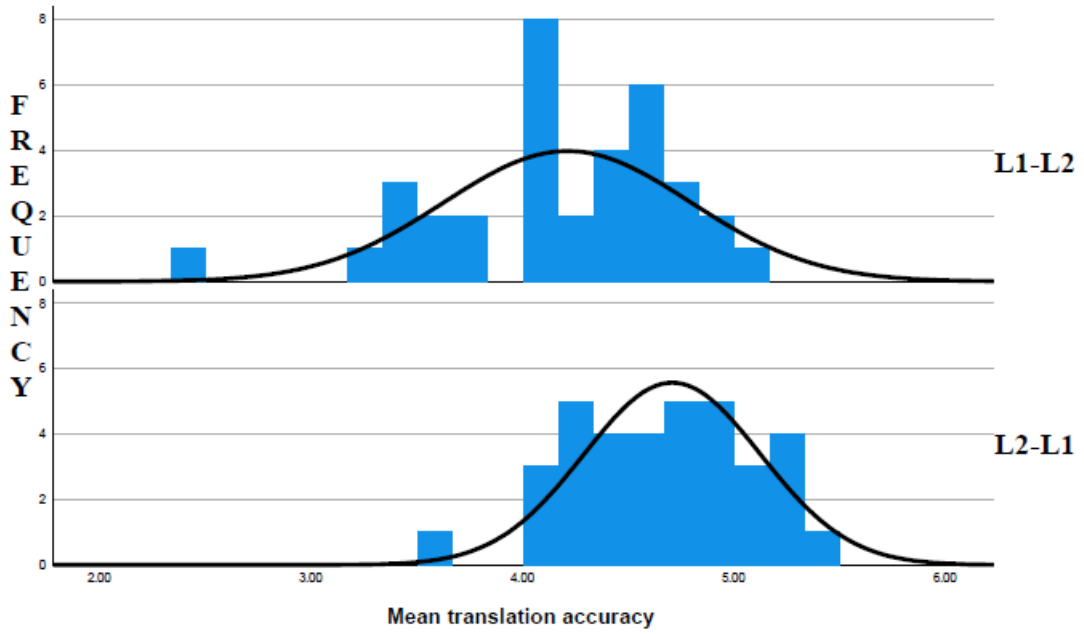


Figure 16. Mean translation accuracy – histograms.

To further verify the results, I first analysed their distribution. A normal distribution was found in both translation directions (L1-L2 p=0.077 and L2-L1 p=0.622); therefore, I conducted a paired t-test. The results reached the level of statistical significance (p<0.001). This means that the accuracy of the L2-L1 translation is significantly greater than the accuracy of the L1-L2 translation. Hypothesis 5 was fully corroborated.

5.6 Linear regression

I was also curious whether the analysed variables predict each other. Therefore, I decided to conduct a linear regression analysis. It begins with the analysis of the r Pearson coefficient since the linear regression can be conducted only when the coefficient reaches the significance level. First, I am going to describe linear regression conducted for the L1-L2 translation direction and next for the L2-L1 translation direction.

5.6.1 L1-L2 linear regression

Unfortunately, for some variables in the L1-L2 translation direction, the r Pearson coefficient was close to zero, meaning that the correlation was too weak, and it was impossible to conduct the linear regression analysis. EKS, average fixation duration,

duration of the orientation phase, translation accuracy and the weighted rating of the NASA-TLX are not good predictors of any of the variables in this translation direction.

The linear regression analysis could be conducted for the total gaze time variable in the L1-L2 translation direction since it correlates with the following variables in the same direction: number of pauses longer than 5s (r Pearson= 0.8) and mean length of pauses longer than 5s (r Pearson= 0.4).

I conducted a linear regression analysis for the total gaze time and the number of pauses longer than 5s. The proposed regression model appears to be a good fit for the data $F(1,23)=41.883$ $p<0.001$. Based on the regression coefficients, it can be stated that the number of pauses is strongly and positively related to the total gaze time (beta= 0.803, $p<0.001$). The more pauses longer than 5s a participant produces, the longer the total gaze time is. The tested model explains as much as 63% of total gaze time variability.

I was also able to conduct a linear regression analysis for the L1-L2 total gaze time and the mean length of pauses longer than 5s in the same translation direction. The proposed model is a good fit for the data $F(1,23)=5.394$, $p=0.029$. The regression coefficients show that the mean length of pauses longer than 5s is moderately and positively related to the total gaze time (beta= 0.436, $p=0.029$). It means that the longer the mean pause length, the longer the total gaze time. The tested model explains 16% of the total gaze time variability.

The variable of the mean length of pauses longer than 5s in the L1-L2 translation direction correlates moderately and positively with the number of pauses longer than 5s in the same translation direction (r Pearson= 0.4). Therefore, a linear regression analysis could be conducted. The proposed regression model can be assumed to be a good fit for the data since $F(1,33)= 6.582$, $p=0.015$. The regression coefficients indicate that the number of pauses longer than 5s is strongly and positively related to the mean length of pauses longer than 5s (beta= 0.4, $p<0.015$). It can be assumed that the higher the number of pauses longer than 5s, the longer the mean length of pauses. The tested model explains 14.1% of the mean length of pause variability.

The linear regression analysis can also be conducted for the duration of the revision phase since it is positively correlated with the duration of the drafting phase (r Pearson= 0.4). The proposed model seems to be a good fit for the data $F(1,33)=5.576$, $p=0.024$. The coefficient indicates that the duration of the drafting phase is moderately and positively related to the duration of the revision phase (beta= 0.38, $p=0.024$). It means that the longer the duration of the participant's drafting phase, the longer the duration of

their revision phase. The tested model explains 11.9% of the variability of the duration of the revision phase.

The duration of the drafting phase was positively correlated with the number of pauses longer than 5s (r Pearson= 0.4). The proposed model was a good fit for the data $F(1,33)=6.044$; $p=0.019$. Based on the regression coefficient, it can be stated that the number of pauses longer than 5s is moderately and positively related to the duration of the drafting phase (beta= 0.393 $p=0.019$). Thus, the more pauses longer than 5s the participant produces, the longer their drafting phase is. The tested model explains 12.9% of the drafting phase variability.

The duration of the drafting phase is also positively correlated with the mean length of pauses longer than 5s (r Pearson= 0.4). The proposed model was a good fit for the data $F(1,33)=5.079$, $p=0.031$. The regression coefficients indicate that the mean length of pauses longer than 5s is moderately and positively related to the duration of the drafting phase (beta= 0.365, $p=0.031$). It means that the longer the mean pause length, the longer the drafting phase. The tested model explains 10.7% of the drafting phase variability.

The L1-L2 total task time variable is positively correlated with three variables in the same translation direction: duration of the drafting phase (r Pearson= 0.8), duration of the revision phase (r Pearson= 0.8), and the mean length of pauses longer than 5s (r Pearson= 0.4). I started with a linear regression analysis for the duration of the drafting phase. The proposed regression model is a good fit for the data $F(1,33)= 71.737$, $p<0.001$. Based on the regression coefficients, it can be stated that the duration of the drafting phase is strongly and positively correlated with the total task time (beta= 0.828, $p<0.001$). It means that the longer the drafting phase, the longer the total task time. The tested model explains as much as 67.5% of the total task time variability.

I was also able to conduct a linear regression analysis for the duration of the revision phase and total task time. The proposed regression model is a good fit for the data $F(1,33)= 73.112$, $p<0.001$. The regression coefficients indicate that the duration of the revision phase is strongly and positively related to the total task time (beta= 0.83, $p<0.001$). Therefore, the longer the participant revises the translated text, the longer their total task time is. The proposed model explains as much as 68% of the total task time variability.

I also conducted a linear regression analysis for the total task time and the mean length of pauses longer than 5s. The proposed regression model appears to be a good fit for the data $F(1,33)= 6.582$ $p=0.015$. Based on the regression coefficient, it can be stated

that the mean length of pauses longer than 5s is strongly and positively related to the total task time (beta= 0.408, p=0.015). It means that the participant whose mean length of pause longer than 5s in the L1-L2 translation direction was high also produced a longer total task time. The tested model explains 14.1% variability of the total task time.

The general number of self-reports correlates strongly and positively solely with the number of self-reports of cognitive effort related to translating collocations (r Pearson= 0.8). The proposed regression model appears to be a good fit for the data $F(1,33)= 91.159$, $p<0.001$. The regression coefficient indicates that the number of self-reports of cognitive effort related to translating collocations is strongly and positively related to the general number of self-reports (beta=0.857, $p<0.001$). It can be assumed that the more self-reports of cognitive effort related to translating collocations the participant verbalises, the more self-reports there are. The tested model explains as much as 72.6% of the variability of the overall number of retrospective reports.

5.6.2 L2-L1 linear regression

In the L2-L1 translation direction, the correlation of some variables was close to zero as well, and thus, they are not good predictors of any other variable. As a result, it was not possible to conduct a linear regression analysis for the EKS, average fixation duration, the duration of the orientation phase, the weighted rating of the NASA-TLX and translation accuracy.

I was able to conduct a linear regression analysis for the L2-L1 total gaze time because it significantly and positively correlates with the number of pauses longer than 5s in the same translation direction (r Pearson= 0.6), the mean length of pauses longer than 5s (r Pearson= 0.5), the length of the revision phase (r Pearson= 0.6) and the drafting phase (r Pearson= 0.6).

The proposed regression model for the total gaze time and the number of pauses longer than 5s appears to be a good fit for the data $F(1,23)=18.842$, $p<0.001$. The regression coefficients indicate that the number of pauses is strongly and positively related to the total gaze time (beta= 0.667, $p<0.001$). It means that the more pauses longer than 5s participants produce, the longer their total gaze time is. The tested model explains 42.1% of the total gaze time variability.

A linear regression analysis could also be conducted for the total gaze time and the mean pause length. The proposed model is also a good fit for the data $F(1,23)=6.802$, $p=0.016$. Based on the regression coefficients, it can be stated that the mean pause length

is strongly and positively related to the total gaze time (beta= 0.478, p=0.016). Therefore, the longer the L2-L1 mean pause length, the longer the total gaze time. The tested model explains 20% of the total gaze time variability.

I was also able to conduct a linear regression analysis for the variables of the total gaze time and the duration of the revision phase in the L2-L1 translation direction since they strongly correlate (r Pearson= 0.6). The proposed regression model appears to be a good fit for the data, F(1,23)=12.591, p=0.002. The coefficients indicate that the duration of the revision phase is moderately and positively related to the total gaze time (beta= 0.595, p=0.002). It can be assumed that the longer the participant's revision is, the longer their gaze time is. The tested model explains 32.6% of the total gaze time variability.

Since there was a strong and positive correlation between the total gaze time and the duration of the drafting phase, a linear regression analysis for the two variables could be conducted. It appeared that the proposed regression model was a good fit for the data, F(1,23)=19.224, p<0.001. According to the regression coefficients, the duration of the drafting phase is moderately and positively related to the total gaze time (beta= 0.675, p<0.001). It means that the longer the participant's gaze time, the longer the duration of their drafting phase. The tested model explains 43.2% of the total gaze time variability.

The variable of the mean length of pauses longer than 5s in the L2-L1 translation direction correlates strongly and positively with the number of pauses longer than 5s in the same translation direction (r Pearson= 0.7). Therefore, a linear regression analysis could be conducted. The proposed regression model can be assumed to be a good fit for the data since F(1,33)=31.655 p<0.001. The regression coefficients indicate that the number of pauses longer than 5s is strongly and positively related to the mean length of pauses longer than 5s (beta= 5.626, p<0.001). It can be assumed that the higher the number of pauses longer than 5s, the longer the mean length of pauses. The tested model explains 47.5% of the mean length of pause variability.

The L2-L1 drafting phase is correlated with two variables in the same translation direction: the number of pauses longer than 5s (r Pearson= 0.9) and the mean length of pauses longer than 5s (r Pearson= 0.7). The proposed regression model for the drafting phase and the number of pauses longer than 5s appears to be a good fit for the data, F(1,23)=155.614, p<0.001. The regression coefficients indicate that the number of pauses longer than 5s is strongly and positively related to the duration of the drafting phase in the same translation direction (beta= 0.908 p<0.001). It means that the more pauses longer

than 5s the participant produces, the longer their drafting phase is. The tested model explains as much as 82% of the duration of the drafting phase variability.

I was also able to conduct a linear regression analysis for the duration of the drafting phase and the mean length of pauses longer than 5s. The proposed regression model appears to be a good fit for the data, $F(1,33)=34.351$, $p<0.001$. The regression coefficients show that the mean length of pauses longer than 5s is strongly and positively related to the duration of the drafting phase in the same translation direction (beta= 0.715 $p<0.001$). It suggests that the longer the participant's mean length of pauses, the longer their drafting phase. The tested model explains 49.7% of the drafting phase variability.

Since there was a strong and positive correlation between the total gaze time and the duration of the drafting phase, a linear regression analysis for the two variables could be conducted. It appeared that the proposed regression model was a good fit for the data, $F(1,23)=19.224$, $p<0.001$. According to the regression coefficients, total gaze time is moderately and positively related to the duration of the drafting phase (beta= 0.675, $p<0.001$). It means that the longer the participant's total gaze time, the longer the duration of their drafting phase. The tested model explains 43.2% of the drafting phase variability.

The L2-L1 total task time variable was either strongly or moderately and positively correlated with as many as five variables in the same translation direction. These are the duration of the drafting phase (r Pearson= 0.9), the duration of the revision phase (r Pearson= 0.4), the number of pauses longer than 5s (r Pearson= 0.9), the mean length of pauses longer than 5s (r Pearson= 0.7), and the total gaze time (r Pearson= 0.8). The proposed regression model for the duration of the drafting phase and the total task time was a good fit for the data, $F(1,33)= 300.38$, $p<0.001$. The regression coefficients show that the duration of the drafting phase is strongly and positively related to the total task time (beta= 0.949, $p<0.001$). It means that the longer the participant's drafting phase, the longer their total gaze time. The tested model explains as much as 89.9% of the total task time variability.

The linear regression analysis was conducted for the duration of the revision phase and the total task time as well. The proposed model appears to be a good fit for the data, $F(1,33)= 6.787$, $p=0.014$. According to the regression coefficients, the duration of the revision phase is moderately and positively related to the total task time (beta= 0.413, $p=0.014$). Therefore, a participant who had a longer revision phase also had a longer total task time. The tested model explains 14.5% of the total task time variability.

It was also possible to conduct a linear regression analysis for the total task time and the number of pauses longer than 5s. The proposed model seems to be a good fit for the data $F(1,33)= 152.363$, $p<0.001$. The regression coefficients indicate that the number of pauses longer than 5s is strongly and positively related to the total task time (beta= 0.907, $p<0.001$). It means that the person who produces more pauses longer than 5s also produces a longer total task time. The tested model explains as much as 81.7% of the total task time variability.

The mean length of pauses longer than 5s is the next variable that correlates with the total task time. Also, in this case, the proposed model seems to be a good fit for the data $F(1,33)= 26.704$ $p<0.001$. The coefficients indicate that the mean length of pauses longer than 5s is moderately and positively related to the total task time (beta= 0.669, $p<0.001$). It can be assumed that the longer the participants' mean length of pauses, the longer their total task time. The tested model explains 43.1% of the total task time variability.

Finally, I was able to conduct a linear regression analysis for the total task time and the total gaze time in the L2-L1 translation direction. Like in the previous cases, the proposed model was a good fit for the data, $F(1,23)=41.3$, $p<0.001$. According to the regression coefficients, the total gaze time is strongly and positively related to the total task time (beta= 0.801, $p<0.001$). It means that the participant who produces a longer total gaze time, produces a longer total task time as well. The tested model explains as much as 62.7% of the total task time variability.

The overall number of self-reports in the L2-L1 translation direction correlates strongly and positively solely with the number of self-reports of cognitive effort related to translating collocations in the same translation direction (r Pearson= 0.8). Therefore, I was able to conduct a linear regression analysis, and the proposed model was a good fit for the data $F(1,33)= 45.79$, $p<0.001$. Based on the results of the regression coefficients, it can be stated that the number of self-reports of cognitive effort related to translating collocations is strongly and positively related to the overall number of self-reports (beta=0.762, $p<0.001$). It can be assumed that the more self-reports of cognitive effort related to translating collocations the participant verbalises, the more overall self-reports there are. The tested model explains 56.8% of the overall number of self-reports variability.

Chapter 6. Discussion

The aim of this chapter is to discuss the results presented in Chapter 5. It is structured into four sections that correspond to four research questions. Each section includes a short summary of the obtained results. I also attempt to interpret the results referring to previous studies and theoretical assumptions related to the issues of directionality and cognitive effort. Finally, I devote the last paragraph of each section to summarise the discussion and provide the answer to the research question.

6.1 Research question 1

In Research question 1, I set out to investigate how directionality influences cognitive effort. I refer to variables that have already been analysed in the context of directionality and the ones that constitute the novum of this study. In Hypothesis 1, I assumed that the L1-L2 translation direction will evoke higher cognitive effort than the L2-L1 translation direction. However, the hypothesis was not fully confirmed. Although the results from descriptive statistics revealed higher cognitive effort for all variables in the L1-L2 translation direction, the hypothesis was corroborated only for five out of seven analysed variables.

Variables like total gaze time, average fixation duration, total task time, and the number of pauses longer than 5s unequivocally indicated that the L1-L2 translation direction evokes significantly higher cognitive effort. Contrary to the results obtained by Pavlovič and Jensen (2009), the L1-L2 gaze time from this study was not only longer, but also the difference reached the level of statistical significance. This translation direction is characterised by a more extended distribution of the variable. Moreover, a few participants produced extremely long total gaze times. A definitely lower cognitive effort in the reverse direction is reflected not only in the lack of outliers but also in a high number of short total gaze times. The results of linear regression indicate that longer L1-L2 total gaze time is predicted by the number and the mean length of pauses longer than 5s. Therefore, it can be concluded that due to an increased cognitive effort, participants were not able to smoothly proceed with the L1-L2 translation process. Instead of that, they frequently had to interrupt their workflow to look at the screen.

The results from the average fixation duration are in line with the data obtained, for example, in the studies by Ferreira et al. (2016) and Pavlovič and Jensen (2009). In the second study, a significantly longer L1-L2 average fixation direction was reported

solely among the group of students. It may suggest that translation trainees are more prone to translation asymmetry, and therefore, they struggle with difficulties imposed by translating into their foreign language. However, there are also cases of extremely low and extremely short outliers in the L1-L2 direction. On the one hand, it may be the result of the participants' individual working styles and the time they devote to looking at the keyboard rather than the screen. On the other hand, there were fewer such outliers when participants were working into their native language. Since the previous gaze measure revealed no outliers in the L2-L1 direction, it can be concluded that the outliers indicate an increased cognitive effort.

Significantly longer L1-L2 total task time obtained in this study confirmed the previous results from the studies by Pavlovič and Jensen (2009), Ferreira et al. (2016) and the descriptive statistics results from the study by Hunziker Heeb (2020). Like in the case of the previously discussed variables, increased cognitive effort in the L1-L2 total task time is characterised by a more extended distribution and extremely long outliers, both of which do not exist in the reverse translation direction. The linear regression indicates that the length of the total task time is predicted by other variables. It appears that the high number of pauses during the drafting phase and the duration of the translation phases are important components of the total task time. As expected, regardless of the direction, most of the translation process was devoted to translating and revising the text. These results are in line with the study by Jakobsen (2002). The detailed participation of each of the phases in the translation process will be discussed in the next section.

Following the adopted hypothesis, participants produced significantly more pauses during their L1-L2 translation. A similar result can be found in the study by Whyatt (2018). However, in her study, the difference in the number of pauses is minor and, thus, did not reach the level of statistical significance. In the case of this study, there are over 200 pauses more in the L1-L2 translation direction, which translates into significant differences. The participants' profiles may also influence the results. Whyatt (2018) gathered data from professional translators having more than 5 years of professional experience. Participants of this study were translation trainees. In line with data from the previously discussed variables, the distribution of the L1-L2 number of pauses is more extended. Since pauses are an observable indicator of cognitive effort, it can be stated that the participants of this study had to overcome a significant number of problem triggers in the L1-L2 translation direction that required them to interrupt their workflow.

The mean length of pauses longer than 5s was only slightly longer in the L1-L2 translation direction, which unfortunately did not translate into statistically significant differences. Interestingly, these results are in line with the data gathered by Hunziker Heeb (2020), who also obtained minor differences between the two translation directions without reaching the level of statistical significance. To the best of my knowledge, this variable was not analysed in the context of directionality in other studies. In fact, there are many similarities between the L1-L2 and L2-L1 pause lengths produced by the participants of this study. For example, both translation directions are characterised by outliers in the form of extremely long pauses and a high number of results close to the mean value. Regardless of the translation direction, the number of pauses is a significant predictor of their length. Nevertheless, a slightly higher number of very short pauses in the L2-L1 translation duration may indicate that participants were able to faster come up with a translation solution in this translation direction.

EKS and the weighted rating of the NASA-TLX constitute the novum of this study, and, to the best of my knowledge, they have not been analysed in the context of directionality before. Both variables revealed interesting results. Although the mean values of the EKS indicated increased cognitive effort in the L1-L2 translation, it was not confirmed by inferential statistics. It was also not possible to conduct a linear regression analysis for this variable. However, the differences between the L1-L2 and L2-L1 EKS cannot be ignored. While translating into their native language, participants frequently produced short EKS. It means that they were able to type the correct translation of collocation almost immediately. While translating into their foreign language, participants produced similar numbers of short and medium-length EKS. Unlike in the L2-L1 direction, some extremely long EKS also appeared this time. It means that participants often had to invest more time translating collocations into their foreign language. It should also be noted that to avoid the white-coat effect, participants were not informed that collocations constitute the object of this study. Therefore, the lack of statistically significant differences between the two translation directions may suggest that trainees are not always able to distinguish collocations as separate units of meaning. Instead of that, they treat the text more holistically and aim at translating the main message of the text unless a collocation causes a significant cognitive overload. At that moment, they have to apply translation techniques and strategies to decrease cognitive effort and proceed with translation. This issue will be discussed in detail in Section 6.3 based on self-reports of cognitive effort.

The weighted rating of the NASA-TLX appeared to be higher in the case of the L1-L2 translation direction. Although the results obtained a level of statistical significance, it was on the verge of the threshold value. It may mean that the results revealed a trend rather than regularity. The results of the NASA-TLX were not predicted by any of the remaining variables analysed in this study. The mean values of the six rating scales provided interesting details of similarities in the students' perception of the two translation directions. Although the L1-L2 direction was indicated as slightly more effortful in all six rating scales, the differences were rather marginal. Since the previously analysed eye-tracking and keylogging variables demonstrated, in almost all cases, significant differences between the translation directions, I expected a more clear-cut difference between the ratings of the NASA-TLX. For example, participants almost identically perceived mental demand, effort and frustration, assessing them as moderate regardless of the translation direction. It shows that neither of the translation directions was devoid of difficulties, and participants had to struggle with significant problem triggers even if they were working into their native language. Moreover, similar results from the mental demand and effort scales may suggest that, in line with the results obtained by Gieshoff and Hunziker Heeb (2022, 2023), participants struggled with distinguishing both concepts despite being provided with detailed instructions, including definitions of each scale.

Interestingly, participants highly evaluated their performance in both L1-L2 and L2-L1 directions. It may mean that they find their translation solutions successful. Although both directions evoked effort and frustration, trainees were satisfied with the outcome of their translation process, regardless of the translation direction. These results are contrary to data obtained from the assessment of translations of collocations, where the accuracy of the L1-L2 translations was significantly lower. However, as pointed out by Bartłomiejczyk (2004), trainees' perception of their L2 performance may be slightly distorted, which probably was the case in this study as well. Due to a smaller awareness of possible translation mistakes, trainees have a less critical approach to their work in the L1-L2 direction.

Moreover, the results of the self-designed questionnaire (Figure 17) distributed at the end of the experimental procedure seem to confirm the data obtained from each variable. Participants were asked if any of the translation directions caused them more difficulties. 24 out of 35 participants found the L1-L2 translation direction more difficult. They explained that the majority of problems resulted from the lack of adequate

equivalents at hand. Participants admitted that they needed more time to find a proper translation solution. Grammar issues, spelling, and stylistics also had an impact on the level of difficulty. Only 5 participants pointed the L2-L1 direction as more demanding. This time, they mentioned problems with Polish grammar and stylistics. Finding Polish equivalents of English words is also not devoid of difficulties. Interestingly, there were 6 participants for whom neither of the translation directions was more difficult. However, the results of four of them unequivocally showed increased cognitive effort in the L1-L2 translation direction. The results of one participant revealed increased effort while translating into their native language, and one participant almost equally struggled with both directions. Participants who found neither of the directions more demanding explained that difficulties can be found in any type of text regardless of the translation direction.

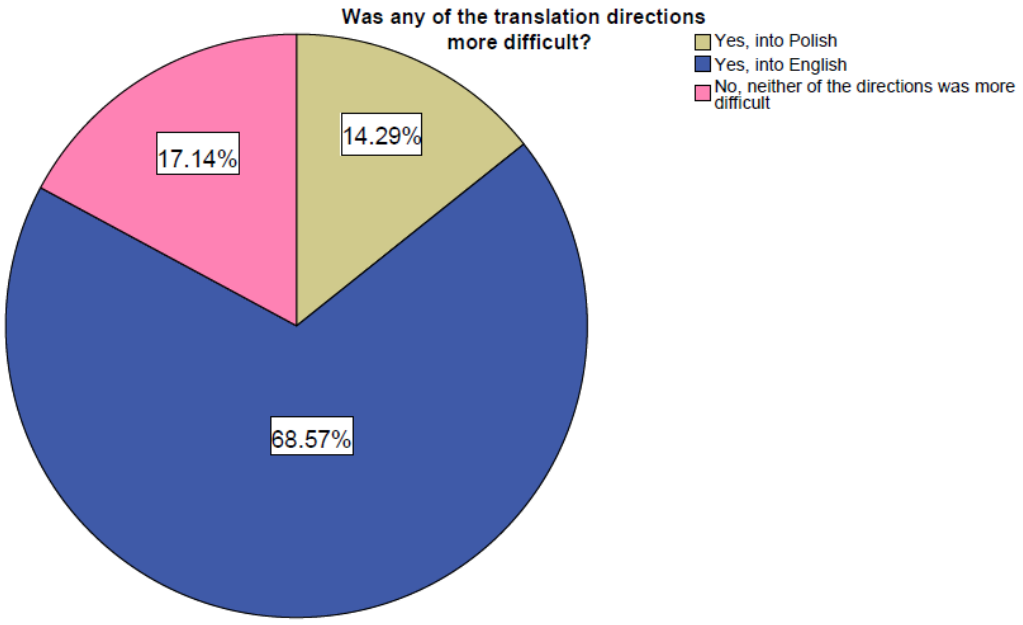


Figure 17. Questionnaire.

To sum up, directionality has an observable impact on cognitive effort. The mean results of eye-tracking and keylogging variables indicated that translation trainees experience higher cognitive effort while translating into their foreign language. It was also corroborated by the questionnaire results, in which more than half of the participants indicated the L1-L2 translation direction as more difficult. However, it should be noted that not all the data reached the level of statistical significance, which may also be related to the design of this study. Surprisingly, the difference in the cognitive effort has a rather

minor influence on participants' perception of their performance. It was assessed as almost equally high regardless of the translation direction.

6.2 Research question 2

In Research question 2, I was interested in how the three stages of the translation process (orientation, drafting, revision) differ in each direction. The study results revealed that apart from crucial differences, there are also some worth noting similarities. Interestingly, only one out of two adopted hypotheses was corroborated.

In Hypothesis 2, I predicted that the orientation phase would be shorter in the L1-L2 translation direction since participants were reading the text in their native language. However, both descriptive and inferential statistics revealed a reverse result. It appeared that not only the orientation phase was longer while translating into English, but also the difference between the two directions was extremely low and did not reach the level of statistical significance. Interestingly, these results are contrary to the results from previous studies by various authors. The eye-tracking results from da Silva et al. (2017) and Whyatt (2018) indicated higher cognitive effort while reading a text in a foreign language.

According to the results of this study, there are significant similarities between the orientation phases in the L1-L2 and L2-L1 directions. In both cases, there was a very high number of participants whose orientation phase was extremely short. Frequently, they either read just a few first words of the source text or omitted the stage of the source text reading and started typing immediately. Following the categorisation introduced by Dragsted and Carl (2013: 142) (see Section 2.4.2), most participants can be characterised as head starters regardless of the translation direction. Only a few participants read the whole source texts in both translation directions. It was mentioned by two participants in the L1-L2 retrospective reports and by four participants in the L2-L1 retrospective reports. It may suggest that translation trainees have not developed the habit of reading the whole source text before translating it. Nevertheless, as Dragsted (2010: 52) pointed out, not all translators prefer reading the whole source text beforehand. It can be assumed that personal preferences may play a crucial role in this case. Translation trainees may also feel time pressure during their translation tasks, especially in laboratory conditions. Thus, they might want to start typing as soon as possible, solving the arising problems in the next phases. However, they were informed that there is no time limit. Moreover, the

linear regression analysis revealed no predictors of the orientation phase length, probably due to its relatively short duration.

Some minor differences in the orientation phases can be detected as well. For example, there is a slightly higher number of the shortest orientation phases in the L2-L1 translation direction. The distribution of the results was a bit greater in the reverse direction, and the outliers were longer than in the L2-L1 direction.

The descriptive and inferential statistics results fully confirmed Hypothesis 3, assuming that participants will produce longer drafting and revision phases in the L1-L2 translation direction. The previously obtained eye-tracking results, for example, by da Silva et al. (2017) and Pavlovič and Jensen (2009) that indicated higher cognitive effort during the L1-L2 drafting and revision, are now confirmed by the duration measure analysed in this study.

The L1-L2 drafting phase is not only longer but also differs in terms of participants' individual behaviour. There is a high number of average results. Moreover, a few participants produced extremely long drafting phases, which cannot be observed in the reverse translation direction. As expected, there was a high number of participants who produced very short drafting phases while translating from English into Polish. Even if some longer drafting phases occur, they are still much shorter than during the Polish-English translation. Thus, it can be assumed that translation trainees had to invest more effort in translating into their foreign language. The number and mean length of pauses are good predictors of the drafting phase length, regardless of the translation direction. Nevertheless, this dependence seems not to be surprising. The more translation disfluencies there are, the longer the participants produce their target texts.

Interestingly, in both translation directions, participants had very short revision phases that lasted, on average, only 5 (L1-L2) and 3 (L2-L1) minutes. However, the difference appeared to be statistically significant, and it can be stated that translation trainees had to invest more effort during the L1-L2 revision. Going into details, the number of very short revisions, the duration of which is close to 0, is higher in the L2-L1 direction. During the L1-L2 translation process, participants tend to revise their target texts either for a very short or medium period of time. Some very long revisions appeared in this translation direction as well. The time devoted to the drafting phase appeared to be a good predictor of the L1-L2 revision phase length. It may also indicate an increased cognitive effort.

Based on these results, it may be assumed that the participants preferred to revise their translations during the drafting phase, performing the so-called online revisions (see Section 2.4.2). Although participants indeed modified their texts on an ongoing basis, the keylogging results revealed that the overwhelming majority of them actively revised their texts also during the revision phase. 26 out of 35 participants introduced some modifications during the L2-L1 revision phase. Usually, they made minor stylistic modifications by changing or adding some words. Sometimes, they also modified parts of the collocations. Another type of revision was introducing punctuation modifications. The remaining nine participants either just read their target text or finished the translation process immediately after the last sentence was translated, not performing the revision at all. The number of participants who introduced some modifications during their L1-L2 revision was slightly higher. 28 out of 35 participants decided to actively revise their texts during this phase. Like in the reverse translation direction, they usually introduced some minor stylistic modifications. However, this time, they also devoted some time to modifying their grammar, with particular attention put on tenses. 8 out of 35 participants either just read their target text or completely omitted the revision phase. Interesting results referring to the revision phases can be found in the retrospective reports. Reading the text once again after drafting the translation was reported three times in the L2-L1 retrospective reports and 0 times in the L1-L2 retrospective reports. Nevertheless, the small number of self-reports describing the revision phase may result from the subjective character of retrospective reports. Participants tend to report significant problems rather than unproblematic parts of the translation process.

To sum up, participants of this study needed slightly more time during the L1-L2 orientation phase. However, they can be characterised as head starters regardless of the translation direction. As expected, both the drafting and revision phases were longer in the L1-L2 translation. Participants tend to implement online revisions during the drafting phase. Moreover, regardless of the translation direction, most participants also revised the text during the revision phase. The character of modifications may slightly differ depending on the translation direction. In both cases, these are rather minor revisions, including some punctuation modifications. In the L1-L2 translation, participants also drew attention to the text's grammatical aspect. In the L2-L1 translation, the revisions were usually related to stylistics.

6.3 Research question 3

In Research question 3, I set out to investigate how participants verbalise cognitive effort related to translating collocations in both directions. First of all, it should be noted that regardless of the translation direction, almost half of all self-reports refer to translating collocations. Bearing in mind that participants usually report problematic aspects (Englund Dimitrova & Tiselius 2009), this result confirms my previous study (Pietryga 2022), suggesting that collocations constitute a significant problem trigger both in the L1-L2 and L2-L1 translation process. Secondly, participants produced twice as many self-reports related to translating collocations in the L1-L2 direction compared to the reverse translation direction. It may indicate that collocations evoked an increased cognitive effort in the L1-L2 direction. Thirdly, as discussed in Section 5.4, the self-reports of cognitive effort related to translating collocations can be divided into three levels. Self-reports belonging to all three levels can be detected in both the L1-L2 and L2-L1 directions. However, their frequency changes depending on the translation direction. I expected that both the number and the content of retrospective reports would indicate higher cognitive effort in translating collocations in the L1-L2 direction. However, this hypothesis was not fully confirmed. It was corroborated only for Level 3 of cognitive effort.

Level 1 of cognitive effort related to translating collocations prevails in the L2-L1 self-reports. Participants produced as many as 30 such self-reports related to this translation direction. They usually provided an explanation of the decision-making that led them to a particular translation solution. The stylistic aspect of the target text also plays an important role. For example, in one of the comments, the participant stated:

> PE03/D/11: In sentence no. 9, I decided to translate ‘this discovery **develops our understanding’** as ‘pomaga nam zrozumieć’ because I totally could not imagine literal translation in this context. It did not sound natural to me. That is why I decided to use ‘pomaga’ here.

Participants also reported using translation strategies and techniques and attempted to explain their decision. Usually, they mentioned literal translation, omission, addition or some textual modifications. For example, one of the participants described translation decision applied to the collocation ‘brain region’ in the following way:

PE31/D/11: In sentence no 9., I omitted the word ‘region’. I figured out that it is about an older and newer part; however, the word ‘regions’ has a geographic connotation for me, and I substituted it with the word ‘part.’

A smaller number of self-reports were classified as Level 1 in the L1-L2 translation direction, but the difference between the directions appeared to be statistically insignificant. In this translation direction, participants usually describe their translation process; however, they less frequently explain their motives. For example,

PE11/T/3: In sentence no.2, instead of ‘**przybliżona data’**, I used a synonym saying that the date is not precisely stated, but I think it has the same sense.

A description of the implemented techniques can also be found. Participants reported the usage of omission, generalisation, and other strategies and techniques like literal translation, reformulation, and idiom usage. For example,

PE37/T/6 : There was omission in the sentence number 9 ‘**odcisk jego łapy**.’ The word ‘odcisk’ was omitted and translated as ‘przedstawiający jego łapę’, ‘zdjęcie’.
PE13/T/3: I definitely did it in a more descriptive form, I did not use the word ‘workers’ or anything similar, just ‘people who work on’. Frankly speaking, I do not know why, probably because I associate ‘workers’ with labourers. – this comment refers to the collocation ‘branża turystyczna.’

Interestingly, regardless of the translation direction, the smallest number of self-reports of cognitive effort related to translating collocations was classified as Level 2, in which participants described the applied modifications. In the L1-L2 self-reports, only two such comments were produced by one participant. Apart from the comment already quoted in Section 5.4, the participant also reported that:

PE04/T/7: **’Rozpiera go energia’** – ‘bursts with energy’. At the beginning, I probably translated it as ‘it bursts with energy’. Even if that is a good translation, it is probably the wrong tense. It is not about bursting with energy now but in general.

There are slightly more comments classified as Level 2 in the L2-L1 direction. This time, participants frequently precisely described their decision-making and motives. The modifications seem to result either from the textual knowledge acquired while translating the subsequent parts of the text or from stylistic or grammatical errors of the target text. For example,

> PE15/D/1: At the beginning I translated '**respond correctly'** as 'odpowiadają prawidłowo', but in fact, dogs do not speak, so I changed the word 'odpowiadają' into 'reagują', probably.
> PE03/D/1: In the first sentence, I wrote 'radosnym, **wysokim głosem'** at first or something like that. But then I saw 'joyful', which I also wanted to translate as 'radosny'. So I left 'joyful,' 'radosny' and I changed the beginning.

Level 3 of cognitive effort related to translating collocations prevails in the L1-L2 self-reports. There are 89 such self-reports in this translation direction. It is worth noting that sometimes it is not a collocation that was a problem trigger but rather its component. For example,

> PE02/T/8: I also had a problem with '**huczne obchody'**, the word 'huczne' caused some problems.
> PE38/T/6: In sentence no. 6, I was not really sure how nicely translate '**skupia uwagę**.'

Another type of self-report classified as Level 3 was reporting a lack of lexical knowledge. Frequently, participants explicitly stated that they either did not know the phrase or could not recall it at the moment. In many self-reports, they also presented a solution they decided to adopt, which was usually descriptive equivalent or omission of an unknown word.

> PE33/T/5: In sentence no. 9, I could not recall how to translate the word '**pamiątkowy**'. That is why I decided to simply omit this word. – This comment refers to the collocation 'pamiątkowy certyfikat.'

PE36/T/6: In sentence no. 4, I was not sure how to name a person who cannot smell, so I did it in a descriptive manner: ‘żółw stracił węch’, so ‘he has lost his ability to smell.’

PE37/T/1: In sentence no. 2, when there is ‘**przybliżona data’**, I said it in a general manner because I did not exactly remember how to say it. I forgot the word, so I said ‘the closest date’.

PE38/T/11: I do not know how to say ‘**odcisk**’, and I am also not sure whether ‘**łapa** żółwia’ is ‘paw’ in English. – this comment refers to the collocation ‘odcisk łapy.’

Participants verbalised only 13 L2-L1 self-reports classified as Level 3, and this time, the difference between the translation directions appeared to be statistically significant. Likewise, participants reported explicit problems with translating collocations and temporal or permanent lack of lexical knowledge. They also provided an explanation of the applied techniques and strategies. Interestingly, in this translation direction, they reported the usage of calque, which cannot be found in the L1-L2 self-reports. For example,

PE06/D/3: There was a problematic phrase in part of the sentence, no 2, ‘that is why scientists **were** made **curious**.’ I translated it as ‘to co bardzo zainteresowało naukowców.’

PE13/D/6: ‘**Acoustic information’**, I am not sure, whether it is just ‘przetwarzanie dźwięków.’

PE17/D/11: In sentence no 9, I did not know how to translate ‘older **brain region’** and ‘newer part’, so I did a calque.

Problems related to translating just one component of a collocation also appeared in this translation direction. However, they had a small scope and frequency. Participants struggled with the phrase ‘high-pithed’, forming a collocation ‘high-pithed voice.’ Four out of 13 comments verbalised in this translation direction draw attention to a problem with this phrase. For example,

PE15/D/1: In the first sentence, I had a problem with the phrase ‘high-pitched’, ‘z podniesionym głosem’ that I came up with sounded a bit aggressively, so I wrote ‘z uniesionym.’

PE27/D/2: I also omitted the phrase ‘high-pitched’ because I was not completely sure of its Polish equivalent, and I did not want to introduce any element of misunderstanding.

One more type of self-reports can be distinguished from retrospections verbalised by the participants. I classified it as ‘connected with collocations’ since the problem did not result directly from the analysed collocation. Participants drew attention to the issue of connecting the collocation with the preceding element. The problem resulted from insufficient lexical or theoretical knowledge in this scope. However, there are only four such self-reports that appeared only in the L2-L1 direction. All of them are related to the collocation ‘brain region’. These are:

PE15/D/12: I was not sure whether something like ‘starszy obszar mózgu i nowszy’ exists. I had no idea what this was really about. I do not have knowledge about the brain regions, so I hope I will present it correctly; we have some older and newer parts of the brain.

PE20/D/8: Later, in the 9th sentence, I had a problem with the following passage, ‘the emotional element with the older brain region’. I am unsure whether I understood it correctly because the next part states that we also have ‘the newer part’. It seems to me that we are talking about the front and back parts of the brain. However, I am not sure.

PE25/D/4: What ‘the older brain region’ is really about? It is a bit unclear to me.

PE35/D/5: Next, in sentence number 9, the brain region was described as older and newer. I was not really sure what it was about, so I translated it literally.

Interestingly, this type of self-reports did not appear in the L1-L2 translation direction. It may mean that the problem of connecting the collocation with the preceding element does not result from the cognitive effort evoked by the direction of translation, but rather, it is an unintended difficulty arising out of the source text.

To sum up, in the L1-L2 translation direction, participants most frequently describe cognitive effort related to translating collocations by reporting an explicit

problem encountered in the translation process. They often draw attention to a temporal or permanent lack of lexical knowledge. However, a problem trigger in the form of just one collocation component is far from being insignificant. In the L2-L1 translation direction, participants most frequently describe cognitive effort related to translating collocations by reporting decisions that led them to implement a particular translation solution. They also provide an explanation of the reasons behind their decisions and report the application of techniques and strategies. Regardless of the translation direction, a description of modifications introduced to the already translated collocations rarely appears in participants' self-reports.

6.4 Research question 4

In Research question 4, I was interested in how directionality influences translation accuracy. First of all, it should be emphasised that the notion of translation accuracy is, unfortunately, a very subjective measure, which may be difficult to apply to the whole text. The assessment of accuracy is also prone to personal bias. To somehow overcome this problem, I decided not only to commission this task to third persons not involved in this study and, in this way, reduce the influence of my personal perspective but also to verify the translation accuracy in a limited, local context of collocations.

In Hypothesis 5, I expected that translation accuracy would be higher in the L2-L1 direction due to the fact that participants are working into their native language. I assumed that based on the linguistic knowledge of their native language, they would have a wider range of solutions and, therefore, they should be able to find a correct translation of collocations. The hypothesis was fully confirmed by the results of both descriptive and inferential statistics. Participants managed to obtain higher scores for their translations of collocations in the L2-L1 direction. The difference of more than 200 points between the two directions is far from being insignificant. The distribution of the percentage of scores shows how the accuracy of both translation directions differs. There are around 13% more highly accurate translations of collocations in the L2-L1 translation direction. The L2-L1 direction is also characterised by almost 12% smaller number of partially accurate translations of collocations. However, there is only a 3% difference in completely inaccurate translations of collocations. Moreover, there is a smaller variability of the results in the L2-L1 translation direction, which means that most participants performed equally well in translating collocations. Participants obtained lower scores in the L1-L2 translation direction. A greater variability and distribution of the results can also be

observed in this case. It means that while some participants managed to produce highly accurate translations of collocations, the accuracy of translations produced by the others was very low. These results are in line with translation asymmetry observed by Kroll and Steward (1994) (see Section 1.4.2) and the findings from the study by Tomczak and Whyatt (2022), according to which translators have less successful lexical choices when translating into their foreign language.

Moreover, I was not able to conduct a linear regression analysis of the translation accuracy since the precondition of a significant correlation with any other variable was not met. It may mean that translation accuracy results solely from the participants' linguistic knowledge in each translation direction rather than being predicted by other conditions like the length of the translation process, time devoted to fixate on the text, and the number of times the participant had to pause.

Experts assessing the accuracy of translations of collocations also had the possibility to explain their decision. However, this part of their task was not obligatory; therefore, it is not possible to precisely assign the types of mistakes to particular collocations. Nevertheless, based on the experts' comments, I am still able to detect general problems that the participants struggled with in each translation direction. When assigning 0 or 1 points to the L1-L2 translations of collocations, experts frequently enumerated issues with incorrect grammar, usually related to articles and tenses. Some spelling mistakes appeared as well. Another frequent problem was a change in the meaning of the sentence. Although participants managed to translate the collocation, its meaning was not conveyed, and the applied translation solution significantly changed the message. Experts also pointed out improper use of translation techniques and strategies like omission, literal translation, and calque. This issue was already discussed in Section 6.3. Participants explained that they used such textual modifications to overcome lexical problems and were usually aware of a possible translation inaccuracy. Finally, experts emphasised problems with incorrect stylistics. Experts also enumerated a lack of equivalence, wrong grammar use, and omitting components of collocations while assessing the L2-L1 translation accuracy. Although this time, they were less frequent. Participants also used words that do not collocate with each other, forming an unnatural collocation. Another problem leading to inaccurate L2-L1 translation was using an existing collocation, which, however, was not an adequate contextual choice.

To sum up, it can be concluded that directionality has a significant influence on translation accuracy. Translation trainees produced more accurate translations when

translating from a foreign language into their native one. The accuracy of translations of collocations in the L1-L2 direction ranges from moderate to high. The accuracy of the L2-L1 translations of collocations is rather high. The most frequent problems leading to translation inaccuracy in both translation directions are the improper cover of meaning, wrong grammar use, and incorrect application of translation techniques and strategies.

Chapter 7. Concluding remarks

This PhD thesis addresses the issues of cognitive effort and directionality in translation process. The aim of this study was two-fold. Its primary objective was to analyse the influence of directionality on cognitive effort among translation trainees. I was interested in whether the results of the study would be in line with the assumptions of the Golden Rule of Translation (Newmark 1988) and the translation asymmetry (Kroll & Steward 1994). The second objective was an attempt to implement new variables to the translation process research on cognitive effort and directionality. I focused on the group of advanced translation trainees since they have developed abilities that allow them to enter the translation market, but at the same time, they have not developed the automatisation of the translation process yet. Therefore, the participants provided valuable information about the cognitive effort evoked by the translation process. The analysed combination of the language of low diffusion (Polish) and contemporary lingua franca (English) enabled gaining insight into the problem of directionality based on the language pair in the case of which the L1-L2 translation seems to be an everyday practice (Whyatt & Kościuczuk 2013).

I adopted a mixed-methods approach to verify five hypotheses and answer four research questions. The data were gathered using four both subjective and objective process methods: eye-tracking, keylogging, retrospective verbal reports, and questionnaires: the NASA-TLX, as well as the self-designed questionnaire. The following variables were adopted as indicators of cognitive effort: average fixation duration, total gaze time, eye-key span, total task time, duration of the orientation, drafting and revision phases, number of pauses longer than 5s, mean length of pauses longer than 5s, self-reports of cognitive effort, the weighted rating of the NASA-TLX and accuracy.

The descriptive statistics results revealed that, in general, L1-L2 translation evokes higher cognitive effort than L2-L1 translation direction. Such observation was confirmed by the mean results from eleven out of twelve variables. However, not all the differences reached the level of statistical significance. Hypothesis 1, assuming that cognitive effort, operationalised by total gaze time, average fixation duration, total task time, the number of pauses longer than 5s, mean pause length of pauses longer than 5s, EKS, and the weighted rating of the NASA-TLX, is higher in L1-L2 direction was partially confirmed. It was corroborated for five out of seven analysed variables. The results from EKS and the mean length of pauses longer than 5s did not reach the level of

statistical significance. Hypothesis 2, assuming that the orientation phase is shorter in the L1-L2 direction, was rejected. Contrary to predictions, the orientation phase was longer in the L1-L2 translation direction. Hypothesis 3, assuming that the drafting phase and revision phase are longer in the L1-L2 direction, was fully corroborated for both variables. Hypothesis 4, assuming that the content and number of retrospective reports would indicate higher cognitive effort in translating collocations in the L1-L2 direction, was partially confirmed. Although the general number of self-reports of cognitive effort reached the level of statistical significance, the detailed results revealed that only the difference in Level 3 of cognitive effort was statistically significant. Hypothesis 5, assuming that translation accuracy will be higher in the L2-L1 direction, was fully corroborated.

Research question 1 covered the topic of the influence of directionality on cognitive effort. Eye-tracking and keylogging data showed higher cognitive effort in the L1-L2 translation direction. This observation was also confirmed by the data obtained through questionnaires. According to the results from the weighted rating of the NASA-TLX and the self-designed questionnaire, participants had more problems when working into their L2. Interestingly, it did not correspond to participants' assessment of their own performance, as regardless of the translation direction and encountered difficulties, they were rather satisfied with the outcome of their translation process. Research question 2 was devoted to differences between the L1-L2 and L2-L1 translation phases. I observed that the drafting and revision phases were significantly longer during the L1-L2 translation. Nevertheless, both directions demonstrate many similarities. First of all, the drafting phase constituted the majority of the translation process. Activities like target text production and implementing online revisions can be identified in this phase. Secondly, translation trainees tend to have a rather short orientation and revision phases. They display a tendency to start drafting a target text immediately without reading the source text beforehand. However, active revision behaviour can be identified in the revision phase. In Research question 3, I was interested in how participants verbalise cognitive effort related to translating collocations. The data obtained from the retrospective verbal reports indicated that the scope of cognitive effort varies depending on the translation direction. In the L2-L1 self-reports of cognitive effort related to translating collocations, participants usually drew attention to the decision-making process and the implemented translation techniques and strategies. The L1-L2 self-reports of cognitive effort mostly covered the subject of explicit problems with translating

collocations. Finally, Research question 4 was devoted to translation accuracy. The L1-L2 translation direction appeared to have inferior translation accuracy resulting from grammar and lexical problems.

To sum up, on the one hand, translation trainees manifest an L2-L1 advantage, but on the other hand, neither of the translation directions is devoid of problems. Although not all hypotheses were confirmed by statistically significant results, the mean results and the differences in their distribution are far from being insignificant. Higher cognitive effort evoked by the L1-L2 translation direction is unequivocally reflected in gazing behaviour, longer drafting and revision phases, greater number of pauses, the number and character of self-reports of cognitive effort, and lower translation accuracy. Therefore, I will conclude that translation trainees struggle with higher cognitive effort when translating from their L1 into their L2.

The results also revealed interesting information about the issue of collocations in the translation process. Collocations were confirmed to be a significant local indicator of cognitive effort. In line with the previously discussed results, translating collocations in the L1-L2 direction evoked higher cognitive effort. First of all, translation trainees managed to verbalise a significant number of self-reports describing various levels of cognitive effort related to translating collocations. However, they frequently reported problems with only one component of the collocation. Moreover, translation accuracy appeared to be significantly lower in the L1-L2 translation direction. However, the difference between the L1-L2 and L2-L1 EKS that was measured based on collocations did not reach the level of statistical significance. Interestingly, despite the reported problems and lower L1-L2 translation accuracy participants highly evaluated their performance in one of the scales that was part of the NASA-TLX. It may suggest that translation trainees do not always perceive collocations as separate units of meaning and take a more holistic approach when translating texts.

These observations may also result in didactic implications. Although advanced translation trainees possess high translation competencies, L1-L2 translation training is still required to minimise the translation asymmetry. Strong emphasis may also be focused on smaller units of meaning like collocations, idioms and metaphors, for example, based on the so-called problem-solving techniques (Piotrowska 2007: 51). Perhaps special attention could be devoted to source text analysis during the orientation phase. According to the obtained results, such practice is needed in both translation directions.

As already mentioned, the second objective was to implement new variables to the translation process research on cognitive effort and directionality. The weighted rating of the NASA-TLX and EKS have never been analysed in the context of directionality before. To the best of my knowledge, the work by Hunziker Heeb (2020) is the only study that attempted to analyse the mean pause length. All three variables appeared to be valuable indicators of cognitive effort in the context of directionality. The NASA-TLX provides detailed information about participants' perception of the translation process. Since the questionnaire is composed of two separate components, the researcher is able to obtain data on various aspects, beginning from the feeling of frustration through assessments of one's own performance to different scopes of cognitive effort. Each of them constitutes an important factor contributing to the experience of cognitive effort evoked by the L1-L2 and L2-L1 directions of translation. Moreover, the weighted average calculated individually for each participant ensures high objectivity of the obtained results. In fact, the researcher can either use both components or solely the rating scales, depending on the study objectives. The analysis of the mean pause length was borrowed from the interpreting process research. Its implementation in translation studies allows for verifying the scope of the experienced cognitive effort. A very long mean pause length can serve as an indicator of significant cognitive overload. This variable can be analysed both in local and global contexts. A very precise measurement of the EKS allows to analyse directionality in reference to the local context and various units of meaning.

I also developed the categorisation in the form of Levels of cognitive effort based on the data obtained from the retrospective reports. To the best of my knowledge, retrospective verbal reports have not been analysed in this context before. I developed the following categories:

•Level 1 – participants describe the process of translating collocations,

•Level 2 – participants describe the process of changing a decision or introducing modifications,

•Level 3 – participants explicitly state that there was a problem during the translation process.

The introduced categorisation allowed me to analyse the characteristics of cognitive effort in the local context of collocations. As a result, I was able not only to identify the more effortful translation direction but also to learn about the reasons behind the experienced cognitive effort.

Obviously, this work is not devoid of limitations. Some of them may result from the number and character of the used methods. For example, since the utilised eye-tracker does not provide scroll compensation for the screen recording mode, I had to significantly reduce the source texts length and the number of analysed collocations. The number of methods definitely had an impact on the length of the experimental procedure. However, I tried to minimise the negative consequences of the prolonged experimental procedure by randomising the order of translation directions. I also struggled with calibration issues and, as a result, had to remove some data sets from the analysis. In my dissertation, I referred to either 25 or 35 data sets. Moreover, in the case of this study, the difference between the L1-L2 and L2-L1 EKS appeared to be statistically insignificant. It can result from the fact that participants were not informed about the interest areas that were used to measure EKS to ensure ecological validity. As a consequence, they may not have detected the objects of analysis. However, it can lead to new avenues for further research. EKS can be analysed in a more restricted context, limited to sentences or phrases. Such an approach was successfully adopted, for example, by Chmiel et al. (2020), in the analysis of time lag measurements in interpreting. I also believe that the remaining adopted variables can be further developed. For example, the analysis of pausing behaviour can be triangulated with eye-tracking analysis. Last but not least, I hope that the developed Levels of cognitive effort could be used in a broader spectrum, not limited to translating collocations.

SUMMARY IN ENGLISH

This thesis addresses the issue of the influence of directionality on cognitive effort of trainees. Directionality, defined as translating into one's foreign or native language (Whyatt 2019), has been a common translation practice for ages. However, translation process researchers drew particular attention to this topic at the beginning of the 21st century. As a result of technological development and the implementation of translation process methods, researchers have been able to analyse translators' self-reports as well as typing and gazing patterns. According to the Golden Rule of Translation (Newmark 1988), translators should work only into their L1. There is also an assumption that the L1-L2 translation direction results in significantly higher cognitive effort (Fonseca 2015; Whyatt 2018, 2019). Nevertheless, this assumption was not fully confirmed by the results of studies on the translation process. Frequently, some of the analysed variables either did not reach the level of statistical significance (e.g., in works by Pavlovič & Jensen 2009 & Hunziker Heeb 2020) or indicated L2-L1 translation direction as more effortful (e.g., in a work by Ferreira et al. 2021).

The aim of my work is two-fold. First of all, I would like to analyse the influence of directionality on cognitive effort in translation. My second objective is to implement new variables to the research on the translation process, directionality and cognitive effort. Variables like EKS and the weighted rating of the NASA-TLX have not been analysed in this context before. To the best of my knowledge, the analysis of the mean pause length was the subject of analysis in only one study. Four methods were used to gather data. These are eye-tracking, in the case of which I utilised the Eyelnk PortableDuo eye tracker; keylogging, in the case of which I used the Translog II programme, retrospective verbal reports and questionnaires. These methods allowed me to gather both subjective and objective data. I was also able to conduct qualitative and quantitative data analysis. Moreover, cognitive effort was measured both globally and locally since collocations served as points of interest. The results indicated that the direction of translation significantly influences the perceived cognitive effort. According to descriptive statistics, higher cognitive effort is evoked by the L1-L2 translation direction. Following the results of statistical tests, two hypotheses were fully corroborated, two hypotheses were partially confirmed, and one hypothesis was rejected. The results of inferential statistics revealed that higher cognitive effort in the L1-L2 translation could be observed based on gazing patterns, typing behaviour, and the length of drafting and

revision phases, the number and character of self-reports cognitive effort, and inferior translation accuracy.

This work is structured into seven chapters. Chapters 1-3 discuss the literature review, and Chapters 4-7 embrace the empirical part of the study. Chapter 1 covers the topic of directionality in translation. I discuss various definitions of directionality that can be found in the literature. Next, I move on to the brief history of attitudes towards directionality, including the emergence of the Golden Rule of Translation (Newmark 1988) and the contemporary perception of translating into L1 and L2. In section 1.3, I focus on the issue of languages of low diffusion. It is worth noting that Polish has been classified within this category. The last section of this chapter discusses the phenomenon of language asymmetry based on the Revised Hierarchical Model (Kroll & Steward 1994).

Chapter 2 covers the topic of process methods. Following the chronological perspective, it opens with a discussion of verbal reports, where particular attention is devoted to retrospective verbal reports and think-aloud protocols. Then, I discuss the method of keylogging, the Translog II programme and the phases of the translation process (Jakobsen 2002). The next section discusses eye-tracking. I refer to the eye-mind hypothesis (Just & Carpenter 1980) and define basic eye-tracking variables, fixations, saccades and pupil dilation. Section 2.6 is devoted to neuroimaging techniques, EEG, PET, and fMRI. Finally, the last section briefly discusses questionnaire studies.

Chapter 3 constitutes the last part of the literature review. It discusses the issue of cognitive effort. I present three models of cognitive effort: the Effort Model (Gile 1995), the Cognitive Load Model (Seeber 2011), and the Cognitive Load Theory (Sweller et al. 1998). The next section refers to the indicators of cognitive effort developed by Chen et al. (2012) and later thoroughly analysed by Ehrenberger-Dow et al. (2020). I also elaborate on the current state of research on cognitive effort and directionality. Finally, I discuss a novelty in the research on cognitive effort, default translation (Halverson 2019), which denotes an uninterrupted translation process resulting in a decreased level of cognitive effort.

The empirical part begins with Chapter 4. It opens with a discussion of the aim of this study. I also devote particular attention to the notion of collocations as problem triggers in the context of directionality because there is a significant research gap in this scope. Since I denote this study as an experiment, I also draw a clear distinction between experimental and quasi-experimental research. Next, I present research questions,

hypotheses, dependent and independent variables, as well as methods and materials used in the experiment. The last section of Chapter 4 covers the details of the pilot study and the main study research design.

Chapter 5 presents the analysis of the results. It is structured into six sections. Five of them correspond to the adopted hypotheses, and the last one discusses the results of linear regression analysis. This chapter includes descriptive statistics results and histograms. I also discuss the results obtained through statistical tests, like the pair test and its non-parametric equivalent, the Wilcoxon test.

In Chapter 6, devoted to a discussion of the results, I make an attempt to provide answers to research questions. It is divided into four sections that correspond to four research questions. I present the obtained results and interpret them based on the theoretical assumptions related to the phenomena of cognitive effort and directionality. Chapter 7 discusses the concluding remarks as well as limitations and potential avenues for further research.

This thesis also includes a list of references with 247 works published between 1980 and 2024, summaries in Polish and English, as well as Appendices including materials used in the experiment, the detailed results and a summary of the statistical tests results.

STRESZCZENIE W JĘZYKU POLSKIM

Niniejsza praca porusza temat wpływu kierunkowości przekładu pisemnego na wysiłek kognitywny studentów tłumaczenia. Jak pokazują badania, kierunkowość rozumiana jako proces tłumaczenia na język obcy lub rodzimy (Whyatt 2019) jest obecna w praktyce przekładowej niemal od samego jej początku. Jednakże zainteresowanie badaczy procesu przekładu wzbudziła na początku XXI w. Postęp technologiczny i pojawiające się w jego rezultacie metody procesualne umożliwiły badaczom wnikliwą analizę zwerbalizowanych komentarzy na temat kierunkowości i wysiłku kognitywnego, procesu powstawania tekstu, a także śledzenie ruchu gałek ocznych. Zgodnie z tzw. Złotą Zasadą Tłumaczenia (Golden Rule of Translation, Newmark 1988) przekład powinien być wykonywany wyłącznie na język rodzimy. W związku z powyższym istnieje również przekonanie, że przekład na język obcy zawsze generuje znacznie większy wysiłek kognitywny niż przekład na język rodzimy (Fonseca 2015, Whyatt 2018, 2019). Niemniej jednak badania nad kierunkowością zdają się nie potwierdzać w pełni tego założenia. Część analizowanych zmiennych nie osiągała poziomu istotności statystycznej (np. w pracach Pavlovič i Jensena 2009, czy Hunziker Heeb 2020), lub też niektóre zmienne wykazywały, że przekład na język rodzimy generuje większy wysiłek kognitywny (np. w pracy Ferreiry i in. 2021).

Celem mojej pracy jest analiza wpływu kierunkowości na wysiłek kognitywny w przekładzie pisemnym. Dodatkowo pragnę rozszerzyć badania nad kierunkowością o nowe zmienne, nieanalizowane dotąd w tym kontekście, takie jakie EKS (*eye-key span*), średnia długość pauz (*mean pause length*), czy też wyniki kwestionariusza NASA-TLX (*weighted rating of the NASA-TLX*). Analizowane dane zostały zebrane przy wykorzystaniu czterech metod procesualnych: okulografii z wykorzystaniem okulografu Eyelink PortableDuo, keyloggingu z wykorzystaniem programu Translog II, protokołów retrospektywnych oraz kwestionariuszy. Wykorzystane metody pozwoliły na zebranie danych obiektywnych i subiektywnych, a także na ilościową i jakościową analizę danych. Wysiłek kognitywny został zmierzony zarówno globalnie, jak i lokalnie. Miary lokalne odnoszą się do nieanalizowanego dotąd pod kątem kierunkowości, przekładu kolokacji. Uzyskane wyniki wskazują, że kierunek tłumaczenia ma znaczny wpływ na generowany wysiłek kognitywny. Wyniki statystyki opisowej jednoznacznie wskazują, że większe obciążenie kognitywne jest wynikiem przekładu na język obcy. Przeprowadzone testy statystyczne pozwoliły na pełne przyjęcie dwóch hipotez, częściowe przyjęcie dwóch

hipotez oraz odrzucenie jednej hipotezy. Statystyka inferencyjna wskazuje, że zwiększony wysiłek kognitywny w przekładzie na język obcy ujawnia się w zmiennych okulograficznych oraz keyloggingowych, długości faz przekładu, zwerbalizowanych komentarzach retrospektywnych oraz w mniejszej dokładności przekładu.

Niniejsza praca składa się z siedmiu rozdziałów. Rozdziały 1, 2 oraz 3 stanowią przegląd literatury. Natomiast rozdziały 4, 5, 6 oraz 7 obejmują część empiryczną pracy. Rozdział pierwszy poświęcony jest tematyce kierunkowości w przekładzie. Autorka pracy omawia różne definicje kierunkowości, funkcjonujące w literaturze przedmiotu. Następnie przedstawiona zostaje krótka historia tego zjawiska, począwszy od powszechnej akceptowalności praktyki przekładu na język obcy, przez publikację Złotej Zasady Tłumaczenia (Golden Rule of Translation), po współczesne postrzeganie tego tematu. Kolejna sekcja rozdziału pierwszego poświęcona jest tematyce języków o ograniczonym rozpowszechnieniu (*low diffusion languages*), do których zaklasyfikowany został również język polski. Rozdział zamyka problem asymetrii językowej, omówiony na podstawie Revised Hierarchical Model (Kroll i Seward 1994).

Rozdział drugi poświęcony jest w całości omówieniu głównych metod procesualnych wykorzystywanych w badaniach nad procesem przekładu. Przyjmując perspektywę chronologiczną, autorka rozpoczyna prezentację od metody protokołów werbalnych, gdzie ze szczególną uwagą przygląda się metodom protokołów retrospektywnych oraz protokołów głośnego myślenia. Następnie omówiona zostaje metoda keyloggingu wraz z programem Translog II, a także podział procesu przekładu na trzy fazy (Jakobsen 2002). W sekcji 2.5 zaprezentowana zostaje metoda okulografii. Autorka omawia stanowiącą podstawę badań okulograficznych eye-mind hypothesis (Just i Carpenter 1980), a także podstawowe zmienne okulograficzne: fiksacje, sakkady oraz rozszerzanie źrenic. Rozdział zamyka prezentacja najnowszych metod neuroobrazowania, takich jak elektroencefalografia (EEG), pozytonowa tomografia emisyjna (PET) oraz funkcjonalne obrazowanie metodą rezonansu magnetycznego (fMRI). Ostatnia sekcja stanowi krótkie omówienie badań kwestionariuszowych.

Rozdział trzeci, a zarazem ostatni części przeglądowej porusza temat wysiłku kognitywnego. W sekcji pierwszej przedstawione zostały najpopularniejsze modele wysiłku kognitywnego, Effort Model (Gile 1995), Cognitive Load Model (Seeber 2011) oraz Cognitive Load Theory (Sweller i in. 1998). Kolejna sekcja poświęcona jest wskaźnikom wysiłku kognitywnego według klasyfikacji Chen i in. (2012) oraz Ehrenberger-Dow i in. (2020). W sekcji 3.4 autorka przedstawia aktualny stan badań na

temat wysiłku kognitywnego oraz kierunkowości. Natomiast ostatnia sekcja rozdziału trzeciego obejmuje novum badań nad wysiłkiem kognitywnym: zjawisko default translation (Halverson 2019). Jest ono definiowane jako krótka faza procesu tłumaczenia, charakteryzująca się niezakłóconym procesem powstawania tekstu.

Rozdział czwarty, będący pierwszym z rozdziałów analitycznych, stanowi prezentację zastosowanej metodologii oraz projektu badania eksperymentalnego. Autorka rozpoczyna rozdział od prezentacji celu pracy oraz uzasadnienia badania. Omówiona zostaje również problematyka przekładu kolokacji. Temat ten pozostaje dotąd niezbadany w odniesieniu do kierunkowości przekładu i wysiłku kognitywnego. Z uwagi na fakt że badanie w sposób uproszczony określane jest mianem eksperymentu przedstawiona zostaje charakterystyka badań eksperymentalnych oraz quasi-eksperymentalnych. Kolejne sekcje poświęcone są omówieniu hipotez i pytań badawczych, analizowanych zmiennych zależnych i niezależnych, a także przyjętych metod, wykorzystanych materiałów oraz problemów etycznych, z którymi autorka mierzyła się podczas eksperymentu. W ostatnich sekcjach rozdziału czwartego, autorka w sposób szczegółowy przedstawia przebieg badania pilotażowego oraz badania głównego.

Rozdział piąty poświęcony jest analizie uzyskanych wyników. Rozdział ten podzielony jest na sześć sekcji, z których pięć pierwszych odpowiada pięciu przyjętym hipotezom, natomiast w ostatniej sekcji omówione zostały wyniki regresji liniowej. W rozdziale piątym autorka prezentuje wyniki statystyki opisowej, dystrybucję zmiennych przedstawioną w postaci histogramów, wyniki testów statystycznych: parametrycznego T-testu studenta dla prób zależnych oraz jego nieparametrycznego odpowiednika, testu Wilcoxona.

W rozdziale szóstym poświęconym dyskusji nad uzyskanymi wynikami autorka podejmuje próbę odpowiedzenia na postawione pytania badawcze. W skład rozdziału wchodzą cztery sekcje odpowiadające czterem pytaniom badawczym. Formułując odpowiedzi na pytania badawcze autorka podsumowuje uzyskane wyniki, a także podejmuje próbę ich interpretacji w świetle literatury związanej z procesem przekładu, kierunkowością oraz wysiłkiem kognitywnym. Ostatni rozdział części empirycznej (rozdział 7) obejmuje podsumowanie uzyskanych wyników, wysunięte wnioski oraz sugestie dotyczące dalszych badań nad zjawiskiem wysiłku kognitywnego w kontekście kierunkowości.

W składy pracy wchodzi również bibliografia składająca się z 247 pozycji opublikowanych w latach 1980-2024, streszczenia w języku polskim i angielskim, 11 załączników zawierających materiały wykorzystane podczas badania eksperymentalnego, szczegółowe wyniki każdej z analizowanych zmiennych oraz podsumowanie wyników testów statystycznych.

REFERENCES

Alves, F. (2006). Cognitive adaptation in translation: an interface between language direction, time and recursiveness in target text production. *Letras de Hoje. Porto Algere*, 41(2), 241-272.

Alves, F. (2015). Translation process research at the interface. Paradigmatic, theoretical and methodological issues in dialogue with cognitive science, expertise studies, and psycholinguistics. In: A. Ferreira, J.W. Schwieter (eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting*. (pp. 17-40). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.115.02alv.

Alves, F. (ed.). (2003). *Triangulating Translation: Perspectives in Process Oriented Research.* Amsterdam; Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/btl.45.

Alves, F., Pagano, A., da Silva, I. (2009). A new window on translators‘ cognitive activity: methodological issues in the combined use of eye tracking, key logging and retrospective protocols. In: I. M. Mees, F. Alves, S. Göpferlich (eds.), *Methodology, Technology, and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. (pp. 267-292). Copenhagen: Samfundslitteratur.

Andermann, G., Rogers, M. (1998). What is that translation for? A functional view of translation assessment from a pedagogical perspective: a response to Hans G. Hönig. In: Ch. Scheaffner (ed.) *Translation and quality*, (pp. 56-63). Multilingual Matters Ltd: Clevedon, Philadelphia, Toronto, Sydney, Johannesburg. DOI: https://doi.org/10.1080/13520529709615479.

Apfelthaler, M. (2019). Directionality. In: M. Baker (ed.) *Routledge Encyclopaedia of Translation Studies, 3rd edition*. (pp. 152-158). London; New York: Routledge.

Balling, L. W., Hvelplund, K.T., Sjørup, A. C. (2014). Evidence of parallel processing during translation. *Meta Translator's Journal,* 59(2), 234-259. DOI: https://doi.org/10.7202/1027474ar.

Bartłomiejczyk, M. (2004). Simultaneous interpreting A-B vs. B-A from the interpreters' standpoint. In: G. Hansen, K. Malmkjær, D. Gile (eds.), *Claims, changes and challenges in Translation Studies: Selected contributions from the EST Congress, Copenhagen 2001*. (pp. 239-249). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.50.20bar.

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). Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Bartłomiejczyk, M., Gumul, E. (2024). Disfluencies and directionality in simultaneous interpreting: A corpus study comparing into-B and into-A interpretations from the European Parliament. *The International Journal for Translation and Interpreting Research*, 16(1), 38-53. DOI: https://doi.org/10.12807/ti.116201.2024.a03.

Beeby Lonsdale, A. (2001). Direction of translation (directionality). In: M. Baker (ed.), *Routledge Encyclopedia of Translation Studies*. (pp. 63-64). London; New York: Routledge.

Bernardini, S. (2001). Think-aloud protocols in translation research. *Target,* 13*(2)*, 241-263. DOI: https://doi.org/10.1075/target.13.2.03ber.

Bernardini, S. (2007). Collocations in translated language: combining parallel, comparable, and reference corpora. In: *Proceedings of the Corpus Linguistics Conference (CL2007)*, (pp. 1-16). Bologna.

Biel, Ł. (2007). Translation of multilingual EU legislation as a sub-genre of legal translation. In: D. Kierzkowska (ed.), *Court interpreting and legal translation in the enlarged Europe 2006.* (pp. 144-163). Warsaw: Translegis.

Biel, Ł. (2016). Tłumaczenie prawa unijnego a centralne pojęcie przekładoznawstwa. *Między oryginałem a przekładem*, 22 (1/31), 105-117. https://doi.org/10.12797/MOaP.22.2016.31.06.

Borghardt, L., Deilen, S., Fuchs, J., Gros, A-K., Hansen-Schirra, S., Nagels, A., Schiffl, L., Sommer, J. (2021). Neuroscientific research on the processing of Easy Language, *Frontiers in Communication*, 6(698044), 1-9. DOI: https://doi.org/10.3389/fcomm.2021.698044.

Brysbaert, M., Duyck, W. (2010). Is it the time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, 13(3), 359-371. DOI: https://doi.org/10.1017/S1366728909990344.

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

Buchowska, M. (2017). Tłumacz w instytucjach Unii Europejskiej: wyzwania współczesnej wieży Babel. *Rocznik Przekładoznawczy: Studia nad teorią,*

praktyką i dydaktyką przekładu, 12, 68-72. DOI: https://doi.org/10.12775/RP.2017.004

Carl, M. (2012). Translog- II: a programme for recording user activity data for empirical reading and writing research. In: Proceedings of the Eight International Conference on Language Resources and Evaluation, European Language Resources Association.

Carl, M., Schaeffer, M., Bangalore, S. (2016). The CRITT translation process research database. In: M. Carl, S. Bangalore, M. Schaeffer (eds.), *New Directions in Empirical Translation Process Research*. (pp. 14-56). Cham; Heidelberg; New York; Dordrecht; London: Springer. DOI: https://doi.org/10.1007/978-3-319-20358-4_2.

Carter, R. (1994). *Vocabulary: Applied linguistics perspective.* London and New York: Routledge.

Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R., Yin, B., Wang, Y. (2012). Multimodal behaviour and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 1-36. DOI: https://doi.org/10.1145/2395123.2395127.

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives*, 25(4), 640-657. DOI: https://doi.org/10.1080/0907676X.2016.1278026

Chen, S. (2018). A pen-eye-voice approach towards the process of note-taking and consecutive interpreting: An experimental design. *International Journal of Comparative Literature & Translation Studies*, 6(2), 1-8. DOI: https://doi.org/10.7575/aiac.ijclts.v.6n.2p.1.

Chen, S. (2020). The impact of directionality on the process and product in consecutive interpreting between Chinese and English: Evidence from pen recording and eye tracking. *The Journal of Specialised Translation*, 34, 100-117.

Chmiel, A., Janikowski, P., Cieślewicz, A. (2020). The eye or the ear? Source language interference in sight translation and simultaneous interpreting. *Interpreting*, 22(2), 187-210. DOI: https://doi.org/10.1075/intp.00043.chm.

Chmiel, A., Janikowski, P., Koržinek D., Lijewska, A., Kajzer-Wietrzny, M., Jakubowski, D., Plevoets, K. (2023). Lexical frequency modulates current cognitive load, but triggers no spillover effect in interpreting, *Perspectives*, 1-19. DOI: https://doi.org/10.1080/0907676X.2023.2218553.

Chmiel, A., Korzinek, D., Kajzer-Wietrzny, M., Janikowski, P., Jakubowski, D., Polakowska, D., (2022). Fluency parameters in the Polish Interpreting Corpus (PINC). In: M. Kajzer-Wietrzny, A. Ferraresi, I., Ivaska, S. Bernardini (eds.), *Empirical investigation into the forms of mediated discourse at the European Parliament*. (pp. 63-91). Berlin: Natural Science Press. DOI: https://doi.org/10.5281/zenodo.6976930.

Chmiel, A., Lijewska, A., (2022). Reading patterns, reformulation, and eye-voice span (IVS) in sight translation. *Translation and Interpreting Studies*, online first, 1-22. DOI: https://doi.org/10.1075/tis.21021.chm.

Chmiel, A., Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In: C. Way, S. Vandepitte, R. Meyleaerts, M. Bartłomiejczyk (eds.), *Tracks and Treks in Translation Studies: Selected papers from the EST Congress, Leuven 2010*. (pp. 189-205). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.108.10chm.

Chmiel. A. (2016). Directionality and context effects in word translation tasks performed by conference interpreters. *Poznań Studies in Contemporary Linguistics,* 52(2), 269-295, DOI: https://doi.org/10.1515/psicl-2016-0010.

Cohen, A. D., Honsenfield, C. (1981). Some uses of mentalistic data in second language research. *Language Learning*, 3(2), 285-313.

Cohen, L., Manion, L., Morrison K. (2007). *Research methods in education 6th ed*. London; New York: Routledge. DOI: https://doi.org/10.4324/9780203029053.

Collard, C. (2019). Predictors of ear-voice span, a corpus based study with special reference to sex. *Perspectives*, 27(3), 431-454. DOI: https://doi.org/10.1080/0907676X.2018.1553199.

Coppieters, R. (1987). Competence differences between native and non-native speakers. *Language*, 63(3), 544-573. DOI: https://doi.org/10.2307/415005.

Council of Europe. (2001). Common European Framework for Reference of Languages Learning, Teaching, Assessment. Cambridge University Press.

da Silva, , I. A. L., Alves F., Schmaltz M., Pagano A., Wong D., Chao L., Leal A. L. V., Quaresma P., Gracia C., da Silva G. E. (2017). Translation, post-editing and directionality: A study of effort in the Chinese-Portuguese language pair. In: A. L. Jakobsen, B. Mesa-Lao (eds.), *Translation in transition*. (pp. 91-117). Amsterdam; Philadelphia: Johns Benjamins. DOI: https://doi.org/10.1075/btl.133.04lou.

da Silva, I. A. L. (2015). On a more robust approach to triangulating retrospective protocols and keyogging in translation process research. In: A. Ferreira, J.W. Schwieter (eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting*. (pp. 175-201). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.115.08sil.

Davies, M. (2004). *British National Corpus (from Oxford University Press*). Available online at https://www.english-corpora.org/bnc/.

De Groot, A. (2002). Lexical representation and Lexical processing in the L2 user. In: V. Cook (ed.). *Portraits of the L2 user*. (pp. 32-63). Toronto: Multilingual Matters Ltd. DOI: https://doi.org/10.21832/9781853595851-004.

de Groot, A.M.B. (1992). Determinants of word translation. *Journal of Experimental psychology: Learning, memory, and cognition*, 26(5), 1283-1296. DOI: https://doi.org/10.1037/0278-7393.26.5.1283.

Doherty, S. (2018). Analysing variable relationships and time-course data in eye-tracking studies of translation processes and products. In: C. Walker, F. M. Federici (eds.), *Eye Tracking and Multidisciplinary Studies on Translation*. (pp. 71-94). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.143.05doh.

Dragsted, B. (2010). Coordination of reading and writing process in translation. An eye on uncharted territory. In: G. M. Shreve, E. Angelone (eds.), *Translation and cognition*. (pp. 41-62). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/ata.xv.04dra.

Dragsted, B. (2012). Indicators of difficulty in translation: Correlating product and process data. *Across Languages and cultures*, 13(1), 81-98. DOI: https://doi.org/10.1556/Acr.13.2012.1.5.

Dragsted, B. and Hansen, I.G. (2008). Comprehension and production in translation: A pilot study on segmentation and the coordination of reading and writing processes. In: S. Göpferlich, A. L. Jakobsen, I. M. Mees (eds.), *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*. (pp. 9-30). Copenhagen: Samfundslitteratur.

Duběda, T., Obdržálková, V. (2021). Stylistic competence in L2 translation: stylometry and error analysis. *The Interpreter and Translator Trainer*, 15(2), 172-186. DOI: https://doi.org/10.1080/1750399X.2020.1868175.

Duchowski, A.T. (2017). *Eye Tracking Methodology. Theory and practice 3rd edition*. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-57883-5.

Ehrensberger-Dow, M. Albi-Mikasa, M., Andermatt, K., Hunziker Heeb, A., Lehr, C. (2020). Cognitive load in processing ELF: Translators, interpreters, and other multilinguals. *Journal of English as a Lingua Franca*, 9(2), 217-238. DOI: https://doi.org/10.1515/jelf-2020-2039.

Englund Dimitrova, B. (2010). Translation process. In: Y. Gambier, L. van Doorslaer (eds.), *Handbook of translation studies.* (pp. 406-411). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/hts.2016.tra6.

Englund Dimitrova, 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. Göpferlich (eds.), *Methodology, technology and innovation in translation process research.* (pp. 109-134). Copenhagen: Samfundslitteratur.

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

Ericsson, K. A., Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture and Activity*, 5(3), 178-186. DOI: https://doi.org/10.1207/s15327884mca0503_3.

Ericsson, K.A., Simon, H.A. (1984/1993). *Protocol analysis: verbal reports as data*. Cambridge: The MIT Press.

Ferreira, A. (2014). Analysing recursiveness patterns and retrospective protocols of professional translators in L1 and L2 translation tasks. *Translation and Interpreting Studies*, 9(1), 109-127. DOI: https://doi.org/10.1075/tis.9.1.06fer.

Ferreira, A., Gries, S. Th., Schwieter, J. W. (2021). Assessing indicators of cognitive effort in professional translators: A study on language dominance and directionality. In: Tra&Co Group (ed.), *Translation, interpreting, cognition: The way out of the box.* (115-143). Berlin: Language Science Press. DOI: https://doi.org/10.5281/zenodo.4545041.

Ferreira, A., Schwieter, J. W. (2014).Underlying processes of L1 and L3 word translation: exploring the semantic relatedness effect. In: J. W. Schwieter, A. Ferreira (eds.) *The Development of Translation Competence: Theories and Methodologies from*

Psycholinguistics and Cognitive Science. (pp. 87-106). Newcastle upon Tyne: Cambridge Scholars Publishing.

Ferreira, A., Schwieter, J. W. (2017). Directionality in translation. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 90-105). Hoboken: Wiley Blackwell. DOI: https://doi.org/10.1002/9781119241485.ch5.

Ferreira, A., Schwieter, J. W., Gottardo, A., Jones, J. (2016). Cognitive effort in direct and inverse translation performance: Insight from eye-tracking technology. *Cadernos de Tradução*, 36(3), 60-80. DOI: https://doi.org/10.5007/2175-7968.2016v36n3p60.

Fielding, N., Fielding J. L. (1986). *Linking data*. Beverly Hills: SAGE Publications. DOI: https://doi.org/10.4135/9781412984775.

Fonseca, N. B., de Lima. (2015). Directionality in translation: investigating prototypical patterns in editing procedures. *The International Journal for Translation and Interpreting Research*, 7(1), 111-125. DOI: https://doi.org/ ti.106201.2015.a08.

Fröster Hegrenæs, C. (2018). *Translation competence development and the distribution of cognitive effort: an exploratory study of student translation behaviour*. Doctoral dissertation. NHH-Norwegian School of Economics.

Garcia, A. M. (2019). *The neurocognition of translation and interpreting*. Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.147.

Garcia, A. M., Ibáñez, A., Huepe, D., Houck, A. L., Michon, M., Lezama, C. G., Chadha, S., Rivera-Rei, Á. (2014). Word reading and translation in bilinguals: the impact of formal and informal translation expertise. *Frontiers in Psychology*, 5, 1-14. DOI: https://doi.org/10.3389/fpsyg.2014.01302.

Garcia, A. M., Mikulan, E., Ibáñez, A. (2016). A neuroscientific toolkit for translation studies. In: R. Muñoz Martín (ed.), *Reembedding Translation Process Research*. (pp.21-46). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.128.02gar.

Gieshof, A. C., Hunziker Heeb, A. (2022, June 15-17). Cognitive load and mental effort. Is there a difference and does that matter for task performance in T&I? [conference presentation]. *1st BTS Symposium on Cognitive Research in Translation and Interpreting*, Stockholm, Sweden.

Gieshoff, A.C. Hunziker Heeb, A. (2023). Cognitive load and cognitive effort. Probing the psychological reality of a conceptual difference. *Translation, cognition & behaviour*, 6(1), 3-28. DOI: https://doi.org/10.1075/tcb.00073.gie.

Gieshoff, A.C., Albl-Mikasa, M. (2022). Interpreting accuracy revised: a refined approach to interpreting performance analysis. *Perspectives*. DOI: https://doi.org/10.1080/0907676X.2022.2088296.

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam; Philadelphia: John Benjamins.

Gile, D., Lei. V. (2020). Translation, effort, and cognition. In: F. Alves, A. L. Jakobsen (eds.) *The Routledge Handbook of Translation and Cognition.* (pp. 263-278). London: Routledge.

Göpferlich, S. (2009). Towards a model of translation competence and its acquisition: the longitudinal study TransComp. In: A. L. Jakobsen, I. M. Mess, S. Göpferlich (eds.), *Behind the mind: methods, models and results in translation process research*. (pp. 11-37). Copenhagen: Samfundslitteratur.

Göpferlich, S., Jääskeläinen, R. (2009). Process research into the development of translation competence: where are we and where do we need to go? *Across Languages and Cultures. A Multidisciplinary Journal for Translation and Interpreting Studies*, 10(2), 169-191. DOI: https://doi.org/10.1556/Acr.10.2009.2.1.

Graves, A., Pascual Olaguíbel, M., Pearson, C. (2022). Conference interpreting in the European Union Institutions. In M. Albl-Mikasa, E. Tiselius (eds.) *The Routledge Handbook of Conference Interpreting*, (pp.104-114). New York. Routledge.

Grucza, S. (2011). Lingwistyka antropocentryczna a badania okulograficzne. *Lingwistyka Stosowana*, 4, 149-162.

Gumul, E. (2006). Conjunctive cohesion and the length of ear-voice span in simultaneous interpreting. A case of interpreting students. *Linguistica Silesiana*, 27, 93-103.

Gumul, E. (2017a). Explicitation and directionality in simultaneous interpreting. *Linguistica Silesiana*, 38, 311-329.

Gumul, E. (2018). Searching for evidence of Gile's Effort Models in retrospective protocols of trainee simultaneous interpreters. *Między oryginałem a przekładem*, 42(4), 17-39. DOI: https://doi.org/10.12797/MOaP.24.2018.42.02.

Gumul, E. (2019a). Evidence of cognitive effort in simultaneous interpreting: Process versus product data. *Beyond Philology*, 4, 11-45. DOI: https://doi.org/10.26881/bp.2019.4.01

Gumul, E. (2019b). O metodzie protokołów retrospektywnych w badaniach nad tłumaczeniem symultanicznym. *Rocznik Przekładoznawczy. Studia nad teorią, praktyką i dydaktyką przekładu,* 14, 171-192. DOI: https://doi.org/10.12797/MOaP.26.2020.50.02.

Gumul, E. (2020a). O metodzie badawczej zapożyczonej z psychologii, czyli jak zaprojektować i przeprowadzić badanie za pomocą metody protokołów retrospektywnych w tłumaczeniu symultanicznym. *Między Oryginałem a Przekładem,* 5(20), 31-54. DOI: https://doi.org/10.12797/MOaP.26.2020.50.02.

Gumul, E. (2020b). Retrospective protocols in simultaneous interpreting: Testing the effect of retrieval cues. *Linguistica Antverpiensia, New Series: Themes in Translation Studies,* 19, 152-171. DOI: https://doi.org/10.52034/lanstts.v19i0.547.

Gumul, E. (2021b). Reporting stress in simultaneous interpreting. The analysis of trainee interpreters' retrospective reports and outputs. *Onomázein*. *Journal of linguistics, philology and translation. Special issue – Emotions in translation and interpreting*, 8, 16-42. DOI: https://doi.org/10.7764/onomazein.ne8.04.

Gumul, E. (2021c). Rola bodźca w retrospekcji z perspektywy tłumaczy symultanicznych. *Rocznik Przekładoznawczy. Studia nad teorią, praktyką i dydaktyką przekładu,* 16, 79-106. DOI: http://dx.doi.org/10.12775/RP.2021.004.

Gumul, E. Herring, R. (2022). Employing retrospective process tracing in an international exploratory study. *Translation Cognition & Behaviour*, 5(2), 221-249. DOI: https://doi.org/10.1075/tcb.00067.gum.

Gumul, E., Pietryga, M. (manuscript under preparation). What do trainee interpreters look at when they explicitate? An eye-tracking and retrospective study of explicitation in sight interpreting/translation (SiT).

Gumul. E. (2017b). *Explicitation in simultaneous interpreting. A Study into expliciatating behaviour of trainee interpreters*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Gumul. E. (2021a). Explicitation and cognitive load in simultaneous interpreting. Product- and process-oriented analysis of trainee interpreters' output. *Interpreting*, 23(1), 45-75. DOI: https://doi.org/10.1075/intp.00051.gum.

Halverson, S., L. (2015). Cognitive Translation Studies and the merging of empirical paradigms: The case of 'literal translation.' *Translation Spaces*, 4(2), 310-340. DOI: https://doi.org/10.1075/ts.4.2.07hal.

Halverson, S., L. (2017). Multimethod approaches. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 195-212). Hoboken: Wiley Blackwell. DOI: https://doi.org/10.1002/9781119241485.ch11.

Halverson, S., L. (2019). 'Default' translation: A construct for cognitive translation and interpreting studies. *Translation, Cognition, and Behaviour*, 2(2), 187-210. DOI: https://doi.org/10.1075/tcb.00023.hal.

Hammarberg, B. (2014). Problems in defining the concepts of L1, L2 and L3. In A. Otwinkowska, G. De Angelis (eds.) *Teaching and Learning in Multilingual Contexts. Sociolinguistic and Educational Perspectives*, (pp. 3-18). Bristol, Buffalo, Toronto: Multilingual Matters. DOI: https://doi.org/10.21832/9781783091263-003.

Hansen-Schirra, S. (2017). EEG and universal language processing in translation. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 232-248). Hoboken: Wiley Blackwell. DOI: https://doi.org/10.1002/9781119241485.ch13.

Hart, S.G., Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock, N. Meshkati (eds.) Human Mental Workload, (pp. 139-183). Amsterdam: North-Holland. DOI: https://doi.org/10.1016/S0166-4115(08)62386-9.

Hegernæs, C., F., Halverson, S., L. (2022, June 15-17). What characterises default translation? An (explorative) study of default translation segments. [conference presentation]. *1st BTS Symposium on Cognitive Research in Translation and Interpreting*, Stockholm, Sweden.

Herring, R. (2018). "*I could only think about what I was doing, and that was a lot to think about": online self-regulation in dialogue interpreting*. Doctoral dissertation. University of Geneva. DOI: https://doi.org/10.13097/ARCHIVE-OUVERTE/UNIGE:108626.

Herring, R., Tiselius, E. (2020). Making the most of retrospective process tracing in dialogue interpreting research. *FITISPOS International Journal*, 7(1), 53-71.

Hervais-Adelman, A., Babcock, L. (2019). The neurobiology of simultaneous interpreting: Where extreme language control and cognitive control intersect.

Bilingualism: Language and Cognition, 23 (4), 740-751. DOI: https://doi.org/10.1017/S1366728919000324.

Hervais-Adelman, A., Moser-Mecer, B., Golestani, N. (2011). Executive control of language in the bilingual brain: integrating the evidence from neuroimaging to neuropsychology. *Frontiers in Psychology*, 2(234), 1-8. DOI: https://doi.org/10.3389/fpsyg.2011.00234.

Hervais-Adelman, A., Moser-Mecer, B., Michel, Ch. M., Golestani, N. (2015). fMRI of simultaneous interpreting reveals the neural basis of extreme language control. *Cerebral Cortex*, 25(12), 4727-4739. DOI: https://doi.org/10.1093/cercor/bhu158.

Hunziker Heeb, A. (2020). *Effects of directionality on professional translators' processes, products and effort*. Hildesheim: Universitätsverlag Hildesheim.

Hunziker Heeb, A., Lehr, C., Ehrensberger-Dow, M. (2021). Situated translators: Cognitive load and the role of emotions. In: R. Muñoz, Martin, S. Sun, D. Li (eds.), *Advances in Cognitive Translation Studies*. (pp. 48-65). Cham: Springer. DOI: https://doi.org/10.1007/978-981-16-2070-6_3.

Hvelplund, K. T. (2017a). Eye-tracking in translation process research. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 248-265). Hoboken: Wiley Blackwell. DOI: https://doi.org/DOI:10.1002/9781119241485.ch14.

Hvelplund, K. T. (2017b). Four fundamental types of reading during translation. In: A. L. Jakobsen, B. Mesa-Lao (eds.), *Translation in Transition: Between cognition, computing and technology*. (pp. 56-78). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.133.02hve.

Hvelplund, K. T.(2016). Cognitive efficiency in translation. In: R. Muñoz, Martin (ed.) *Reembedding Translation Process Research*. (pp. 149-169). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.128.08hve.

Hyönä, J., Lorch Jr., R. F., Rinck, M. (2003). Eye movement measures to study global text processing. In: J. Hyönä, R. Radach, H. Deubel (eds.), *The Mind's Eye – Cognitive and Applied aspects of Eye Movement Research*. (pp. 313-334). Amsterdam: Elseveier. DOI: https://doi.org/10.1016/B978-044451020-4/50018-9.

Immonen, S. (2006). Translation as a writing process. Pauses in translation versus monolingual text production. *Target*, 18(2), 313-335. DOI: https://doi.org/10.1075/target.18.2.06imm.

Immonen, S., Mäkisalo, J. (2010). Pauses reflecting the processing of units in monolingual text production and translation. *Hermes – Journal of Language and Communication Studies, 44,* 45-61. DOI: https://doi.org/10.7146/hjlcb.v23i44.97266.

Ivanova, A. (2000). The use of retrospection in research on simultaneous interpreting. In: S. Trikkonen-Condit, R. Jääskeläinen (eds.), *Tapping and mapping the process of translation and interpreting: Outlooks on empirical research*. (pp. 27-52). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.37.05iva.

Jääskeläinen, R. (1993). Investigating translation strategies. In: S. Trikkonen-Condit, J. Laffling (eds.), *Recent trends in empirical translation process research.* (pp. 99-120). Joensuu: University of Joensuu.

Jääskeläinen, R. (2000). Focus on methodology in think-aloud studies on translating. In: S. Trikkonen-Condit, R. Jääskeläinen (eds.), *Tapping and mapping the process of translation and interpreting: Outlooks on empirical research*. (pp. 71-82). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Jääskeläinen, R. (2001). Think-aloud protocols. In: M. Baker (ed.), *Routledge encyclopaedia of translation studies*. (pp. 265-268). London; New York: Routledge.

Jääskeläinen, R. (2010). Think-aloud protocol. In: Y. Gambier, L. van Doorslaer (eds.), *Handbook of translation studies.* (pp. 371-373). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/hts.1.thi1.

Jääskeläinen, R. (2017). Verbal Reports. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition.* (pp. 213-232). Hoboken: Wiley Blackwell. DOI: https://doi.org/10.1002/9781119241485.ch12.

Jakobsen, A. L. (1999). Logging target text production with Translog. In: G. Hansen. (ed.), *Probing the Process in Translation: Methods and Results*. (pp. 9-20). Copenhagen: Samfundslitteratur.

Jakobsen, A. L. (2002). Translation drafting by professional translators and by translation students. In: G. Hansen (ed.), *Empirical translation studies. Process and product*. (pp. 191-204). Copenhagen: Samfundslitteratur.

Jakobsen, A. L. (2003). Effects of think aloud on translation, speed, revision and segmentation. In: F. Alves (ed.), *Triangulating translation. Perspectives in process oriented research*. (pp. 69-95). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.45.08jak.

Jakobsen, A. L. (2006). Research methods in translation – Translog. In: K. P. H. Sullivan, E. Lindgren (eds.), *Computer keystroke logging and writing*: *Methods and applications*. (pp. 95-106). Amsterdam: Elsevier.

Jakobsen, A. L. (2011). Tracking translators' keystrokes and eye movements with Translog. In: C. Alvstad, A. Hild, E. Tiselius (eds.), *Methods and Strategies of Process Research: Integrative approaches in translation studies*. (pp. 37-55). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.94.06jak.

Jakobsen, A. L., Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. In: S. Göpferlich, A. L. Jakobsen, I. M. Mees (eds.), *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing.* (pp. 103-124). Copenhagen: Samfundslitteratur.

Jakobsen, A. L., Schou, L (1999). Translog documentation. In: G. Hansen (ed.), *Probing the Process in Translation: Methods and Results*. (pp. 151-186). Copenhagen: Samfundslitteratur.

Jankowiak, K., Lehka-Paul O. (2021). Novel metaphor translation is modulated by translation direction. *Applied Psycholinguistics*, 43(1), 1-16. DOI: https://doi.org/10.1017/S0142716421000461.

Janyan, A., Popivanov, I., Andonova, E. (2009). Concreteness effect and word cognate status: ERPs in single word translation. In: K. Alter, M. Horne, M. Lindgren, M. Roll, J. von Kos Torkildsen (eds.), *Brain Talk: Discourse with and in the brain*. (pp. 21-30). Lund: Lundus Universitet.

Jensen, K. T. H., Sjørup, A. C., Balling, L. W. (2009). Effects of L1 syntax on L2 translation. In: I. M. Mees, F. Alves, S. Göpferlich (eds.), *Methodology, Technology, and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. (pp. 319-336). Copenhagen: Samfundslitteratur.

Just, M. A., Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354. DOI: https://doi.org/10.1037/0033-295X.87.4.329.

Kajzer-Wietrzny, M., Whyatt, B., Stachowiak, K. (2016). Stylistic simplification in inter- and intralingual translation – combining corpus linguistics, key logging and eye-tracking. *Poznań Studies in Contemporary Linguistics*, 52(2), 235-267. DOI: https:/doi.org/10.1515/psicl-2016-0009.

Kemmerer, D. (2015). *Cognitive neuroscience of language*. New York; London: Psychology Press. Taylor & Francis Group. DOI: https://doi.org/10.4324/9781315764061.

Korpal, P. (2015). Eye-tracking in Translation and Interpreting Studies: The growing popularity and methodological problems. In: Ł. Bogucki, M. Deckert (eds.), *Accessing audiovisual translation*. (pp.199-212). Łódź: Peter Lang.

Korpal, P. (2016a). Interpreting as a stressful activity: physiological measures of stress in simultaneous interpreting. *Poznań Studies in Contemporary Linguistics,* 52(2), 297-316. DOI: https://doi.org/10.1515/psicl-2016-0011.

Korpal, P. (2016b). *Linguistic and psychological indicators of stress in simultaneous interpreting*. PhD thesis. Poznań.

Korpal, P., Jankowiak, K. (2021). On the potential impact of directionality on emotion processing in interpreting. *Onomázein*. *Journal of linguistics, philology and translation. Special issue – Emotions in translation and interpreting*, 8, 43-60. DOI: https://doi.org/10.7764/onomazein.ne8.05.

Kościuczuk, T. (2016). L2 translation: to teach or to discourage. *Journal of Translator Education and Translation Studies*, 1(1), 4-22.

Krings, H. P. (2005). Wege in Labyrinth – Fragestellungen und Methoden der Übersetzungsprocessforschung in Überblick. *Meta Translator's Journal,* 50(2), 342-358. DOI: https://doi.org/10.7202/010941ar.

Kroll, J. F., Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In: M. Gruneberg, P. Moris, R. Sykes (red.), *Practical aspects of memory vol 2*. (pp. 389-395). London: John Wiley and Sons.

Kroll, J. F., Michael E., Tokowicz, N., Dufour, R. (2002). The developmental of lexical fluency in a second language. *Second Language Research*, 18(2), 137-171. DOI: https://doi.org/10.1191/0267658302sr201oa.

Kroll, J. F., Steward. E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.

Kroll, J. F., van Hell, J. G., Tokowicz, N., Green, D. W. (2010). The Revised Hierarchical Model: A critical Review and assessment. *Bilingualism: Language and Cognition*, 13(3), 373-381. DOI: https://doi.org/10.1017/S136672891000009X.

Kruger, H. (2016). What's happening when nothing's happening? Combining eyetracking and keylogging to explore cognitive processes during pauses in translation production. *Across Languages and Cultures*, 17(1), 25-52. DOI: https://doi.org/10.1556/084.2016.17.1.2.

Kumpulainen, M. (2015). On the operationalisation of 'pauses' in translation process research. *The International Journal for Translation and Interpreting Research*, 7(1), 47-58.

Kussmaul, P., Trikkonen-Condit, S. (1995). Think-Aloud Protocol Analysis in Translation Studies. *TTR*, 8(1), 177-199.

Lacruz, I. (2017). Cognitive effort in translation, editing and post editing. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 386-401). Hoboken: Wiley Blackwell. DOI: https://doi.org.10.1002/9781119241485.CH21.

Lauffer, S. (2002). The translation process: an analysis of observational methodology. *CAD. De Traducao*, 2(10), 59-75.

Leijten, M., Van Waes, L. (2006). Inputlog: New perspectives on the logging on-line writing processes in a Windows environment. In: K. P. H. Sullivan, E. Lindgren (eds.), *Computer keystroke logging and writing*: Methods and applications. (pp. 73-93). Amsterdam: Elsevier. DOI: https://doi.org/10.1163/9780080460932_006.

Leijten, M., Van Waes, L. (2013). Keystroke logging in writing research: using Inputlog to analyse writing process. *Written Communication*, 30(3), 358-392. DOI: https://doi.org/10.1177/0741088313491692.

Lemhöfer, K., Broersma,M. (2012). Introducing LexTale: a quick and valid lexical test for advanced learners of English. *Behaviour Research Methods*, 44, 325-343. DOI: https://doi.org/10.3758/s13428-011-0146-0.

Lewis, M. (ed.) (2000). *Teaching collocation. Further developments in lexical approach.* Hove: LTP Teacher Training.

Lijewska, A., Chmiel, A., Koržinek, D., Dutka, Ł., Marasek, K., Szarkowska, A., Brocki, Ł. (2017). Ear-voice span and pauses in intra- and interlingual: An exploratory study into temporal aspects of the respeaking process. *Applied Psycholinguistics*, 38(5), 1201-1227. DOI: https://doi.org/10.1017/S0142716417000108.

Liu, M. (2010). Methodology in interpreting studies. A methodological review of evidence-based research. In: B. Nicodemus, L. Swabey (eds.), *Advances in Interpreting Research*, (pp.85-120), Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.99.08liu.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-285. DOI: https://doi.org/10.1017/S0272263100009165.

Lörscher, W. (1992). Process-oriented research into translation and implications for translation teaching. *TTR*, 5(1), 145-161. DOI: https://doi.org/10.7202/037110ar.

Malkiel, B. (2009). From Ántonia to my Ántonia: Tracking self-corrections with Translog. In: S. Göpferlich, A. L. Jakobsen, I. M. Mees (eds.), *Behind the mind: Methods, models, and results in translation process research*. (pp. 149-166). Copenhagen: Samfundslitteratur Press.

Mead, P. (2002). Exploring hesitations in consecutive interpreting: An empirical study. In: G. Garzone, M. Viezzi (eds.), *Interpreting in the 21st century: Challenges and opportunities*. (pp. 73-82). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.43.08mea.

Mellinger, Ch., D., Hanson, T., A. (2020). Methodological considerations for survey research: Validity, reliability, and quantitative analysis. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 19, 172-190. DOI: https://doi.org/10.52034/lanstts.v19i0.549.

Mikolič Južnič, T., Zlatnar Moe, M., Žigon, T. (2021). Literary translators for languages of low diffusion: market needs and training challenges in Slovenia. *The Interpreter and Translator Trainer*, 15(2), 243-259. DOI: https://doi.org/10.1080/1750399X.2020.1868174.

Miller, K. S., Sullivan, K. P.H. (2006). Keystroke logging: an introduction In: K. P. H. Sullivan, E. Lindgren (eds.), *Computer keystroke logging and writing*: Methods and applications. (pp. 1-10). Amsterdam: Elsevier. DOI: https://doi.org/10.1163/9780080460932_002.

Mraček, D. (2018). Inverse translation: the more challenging direction. *Linguistica Pragensia*, 2, 202-221.

Mraček, D. (2019). The role(s) of native speakers in L2 translation. *Studie z Aplikovaně Lingvistiky*, 1, 7-25.

Mueller, S. T., Piper, B. N. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of neuroscience method*, 222, 250-259. DOI: https://doi.org/10.1016/j.jneumeth.2013.10.024.

Muñoz, Martin, R. (2009). The way they were: subject profiling in translation process research. In: I. M. Mees, F. Alves, S. Göpferlich (eds.), *Methodology, Technology, and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. (pp. 87-108). Copenhagen: Samfundslitteratur.

Muñoz, Martin, R. (2017). Looking towards the future of cognitive translation studies. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 555-572). Hoboken: Wiley Blackwell. DOI: https://doi.org/10.1002/9781119241485.ch30.

Muñoz, Martin, R. (2021, June 14-25). It took 50 years to look this good. Situating cognitive translation and interpreting studies. [Summer School presentation]. *First International Summer School on Cognitive Translation & Interpreting Studies, Forli, Italy.*

Muñoz, Martin, R., Apfelthaler, M. (2022). A task segment framework to study keylogged translation processes. *The International Journal for Translation and Interpreting Research*, 14(2), 8-31. DOI: https://doi.org/10.12807/ti.114202.2022.a02.

Muñoz, Martin, R., Cardona, Guerra, J. M. (2019). Translating in fits and starts: Pause thresholds and roles in the research of translation process. *Perspectives*, 27(4), 525-551. DOI: https://doi.org/10.1080/0907676X.2018.1531897.

Newmark, P. (1988). *A Textbook of Translation.* New York: Prentice Hall.

Newmark, Peter. (1981). *Approaches to translation*. Oxford, New York, Toronto, Sydney, Paris, Frankfurt: Pergamon Press.

Nicodemus, B., Emmorey, K. (2015). Direction asymmetries in spoken and signed language interpreting. Bilingualism, Language and Cognition 16(3): 624-636. https://doi.org/10.1017/S1366728912000521.

O'Brien, S. (2008). Processing fuzzy matches in Translation Memory tool: an eye tracking analysis. In: S. Göpferich, A. L. Jakobsen, M. Inger (eds.), *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. (pp. 79-102). Copenhagen: Samfundslitteratur Press.

O'Brien, S. (2009). Eye tracking in translation process research: methodological challenges and solutions. In: I. M. Mees, F. Alves, S. Göpferlich (eds.),

Methodology, technology and innovation in translation process research. (pp. 251-266). Frederiksberg: Samfundslitteratur.

PACTE. (2000). Acquiring Translation Competence: Hypotheses and methodological problems in a research project. In: A. Beeby, D. Ensinger and M. Presas (eds.), *Investigating Translation.* (pp. 99-106). Amsterdam; Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/btl.32.

PACTE. (2003). Building a translation competence model. In: F. Aves (ed.), *Triangulating translation: perspectives in process oriented research.* (pp. 43-66). Amsterdam; Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/btl.45.06pac.

Pagano, S., Magalhães, C., Alves, F. (2004). Towards the construction of a multilingual, multifunctional corpus: factors in the design and application of CORIDALL. *TradTerm,* 10, 143-161. DOI: https://doi.org/ 10.11606/issn.2317-9511.tradterm.2004.47160.

Pavlovič N., Jensen, K. T. H. (2009). Eye tracking translation directionality. In: A. Pym, A. Perekrestenko (eds.), *Translation Research Projects 2*. (pp. 93-109). Tarragona: Intercultural Studies Group.

Pavlovič Natasa. 2007a. Directionality in translation and interpreting. Report on a questionnaire survey in Croatia. *International Journal of Interpretation and Translation 5,* 77-99.

Pavlovič, N. (2007b). *Directionality in collaborative translation processes. A study of novice translators.* Doctoral dissertations. Universitat Rovira i Virgili, Spain. Universitat of Zagreb, Croatia.

Pavlovič, N. (2010). What were they thinking?! Students' decision making in L1 and L2 translation process. *Hermes – Journal of Language and Communication Studies* 44, 63-87.

Pellicer- Sánchez, A., Siyanova-Chanturia, A., Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics,* 43, 727-756. DOI: https://doi.org/10.1017/S014271642200011X.

Pęzik, P. (2012). Wyszukiwarka PELCRA dla danych NKJP. In: A. Przepiórkowski, M. Bańko, R. Górski, B. Lewandowska-Tomaszczyk (eds.), Narodowy Korpus Języka Polskiego. (pp. 253-274). Warszawa: Wydawnictwo Naukowe PWN.

Pietryga, M. (2022). Perception of directionality among translation students. *Acta Neophilologica*, 24:1, 87-109. DOI: https://doi.org/10.31648/an.6807.

Pietryga, M. (under review). Eye-key span as a measure of cognitive effort in translation. A study on the influence of directionality on cognitive effort.

Piotrowska, M. (2000). Protokoły głośnego myślenia – TAPs. In: U. Dąbska-Prokop (ed), *Mała encyklopedia przekładoznawstwa.* (pp.163-166). Częstochowa: Wydawnictwo Wyższej Szkoły Języków Obcych i Ekonomii.

Piotrowska, M. (2002). *A compensation model for strategy and techniques in teaching translation*. Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej.

Piotrowska, M. (2007). *Proces decyzyjny tłumacza.* Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej. DOI: https://doi.org/10.7146/hjlcb.v23i44.97267.

Płużczyka, M. (2019a). Eyetracking jako poszerzenie translatorycznej perspektywy badawczej. In: J. Lubocha-Kruglik, O. Małysa, G. Wilk (eds.), *Przestrzenie przekładu 3*. (pp. 13-28). Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Płużyczka, M. (2011). Okulograficzne wsparcie badań nad procesem tłumaczenia avista. *Lingwistyka Stosowana,* 4, 181-189.

Płużyczka, M. (2013). Okulograficzne spojrzenie na trudności translacyjne. *Rocznik Przekładoznawczy. Studia nad teorią, praktyką i dydaktyką przekładu,* 8, 59-76. DOI: https://doi.org/10.12775/RP.2013.004.

Płużyczka, M. (2016). Przestrzenne ruchy sakkadowe a pamięć długotrwała. *Lingwistyka Stosowana*, 20(5), 101-118.

Płużyczka, M. (2018). The first hundred years: a history of eye tracking as a research method. *Applied Linguistic Papers*, 25(4), 101-116.

Płużyczka, M. (2019b). Weryfikując modele mentalne tłumaczenia: atuty i pułapki okulografii. *Studia Translatorica*, 10, 41-63. DOI: https//:doi.org/10.23817/strans.10-3.

Pokorn, N. K. (2004). Challenging the myth of native speaker competence in translation theory. In: G. Hansen, K. Malmkjær, D. Gile (eds.), *Claims, Changes and Challenges in Translation Studies: Selected contributions from the EST Congress, Copenhagen 2001*. (pp. 113-124). Amsterdam; Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/btl.50.10pok.

Pokorn, N. K. (2005). *Challenging the traditional axioms. Translation into a non-mother tongue.* Amsterdam; Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/btl.62.

Pokorn, N. K. (2011). Directionality. In: Y. Gambier, L. van Doorslaer (eds.), *Handbook of translation studies.* (pp. 37-39). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/hts.2.dir1.

Potter, M. C., Kwok-Fai, S., von Eckardt, B., Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behaviour*, 23, 23-38. DOI: https://doi.org/10.1016/S0022-5371(84)90489-4.

Przybylska, R. (2020). Kolokacje a analiza semantyczna wyrazu, *LingVaria* 2(30), 43-51. DOI: https://doi.org/10.12797/LV.15.2020.30.04.

Religa, L. (2009). O pojęciu kolokacji słów kilka. In: M. Łukasik (ed.) *Debiuty naukowe 3. Leksykon-tekst-wyraz*. Warszawa: Katedra Języków Specjalistycznych Uniwersytetu Warszawskiego.

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Teräs, M., Sipilä, H., Sunnari, M. (2000). The translating brain: cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters*, 294, 85-88. DOI: https://doi.org/10.1016/s0304-3940(00)01540-8.

Rodríguez-Inés, P. (2022). Directionality *@ ENTI (Encyclopedia of translation and interpreting*). AIETI. DOI: https://doi.org/10.5281/zenodo.6366327.

Rojo López, A. M., Foulquié-Rubio, A-I., Espín López, L., Martínez Sánchez, F. (2021). Analysis of speech rhythm and heart rate as indicators of stress on student interpreters. *Perspectives*, 29(4), 591-607. DOI: https://doi.org/10.1080/0907676X.2021.1900305.

Russell, D., Winston, B. (2014). Tapping into the interpreting process: Using participant's reports to inform the interpreting process in educational settings. *The International Journal for Translation and Interpreting Research,* 6(1), 102-127. DOI: https://doi.org/10.12807/ti.106201.2014.a06.

Saldanha, G., O'Brien, S. (2013). *Research methodologies in translation studies*. London; New York: Routledge. DOI: https://doi.org/10.4324/9781315760100.

Schaeffer, M., Carl, M. (2013). Shared representations and the translation process. A recursive model. *Translation and Interpreting Studies*, 8(2), 169-190. DOI: https://doi.org/10.1075/tis.8.2.03sch.

Schaeffer, M., Carl, M. (2014). Measuring the cognitive effort in literal translation process. In: *Proceedings of the EACL 2014 Workshop on Humans and Computer-*

assisted Translation, (pp. 29–37), Gothenburg. Association for Computational Linguistics. DOI: https://doi.org/10.3115/v1/W14-0306.

Schaeffer, M., Carl, M. (2017). Why translation is difficult? A corpus-based study of non-literality in post-editing and from-scratch translations. *Hermes – Journal of Language and Communication in Business*, 56, 43-57. DOI: https://doi.org/10.7146/hjlcb.v0i56.97201.

Schaeffer, M., Dragsted, B., Hvelplund, K., T., Balling, L., W., Carl., M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In: M. Carl, S. Bangalore, M. Schaeffer (eds.) *New Directions in Empirical Translation Process Research*. (pp. 183-209). Cham; Heidelberg; New York; Dordrecht; London: Springer. DOI: https://doi.org/10.1007/978-3-319-20358-4_9.

Schou, L., Dragsted, B., Carl, M. (2009). Ten years of Translog. In: I. M. Mees, F. Alves, S. Göpferlich (eds.), *Methodology, Technology, and Innovation in Translation Process Research. A Tribute to Arnt Lykke Jakobsen*. (pp. 37-50). Copenhagen: Samfundslitteratur.

Seeber, K. G. (2011). Cognitive Load in simultaneous interpreting: Existing theories – new models, *Interpreting*, 13(2), 176-204. DOI: https://doi.org/10.1075/intp.13.2.02see.

Seeber, K., G. (2013). Cognitive load in simultaneous interpreting. Measures and methods. *Target*, 25(1), 18-32. DOI: https://doi.org/10.1075/target.25.1.03see.

Seleskovitch, D., Lederer, M. (1989). *Pédagogie raisonnée de l'interprétation.* Paris: Didier.

Serbina, T., Niemietz, O., Neuman, S. (2015). Development of a keystroke logged translation corpus. In: C. Fantinuoli, F. Zanettin (eds.), *New direction in corpus-based translation studies*. (pp. 11-34). Berlin: Language Science Press.

Shamy, M., de Pedro Ricoy, R. (2017). Retrospective protocols: Tapping into the minds of interpreting trainees. *The International Journal for Translation and Interpreting Research*, 9(1), 51-71.

Sharmin, S., Špakov, O., Karri-Joukko, R., Jakobsen, A. L. (2008). Where and for how long do translators look at the screen while translating? . In: S. Göpferlich, A. L. Jakobsen, I. M. Mees (eds.), *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing.* (pp. 31-52). Copenhagen: Samfundslitteratur

Shreve, G. M., Lacruz, I. (2017). Aspects of a cognitive model of translation. In: J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*. (pp. 127-143). Hoboken: Wiley Blackwell. DOI :https://doi.org/10.1002/9781119241485.ch7.

Sinclair, J., Jones, S., Daley, R. (2005). *English collocation studies: The OSTI report*. London: Continuum.

Sjørup, A. C. (2013). *Cognitive Effort in Metaphor Translation: An Eye-tracking and key-logging Study, PhD series*. Copenhagen Business School (CBS): Frederiksberg.

Sonbul, S., El-Dakhs, D., A., S., Al-Otabi, H. (2022). Translation competence and collocation knowledge: Do congruency and word type have an effect on the accuracy of collocations in translation. *The Interpreter and Translator Trainer*, 16(4), 409-427. DOI: https://doi.org/10.1080/1750399X.2022.2084251.

Stachowiak-Szymczak, K. (2019). *Eye movements and gestures in simultaneous and cognitive interpreting.* Cham: Springer.

Stachowiak-Szymczak, K., Korpal, P. (2019). Interpreting accuracy and visual processing of numbers in professional and student interpreters. An eye-tracking study. *Across languages and cultures*, 20(2), 235-251. DOI: https://doi.org/10.1556/084.2019.20.2.5.

Sun, S. (2011). Think-Aloud-Based Translation Process Research: Some Methodological Considerations. *Meta Translator's Journal*, 56(4), 928-951.

Sun, S. (2015). Measuring translation difficulty: theoretical and methodological considerations. *Across languages and cultures*, 16(1), 29-54. DOI: https://doi.org/10.1556/084.2015.16.1.2.

Sun, S., Li, T., Zhou, X. (2020). Effect of think aloud on cognitive effort in translation. *Linguistica Adverpiensia, New Series: Themes in Translation Studies*, 19, 132-151. DOI: https://doi.org/10.52034/lanstts.v19i0.556.

Sun, S., Shreve, G. M. (2014). Measuring translation difficulty: An empirical study, Target, 26(1), 98-127. DOI: https://doi.org/10.1075/target.26.1.04sun.

Sweller, J., van Marrienboer, J. J. G., Paas, F. G. W. C, (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296.

Teixeira, C. S. C., O'Brien, S. (2017). Investigating the cognitive ergonomic aspect of translation tools in a workplace setting. *Translation Spaces*, 6(1), 79-103. DOI: https://doi.org/10.1075/ts.6.1.05tei.

Teubert, W. (2004). Units of meaning, parallel corpora and their implications for language teaching. In: U. Connor, T. A. Upton (eds.), *Applied Corpus Linguistics. A multidimensional perspective*, (pp.171-190). Amsterdam, New York: Rodopi.

Texeira, C. S. C., O'Brien, S. (2018). Overcoming methodological challenges of eye tracking in the translation workplace. In: C. Walker, F. M. Federici (eds.), *Eye Tracking and Multidisciplinary Studies on Translation*. (pp. 33-54). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.143.03tei.

Timarová, Š., Dragsted, B., Hansen, I.G. (2011). Time lag in translation and interpreting. A methodological exploration. In: C. Alvstad, A. Hild, E. Tiselius (eds.), *Methods and strategies of process research: Integrative approaches in translation studies*. (pp. 121-146).). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.94.10tim.

Tiselius, E., Jenset, G. B. (2011). Process and product in simultaneous interpreting. Why they tell us about experience and expertise. In: C. Alvstad, A. Hild, E. Tiselius (eds.), *Methods and strategies of process research: Integrative approach in translation studies*. (pp. 269-300). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.94.20tis.

Tiselius, E., Sneed, K. (2020). Gaze and eye movement in dialogue interpreting: An eye-tracking study. *Bilingualism: Language and Cognition*, 23, 780-787. DOI: https://doi.org/10.1017/S1366728920000309.

Tomaszkiewicz, T. (2013). Tłumaczenie w tandemach: nowe możliwości w dydaktyce tłumaczeniowej. *Między Oryginałem a Przekładem*, 19-20, 121-140. DOI: https://doi.org/10.12797/MOaP.19.2013.19-20.09.

Tomczak, E., Whyatt, B. (2022). Directionality and lexical selection in professional translators: Evidence from verbal fluency and translation tasks. *The International Journal for Translation and Interpreting Research*, 14(2), 120-136. DOI: https://doi.org/10.12807/ti.114202.2022.a08.

Trikkonen-Condit, S. (2005). The monitor model revised: Evidence from process research. *META*, 50(2), 405-414. DOI: https://doi.org/10.7202/010990ar.

Venuti, L. (1995). *The Translator's Invisibility. A history of translation.* London; New York: Routledge.

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). Amsterdam; Philadelphia: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/btl.43.07vik.

Vinay, J., P., Darbelnet, J. (1958). *Stylistique comparée du français et de l'anglais*. París: Georgetown University Press.

Vottonen, E., Kujamäki, M. (2021). On what grounds? Justifications of students translators for their translation solutions. *The Interpreter and Translator Trainer*, 1-20. DOI: https://doi.org/10.1080/1750399X.2021.1891514.

Wenchao, S. (2023). Eye-voice span in sight interpreting: an eye-tracking investigation. *Perspectives,* 1-17. DOI*:* https://doi.org/10.1080/0907676X.2023.2171800.

Whyatt, B. (2012). *Translation as a human skill. From predisposition to expertise.* Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu.

Whyatt, B. (2018). Old habits die hard: Towards understanding L2 translation. Między Oryginałem a Przekładem, 24 (3: 41), 89-112. DOI: https://doi.org/10.12797/MOaP.24.2018.41.05.

Whyatt, B., Kościuczuk, T . (2013). Translation into a non-native language: The double life of the native-speaker axiom. *mTm Journal*, 5, 60-79.

Whyatt, B., Pavlovič, N. (2021). Translating languages of low diffusion: current and future avenues. *The Interpreter and Translator Trainer*, 15(2), 1-13. DOI: https://doi.org/10.1080/1750399X.2021.1917172.

Whyatt, B., Witczak, O., Tomczak, E. (2021). Information behaviour in bidirectional translators: focus on online resources. *The Interpreter and Translator Trainer*, 15(2), 1-18. DOI: https://doi.org/10.1080/1750399X.2020.1856023.

Whyatt. B. (2019). In search of directionality effects in the translation process and in the end product. *Translation, Cognition and Behaviour*, 2(1), 79-100. DOI: https://doi.org/10.1075/tcb.00020.why.

Wickens, C. D. (1984). Processing resources in attention. In: R. Parasuraman, D. R. Davies (eds.), *Varieties of attention*. (pp. 63-102). New York: Academic Press

Wolter, B., Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics* 36, 1193-1221. DOI: https://doi.org/10.1017/S0142716414000113.

Zagórska, A. (2022). Defining collocations for the purpose of LSP lexicology. *Kwartanik neofilologiczny*, 69(1), 90-103. DOI: https://doi.org/10.24425/kn.2022.141252.

Zhou, X., Wang, X., Liu, X. (2022). The impact of task complexity and translating self-efficacy belief on students' translation performance: Evidence from process and product data, *Frontiers in psychology*, 13(911850), 01-13. DOI https://doi.prg/10.3389/fpsyg.2022.911850.

Zwierzchoń-Grabowska, E. (2011). Okulograficzne wsparcie badań nad procesem tłumaczenia pisemnego. *Lingwistyka Stosowana*, 4, 199-210.

ELECTRONIC SOURCES

Morell, V. (2020). Dogs understand praise the same way we do. Here's why that matters. Available at: https://www.nationalgeographic.com/animals/article/dogs-praise-brains-human-language.

Łysiak, M. (2022). Jonathan to najstarszy żółw na świecie. Urodził się przed wynalezieniem żarówki. Available at: https://www.national-geographic.pl/artykul/jonathan-to-najstarszy-zolw-na-swiecie-urodzil-sie-przed-wynalezieniem-zarowki.

APPENDIX 1

Informed consent in Polish

DEKLARACJA DOBROWOLNEJ I ŚWIADOMEJ ZGODY NA UDZIAŁ W BADANIU NAUKOWYM

Imię i nazwisko uczestnika badania (drukowanymi literami)

……

Niniejszym oświadczam, że:

1. Jestem osobą pełnoletnią i nieubezwłasnowolnioną całkowicie.
2. Zostałem/am poinformowany/a przez mgr Marcelinę Pietryga o planowanym badaniu naukowym, w szczególności o jego założeniach, celach, przebiegu i sposobie przeprowadzania oraz o przewidywanych korzyściach i ryzyku związanym z udziałem w tym badaniu, a także że otrzymałem, przeczytałem i zrozumiałem pisemną informację dla uczestnika tego badania naukowego.
3. Otrzymałem/am zadowalające odpowiedzi na wszystkie zadane przeze mnie pytania i rozumiem wszystkie przekazane mi informacje dotyczące tego badania naukowego.
4. Zostałem/am poinformowany/a, że udział w badaniu naukowym jest absolutnie dobrowolny.
5. Zostałem/am poinformowany/a, że mogę wycofać się z udziału w tym badaniu naukowym w czasie dwóch tygodni od daty wzięcia udziału w badaniu, bez podania przyczyn, a moja decyzja nie pociągnie za sobą żadnych kar ani utraty praw, które mi przysługują z innych tytułów, w szczególności prawa do opieki zdrowotnej.
6. Zostałem/am poinformowany/a, że jeśli w trakcie trwania badania naukowego będę miał jakieś pytania lub wątpliwości, mogę się z nimi zwracać do osób wskazanych w informacji.
7. Dobrowolnie wyrażam zgodę na udział w badaniu naukowym.

…… …..……

(miejscowość i data) (czytelny podpis)

Zgoda na przetwarzanie danych osobowych:

Wyrażam zgodę na przetwarzanie moich danych osobowych

w zakresie: imię i nazwisko; płeć; wiek; adres e-mail, ilość ukończonych godzin ćwiczeń przekładu pisemnego oraz zajęć z języka polskiego dla tłumaczy; rok studiów; uniwersytet; dane głosowe; zapis ruchu gałek ocznych; zapis procesu przekładu, informacje na temat procesu powstawania przekładu, przekład tekstu

przez**: Uniwersytet Śląski w Katowicach, ul. Bankowa 12, 40-007 Katowice**, administrator.danych@us.edu.pl

w celach: zbadania procesu przekładu wśród studentów kierunków tłumaczeniowych

Oświadczam, że wyrażam zgodę dobrowolnie. Jestem świadomy/a, że mam prawo wycofać zgodę w czasie dwóch tygodni od wzięcia udziału w badaniu bez podawania przyczyny. Przyjmuję do wiadomości, że wycofanie zgody nie wpływa na zgodność z prawem przetwarzania, którego dokonano na podstawie zgody przed jej wycofaniem.

……. ……

(miejscowość i data) (czytelny podpis)

Zgoda na wykonanie nagrania głosowego

Wyrażam zgodę na wykonanie nagrania głosowego moich sesji retrospektywnych. Zostałem/am poinformowany/na, że nagranie zostanie następnie poddane transkrypcji, a spseudonimizowane przykłady, niepozwalające na identyfikację mojej osoby mogą być podawane w pracach naukowych autorstwa mgr Marceliny Pietryga.

…… ……

(miejscowość i data) (czytelny podpis)

Zgoda na ponowne wykorzystanie zebranych danych

Wyrażam zgodę na ponowne wykorzystanie moich danych osobowych przez kierownika projektu, mgr Marcelinę Pietryga, zgodnie z zasadami RODO, do opracowania innych prac naukowych, np. artykułów, dotyczących procesu przekładu.

…… ……

(miejscowość i data) (czytelny podpis)

APPENDIX 2

GDPR form in Polish

KLAUZULA INFORMACYJNA

1. Administratorem danych osobowych jest Uniwersytet Śląski w Katowicach, ul. Bankowa 12, 40-007 Katowice, *administrator.danych@us.edu.pl.*
2. Administrator danych osobowych posiada również inspektora ochrony danych osobowych, z którym można skontaktować się poprzez adres mailowy: iod@us.edu.pl
3. Przetwarzaniu danych osobowych podlegają dane wymienione w *Zgodzie na przetwarzanie danych osobowych.*
4. Posiada Pan/Pani prawo do wycofania zgody na przetwarzanie danych osobowych w czasie dwóch tygodni od daty wzięcia udziału w badaniu, bez podawania przyczyny. Wycofanie zgody nie wpływa na zgodność z prawem przetwarzania, którego dokonano na podstawie zgody przed jej wycofaniem.
5. Udział w badaniu jest dobrowolny i świadomy. Przysługuje Panu/Pani prawo do wycofania się z uczestnictwa w badaniu w ciągu dwóch tygodni od daty wzięcia udziału w badaniu, bez ponoszenia konsekwencji, zarówno teraz jak i w przyszłości. Po wycofaniu się Pana/ Pani z udziału w badaniu, Pani/Pana dane zostaną trwale usunięte. W przypadku chęci wycofania się z badania bardzo proszę o kontakt na adres mailowy: marcelina.pietryga@us.edu.pl
6. Zebrane dane są poufne, nie będą rozpowszechnione w sposób pozwalający na identyfikację uczestnika badania. Pani/Pana imię i nazwisko zostanie zastąpione numerem, a uzyskane dane będą przetwarzane w celach badawczych, do opracowania pracy doktorskiej oraz artykułów naukowych.
7. Celem przetwarzania danych osobowych jest zbadanie procesu przekładu wśród studentów kierunków tłumaczeniowych. Podstawą prawną do przetwarzania danych osobowych jest wyrażona przez Pana/Panią zgoda.
8. Każdy uczestnik badania otrzyma dwa ulgowe bilety autobusowe uprawniające do przejazdu na terenie dwóch miast. Otrzymają Państwo również drobne gadżety uniwersyteckie. Sponsorem biletów autobusowych oraz gadżetów uniwersyteckich jest Uniwersytet Śląski.
9. Dane osobowe będą przechowywane przez co najmniej 10 lat lub tak długo jak będzie to wymagane.
10. Posiada Pan/Pani prawo do żądania od administratora usunięcia danych osobowych, a także prawo do ich przenoszenia.
11. Posiada Pan/Pani prawo do wniesienia skargi do organu nadzorczego.

12. Podanie danych osobowych jest warunkiem uczestnictwa w badaniu. Konsekwencją niepodania danych osobowych, będzie brak możliwości uczestnictwa w badaniu.
13. Dane osobowe nie będą podlegały zautomatyzowanemu podejmowaniu decyzji, a więc profilowaniu.
14. Uczestnictwo w badaniu wiąże się z ryzykiem narażenia na działanie ekranu komputera.
15. W celu uzyskania dalszych informacji uprzejmie proszę o kontakt na adres mailowy: marcelina.pietryga@us.edu.pl.

APPENDIX 3

Self-assessment grid Table in English

ENGLISH

numer uczestnika:

		A1	A2	B1	B2	C1	C2
UNDERSTANDING	Listening	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.
	Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.
SPEAKING	Spoken Interaction	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.
WRITING	Writing	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.	I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

APPENDIX 4

Self-assessment grid table in Polish

numer uczestnika:

Tabela samooceny. JĘZYK POLSKI

	A1	A2	B1
Słucham i rozumiem	Rozumiem słowa i bardzo proste zdania na mój temat, na temat mojej rodziny, na temat tego co mi bliskie i znane, jeśli moi rozmówcy mówią do mnie powoli i wyraźnie	Potrafię zrozumieć bardzo często używane słowa i wyrażenia na mój temat, mojej rodziny, na temat tego co mi bliskie i znane, zakupów, szkoły, kolegów. Potrafię zrozumieć główną informację komunikatów i wiadomości przekazywanych ustnie.	Rozumiem informacje podane w języku standardowym na temat domu, szkoły, pracy wolnego czasu, itd. Potrafię zrozumieć główną treść większości programów telewizyjnych i radiowych dotyczących bieżących wydarzeń, interesujących mnie tematów jeśli rozmówcy mówią względnie wolno i wyraźnie.
Czytam i rozumiem	Potrafię zrozumieć nazwy, słowa i bardzo proste zdania, na przykład w ogłoszeniach, na plakatach, w katalogach.	Rozumiem teksty krótkie i bardzo proste. Potrafię domyślić się sensu informacji podanej w takich tekstach jak: reklamy, broszurki, menu, rozkłady jazdy. Potrafię zrozumieć krótkie i proste teksty osobiste.	Potrafię zrozumieć głównie teksty zredagowane w języku codziennym lub odnoszące się do mojej pracy. Potrafię zrozumieć opis wydarzeń, opis przeżyć a także życzenia zawarte w prywatnych listach.
Porozumiewam się	Potrafię porozumieć się jeśli mój rozmówca zgadza się powtórzyć wolno lub uprościć swoje zdanie, a także pomóc mi w mojej wypowiedzi. Potrafię postawić proste pytania na znany mi temat, lub na temat tego co jest mi niezbędnie potrzebne, a także odpowiedzieć na podobne pytania.	Potrafię porozumieć się w trakcie wykonywania znanych mi i prostych prac wymagających jedynie prostej i bezpośredniej wymiany informacji na znane tematy. Potrafię się porozumieć w trakcie krótkiej wymiany zdań, jeśli nawet nie umiem prowadzić dłuższej rozmowy.	Potrafię porozumieć się w większości sytuacji w jakich mogę się znaleźć podróżując po kraju, którego języka się uczę. Potrafię bez przygotowania wziąć udział w rozmowie na tematy osobiste lub dotyczące życia codziennego np. rodziny, czasu wolnego, pracy, podróży i bieżących wydarzeń.
Mówię	Potrafię użyć słów i wyrażeń bardzo prostych po to aby opisać siebie, swoje mieszkanie i ludzi, których znam.	Potrafię w kilku zdaniach opisać w prosty sposób moją rodzinę lub innych ludzi, opisać to co robię, moje studia, mój obecny i poprzedni zawód.	Potrafię w prosty sposób zrelacjonować wydarzenia, opowiedzieć o swoich doświadczeniach, marzeniach, nadziejach i celach. Potrafię krótko udzielić wyjaśnień, podać przyczyny, wyrazić opinię lub przedstawić swoje plany. Potrafię opowiedzieć krótką historię, główną intrygę filmu, i wyrazić moje reakcje na jego temat.
Piszę	Potrafię napisać krótką kartkę, np. z wakacji. Potrafię wypełnić prosty formularz wpisując moje imię, nazwisko, narodowość, wiek, adres.	Potrafię zanotować i przekazać na piśmie proste i krótkie informacje. Potrafię napisać osobisty, bardzo prosty list, np. podziękowania.	Potrafię napisać prosty zwięzły tekst na znany mi lub interesujący mnie temat. Potrafię napisać list opisujący moje osobiste doświadczenia, przeżycia i wrażenia.

	B2	C1	C2
Słucham i rozumiem	Rozumiem stosunkowo długie wypowiedzi i wykłady , a nawet potrafię śledzić złożony wywód jeśli dotyczy tematu, który nie jest mi obcy. Potrafię zrozumieć dziennik oraz większość programów telewizyjnych dot. aktualnych tematów. Potrafię zrozumieć większość filmów jeśli ich język jest standardowy.	Potrafię zrozumieć długą wypowiedź nawet jeśli jej struktura nie jest jasna i gdy zawiera ona niedopowiedzenia. Rozumiem programy telewizyjne i filmy bez specjalnego wysiłku.	Nie ma żadnych trudności w rozumieniu języka mówionego, tak w przekazach bezpośrednich jak i poprzez media, ani gdy mówi się szybko. Potrzebuję tylko trochę czasu na oswojenie się z nowym akcentem.
Czytam i rozumiem	Potrafię zrozumieć artykuły i teksty opisujące problematykę współczesną, których autorzy przyjmuję konkretną postawę lub szczególny punkt widzenia. Potrafię zrozumieć współczesny tekst literacki napisany prozą.	Potrafię zrozumieć długie i złożone teksty prasowe lub literackie o zróżnicowanym stylu. Potrafię zrozumieć teksty specjalistyczne i długie instrukcje techniczne nawet jeśli nie są związane z moją specjalizacją.	Czytam bez wysiłku wszelkie teksty, nawet takie, których forma i treść posiadają duży stopień abstrakcji, np. podręcznik, artykuł na tematy specjalistyczne, dzieło literackie.
Rozmawiam	Potrafię porozumieć się w miarę swobodnie i spontanicznie w taki sposób, że interakcje z rdzennym użytkownikiem języka stają się naturalne. Potrafię uczestniczyć czynnie w rozmowie na tematy codzienne, przedstawiać swoje poglądy i ich bronić.	Potrafię porozumiewać się swobodnie i spontanicznie nie szukając słów. Potrafię skutecznie używać języka w kontaktach społecznych i zawodowych. Potrafię wyrażać moje myśli i poglądy oraz nawiązać do wypowiedzi moich rozmówców.	Potrafię uczestniczyć bez żadnego wysiłku we wszelkich rozmowach i dyskusjach swobodnie stosując wyrażenia idiomatyczne i potoczne. Potrafię wypowiadać się płynnie i wyrażać precyzyjnie niuanse znaczeniowe. W przypadku trudności potrafię zręcznie naprawić potknięcie tak, że manewr ten zostanie niemal niezauważony.
Mówię	Potrafię wypowiadać się jasno i szczegółowo na wiele tematów dotyczących moich zainteresowań, potrafię przedstawić swój pogląd na aktualny temat oraz wyjaśnić korzyści i niedogodności różnych rozwiązań.	Potrafię opisać jasno i szczegółowo tematy złożone, odnosząc się przy tym do tematów z nimi związanych, rozwijając niektóre aspekty i odpowiednio kończąc moją wypowiedź.	Potrafię opisać i przeprowadzić jasną i płynną argumentację w stylu dostosowanym do kontekstu, przedstawić w sposób zwięzły i logiczny dany temat i przypomnieć moim słuchaczom główne punkty prezentacji.
Piszę	Potrafię napisać teksty jasne i szczegółowe na wiele tematów związanych z moimi zainteresowaniami. Potrafię napisać esej lub sprawozdanie przekazując informację lub przedstawiając mój pogląd za lub przeciw wyrażonej opinii. Potrafię napisać listy, które przekażą znaczenie jakie osobiście przypisuję danym wydarzeniom i doświadczeniom.	Potrafię wypowiedzieć się poprzez jasno i dobrze skonstruowany tekst oraz rozwinąć w nim mój punkt widzenia. Potrafię wypowiedzieć się pisemnie w liście, eseju lub sprawozdaniu na tematy złożone podkreślając ważne z mego punktu widzenia aspekty. Potrafię dostosować swój styl do stylu nadawcy.	Potrafię napisać tekst jasny, płynny i Stylistycznie dostosowany do kontekstu sytuacyjnego. Potrafię zredagować listy, sprawozdania, raporty i artykuły na złożone tematy, o jasnej strukturze pozwalającej czytelnikowi zrozumieć ich najważniejsze punkty. Potrafię streścić lub przedstawić pisemną krytykę tekstu naukowego lub literackiego.

APPENDIX 5

Polish source text

Swoje 190 urodziny świętuje w tym roku żółw Jonathan, który jest najstarszym zwierzęciem lądowym na świecie, gdyż urodził się jeszcze przed wynalezieniem żarówki. Oficjalnie było to 200 lat temu, jest to jednak przybliżona data, ponieważ w rzeczywistości nikt nie wie, kiedy przyszedł na świat. Pracownicy branży turystycznej na wyspie przypuszczają jednak, że żółw może być znacznie starszy. Leciwy żółw stracił węch i jest ślepy, dlatego jest ręcznie karmiony przez swoich opiekunów, inaczej nie zorientowałby się nawet, że ktoś przyniósł mu jedzenie. Choć część zmysłów go zawodzi, to nadal ma bardzo dobry słuch. Weterynarz twierdzi, że pacjent dobrze reaguje na jego głos i rozpiera go energia. Sławny żółw skupia uwagę mieszkańców wyspy, dlatego niektórzy żartobliwie nadali mu miano celebryty. Coś w tym jest – władze wyspy już od dłuższego czasu przygotowują huczne obchody jego urodzin, które zaplanowano na drugą połowę roku. Z tej okazji ukaże się limitowana seria znaczków pocztowych. Dodatkowo każdy, kto odwiedzi żółwia w roku jego jubileuszu, otrzyma pamiątkowy certyfikat z pierwszym znanym zdjęciem przedstawiającym odcisk jego łapy.

APPENDIX 6

English source text

Every dog owner knows that saying Good dog! in a happy, high-pitched voice will provoke a reaction of joyful tail wagging in their pet. That is why scientists were made curious: What exactly happens in dog's brain when it hears praise? Is it similar to the hierarchical way human brain processes acoustic information? In 2016, scientists discovered that dogs' brains analyse the intonation and meaning of a word separately. They use their right brain hemisphere to do so, whereas people use their left hemisphere. So scientists raise a question: "Do dogs' brains go through the same steps to process approval? Dogs are a speechless species and they respond correctly to our words," says Attila Andics. Scientists found that dogs process the sounds of spoken words hierarchically. Firstly, they analyse the emotional element with the older brain region and then the words' meaning with the newer part. This discovery develops our understanding of how human language evolved. Especially, that dogs and humans last shared a common ancestor some 100 million years ago.

APPENDIX 7

List of collocations analysed in the study

Polish ST	English ST
przyjść na świat	to provoke a reaction
stracić węch	to be made curious
być ślepym	to analyse intonation
przynosić komuś jedzenie	to raise a question
kogoś rozpiera energia	to respond correctly
skupiać uwagę	to develop understanding
przybliżona data	a high-pitched voice
branża turystycza	an acoustic information
huczne obchody	spoken words
limitowana seria	a brain region
pamiątkowy certyfikat	a human language
odcisk łapy	a common ancestor

APPENDIX 8

The questionnaire in Polish

numer uczestnika:

1. Czy któryś z kierunków tłumaczenia był dla Pani/Pana bardziej wymagający/ trudniejszy?
   a) tak, tłumaczenie NA JĘZYK POLSKI
   b) tak, tłumaczenie NA JĘZYK ANGIELSKI
   c) oba kierunki były dla mnie bardzo trudne/wymagające
   d) nie, żaden z kierunków nie był dla mnie wymagający/trudny

2. Proszę o krótkie uzasadnienie swojej odpowiedzi, a w szczególności o wskazanie elementów, które sprawiły że dany kierunek lub oba kierunki sprawiły szczególną trudność. (można wymienić te elementy)

……

……

……

……

……

……

……

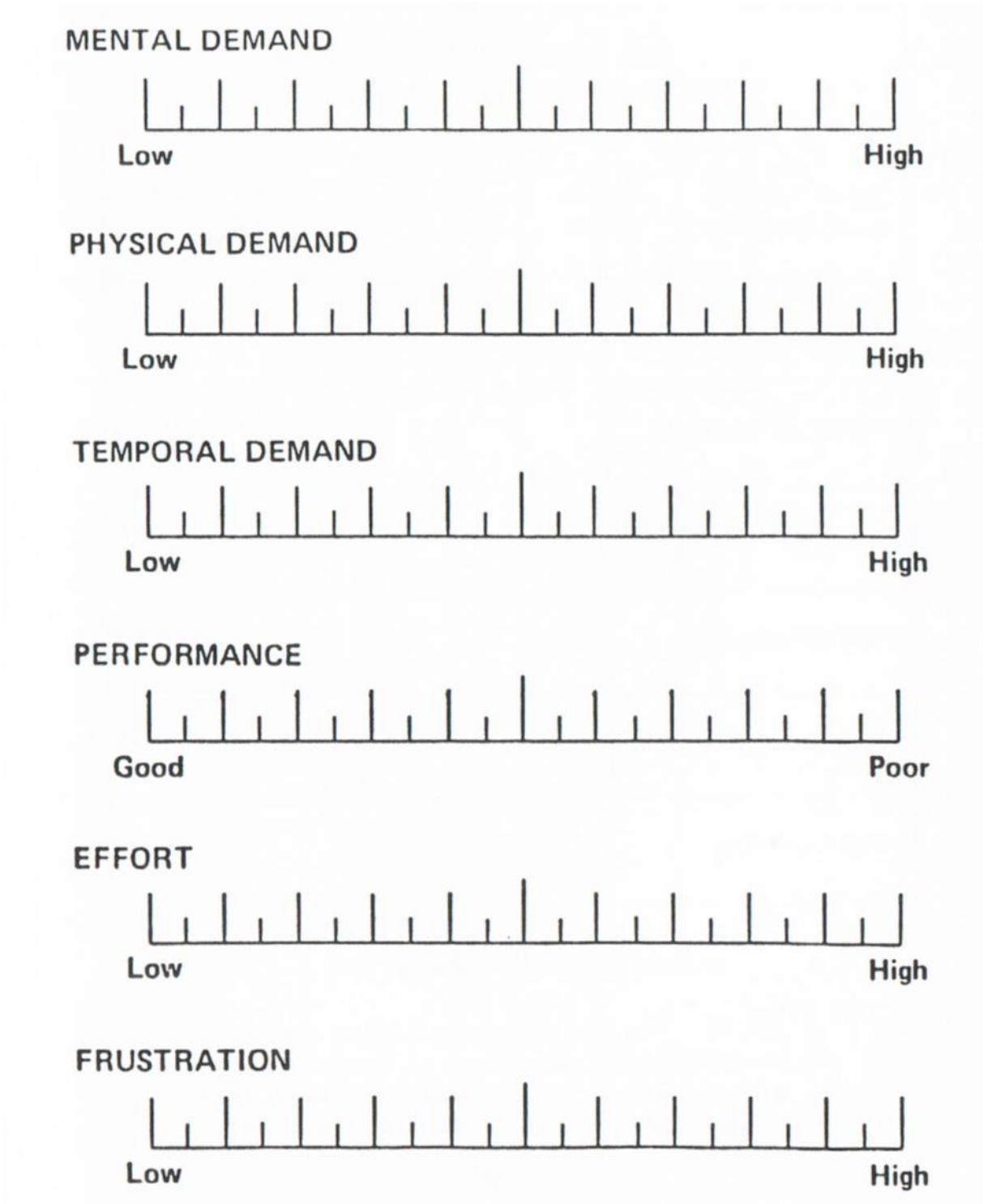
…….

APPENDIX 9

Rating scale definitions in English

Title	Endpoints	Descriptions
MENTAL DEMAND	**Low/High**	**How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exact or forgiving?**
PHYSICAL DEMAND	**Low/High**	**How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?**
TEMPORAL DEMAND	**Low/High**	**How much time pressure did you feel due to the rate or pace at which the tasks or elements occurred? Was the pace slow and leisurely or rapid and frantic?**
PERFORMANCE	**Good/ Poor**	**How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing the goals?**
EFFORT	**Low/High**	**How hard did you have to work (mentally and physically) to accomplish your level of performance?**
FRUSTRATION LEVEL	**Low/High**	**How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?**

Rating scales in English



APPENDIX 10

RESULTS

Average fixation duration

AVERAGE FIXATION DURATION		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	300.91 ms	289.29 ms
PE03	534.26 ms	515.25 ms
PE04	273.09 ms	263.88 ms
PE06	280.92 ms	305.24 ms
PE07	274.25 ms	265.88 ms
PE09	269.18 ms	283.81 ms
PE10	246.79 ms	252.68 ms
PE11	170.8 ms	192.55 ms
PE12	246.75 ms	206.88 ms
PE13	301.11 ms	320.72 ms
PE14	314.05 ms	245.43 ms
PE17	296.17 ms	267.13 ms
PE18	252.16 ms	254.43 ms
PE20	281.19 ms	301.57 ms
PE23	335.62 ms	292.78 ms
PE27	258.64 ms	257.82 ms
PE28	354.76 ms	308.99 ms
PE29	365.1 ms	324.1 ms
PE30	303.44 ms	290.8 ms
PE31	237.97 ms	233.94 ms
PE33	285.57 ms	267.69 ms
PE34	312.93 ms	299.05 ms
PE35	319.11 ms	277.79 ms
PE36	374.45 ms	340.06 ms
PE38	186.96 ms	203.97 ms

Total gaze time

TOTAL GAZE TIME		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	572029.91 ms	749261.1 ms
PE03	743155.66 ms	623850.81 ms
PE04	901470.09 ms	915663.6 ms
PE06	1108229.4 ms	1036595.04 ms
PE07	805746.5 ms	496397.96 ms
PE09	593003.54 ms	538103.76 ms
PE10	452859.65 ms	410352.32 ms
PE11	401721.6 ms	379901.15 ms
PE12	519655.5 ms	274529.76 ms
PE13	1307118.51 ms	1060621.04 ms
PE14	823439.1 ms	465580.71 ms
PE17	784258.16 ms	423935.31 ms
PE18	1311988.48 ms	811122.84 ms
PE20	456933.75 ms	528652.21 ms
PE23	797433.12 ms	1010091 ms
PE27	656428.32 ms	502491.18 ms
PE28	922376 ms	623850.81 ms
PE29	487773.6 ms	608335.7 ms
PE30	1483518.16 ms	960221.6 ms
PE31	579932.89 ms	394188.9 ms
PE33	797025.87 ms	610868.58 ms
PE34	2474337.51 ms	1246739.45 ms
PE35	872127.63 ms	656139.98 ms
PE36	1533747.2 ms	631491.42 ms
PE38	436364.64 ms	601099.59 ms

Eye-key span

EKS		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	2910 ms	10755.33 ms
PE03	7850.63 ms	5385.43 ms
PE04	12307.17 ms	5762.63 ms
PE06	2045.29 ms	2386.57 ms
PE07	8274.67 ms	3899.75 ms
PE09	7537.57 ms	6785.33 ms
PE10	25771.63 ms	11109.45 ms
PE11	3143.25 ms	14451.5 ms
PE12	5203 ms	6809.71 ms
PE13	3739.22 ms	3383.5 ms
PE14	9335 ms	5398.78 ms
PE17	6626 ms	3180.86 ms
PE18	5911 ms	6685.2 ms
PE20	5411.6 ms	13029 ms
PE23	2208 ms	9552 ms
PE27	3849.14 ms	3386.17 ms
PE28	4818.83 ms	3666.38 ms
PE29	4533.91 ms	5497.71 ms
PE30	2449.88 ms	1867.33 ms
PE31	2488 ms	7209.63 ms
PE33	13357.25 ms	5598.22 ms
PE34	5752.17 ms	4241.2 ms
PE35	8489.25 ms	5570.5 ms
PE36	5655.38 ms	4668 ms
PE38	5114.43 ms	5976.56 ms

Total task time

TOTAL TASK TIME		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	1000044 ms	1163886 ms
PE02	974679 ms	696481 ms
PE03	1269356 ms	1306804 ms
PE04	1361062 ms	1397046 ms
PE06	1455439 ms	1260874 ms
PE07	1228087 ms	884164 ms
PE08	1509120 ms	1212408 ms
PE09	919523 ms	720143 ms
PE10	1312573 ms	1077846 ms
PE11	1036707 ms	758063 ms
PE12	860137 ms	647838 ms
PE13	1590740 ms	1270797 ms
PE14	1252477 ms	846039 ms
PE15	1247747 ms	1362818 ms
PE16	853040 ms	848097 ms
PE17	1345338 ms	963414 ms
PE18	1963893 ms	1194660 ms
PE20	706585 ms	765251 ms
PE21	1175202 ms	1297754 ms
PE22	636717 ms	521665 ms
PE23	1031461 ms	1278439 ms
PE25	1397842 ms	1037411 ms
PE26	1457658 ms	1751625 ms
PE27	1018534 ms	708349 ms
PE28	997262 ms	978250 ms
PE29	1349176 ms	764088 ms
PE30	630418 ms	1126312 ms
PE31	1770503 ms	762365 ms
PE32	1311988 ms	964931 ms
PE33	1060774 ms	993082 ms
PE34	1340191 ms	1815716 ms
PE35	3492341 ms	781141 ms
PE36	999393 ms	873613 ms
PE37	2065345 ms	1233901 ms
PE38	1996160 ms	1173542 ms

Duration of the orientation phase

DURATION OF THE ORIENTATION PHASE		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	5035 ms	14111 ms
PE02	55731 ms	60128 ms
PE03	55558 ms	60929 ms
PE04	80064 ms	98537 ms
PE06	43498 ms	35105 ms
PE07	13750 ms	15083 ms
PE08	49121 ms	61338 ms
PE09	36077 ms	46645 ms
PE10	9647 ms	9248 ms
PE11	70596 ms	54449 ms
PE12	33828 ms	2794 ms
PE13	17452 ms	7887 ms
PE14	1130 ms	9926 ms
PE15	10617 ms	62919 ms
PE16	17145 ms	6900 ms
PE17	9311 ms	4510 ms
PE18	119481 ms	52529 ms
PE20	62048 ms	53473 ms
PE21	12780 ms	11055 ms
PE22	44010 ms	44504 ms
PE23	5439 ms	11856 ms
PE25	20998 ms	7053 ms
PE26	90835 ms	91892 ms
PE27	6170 ms	10971 ms
PE28	5266 ms	5027 ms
PE29	115115 ms	10819 ms
PE30	5529 ms	2083 ms
PE31	55768 ms	25610 ms
PE32	9585 ms	12857 ms
PE33	39732 ms	9705 ms
PE34	15393 ms	7760 ms
PE35	15535 ms	110389 ms
PE36	129526 ms	10840 ms
PE37	24074 ms	8793 ms
PE38	40642 ms	4266 ms

Duration of the drafting phase

DURATION OF THE DRAFTING PHASE		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	806801 ms	947984 ms
PE02	895911 ms	627480 ms
PE03	970198 ms	902679 ms
PE04	823752 ms	1071573 ms
PE06	1139684 ms	888983 ms
PE07	903877 ms	733135 ms
PE08	1270547 ms	986159 ms
PE09	737817 ms	628069 ms
PE10	1068628 ms	992228 ms
PE11	811184 ms	604465 ms
PE12	733202 ms	603818 ms
PE13	1064589 ms	1053877 ms
PE14	746938 ms	662687 ms
PE15	715030 ms	808134 ms
PE16	534713 ms	622313 ms
PE17	1044252 ms	829184 ms
PE18	1100479 ms	813335 ms
PE20	542631 ms	546662 ms
PE21	1064086 ms	992788 ms
PE22	461859 ms	415668 ms
PE23	786805 ms	1200931 ms
PE25	1129696 ms	901043 ms
PE26	1279213 ms	1503999 ms
PE27	895607 ms	612506 ms
PE28	761174 ms	737095 ms
PE29	803442 ms	508607 ms
PE30	530697 ms	973643 ms
PE31	1347301 ms	596223 ms
PE32	1028056 ms	767580 ms
PE33	682050 ms	607736 ms
PE34	913878 ms	979014 ms
PE35	1594122 ms	529395 ms
PE36	762488 ms	686869 ms
PE37	1824169 ms	1193593 ms
PE38	1563932 ms	862453 ms

Duration of the revision phase

DURATION OF THE REVISION PHASE		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	188208 ms	201791 ms
PE02	23037 ms	8873 ms
PE03	243600 ms	343196 ms
PE04	457246 ms	226936 ms
PE06	272257 ms	336786 ms
PE07	310460 ms	135946 ms
PE08	189452 ms	164911 ms
PE09	145629 ms	45429 ms
PE10	234298 ms	76370 ms
PE11	154927 ms	99149 ms
PE12	93107 ms	41226 ms
PE13	508699 ms	209033 ms
PE14	504409 ms	173426 ms
PE15	522100 ms	491765 ms
PE16	301182 ms	218884 ms
PE17	291775 ms	129720 ms
PE18	743933 ms	328796 ms
PE20	101906 ms	165116 ms
PE21	98336 ms	293911 ms
PE22	130848 ms	61493 ms
PE23	239217 ms	65652 ms
PE25	247148 ms	129315 ms
PE26	87610 ms	155734 ms
PE27	116757 ms	84872 ms
PE28	230822 ms	236128 ms
PE29	430619 ms	244662 ms
PE30	94192 ms	150586 ms
PE31	367434 ms	140532 ms
PE32	274347 ms	184494 ms
PE33	338992 ms	375641 ms
PE34	410920 ms	828942 ms
PE35	1882684 ms	141357 ms
PE36	107379 ms	175904 ms
PE37	217102 ms	31515 ms
PE38	391586 ms	306823 ms

Number of pauses longer than 5s

NUMBER OF PAUSES LONGER THAN 5S		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	19	35
PE02	28	20
PE03	23	23
PE04	35	41
PE06	40	26
PE07	25	16
PE08	41	27
PE09	20	17
PE10	41	26
PE11	28	18
PE12	16	8
PE13	34	35
PE14	24	20
PE15	10	23
PE16	9	14
PE17	29	31
PE18	35	16
PE20	6	8
PE21	39	37
PE22	5	3
PE23	18	34
PE25	35	26
PE26	50	64
PE27	30	14
PE28	44	24
PE29	3	5
PE30	61	37
PE31	20	16
PE32	33	23
PE33	27	20
PE34	67	42
PE35	22	19
PE36	52	10
PE37	70	37
PE38	25	33

Mean length of pauses longer than 5s

MEAN LENGTH OF PAUSES LONGER THAN 5S		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	7.9 ms	9.43 ms
PE02	14.43 ms	9.29 ms
PE03	11.11 ms	10.44 ms
PE04	8.41 ms	11.91 ms
PE06	11.45 ms	9.45 ms
PE07	9.31 ms	7.89 ms
PE08	11.27 ms	10.3 ms
PE09	9.16 ms	9.74 ms
PE10	10.25 ms	12.07 ms
PE11	10.55 ms	8.95 ms
PE12	9.67 ms	6.84 ms
PE13	9.65 ms	12.18 ms
PE14	9.41 ms	8.92 ms
PE15	12.09 ms	8.9 ms
PE16	7.88 ms	8.5 ms
PE17	11.95 ms	9.16 ms
PE18	16.45 ms	9 ms
PE20	9.71 ms	6.77 ms
PE21	11 ms	8.66 ms
PE22	6.33 ms	6.83 ms
PE23	8.23 ms	11.26 ms
PE25	9.32 ms	10.31 ms
PE26	13.08 ms	16.08 ms
PE27	11.07 ms	12.96 ms
PE28	8.17 ms	7.26 ms
PE29	6.99 ms	9.12 ms
PE30	11.34 ms	11.07 ms
PE31	7.85 ms	7 ms
PE32	10.82 ms	11.96 ms
PE33	10.76 ms	7.48 ms
PE34	12.11 ms	10.81 ms
PE35	13.91 ms	9.22 ms
PE36	10.66 ms	8.2 ms
PE37	11.4 ms	10.06 ms
PE38	9.19 ms	9.73 ms

Self-reports of cognitive effort – general results

NUMBER SELF-REPORTS OF COGNITIVE EFFORT		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	1	2
PE02	6	0
PE03	5	2
PE04	6	3
PE06	7	7
PE07	2	1
PE08	2	0
PE09	3	2
PE10	3	1
PE11	3	1
PE12	3	0
PE13	7	3
PE14	6	3
PE15	7	2
PE16	4	3
PE17	2	1
PE18	4	2
PE20	2	1
PE21	6	6
PE22	1	1
PE23	0	0
PE25	5	0
PE26	0	0
PE27	0	2
PE28	0	2
PE29	0	0
PE30	5	1
PE31	4	4
PE32	4	2
PE33	1	0
PE34	0	1
PE35	3	0
PE36	4	0
PE37	3	1
PE38	4	0

Self-reports of cognitive effort – Level 1 of cognitive effort

LEVEL 1 OF COGNITIVE EFFORT		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	0	1
PE02	0	0
PE03	2	1
PE04	2	3
PE06	1	3
PE07	0	1
PE08	0	0
PE09	2	2
PE10	0	0
PE11	2	1
PE12	0	0
PE13	1	0
PE14	1	0
PE15	3	0
PE16	0	2
PE17	0	0
PE18	0	0
PE20	0	0
PE21	4	4
PE22	0	1
PE23	0	0
PE25	0	0
PE26	0	0
PE27	0	1
PE28	0	2
PE29	0	0
PE30	0	1
PE31	1	3
PE32	1	2
PE33	0	0
PE34	0	1
PE35	0	0
PE36	0	0
PE37	2	1
PE38	0	0

Self-reports of cognitive effort – Level 2 of cognitive effort

LEVEL 2 OF COGNITIVE EFFORT		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	0	0
PE02	0	0
PE03	0	1
PE04	2	0
PE06	0	1
PE07	0	0
PE08	0	0
PE09	0	0
PE10	0	0
PE11	0	0
PE12	0	0
PE13	0	1
PE14	0	0
PE15	0	1
PE16	0	1
PE17	0	0
PE18	0	0
PE20	0	0
PE21	0	1
PE22	0	0
PE23	0	0
PE25	0	0
PE26	0	0
PE27	0	0
PE28	0	0
PE29	0	0
PE30	0	0
PE31	0	1
PE32	0	0
PE33	0	0
PE34	0	0
PE35	0	0
PE36	0	0
PE37	0	0
PE38	0	0

Self-reports of cognitive effort – Level 3 of cognitive effort

LEVEL 3 OF COGNITIVE EFFORT		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	1	1
PE02	6	0
PE03	3	0
PE04	2	0
PE06	6	3
PE07	2	0
PE08	2	0
PE09	1	0
PE10	3	1
PE11	1	0
PE12	3	0
PE13	6	2
PE14	5	3
PE15	4	1
PE16	4	0
PE17	2	1
PE18	4	2
PE20	2	1
PE21	2	1
PE22	1	0
PE23	0	0
PE25	5	0
PE26	0	0
PE27	0	1
PE28	0	0
PE29	0	0
PE30	5	0
PE31	3	0
PE32	3	0
PE33	1	0
PE34	0	0
PE35	3	0
PE36	4	0
PE37	1	0
PE38	4	0

Self-reports of cognitive effort – lack of cognitive effort

LACK OF COGNITIVE EFFORT		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	0	0
PE02	0	0
PE03	0	0
PE04	0	0
PE06	1	1
PE07	0	0
PE08	0	0
PE09	0	0
PE10	0	0
PE11	0	0
PE12	0	0
PE13	2	2
PE14	1	1
PE15	0	2
PE16	1	2
PE17	0	1
PE18	0	0
PE20	1	1
PE21	3	2
PE22	1	0
PE23	0	0
PE25	0	0
PE26	0	0
PE27	0	0
PE28	0	0
PE29	0	0
PE30	2	1
PE31	0	1
PE32	0	0
PE33	0	0
PE34	0	0
PE35	0	0
PE36	0	0
PE37	0	0
PE38	0	0

Weighted rating of the NASA-TLX

WEIGHTED RATING OF THE NASA-TLX		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	33.67	30.33
PE02	45.67	50
PE03	36.67	34
PE04	44	25
PE06	58.33	84.67
PE07	40.33	39.67
PE08	54.67	52.33
PE09	53.33	31.33
PE10	80	80
PE11	61.33	62
PE12	47	62.33
PE13	56.67	56
PE14	40.33	35.67
PE15	73.33	68
PE16	52.67	53.33
PE17	55	50.33
PE18	52	50.33
PE20	52.67	44.67
PE21	56	54
PE22	41.67	46.33
PE23	55.33	60
PE25	32.67	12
PE26	54	41.33
PE27	44	43.33
PE28	44.67	37.33
PE29	71.33	45.33
PE30	67.33	72
PE31	56	41.33
PE32	29	29
PE33	54	52.67
PE34	49.33	38.33
PE35	69.67	46.33
PE36	40	37.33
PE37	43.33	49.33
PE38	29	29

Accuracy of translation

ACCURACY OF TRANSLATION		
PARTICIPANT NUMBER	**L1-L2**	**L2-L1**
PE01	2.42	4.33
PE02	4.67	5
PE03	4.67	4.5
PE04	3.75	4.92
PE06	4.25	4.33
PE07	4.58	5.25
PE08	4.42	5,5
PE09	4.83	3.67
PE10	4.58	4.83
PE11	4.33	4.58
PE12	3.42	4.33
PE13	4.67	5
PE14	5	4.83
PE15	4	4.83
PE16	4.17	4.58
PE17	4.5	4.08
PE18	3.58	4.83
PE20	3.67	4.17
PE21	4.42	4.92
PE22	4.08	5.25
PE23	3.83	4.83
PE25	4.67	4.67
PE26	4.83	4.25
PE27	4.17	4.42
PE28	4.08	5.25
PE29	5	4.5
PE30	4.08	4.58
PE31	3.33	4.5
PE32	4.83	4.25
PE33	4.42	5.17
PE34	4	4.08
PE35	5.08	5.08
PE36	4.08	4.92
PE37	3.42	5.08
PE38	3.42	5.25

APPENDIX 11

Summary of statistical tests results

VARIABLE	L1-L2 MEAN VALUE	L2-L1 MEAN VALUE	IS THE DIFFERENCE STATISTICALLY SIGNIFICANT?
AVERAGE FIXATION DURATION	295.05 ms	282.5 ms	YES
TOTAL GAZE TIME	872906 ms	662403.43 ms	YES
EKS	6591.3 ms	6250.3 ms	NO
TOTAL TASK TIME	1303357.5 ms	1041109 ms	YES
DURATION OF THE ORIENTATION PHASE	37899.6 ms	29771.17 ms	NO
DURATION OF THE DRAFTING PHASE	952537.4 ms	811197.4 ms	YES
DURATION OF THE REVISION PHASE	312920.5 ms	200140.4 ms	YES
NUMBER OF PAUSES LONGER THAN 5S	30.4	24.2	YES
MEAN LENGTH OF PAUSES LONGER THAN 5 S	10.4s	9.7s	NO
WEIGHTED RATING OF THE NASA-TLX	50.7	47	YES
SELF-REPORTS OF COGNITIVE EFFORT	9.4	7.4	YES
LEVEL 1 OF COGNITIVE EFFORT	0.6	0.8	NO
LEVEL 2 OF COGNITIVE EFFORT	0.05	0.2	NO
LEVEL 3 OF COGNITIVE EFFORT	2.5	0.5	YES
LACK OF COGNITIVE EFFORT	0.3	0.4	NO