Olga Witczak

365473

Information searching in the post-editing and translation process

[Poszukiwanie informacji w procesie post-edycji i tłumaczenia]

Praca doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza w Poznaniu
pod kierunkiem prof. UAM dr hab. Bogusławy Whyatt

Poznań, 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my supervisor, Prof. Bogusława Whyatt, for her expert advice, unwavering support, and endless patience. Thank you for believing in me and this project from its inception to the laborious completion. I could not be more grateful for your help and guidance with carrying out this research project and motivating me along the way.

Secondly, my sincerest thanks to all the volunteer participants from the Faculty of English, Adam Mickiewicz University, Poznań – for generously giving their time and effort to take part in this project. If it had not been for them, this thesis would not have been possible.

Thirdly, I would like to thank all who helped me with designing the experiment and at various stages of this project. I am indebted to all my teachers at the Faculty of English PhD Studies Programme for their insightful advice throughout the planning stage of my experiment. Also, my sincerest thanks to the organisers of the Summer School on Process and Product Methodologies in Translation and Interpreting Studies in 2017, especially Dr. Joke Daems and Prof. Gert De Sutter. Participating in the summer school was an amazing opportunity to discuss my project and an invaluable help in the process of designing the experimental study. I would also like to express my gratitude to Prof. Luuk Van Waes and Prof. Mariëlle Leijten for introducing me to the world of Inputlog and for our cooperation with developing the Polish version of the Copy Task. Many thanks to Maximiliana Behnke for her help with extracting corpus data for the Inputlog Copy Task. Many thanks to Ewa Tomczak and Mateusz Kołek for helping with the statistical analyses and their patience with my numerous questions along the way. Finally, I would also like to thank my colleagues from the Department of Translation Studies at the Faculty of English for their support and help at various stages of this project.

Completing this thesis would not have been possible without my family and friends. I am most grateful to my parents and grandparents for always believing in me and supporting my dreams. And last but not least, my sincerest thanks to my dear friends, Bartosz Brzoza, Katarzyna Jusiak, Marcin Naranowicz, and Joanna Wojtaszek – thank you for cheering me and being the occasional sounding board. I am incredibly grateful to all of you.

OŚWIADCZENIE

Ja, niżej podpisana

Olga Witczak

przedkładam rozprawę doktorską

pt. Information searching in the post-editing and translation process
(Poszukiwanie informacji w procesie post-edycji i tłumaczenia)

na Uniwersytecie im. Adama Mickiewicza w Poznaniu
i oświadczam,
że napisałam ją samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałam z pomocy innych osób, a w szczególności nie zlecałam opracowania rozprawy lub jej części innym osobom, ani nie odpisywałam tej rozprawy lub jej części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

Poznań, 8 kwietnia 2021 r.

.....

(miejscowość, data)

.....

(czytelny podpis)

Table of contents

List of tables

List of figures

List of equations

List of abbreviations

ASL	average segment length
CAT	computer-assisted translation
EFL	English as a foreign language
FRES	Flesch reading ease score
IB	information behaviour
IS	information searching
LHQ	language history questionnaire
LSP	language service provider
LTM	long-term memory
MAHT	machine-aided human translation
MT	machine translation
NMT	neural machine translation
OR	online resource
SERP	search engine results page
SL	source language
SMT	statistical machine translation
ST	source text
TAP	think-aloud protocol
TC	translator competence
TCI	translator-computer interaction
TII	translator-information interaction
TL	target language
TPR	translation process research
TT	target text

Introduction

The translation profession has already long ago transitioned from the analogue world of typewriters and paper dictionaries into the digital realm of computers, computer-assisted translation, and online resources, although paper dictionaries have not yet become entirely obsolete. Thus, effective consultation of the wealth of available resources on the Internet for translation requires developing specialised skills related to information searching. The profession has become irrevocably intertwined with technology as automation wormed its way into various aspects of translation (e.g. machine translation and other CAT features). Post-editing – i.e. correcting machine translation errors by a human – is a task that a translator can be requested by the client or it may supplement an array of other translation aids. However, translators often have mixed feelings towards this way of obtaining a target text and the reasons for it are complex. The relationship between those attitudes and actual performance as well as the concurrent use of online resources is still an underdeveloped area in Translation Studies. Also, technological and information-mining skills have become a fixed part of syllabuses across translation training programmes (e.g. European Master's in Translation as per EMT framework 2017). Future translators who belong to the demographic of the so-called digital natives already born into the world of ubiquitous technology may have it easier to acclimatise themselves with the digital reality of the profession. This is why their information behaviours are particularly interesting from the point of view of translation process as new translation trainees are usually already competent users of various technological aids and online resources.

In the light of the above, this thesis aims at investigating the effects of early translation training with reference to machine translation use and information searching behaviours. In order to test for potential effects and correlations, a mixed-method experimental study was conducted. The main interest of this dissertation is to gauge the interconnections between the intuitive use of machine translation and online resources with accuracy in selecting translation equivalents and attitude towards machine translation as well as effort put into the translation and post-editing process.

This dissertation is divided into two parts: theoretical (Chapters 1–3) and empirical (Chapter 4). The theoretical part starts with exploring the main concepts connected with machine translation, post-editing, and translators' attitudes towards them (Chapter 1). Then, information behaviour along with translator competence models are detailed (Chapter 2).

Finally, the theoretical part concludes with issues connected with effort in both translation and post-editing process, with the focus on methodology in eye-tracking and keylogging studies. Next, the empirical part (Chapter 4) provides a detailed report on the experimental study on the effects and correlations between aspects of translation training and information behaviour during translation and post-editing.

Chapter 1 opens with an overview of characteristics of machine translation with reference to its strengths and challenges as a translation aid. Two main types currently popular in professional translation are considered: statistical and neural machine translation. Then, the process of translation is examined from two perspectives – in the broad sense from the point of view of Situated Translation (Risku 2010) and in the narrow sense focusing on the cognitive processing involved in translation. This is then followed with an overview of the process of post-editing as seen through the lens of CAT tool integration as well as the information processing and dynamic connectionist/embodied models. Finally, the chapter concludes with the exploration of the relationship between translators and technology as well as how it impacts the translator-information interaction (Zapata 2016).

Chapter 2 presents an in-depth analysis of information behaviour in translation and post-editing process with the main focus on strategies and types of online resources employed by translators to satisfy their information needs. The types of online resources are provided as classified by various studies (e.g. by Raído 2014 and Gough 2017). At the of the chapter, translator competence is considered in relation to information behaviour and technological skills. Models by the PACTE group (Beeby et al. 2009: 208), EMT (EMT Board 2017; EMT expert group 2009), Göpferich (2009), and Kiraly (2013) are presented. Finally, the interplay between expertise and information searching behaviours is explored. The chapter concludes with an exploration of factors affecting information searching as various translation and post-editing studies have examined them (e.g. Daems et al. 2016, 2017; Gough 2017; Hvelplund 2017; Kuznik 2017; Raído 2014).

Chapter 3 zooms on effort in the information searching for translation and post-editing – its operationalisation in keylogging and eye-tracking studies. Starting with the eye-mind assumption (Just and Carpenter 1980) and examining its flaws, gaze-based correlates of effort are then introduced along with the consideration of the traditional division into cognitive, temporal, and technical effort (Krings 2001). The methodological aspects of eye-tracking are presented in preparation for the empirical chapter proceeding this one. These include data quality filters in eye-tracking and measures connected with operationalising effort and cognitive

load: objective (fixation-based measures and text readability) as well as subjective (such as effort perception).

The final chapter of this dissertation, Chapter 4, is a detailed report on an experimental study conducted for the purpose of this thesis. The design of the experiment is the author's attempt at an original contribution to the translation process research on translator-information interaction. The study consisted in recording the translation and post-editing process of twenty students (11 translation trainees and 9 EFL students) with the use of eye-tracking and key-logging. It was designed to test for potential differences regarding effort (cognitive and temporal) between the groups and tasks. The study also aimed at looking for correlations between effort and other measures pertaining to both the process and product (such as the range of consulted online resources and accuracy). The recordings were supplemented with questionnaire data which provided group characteristics and investigated attitudes towards MT. The chapter begins with enumerating the aims of the study, variables, and hypotheses. Then, participants, materials, and tools are described. Data analysis is divided into the sections about the process, product, and questionnaire data. Finally, results of the experiment are provided for all seven hypotheses along with a qualitative analysis of open questionnaire answers, followed by a discussion for each of them in separate sections. At the end of the chapter, there is a reflection on the study limitations as well as possible future research avenues to be explored. The chapter concludes with a general discussion of results in the light of the whole project and the pedagogical implications of the study. The closing remarks are followed by the list of references and appendices with source texts, translation and post-editing briefs, informed consent form, and empty questionnaires used in the study.

Chapter 1: Translating with the machine: neural machine translation, post-editing, and translation from scratch

1.1. Introduction

This chapter provides a theoretical overview for an experimental study of Information Searching and the process of translation as well as post-editing. Strengths and challenges of using machine translation (henceforth MT) for translation and post-editing are presented, along with two main types of MT – statistical and neural MT – and translators' attitudes towards the technology. The chapter then proceeds with a review of relevant studies on the translation and post-editing process to establish the theoretical concepts important for the empirical study presented later in this thesis. As far as the process of translation is concerned, these concepts include Situated Translation, cognitive processing, computer-assisted translation (henceforth CAT) tools for integration of MT. Furthermore, subsequent sections focus on the information processing model of post-editing along with the dynamic connectionist/embodied model, in parallel to the concepts underpinning the translation process. At the end, the chapter describes the concepts of translator-computer and translator-information interaction as a prelude to Chapter 2 which zones in on Information Searching.

1.2. Statistical Machine Translation

The inception of statistical machine translation (henceforth SMT) dates back to 1949 when Warren Weaver had the idea that statistics might provide a way to enable automatic translation between languages. However, earlier computers lacked the required computational capacity to process large databases and it was not until 1991 when IBM applied SMT in the "Candide" project (Yang and Min 2015: 201). Furthermore, although SMT is fundamentally different from the rule-based paradigm, the latter played a key role in the development of mainstream online MT services. This section will provide an outline of how Google Translate, probably the best known and most easily available MT system, came to be the way it is today and how it transitioned to the latest neural-based system.

The popularisation of the Internet since the mid-1990s was the catalyst for MT popularity and development in general. A popular IT magazine, *Computerworld*, in a 1995 issue

stated that the translation market is flourishing and that MT might be useful alongside other computer aids for translators, e.g. word processors (Richman 1995). According to the article, MT generated by such products as Logos, Transcend, or Systran was considered useful only for words and phrases in contexts such as parts lists in technical texts (Richman 1995).

In 1994 Systran made its MT system widely available on the Internet (Boitet et al. 2010: unpaginated). Later, based on Systran technology, the Babelfish website was launched in 1997 via the AltaVista service to allow free automatic translation of texts or whole websites (Yang and Lange 1998). Babelfish used the rule-based paradigm to deliver its automatic translations; this paradigm utilised linguistic rules and huge dictionaries (Hutchins and Somers 1997; SYSTRAN [n.d.]; Choi 2002). The service was one of landmark developments in the mainstream MT. In 2009, it was acquired by Yahoo! and later in 2012 merged with Microsoft's Bing Translator,[1] which is a contemporary titan of popular online MT besides Google.

Systran used to be the core of the MT engine for Google until 2007, when Google Translate fully switched to its own proprietary SMT. Between 2006 and late 2007, Google used its SMT engine only for Arabic, Chinese, and Russian (Och 2006; Chitu 2007). Och (2006) described the workings of SMT in the following way: "[W]e feed the computer with billions of words of text, both monolingual text in the target language, and aligned text consisting of examples of human translations between the languages. We then apply statistical learning techniques to build a translation model." What it means is that statistical systems like Google's former system use monolingual data to model target language (henceforth TL) structures and bilingual corpora to align source language (henceforth SL) positions (i.e. their placement) to their counterparts in TL positions (Yang and Min 2015; Och 2005). Google moved from word- and phrase-based units to syntax-based structures that reflect hierarchy (Yang and Min 2015; Och 2005). According to Google (Inside Google Translate 2010), in order to produce an automatic translation, their system looked "for patterns in hundreds of millions of documents" and through a two-step process using English as interlingua (intermediary) language, delivered a TL text (Boitet et al. 2010). After almost a decade of SMT in Google-provided translations, in 2016 the platform "went neural," i.e. started utilising neural networks in their MT system.

[1] https://blogs.msdn.microsoft.com/translation/2012/05/30/welcoming-yahoo-babel-fish-users/ (date of access: 25 Jan. 2021).

1.3. Neural machine translation

In November 2016, Google announced a new MT system available for eight language pairs (Turovsky 2016) and in March 2017 also for Polish (Zimowska 2017) – Google Neural Machine Translation or GNMT. Neural machine translation (henceforth NMT) is a form of SMT that utilises neural networks. Such a network is a "machine learning technique that takes a number of inputs and predicts outputs" (Koehn 2017: 6). Furthermore, it means that computers learn from experience via machine learning in such a way that they are able to "modify its processing on the basis of newly acquired information" (OED Online 2018). Input can be any dataset fed into layers of interconnected processors, which are said to imitate the structures in the human brain (Matacic 2016). The following sections outline strengths and challenges of NMT.

1.3.1. Strengths of NMT over SMT

The strength of NMT lies in its capacity for processing whole sentences instead of phrases or n-grams, as it used to be the case with Google's phrase-based SMT. A typical NMT architecture involves two recurrent neural networks; one processes the input sentence, while the other produces the translation – output. All this is usually facilitated by attention mechanism which allows effective processing of long sentences (Wu et al. 2016: 1). NMT is also able to process broader context, thus producing output which reads coherently and more human-like (Turovsky 2016). The system "understands" the co-text because every output word is conditioned by the network on all preceding words (Läubli 2017). In other words, a type of recurrent neural networks called Long Short-Term Memory Models used by Google is able to process long sentences and more accurately manage their attention mechanism to certain input words (e.g. in the case of sentences with subordinate clauses which separates the predicate from the agent) (Koehn 2017: 41). Therefore, neural networks generate fluent translated sentences thanks to their capacity to encompass the whole sentence, as opposed to much shorter n-grams in classic phrase-based SMT (Läubli 2017).

Furthermore, input words are transformed into vectors which represent the relatedness to all other words contained in the training data, e.g. the fact that *dog* is more closely related to *cat* than *car* (Quoc Le, as quoted in Matacic 2016). Neural networks are able to capture more nuance in word similarity than classic SMT did. For instance, Läubli (2017) explained that

thanks to the way words are processed by the network (word embeddings), *except* and *but* are assigned similar numerical representation in the following sentences:

(1) I can resist everything except temptation.

(2) I can resist everything but temptation.

The architecture of a recurrent neural network like GNMT is simpler than the one employed in traditional SMT systems which relied on multiple components (Bentivogli et al. 2016: 257). As opposed to the previous statistical paradigm which required multiple systems for multiple language pairs, GNMT encodes semantics of a given sentence instead of memorising translated phrases. That is, the network creates an interlingua for all pairs in the system thus also enabling so-called zero-shot translation, i.e. "translation between language pairs never seen explicitly by the system" (Schuster et al. 2016). In other words, the data existing for other pairs can be utilised in combinations not seen by the system during training. For instance, the transfer learning means that a multilingual NMT model that has been trained with Portuguese into English and English into Spanish data will be able to produce translations from Portuguese into Spanish (Johnson et al. 2016: 2). Similarly, this can be done for any language combination, just like in Fig. 1.

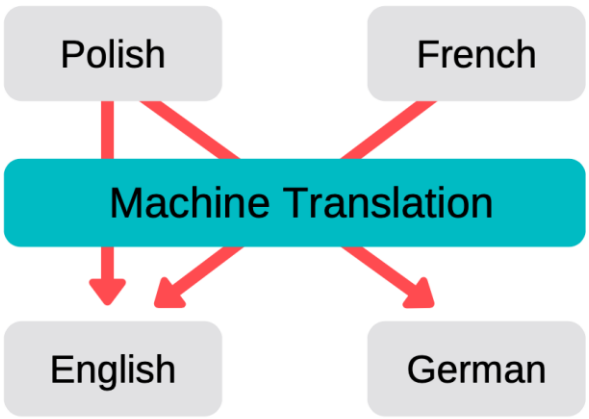


Fig. 1. Multi-language MT, adapted from Koehn (2017: 82)

In system experiments, GNMT proves to work better with related languages. A significant drop in quality has been observed for translation from Spanish into Japanese, which can be attributed to the fact that the two are unrelated (Johnson et al. 2016: 9). Interestingly, a single model is used for all pairs. As a side effect, it improves the translation quality of languages with scarce training data (Johnson et al. 2016: 15). This means that there

are positive implications of the single model for Polish in terms of output quality, as it is a language of limited diffusion.

A study by Junczys-Dowmunt et al. (2016: unpaginated) compared phrase-based SMT with NMT BLEU scores. BLEU is an automatic MT evaluation metric based on how close it is to a human reference translation as calculated by a numerical metric. The BLEU scores are reported to correlate well with human judgements of MT quality (Papineni et al. 2002). According to the study, NMT was either comparable or outperformed SMT for 29 out of 30 language pairs, however, the BLEU scores varied depending on the language pair (Junczys-Dowmunt et al. 2016: unpaginated). As for NMT quality in GNMT, it is reported (Wu et al. 2016: 2) that there are improvements for both BLEU scores when compared to reference human translations. There are 60% fewer translation errors than in Google's previous phrase-based SMT for language pairs such as EN<>FR, EN<>ES, EN<>CH, when compared to human translations.

The above-mentioned experiments were conducted on BLEU scores, however, there is another study that incorporated reference post-edited sentences using TER scores[2] (Snover et al. 2006). For the sake of clarity, to post-edit means to "edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)" (Allen 2003: 297). Post-editing is described in detail in Section 1.5. Betivogli et al. (2016: 9) compared three phrase-based SMTs with an NMT system[3] in terms of MT errors (morphological, syntactic, lexical, word order) and they found that NMT outperforms SMT in all respects. Also, a TL with rich morphology and requiring word order shifts (i.e. German in this case) does not imply a decrease in MT output quality. The study found that in terms of processing longer input sentences, NMT is superior to phrase-based SMT. Their results also showed that NMT-generated output requires less post-editing effort than SMT output, i.e. human translators need to put less effort to correct NMT output.[4]

Stefaniak (2020) conducted a study which evaluated NMT performance for the English>Polish language pair. Nine translators from the Polish Language Department of DGT (Directorate General for Translation) translated between 1 and 13 texts (from 1 to 150 pages) using TM and NMT matches. Text types in this study included legislative and non-legislative

[2] TER (translation errors rate) measures the number of edits needed to change MT output into one of the references, post-edited in the case of this study.
[3] Source texts were English TED talks, reference sentences were post-edited into German by professional translators.
[4] More on post-editing and effort in Section 3.1.

texts (e.g. public consultation, report of an audit, notification of a concentration). Productivity tests based on task time were conducted on six translators on 12 random sentences, who were divided into two groups (as to why, it is unclear in the paper) to translate 6 sentences from scratch and post-edit the remaining 6 sentences. The translators worked in an MS Word document. Stefaniak (2020: 265) used TER scores for comparison with task time which the author measured with a stop watch – an unorthodox choice with the keylogging software currently available. However, the most valuable take-away from the study is the evaluation of NMT output in Polish. Stefaniak (2020: 265) reports fluent NMT output with missing information from the ST, errors in terminology (deprecated, obsolete terms chosen by the system), inconsistencies occurring even within the same sentence. Also, wrong (calqued) word order along with mistakes in verb forms and pronouns were observed in the NMT output. What Stefaniak (2020: 266) mentions to be specific for DGT output are mistranslations of legal act titles and quotations. Finally, infrequent words prompted the NMT engine to produce creative choices, e.g. mash-ups of surnames ("Łukasz Brasszek" vs. "Łukasz Brzenczek") that should have been transferred verbatim into the TT. The system also coined new non-existent words like *femzabójstwa* for *femicides* (literally in English *fem-homicides*, such a prefix cannot be attached this way to form new words in Polish) or *nowe borówki* for *newborns* (literally in English: *new berries*). The author explained that these issues were probably caused by the fact that the engine was trained by legal corpora (Stefaniak 2020: 266). The results of the analysis based on TER scores show that NMT performed better for legal texts – most likely due to their repetitive and standard language. In those texts, infrequent and thus more problematic words were not as numerous as in the non-legal texts. The quantitative comparison of median and average TER scores, however, was not reported to be significantly different – only descriptive statistics were reported. In general, NMT output performed well and the participants did not perceive it as cumbersome, but less than 20% of NMT segments were without errors. The correlational analysis for TER scores and post-editing/translation time, however, should be interpreted with extreme caution since no statistical tests were reported to account for any significance. The author provided only a weak correlation between the TER score and post-editing time, but whether negative or positive – it is not stated (Stefaniak 2020: 268).

All in all, despite a number of strengths, NMT is not ideal and comes with some shortcomings as well, which are outlined in the next section.

1.3.2. Challenges in NMT

According to Wu et al. (2016: 1f), NMT systems were less accurate than phrase-based SMT until recently, particularly systems working on large data sets. Three weaknesses of NMT were reported: slow training time, problems with processing rare words, and failure to translate all input words from the source sentence.

Betivogli et al.'s (2016: 9) study shows that long sentences are better handled by NMT, but the longer the input sentence, the more drastic the decrease in output quality when compared with the SMT systems. Another challenge reported in the same study is translation choices dependent on deeper understanding of semantics in the input, as it was the case with the placement of the negation particle in German (*nicht*) or prepositional phrases (*in my life*), which would need to be reordered to conform with the rules of German syntax. While NMT is said to process the semantics of the input data to a certain degree, this "understanding" remains to be researched.

Koehn (2017: 90–100) also mentioned some challenges of NMT models. Firstly, NMT models often do not perform well when input differs significantly from the training data, e.g. in a German sentence from a subtitle corpus *Schaue dich herum* (reference: *Look around you*). It was translated by NMT trained on medical texts as the following incomprehensible string: *EMEA / MB / 049 / 01-EN-Final Work progamme for 2002*, which showed that NMT is helpless when facing a mismatch between the training data and the input provided later. This was shown to be the case in the study by Stefaniak (2020), as reported in the previous section. Most of the time, however, NMT output is deceivingly fluent. If MT is used for gisting, Koehn (2017: 93) stated that fluency oftentimes has nothing to do with the accuracy of NMT: "the user will be misled by hallucinated content in the neural machine translation output." This was observed in reference to the output produced by an NMT system trained on the Quran corpus, formulating a very appropriate and coherent sentence *Take heed of your own souls*. Furthermore, the amount of training data plays a role in the accuracy of NMT output, as corpora containing a few million words or less will produce inaccurate or even unrelated output. Other challenges include noisy data (e.g. misaligned sentences in the input) and issues with word alignment which SMT appears to have better ways of dealing with. Additionally, in terms of analytics of NMT models, the inner workings of the decoding process are said to be opaque to the analysis when compared to other MT approaches (Bentivogli et al. 2016: 257). This opaqueness is sometimes even compared to magic (Kenny 2017). Thus, if a certain error pattern is discovered in the output, it

is not immediately obvious what caused it in the network. Despite a number of challenges that the NMT poses, its advantages make it a system rapidly growing and becoming more present in human translation.

1.4. Translators vs. machine translation

Post-editing of MT has been dubbed one of the fastest-growing segments of the language industry, according to Common Sense Advisory 2016 survey (Common Sense Advisory 2016). Language Service Providers (henceforth LSPs) who implemented MT between 2013 and 2015 were reported to grow 3.5 times more quickly than LSPs with a more conservative approach towards MT usage.

Furthermore, the Common Sense Advisory survey deemed MT to soon become a mainstream solution among LSPs, but replace human translators only in some types of translation jobs. Implementing MT is the only solution, according to the report, to provide high quality translations of increasingly higher volumes of text: "Large enterprises expect double-digit annual growth rates in translation, growth that present methods cannot possibly keep up with, even if the language industry were to add new translators at a historically unprecedented rate" (Common Sense Advisory 2016). This is an important point, since MT is often perceived as a threat to human translators, who fear uncertainty about the future of the profession (see further in this section). Lorenzo and Franceschi (MateCat 2018) described MT post-editing as "a way to reinvent yourself as a professional," thus pointing to MT as more of an asset to individual translators. In 2021, MT integrations have become a standard in all translation workstations available on the market (e.g. Memsource, SDL, memoQ, etc.).

Most importantly, as the technology develops, the post-editor will be more in control of the constantly-changing process of post-editing – new technologies are expected to remedy the translators' dislike of post-editing (Common Sense Advisory 2016). The same survey also featured the popularity of post-editing around the world and, according to Fig. 2, 70% of respondents from Asia and Latin America admitted that they were post-editing, compared to 45–50% of European and North American respondents and 35% in Africa.

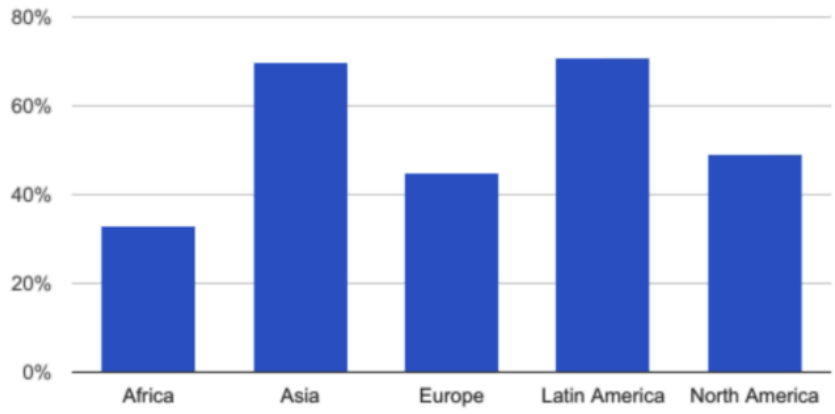


Fig. 2. Adapted from Common Sense Advisory 2016 survey: "Post-editing around the world" from MateCat (2018)

The main reasons to use MT are that it is "directly related to speed (reaching markets more quickly), volume (more content in more languages) and saving costs" (Guerberof Arenas 2013: 87). However, ever since translation technology started to become more widespread, MT has been a controversial issue among translators. In 1993, Meijer found that translators thought the quality of MT output was bad, MT prompted translators to use constructions unnatural for them and that the product was also inferior to the content produced from scratch (1993: 11f). For translators employed in companies frequently using MT, half of the respondents thought that MT boosts translation speed (Meijer 1993: 11), but this varied for different companies. Even though the negativity was prevalent in attitudes at the time, Meijer (1993: 12) reported that a large proportion of the respondents expressed their willingness to find out more about MT and start working with it, thus being open to the opportunities offered by automation. More recent studies have shown similar nuance in attitudes.

Lagoudaki (2008: 265, 268) reported positive reception of MT as a feature of a translation memory system and noted that "translators also *seem to be coming to terms* with machine translation as an alternative means of translation production" [emphasis mine, OW]. In a later study by Tatsumi (2010: 185), survey responses suggested that translators' attitude towards post-editing was flexible. These studies explored attitudes of professionals, while a report published by TAUS (2010) suggested that translators were resistant to embrace post-editing as a part of their workflow. Guerberof Arenas (2013) pointed out that the TAUS report reflected the viewpoint of companies instead of the attitudes of actual translators. Her study investigated professional translators' opinions on post-editing and MT. Participants of the study were professional translators and some of them had experience with post-editing. According to

the survey (Guerberof Arenas 2013: 78), participants admitted exposure to MT helps them spot errors when post-editing rather than making them more tolerant of errors. Survey participants liked MT output for a variety of reasons, e.g. no need to type the translation continuously, ensuring accuracy, consistency, and suitability for formulaic texts (Guerberof Arenas 2013: 84). However, the majority of participants had mixed feelings about post-editing and MT output generally, but those who worked with MT had knowledge of the general principles of MT and typical errors (Guerberof Arenas 2013: 88).

Interestingly, while the primary goal of machine translation post-editing is to reduce effort, Guerberof Arenas' participants thought working with MT required more effort[5] than editing human translations, which could be attributed to cognitive effort being higher than temporal effort for post-editing (2013: 87). They also admitted that rates for MT post-editing were not adequate (2013: 78–79). The rates for this task are reported to be "almost always lower than the standard rates paid for translation, sometimes equivalent to the rates paid for editing fuzzy matches from a translation memory (TM) system, and sometimes lower than TM editing rate" (O'Brien 2017: 320). Although the actual rate is dependent on a variety of factors (e.g. domain, language pair), the general tendency is towards keeping post-editing cheaper than human translation from scratch, at the same time expecting faster delivery.

Translators, who in their professional jobs did not post-edit and reported a dislike towards MT, showed in their performance in the process part of Guerberof Arena's study that MT boosted their productivity. There exists, therefore, a complex relationship between the actual performance and attitude or personal preference. As Guerberof Arenas (2013: 83) put it: "these translators did show productivity increases when working with MT, but of course this does not mean that they were actually 'enjoying' it." Most importantly, despite sometimes ambivalent attitudes, those translators were aware that MT is what the current translation market expects them to be familiar with. They displayed a very open-minded and practical attitude towards it, albeit without embracing it wholeheartedly (2013: 88).

In a study of professional and novice post-editors, Moorkens and O'Brien (2015) examined the post-editing process and attitudes of experts and novices. The study was conducted in PEARL, i.e. a web-based post-editing interface and the objective was to investigate features specific to MT post-editing software, e.g. change case, reject MT output, copy punctuation. Data sets contained 50 English segments each (from Norton Security helpdesk documentation), but only one of the sets had the above features turned on. MT was

[5] Effort is explored in Chapter 3.

provided by the Moses SMT. In a between-subjects design, participants post-edited one of the data sets into German without switching off the PEARL interface, i.e. they were not allowed to consult any resources whilst completing the task. The participants were translation professionals (group 1) and translation trainees (group 2). Apart from the post-editing speed, the researchers measured the participants' attitudes and user engagement (i.e. willingness to try out new features of the new environment, as measured with average number of interface button presses). It turned out that the professionals were less interested in engaging in the task – perhaps because of their automatic responses to the task or negative attitude towards post-editing/MT. Conversely, the trainees interacted in a more engaged way with the task and environment features. The researchers concluded (2015: 80) that novices are potentially more positive and enthusiastic about the new tasks involved in the research, but the results and conclusions could not be generalised towards expert translators.

A later study by Cadwell et al. (2018) on translators' attitudes towards MT concluded that although translation technology has significantly progressed, there is still a strong reluctance towards using MT as a translation aid. The study employed agency theory (Pickering 2008) and focus group interviews as a methodology to gauge attitudes from professional translators employed in two agencies (EU's DGT and Alpha CRC). As concluded by Cadwell et al. (2018: 302), agency in translation should not be considered independently from sociological and cultural contexts, among other things, in which the translator is operating. Agency in translation as well as post-editing is discussed in detail in Section 1.6.1. Both groups of the study reported frequent usage of MT (majority of DGT translators and a quarter of Alpha translators). Cadwell et al. (2018: 301) mention the translators' "concerns about the impact [MT, addition mine, OW] might have on their long-term work practices and skills." MT was considered by both groups a potential constraint of creativity and a source of inspiration "to kick-start the translation process or to get ideas" (Cadwell et al. 2018: 312). Interestingly, the proportions differed for the groups regarding MT as a source of inspiration (76% for DGT and 25% for Alpha), which means that the environment – the mentioned sociological and cultural context – plays a key role in determining the translators' attitude towards MT (Cadwell et al. 2018: 312).

Translators may see post-editing as a nuisance and form their bias towards MT as a result of incorrect information, thus the affective aspect of MT attitude is an important factor to consider in research on effort (Cadwell et al. 2018: 303). In the reported study, the interviewed participants, mainly from the Alpha group, thought that MT made them work

slower than without any machine assistance and that using MT was not enjoyable (Cadwell et al. 2018: 311). As they reported relatively frequent use of MT in their work, it is not surprising that they also were aware of terminology in MT being potentially misleading. They also used MT as a source of terminological suggestions when supplied by the client with information about verified terminology (Cadwell et al. 2018: 312). Finally, the setting (in-house translators) might also play a key role in a relatively positive attitude toward using MT, as such translators do not need to worry about potential monetary losses or confidentiality issues as opposed to freelancers.

In terms of quality, both regarding the source text (henceforth ST) and target text (henceforth TT), Cadwell et al. (2018: 313ff) report differences between the interviewed groups, which were dependent on contact with the clients and the level of post-editing required from the translators (light vs. high quality TTs). When it comes to trust when MT is involved, translations produced by humans were deemed more trustworthy than MT suggestions, but also not all engines were reported to be trusted to the same degree (Cadwell et al. 2018: 314ff). One participant mentioned that if a suggestion came from Google, they did not know anything about it as opposed to the direct contact they had with the proprietary MT engine owned by the Alpha company. Alternatively, the DGT group expressed trust in both human-generated TM suggestions and MT output, as in their case they knew that MT is synthesised from the translations produced in-house. Trust is a key issue with reference to technology, because as Cadwell et al. (2018: 315) argued, lack of trust could result in refusal to adopt MT.

Another study that gauged attitudes towards MT was a social media sentiment analysis by Läubli and Orrego Carmona (2017). They conducted a qualitative and quantitative analysis of 13,000 tweets that referenced MT. The qualitative part concluded that MT-generated mistranslations are used as examples to reinforce the conviction that human translators are not yet redundant. According to their analysis, 48% of the analysed posts reference Google Translate. In these posts, translators express their doubts regarding the quality improvements of MT as announced by developers, which indicates a lack of efficient communication between translators and developers. Fig. 3 below shows an example of MT's failure shared on Twitter. The same example was reposted on different websites also later in 2017, e.g. on LinkedIn[6] (37

[6] https://www.linkedin.com/feed/update/urn:li:activity:6336501045904183297 (date of access: 18 Jan. 2021)

recommendations as of 18 January 2021) and Tumblr[7] (almost 15,000 interactions with the post as of 18 January 2021), not counting reposts on Twitter.[8]

Fig. 3. Example of Google Translate failure posted on Twitter

The negative view of MT can also be seen in how Google Translate is used as an indicator of low quality in comparison to bad human translators, e.g. some are "such poor examples of translations that 'not even Google translate [sic] is that bad'" (Läubli and Orrego Carmona 2017: 63). Another example of negative MT opinion among translators on social media is conviction that it can give a false sense of competence or that amateurs can use MT without honing their translation skills. Some of the analysed opinions mention that allowing readers to get the gist of the message is not acceptable and it may lead to the lowering of profession standards. They are aware of MT's limitations regarding certain language combinations and domains, recognising that MT could be an asset as a translation aid in rapidly developing translation market.

In the quantitative analysis, authors used a web crawler on 13,150 tweets that contained "machine translation" and/or "machine translated" written in English between 1 January 2015 and 31st July 2017. Then, a small portion was annotated manually which then was used to train the automatic sentiment classifier for the rest of the data (Läubli and Orrego Carmona 2017: 65). The results show that translators on Twitter are predominantly negatively disposed towards MT – negative tweets are three times more frequent than positive ones. The following tweet was unanimously deemed negative by both human annotators and the classifier: "Six reasons why machine translation can never replace good human translation: https://t.co/JzLYbXO6yJ #xl8 #t9n[9]" (Läubli and Orrego Carmona 2017: 67). The tweet echoes the topics from the

[7] https://allthingslinguistic.com/post/167465751184/the-best-machine-translation-fail-ive-seen-in (date of access: 25 Jan. 2021)
[8] One example of such repost was retweeted over 10,000 times https://twitter.com/mikithebunny/status/929359446500954113 as of 18 January 2021.
[9] The tweet appears to have been deleted as of 18 Feb. 2021. It was a title of an article on the *Glokalize* website: http://glokalize.com/2017/07/09/machines-translate-humans-write/?platform=hootsuite (date of access: 18 Feb. 2021), also linked in the tweet (hence the link after the colon).

qualitative analysis – the fear of being replaced by machines. The fear of technology in translation is not new, Pym described it as "usually a defence of old accrued power, dressed in the guise of quality" (2011: 4). With the inevitable automation of the translation process, according to Läubli and Orrego Carmona, it can be diminished by involving translators in the development process, starting with creating spaces on social media which are shared by both translators and developers.

In a recent study, Vieira (2020) analysed the automation anxiety among translators in a corpus-based exploration of translators' blog and forum postings and juxtaposed them with the current trends in the areas of employment, work automation, as well as the economy in general. An interesting prediction on the future of the profession as a result of automation include the shift of human translators towards the more creative areas and abandoning those domains that are the most likely to fall prey to automation, e.g. technical translation (Vieira 2020: 17). He used the Sketch Engine's built-in tool (Baroni et al. 2006) to crawl the data from websites such as TranslatorsCafé.com and blogs. The results showed that there is a lot more nuance to translators' attitudes towards MT and worries about job displacement or pay rates were not prominent for most, often focusing on criticizing business practices and MT's limitations (Vieira 2020: 15) which was also the conclusion of the studies already recapped in this section (e.g. Guerberof Arenas 2013; Cadwell et al. 2018; Läubli and Orrego Carmona 2017). MT could only threaten the profession if the translation process is regarded as a mere linguistic operation rather than a comprehensive service, often relying on extensive extralinguistic knowledge and skills (Vieira 2020: 15f). The current translation market trends stray towards segmentation and some non-creative areas of specialized translation such as technical translation could become less prestigious. According to Vieira (2020: 16), these trends indicate that there is a gap between the industry and Translation Studies in terms of what is actually happening in the translation process and what skills translators have. Furthermore, while translators' attitudes do not uniformly dislike MT for fear of being replaced, the negativity towards it stems from detrimental business practices and the fact that it is still an imperfect translation aid. Technology in translation cannot be perceived in vacuum and must be considered in conjunction with its effect on market practices (Vieira 2020: 16). Another conclusion that Vieira (2020: 17) came to is that any dystopian predictions regarding translation should be considered with caution and leaving less creative domains (such as technical translation) to non-translator professionals might lead to further erosion of the concept of the translation process and of translation as a profession.

There is a scarcity of studies surveying students' attitudes towards machine translation. One such study by Daems et al. (2017) compared post-editing and translation between professional translators and students, investigating attitudes and other issues. Translation trainees participating in the study confused working with MT output with using CAT tools in general. A majority of those who reported some experience with post-editing thought it was "equally rewarding as human translation, or preferred human translation to a small degree" (Daems et al. 2017: 261). Professionals thought translation from scratch was more rewarding, but they did not dislike working with MT, recognising its benefits (saving time when MT quality was sufficient). Some professionals valued creativity and thus did not believe MT to have the capacity to boost their translation speed. Lower rates for using MT were also an issue. After a series of post-editing and translation tasks, most participants expressed a preference of translation without MT, with only one student and four professionals admitting the opposite (out of 10 students and 13 professionals in total). It is also maintained that cognitive effort[10] plays a significant role in both productivity as well as attitude towards working with MT (Daems et al. 2017: 262). In general, both groups displayed similar attitudes – more positive after the task – and there were individual variations in preferences rather than group differences (Daems et al. 2017: 264). Interestingly, student post-editors deemed working with MT less tiring, which was in line with Tirkkonen-Condit (1990) who found that for trainees, translation is a linguistic task. As reported by Daems et al. (2017: 264), the impression of being less tired after working with MT may be due to the lessened need to search for information, which is typical for students (Jensen 1999). The sometimes seemingly correct equivalents suggested by MT may relax inexperienced translators and thus make them less vigilant as they correct MT. This issue is crucial for the purposes of this thesis which is explained in detail in Chapter 4.

The author of this thesis conducted two studies gauging attitudes towards MT. The first one (Witczak 2016a) was carried out in 2015 during a series of classes on post-editing and involved 21 Master's programme translation students in their first year of the programme. Most of those students had never post-edited MT output before that class exercise, but were introduced into it and the general workings of SMT by the teacher. Having post-edited a medical leaflet, the students were enthusiastic about the general MT quality and accuracy. One student wrote:

[10] For more details about cognitive effort in post-editing, see Chapter 3.

> I didn't expect MT to be precise to such an extent as it was. I was prepared for a lot of editing and looking for terminology on my own. Surprisingly and fortunately to me, MT did that part for me and I only had to challenge a few of its choices. (Witczak 2016a: 44)

The students knew what to expect from MT and were generally positively attuned towards it, but there were also expressions of dislike towards MT and general preference of translation from scratch. After working with a more creative text (a *New York Times* article) they were a bit disillusioned with MT capabilities but in general, having peer-reviewed texts produced by other classmates, they became convinced that MT is useful, albeit in limited capacity.

Another study (Witczak 2016b) – a survey – was conducted to explore the practices related to translation technology among translators whose one working language was Polish. According to the survey, most of them had some experiences with MT post-editing, but they rarely used it in their practice or work. The surveyed groups included 56 professional translators and 24 translation trainees. It was an exploratory study with a small sample and the results can by no means be generalised to apply for all trainees and professional translators working with Polish, but the observations were nevertheless interesting. Most of the respondents had learned to work with MT on their own, but many admitted that they did not know how to post-edit (36% of professionals and 17% of trainees), adding their lack of interest in using it for translation assignments. Professionals and trainees neither agreed nor disagreed with the statement "Machine translation is a useful tool for translators." Interestingly, one respondent elaborated on their choice regarding frequency of doing post-editing by saying: "I do not post-edit as such, but I sometimes check what word (not sentence) is prompted by Google Translate for a given expression. Then I judge for myself if this is a good direction." The study showed that the translators sometimes used MT as a dictionary rather than a source of full-sentence drafts that would constitute classic post-editing.

To sum up, both Cadwell et al. (2018) and Daems et al. (2017) indicate that attitude towards MT can be largely dependent on the level of involvement in the development process and familiarity with the workings of a given system. Some of the findings of these studies echoed responses reported by Guerberof Arenas (2013), especially in terms of concerns regarding increased perceived effort and lack of enjoyment when working with MT. According to Cadwell et al. (2018: 317), the only remedy to this is personal development (webinars, university courses, etc.), especially in the case of freelancers who lack the privilege of direct contact with developers within the same work environment. Also, the lack of metadata or

general awareness that the training data used for the MT engine came from good translators is a factor contributing to the lack of trust and, as a consequence, negative attitude. Cadwell et al. (2018: 317) referred to this as a sense of agency and reported that it differed depending on the company/translation environment and social structures inherent in such workplaces. In general, therefore, these factors seem to play a key role in attitudes towards MT: translation environment, experience with using MT, contact with developers, and the current state of technology. Furthermore, negativity in Guerberof Arenas' study (2013) may in part stem from the fact that a difference of four years between the studies when it comes to technology is substantial. Also, the fact that her participants did not usually perform post-editing may have contributed to the negative attitudes. A comprehensive analysis by Vieira (2020) provided a fresh outlook on the source of anxiety about automation among translators, emphasising the need to regard MT not only through the lens of its own limitations, but together with business practices and the trend towards redefining the profession.

This section already outlined some studies focusing on attitudes towards MT and some of them included post-editing tasks, hence the next section further elaborates on this particular type of translation involving MT.

1.5. Post-editing

Post-editing consists in correcting errors in a text translated via MT from one language into another (O'Brien 2006; ISO 2015). One of the first extensive studies on the process of post-editing was the one by Krings (2001) conducted in 1994. As Lorenzo and Franceschi reported in the MateCat webinar about post-editing (2018), in 2017 MT post-editing accounted for 4.2% of the total language services market, which was worth $ 1.6 billion (out of the total $ 24 billion for the entire translation market), but the percentage is likely to have increased by now (MateCat 2018). Furthermore, they claimed that it is more difficult to establish the actual percentage of translations carried out with some degree of MT – it may be used despite not being declared. Interestingly, Lorenzo and Franceschi referred to post-editing as the "dirty little secret of the industry" (MateCat 2018), which is not that surprising, considering the still prevalent negative attitude towards MT among professional translators. The discourse surrounding post-editing has been revolving around developers and MT service providers attempting to convince

language professionals that post-editing can be an asset during translation (see Section 1.4 about attitudes towards MT).

The demand for post-editing mainly comes from the software/consumer electronics, manufacturing, heavy machinery and equipment industry, as well as consumer products, as reported by Lommel and DePalma (2016). Other areas that employ post-editing to a lesser degree include pharmaceutics, life sciences, legal services, advertising and marketing, and the public sector. The least amount of demand, according to the report, comes from the tourism industry, finances, and education. The data came from 2016 and it is likely that the demand has increased to higher levels.

The main goal of post-editing MT output is improving it, but – as Doherty and Gaspari (2013) put it – not always making it perfect. In terms of the degree of automation, post-editing is a type of machine-aided human translation (henceforth MAHT), as specified by Hutchins and Somers (1997: 148) in Fig. 4.

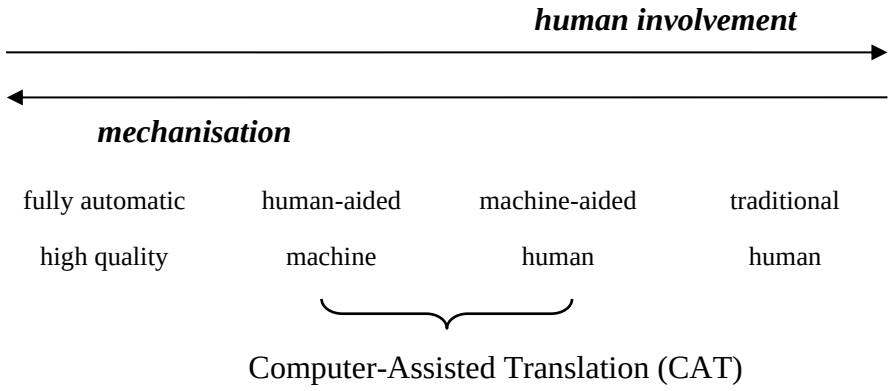


Fig. 4. Human and machine translation (Hutchins and Somers 1997: 148)

Even current state-of-the-art MT systems are still far from the FAHQT (fully automated human quality translation in Fig. 4) ideal criticised in the 1950s and 1960s (Bar-Hillel 1960). Thus, human translators are still indispensable to control and improve the quality of the output, as raw MT output is only occasionally usable and understandable.

Doherty and Gaspari (2013) mention three features of post-editing:

1. Making MT output more usable/understandable
2. Achieving it via least amount of effort (saving time and money)
3. Tailoring the accuracy and extent of post-editing to the needs of specific projects

Accuracy and the degree of editing are related to different types or levels of post-editing, which are summarised in the next section.

1.5.1. Post-editing guidelines

There are usually two or three basic levels of post-editing. For instance, they could include no post-editing, minimum or medium post-editing, and full or complete post-editing (Doherty and Gaspari 2013). The first two types (no post-editing, minimum/medium) usually are intended for internal circulation as opposed to the third type, full post-editing, which is intended for publication. Allen (2003: 297) discerned two factors determining post-editing level: inbound (internal use only) vs. outbound (publishable).

Post-editing is done by following specific guidelines in line with the expected final quality. For instance, TAUS (2016) published such guidelines in 2010 in cooperation with Centre for Next Generation Localization (CNGL). Despite updating them in 2016, they retained the core set of seven guidelines first proposed in 2010 for full post-editing. According to TAUS (2016), similarly to Allen (2003), there are two basic types of post-editing: "good enough" and human translation quality, also referred to as "similar or equal to human translation." Other levels can be created, depending on one's needs. The main difference between these two types of post-editing is that human-like quality apart from being comprehensible and accurate is also stylistically appropriate.

DePalma (2013) defined light post-editing as creating understandable and usable text which is not perfect in terms of style and language and it is easily discernible as created by a machine with minimal human correction. When it comes to full post-editing, it "is meant to produce human-quality output. The goal is to produce stylistically appropriate, linguistically correct output that is indistinguishable from what a good human translator can produce" (DePalma 2013). It is important to stress that such guidelines are usually tailored for not only specific projects but individual LSPs would have their proprietary set of dos and don'ts regarding post-editing. As Hu and Cadwell (2016: 348) posit, only a few sets of guidelines have been published online. From those that are available, Hu and Cadwell (2016) compiled a comparison, including O'Brien (2010), Mesa-Lao's (2013b), Densmer (2014), Flanagan and Christensen (2014), and TAUS (2016) guidelines. According to Hu and Cadwell (2016: 347), most resources dealing with translation propose similar descriptions of light post-editing. However, those about full post-editing differ with respect to certain key issues, e.g. O'Brien and Mesa-Lao do not expect the style to be human-like. Interestingly, it is Densmer and TAUS (as well as DePalma) – those representing the industry – who are inclined to expecting human-

like quality, while scholars lower the quality bar for full post-editing to "medium" (Hu and Cadwell 2016: 351).

The TAUS guidelines (2016) for full post-editing (used for the empirical experiment in this thesis) are as follows:

1. Aim for grammatically, syntactically and semantically correct translation.
2. Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.
3. Ensure that no information has been accidentally added or omitted.
4. Edit any offensive, inappropriate or culturally unacceptable content.
5. Use as much of the raw MT output as possible.
6. Basic rules regarding spelling, punctuation and hyphenation apply.
7. Ensure that formatting is correct.

A question of profitability arises as the quality is expected to equal human translations. Densmer (2014) mentioned that with full post-editing there is the possibility that it would require more effort than translation from scratch, which would be the opposite of what post-editing is intended to provide – effort decrease. Densmer (2014) also argued that when clients request the quality of full post-editing but want it as quick and cheap as light post-editing, a possibility of exploitation and damaging rates arises.

In summary, there are different types of post-editing, but the one feature that they all have in common is that they are different from translating from scratch. How different the two processes are is outlined in the next sections.

1.6. The process of translation

This section outlines the relevant cognitive theoretical frameworks describing the translation process, bridging the gap between cognitive processing in the narrow and broad sense through discussing such models as the recursive model of translation (Schaeffer and Carl 2013) and Situated Translation (Risku 2010; Risku et al. 2013; Krüger 2016). In the broad sense, translation is a series of tasks leading to the creation of TT, while in the narrow sense it refers to the translator's mental operations or cognitive processes, as defined by Hvelplund (2011: 11).

1.6.1. The broad sense: Situated Translation

The paradigm most comprehensively encompassing the broad context of a translational situation within cognitive translation studies is the theory of Situated Translation (Risku 2010; Risku and Windhager 2013). Translators work in context and it is therefore crucial to consider both what happens in the mind and outside of it as one system. Cognition is embedded in the brain, body, and environment (Muñoz Martín 2016, 2010; Jakobsen 2017: 41).

Risku (2010) defines the relationship between cognition and external factors in the following way:

> The mind is only one part of the story. We need to find out not only what happens in a translator's mind, but also what happens elsewhere, e.g. in their hands, in their computers, on their desks, in their languages or in their dialogues. Translation is not done solely by the mind, but by complex systems. These systems include people, their specific social and physical environments and all their cultural artefacts. (Risku 2010: 103)

These artefacts, as put by Risku and Windhager (2013: 36), are the tools included in the modern translator's workplace, i.e. text processors, online research tools, among others, as they co-create the translational ecosystem (Strohner 1995; Krüger 2016) and are a part of the extended cognition system. Jakobsen (2017: 41) referred to translation with the aid of technology as "the result of the joint efforts of many people but also as the outcome of a translator's dialogue with a technological system that communicates with the translator like a kind of colleague." Risku and Windhager (2013: 36) consider Situated Translation a methodological consequence of the mind "leaking" into its environment, both social and technical. Situated Translation, according to Krüger (2016: 118), is emergent from the interaction between humans and artefacts present in their translational ecosystems. The Cologne Model of the Situated LSP Translator (Krüger 2016: 119) includes MT in the artefact group of translation technology in the narrow sense, while Internet resources are listed as part of the digital research and communication resources.

There are four theoretical assumptions of Situated Translation: situated cognition, actor-network theory, activity theory, and agency theory. Hutchins (1995, 2000) viewed extended or situated cognition as similar to the environment of an aircraft cockpit, in which the instruments and co-pilots function as an interrelated cognitive unit. Risku and Windhager (2013) list the actor-network theory (Latour 2005) as a framework supporting situated cognition as it takes into account the interaction of human actors and non-human actants, exploring the

interdependencies of people and things. Apart from the actor-network theory, activity theory also supports situated cognition, emphasising tool-mediated and object-oriented aspects of human actions (Leontiev 1977[1972]; Risku and Windhager 2013: 37).

Another crucial theory for the emergent cognition in Situated Translation is the agency theory (Koskinen 2010) which is connected with the notions of "mangle of practice" as well as "dance of agency" (Olohan 2011; Pickering 1993). This "mangle of practice" as understood by Olohan (2011), following Pickering (2010: xi), was the "temporal structuring of practice as a dialectic of resistance and accommodation". Furthermore, the "dance of agency" in translation is defined by Pickering as the interplay of interacting human and material agency, striving to stabilise each other very much like in the case of engineers working to stabilise a river – both parties engage in a negotiation of agency in their interaction (Olohan 2011: 344; Pickering 2008). These notions are crucial in analysing the interdependence of the translation process and attitude towards artefacts, i.e. technology. Olohan's (2011) study focused on the material and human agency in interactions with TMs. She analysed forum posts about TM technology from an online forum and concluded that Pickering's terms can be applied to the interrelation of translation technology and translators, but more direct methodology (e.g. keylogging and eye-tracking from Translation Process Research, henceforth TPR) is crucial to capture the emergent interaction of humans and technology (Olohan 2011: 353f).

Based on these theoretical frameworks within Situated Translation, technology and especially MT, word processors, and online resources can be regarded as artefacts, scaffolding the translation process and being an integral part of extended cognition. While Risku and Windhager (2013: 43) stressed that Situated Translation requires a more natural and less controlled setting of a real translation project instead of laboratory settings of TPR studies, this thesis sets out to implement the main tenets of Situated Translation in an experimental and controlled setting. The reason for this is to increase the ecological validity by integrating artefacts and subjective perceptions of participants without compromising the variables that can be controlled.

The broad understanding of the translation process is described by Hansen (2003: 26) in the following way: "from every pencil movement and keystroke, to dictionary use, the use of the internet and the entire thought process that is involved in solving a problem or making a correction – in short everything a translator must do to transform the source text to the target text." Thus, the context of the translation task encompasses much more than the mental

processes involved in working with the source and target texts. The next section focuses on the micro-level, i.e. the narrow sense of cognitive processing in translation.

1.6.2. The narrow sense: Cognitive processing

As mentioned previously, on the micro-scale level, cognition in translation is regarded in this thesis from the point of view of the cognitive (Shreve and Lacruz 2017) and recursive model of translation (Schaeffer and Carl 2013) with reference to the classical phases of translation defined by Jakobsen (1999).

The cognitive model of translation integrates activities involved in the translation process, such as reading, writing, and transfer, relying primarily on translation expertise (Shreve and Lacruz 2017). Translation expertise is discussed in detail in relation to competence in Chapter 2. The central notion in the process is transfer which is defined in the following manner:

> [T]ransfer can be seen as the cognitive process of selecting the most appropriate linguistic structures from a target language system in such a way as to ensure that all relevant meaning elements and structures present in a source language text or that are required for text comprehension [...] are represented explicitly or implicitly in an acceptable and coherent (to the target culture) target text. (Shreve and Lacruz 2017: 129)

The selection of appropriate linguistic structures in the TT, according to Shreve and Lacruz (2017: 130), relies on the translator's linguistic, textual, and cultural knowledge of similarities and differences of the two working languages as well as ST comprehension. Expertise plays a key role in the selection process, factoring such aspects as metacognitive awareness of the task (Roberts and Erdos 1993; Shreve and Lacruz 2017: 130). The selection in the translation process happens through the processing of perceptual input when the source text is being read with activation within long-term memory (henceforth LTM). Shreve and Lacruz (2017: 130) state that accessing knowledge happens during "a bundled sequence of iterated higher-order cognitive activities that comprise what we understand as the translation task: reading, transfer, and writing." They explain that text translation is usually a linear sequence of these behavioural segments, i.e. translation units: text segment reading, activating transfer, and writing the target text segment (Shreve and Lacruz 2017: 130). These units start with reading activity and end in writing activity, which makes them both cognitive and

behavioural in nature. Units are delimited with pauses and this suggests that text processing within translation is bundle-like (Shreve and Lacruz 2017: 130f). In other words, as Dragsted and Hansen (2008) also observe, production in translation co-occurs with reading, comprehension, and monitoring processes.

The recursive model proposed by Carl and Schaeffer (2013) connects the linguistic forms in L1 and L2 to their conceptual representations. The recursive model is complimentary to the basic notions presented earlier from Shreve and Lacruz (2017). In this model, shared bilingual representations (de Groot 1992) are activated through early priming processes and subsequently followed by monolingual vertical monitoring processes (Carl and Schaeffer 2017: 62; Schaeffer and Carl 2013). In other words, Schaeffer and Carl (2013: 185) posit that during the early processing stage shared representations are accessed automatically in the horizontal process (black arrows in Fig. 5) and the vertical processes monitor the acceptability of output from the previous stage as more context becomes available (grey arrows in Fig. 5). The model is illustrated in Fig. 5.

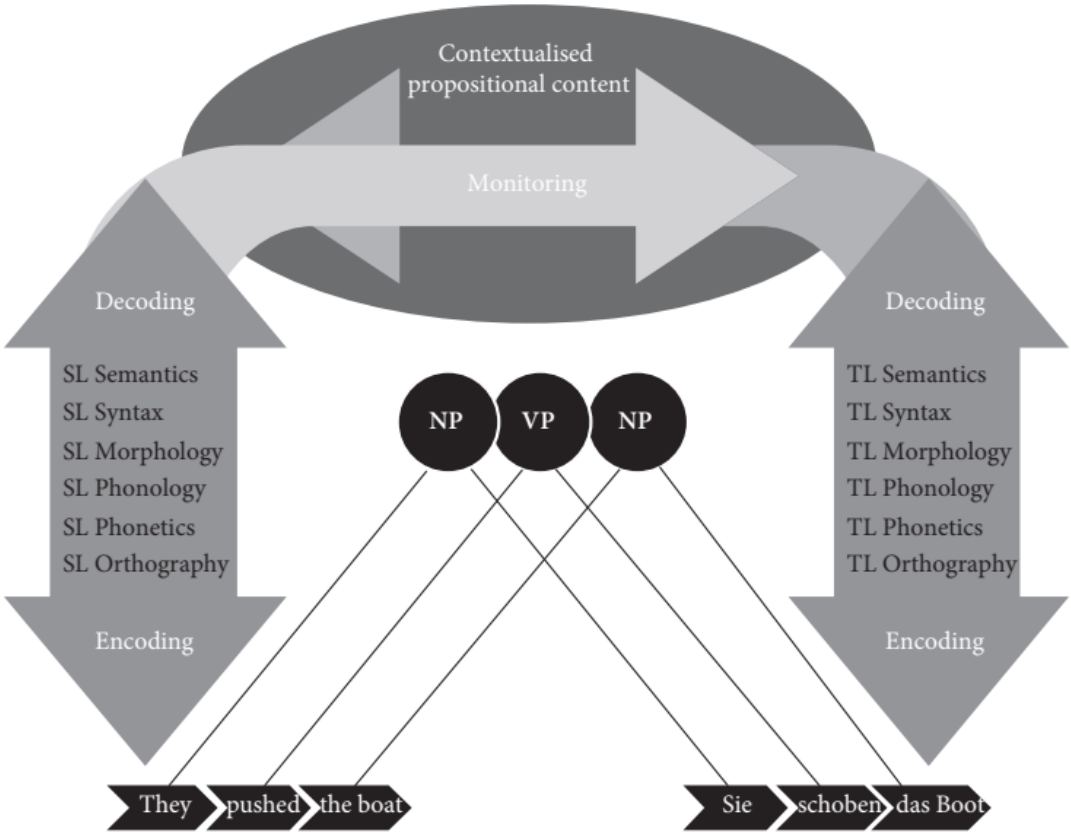


Fig. 5. Recursive model of translation (adapted from Schaeffer and Carl 2013: 182)

From this micro level of cognitive processing in translation, phases of the TT production emerge, as it was established thanks to the keystroke logging studies in TPR (Jakobsen 1999, 2017). The three main phases of translation can be distinguished: initial orientation, drafting, and end revision (alternatively pre-drafting, drafting, and post-drafting). A more detailed taxonomy of translation styles (as briefly explained below) has been provided in a study by Carl et al. (2011: unpaginated), in which 12 professional and 12 MA student translators participated. In Translog (Jakobsen 1999), the participants translated three texts from English into Danish, without access to online resources. The first phase of translation is the initial orientation which consists in the reading of ST before any typing activity occurs. It is subject to individual variation, i.e. some will read the entire text prior to drafting (systematic initial orientation), while others would only scan the text (skimming), read the first phrases/sentences (quick planning), or not read any of the ST at all (head start) (Carl et al. 2011: unpaginated). Carl et al. (2011: unpaginated) reported that the majority of their study participants leaned towards limited context rather than the whole ST.

Once the first character is typed, the drafting phase starts. Carl et al. (2011: unpaginated) distinguishes four drafting styles: large-context planning, small-context planning, backtracking, and non-backtracking. Large-context planning during drafting means that the translator reads a broader context up to a few sentences ahead in the ST. In small-context planning a translator concentrates on up to a few words ahead. When backtracking, translators re-read already translated words in ST. Conversely, non-backtracking means that the translator does not go back to the already translated ST words in a systematic manner.

During the drafting phase, online revisions can be carried out and they may include "the number of text elimination keystrokes [...] [for instance, addition mine, OW] correction of typos, rephrasing of words, phrases and sentences, [...] change of word order" (Carl et al. 2011: unpaginated). End revisions are implemented after the drafting phase has been completed (Jakobsen 2017: 30; Carl et al. 2011: unpaginated). Some translators in Carl et al.'s (2011: unpaginated) study were reported not to make any corrections after the drafting phase, only reading through the typed text, which in general resulted in more time being spent in the drafting phase than during the end revision. However, eight of their participants spent at least 20% of the total task time on end revision – their behaviour was referred to as end revision behaviour. Four of those translators also exhibited so-called constant revision behaviour as they also tended to delete text in online revisions. Online revision behaviour was assigned to those who spent less than 20% of total task time on end revision.

This outline of the translation process both in the narrow sense (translation phases, etc.) as well as in the broader sense (Situated Translation) shows that translation is a complex cognitive task. While some patterns of behaviour in the translation process can be distinguished, individual variation needs to be taken into consideration. The complexity of the task also is reflected in a different kind of effort, i.e. cognitive, technical, and temporal. These types of effort in the translation process, which can be captured by gaze activity and keystroke logging are discussed in detail in Chapter 3. Next section discusses the process of post-editing from the cognitive perspective.

1.7. The process of post-editing

This section summarises the process of post-editing, focusing on the fact that it is a complex cognitive process and mapping its similarities and differences from the process of translation from scratch, outlined in Section 1.6. Post-editing can be seen as a type of translation that includes one more artefact than translation from scratch, i.e. MT output as a draft of the TT. Therefore, the previously discussed Situated Translation on the macro-level applies to post-editing as well. However, a number of differences from the manual translation need to be outlined.

1.7.1. Post-editing in CAT tools

Firstly, from a procedural point of view, post-editing has become much more accessible and user-friendly than it used to be around ten years ago. Luca De Franceschi in MateCat webinar (2018) outlined the difference between post-editing then and now. Having received the ST document from a client, the translator had to create a monolingual TMX[11] file and populate it with MT output. Afterwards, that TMX had to be transformed into a bilingual one and only then could it be uploaded to a CAT tool. De Franceschi (2018) described these five steps as cumbersome for the translation process. Nowadays these intermediate steps are taken care of by the CAT tool that can access MT suggestions directly from the provider.

[11] TMX, or Translation Memory eXchange format, is an XML-based method of exchanging TM data between different CAT tools that ensures little or no data loss (GALA Global 2016).

The use of MT for post-editing was simplified as a CAT tool procedure and, interestingly, the task itself is often considered to be much simpler than translation from scratch (O'Brien 2017: 320). A cognitive model combining the information-processing view and a dynamic connectionist/embodied view of post-editing is currently the most comprehensive way of describing this complex process, as both models provide a different insight into the relevant cognitive processes (O'Brien 2017: 321). Both of these views are described and connected in the next section.

1.7.2. Information processing model of post-editing

From the information-processing view of translation (Winograd 1972), during the production stage, translators behave like code-switchers between the ST and TT. MT output in the post-editing process acts like another "assistant" code switcher (O'Brien 2017: 321). They have to work on two STs, still performing their own code switching, reading, comprehension, and monitoring when fixing MT errors.

The universal model of the post-editing process is described by O'Brien (2017), but it can also be chunked into post-editing styles, according to Mesa-Lao (2013a; Schaeffer and Carl 2017: 150):

1. Reading TT segment → detecting MT error → reading ST segment → fixing MT error
2. Reading ST segment → reading TT segment → detecting MT error → fixing MT error
3. Reading TT segment → detecting MT error → fixing MT error

There is also a fourth type of style in which, as Schaeffer and Carl (2017: 150) report it, before fixing the MT error, the post-editor reviews the previous segment.

O'Brien (2017: 321) lists first the semantic and syntactic analysis of the source sentence and then of the MT output as first two stages of the post-editing process (respectively steps 1 and 2 in Fig. 6 below).

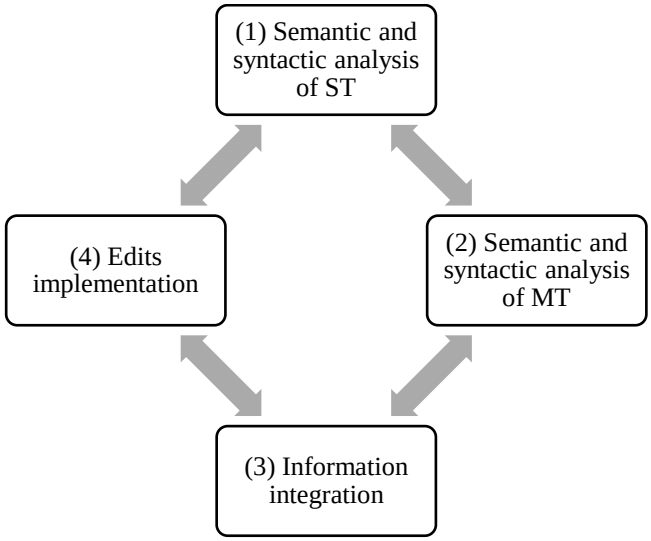


Fig. 6. The process of post-editing based on Winograd (1972) and O'Brien (2017)

Then the translator judges the grammaticality and idiomaticity of ST and MT. Studies suggest that an ST sentence is read before MT output (e.g. Carl et al. 2016), but that is not the only strategy adopted by post-editors – like phases in translation, the process of post-editing is also subject to individual variation. In the case of semantic and/or syntactic issues, re-reading of ST sentences or turning to MT output for clarification are possible. The post-editor will resolve the issues spotted in the MT segment based on their knowledge and information at their disposal and/or contact a third party who may be involved in the translation process (e.g. project manager) to clarify. The processing of ST may be chunked into smaller information units due to human information-processing capacity limitations and carried out in parallel with processing the corresponding MT output. These information units may be compared in smaller chunks, depending on the length of the segment. According to the model, the translator proceeds to the semantic and syntactic analysis of the MT output, considering its idiomaticity as well as grammaticality and comparing it with the ST segment meaning. After that, the information integration stage takes place. The whole process is outlined in Fig. 6. Schaeffer and Carl (2017: 154) also posit that post-editing, similarly to translation from scratch, requires referring to the ST whenever there is semantic and/or syntactic complexity or non-literal language in the TT. This means that errors in MT make post-editors behave like translators do in manual translation. The information processing model is, however, not enough to account for the complexity of post-editing as a cognitive process. The model outlined here is supplemented with the dynamic connectionist/embodied view in the next section.

1.7.3. Dynamic connectionist/embodied model of post-editing

The information integration stage requires the dynamic connectionist perspective. The eponymous dynamic connectionist system is "any system that evolves over time in a law-governed way" (O'Brien 2017: 324). In other words, the system's internal rules influence its development. This is crucial to understand cognition during post-editing, as the information-processing model does not take into account the interdependencies of the situational context. This context in the post-editing process can be seen as the "double dance of agency," in which human agents are involved in a symbiotic relationship with the material agent (Cadwell et al. 2018).

The information integration, according to the dynamic field model (Bermúdez 2014: 417), depends on three sources of input: environmental, task-specific, and memory input (step 3 in Fig. 6). O'Brien (2017: 325) suggests that the environmental input could be the interface used for post-editing and how it constrains or facilitates the process with various visual or functional features (e.g. interactive MT, confidence scores, etc.). Another aspect of environmental input may be related to the physical and organisational features of the workplace, described by Ehrensberger-Dow and O'Brien (2015) as cognitive ergonomic issues. Additionally, environmental input potentially influencing the process of post-editing could be codes of professional practice as well as the status of a post-editor at a given workplace. O'Brien (2017: 325) notes that attitude towards MT and general emotional disposition are also important besides the workplace-related issues. As mentioned before in Section 1.4, requiring "good enough" quality from post-editors may have an impact on the status of translators/post-editors. All these aspects could potentially influence the post-editing process at a cognitive level, according to O'Brien (2017: 325).

The task-specific input refers to the demands created by a given task (O'Brien 2017: 326). Firstly, all the information will be integrated in relation to the post-editing brief, i.e. the target audience expectations, text function, payment, information half-life (i.e. its perishability), the deadline and client, among others. The post-editor will make decisions based on these pieces of information, depending on, for instance, how soon the translation will become outdated. Translation style guides also influence the decisions during the post-editing process, sometimes replacing or supplementing a brief.

Lastly, the third type of input is the memory input which focuses on the connection with one's previous behaviour, e.g. past post-editing tasks or, on a smaller scale, previous sentences

or phrases (co-text) (O'Brien 2017: 326). MT as an artefact in the post-editing process also has the potential to lessen the scaffolding effect of co-text. If the quality of MT is poor, it can have a disruptive effect on cognitive segmentation in a segment, which is similar to increased difficulty of texts (O'Brien 2017: 326; Dragsted 2006).

The final step of the process (step 4 in Fig. 6), as O'Brien (2017: 327) suggests, is the actual decisions taken by the post-editor, i.e. additions, omissions, revisions chosen in accordance with the information integrated from the semantic and syntactic evaluation of the ST and MT. Then, the decisions will be implemented in the form of actual edits with more decisions regarding the order of implementation. Post-editors would sometimes delete the phrase and retype it instead of moving it from one sentence part to another or delete and retype the whole word instead of making a small change, e.g. adding a suffix. This apparent technical inefficiency has a few potential reasons (O'Brien 2017: 327). Firstly, it may be dictated by the cognitive ease of retyping rather than integrating complex operations with mouse and keyboard in another part of the segment or sentence. Another reason may be the lack of technical editing skills or the limitations of the post-editor's information-processing system.

The process of translation has been described in line with cognitive models in TPR as well as Situated Translation. Post-editing process has been summarised in accordance with O'Brien's (2017) dynamic connectionist/embodied view, as it takes full account of MT output as an artefact or an assistant code switcher in this process. The next section introduces a broader notion of translator-computer interaction and translator-information interaction to situate translation and post-editing process in the context of information searching.

1.8. Translator-computer and translator-information interaction

The term translator-computer interaction (henceforth TCI) was coined by O'Brien (2012) and is based on human-computer interaction (henceforth HCI), or "the study of the interaction between people, computers and tasks" (Johnson 1992: 1). While the technology as artefacts has already been extensively discussed in reference to Situated Translation, this section focuses on technology as a means of reaching information in translation. Translator-information interaction (henceforth TII) complements O'Brien's TCI. Zapata (2016: 136) defined TII as "the field of study that investigates translators' interaction with (digital) information and information tools". A key term in TII is information behaviour (henceforth IB), which

investigates "information research strategies, information evaluation criteria, and the modalities and contexts of information use" (Zapata 2016: 140). Translators do not operate in vacuum; they are constantly connected to various information resources, consulting dictionaries, specialised websites, terminology banks available online, etc. As translation and post-editing are both problem-solving tasks, introducing an IB perspective into TPR is essential to understand the broader cognitive perspective of an extended mind. It is thus crucial to explore "the ways of browsing the different sources of information, and of evaluating if the information found is adequate for solving a given problem in order to use it according to the constraints set by the context" (Fidel 2012: 35–37; see also Zapata 2016). With the rapid development of the Internet, translators primarily turn to it for information retrieval purposes (Borja Albi 2008; Simard 2013; Zapata 2016: 149).

It is a fact that technology has an impact on the translation process. Pym (2011: 2) discussed the way technology influences the processing of a linear text, i.e. turns the translation process into a selection of items found in external resources. He described using a search engine as "eliminating items in a vertical movement, searching the one option than might be of help" (Pym 2011: 3). Then, if the search reaches a relevant linear text, reading it in its entirety is likely to be foregone in favour of skimming for relevant information.

Translators already process STs in a segmented way which is imposed by the tool they use and their cognition, as described in the previous sections. This segmented processing proceeds thanks to and in spite of the technological environment of the ST and MT in the case of post-editing. While post-editing the translators have to assess the quality of the MT output and NMT has significantly improved in quality, but it is still imperfect. In the near future, MT is most likely to become a more frequent part of the translation process as an aid for the translator – as one of the many tools used during the process, not just a separate task commissioned by a client.

1.9. Chapter summary

This chapter has reviewed the strengths and challenges introduced by MT into the translation and post-editing process, contextualising both of these processes in relation to relevant theoretical frameworks: Situated Translation, actor-network theory, and agency theory, as well as the dynamic connectionist/embodied model for post-editing in particular. Attitude towards

MT has been established as an important element in the post-editing process, albeit its potential effect or relationship with certain aspects of that process still remain to be thoroughly researched. It is not easy to determine at this point whether post-editing gives advantage over from-scratch translation in general terms. There is much controversy surrounding automated translation and sometimes ambivalent attitudes towards MT persist among translators. Thus, translator competence in relation to technology and information behaviour is the focus of the next chapter. Translators' skills and competence develop and emerge in symbiosis with technology and external resources available primarily online.

Chapter 2: Translator competence and Information Behaviour

2.1. Introduction

The previous chapter explored the essential interaction between translators and technology. This chapter will scrutinise the translation and post-editing process from the viewpoint of Information Behaviour (henceforth IB), but most discussion will be devoted to translation. This is because Information Behaviour in the post-editing process has only been investigated in a few process studies to date, as Section 2.7.2 shows. Key concepts from the IB field connected with web searching behaviour during the translation and post-editing process will be introduced. Also, translation competence and web searching expertise will be described in terms of how they factor into problem-solving via information obtained online. While professional translators' interaction with online resources has been studied to a degree, translation trainees' online search behaviours have been overlooked in TS, as Raído (2014: 4) pointed out. In a more recent study by Hvelplund (2017: 72) on the use of digital resources by professional translators, the author addressed this gap in research. Specifically, he claimed that lexicographic studies have explored dictionary usage in non-translation context, but "it is not known which types of translation tools are used [in the process of translation, addition mine, OW], if there are behaviours that are specific to certain kinds of problems, how professionals use digital resources differently than novices, let alone how much time is actually spent on resource consultation during the translation process." According to Hvelplund's study (2017: 84), web searching during translation takes up around 20% of the total task time (around 25% for specialised texts and 11% for literary texts). He also stated that by investigating the use of digital resources, studies will be more complete in their description of the translation process.

Since this thesis aims at addressing this gap in research, focus will be on selected components of translation competence with emphasis on translation trainees and how they differ from seasoned professionals as well as advanced English as a foreign language (henceforth EFL) students.

2.2. Information Behaviour in the post-editing and translation process

The abundance of often contradictory and unreliable information available online is the reason why searching for information is a specialised skill that translators need to acquire. This thesis adopts the nested view of IB after Wilson (1999), Raído (2014), and Gough (2017), as illustrated in Fig. 7.

Information Seeking encompasses various methods employed to access information resources for translation purposes, including paper and electronic, both offline and online ones. However, Information Searching (henceforth IS) includes only those resources available via computers (Gough 2017: 32; Wilson 1999). Gough (2017: 32) argued that nowadays most resources are accessed via computers and simplified the terminology with an umbrella term *online information behaviour*. In the experimental chapter, the focus is on online resources which for the sake of clarity will be treated synonymously with the category *information searching*. However, this chapter recaps various studies that sometimes involve resources other than those available online: paper resources as well as offline and electronic ones. Hence, the activities involved in using them will be referred to with the broadest umbrella term of information behaviour – unless specified otherwise.

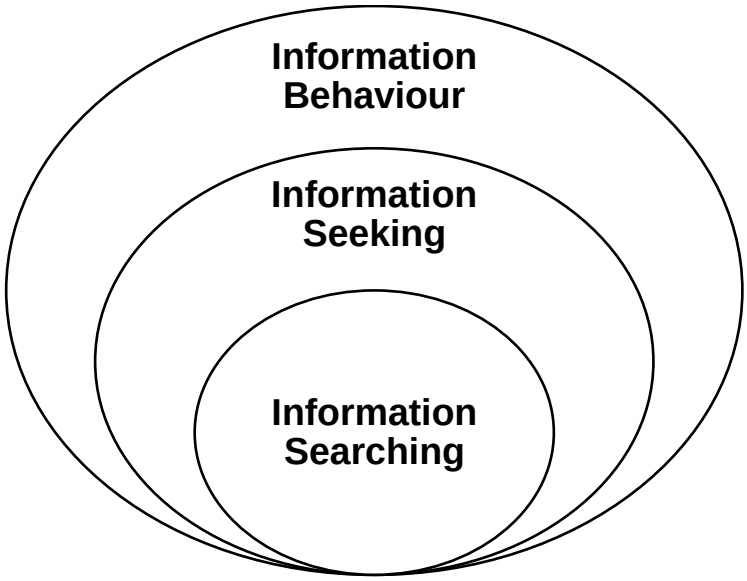


Fig. 7. Nested view of Information Behaviour (after Wilson 1999)

Raído mentioned that the "ubiquity and structure, along with the dispersed and dynamic nature of the information available on the web, pose a set of challenges for the critical evaluation, selection, and use of credible sources of information" (2014: 2). Translators looking

for information online are subject to so-called *infoxication* (Cornellà 2000; Sales Salvador 2006). She also stressed that understanding the process of IS is essential to translate successfully (Raído 2014: 49). Researching information for translation purposes is, after all, a crucial part of the process. Illich (1981: 100) referred to this as unpaid labour in the form of self-service or so-called shadow work, i.e. a "form of unpaid work which an industrial society demands as necessary complement to the production of goods and services." Translation is a form of service, therefore any additional work that is remunerated at a word or page count basis is considered shadow work. Gough suggested that searching for information in translation is shadow work, especially for freelance translators, as one of her participants complained that "translators […] are seen as mere machines. Clients just think that with a click of the mouse the translation is done" (2017: 244).

The latest Eurostat data on Internet use and activities show that the number of people in Poland who go online on a daily basis has increased from 27% in 2007 to 68% in 2019[12] (and to 72.3% in 2020 as per GUS[13] data; in comparison, the increase is from 38% to 79% for the same years in the EU-28 area).[14] In the US in early 2018, 77% respondents reported going online every day.[15] Furthermore, the Pew Research Center examined the use of search engines in the US and, according to a decade-long data collection, it is one of the most popular Internet activities. In 2002, 52% of the US Internet users reported using search engines, while in 2012 the percentage rose to 73%. There is no such data for the European Union populations, unfortunately. Internet use for translation purposes is, however, very different from casual browsing or social media use, but these data for casual Internet users show how ingrained in everyday life the online resources have become.

Also, there is no such statistical data about Internet use among translators on this scale, but it is safe to assume that most of translators nowadays use digital online resources and they do it frequently when compared to offline and paper resources. Zapata (2016: 149) argued that the Internet in general is the El Dorado of knowledge (see Duval 2012: 50) and that:

> The Internet is arguably becoming translator's primary source for information retrieval (Borja Albi 2008; Simard 2013). Few translators still take the time to open, even to carry along their

[12] Source: http://ec.europa.eu/eurostat/web/products-datasets/-/isoc_bde15cua (date of access: 20 Jan. 2021)
[13] Source: https://stat.gov.pl/en/topics/science-and-technology/information-society/ (date of access: 20 Jan. 2021)
[14] The EU-28 area from the years 2013–2020. Source: http://ec.europa.eu/eurostat/web/products-datasets/-/isoc_bde15cua (date of access: 20 Jan. 2021)
[15] Source: http://www.pewresearch.org/fact-tank/2018/03/14/about-a-quarter-of-americans-report-going-online-almost-constantly/ (date of access: 20 Jan. 2021)

> (huge) paper dictionaries, paper term records and language books, to name only a few "traditional" informational resources. On the web, translators can find hundreds of monolingual and bilingual dictionaries, concordancers and biconcordancers, terminology databases, grammar and conjugation guides, encyclopaedia and other documentation; in sum, practically all the information that may be useful when producing a translation. (Zapata 2016: 150)

Digital resources can be defined as aids which are not part of the text processing software and they can be divided into internet and non-internet based ones (online and offline resources, respectively) (Hvelplund 2017: 72). Fig. 8 illustrates the use of both paper and electronic resources as reported by the questionnaire already mentioned in Chapter 1, Section 1.4[16] (Witczak 2016b), which was conducted by the author of this thesis. Translation trainees and professionals reported using online resources very often, while offline resources were used less frequently. Paper resources were even less popular.

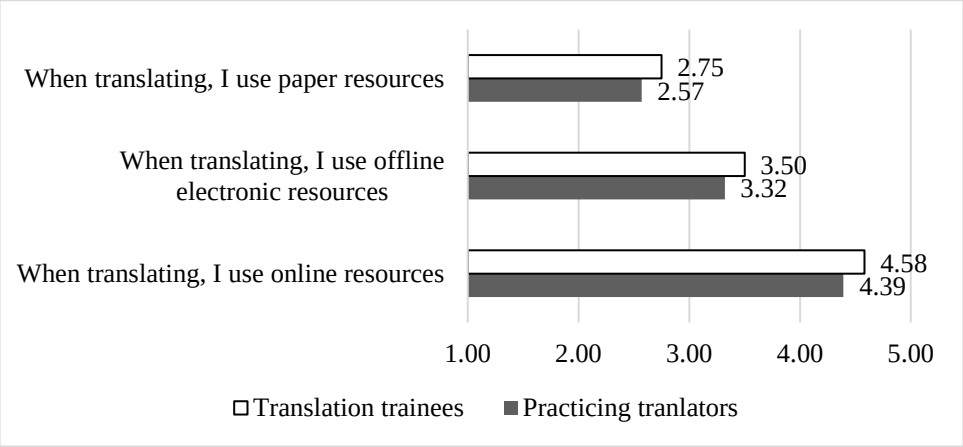


Fig. 8. Paper and electronic resources use among translation trainees and practicing translators

These results are also in line with a later study by Gough (2017: 137), in which she surveyed 540 participants on their preference regarding resources for translation. She found that online resources (henceforth ORs) are the first choice for translators. Following a 2012 study on resources used by translation trainees, Hirci (2013: 162) reported that "[t]rainee translators generally find the use of electronic aids an absolute necessity when translating. […] Search results found on the Internet are in general taken as an absolute authority, as trainee translators seem to trust search results indiscriminately." The absolute trust, however, may lead to incorrect translation solutions.

[16] This information was not included in the published chapter from the book.

Gough drew a clear distinction between resources and tools, defining the former as the content accessed by the translator (e.g. dictionaries, glossaries, term banks, corpora, bitexts, web pages, knowledge-based resources), while the latter is the mechanism or software (e.g. TM, terminology management, term extractors, alignment tools and localisation tools, etc.). Interestingly, she classified MT as a hybrid entity. In the post-editing mode, when MT is consistently applied to the whole process and becomes a task in itself, it is a tool, but it can also be used ad hoc. Gough (2017: 89), therefore, assumed tools to be technologies used during the whole task, but resources assist individual research needs in an ad hoc manner. In her study, which is discussed in the next sections in more detail, MT was used as a tool in a post-editing task by two participants and as an ad hoc resource by the remaining three. The latter type of MT use is also referred to as search and discovery mode, resembling querying a dictionary with words or longer phrases.

Web searching, as defined by Raído (2014: 40), encompasses "goal-driven actions aimed at meeting the research participants' information needs for translation problem solving." She described translation problems as translation units that the participants of her study explicitly identified (Raído 2014: 39). It is very important to consider translation problems as dependent on the translating participant because, as Séguinot (2000: 90) emphasised, "problems do not actually exist 'out there'. It is our perception that identifies something as a problem. In other words, it is the construct of an individual."

There are four main stages of problem solving and IS (Raído 2014: 40):

1. **Search need** (an information need is identified based on the encountered problem);
2. **Search goal** (information type is recognised as potentially being able to satisfy a given information need);
3. **Search process** (all online actions in one or more search sessions in reference to one or more information needs);
4. **Search outcome** (information type potentially used to both satisfy a search need and find a solution to a translation problem).

A single information need involves a web search task which, in turn, consists of at least one search session (Raído 2014: 40). A search session is a key concept in web searching, as it is a "temporal series of online actions aimed at satisfying a specific information need" (Raído 2014: 40; Jansen et al. 2007). Furthermore, Gough's (2017: 65) *research need* was the term equivalent to Raído's search need and referred to the type of information required, i.e.

comprehension, equivalent retrieval, etc. According to Gough (2017: 65), this need is related to the research unit from the ST and is followed by a *research episode*. A *research unit* is a lexical unit, i.e. word or phrase, which triggers a research need. In her analysis, Raído differentiated between common and individual information needs, i.e. those ST units that were selected by at least two participants vs. those chosen only by individual participants.

Paradowska (2015: no pagination) devised a typology of information needs which include the following:

1. Checking the meaning of unknown words
2. Checking the extralinguistic knowledge
3. Checking the accuracy of translated phrases
4. Searching parallel texts
5. Other needs (e.g. accessing an e-learning platform to download the ST)

Once an information need is expressed, a web searcher specifies an information goal which could be a definition, an equivalent, acronym decoding, resolution of contextual meaning, usage in a given context, lexico-syntactic, or spelling issues, among others (Raído 2014: 121). Depending on the nature of the information need and its goal, there are two main types of search goals, i.e. close-ended and open-ended search goals. Close-ended ones yield unambiguous answers, while open-ended searches are exploratory and the nature of their outcomes is not known in advance (Raído 2014: 47). The phrasing of those questions determines the system's response to it, i.e. information retrieval (White and Iivonen 2001: 723). The choice of a search goal determines a strategy adopted to access a search outcome through a search process. The next subsection will address the different strategies that can be adopted in web searching for translation.

2.3. Information Searching strategies

When it comes to web search strategies, there are three main types, i.e. *institutional*, *thematic*, and *keyword searches* (Austermühl 2014: 52). Google is usually the search engine of choice for translators (Gough 2017: 155). *Institutional searches* are carried out via URL addresses and consist in guessing (or knowing in advance) the URLs and visiting the websites of online newspapers or international organisations and then accessing the expert resources like glossaries, databases, documents, etc. via, for instance, that website's internal search engines.

According to Austermühl (2014: 55), the information accessed there may be initially very general, but digging deeper into the contents of such a website will yield more specific information on a given topic. Raído called these internal searches *site queries* and treated them separately from Google searches on account of their different nature and lower frequency of use (2014: 133) Another type of search strategies are *thematic searches* via so-called subject trees, which are based on thematic categories like those that Yahoo used to list[17] (Austermühl 2014: 56). This is an almost obsolete means of searching for information, but there are still websites providing categories listed in a tree-like structure, such as library catalogues (Raído 2014: 50). Finally, *keyword searches* involve typing keywords into search engines in order to access information on a particular topic (Raído 2014: 50). Such searches are said to cause "web blindness," i.e. "a sense that we know there's stuff we want to find, but have no idea how to find it" (Battelle 2005: 32). It is especially in reference to browse searches which involve navigating the web via links, not necessarily on a search engine results page (henceforth SERP), but on websites as well. When browsing, web searchers quickly decide where to click based on swift searches and backtracking (White and Iivonen 2001: 724). This search strategy is, according to Raído (2014: 122), emphasising the changing nature of information needs because of how the queries evolve and become triggers to formulate new searches. Conversely to these exploratory searches, *navigational queries* are used to find an already known website, usually within two steps: typing the keyword and clicking on the link to the website (Raído 2014: 122; Battelle 2005: 31; Hvelplund 2017: 81). These, next to the strategy of directly visiting a website via typing its URL into the address bar, are among the shortest means to access a website.

Keyword searches are considered more powerful, as they rely on a search engine's index put together thanks to web crawlers which automatically connect the web addresses with their text-based content. The index is a database comprising all the pages compiled by a crawler and the search and matching algorithm connects user's query to the index (Raído 2014: 55). Information retrieval in search engines is fully automatic and character-based, therefore not taking into account semantic differences like homographs (e.g. *java* – the language, *Java* – the island) (Austermühl 2014: 60). Finally, the ranking algorithm puts the search results in order based on the location and frequency of keywords on a website (Raído 2014: 55). Knowing the basic mechanics of search engines is crucial for a more conscious and effective web searching (Raído 2014: 57).

[17] Yahoo Directory closed in December 2014 (https://searchengineland.com/yahoo-directory-closes-211784).

Thus, due to their mechanics, search engines are most effective with providing factual and unambiguous information, i.e. proper names or numbers, but typing in open-ended questions as queries will yield few relevant results (Raído 2014: 51). This is due to how artificial intelligence processes language and how it differs from the way humans actually use it. It is referred to as the *vocabulary problem* – if the right indexing term is not known, the desired information cannot be accessed (Peters and Becker 2009: 286; Furnas et al. 1987: 964). Features of natural language, such as polysemy (e.g. *TV* vs. *television*), synonymy (e.g. *java*), morphology (e.g. *television* vs. *televisions*) come in the way of effective information retrieval from search engines (Raído 2014: 51; Carpineto and Romano 2012: 2). Hence, it is vital to narrow down the possible search results with not only precise wording, but also by implementing the correct syntax or search operators. Short queries are less problematic for web searchers, as Aula argued (2005: 17f), and the shorter ones are preferred without operators or modifiers. She also stressed that term choice in queries is crucial, as it determines the relevance of search outcomes.

Relevance in web searching is, however, relative. Depending on the stage of a search task or session, partially relevant documents may prove to be useful to either understand the information problem at the initial stage of searching or give a link to a more relevant document (Spink et al. 1998: 612; Aula 2005: 20). When it comes to SERPs, only the first page per query tends to be looked at, usually checking out the first two results before clicking (Jansen and Pooch 2001; Granka et al. 2004). The more there are results on a SERP, the more efficient it is to sample them instead of examining the retrieved documents in detail (Aula 2005: 21). There are three main types of sampling, according to Sutcliffe and Ennis (1998: 329): serial search (looking at the results one by one), scanning, and systematic sampling of results.

Another web searching strategy as regards search engine use is *Google image search*. This strategy, serving either as a complement to or replacement of regular text-based search, was reported by Hvelplund (2017: 81) to seemingly "identify the meaning of a ST item and to pair a potential TT equivalent with the ST item." Only 4 out of 18 translators employed this strategy in his study, which may be connected to personal preferences in the choice of search strategies and the type of source text.

Furthermore, web searching is said to be non-sequential, which means that the search behaviour is iterative, i.e. users start new searches, when still looking at a given web page from the previous search (Raído 2014: 63). As cognitive effort increases during extended web searching, the whole process is thought to become less effective because of so many browser

windows/tabs are opened at the same time (Kirsh 1995, 2000; Lee 2003, 2005). Thus, a general information-search model can involve the three basic steps:

1. Typing an initial query
2. Examining SERPs
3. Clicking a link or modifying a query

This basic division can be appended with site queries, internal links, and external links, thus allowing to obtain a richer picture of search behaviour and to determine whether it is shallow or deep (Raído 2014: 137). *Deep queries*, as Hvelplund (2017: 81) calls them, serve the purpose of reaching relevant information on relevant web pages via search engine, usually Google. Conversely, when *searching shallowly*, the translator does not go beyond the SERP. Through the use of inverted commas or expression frequency and investigating no further than in the text fragment displayed on the SERP, they will find the information needed. These queries, according to Hvelplund (2017: 81), are the most frequent ones (91% of the queries in his study were shallow). Interestingly, Rowland et al. (2008: 300) pointed out that shallow searching seems to be a more general tendency, not only to be associated with the so-called Google generation (i.e. people born after 1993). To quote, "from undergraduates to professors, people exhibit a strong tendency towards a shallow, horizontal, 'flicking' behaviour in digital libraries. Power browsing and viewing appear the norm for all." The depth of search behaviour as well as query construction will be discussed in detail with reference to translation competence and types of preferred online resources in Section 2.5. However, first a typology of online resources will be outlined in the next section.

2.4. Types of online resources used by translators

Based on studies exploring the use of online resources during translation (Hvelplund 2017: 80ff; Gough 2017), the following types can be distinguished: *bilingual dictionaries and term bases*, *glossaries*, *search engines*, *monolingual dictionaries and term bases*, *reference works*, and *conversion tools*. The resource categories in a study by Daems et al. (2017: 257) include a more detailed division: *dictionary*, *concordancer* (or biconcordancer like *Linguee.pl*), *search*, *encyclopaedia*, *MT*, and *other* (i.e. grammar/spelling websites, forums, news sites, term banks, thesauri). Search engines involve not only the use of text-based Google search, but also Google

Image search. The use of search engines and the strategies involved in Information Searching and retrieval have been described in the previous section.

2.4.1. Dictionaries, term bases, and glossaries

Bilingual dictionaries and term bases can include such websites as *IATE*[18] or, in the case of Polish and English, *Dict.pl*. Hvelplund (2017: 80) indicated in his study that 75% of external consultations were conducted on such resources. Furthermore, monolingual dictionaries and term bases could be used for synonyms and resolving issues with spelling, but the latter is usually taken care of by the word processing software (Hvelplund 2017: 81). Those are not that frequently consulted (17% of study participants used them). While monolingual resources used to be regarded as those favoured by the more experienced translators (Jääskeläinen 1989: 186ff), Gough's data indicated that bilingual dictionary consultations outnumber monolingual resources by a large margin – 152 vs. 28 instances, respectively (2017: 155). Other dictionaries, such as thesauri and idiom dictionaries or specialised ones were consulted even less frequently (Gough 2017: 155).

Whyatt et al. (2021)[19] investigated effort and directionality in IS among 30 professional translators. The participants translated four STs (each around 160 words), two into Polish and two into English (product description and film review). The STs were balanced for readability (Gunning Fog index). Two of the OR categories proposed by the authors are bilingual and monolingual resources. Dictionaries and corpora belonged to both categories, while thesauri and language advice websites belonged in the monolingual category (Whyatt et al. 2021: 10f). They found that, in line with Hvelplund (2017), bilingual resources satisfy most information needs during translation, more so for L2 than L1 translation (on average 8.38 vs. 5.52 number of Google searches, respectively) (Whyatt et al. 2021: 12). The participants used bilingual resources the most when translating the technical texts.

Similarly, glossaries are another type of a concise resource, either mono- or bilingual, sometimes supplemented with visual elements, i.e. diagrams or pictures. Gough (2017: 96) refers to them as picture glossaries. An example of such a picture glossary is visible in Fig. 9 below.

[18] Source: https://iate.europa.eu (date of access: 20 Jan. 2021).
[19] More details about results are described later in Section 3.5.3 about TPR eye-tracking studies investigating effort.

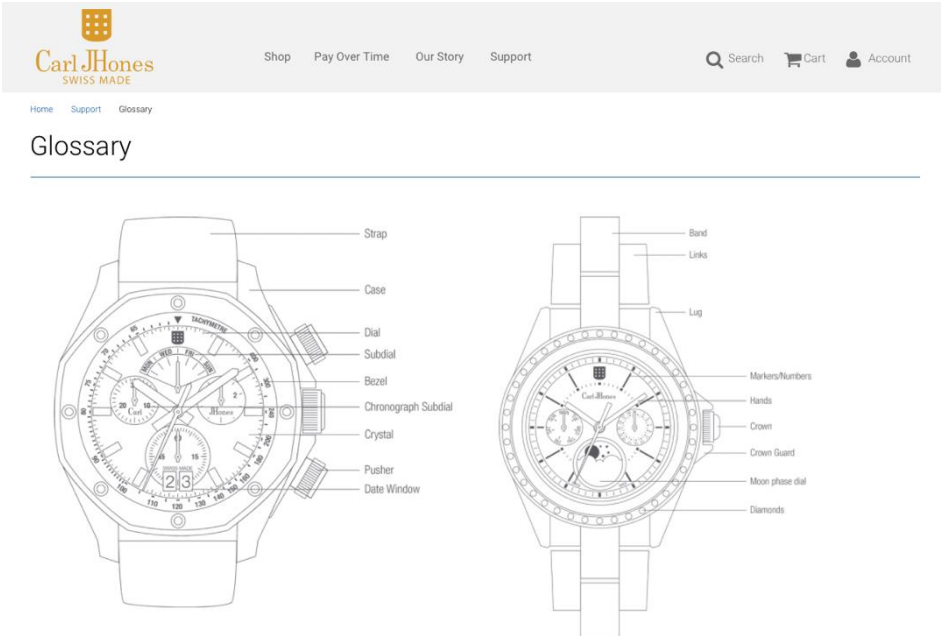


Fig. 9. An example of a visual glossary of watch anatomy from http://carljhones.com/glossary/

Glossaries can also be a collaborative resource, just like in the case of *Proz.com* and *Wikipedia*. The term search function on *Proz.com* allows to browse terminological contributions from the translator community, with confidence votes from member contributors. For instance, an internal search for *closed-back headphones*[20] would yield the result shown in Fig. 10, in which users vote and comment on the posted answer.

[20] https://www.proz.com/search/?term=%5C%22closed-back+headphones%5C%22&from=eng&to=pol&es=1 (date of access: 25 Jan. 2021).

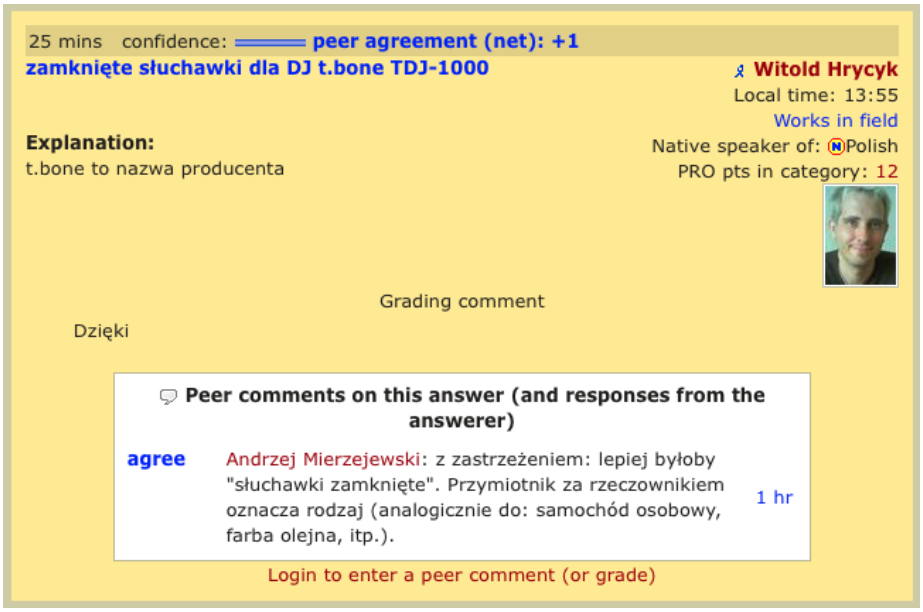


Fig. 10. Proz.com terminology search: peer comments and confidence votes

The above figure contains a peer comment in Polish on the suggested term translation, explaining a terminological rule of putting the adjective after the noun, i.e. *słuchawki zamknięte* instead of *zamknięte słuchawki*. This rule differentiates regular adjectival meanings from categories of nouns (like *closed-back* in *closed-back headphones*).

2.4.2. Reference works

Reference works and websites are another type of online resources. Interestingly, in Hvelplund's study (2017: 82), a third of participants used them, but their use accounted for only 3% of all consultations. Such websites most of the time feature general information texts (like language versions of *Wikipedia*) or texts on a specialised topic and translators use them to access specialised terminology in context (like the website of an international accounting company used by participants in Hvelplund's study, *PricewaterhouseCoopers*, *pwc.dk*), for instance, to gauge its meaning. Raído (2014: 141) had a different understanding of what reference works implied. She called classical reference works all dictionaries (*Meriam Webster*, *The Free Dictionary*, *Word Reference*) and encyclopaedias (*Wikipedia*) beside what Hvelplund would count among reference works, i.e. organisational or academic websites. Gough (2017: 103) used the term *knowledge-based resources* and those included only encyclopaedic resources, i.e. encyclopaedias, wikis, compendia, information databases, but also websites providing computational knowledge. An example of such a computational knowledge engine

is *Wolfram Alpha*. Furthermore, in Whyatt et al.'s study (2021: 10), *knowledge resources* included websites such as *Wikipedia* and *Google Search*. This thesis adopts the category of reference works after Gough and Hvelplund, i.e. as organisational, academic, and knowledge-based websites that are not primarily term bases or dictionaries.

Interestingly, reference works in the form of websites (and their language versions) have become a more frequently consulted type of resource in the last few years. Hirci (2013) conducted a longitudinal study into the use of translation resources among trainees (in 2005 and then in 2012 on a different group of students). She found out that the younger generation in 2012 exclusively used the Internet for consultations, while the 2005 participants also listed paper resources as their go-to aids (Hirci 2013: 155). Furthermore, the most striking shift was therefore from paper to OR, to such examples like monolingual corpora and parallel texts found through Google searches (Hirci 2013: 155f). Moreover, Gough's (2017: 143) quantitative analysis of resource types accessed in her experiment suggest that web pages have become even more important in IS for translation.

Wikipedia as a resource warrants a bit more commentary. It is a reference website in principle, but translators use it not only as an encyclopaedia, but also in the capacity of visual reference, multilingual corpus, dictionary, and a means to reach consensus with clients, for instance on the correctness of a given specialised term (Alonso 2015: no pagination). *Wikipedia* articles are the source of deciphering cultural references and general research on a variety of topics in the context of translation. Interestingly, professional translators considered this resource as positive. As Alonso (2015: no pagination) reported, they described it as free, multilingual, available online, covering many topics, containing images, references, and cultural information, links to other articles and concepts, etc. The translators also mostly agreed that it should be used by people in general as well as translators and students (Alonso 2015: no pagination). From the point of view of web searching strategy, *Wikipedia* in Alonso's study (2015: no pagination) was reported as mostly chosen from a variety of other search results on a SERP, which could be seen as "not usually a part of a pre-planned strategy". Gough (2017: 103) reported *Wikipedia* to be in the top ten most frequently accessed websites in the world. One of the most intriguing aspects of *Wikipedia* as a translation resource is that it combines documentation with terminological, lexicographical, and visual aspects, thus being a comprehensive source of background information on a topic for translation. *Wikipedia* also dominated the category of knowledge-based resources in Gough's (2017: 155) study (109

consultations out of 114 for this type of resource). Paradowska (2015: no pagination) grouped *Wikipedia* among other online databases.

Since online directories are now an obsolete resource, Gough (2017: 105ff) pointed out that the emergence of knowledge portals may be regarded as a form of replacement. She provided *About.com* as an example, but at the time of writing this thesis, the website does not exist anymore. User-curated knowledge bases include portals like *Anwers.yahoo.com* or *Answers.com* or *Quora.com*. An example of such a contribution is provided in Fig. 11 below.

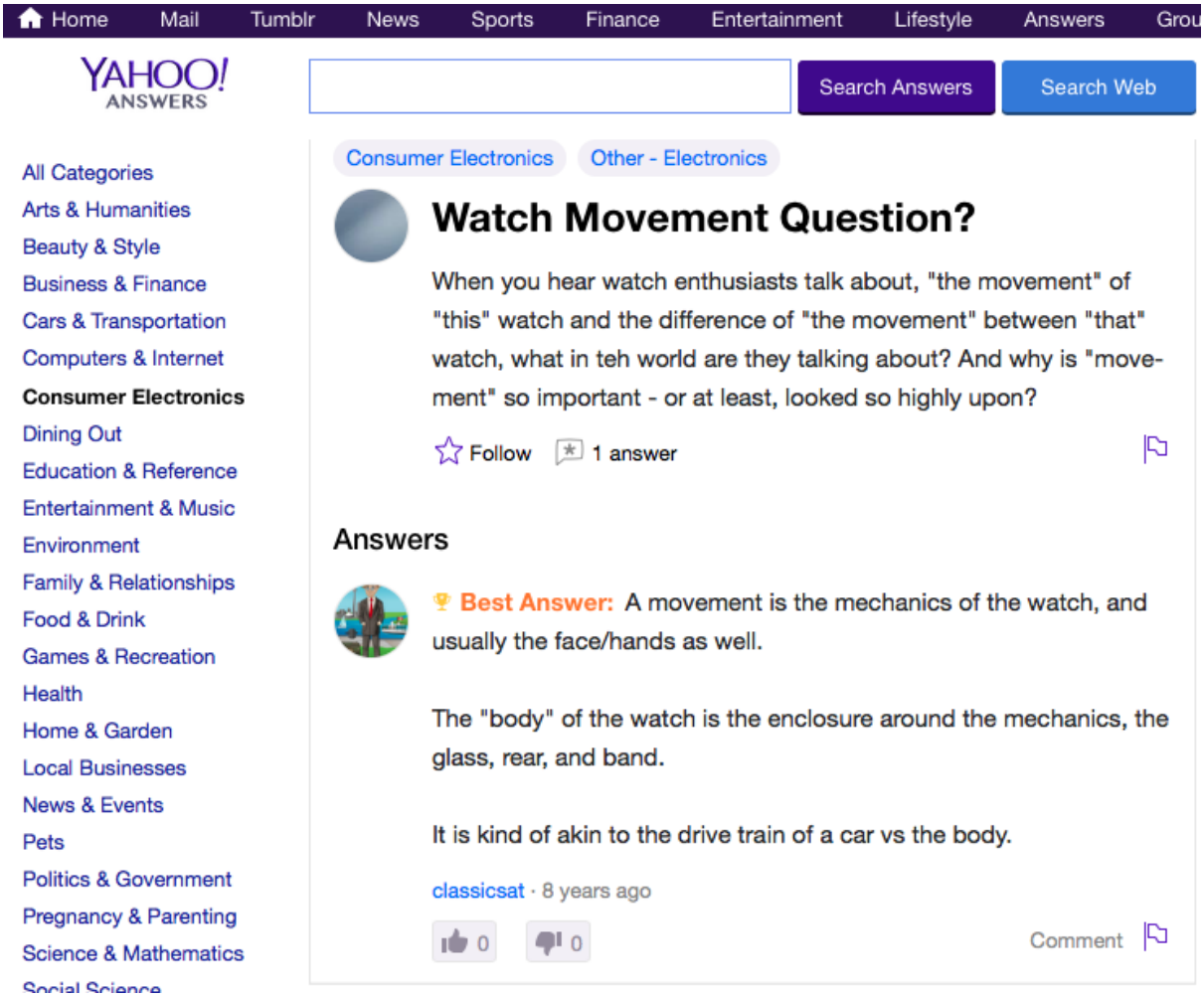


Fig. 11. Explanation of watch movement on Answers.yahoo.com

The reliability of answers provided on such portals is rather low. The community aspect can also be found on *Proz.com*, *WordReference*, and *TranslatorsCafé*, which Gough (2017: 106) described as "'consultation rooms' for peers who often provide expert advice on terminology-related questions or indeed, any other subject." Thus, depending on the perspective, sources like *Proz.com* can be classified as glossaries or discussion forums.

Parallel texts are a tricky subtype of online resources. Raído (2014: 141) counted them as a separate category of ORs. Depending on the type of website that a given parallel text is displayed on, it can be counted as either a stand-alone category or a part of a reference work, e.g. a parallel text that one of her participants accessed via the Greenpeace New Zealand website. Gough (2017: 110f) discussed web pages which can be used by translators in the capacity of parallel texts in order to extract terminology or get background information. Gough also claimed that web pages of various kinds can be consulted to resolve spelling issues, check the use of a given term (whether it is translated or directly borrowed). These types of resources should probably be referred to as comparable texts instead, not to confuse multilingual websites and unaligned documents with aligned texts used for parallel corpora. Kit and Nie (2015: 509) defined them as: "bilingual or multilingual texts about the same topics, […] e.g., Wikipedia (or news) articles in different languages about the same concepts (or events)." For the sake of clarity, however, both of these types of resources will be referred to as parallel texts, as it is done by both of the above mentioned scholars. Corpora are another type of OR, however.

For specialised registers and translating into languages of limited diffusion reference works might be the only source for mining terminology, mainly because glossaries or term bases are scarce or non-existent for either certain fields or language pairs. However, using web pages to search for equivalents can be problematic, because of difficulties with discerning original language content from translated text. Concordancers, which are described in the next section, retrieve key words and phrases from parallel texts and do that extraction for the user.

2.4.3. Concordancers

When it comes to concordancers, *Linguee*[21] is a popular resource among translators and, according to Zapata (2016: 146), it is preferred over dictionaries or term banks. With the proliferation of different types of resources, it is not surprising that dictionaries become decentralised and, in some cases, quickly outdated (Gough 2017: 95; cf. Pastor and Alcina 2009). Gough (2017: 93ff) talked about dictionaries becoming more multipurpose and blurring the difference between tools and resources. This is the case with *Linguee* and *Reverso Context*,[22] for instance. *Linguee* is both a biconcordancer and a dictionary, while *Reverso Context*

[21] https://www.linguee.pl (date of access: 20 Jan. 2021).
[22] https://context.reverso.net/tłumaczenie/ (date of access 20 Jan. 2021).

combines “an example-based dictionary, a concordancer, a search engine, a bilingual aligner and an analysing tool” (Gough 2017: 88).

2.4.4. Other resources

Online documents in various file formats are also a valuable resource for translators. Gough (2017: 112) listed *SlideShare*[23] as an example of repository containing presentations and other files that are easily accessible via simple searches. Resources that could possibly be utilised by translators are also the professional content on websites of companies, organisations, academic institutions, and government bodies – infographics, presentations, reports, text files, spreadsheets, etc.

Finally, conversion tools (e.g. Google unit converter) can be used to convert imperial units into the metric system, e.g. feet into metres. If translators use them, it shows that they are aware of what a potential target audience expects in terms of units (domestication vs. foreignization) and that they know how to solve this specific translation problem (Hvelplund 2017: 82).

Skilfully navigating OR to successfully address information needs is closely connected with a translator’s experience. Thus, the next section will examine the relationship between Translator Competence and IB.

2.5. Translator Competence and Information Behaviour

This section focuses on reviewing relevant subcompetences in relation to IB and MT. Knowledge Integration Network (henceforth KIN) (Whyatt 2012) is a concept drawing attention to the importance of expertise in the translation process. KIN was based on the notion that expertise in translation is connected with the ability:

> […] to draw on all the necessary cognitive resources which are kin to the translation task at hand. These cognitive resources include all kinds of knowledge that the translator has at his/her disposal as well as the knowledge which if needed can become available through external factual research or the use of dictionaries. (Whyatt 2012: 199)

[23] https://www.slideshare.net (date of access 20 Jan. 2021).

With growing experience, translators see the text as a whole and skilfully integrate various kinds of knowledge with their vast bilingual knowledge, so that their choices in the TL become more confident.

Another important source defining competence within the context of IB and MT is the newest ISO (2015) standard regarding professional translation. According to the standard, professional translators have research, information acquisition, and processing competence, i.e. "the ability to efficiently acquire the additional linguistic and specialized knowledge necessary to understand the source language content and to produce the target language content." They are proficient with the research tools and have the ability to efficiently use the information resources they have access to. This is also in conjunction with the skill of using the technical resources scaffolding the translation process. The skills which both KIN (Whyatt 2012) and ISO (2015) emphasise, have been conceptualised in Translator Competence (henceforth TC) models.

There are four main competence models in Translation Studies literature, i.e. by PACTE (Beeby et al. 2009: 208), EMT (European Master's in Translation) (EMT Board 2017; EMT expert group 2009), Göpferich (2009), and Kiraly (2013). The next subsection will explore these models, particularly from the view point of instrumental competence (Beeby et al. 2009: 208; Hurtado Albir ed. 2017), thematic and information mining competence, as well as technological competence (EMT expert group 2009) later consolidated in the EMT Competence Framework (EMT Board 2017). Then, IB as displayed by translators and non-translators will be examined in relation to competences and expertise in TC models.

2.6. Competence models and Information Behaviour

Massey and Ehrensberger-Dow (2011b: 10) pointed out that behaviours related to IB and technological literacy have only recently started being investigated more within TS. Before key concepts and finding related to competence are relayed in this section, the definition of competence itself ought to be established. The EMT Competence Framework (EMT Board 2017) defines competence as "the proven ability to use knowledge, skills and personal, social and/ or methodological abilities, in work or study situations and in professional and personal development." According to PACTE, the instrumental competence is "predominantly procedural knowledge related to the use of documentation resources and information, and

communication technologies applied to translation" (Beeby et al. 2009: 208). The PACTE group described internal and external support in the decision-making process of translation. Internal support concerns using the translator's automatic and non-automatic cognitive resources, while external support entails documentation sources (PACTE group 2005: 612). It is important to note that instrumental competence is indicative of expertise and can be considered one of the characteristics of professional translation (PACTE group 2005: 615ff; Beeby et al. 2009: 227f; Kuznik 2017: 241). Kuznik (2017) in a later study as a part of empirical validation of the PACTE model confirmed that the instrumental sub-competence is a distinct feature of more experienced translators. Translators use more resources, take more time to research information, use different sequences of types of resources than non-translators, i.e. teachers in that study (Kuznik 2017: 241).

Similarly, Göpferich developed a competence model which included the tools and research competence as well as external sources of information and tools available, as visible in Fig. 12 below.

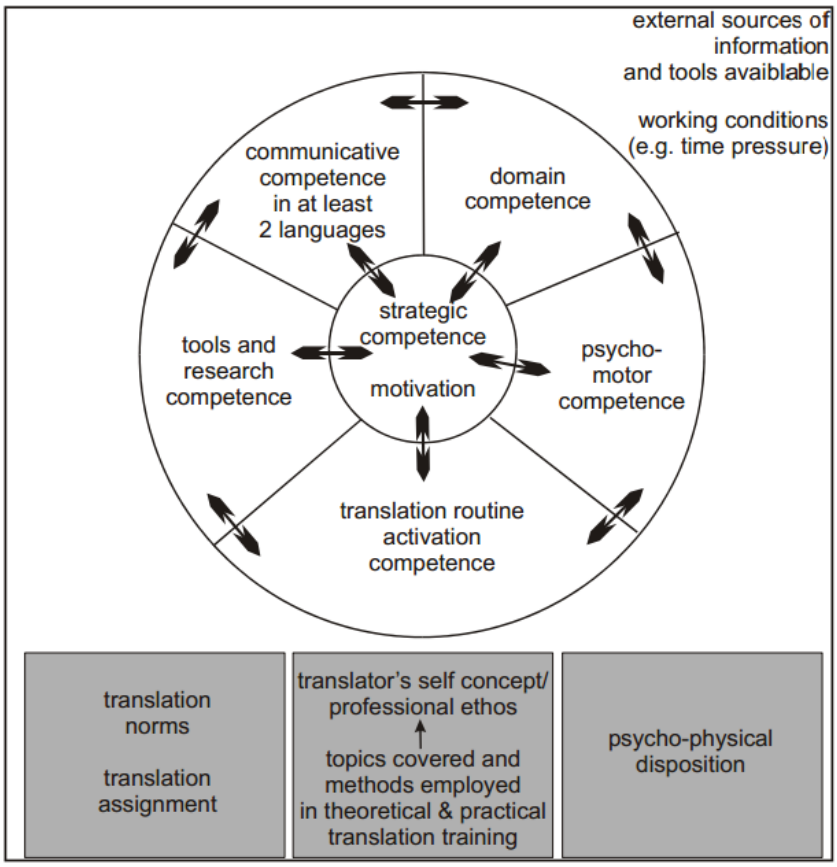


Fig. 12. Göpferich's translation competence model (2009: 21)

The most important aspect of Göpferich's model is that all the subcompetences are interrelated and do not develop independently of each other. The tool and research sub-competence consisted in being able to use (electronic) tools specific to translation (Göpferich 2009: 22). The implicit assumption of both the PACTE and Göpferich's model is that, by definition, professional translators are more competent than novices, for instance, regarding reliance on more global strategies instead of over fixating on the surface level of the text (Daems et al. 2017: 247; Göpferich 2009).

EMT, however, enumerated thematic, information mining, and technological competences as relevant in the translator profession and related to MT use and IB. Thematic competence entails "knowing how to search for appropriate information to gain a better grasp of the thematic aspects of a document" (EMT expert group 2009: 6), whereas information mining competence involves:

C1. Knowing how to identify one's information and documentation requirements
C2. Developing strategies for documentary and terminological research (including approaching experts)
C3. Knowing how to extract and process relevant information for a given task (documentary, terminological, phraseological information)
C4. Developing criteria for evaluation vis-à-vis documents accessible on the Internet or any other medium, i.e. knowing how to evaluate the reliability of documentary sources (critical mind)
C5. Knowing how to use tools and search engines effectively (e.g. terminology software, electronic corpora, electronic dictionaries)
C6. Mastering the archiving of one's own documents

What is directly connected with most of the above, is what technological competence entails – apart from the know-how of TM software and the ability to learn new tools – "knowing the possibilities and limitations of machine translation" (EMT expert group 2009: 7). This is one of the aspects that are of key interest for this thesis, namely trust towards MT and how that manifests in IB.

In the updated EMT Competence Framework (EMT Board 2017), areas of competence are specified. Most importantly, in one such area referred to as *translation* (made up of strategic, methodological and thematic competence) it is stressed that "the ability to interact with machine translation in the translation process is now an integral part of professional translation competence" (EMT Board 2017: 7). Similarly to the wording from 2009, this area is also connected with recognizing the relevance as well as reliability of sources and being able to post-edit on the required level in line with the quality and productivity objectives (EMT Board

2017: 8). What is more, according to the 2017 Competence Framework, the area of technology (connected with tools and their application) includes effective use of search engines and corpus-based tools. It also mentions mastering "the basics of MT and its impact on the translation process" (EMT Board 2017: 9). Moreover, the same area lists the ability to tell whether an MT system is relevant in the translation workflow and whether it is appropriate to use it in a given context.

Furthermore, the issue of trust towards MT and the ability to assess (online) resources echo Kiraly's (1995: 44f) question concerning the strategies behind using dictionaries: "[D]o translators uncritically accept translation equivalents proposed by bilingual dictionaries, or do they use collocations or connotation knowledge (if they have it) to evaluate the proposed equivalents?". Kiraly (2013) developed a dynamic model of translator competence, which stressed the emergent nature and co-existence of all sub-competences. He specifically decided "not to specify the particular sub-competences in the model as there is no consensus on which ones actually exist" and proposed a four-dimensional model, envisioning it with the vortex-like depiction visible in Fig. 13. This emergentist model is distinctly different from the two-dimensional set of sub-competences described above (cf. Göpferich 2009; PACTE group 2005).

The dynamic and emergentist model appears to provide a viable approach as the profession evolves and adjusts to the requirements of the market. Piotrowska (2015: 16) talks about the dispersion of professional roles for translators which include the following: proof-readers, bilingual editors, multimedia designers, research and information specialists, cultural mediators, data processors, product localisers, post-editors, terminologists, project managers, among others. According to Piotrowska (2015), Kiraly's (2013) emergentist model makes sense from a didactic point of view because it accounts for "competing expectations, demands and standards, interdependencies among myriad actors in authentic situations of interlingual, intercultural communication" (Kiraly 2014). Thus, one of the conclusions relevant for this study and its pedagogical implications is that holistic emergentism rather than compartmentalization is important in a translation pedagogy that accounts for the constantly changing reality and skills of the profession (Piotrowska 2015).

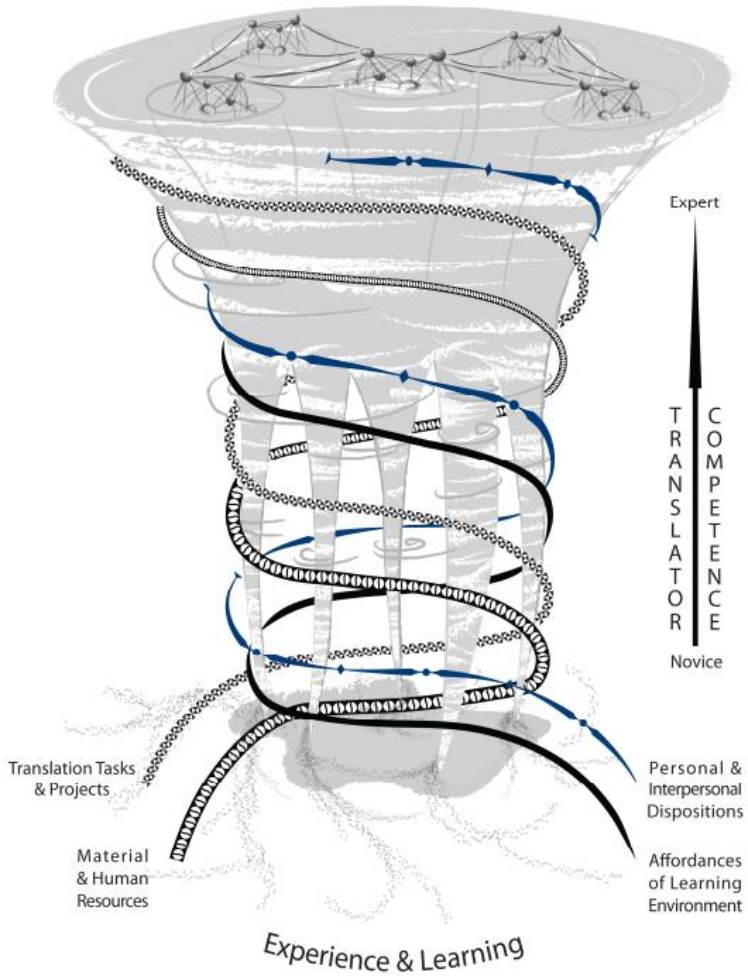


Fig. 13. Four-dimensional model of emergence of translator competence after Kiraly (2013: 211)

The relationship between competence and translation product quality was investigated in a study by Pokorn et al. (2020). They researched three competences (language, thematic, and information-mining competences) in relation to directionality and errors. An analysis of 112 translations by 14 translation trainees was conducted with reference to accuracy (semantics) and style (register and collocations). The students translated four texts, two of which were on a familiar (general) topic and the other two were on an unfamiliar one (specialised) – in L1 and L2. The main finding was that directionality does not have as much impact on quality as L2 proficiency and extralinguistic knowledge have. For this study, the most important result was the one in reference to the information-mining competence. The design in Pokorn et al.'s (2020) study relied on providing access to OR only after the translation was already drafted (after 60 minutes). This introduced a certain limitation in evaluating the information-mining competence – the participants admitted that having produced the draft already, they were selective with their OR access once they could consult them. The evaluation of translation quality based on

solutions selected before and after using OR showed that the best solutions were already selected by student translators without consulting OR (Pokorn et al. 2020: 17).

Having reviewed the TC models above, one may conclude that IB is likely to develop alongside the emergent translation expertise. Whether this development is visible early in the translator training context is an important question for this thesis. While this section discussed the sub-competences involved in IB and interaction with tools such as MT, the next one focuses on behaviours displayed by professional translators and non-translators when performing IS in the translation and post-editing process.

2.7. Factors affecting Information Behaviour in translation

The following sections enumerate all the factors influencing IB in translation.

2.7.1. Review of research in information searching and translation

Tirkkonen-Condit (1990) found that non-professionals rely heavily on dictionary use, treating the translation process as a lexical task. However, with experience in translation, dictionary use is expected to decrease (Jensen 1999). Also, more experienced student translators have a more global approach towards translation and focus on coherence and structure (Tirkkonen-Condit 1990). As a result, internal resources are said to support most of the decisions made by more experienced translators, although external resources are frequently consulted (Alves and Liparini Campos 2009). In general, studies found that professionals are expected to be faster translators than trainees (Tirkkonen-Condit 1990) and to process texts at higher levels (Séguinot 1991), but not resort to external resources as frequently as students do (Jensen 1999). Also, translation trainees tend to use frequency checks in Google SERPs as the basis for decisions in translation, which can be a double-edged sword (Gile 2004: no pagination). While it shows that students may display a descriptive approach towards adopting actual terminological choices used within a given specialised community, it also means that the verification is only superficial. According to Gile (2004: no pagination), such subtleties in filtering results as regarding the cultural context for a given phrase or term are ultimately lost. This means, for instance, the ability to distinguish the varieties of the language used in a particular resource, e.g. American vs. British vs. lingua franca English or Canadian vs. Swiss French.

Vanessa Enríquez Raído (2014) conducted a mostly qualitative study of web searching in translation, and her results will be briefly summarised here in reference to other studies on IB. The study consisted of two translation tasks completed by four translation trainees and two more experienced translators. The ST was a Greenpeace guide to GMO in Spanish and the participants were asked to translate it into English. Two of the students were English native speakers. Task 2 was only completed by the translation trainees. It was a popular science text – a press release by the Spanish National Research Council reporting on the discovery of two new enzymes possibly able to treat AIDS. This text was described as more specialised than the first one.

Raído's (2014: 59) study participants reported using various elements of a SERP to evaluate the relevance of results, i.e. a web page title, the snippet underneath it, the URL address. It is, however, typical of most search users to only look at the first SERP without refining the query (Jansen and Pooch 2001), because it would be "inherently difficult, as cognitive processes are much more effective in handling information about what is present than they are in dealing with information on what is not present" (Aula 2005: 19). Raído (2014: 64) also mentioned that students and more experienced participants navigated the resources differently, i.e. trainees used the back button, whereas expert translators had multiple browser windows or tabs open, which served the purpose of navigating between different topics or reaccessing information.

Multitasking is concept tightly connected with navigating different resources. After Raído (2014: 103), the extended definition of an online action included actions outside of the web browser or other resources, i.e. involving the translation proper. As it turns out, task switching behaviour differs, depending on the level of translation and search expertise. One of Raído's participants, Bob (an experienced professional), switched between tasks the least number of times (29), while a student who produced the lowest quality translation, Maria, displayed "a highly ricocheting behaviour by frequently switching between online tasks" (Raído 2014: 109). An expert translator and searcher, on the other hand, did the following: "ST reading, background research, translation interspersed with selected research, and problem-solving reporting"[24] (Raído 2014: 109). What the inexperienced trainee did, was switch between these tasks far more often than Bob did, also allowing the reporting to interfere with the translation process. Massey and Ehrensberger-Dow (2011b: 6ff) reported on the ergonomic

[24] Problem-solving reporting was a part of the study and was supposed to be carried out at some point in the task, either concurrently or after its completion.

aspects of multiple window management and user interface of OR like dictionaries, encyclopaedias, and search engines. They concluded that these ergonomic aspects are connected with inefficiency of interaction with these interfaces and management of resources, among others. In another study, Massey and Jud (2015) found that students experience cognitive overload when working with novel tools and tasks. In line with that, experience might be connected with strategies during resource consultation. According to Donald and Andreassen (2007: 75), tasks exceeding capacity for conscious control will cause trading off accuracy and speed, whenever multitasking. Thus, inexperienced translators have not yet automatised certain process aspects of translation and it is the experts who will tolerate disturbances to the process without having to compromise either accuracy or speed. Researching information for translation purposes can be regarded as such a disruption.

Gough (2017: 248) reported the development of coping mechanisms with the disruptive nature of interacting with resources: "[S]ome translators plan their research, whilst others simply 'absorb' the information they encounter through meandering behaviour. Some might manage by fast and shallow processing and others by simply avoiding any research that is not deemed absolutely necessary thus reducing their research activities to a minimum." Having compared the results of her study with a pre-internet era study by Nord (2005), Gough concluded that contemporary translation process is interrupted by consulting external resources twice as often (every 1.5 minutes vs. 3.5 minutes in 1997) (2017: 256). The fragmentation of the translation process is also most likely caused by the instant availability of various types of information, i.e. "[j]ust knowing that a piece of information is readily available anywhere and anytime leads humans not to memorise" (Duval 2012; Zapata 2016: 152).

Another information behaviour reported by Raído (2014: 123f) was connected with initial search actions. The expert searcher in the study started research with search engine queries, whereas the inexperienced searchers began with known resources, i.e. dictionaries, regardless of their information goals and needs. This led to the misuse of dictionaries. Search behaviour of the trainees in her study was loyal, sticky, and limited, which manifested in visiting only one website per a given search need and returning to that same website more than once (Raído 2014: 129). Expert search behaviour is thus characterised by validation of solutions through multiple websites. When it comes to further searches, the direct address searches were conducted by inexperienced searchers in dictionaries only, such as *WordReference* (Raído 2014: 131). However, when dictionaries did not provide satisfactory solutions, students resorted to search engine queries (2014: 176). In Task 2, student participants accessed a wider range of

dictionaries and encyclopaedias, resorting less frequently to direct address searches and engaging with the content more (Raído 2014: 159ff). In Gough's (2017: 250f) study, the most prevalent strategy with resources was a bottom-up approach (almost half of her study participants did that) and she remarked that it indicated the translators' knowledge and confidence of their trusted and familiar resources.

The depth of search behaviour is another important aspect of the search process and can be determined by the number of items viewed in a single search session. As far as general web-searching behaviour is concerned, average web searchers would engage with websites in a cursory and brief manner (Nicholas et al. 2006: 210). Jansen, Spink, and Saracevic (2000) claimed that on average, users access 2.35 pages and, in most cases, they click on the results on the first SERP, which can be described as shallow, lazy, unsuccessful (see Nicholas et al. 2006: 210). Another study which Raído references is the one carried out by Désilets (2010: no pagination) investigating whether translators are averse to technology by evaluating their attitudes and work practices regarding translation resources and collaboration/crowdsourcing.

Désilets referred to this shallow search approach as problem coverage, which he defined as "probability that at least one relevant solution [would be, addition mine, OW] found in top 10" and it was more important to student participants than precision (i.e. a relevant solution) (Raído 2014: 177f). Désilets' professional participants felt the same, valuing problem coverage over precision and recall ("percentage of all relevant solutions that is actually proposed by the resource", Désilets 2010: no pagination). Interestingly, Raído points out that her study participants often did succeed in retrieving the desired equivalents while searching shallowly, treating the web as a sort of a metadictionary. It is also important to stress that both search behaviours and their usefulness depend on whether (a) given language(s) in a pair belong to a minority online. Online content is in 60.5% in English, while Russian accounts for only 8.5%, Spanish for 4%, which are the first three, and Polish as a language of low diffusion accounts for only 0.6% of the online content.[25] According to Pavlović (2007: 182), the usefulness of the Internet is dependent on the language pairs involved. In Gough's (2017: 247) study, participants translating into Polish, Hungarian, and Dutch expressed some degree of dissatisfaction with the available resources.

Shallow search behaviour was displayed by most of Raído's inexperienced participants except one. Deep search behaviour is characterised by a preference for browse searches rather

[25] From W3Techs.com, percentages of websites using various content languages as of 21 Jan. 2021 (https://w3techs.com/technologies/overview/content_language/all) (date of access: 21 Jan. 2021).

than direct searches. Seeking thematic rather than linguistic information is also attributed to deep search behaviour. An example of such behaviour (deep and wide) can be that displayed by Bob, an expert searcher, whose "searches aimed at acquiring knowledge on the topic dealt with in the ST" (Raído 2014: 135). Students resorted to search engines when they failed to retrieve satisfactory answers from dictionaries, which Raído reported to be in line with Domas White et al. (2008: 591).

Another feature of deep searching behaviour is query reformulation, which was also something that the expert searcher in the study did and what the students did more in Task 2. Some students did refine their queries by replacing, adding, or removing terms, but only occasionally used operators or modifiers (Raído 2014: 177). In general, deep search behaviour relies on the interactionistic style of IS, which assumes searching for information to be a fundamentally interactive process between people and texts, during which information needs change (Raído 2014: 139; Vakkari 1999: 823).

Shallow searchers do not engage with the visited websites as much as deep searchers do, as they rely on "easy, fast, and more or less cursory visits to a few selected websites" (Raído 2014: 140). As already mentioned, students in Task 2 engaged more with the websites that they accessed and one of the participants adopted more mature search strategies, similar to the expert searcher's depth-oriented behaviour. This may have been caused by the more specialised nature of the STs and their learning experience. On average, the trainees in Task 2 visited three times as many pages when compared to Task 1 (33.5 vs. 10.5) and conducted more browse searches (Raído 2014: 164ff). Also, one student whose search expertise increased most significantly between tasks relied mostly on SERPs in Google to test her terminological hypotheses (Raído 2014: 168).

In Gough's (2017: 153) main task, professional translators on average 54 times accessed 10.2 resources, when translating a 412-word text. Moreover, to solve a single translation problem, on average a participant took 2.6 steps and used 1.8 resources, while in a different study from 2009, there were 1.05 consultations per problem (Désilets et al. 2009). Similarly, Gough reported 34% of research episodes to be one-step processes.

Whyatt's (2012: 343) Translog investigation of the translation process found that inexperienced translators tend to use external resources to make up for their lack of confidence in their choices. Participants translated into their L2 language (English) from their L1 (Polish). Trainees would put a lot of time and effort into consulting external resources to deliver the best quality they can, which is something they are taught and expected to do for classes. While

experts make use of their cognitive resources, novices and EFL students would consciously attempt to integrate knowledge as they go on translating and researching. Whyatt (2012: 344) concluded that for experienced translators individual words generally do not pose problems (albeit with exceptions), because "[t]hey can also resort to automatic interlingual associations which have left memory traces which speed up lexical access." This is in contrast to novices and EFL students, who resort to external help "to find confirmation for their hunches and hypotheses due to low self-confidence on how to best solve the subsequent translation problems" (Whyatt 2012: 343). The study also established an inverse correlation between the years of translation experience and the number of dictionary look-ups, i.e. the more experienced translators resorted less frequently to external resources (Whyatt 2012: 346). Trainees were reported to more frequently resort to external help, which again probably reflected training, thus pointing to the aim of delivering a high-quality text with having as many items double-checked as possible. The results of the ParaTrans project (Whyatt 2018) showed that translation trainees consulted external resources significantly more often than both professional translators and EFL students even when they paraphrased a text in their L1. The study explored the process of translation and paraphrase as performed by translation professionals, trainees, and language students. This finding shows that transfer of training occurs in translation-like tasks as far as consulting resources is concerned, which is relevant for this thesis as well in terms of post-editing. All in all, however, there is the potentially disruptive effect of consulting resources, mentioned earlier in this chapter. Inexperienced translators are particularly vulnerable to such disruptions, according to Whyatt (2012: 347).

Furthermore, query construction also differentiates experts from novices, i.e. their complexity, length, and types (as first described in Section 2.3). However, this issue is beyond the scope of this thesis.

In an empirical investigation of instrumental sub-competence, Kuznik (2017) was able to establish tendencies among translators and foreign language teachers as regards search behaviour and acceptability of the translation solutions. The participants of the study translated English texts about computer viruses either into Spanish or Catalan and were recorded with screen recording software, Camtasia. While the results do not reveal distinct search profiles of translators, some general tendencies can be drawn from the search behaviour displayed in the experimental task. In direct translation (L2 into L1), teachers used fewer different resources (Kuznik 2017: 225). Also, translators devoted more time for searching than teachers (8.06 min vs. 4.41 min on average per rich point in direct translation). Translators whose acceptability

score (translation product quality) was low, spent more time on searches, but for inverse translation the less time was devoted on searching, the better the product quality turned out to be. However, for the purpose of this thesis, direct translation results are mostly of interest. Kuznik found that translators spent more time searching when translating into their L1 than into L2, while teachers spent less time when translating into L1. Translators used mostly keyword, equivalent, and definition searches, while teachers preferred keyword and equivalent searches (Kuznik 2017: 236). Both the difference in the number of different resources for translators (fewer for teachers) and the higher frequency for keyword and equivalent searches were statistically significant, thus making these preferences characteristic of translators (Kuznik 2017: 236). When it comes to the combinations of search types for direct translation, the most frequent one was the combined type (40%), while 41.6% of the teachers did not search at all or counted as the simple type (20%) (Kuznik 2017: 237). Kuznik concluded that based on the different tendencies of the two groups, the final translation quality (especially for translators) increased with the number of resources used. She also concluded that it was distinctly a characteristic of translators to spend a lot of time on researching (on average 16.24% vs. 8.37% of the total task time for translators and teachers respectively) and carrying out many searches.

Another study about the differences in the translation process as regards the level of translation experience along with translation direction, TL, and setting (lab vs. natural) is an eye-tracking, keylogging, and retrospective interview study by Massey and Ehrensberger-Dow (2014). The authors set out to provide a process-based explanation of different levels of competence in their participants. There were 15 beginner translator trainees, 8 MA translation programme students, and 8 professional translators. The results of the study showed that MA students and professionals tend to rely more on their internal resources than beginner translators do. In terms of consulting external resources (whether online only or both offline and online is not explicitly stated by the authors), MA students and professionals were able to tailor their selection of resources to the problem identified in the text, not over relying on dictionaries or linguistic aids (Massey and Ehrensberger-Dow 2014: 94). Thus, based on the results of this study, the link between translation experience and IB skills is clear.

Mutta et al.'s (2014) study also concentrated on the use of external resources, but this time solely focusing on EFL students at university level. What this thesis has in common with Mutta et al.'s study besides IB and digital literacy, is the group of participants – EFL students. The study found that students formulated more queries when searching for information in their

L2 and their tasks ranged from retrieving easily accessible factual information (i.e. the height of the tower of Pisa) to much less obviously accessible information (i.e. the frequency of a given radio station). The results reported high individual variation in the choice of search strategies (Mutta et al. 2014: 14).

Interestingly, professionals might display less than stellar performance in experimental conditions due to possible incongruences with their routine tasks, i.e. the experimental text(s) might fall outside of their specialised fields (Daems et al. 2017: 248; Jääskeläinen 2010). Earlier studies found that both expertise level and familiarity with a given domain affect the choice of search terms (Hsieh-Yee 1993) and overlap between own cognitive resources and content from the search results leads to more successful retrieval of relevant solutions (Spink and Saracevic 1997). Hvelplund and Dragsted (2018) in a study on research strategies of literary and Language for Special Purposes (henceforth LSP) translators found that genre familiarity plays a key role in the translation process, i.e. introduces automated behaviours. Also, the study established that LSP translators translate faster than literary translators, as well as employ more advanced search strategies.

A multiple case study by Paradowska (2015: no pagination) investigated undergraduate student web-searching skills. She conducted a repeated measures quantitative and qualitative study with screen recording capturing her students' performance during a CAT tools course. Between the recording sessions, students were instructed on expert searching techniques and other skills relevant for information searching. Paradowska found that in the first recording session, students relied mostly on bilingual dictionaries and wanted to increase their domain knowledge. After receiving training in expert searching techniques, students chose more parallel texts and checked the accuracy of their predictions. They also used search operators more often. Interestingly, the average search time increased for almost all students except one, who also exhibited a deterioration of their web searching skills.

Participants in Hvelplund's study favoured dictionaries and term bases (like *IATE*), but did not resort to specialised dictionaries as much as expected (Hvelplund 2017: 80f) and when it comes to their search behaviour, 91% of their queries were of shallow nature (i.e. they remained on the SERP on Google). Interestingly, one expert searcher from Raído's study (2014: 166) was very meticulous when evaluating the results on a SERP in order to gauge the context of usage and quality of the resource. Students in Raído's Task 2, on the other hand, usually opted for one of the first search results and ignored those further down the list. Image search was used by only four translators, while three accessed monolingual resources (for spelling and

synonyms), five translators used conversion tools (feet into metres), and four translators used only four different reference websites. A possible explanation for this low frequency of access for such resources is that they opt for guessing the equivalent for a given difficult term, which may have negative consequences for the quality of the translation (Hvelplund 2017: 82). An important conclusion from Hvelplund's study is that individual variation is a key feature of information searching behaviour. Translators come up with different solutions and search strategies to the same translation problems, and these solutions are all acceptable, as the study concluded (Hvelplund 2017: 83).

When it comes to the types of external resources favoured by translation novices and experts, Massey and Ehrensberger-Dow (2011a: 198) point out that novices prefer online dictionaries (as well as offline and printed ones), while experts choose search engines or access parallel texts. This is in line with Raído's (2014: 139) findings from the results of Task 1 in her study, i.e. the expert searcher, Bob, almost never used dictionaries and relied on browse searches instead. In Raído's Task 2, students still persisted in accessing the same types of reference websites, which might have prevented them from broadening their search towards more specialised resources (2014: 165). However, there were fewer types of resources accessed (eight in Task 1 vs. six in Task 2). The type of resources accessed in the IS process also seems to determine the accuracy of translation solutions adopted in reference to these sources. Numerous choices provided by dictionaries, especially when the queried term is a polysemous word, may lead to choosing incorrect solutions (Raído 2014: 144). In contrast, relying on search engine queries, trusted sources (like official websites), and encyclopaedias contributed to higher accuracy of retrieved solutions (Raído 2014: 144). In general, the type of resources consulted by different participants of Raído's study depended on their awareness of available translation resources (dictionaries, glossaries, term bases, parallel texts, etc.). As another interesting measure for examining IS behaviour in OR, Gough (2017: 204) measured the variety of types of resources like dictionaries, search engines, and parallel texts. She categorised participants based on the number of resource categories accessed in the course of the task – low, medium, or high number of types.

Wikipedia again deserves a little more attention, as Raído's student participants treated it as a source of background knowledge and a sort of a bilingual corpus. They switched between language versions in articles to find their TL equivalent. The expert searcher, however, employed *Wikipedia* for background knowledge only. In Task 2, student participants used *Wikipedia* even more to solve their terminological problems and to seek background

knowledge. Raído attributed this choice to higher specialisation of the ST in Task 2. The shift towards seeking more extralinguistic information, therefore, could not be entirely attributed to the expertise level, but also to the features of the ST. Gough (2017: 251) pointed out that her participants used *Wikipedia* with full knowledge of its community-created content, as one of the translators believed it to be a useful source for accessing general information.

This section reviewed studies exploring IB within TPR. While some of the reported studies focus on issues outside of the scope of this thesis such as formulating search queries or depth of searching, it was relevant to briefly review them so that the main points of interest in the field are sketched out. The next section provides an overview of studies targeting post-editing process either primarily or in comparison with translation.

2.7.2. Review of research in information behaviour and post-editing

The studies recapped in the previous section focused mainly on translation from scratch, while this thesis predominantly zones in on a very under researched area, i.e. IB in the process of post-editing. Below, three studies on external resources in post-editing are presented (Daems et al. 2016, 2017; Zapata 2016). These are the studies targeting this process from the viewpoint of IB.

In the study by Daems et al. (2016), 10 MA students post-edited four texts and translated four different texts from English into Dutch, working in the CASMACAT[26] tool combined with Inputlog. It was a keylogging study exploring external resource consultation in translation students in translation and post-editing tasks. The study found that types of the resources consulted have less influence on quality than participant variation and differences between texts (Daems et al. 2016: 126). The authors provide interesting insights into online encyclopaedia use and acceptability, i.e. the quality of the text as regarded from the TL point of view. Consulting an encyclopaedia proved to lead to unsuccessful translation solutions, which could be attributed to its factual nature that does not necessarily help with quick retrieval of equivalents (Daems et al. 2016: 130).

[26] More information about the CASMACAT project is available at: https://www.prhlt.upv.es/wp/project/2016/casmacat.

In another study, Zapata (2016) examined translators interacting (in the field of TII) with a biconcordancer tool (BiConc) embedded into the CASMACAT workbench. Seven professional translators post-edited two 4,500-word English medical texts into Spanish in two modes: interactive and traditional post-editing. Similarly to Daems et al. (2016, 2017), Zapata (2016) observed heavy reliance on concordance tools for post-editing tasks. In Daems et al.'s (2017: 257) study, both groups during both tasks frequently used Google search, concordancers, and dictionaries (in total, they all used 22 types of dictionaries). There were six dictionaries that only students used, and nine that only professionals used (Daems et al. 2017: 258). When it comes to concordancers, which were frequently used (similarly to Zapata 2016), *Linguee* was the choice for professionals, whereas students additionally consulted *Glosbe* (Daems et al. 2017: 258). As for search queries within a popular dictionary for Dutch language, *Van Dale*, a higher percentage of students formulated their queries in English (SL in the study) than professionals (82% and 76%, respectively; the authors did not report this difference to be statistically significant, however) (Daems et al. 2017: 258).

Adding to the above observations in relation to Zapata (2016), the Daems et al.'s (2017: 257) study[27] established that while Google search, concordancers, and dictionaries were the most frequently used resources, student translators favoured dictionaries. MT was used more by professionals than students and even during post-editing. The reason for such, seemingly odd double use of the same resource, was to obtain translations for single words and alternative translations in an ad hoc manner, as Raído put it. For students, thesauri rather than MT were used for such purposes (Daems et al. 2017: 257).

The above studies, to the best of the author's knowledge, are the only ones to date that comprehensively explored certain aspects of IB in post-editing. More discussion about the effect of IB on the translation and post-editing process is provided in Section 3.4 from the perspective of time spent consulting external resources, albeit the number of studies is still low when compared to the research on IB in translation from scratch. The contrast in the number of studies shows how under researched this area remains, despite the growing popularity of MT post-editing. Pym (2013) listed a number of – still relevant – questions about IB regarding post-editing. According to Pym (2013: 490f), the issue of knowing what to trust when consulting ORs and evaluating MT suggestions is more of a macro-skill rather than an intrinsically technological skill. The decision-making in IS behaviours can be regarded as a much more global skill rather than mere procedural knowledge of using one specific tool.

[27] The study was introduced with more details in Section 1.4.

2.8. Chapter summary

This chapter reviewed issues related to Information Behaviour, focusing on select IS strategies, types of OR. Another key issue which this chapter explored was Translator Competence with special attention paid to these competences that included IS and interaction with MT – starting with PACTE (Beeby et al. 2009: 208), EMT (European Master's in Translation) (EMT Board 2017; EMT expert group 2009), and Göpferich (2009) and concluding with Kiraly's (2013) model of emergent competence vortex. Finally, research on IS in translation as well as post-editing was reviewed, emphasising how scarce the latter still is.

IB in both post-editing and translation is not only connected with the types of external resources accessed in the process, but also with the effort put into the task. The next chapter focuses on the nature of effort and its operationalisation as well as on methodological solutions in TPR.

Chapter 3: Effort in post-editing and translation from scratch

3.1. Introduction

The previous chapters focused on theoretical aspects and relevant studies pertaining to MT, TPR, post-editing process, and IS. This chapter explores the construct effort in TPR – its operationalisation in post-editing research, as well as in the keylogging and eye-tracking TPR studies. In conjunction with eye-tracking as a methodology which is used for the experimental study in this thesis, the eye-mind assumption is described, along with its strengths and drawbacks. Furthermore, the subsequent section explores different eye-tracking measures, along with relevant studies focusing on effort. Then, data quality control is reviewed before other measures related to effort relevant for this thesis are described at the end: text readability and perception of effort.

3.2. Operationalising effort in post-editing research

In his seminal work on the process of post-editing, Hans Krings (2001: 178ff) described three types of effort, i.e. temporal, cognitive, and technical. Before Krings' classification, research on post-editing included temporal and technical effort (Moorkens 2018: 56). This section focuses on effort in post-editing studies, but in line with Lacruz and Jääskeläinen (2018: 221), "the same processes are involved in post-editing as in translation […], with the difference that post-editing also involves comprehension of a draft target text (the machine translated text) that is available alongside the source text." Thus, the operationalisation of effort discussed in this chapter will be applicable for both post-editing and translation from scratch.

The temporal dimension is the most direct and economically important one, as it translates explicitly into the cost of performed labour (e.g. hourly rates). It can be defined as "the speed or productivity rate of post-editing" (Moorkens 2018: 56) and can be measured with words per second or minute. Hvelplund (2017: 72) operationalised temporal effort as the amount of time per given activity in the translation process. He refers to temporal effort as the indicator of cognitive attention distribution, which involves activities such as dictionary look-ups, searching in *Google* or *Wikipedia*, among others. In a study by Hvelplund and Dragsted (2018), it was found that text type significantly influences the way one's attention is distributed

throughout the task and that specialised texts involve more time spent in online resources (25.3% of total task time as opposed to 11.8% for literary texts). Managing attention in translation can be ergonomically difficult and demanding, even for digitally literate people, due to managing multiple windows (or tabs) which require increased cognitive effort and slow down the task. In line with Kirsh (1995, 2000) and Lee (2003, 2005) the more extensive the online search, the more strain is put on cognitive resources and the less effective the searching becomes. A similar issue with managing multiple windows was also reported by Massey and Ehrensberger-Dow (2011b: 5f), although mainly for beginner translators, showing that even translation students with some experience cope better with keeping track of their online resources. The temporal effort in relation with external resources in translation and post-editing studies is examined in detail in Section 3.4.

Technical effort is another type of effort and consists in deletion, insertion, reordering of text or all three (Krings 2001: 178ff). In other words, it is "the number of actual edits performed by the post-editor, either measured using keylogging software or approximated using automatic metrics, e.g. the hTER metric, developed by Snover et al. (2006), which calculates the fewest possible edits required from a pre- to post-edited segment" (Moorkens 2018: 56). The most popular ways of recording and quantifying technical effort in TPR has been keystroke logging software such as Translog-II (Carl 2012) or Inputlog (Leijten and Van Waes 2013). The latter is described in detail in Section 4.9.

Finally, cognitive effort is a type of temporal effort which cannot be directly observed. As Krings (2001: 179) put it, it depends on the correction effort required to parse and remedy any encountered errors in the raw MT output. The relationship between the three types of effort is illustrated in Fig. 14.

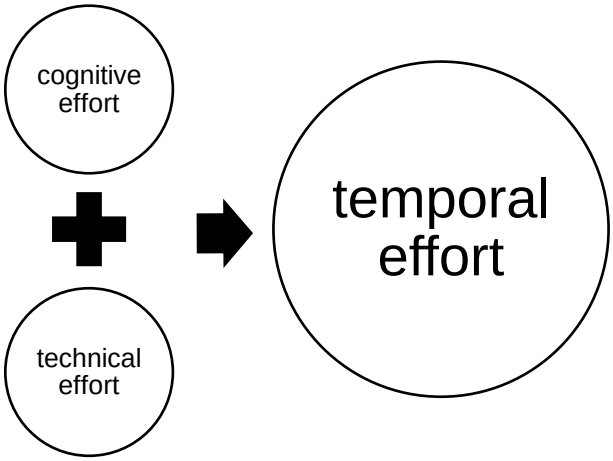


Fig. 14. The dimensions of effort after Krings (2001: 178ff)

It was Krings who first introduced cognitive effort into post-editing research and employed Think-Aloud Protocols[28] (henceforth TAPs) to measure it. Prior to Krings, cognitive effort has been explored in psycholinguistic research, functioning as "demands for controlled information processing or executive function" (Kool et al. 2010: 667) or "fraction of limited attentional resources that are momentarily allocated to a process" (Piolat et al. 2004). Furthermore, cognitive effort has been referred to in literature with various terms. Muñoz (2012: 171) describes the construct of mental load as "common grounds of several theories [...] which use terms such as cognitive load, mental workload", explaining "[...] that performance may be affected once task demands surpass a given threshold of resource availability." According to Halverson (2017: 199), mental load and cognitive effort have been used interchangeably within TPR. However, others point out that cognitive load is more connected with input, while effort is concerned with the reaction of a participant (cf. Paas et al. 2003; Sweller 1988).

This thesis, therefore, uses exclusively the term cognitive effort to refer to the participant's response to the task demands, acknowledging the distinction between mental load and cognitive effort. The construct of cognitive effort is measured and understood as produced in the activity performed by the participant, i.e. as their observable reaction. This is in line with the following three assumptions (Jakobsen 2014) which have been incorporated to study cognitive effort in TPR and this thesis also relies on them in the experimental design and data analysis, as described in Chapter 4:

[28] Think-Aloud Protocols are verbalisations supplied by participants, reflecting on their thought processes, either throughout the task or after it.

1) […] cognitive ("mental") activity has observable and measurable behavioural correlates

2) […] the latencies ("pauses") between such behavioural and microbehavioural manifestations in the UAD [user activity data, addition mine, OW] are as important cues to cognition as the recorded manifestations themselves

3) […] triangulation of quantitative, machine-recorded data with qualitative data elicited from the same event (...) has the potential to lead to stronger hypothesis generation (Jakobsen 2014: 75–77).

It is important to emphasise that these three types of effort usually are operationalised differently. Untangling them and dividing into neat boxes is not as straightforward as the traditional division suggests. Ultimately both the cognitive and technical dimensions contribute to the final temporal aspect of the whole activity. That being said, Section 3.3 illustrates that time can be used as an indicator of cognitive effort, which interestingly shows that the cognitive and temporal dimension are closely intertwined, albeit various studies operationalise them differently. Similarly to psycholinguistic studies using reaction time as an indicator or cognitive effort, TPR studies often employ temporal measures as a proxy for cognitive effort. This thesis focuses on temporal and cognitive aspects of effort, acknowledging their inherent qualitative difference. This is illustrated by the fact that short and complex texts will generate a lot of cognitive and temporal effort, but possibly require little technical effort (i.e. typing) (Lacruz 2017: 387). Conversely, easy but long texts will take long with a lot of typing, but are less effortful in terms of time and cognitive processing (Lacruz 2017: 387).

Also, individual variations in technical effort especially in post-editing can be attributed to different post-editing styles or levels of expertise (Lacruz 2017: 387). Individual variation in translation process is an important variable regarding the cognitive and temporal dimensions, in the case of technical effort, it could be connected with a preference for typing. In particular, during post-editing when the post-editor may prefer to navigate the TT with keyboard shortcuts, thus increasing the technical effort significantly without it meaning that substantial parts of the texts were reworked. Such navigational use of keyboard "is akin to moving a pen or a finger under printed words when reading or revising. When this activity is ongoing, it is likely that cognitive processing is also ongoing" (O'Brien 2006: 16). Conversely, it is important to note that "high temporal effort may be associated with high cognitive effort, but little technical effort" (Lacruz 2017: 387).

A range of methods has been used to measure cognitive effort with thinking aloud (TAPs) as the first and most indirect method. While TAPs proved to be obtrusive to the very

process, they meant to objectively measure (Krings' study reported 30% faster processing speed for tasks without thinking aloud), other less invasive methods of measuring cognitive effort have been employed in various studies on post-editing and translation process. These methods include keystroke logging (Jakobsen 1999; O'Brien 2005) with main focus on pauses (O'Brien 2006; Lacruz et al. 2014), fMRI (Chang 2009), EEG (Doherty 2016; J.-L. Kruger 2016; Tra&Co Team 2016; Hansen-Schirra 2017) and eye-tracking (e.g. O'Brien 2007; Hvelplund 2011; H. Kruger 2016). Eye-tracking and keystroke logging, or keylogging in short, are of particular interest for this thesis.

As Halverson (2017: 201f) reported, more recent studies do not seek out to calculate cognitive effort as an end in itself, but employ it as an additional measure, which is in line with Jakobsen's (2014) third assumption, mentioned earlier in this section. In this thesis, in the spirit of the multimethod approach, the behavioural cognitive effort measurements were combined with other quantitative and qualitative measures, including the characteristics of STs and TTs as well as participant attitudes.

The following sub-sections focus on methodological aspects of both eye-tracking and keylogging, reporting on selected relevant studies that combine these methods to examine the effort in the post-editing and translation process.

3.3. Effort in keystroke logging studies

TPR studies have employed a few keylogging programs to record the process of translation and post-editing. Popular choices have been Translog-II (Carl 2012) and Inputlog (Leijten and Van Waes 2013), but other programmes such as PET[29] and CASMACAT[30] were used in some studies. Keyloggers are able to track the writing activities, logging keystrokes and marking them with time stamps. Such information can be used to extract data about the speed of text production, pauses, and revisions (H. Kruger 2016: 26). For the purposes of this study, the most important feature of keyloggers was the option to track time and activities spent in particular windows (ST, TT) and, for some keyloggers, in OR (Internet browser).

In terms of operationalising effort, keyloggers have been used in post-editing process studies to measure all three types of effort. O'Brien (2005) used Translog (Jakobsen 1999) to

[29] A tool developed by Aziz et al. (2012), available at: http://wilkeraziz.github.io/dcs-site/pet/index.html.
[30] See Section 2.7.2 for the study with CASMACAT.

measure technical effort as the number of deletions, insertions, cutting and pasting actions. She also used Translog to capture cognitive effort in the form of pauses – as pause ratio, defined as "the total pause time as a percentage of the total processing time" (O'Brien 2005: 49). A later study by Lacruz et al. (2012) also argued that pauses correspond to cognitive activity during reading, identifying problems, decisions to post-edit, evaluating solutions, and monitoring processes. In that study, participants worked in SDL Trados Studio and their activity was logged with Inputlog (Leijten and Van Waes 2013). Average pause ratio (APR) was selected as a measure of effort for post-editing (average time per pause divided by average time per word). Furthermore, processing time per word (average word time, AWT) is also employed to gauge cognitive effort for post-editing tasks in keyloggers, as reported by Koponen et al. (2012), who used PET as the keylogger – more time per word suggests increased cognitive effort. Another type of temporal operationalisation of cognitive effort is event to word ratio (EWR), e.g. as reported in Lacruz and Shreve (2014) who recorded the data in Translog. The event here is understood as a complete editing event in the keylogging report, so a deletion or typing of a character. The higher the ratio, the higher the effort as well (Lacruz and Shreve 2014: 257). Lacruz (2018: 237) reported a post-editing experiment employing processing time per word and pause to word ratio to gauge cognitive effort and the results showed a significant difference between processing MT segments without errors (shorter times) than segments with errors (longer times), but found no differences between types of MT errors (mechanical vs. transfer) or performance by professional translators and translation students. They used an online keylogging environment, TransCenter (Denkowski and Lavie 2012). Only O'Brien's (2005) study out of all five studies mentioned in this paragraph allowed the use of external resources. Yet, in O'Brien (2005), Translog built-in dictionary is the resource used by participants, not OR available on the Internet or anywhere outside of Translog. Not allowing the participants to use the Internet or limiting their choice to just one dictionary lowers the ecological validity of the studies.

Ortiz-Boix and Matamala (2015: unpaginated) conducted a study on post-editing audiovisual materials, i.e. producing dialogue lists in Spanish for two excerpts of documentary voice-overs and their keylogging tool was Inputlog (Leijten and Van Waes 2013). Inputlog was chosen due to its unobtrusive nature – it did not interfere with the audiovisual translation software and video playback as it works in the background of other applications. They analysed all three dimensions of effort in accordance with Krings (2001). In general, Ortiz-Boix and Matamala found that post-editing effort is lower, but the difference is only significant for the

first post-edited excerpt of the two – it is faster, requires less keyboard and mouse activity, the difference between average pause ratio (APR) and pause to word ratio (PWR) is smaller. Interestingly, a detailed analysis of time devoted to tasks during post-editing and translation of two video excerpts shows the percentage for consulting online resources. For translation of excerpt 1, it is 67.6% in the main document, 20.7% online, and 6.6% in the video software. During post-editing of the same fragment, 61.4% in the main document, 20.4% online, 8.9% in the video software. The difference is significant between the research and video time. When it comes to excerpt 2, during post-editing 81% was devoted to the main document, 7.1% was spent online, and 7.3% on the video. Translators spent only 59.5% in the main document, 16% online, 13.7% on video. For this excerpt, there are no significant differences, however. The excerpts selected for the study were both comparably long and had similar content features (i.e. narrators vs. expert guests speaking in the voice-over), but the non-significant results might be attributed to the fact that they were not identical in terms of terminology and syntax and BLEU as well as TER scores for MT output. Also, it was a between-subjects design, and with a limited number of participants (12 students).

The first study to use Inputlog in recording activity in external resources in addition to translation and post-editing process was Daems et al. (2016). They used Inputlog in combination with CASMACAT, two keyloggers. The latter programme logs only the activity that occurs within its interface, which is similar to Translog, therefore Inputlog was required to capture online research. This study was already recapped with reference to IB in Section 2.7.2, but conclusions regarding temporal effort are summarised in Section 3.4 below. This thesis used Inputlog to capture keystroke data and record both translation and post-editing as well as research in OR, therefore more details on Inputlog are given in Section 4.9. The next sections focus on temporal effort and OR, as well as the other methodology measuring effort in the experiment conducted for the purpose of this thesis, i.e. eye-tracking.

3.4. Temporal effort and consulting online resources in translation and post-editing studies

Chapter 2 was primarily concerned with more general aspects of IB and competence, hence the conclusions from the literature review regarding temporal effort and consulting OR are recapped here – in translation from scratch and post-editing.

Firstly as far as translation from scratch is concerned, Hvelplund and Dragsted (2018) conducted a study on external resources in the translation process, focusing exclusively on professionals (with experience ranging 4–51 years). Participants translated four texts from English into Danish (two literary texts, 1,984 characters with spaces, and two LSPs, 1,877 characters with spaces – a technical report and a report from a tobacco company). The study established that text type does influence the number of times external resources are consulted and how long it takes to do so. They established that translating LSP texts involves thrice as many consultations of external resources as literary translation (one such consultation per 34 words of LSP and per 108 words of a literary text), but that average time of one consultation is quite similar for both (8.9 s for literary texts and 7.5 for LSPs). Such a difference could be potentially accounted for by the presence of low-frequency terms in LSPs. In general, the temporal measurements in the study were related to the total task time, not specific information needs or rich points.

Furthermore, Gough (2017) argued that some translators work with short translation episodes, but others favour translating without switching to resources for longer times. However, on average, translation episodes lasted for 1 min 27 sec, while research episodes took 1 min 1 sec. She compared these averages with Raído's (2014: 156, 256) 59 seconds average session time for Task 1 and 1 min 51 sec in Task 2 for translation trainees, explaining that task (text) complexity might have an effect on research time (Gough 2017: 168). Admittedly, the two are difficult to compare, as both the texts and experiment designs also differ.

When it comes to the rhythm of the research episodes, they can either be clustered in one particular phase of the translation process or evenly distributed over the whole task (Gough 2017: 247). Gough (2017: 243ff) also found that 30% and 36% of total task time is spent researching when the ST is from a familiar and unfamiliar domain, respectively. She observed that translators tend to switch between the translation and the resources, i.e. backtrack and that they research either retrospectively long after coming across a problem or prospectively – before encountering one. Such a research episode may consist of one or many steps, thus forming a composite episode. The study also established, that translators sometimes return to their research session with new research episodes for the same ST research unit. Another strategy adopted by her participants is indirect research which consisted in detecting potential terms when reading parallel texts. It is also known as information encountering or incidental knowledge acquisition. Such incidental detection of terms happens during the pursuit of another

information need and is stored for later, either in the translator's memory or a term base/glossary (Gough 2017: 170).

Hvelplund (2017: 75ff) also found that almost 1/5 of the translation time was devoted to consulting external resources. Furthermore, participants spent 11.8% of the total task time consulting OR for literary translation, while their online research during LSP translation took considerably more time, i.e. 25.3% of total task time. The author explains this difference by the fact that LSP texts pose more potential terminological issues that require resorting to external resources. What is also very intriguing is that when consulting OR, the participants tended to fixate for longer periods than they did on the ST and TT areas during either of the translation phases. This is a strong indicator of heavier cognitive processing than during translation drafting and revision. The author explained that such a difference may result from the complex nature of resource consultation, as it is an activity consisting of a number of varied tasks, i.e. "looking up words in the dictionary, browsing websites, performing search engine searches, reading encyclopaedia articles, viewing images, and so on" (Hvelplund 2017: 77). For this reason, Hvelplund (2017: 79) also argued that in the case of consulting external resources, it is much more difficult to develop automaticity than it is for any of the phases of translation proper (drafting, revision).

In another study, Daems et al. (2016) concluded that when translating from scratch, spending more time in external resources results in the decrease in errors (2016: 125). This was thought to have been connected with the fact that the participating students either had successful search strategies (having received training in translation from scratch) or that they simply look up easy phrases that would be less error-prone anyway (Daems et al. 2016: 126). They also observed that more temporal effort was devoted to using ORs during translation than post-editing and in general in terms of total task time, translation from scratch required more temporal effort than post-editing (Daems et al. 2016: 121ff). It is crucial to note that time in OR was measured in relation to total task time and the analyses performed in the study involved a normalised ratio of time spent in OR to the number of ST tokens.

Most importantly, however, the study found that translation from scratch required more time to consult external resources than post-editing and that there was no significant difference between the type and number of external resources in translation from scratch and post-editing (Daems et al. 2016: 130). Participants spent more time using dictionaries, concordancers, and search engines than they did using encyclopaedias for both translation from scratch and post-editing (Daems et al. 2016: 130). When it comes to the total task time, it was again translation

from scratch that required more time than post-editing. Raído's (2014: 175) participants spent on average more time translating from scratch the text in Task 2, the reportedly more specialised text among the two in the study (69 min. 16 sec vs. 95 min. 37 sec for Task 1 and 2, respectively). This increase in total task time could be indicative of increased task complexity, which is also connected with an increase in the number of information resources accessed (there were more pages accessed during Task 2) (Raído 2014: 175). When it comes to time devoted to IS, Daems et al. (2016: 122) found that the post-editing set up resulted in less time spent researching information. Also interestingly, no significant difference was found when it comes to types of resources consulted in the course of both post-editing and translation from scratch (Daems et al. 2016: 121–130). Finally, Raído (2014: 144) concluded that higher quality of TT might also be connected with the increase in total task time and effort and information searching is an important part of the whole task (cf. Gerloff 1988; Jääskeläinen 1990; Krings 2001).

However, in contradiction with the previous results from the 2016 study, Daems et al. (2017: 257) found that there was no significant difference in time devoted to external resources, neither between groups (professionals and students) or between tasks (translation and post-editing). This was a different finding than in their earlier study (Daems et al. 2016) – as reviewed above. The study from 2017 applied the same way to measure time spent in OR – the total amount it was normalised per ST token. Their results have also shown that professional translators rely less on dictionaries than students do, which was in line with Jensen (1999) as well as Massey and Ehrensberger-Dow (2014).

The above studies all included temporal effort measures to investigate IS, some of them exploring the time spent in OR in relation to total task time (e.g. Gough 2017, Hvelplund 2017 as well as Hvelplund and Dragsted 2018 or Whyatt et al. 2021 described earlier) or focusing on search sessions or rich points (e.g. Raído 2014 and Gough 2017 as well). Such decisions in terms of operationalising time used in OR might potentially affect the results. The next sections zoom in on eye-tracking studies focusing on cognitive effort measures in post-editing and translation as well as IS.

3.5. Cognitive effort in eye-tracking studies

3.5.1. The eye-mind assumption

Translation and post-editing are both processes involving a combination of writing and reading. To examine effort in reading during translation and post-editing, eye-tracking has been used in TPR for some time. Eye movements recorded by an eye-tracker trace the mechanical movement of the eye on screen and are believed to index cognitive effort involved in processing textual input during reading (Rayner 1998; H. Kruger 2016; Radach and Kennedy 2013). The basis for eye movement data being a correlate of cognitive effort is the eye-mind assumption formulated by Just and Carpenter (1980). In line with this assumption, "there is no appreciable lag between what is being fixated and what is being processed" (Just and Carpenter 1980: 331). However, it is much safer to assume that the visual focus provides only an approximation of cognitive focus (Hvelplund 2014: 209). Due to mind drifting (shifting attention to something unrelated to what is being fixated) and potential disagreement between visual and cognitive focus (looking at the ST and considering possible equivalents), the eye-mind assumption needs to be interpreted with caution (Hvelplund 2014: 209f). Eye movements are said to lag behind the mind focusing on an object up to 250 milliseconds (Holmqvist et al. 2011: 379) and the recorded gaze data may be compromised due to the technical issue of drift, i.e. gradual desynchronization of the recorded eye position and actual eye position, which can decrease the quality of longer recording sessions (Hvelplund 2014: 210f). Also, two issues are likely to affect accuracy of assigning gaze data to single words in larger texts – spill-over effect and increased perceptual distraction to the right (Rayner 2009; Jarodzka and Brand-Gruwel 2017). The former is connected with the reader being likely to fixate on a word while still processing the previous one, as peripheral vision allows one to process and perceive words without fixating on them. The latter, i.e. the perceptual distraction to the right, means that perception of words via peripheral vision occurs without being fixated on (Jarodzka and Brand-Gruwel 2017: 195). Moreover, when it comes to perceiving without fixating, so-called banner blindness is another relevant concept – it is the assumption that when browsing the Internet, people tend to avoid looking at ads. Hervet et al. (2011) investigated it and found that the participants fixated on the ads at least once, but the ads that were congruent with the content of the website turned out to be better memorised than those that were incongruent.

Despite these potential issues, reading research supports the link between visual and cognitive focus in eye-tracking data. For instance, reading less frequent and less predictable words is connected with increased fixation duration (Inhoff and Rayner 1986; Rayner and Pollatsek 1989). But most importantly, Hvelplund (2014: 211) argues that translation is cognitively demanding, which means that "there is arguably little room for much mind wandering, and we may cautiously assume that the majority of eye movements during translation relate to on-going, conscious, synchronous processing of the translation task." Hvelplund argues also that mind wandering is most likely to occur in simple or automatic tasks, according to psychology research (Smallwood and Schooler 2006: 947, 956), thus concluding that the eye-mind assumption with reference to translation is necessary, reasonable, and has been validated by research from both reading and psychology studies.

3.5.2. Eye-tracking measures

To gauge cognitive effort, a number of eye-tracking measures have been used in TPR. They can be divided into four categories: fixation-, pupil-, saccade-, and transition-based measures (Hvelplund 2014: 212). The most popular measures have been those based on fixations and pupillometry. Since this thesis focuses on fixation-based measures of cognitive effort, those measures are recapped in the next paragraphs.

Already in 1976, Just and Carpenter (1976) found that a fixation represents the time to process the fixated word and the longer it takes, the more effort is involved in the processing. According to Duchowski (2007: 46), fixations bring an object of interest into visual focus and they can be defined as a "period of time during which the eye is relatively stable" (Hvelplund 2014: 212). Similarly to temporal effort in keylogging studies, increased fixation duration and fixation count are indicative of more effortful cognitive processing involved in the task performance, but there is a link between the type of task and fixation duration (Hvelplund 2014: 213). Mean durations for different tasks are summarised in Fig. 15 below.

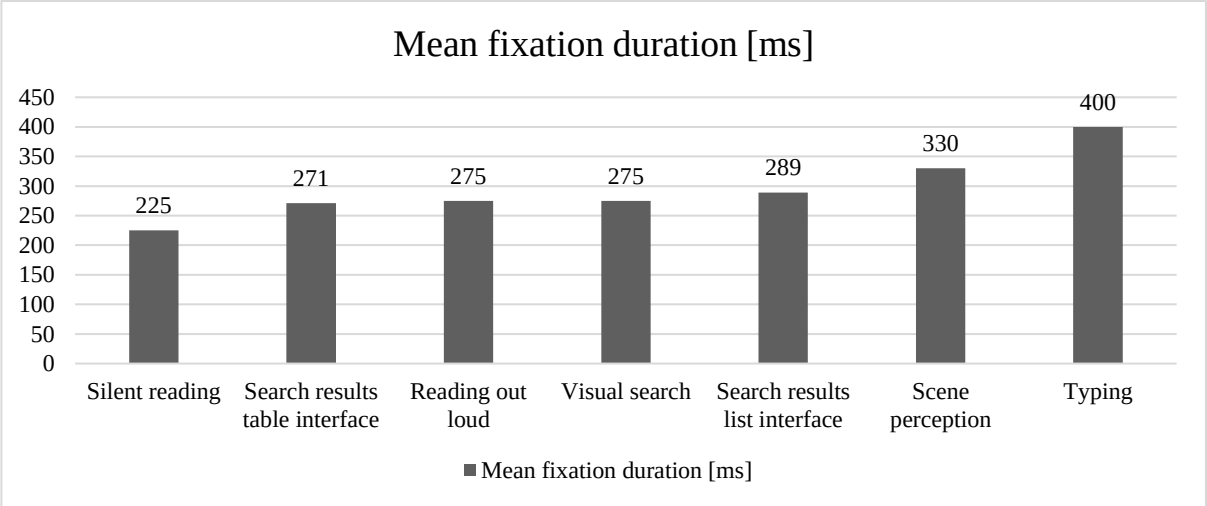


Fig. 15. Mean fixation durations in milliseconds [ms] for different tasks after Rayner (1998: 373) and Rele

Fixations have been found to be longer for visual search[31] (Rayner 1998: 373) or reading search engine results (Rele and Duchowski 2005), but longest for coordinating a visual and motor task, such as typing. Based on the mean durations in Fig. 15, it is important to note that increased mean fixation duration is not exclusively connected with cognitive effort, but the physical action of typing and concurrently monitoring the typing process may affect the length of fixations. More cognitive effort is needed to coordinate these tasks, which illustrates how difficult it is to tease apart the three kinds of effort as traditionally divided into cognitive, temporal, and technical.

The following sections recap the relevant studies from TPR that employed eye-tracking measures in capturing cognitive effort.

3.5.3. Relevant eye-tracking studies investigating cognitive effort

O'Brien's (2007) study was one of the first to include eye movements as a correlate of cognitive effort in TPR, although she used pupil dilation as the index of cognitive effort, not fixation-based measures. Four professional translators edited TM matches (of varying types) and post-edited MT suggestions (from Systran) in SDL Trados. The English ST was a collection of segments, some of them had no matches from either TM or MT. The TL was either French or

[31] Visual search consists in active scanning of a space in order to locate a target among non-targets (distractors) (Burack et al. 2012).

German, depending on the L1 of the participant. Despite being a very small-scale study, it allowed to establish a correlation between processing speed and pupil dilation of different translation memory match types (O'Brien 2007: 199). It turned out that exact matches require the least amount of effort and in general, the effort increases with the match percentage decreasing. Interestingly, MT matches required a similar amount of effort as 80–90% fuzzy matches did (O'Brien 2007: 200). Currently, TPR studies involving eye-tracking usually triangulate the data with other measures, including keylogging and participants' subjective judgements. Lacruz (2017: 389) summarises the triangulation strategy as follows: "More information could be gained by a combination and comparison of different metrics: the whole was greater than the sum of its parts."

According to the already discussed study by Hvelplund (2017: 76), which combined eye-tracking with keylogging in Translog-II, fixations tend to be longer when processing ORs than when processing the text. Fixation durations tend to be 9–54% longer for ORs when compared to the reading in the ST and typing in the TT area. Thus, significantly more effort is put into processing ORs than translation drafting and revision. The only non-significant comparison in the study was between translation drafting and resource consultation for specialised texts. Hvelplund (2005: 5) suggests that the reason for increased cognitive effort is that "the activity of digital resource consultation is composed of a greater variety of underlying tasks than the activities of drafting and revising". Hvelplund (2017: 79) concludes that resource consultation requires more cognitive effort due to its complex and heterogenous nature and may include "switching between a range of tasks such as image viewing, vertical and horizontal reading and writing, these switches of attention will incur some cost in terms of increased processing load."

Hvelplund (2017: 79) used pupil dilation to examine effort in translation and resource consultation and had interesting conclusions regarding automaticity. Automaticity in translation and post-editing is also connected with effort. With experience, some tasks during translation can be performed with less amount of effort due to developing automaticity in certain patterns and schemas. For instance:

> [T]ranslators are more likely to develop and apply automated processing for translation drafting and revision than for digital resource consultation, since 'translation proper' constitutes the bulk of the translation time […]. In addition, the variety of tasks associated with resource consultation makes automation more difficult and less likely for this complex activity. (Hvelplund 2017: 79)

This means that when less effort appears to be exerted in tasks such as these enumerated in the quote above, expertise level is a likely explanation, as also corroborated in Hvelplund (2016: 166) – experienced translators are more flexible in their allocation of cognitive resources and have automated their translation process more than the less experienced translators. This study is one of few studies exploring cognitive effort in relation to the use of ORs.

Another eye-tracking study of OR conducted by Whyatt et al. (2021) tested if the use of OR affected the process of translation in terms of time and cognitive effort. First, they found that the use of OR does affect the temporal and cognitive effort in the translation process. A strong positive correlation was found between total task duration and time spent in OR, albeit stronger for L2 than L1 translation and stronger for the product description than the film reviews. There was also a significant strong positive correlation between the number of searches in *Google* and the number of pauses longer than 10 s and slightly weaker positive correlation between the number of searches and the number of pauses longer than 5 s. The correlation coefficients were similar for both directions and text types. Moreover, there was a significant effect of area of interest (ST, TT, Internet browser) on the cognitive effort measure of average fixation duration – the longest fixations in the browser (M = 320.30 ms) and the shortest in the ST area in the Translog window (M = 222.03 ms). There were no significant differences for browser and the TT area (M = 313.81 ms). Whyatt et al. (2021: 9) concluded that cognitive effort increases during consulting OR regardless of the directionality or text type, but slightly more so when translating into L2. Bilingual resources are most frequently used, significantly more often when translating into English – L2.

The next section details ways to control for data quality in eye-tracking studies and to ensure that less effort in the form of a decreased fixation count is actually due to less effort rather than to bad data quality.

3.5.4. Data quality in eye-tracking studies

Obtaining good quality eye-tracking data in TPR studies is not easy, mostly because the experimental sessions can be rather long and involve working with text, rather than with single words or sentences. Data quality in fixation-based measures of cognitive effort is dependent on a number of factors. Mean fixation duration is one of such measures to calculate thresholds of unacceptably low quality data sets. In TPR, the acceptability threshold ranges from 180 ms

(Sjørup 2013: 105) to 200 ms (Pavlović and Jensen 2009: 99; Hvelplund 2011: 106). Hvelplund (2014: 217), however, points out that also completeness of the whole recording needs to be assessed, i.e. portion of the recording showing no eye movements at all, potentially due to issues such as eye glasses or contact lenses. Also, a participant could just move around and compromise the eye-tracker setting, thus losing focus needed for a significant portion of the recording session.

Another eye-tracking study which investigated cognitive rhythm and effort (Whyatt et al. 2016) reported substantial data loss. The researchers list the following issues decreasing the recording quality: transitions between keyboard and screen, diminished accuracy of remote eye-tracking, long duration of a recording session, and visual impairments (Whyatt et al. 2016: 201f). The faulty data were therefore excluded if the gaze pattern was distorted as could be seen in the data viewing software and based on average fixation duration below 190 ms (Whyatt et al. 2016: 201f).

To control the data quality more accurately than only with mean fixation duration, gaze time on screen (GTS) and gaze sample to fixation percentage are proposed. High GTS score suggests that the participant looked at the screen for most of the task time or the eye-tracker successfully recorded all the actual eye movements (Hvelplund 2014: 217). This is not an entirely reliable measure if the participant spends time using paper or offscreen resources or often looks down onto the keyboard while typing. It can be calculated as follows:

$$GTS = \frac{total\ fixation\ duration}{total\ task\ time} \times 100\%$$

Equation 1. Gaze time on screen (GTS) after Hvelplund (2014)

The other measure, gaze sample to fixation percentage (GSF) relies on the ratio between saccades and fixations. Hvelplund (2014: 218) reports that saccades are 5–15% of all eye movements during reading, so the GSF does the following with the total gaze sample and fixation counts:

$$GSF = \frac{number\ of\ fixation\ gaze\ samples}{total\ number\ of\ gaze\ samples} \times 100\%$$

Equation 2. Gaze sample to fixation percentage (GSF) after Hvelplund (2014)

This measure, according to Hvelplund (2014: 218f), has the advantage over GTS in that it does not set the same gaze time on screen percentage threshold for all participants, although it may be laborious to calculate it from the raw eye-tracking data.

So far, the keylogging and eye-tracking measures of effort have been summarised, but the next section recaps two other measures of effort, i.e. text readability and subjective perceptions of effort.

3.6. Other measures related to effort

As pointed out by Halverson (2017), examining cognitive effort only for its own sake without combing the measurements with different factors is not enough, as numerous TPR studies indicate. This section examines the measures related to the properties of the ST that also affect the cognitive load related to processing the text, i.e. text readability and perceived effort. Readability is a predictor of effort. It is assumed that increased cognitive load will require increased cognitive effort from the translator. The actual effort can also be measured by asking the participants to self-assess the effort they put into the performance of the task. The subjective measure, thus, is the perception of effort in translation and post-editing studies.

3.6.1. Text readability

Text readability can relate to either ST or TT. There is a number of tests than can be applied to assess text readability, e.g. Flesch reading ease score (henceforth FRES) (Flesch 1948), Flesch-Kincaid score (Kincaid et al. 1975), Gunning fog index (Gunning 1952), etc. However, for the purposes of this thesis, only the FRES will be detailed in this section.

OED Online (2019) defines readability as: "[t]he ease with which a text may be scanned or read; the quality in a book, etc., of being easy to understand and enjoyable to read". The crucial keyword in the context of TPR studies is the ease of processing information, i.e. to understand and read. Enjoyability would entail an entirely separate construct requiring different operationalisation. One of the ways to calculate FRES is to use an in-built readability calculator for MS Word. FRES is a readability test based on the average number of syllables and words per sentence (Office Support 2017). It rates a given text on a 100-point scale and the lower the

score, the more difficult the rated text is. The formula relies on the average sentence length (ASL) and average number of syllables per word (ASW). The formula is as follows:

$$FRES = 206.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

Equation 3. FRES formula after Office Support (2017)

Here, ASL is the total number of words over the number of sentences in the rated text and ASW stands for the number of syllables over the number of words (Office Support 2017). FRES was chosen because it does not rely on the US school grade system and only weighs words, sentences, and paragraphs in the text fragment without introducing semantic or terminological variables into the scoring system. It is vital to emphasise that text readability formulas do not determine the text complexity or translation difficulty. They are based on shallow text properties, as discussed above, but are nevertheless used for selecting texts in experimental studies (Hvelplund and Dragsted 2018; Whyatt et al. 2021). Therefore, only a comprehensive approach to measuring effort in the actual translation and post-editing performance can show how much effort was needed for each task respectively.

This section explored objective measures of cognitive load, therefore, the subjective measure of cognitive effort – perception of effort – is described in the following section.

3.6.2. Perception of effort

To supplement the objective measures of cognitive effort, including the effort measurements as perceived by participants themselves is crucial to complete the picture.

Gaspari et al.'s (2014: unpaginated) study integrates objective and subjective measures of productivity, which includes effort, for post-editing as well as translation from scratch. They refer to the distinction as real vs. perceived productivity. Twenty professional translators translated and post-edited wiki texts from two media organisations (*Deutsche Welle* and the *Netherlands Institute for Sound and Vision*). The languages involved were German ↔ English as well as Dutch ↔ English. The tasks were performed bidirectionally, however it was not stated which one was the participants' L1. The participants also filled in a questionnaire about their previous post-editing experience and their impressions of the MT quality that they worked with during the experiment as well as how they perceived post-editing in relation to translation

from scratch. The post-editing product quality was to be publishable on the website. None of the participants had professional post-editing training. The results showed that participants favoured translation from scratch for all examined categories, i.e. speed, effort, as well as their favourite working method. The favoured working method was consistently translation from scratch, irrespective of actual gains in terms of speed and effort. After completing the task of post-editing, the participants were asked which method, i.e. translation or post-editing, involved more effort and they provided their answers on a 5-point Likert scale. When it comes to German-English translations, there was a lot of variation in responses, including neutral opinion regarding post-editing as well as a preference for it or dislike. For English-German, there was a clear preference for translation from scratch. However, for this language combination and direction, 20% participants expressed their preference for post-editing in general, which according to the authors again pointed towards the importance of individual variation regarding such preferences. For the participants who worked with the Dutch-English texts, there was a slight preference for post-editing over manual translation as regards the task speed, but the English-Dutch participants stated that translation was faster. Participants for both directions with Dutch expressed their preference for manual translation in general.

Moorkens et al. (2015: unpaginated) explored two questions regarding the perception of effort in post-editing: whether human estimates of post-editing effort accurately predict actual effort and whether effort indicators for post-editing influence actual post-editing. For the purposes of this thesis, the most relevant research question is the first one. To establish whether human effort estimates are consistent with actual effort for post-editing, participants and independent raters provided their estimates of effort on a 3-point scale (adapted from Specia et al. 2009), deciding whether the MT output required complete retranslation, some editing or (almost) no editing.

The focus was on the MT output rather than ST difficulty. Each MT segment that was later post-edited was rated and received a mean effort score. When it comes to the correlation between the individual effort expectations and the mean derived from all participant scores, there was a moderate positive correlation, which the authors explained to indicate subjectivity and lack of general agreement as to the amount of post-editing effort. Then, the authors compared actual mean temporal effort scores with the estimates and there was only a moderate positive correlation between the respective averages, both for segments as well as for individual participants and both for the expert group of translators as well as the translation students.

For cognitive effort, operationalised as fixation-based measures, the correlations between the actual and predicted scores were weak or very weak. Technical effort and predicted effort also correlated only moderately. Moorkens et al. (2015: unpaginated) concluded that "humans' ratings for predicted PE effort are moderately, but not very strongly correlated to actual post-editing effort, when measured through fixation data" and "do not correlate strongly with the actual time required during post-editing". There was no significant behavioural change in their actual effort when presented with indicators based on predictions in comparison with effort measured without these indicators. An explanation for their results was that the phrasing of the rating descriptions might have biased the participants in favour of assessing technical effort and the amount of editing rather in the first element of the scale and temporal effort in the second ("Requires some editing but PE is *still quicker* than retranslation," emphasis mine, OW).

Vieira (2016) investigated various cognitive effort measures and how they relate to each other, among which subjective ratings were also included. Ten professional and student translators were asked to post-edit two news articles from French into English (from a subset of the news translation task test set at the Workshop on Statistical Machine Translation [WMT] 2017). MT output was obtained from a variety of MT systems to introduce varying MT quality (Vieira 2016: 46). Vieira (2016) based the subjective rating measure on the scale adapted from education psychology that was supposed to capture "the perceived intensity of mental effort" (Paas 1992: 429). The scale ranged from 1 ("very, very low mental effort") to 9 ("very, very high mental effort") (Paas 1992: 430) without labels on the internal levels. Participants filled it in in the PET[32] program when working on the post-editing task. The subjective measure was correlated with average pause ratio, pause ratio, average fixation duration, and seconds per word. The correlations were different in strength and direction, but subjective rating proved to be more strongly associated with objective measures for the group of professionals, i.e. that they might connect temporal effort with subjective ratings of cognitive effort (Vieira 2016: 52). The study also showed that all measures of cognitive effort included in the analysis are connected to each other, albeit to a different degree (Vieira 2016: 59). The participants were not allowed to use any external resources which is a substantial drawback in terms of ecological validity.

Herbig et al. (2019) devised a multimodal study investigating post-editing effort, which integrated subjective and objective measures, involving both behavioural and physiological

[32] See section 3.3, footnote 29 for more information about this tool.

measures. The subjective cognitive effort rating scale was based on the same study by Paas (1992). Participants of the study were 10 native German speakers and translation trainees. The interface used in the experimental setting was SDL Trados. The materials used for the study were from the subset of the WMT 2017 news translation task test set, similarly to the previously reported study by Vieira (2016). The majority of ratings (89.7%) clustered around 3–7 on the 9-point scale. Herbig et al. (2019: 101) suggest that the tendency to choose non-extreme ratings might have been prompted by the wording of the scale. Although the authors claim that their choice of segments included ones that they definitely believed to require "very, very low mental effort" or "very, very high mental effort", the participants might have had trouble labelling the effort with these extreme ratings (Herbig et al. 2019: 101). The objective data gathered from various modalities was correlated with subjective ratings. They found that "using a combination of multiple modalities improves results considerably compared to each modality used alone" and hence the results of correlational analysis for combined features and perceived effort will be reported here. Herbig et al. (2019: 108) found that the only strong correlations (Spearman's correlation at $p<0.001$) were for the perceived effort with post-editing time and the amount of blinking. A moderate correlation was found for the subjective ratings and TER,[33] average fixation duration, galvanic skin response per participant, as well as average pause ratio. Finally, a weak to moderate correlation – for the heart modality, i.e. average root mean square of successive RR interval[34] differences as well as for the average saccade durations. Based on these correlational results, the conclusion is that combining modalities in effort measurement can more reliably estimate cognitive effort, since subjective perception focuses more on stress and exhaustion (Herbig et al. 2019: 111f).

3.7. Chapter summary

This chapter focused on effort and its operationalisation within three dimensions, i.e. temporal, technical, cognitive, in TPR studies on post-editing process. The construct of effort in TPR is complex and needs a clear operationalisation in order to be successfully measured. The complexity is well-illustrated with how difficult it is to find correlations between objective and subjective measures, as the latter might be understood differently by participants. The best

[33] See section 1.3.1, footnote 2 for more information about this tool.
[34] The RR interval is the length between two peaks in the ECG signal

example of such bias was provided by Moorkens et al. (2015: unpaginated): participants might have thought of technical effort of typing rather than temporal or cognitive effort exerted in the course of a task. While it is clear that consulting OR takes up a considerable amount of time spent on translation and post-editing, the studies focusing on the use of OR also sometimes report contradictory results in terms of cognitive and/or temporal effort (e.g. the two studies by Daems et al.: first from 2016 and 2017). Finally, as established in Chapter 1 – attitude towards using MT output, especially negative, also might be related to the use of OR when post-editing. The following chapter describes the experimental study conducted for the purpose of this thesis. The focus is on effort allocated to information searching in translation and post-editing and how it intersects with the participants' attitude and with the translation product (effectiveness).

Chapter 4: Information searching in translation and post-editing: An empirical investigation

4.1. Introduction

This chapter introduces the experimental study exploring information searching in translation and post-editing. Study design along with independent and dependent variables are recapped. Seven hypotheses which are tested in the experiment are briefly introduced. Then, participants, materials, and tools are described. Data analysis is divided into four sections: focusing on process, product, questionnaire data, and statistical analyses. The results section is arranged with respect to each hypothesis and supplemented with a qualitative analysis of open questionnaire answers about MT and consulting OR. The results of each hypothesis are discussed and related to theories and studies introduced in the previous chapters. Study limitations are the focus of the next sections, listing weak points of the study as well as potential avenues to be explored in further research. Furthermore, didactic implications based on study results are explored in this chapter, zooming in on Translator Competence in relation to consulting OR. The chapter concludes with a general discussion on the study's implications.

4.2. Aim of the study

The study aims at examining effort, range of resources, effectiveness, and attitude towards MT in translation and post-editing. Statistical analyses are made to establish potential effect of task type and translation training on temporal as well as cognitive effort (as well as technical effort in addition), and resource range during IS. Correlational analyses examine potential relationships between perceived difficulty and temporal effort as well as range of resources. Additionally, a qualitative analysis of student impressions supplements the quantitative analyses.

4.3. Study design

The study adopted a mixed factorial design: the within-subjects and between-groups design to investigate the translation and post-editing process. The within-subjects design was used to gauge participants' *effort* (temporal and cognitive as well as perceived effort), *resource range*, and *accuracy* depending on text type (*technical* vs. *medical*) and task type (*translation* vs. *post-editing*). As for the between group design, *effort* put into text types and task types as well as *resource range* and *accuracy* were studied to investigate potential differences between translation trainees and EFL students (non-trainees). In other words, the effect of independent variables was tested on the following dependent variables: *effort* (temporal and cognitive measures, technical additionally), *resource range,* and *accuracy*.

Furthermore, to examine the translation and post-editing process with their product, correlational analyses were conducted on such variables as *temporal effort*, *accuracy*, *attitude*, *resource range,* and *perceived difficulty*. The experimental study was also supplemented with a pre- and post-task questionnaires.

4.4. Independent variables

The independent variables in this study are *experimental group membership*, *text*, *task,* as well as *research unit* (adapted after Gough 2017: 65). There are two experimental groups in the study: translation trainees and EFL students (non-trainees). The measurements of *effort, attitude, resource range*, *accuracy*, *researched %*, and *percentage of time* are taken from trainees and non-trainees to check for potential differences due to the effect of translation training in the course of graduate studies and correlations. Details concerning the groups are provided in Section 4.7 and operational definitions for variables are provided in the next sections.

4.5. Dependent variables

The main dependent variables (i.e. outcomes) in this study include:

(1) *Effort*: measured with two objective effort indicators, i.e. for cognitive and temporal effort, and one subjective effort indicator, i.e. *perceived difficulty*. For *cognitive effort*,

eye gaze data was used: average fixation duration was measured during the translation and post-editing tasks. The rationale behind gauging cognitive effort with eye-tracking measures was explained in detail in Section 3.5. *Temporal effort* was captured through time measurements recorded by Inputlog in various application windows as well as different types of OR. *Perceived difficulty* was measured through Likert-scale judgments in a post-task questionnaire.

(2) *Attitude*: measured in the pre-task questionnaire with statement ratings on a visual-analogue scale.

(3) *Resource Range*: calculated as the number of OR types consulted by a participant per given text or task. The variable is discrete, i.e. consists of non-negative integers.

(4) *Accuracy*: measured as a binary ordinal variable (0 for incorrect and 1 for correct) on *Research Units* (terminological rich points) in the TT.

(5) *Researched %:* measured as ratio: number of consulted *research units* (looked-up deliberately or indirectly with the gaze cursor) divided by the total number of *research units* in a given text or task type.

(6) *Percentage of time* (in OR): time spent in OR in relation to total task time.

4.6. Hypotheses

Seven hypotheses were tested in the study described in the following sections.

4.6.1. Hypothesis 1

Both groups put more effort into information searching when translating than when post-editing.

Two indicators have been used as operational definitions of effort in this study:

(1) *Temporal effort* as time spent in applications (measured with Inputlog);

(2) *Cognitive effort* as average fixation duration (measured with the eye-tracker);

(3) (Additionally) *technical effort* as the number of typing events (measured with Inputlog).

This division reflects the operationalisation in TPR literature, as established in Chapter 3. The complex relationship between temporal and cognitive effort does not allow for neat divisions or separating them (the latter is a subtype of the former in line with Krings 2001), but for practical purposes these operational definitions are the most logical and straightforward.

It is hypothesised that the task type (translation and post-editing) will significantly influence the amount of effort (temporal, cognitive, and additionally technical) put into the process of IS by both of the above indicators, regardless of the group (and the effect of translation training).

4.6.2. Hypothesis 2

Translation trainees put more effort into information searching than non-trainees in translation and post-editing.

The same two effort indicators have been used as operational definitions of effort as above in Section 4.6.1 to explore Hypothesis 2. This hypothesis posits that regardless of the task type, the trainees will put more effort into information searching than non-trainees as a result of the experience and awareness which emerged through translation training. In other words, there will be a between-subjects effect of translation training on the amount of cognitive and temporal effort (as well as technical effort) regardless of the task.

4.6.3. Hypothesis 3

The range of consulted resources is narrower when post-editing than when translating from scratch for both groups.

The third hypothesis of this study assumes that regardless of group membership, the participants will search for information in OR depending on the task type. The independent variable is *task* (with two levels: translation and post-editing), the dependent variable the *resource range*. In other words, during post-editing the number of consulted OR categories is expected to be lower than when translating from scratch. The *resource range* variable is operationalised as the number of resource categories, calculated from the preprocessed Inputlog data.

4.6.4. Hypothesis 4

Translation trainees are more effective in information seeking than non-trainees.

In this hypothesis, the data about the product are analysed. It is expected that trainees will be more effective in their information searching than non-trainees, regardless of the task or text type. The independent variable is *group* and it has two levels: trainees and non-trainees. Terminological *accuracy* for the selected *research units* is used as a proxy for *effectiveness*. It is assumed that the trainees, due to the effect of their training will display higher accuracy scores regardless of text or task type, i.e. they will be more effective than EFL students.

4.6.5. Hypothesis 5

Translation trainees' and non-trainees' attitude towards MT correlates with the percentage of time in online resources when post-editing.

For this hypothesis, the questionnaire and process data have been used. It is posited that regardless of the group membership, attitude towards MT will correlate with the percentage of time spent in OR when post-editing, thus assuming task type contributes towards the amount of *temporal effort* exerted in the process. *Attitude* here is operationalized as a mean score from participants' rating on the visual analogue scale.[35] The *percentage of time* in ORs is based on temporal effort indicator and operationalised as time spent in OR in relation to total task time as per Inputlog data, expressed as a percentage.

4.6.6. Hypothesis 6

There is a correlation between perceived difficulty and temporal effort

Similarly to the previous hypothesis, Hypothesis 6 draws on the questionnaire and process data. It is an attempt to establish whether there is a relationship between *perceived difficulty* and *temporal effort*, i.e. the subjective and objective measures. To operationalise *perceived*

[35] Statement ratings were included in the pre-task questionnaire (see Appendix C: Pre-task questionnaire).

difficulty, ratings from post-task questionnaires regarding the completed task have been collected – it is a rating from each participant. *Temporal effort* indicators have been operationalised as in the previous hypotheses, i.e. as time spent in OR.

4.6.7. Hypothesis 7

The range of consulted resources correlates with the perceived difficulty

The last hypothesis explores the relationship between the questionnaire and process data as well – it seeks to establish whether there is a relationship between the subjective measure of effort and the range of consulted OR. The *resource range* is the number of consulted OR categories and its operationalisation is the same as stated in the previous hypothesis. Likewise, the *perceived difficulty* is operationalised in the same way. This hypothesis attempts to determine whether there is a statistically significant relationship between these two variables through a correlational analysis.

4.7. Participants

Two groups of participants were recruited for this study, translation trainees (henceforth T) and EFL students (henceforth G) from the Faculty of English and Faculty of Modern Languages and Literatures at Adam Mickiewicz University, Poznań, Poland. In total, 22 people participated in the experimental procedure, but 20 data sets were analysed (11 T and 9 G). Two datasets were rejected from the G group in their entirety due to technical issues during the recording sessions that rendered the data incomplete, e.g. Internet access malfunction. The participants participated voluntarily and received no remuneration except for extra course credits awarded for research participation. The project had no additional funding and was time-consuming for the participants, hence the relatively low number of participants.

The participants comprised four men and sixteen women, their age spanning from 21 to 32 (M = 23.24, SD = 2.54). Apart from the actual experimental task, both groups completed

a simplified version of a Language History Questionnaire[36] (henceforth LHQ) to establish their linguistic background and LexTALE (Lemhöfer and Broersma 2012) to objectively gauge their L2 proficiency. All 20 analysed participants had Polish as their L1 and English, on average, was their longest used foreign language (M = 16 years, SD = 3 years). Within the LHQ, they also self-evaluated their language proficiency on the scale from 1 (poor) to 7 (native-like) for each language they have ever learned, which was supposed to supplement the objective evaluation of language proficiency via the LexTALE test. Finally, all participants also completed a copy task from which their typing speed was measured as characters per minute.

The following dependent variables then were compared with respect to group membership and declared foreign language to test for potential differences: mean years of use, mean proficiency, LexTALE, characters per minute. The independent variable of language had three levels: English, German, and Other (which was an aggregate variable from all other languages reported by a given participant: French, Spanish, Russian, Japanese, Czech, Irish, Welsh, and Latin). The LexTALE score was only calculated for English, but it was included in the model nevertheless – only the means and tests for other levels have been disregarded. Similarly, the copy task results (characters per minute) were only calculated for one language, but for the sake of convenience they were also included in the model. Table 1 below shows the descriptive statistics:

[36] The version of the questionnaire was a version standardized by the Language and Communication Laboratory at the Faculty of English, Adam Mickiewicz University, Poznań, Poland.

Table 1. Descriptive statistics for participants info: mean years of use, mean proficiency, LexTALE, characters per minute by group (G – general and T – translation trainees)

Variable	Group	Language	*M*	*SD*	*N*
Mean years of use	G	English	16.44	3.81	9
		German	7.43	3.10	7
		Other	2.67	1.15	12
	T	English	15.27	2.15	11
		German	7.00	3.70	8
		Other	3.25	1.99	14
Mean proficiency	G	English	5.92	0.56	9
		German	2.32	1.27	7
		Other	2.79	1.54	12
	T	English	6.16	0.57	11
		German	2.84	1.00	8
		Other	2.89	1.44	14
LexTALE	G	English	81.39	9.55	9
	T	English	84.32	9.23	11
Characters per minute	G	Polish	453.00	145.94	9
	T	Polish	478.18	173.69	11

A MANOVA model was calculated to test for differences between groups for each of the above variables. In the multivariate tests, the following effects turned out non-significant: group ($F(4, 52) = 0.616$; $p = 0.949$; $\eta_p^2 = 0.16$), interaction between declared language and group ($F(8, 106) = 0.616$; $p = 0.724$; $\eta_p^2 = 0.16$). However, differences in terms of language as main effect (see Table 2) were significant ($F(8, 106) = 15.01$; $p < 0.001$; $\eta_p^2 = 0.531$).

Table 2. MANOVA results for participant info variables: main effect and interaction

Effect	Variable	*F*	*p*	η_p^2	Observed Power
Group	Mean years of use	0.24	0.627	0.00	0.077
	Mean proficiency	0.88	0.354	0.02	0.151
	LexTALE	0.32	0.575	0.01	0.086
	Characters per minute	0.00	0.958	0.00	0.050
Language	Mean years of use	135.71	**< 0.001**	0.83	1.000
	Mean proficiency	52.85	**< 0.001**	0.66	1.000
	LexTALE	0.07	0.932	0.00	0.060

	Characters per minute	0.02	0.979	0.00	0.053
Group*Language	Mean years of use	0.63	0.534	0.02	0.151
	Mean proficiency	0.15	0.858	0.01	0.072
	LexTALE	0.17	0.845	0.01	0.075
	Characters per minute	0.11	0.892	0.00	0.067

According to the post-hoc analysis for the language effect, the only significant differences were observed for mean years of use and mean proficiency which are both self-reported measures. For both of these variables post-hoc HSD Tukey test was calculated. For mean years of use, all groups differed significantly. However, for mean proficiency, the only statistical difference was between German and Other. These differences are illustrated with Fig. 15 and 16 below.

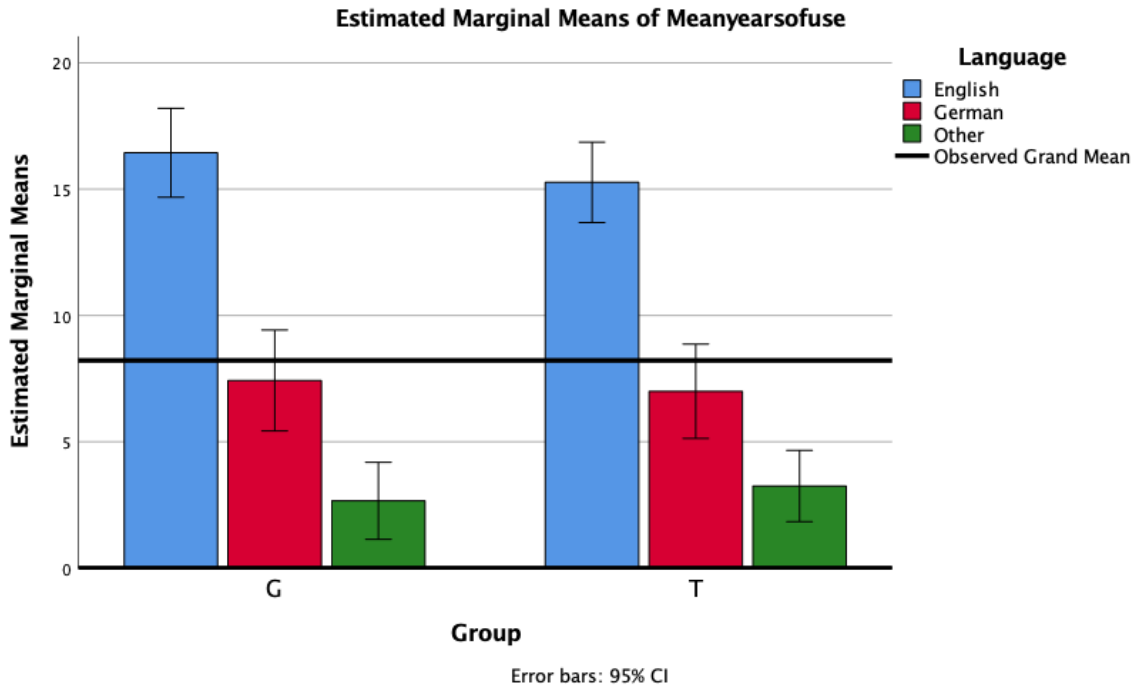


Fig. 16. Estimated marginal means for Mean years of use for language by group (G – general, T – translation trainees)

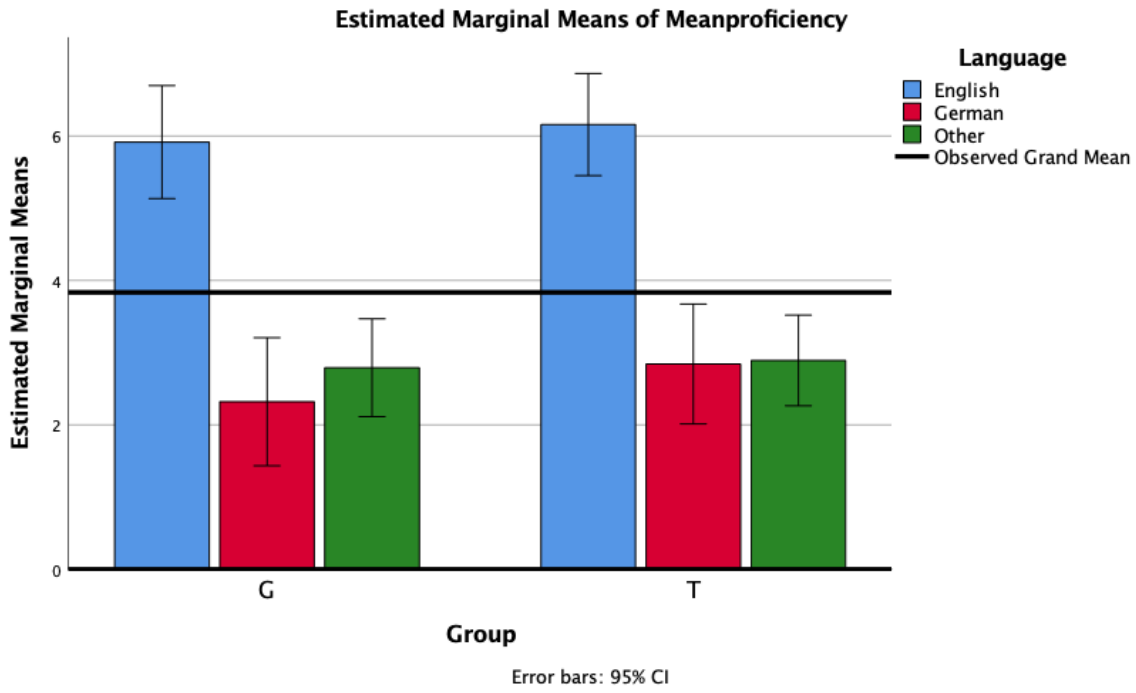


Fig. 17. Estimated marginal means for Mean proficiency for language by group (G – general, T – translation trainees)

These significant differences show that the participants in both groups only differ significantly in terms of subjective and self-reported measures (mean proficiency) between languages that were not involved in the experimental procedure and number of years of all languages used by participants. These differences should not influence the outcomes of the experiment.

Most importantly, the differences between groups for the LexTALE scores were not significant (for the G group $M = 81.39$, $SD = 9.55$; for the T group $M = 84.32$, $SD = 9.23$). Their mean scores place the groups within the range of C1–C2, i.e. upper and lower advanced/proficient user (Lemhöfer and Broersma 2012: 341). This means that in terms of English proficiency as objectively calculated through LexTALE, there were no significant differences between the groups. There was also no significant difference between the groups in terms of their typing speed in Polish (characters per minute). If there had been significant differences between groups for these two variables, they might influence the outcomes of the experiment in addition to the hypothesised effect. Since there are no such differences, the groups can be considered comparable, except for the translation training for the group of trainees. The results for the LexTALE and characters per minute from the copy task are summarised in Fig. 18 below:

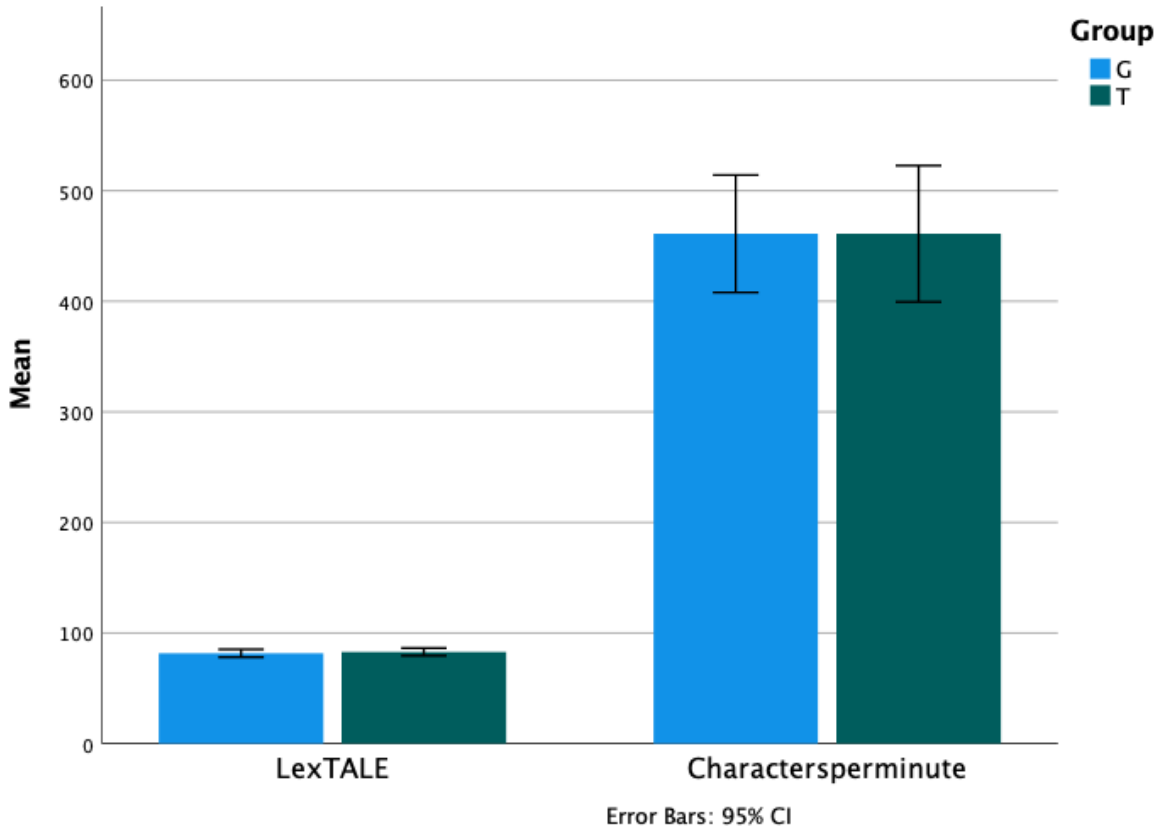


Fig. 18. Mean LexTALE scores by group (G – general, T – translation trainees)

The T group were students of the first semester of the Master's programme first year (i.e. 1MA). This group participated in the experiment from November 2017 to early January 2018 to ensure that they received minimal training with respect to MT and that their post-editing would be based on general guidelines and transferred skills from translation and editing strategies. The guidelines were abridged adapted from TAUS (2016) discussed in Section 1.5.1: the points 2 and 7 were removed. These were pertaining to retaining key terminology and preserving the formatting. The point about terminology was omitted in the instructions for participants because their attention was being purposefully averted from consciously thinking about terminology. It was one of the aims of the study to check whether or not the participants would trust MT suggestions enough not to check them in OR. Furthermore, the point pertaining to formatting was also disposed of because the research objectives of the experiment did not target any formatting skills of the participants and formatting was not analysed at any point.

The G group comprised students from both the BA and MA programme, but none of them had been formally instructed with respect to specialised translation except for an introductory course that all students at the Faculty of English receive during the second year of

the BA programme. The G group is a control group that would not have the effect of the translation training.

4.8. Materials

The materials in the study were two texts for translation as well as two for post-editing, the pre- and post-task questionnaires, the LexTALE test, and the Inputlog copy task. Each participant translated and post-edited the same texts. To minimise the task order effects (Mellinger and Hanson 2018), the text and task order was based on independent random assignment, so that each order was repeated no more than on two participants per group. The experiment set-up was as shown in Fig. 19. The MS Word ST along with MT output below for post-editing were on the left side of the screen. The right half of the screen was the Chrome browser. Fig. 19 also features the blue dot of the gaze cursor, fixating on the Polish *Wikipedia* article about the Apollo Programme mentioned in one of the STs, i.e. the WAT text (see Appendix A: Source texts and MT output). The participants were allowed to access any and all websites that they preferred during the translation and post-editing tasks.

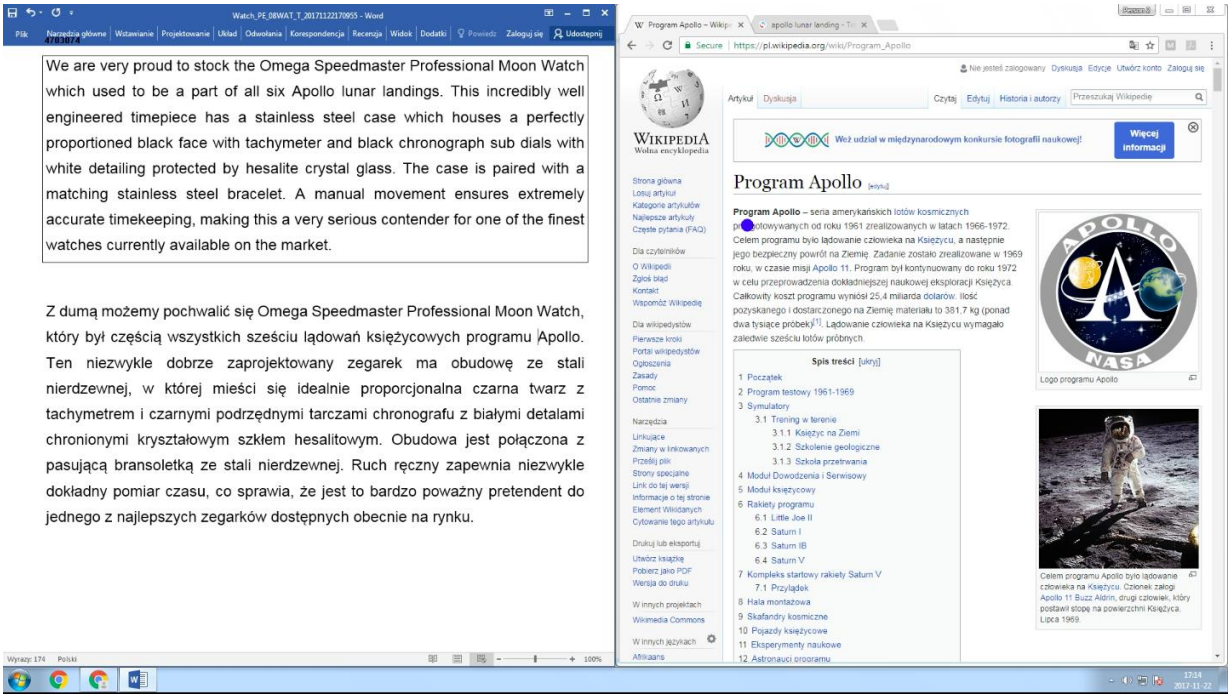


Fig. 19. Experiment set-up: Post-editing in MS Word (left) and Internet browser (right)

4.8.1. Texts

Each participant worked on four texts in total (two for translation, two for post-editing). With respect to their function (Reiss 1976), these were two informative and two operative texts. The rationale for selecting two text types for two tasks was to improve generalizability of results (Clark 1973). The informative texts were from the medical domain, while the operative texts were both technical and commercial. Each text contained a balanced number of terms that were expected to be researched by participants (around 10 terminological items). The texts were all selected based on their readability scores, provided in Table 3 below. These are all authentic texts with only slightly abridged word count so that it would not be too taxing to translate and post-edit them in two sessions.

Table 3. Source text readability

Text	FRES	WORD COUNT	CHARACTER COUNT	PARAGRAPH COUNT	SENTENCE COUNT
APP	31.7	94	486	1	4
WAT	36.4	90	486	1	4
MMR	31.3	97	540	1	4
HEA	37.2	92	485	1	4
Informative-medical	31.5	191	1025	2	8
Operative-technical	36.8	182	971	2	8

APP and MMR are the two informative-medical texts about appendicitis and MMR vaccine respectively. APP was adapted from the Scottish *NHS Inform* website[37] and MMR from the Welsh *NHS Direct* website.[38] The two operative-technical texts, WAT and HEA, are about a wristwatch and headphones, respectively. The HEA text was adapted from an online store product description,[39] whereas the WAT text came from an *Goldsmiths.co.uk*, an online store product description and was abridged for the study.[40]

The objective text-based measures were also supplemented with participant-based difficulty judgements for each text (5-point Likert scale, 1 – Very easy, 2 – Easy, 3 – Medium,

[37] The text remains unchanged on the NHS website as of the time of access (https://www.nhsinform.scot/illnesses-and-conditions/stomach-liver-and-gastrointestinal-tract/appendicitis, date of access: 23 Nov 2019).
[38] The text remains unchanged on the NHS website as of the time of access (https://www.nhsdirect.wales.nhs.uk/livewell/vaccinations/MMR/, date of access: 23 Nov 2019).
[39] The original description has been changed since the time of the original access (https://www.amazon.co.uk/Beyerdynamic-Pro-Headphones-Black-Limited/dp/B01ERLN180, date of access: 6 Nov 2017).
[40] The original description has been completely changed since the time of the original access (https://www.goldsmiths.co.uk/Omega-Speedmaster-Moonwatch-Professional-Chronograph-42mm-Mens-Watch/p/17331157/, date of access 6 Nov 2017).

4 – Difficult, 5 – Very difficult) provided in the post-task questionnaire. The descriptive statistics for each group as well as for both groups together are provided in Table 4 below. There was a statistically significant difference in *perceived difficulty* depending on text, $\chi^2(3) = 10.510$, $p = .015$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.0125$. Median (IQR) perceived difficulty scores for WAT, HEA, MMR, APP were 3 (2.25 to 4), 3 (3 to 4), 3 (2 to 3), and 3 (2 to 3), respectively. There were no significant differences between WAT and HEA ($Z = -1.604$, $p = 0.109$), WAT and MMR ($Z = -1.706$, $p = 0.088$), WAT and APP ($Z = -1.032$, $p = 0.302$), APP and HEA ($Z = -2.209$, $p = 0.027$), or between APP and MMR ($Z = -0.513$, $p = 0.608$) despite an overall reduction in *perceived difficulty* in operative-technical texts vs. informative-medical texts. This lack of significance in terms of difference suggests that to a certain degree, the applied objective measures were reflected in what the participants subjectively thought about the texts.

Table 4. Source texts: Participant subjective judgements of difficulty by group and treated as one

Group		N		Mean	Median	Std. Deviation	Minimum	Maximum	Percentiles		
		Valid	Missing						25	50	75
G	WAT_DIFF	9	0	3.22	3.00	.97	2.00	5.00	2.50	3.00	4.00
	HEA_DIFF	9	0	3.33	3.00	.87	2.00	5.00	3.00	3.00	4.00
	MMR_DIFF	9	0	2.78	3.00	.44	2.00	3.00	2.50	3.00	3.00
	APP_DIFF	9	0	2.78	3.00	.97	1.00	4.00	2.00	3.00	3.50
	TECH_DIFF	9	0	3.28	3.00	.87	2.00	5.00	2.75	3.00	3.75
	MED_DIFF	9	0	2.78	3.00	.51	2.00	3.50	2.25	3.00	3.00
T	WAT_DIFF	11	0	3.00	3.00	.77	2.00	4.00	2.00	3.00	4.00
	HEA_DIFF	11	0	3.45	3.00	.52	3.00	4.00	3.00	3.00	4.00
	MMR_DIFF	11	0	2.73	3.00	.65	2.00	4.00	2.00	3.00	3.00
	APP_DIFF	11	0	2.91	3.00	.70	2.00	4.00	2.00	3.00	3.00
	TECH_DIFF	11	0	3.23	3.50	.47	2.50	4.00	3.00	3.50	3.50
	MED_DIFF	11	0	2.82	3.00	.60	2.00	3.50	2.00	3.00	3.50
All	WAT_DIFF	20	0	3.10	3.00	.85	2.00	5.00	2.25	3.00	4.00
	HEA_DIFF	20	0	3.40	3.00	.68	2.00	5.00	3.00	3.00	4.00
	MMR_DIFF	20	0	2.75	3.00	.55	2.00	4.00	2.00	3.00	3.00
	APP_DIFF	20	0	2.85	3.00	.81	1.00	4.00	2.00	3.00	3.00
	TECH_DIFF	20	0	3.25	3.25	.66	2.00	5.00	3.00	3.25	3.50
	MED_DIFF	20	0	2.80	3.00	.55	2.00	3.50	2.13	3.00	3.00

However, there was a statistically significant difference in perceived difficulty between MMR and HEA (Z = -2.829, p = 0.005), which were deemed easiest and most difficult as per means in Table 4. A separate Wilcoxon signed-rank test was conducted for informative-medical (averaged APP and MMR) vs. operative-technical (averaged HEA and WAT) and the comparison reached significance (Z = -2.251, p = 0.024). Median (IQR) perceived difficulty scores for informative-medical and operative-technical were 3 (2.125 to 3) and 3.25 (3 to 3.5), respectively. Thus, the operative-technical set was deemed slightly more difficult than the other set.

4.8.2. Procedure

The procedure involved the following steps:

(1) General written instructions for the whole experiment (Appendix F: General instructions for participants);
(2) Informed consent form (Appendix G: Informed consent form);
(3) Copy Task (Van Waes et al. 2019) (see Section 4.9.2);
(4) Pre-task questionnaire (Appendix C: Pre-task questionnaire);
(5) Translation/post-editing brief (Appendix H: Translation and post-editing briefs) – depending on the task order combination;
(6) Software launch: Morae Recorder;[41]
(7) Software launch: EyeLink software (Pop-up Calibration) – eye-tracker calibration and validation, start of the eye-tracking recording;
(8) Software launch: Inputlog;
(9) Translation or post-editing task;
(10) Repeat of steps 5–9 with another text from the pair;
(11) Post-task questionnaire (Appendix D: Post-task questionnaire (WAT+APP) and Appendix E: Post-task questionnaire (HEA+MMR));
(12) LexTALE test (Leijten and Van Waes 2013) via the online form;[42]
(13) LHQ test (Appendix B: Language History Questionnaire);

[41] The reason for using Morae Recorder is provided in Section 4.9.4.
[42] http://www.lextale.com/takethetest.html (date of access: 30 January 2020).

There were two sessions to avoid fatigue – one session took as long as 90 minutes when all the questionnaires, forms, calibrations, and other preparations were included. The above steps (without steps 1–4 and 12–13) were repeated during the second session which was scheduled with at least a day's break to ensure participants had a fresh look and got some rest between the sessions. As indicated in the procedure, the process data from the translation and post-editing tasks were recorded by three independent tools: the keylogging programme (Inputlog), the eye-tracker (EyeLink 1000 Plus) and a screen capture programme (Morae). All the tools are described in detail in the following section.

4.9. Tools

The tools used in the study included Inputlog, the Copy Task, EyeLink eye-tracker and capture software, Morae Recorder for screen recording, and the *1ka.si* survey website. Each of these tools will be described in more detail in the next sections.

4.9.1. Inputlog

As explained by the designers themselves, Inputlog is an MS Windows-based "keystroke logging program enabling you to observe writing process dynamics and collect fine grained data. The program also provides a wide range of analyses opening new perspectives to a better understanding of the (cognitive) complexity of writing" (Leijten and Van Waes 2019). The first version was developed in 2003, the version used for this thesis is Inputlog 7.1.0.47.

The program's primary feature is to record, i.e. log keyboard, mouse, and speech events in Windows applications and assign them with a time stamp in milliseconds. All the information pertaining to the event, timestamp, character position, document length, and copy/paste/move actions (the last three only when used with MS Word) "are continuously and unobtrusively stored for later processing. This continuous data storage does not delay or interfere with the normal use of the computer" (Leijten et al. 2019: 6). It is particularly important to note that Inputlog is unobtrusive for recording post-editing and translation process tasks, as it functions in the background and captures activity from any application, including Internet browsers. The advantage of Inputlog over Translog, which was originally designed for the study of the

translation process by Jakobsen (1999), is that it does not require the translator to work in a different environment, but a familiar word processor.

Inputlog also allows pre- and post-processing of recordings and has an in-built analysis as well as replay tool (Leijten et al. 2019: 6). The pre-process module enables process recordings with respect to certain keyboard/mouse/speech events, time stamps, or certain application windows. These are described in detail in the analysis Section 4.10.1.2. Copy task data is a very useful Inputlog feature that allows for a more comprehensive picture of process recordings. It is described in detail in the next section.

4.9.2. Copy task in Inputlog

Under the assumption that typing skills could have influence over text production (Leijten et al. 2019: 27) and that translation and post-editing are both writing activities that usually involve typing on a keyboard, a copy task was also used in the experimental procedure in this thesis. It was intended to both acclimatise the participant to the keyboard in the lab and check their typing skills in a set of controlled typing tasks.

The copy task developed within Inputlog "is designed to create a set of measures that allow a fine grained analysis of low level typing and motor skills" (Leijten et al. 2019: 27). The copy task used for this thesis is the Polish language version, the TL for this study. The Polish version is available online[43] in a java-based web interface. The copy task was translated into Polish by the author of this thesis in late 2017 in cooperation with the Inputlog research team (see Van Waes et al. 2019).

The copy task involves seven typing tasks. The first task is the tapping task which consists in pressing two keys with alternating hands, i.e. "d" and "k" for 15 seconds. Second task measures copying high frequency words in a sentence for 30 seconds (as many times as possible). For Polish, this sentence is: "kot siedzi pod domem i wolno macha ogonem" ("the cat is sitting under the house and slowly swishing its tail"). After that, three sets of word combinations are presented for copying (numeral + adjective + noun). The words were carefully selected based on word length, lemma, and bigram frequency in SUBTLEX-PL (Mandera et al. 2015) and balanced for each hand combination position on the keyboard, adjacency, and no repetitiveness for keys. These sets for Polish were: "trzy relaksacyjne techniki" (three relaxation

[43] http://inputlog.ua.ac.be/Website/copytask/tasks.html (date of access: 27 Jan. 2021).

techniques), and “dwie niespotykane poprawki” (two unanticipated corrections), “oba nieszkodliwe narkotyki” (both harmless drugs). Next task involves a word combination of indefinite determiner/countable + adjective + noun. The combination was again carefully selected based on word length and balanced for minimal or no key repetitiveness. Bigrams in this combination were selected from the low frequent <50% percentile in SUBTLEX-PL (Mandera et al. 2015). The set for Polish was: “trzy tryumfalne gzymsy” (three triumphant cornices). These are all grammatical albeit nonsensical sentences, but the combinations with the adjectives needed to be unique, due to the control constraints enumerated above. Bigram frequencies were extracted using a Python script (Behnke 2017). Then, the consonant groups task is the same for all language versions, as is the tapping task. The choice for word combination and sentence was aided by the coded Excel file provided by the Inputlog research team.

4.9.3. EyeLink 1000 Plus

The eye-tracking device used for capturing gaze data was EyeLink 1000 Plus developed by SR Research. The software used for analysis was EyeLink Data Viewer[44] and the details of the analysis are provided in Section 4.10.1. The set-up described in Fig. 20 below is the one used for this study. The participant completed the tasks in the same experimental se-up working on the same computer, i.e. the Display PC (with Windows 7) with a 24-inch monitor (53.3 × 30 cm). The Host PC with the eye-tracking software for data collection (not displayed in Fig. 20) was placed behind the participant. It was on the Host PC that any adjustments and calibrations for the eye-tracker were done by the researcher.

[44] https://www.sr-research.com/data-viewer/ (date of access 17 Feb. 2021).

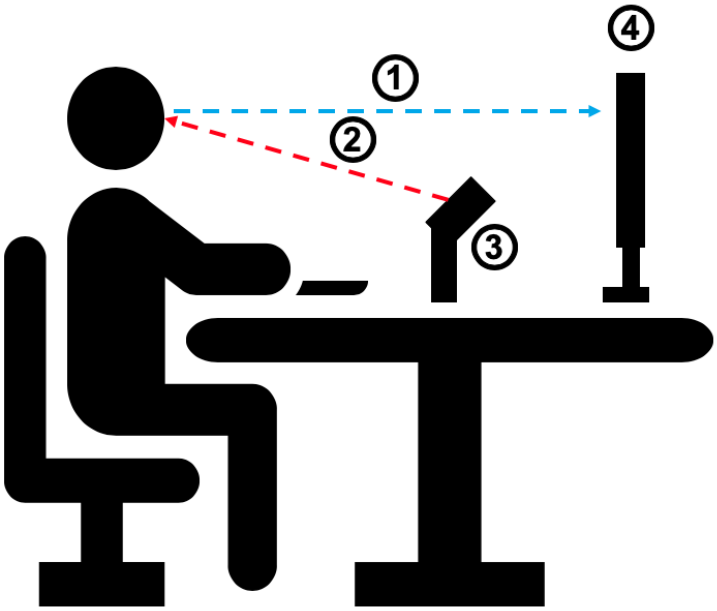


Fig. 20. Equipment set up on the participant's desk

The eye-tracker, (3) in Fig. 20, was set up in head free-to-move remote mode and placed as a desktop mount (SR Research 2018), between the participant and the computer screen, without obscuring any part of the screen. The screen, (4) in Fig. 20, was adjusted to the participant's height, i.e. line (1) in Fig. 20. The eye-tracker's long range illuminator emitted infrared light depicted as (2) in Fig. 20. (2) also depicts the distance from the eye-tracker's fibre optic camera to the participant's eye which is between 40 to 70 cm. The room with the eye-tracking equipment had fixed overhead lighting and natural light blocked with roller shades to create the same conditions for all participants. The door to the other room with the researcher was also closed for the duration of the task so that the light from outside the recording room would not interfere with gaze capture.

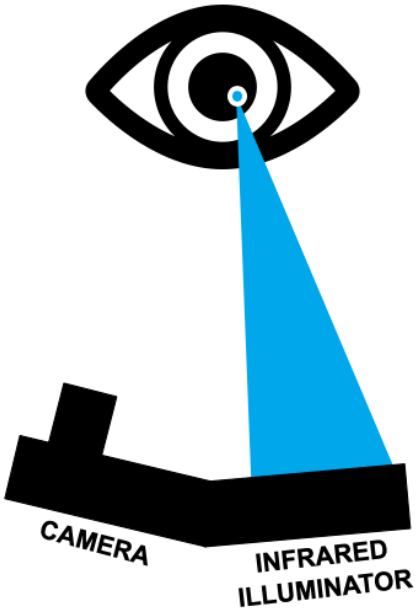


Fig. 21. Pupil and corneal reflection adapted from SR Research website (2018)

Fig. 21 illustrates the bird's eye view of how the eye-tracker works. In video-based eye-tracking, the camera – visible on the left in Fig. 21 – captures the eye movements (SR Research 2018). The software determines two locations on the captured images: the centre of the pupil and corneal reflection (the blue-white dot in Fig. 21) produced by the infrared illuminator on the eye. While the corneal reflection usually remains fixed in relation to the camera sensor, the centre of the pupil moves with each eye movement as the participant looks at different parts of the screen. The camera captured the movement of right eye with the frequency of 500 Hz via the 16 mm lens (optimal for the remote set-up).

The study required obtaining four eye-tracking recordings per participant (one per each text), two per session. Total task duration varied among participants and text types, so it was safer to calibrate (i.e. train the eye-tracker for each participant's eye in relation to 13 points on the computer screen) the tracker for each text anew. The calibration screen set-up is presented in Fig. 22, 13 points were used to ensure maximum accuracy. Each calibration was always followed by validation to ensure minimal drift.

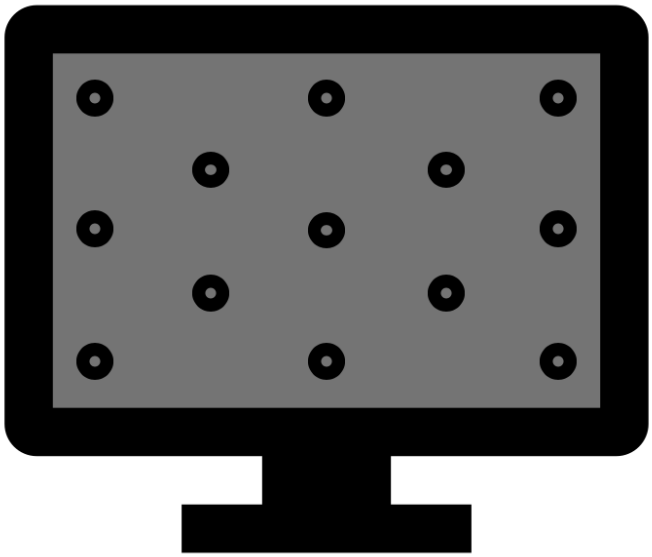


Fig. 22. Calibration/validation points

The time needed for the 13-point calibration was also subject to individual variation, ranging from a few minutes up to 15 minutes or more, but never exceeding three takes. Participants had either normal or corrected to normal vision (glasses or contact lenses) and reported no sight disorders which would compromise the recording reliability.

4.9.4. Morae Recorder

Morae Recorder by TechSmith Corporation was used to triangulate the process recording data. Release 3.3.0 was used. The programme operated simultaneously with Inputlog and eye-tracking capture, but neither video recordings nor keylogging data were used from Morae in any of the analyses. Morae served as a screen capture backup in case any of the data from Inputlog or EyeLink were damaged.

4.9.5. Survey website 1ka.si

The website *1ka.si*[45] was used to gather supplementary questionnaire data. It is an open source online application that allows creating and sharing online surveys. It was developed and is

[45] https://www.1ka.si/d/en (date of access: 17 Feb. 2021).

maintained by the Centre for Social Informatics, at the Faculty of Social Sciences, University of Ljubljana.

4.10. Data analysis

After collecting the data, the analysis proceeded in the following order: process data analysis, product data analysis, questionnaire data analysis, and statistical analysis.

4.10.1. Process data analysis

The following sections provide the information about process data analysis for eye-tracking and keylogging measures.

4.10.1.1. Eye-tracking data analysis

First, the eye-tracking data was run through the Data Viewer. The programme allows to view, filter, and process gaze data. Each recording (e.g. 01APPT_T) was first viewed to filter out any inconsistencies or errors in the video recording. Then, it was trimmed to establish task start time and end time. The reason for trimming was because the recording had been started before the ST window was opened to prevent the participant from focusing on reading while the set up was still being prepared. Finally, two rectangular Interest Areas have been drawn: Text (MS Word application window on the left hand side of the screen) and Browser (Google Chrome browser on the right). They were drawn for each recording separately to ensure filtering out the stray fixations in the middle of the screen. Then, Interest Area Reports have been generated with the following variables relevant to calculate average fixation durations per IA, as provided in Table 5 below:

Table 5. Variables included in the Interest Area Report from Data Viewer

Field	Contents
RECORDING_SESSION_LABEL	Data file label
IA_LABEL	Current interest area label (i.e. Text or Browser)
IA_DWELL_TIME	Dwell time (i.e., summation of the duration across all fixations) in a given interest area, in milliseconds
IA FIXATION COUNT	Total number of fixations in a given interest area, integer number

The reports have been merged into a single MS Excel spreadsheet and the cognitive effort variables have been calculated based on the two main IAs: Text and Browser as the average fixation duration from the IA_DWELL_TIME divided by the IA_FIXATION_COUNT. Also, an aggregate variable, Total, has been calculated from adding IA_DWELL_TIME from Text and Browser as well as IA_FIXATION_COUNT for the same variables. Thus, average fixation duration has also been calculated.

To filter out compromised recordings, quality filter has been applied. The eye-tracking data quality filter included average fixation duration threshold and GTS (gaze time on screen), as described in Section 3.5.4 (Hvelplund 2014: 216ff). The acceptable threshold for the mean fixation duration was 200 ms and GTS was 50%. Both of these minimums had to be achieved and anything below them resulted in the eye-tracking data set being excluded from the statistical analysis.

To prepare the data for statistical analyses in hypotheses 1 and 2, group and task averages have been calculated. For hypothesis 1, the groups were counted as one, so average fixation durations per participant per variable have been calculated for both groups averaged. For hypothesis 2, the tasks were counted as one, so average fixation durations per participant per variable have been calculated for both tasks averaged.

4.10.1.2. Inputlog data analysis

For process data analysis, the Preprocess tool was used to trim the recording after participants pressed the lowercase “f” key at the very end of their post-editing or translation task. The character “f” stands for “finished” and it was meant to be pressed only when all changes and research activities were finished, so that there would not be any redundant pause time in the recording between finishing the task by the participants and pressing save in Inputlog by the researcher. Some participants continued to add changes to their translations after hitting “f” at

the end. In that case, the recording was trimmed either right after their cursor stopped moving completely or right before clicking on the "save" button in MS Word by the researcher.

The analyse module offers three representations of the process (general and linear logging file and the s-notation of the text) and produces four aggregated levels of analysis from the recordings – summary, pause, revision, and source analysis, as well as a process graph (Leijten et al. 2019: 6). This thesis focused on the data extracted from the summary and source analysis components, which were then merged via the post-process module into a single log file for statistical analysis.

The analysis in the Preprocess tab also involved bundling single website visitations into OR categories. The preprocessing involved aggregating events based on window changes (MS Word vs. Google Chrome websites). All other windows were excluded from the analysis. Then, to input the data into statistical analysis software, the Postprocess function was used in order to merge single data points into a single spreadsheet.[46] This was used to calculate temporal effort for each resource category from very specific ones (e.g. *WS_TT_CORPORA_PL*[47] – Polish corpora) to the most general aggregate categories (e.g. *TT_Combined_OR* – time spent in all OR). The categories are all provided in Table 6 below. This thesis adapted the category of reference works after both Gough and Hvelplund, i.e. as organisational, academic, and knowledge-based websites that are not primarily term bases or dictionaries.

Table 6. Inputlog Preprocess tab online resource categories

No.	Variable (Category)	Description
1	WS_T_totalTime	Time spent on the entire task = total task time (in seconds)
2	WS_TT_INPUTLOG_MS_Word	Time spent in MS Word (in seconds)
3	TT_Combined_OR	Time spent in all OR (sum of rows 4, 5, 9, 9) (in seconds)
4	TT_All_EN_PL	Time spent in all OR (sum of rows 6 and 7) (in seconds)
5	WS_TT_Google	Time spent in Google, including images, translate, and search (sum of rows 10–12) (in seconds)
6	TT_EN_SUM	Time spent on English language websites (e.g. monolingual dictionaries or corpora) (sum of rows 18–21) (in seconds)
7	TT_PL_SUM	Time spent on Polish language websites (e.g. monolingual dictionaries or corpora) (sum of rows 22–25) (in seconds)
8	WS_TT_CONCORDANCER	Time spent in concordancers (e.g. *Linguee* or *Glosbe*) (in seconds)
9	WS_TT_BILINGUAL_DICTIONARIES	Time spent in Polish and English bilingual dictionaries (in seconds)
10	WS_TT_Google_Search	Time spent in Google search (in seconds)
11	WS_TT_Google_images	Time spent in Google images (in seconds)

[46] More details about Pre- and Postprocessing in Inputlog are available in the manual and official website: https://www.inputlog.net/overview/ (date of access: 14 Aug 2020).
[47] The labels for most of the categories are left as they were generated by Inputlog.

12	WS_TT_Google_Translate	Time spent in Google Translate (in seconds)
13	WS_TT_REFERENCE	Time spent on specialised websites in Polish and English (sum of rows 18 and 22) (in seconds)
14	WS_TT_WIKIPEDIA	Time spent on Polish and English *Wikipedia* (sum of rows 19 and 23) (in seconds)
15	WS_TT_MONOLINGUAL	Time spent in Polish and English monolingual dictionaries (sum of rows 20 and 24) (in seconds)
16	WS_TT_CORPORA	Time spent in Polish and English corpora (sum of row 21 and 25) (in seconds)
17	WS_TT_LANGUAGE_REF[48]	Time spent on language reference websites (e.g. PWN language advice centre) (in seconds)
18	WS_TT_REFERENCE_EN	Time spent on English specialised websites (in seconds)
19	WS_TT_WIKIPEDIA_EN	Time spent on English *Wikipedia* (in seconds)
20	WS_TT_MONOLINGUAL_EN	Time spent in monolingual English dictionaries (in seconds)
21	WS_TT_CORPORA_EN	Time spent in English corpora (in seconds)
22	WS_TT_REFERENCE_PL	Time spent on Polish specialised websites (in seconds)
23	WS_TT_WIKIPEDIA_PL	Time spent on Polish *Wikipedia* (in seconds)
24	WS_TT_MONOLINGUAL_PL	Time spent in Polish monolingual dictionaries (in seconds)
25	WS_TT_CORPORA_PL	Time spent in Polish corpora (in seconds)

Thanks to these categories, the variable *resource range* was also calculated. The websites that Inputlog collected in the log file were aggregated into categories and the variable *resource range* contains the number of resource types that a given participant consulted. This particular variable counted whether participants accessed a given resource category for any number of time (could be a few or few hundred seconds) and if they did – counted the category as 1. If there was no time spent in a given category, 0 was assigned to that category. The following 14 categories were used:

(1) WS_TT_CONCORDANCER
(2) WS_TT_BILINGUAL_DICTIONARIES
(3) WS_TT_Google_Search
(4) WS_TT_Google_images
(5) WS_TT_Google_Translate
(6) WS_TT_REFERENCE_EN
(7) WS_TT_WIKIPEDIA_EN
(8) WS_TT_MONOLINGUAL_EN
(9) WS_TT_CORPORA_EN
(10) WS_TT_REFERENCE_PL
(11) WS_TT_WIKIPEDIA_PL

[48] There is only the Polish category because none of the participants visited any such equivalent websites in English.

(12) WS_TT_MONOLINGUAL_PL

(13) WS_TT_CORPORA_PL

(14) WS_TT_LANGUAGE_REF

The descriptive statistics for these 14 categories with respect to the division into groups and tasks are provided in Table 17 in 4.11.3.

Inputlog also recorded technical effort with the variable *WS_totalKeypresses*, but although this type of effort is beyond the scope of this study and the main hypotheses, the variable was included in the analyses as it does contribute to the overall effort in the task. This variable was not taken into account during hypothesis formulation partly because of potential technical inefficiency in post-editing (O'Brien 2017: 327). This inefficiency is connected with lack of technical editing skills, post-editor's limitations, and/or the cognitive ease of deleting and retyping whole words or phrases rather than changing a few characters. Some translators might even resort to navigating the text with keyboard arrows to facilitate reading. Thus, the results regarding this type of effort will be considered in conjunction with temporal and cognitive effort to which it contributes.

4.10.2. Product data analysis

The process data analysis was supplemented with product analysis. As a proxy for IS effectiveness, the *accuracy* of TTs was determined by the author by means of a binary variable (1 = correct, which included minor typos that did not change the meaning, 0 = incorrect or unacceptable typo, e.g. **loparoskopię* instead of *laparoskopię* (*laparoscopy*). The texts, as described in Section 4.8.1, were selected also partially with this in mind – not only not too difficult for the participants but also for the researcher to be able to confidently and reliably assess the translation solutions of the participants. What is more, the texts were selected and prepared (slightly abridged for readability, for more details see Section 4.8.1) with a set number of terminological rich points in mind. Henceforth, these rich points will be referred to as *research units* (henceforth RUs). At the point of study design, those RUs were anticipated to be researched by each participant, albeit their number was expected to increase in analysis. More terms/phrases needed to be treated as RUs based on actual participants' behaviour. In the end, thanks to the screen recording data and terminological relevance, there were 8, 9, 9, 8 RUs for APP, MMR, HEA, WAT, respectively. The RUs were only used for the product analysis,

the process analyses of effort were not conducted with reference to these RUs – these were done for whole tasks.

Table 7 presents the percentage of the entire group that had their translation marked as correct for this particular RU (*Acc General* and *Acc Trainees*) as well as the percentage of the group that researched each unit (*Researched % in General* and *Researched % in Trainees*).

Table 7. Research units in ST with their *Accuracy* and *Researched %* by group (G – general, T – translation trainees)

Text	Research unit	Acc General	Researched % in General	Acc Trainees	Researched % in Trainees
APP	Abdomen	88%	100%	100%	82%
	appendectomy	75%	100%	91%	100%
	appendicitis	100%	86%	100%	100%
	burst (appendix)	100%	86%	91%	82%
	keyhole surgery (laparoscopy)	100%	100%	82%	100%
	open surgery	88%	100%	100%	91%
	success rate	63%	57%	100%	91%
	surgical instruments	63%	71%	100%	55%
MMR	combined vaccine	67%	75%	91%	91%
	Deafness	100%	25%	100%	36%
	develop (serious) conditions	78%	50%	100%	55%
	Measles	100%	75%	100%	82%
	meningitis	89%	50%	91%	91%
	MMR	100%	38%	100%	55%
	Mumps	100%	75%	100%	91%
	rubella (German measles)	100%	100%	100%	91%
	swelling of the brain (encephalitis)	89%	75%	100%	91%
HEA	(excellent) sound reproduction	89%	38%	73%	73%
	closed backed headphones	89%	100%	100%	100%
	closed dynamic headphone	89%	75%	82%	73%
	critical music and sound monitoring	67%	88%	27%	100%
	diaphragms	89%	100%	91%	100%
	drivers	22%	88%	55%	100%
	ear pads	67%	88%	91%	91%
	professional studio applications	100%	75%	100%	91%
	single-sided cable	100%	100%	73%	82%
WAT	Apollo lunar landings	33%	38%	45%	64%
	case	100%	0%	100%	36%
	chronograph sub dials	22%	75%	36%	73%
	face	67%	88%	100%	73%
	hesalite crystal glass	0%	25%	18%	64%
	manual movement	33%	63%	55%	100%
	tachymeter	89%	63%	91%	82%
	timepiece	100%	38%	100%	27%

Researching (as referred here by the *Researched %*) here means explicitly typing the phrase/term directly related to the research unit in question into a dictionary or search engine. It also counted as researching when in the eye-tracking screen recordings the gaze cursor moved over a certain equivalent and information pertaining to it when reading for another RU. Gough (2017) referred to this type of search as indirect. For instance, when translating APP, participants numbered as T2, T12, T15, and T16 read an article on *Wikipedia* about *laparoskopia* which included the phrase *jama brzuszna*/*powłoki brzuszne*, which could be used as equivalents for *abdomen*). G15 also accessed an *ang.pl* forum post[49] about translating *keyhole surgery* where they read about *jama brzuszna* in this context (mouse hover visible over the relevant phrase in Fig. 23).

[49] https://www.ang.pl/forum/pomoc-jezykowa-tlumaczenia/42681, date of access: 13 Jul 2020, now archived.

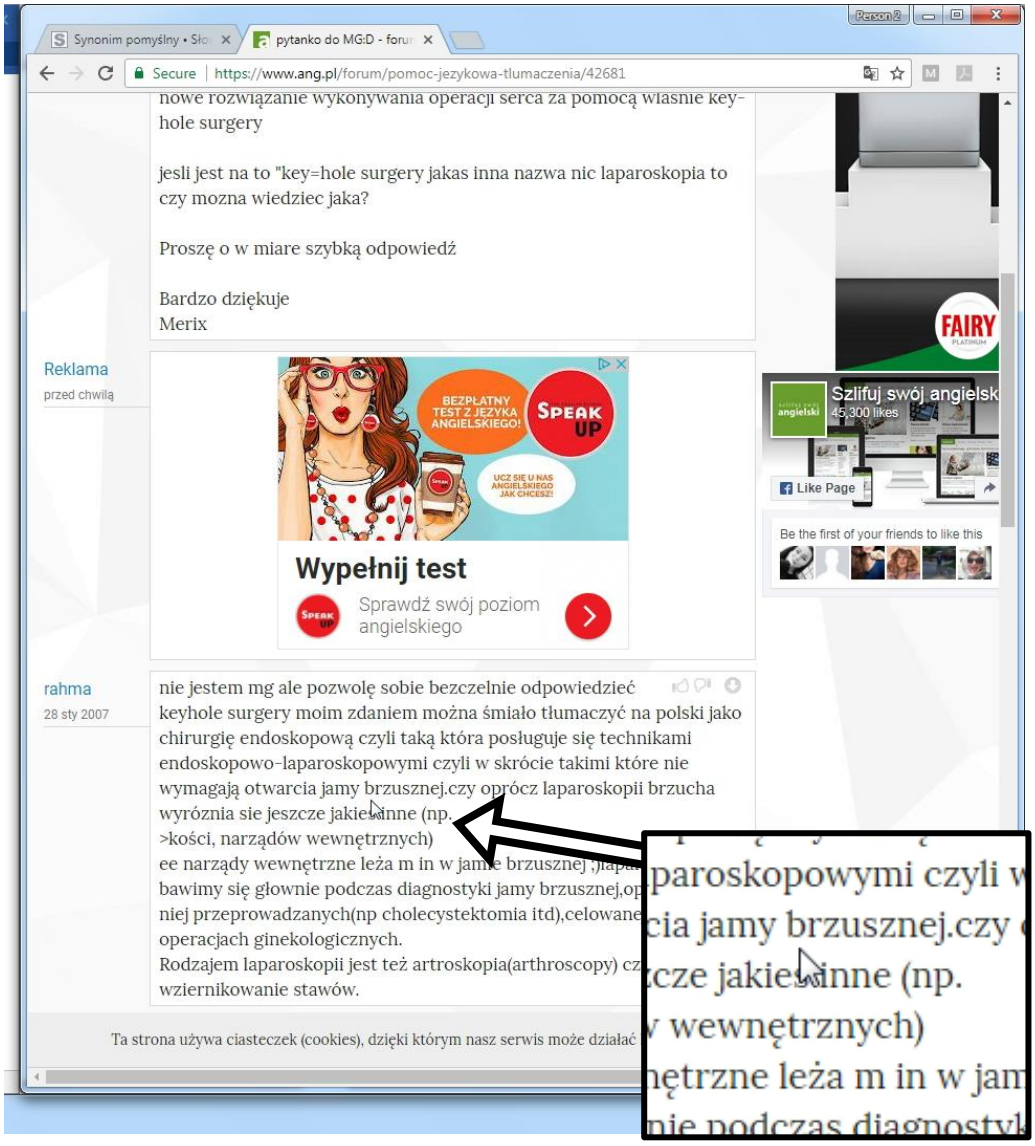


Fig. 23. Indirect research (*abdomen*) via text directly pertaining to another term (*keyhole surgery*)

The same situation happened with participant T2, who accessed the Google *Wikipedia* blurb for *laparoskopia* and read the words *powłoki brzuszne* pertaining to *abdomen* in the ST. Whether the participant knew the Polish equivalent beforehand and typed it from memory or the fragment shown in Fig. 24 jogged their memory does not mean that this indirect research did not occur. Here, only this can be objectively measured via keylogging data and screen recordings showing such cases of indirect research and more detailed profiling of the researching is not necessary, therefore participants were not asked to further elaborate about their process decisions. It may be a limitation, as Massey and Ehrensberger-Dow (2011b: 9f) found, that due to either cognitive or ergonomic issues translators of all levels of expertise often fall prey to the phenomenon called “looking but not seeing”. Usability of online interfaces in dictionaries and other resources for beginners (or non-translators as well) might facilitate or

hinder the process of consulting as well. However, in the context of this study, the question of whether the person only looked but failed to see something relevant to the final solution or was hindered/facilitated by the UI is beyond the scope of the present research aims.

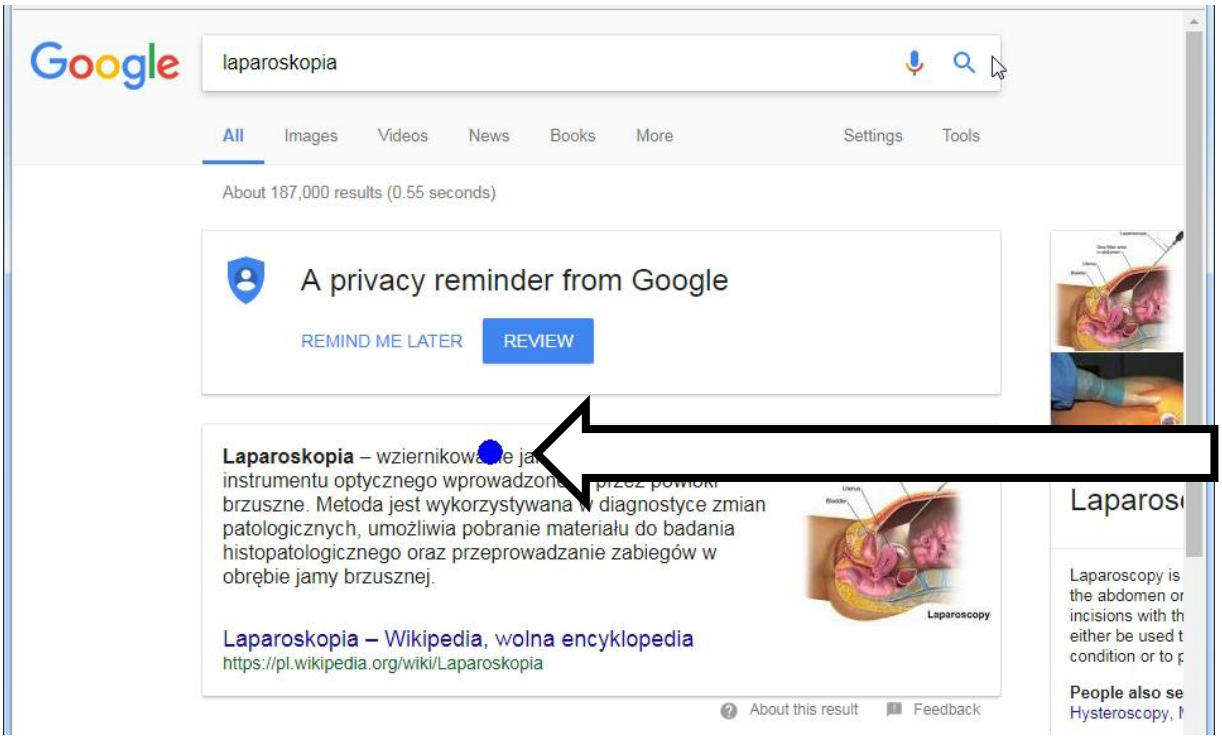


Fig. 24. Indirect research (*abdomen*) via text directly pertaining to another term (*keyhole surgery*)

Another example of indirect research happened with G13 whilst translating the HEA text. Fig. 25 shows the gaze cursor hovering over *jednostronny przewód* (=*one-sided cable*) which then was typed into the TT in MS Words, although the RU directly addressed in the search was *drivers*.

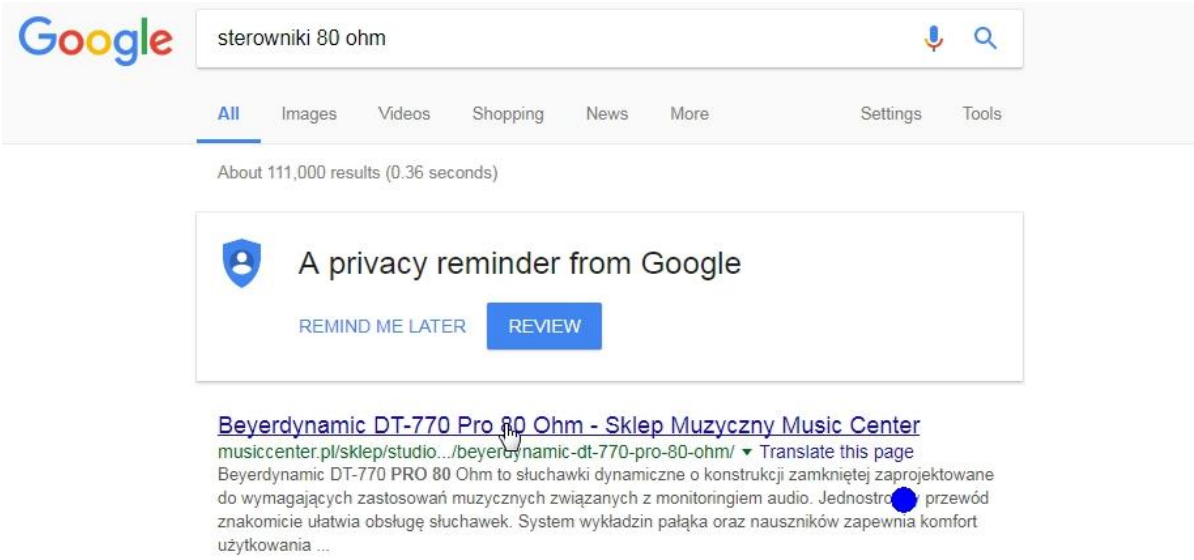


Fig. 25. Indirect research (*one-sided cable*) via search directly pertaining to another term (*drivers*)

One more example of indirect research was in the text MMR for two participants, namely T17 and G22. In Fig. 26, participant T17 through deciphering the meaning of *German measles* and *rubella* stumbled across equivalents for all three diseases from the ST. This instance was also counted as "researched" for all three *research units* (*measles*, *mumps*, and *rubella*).

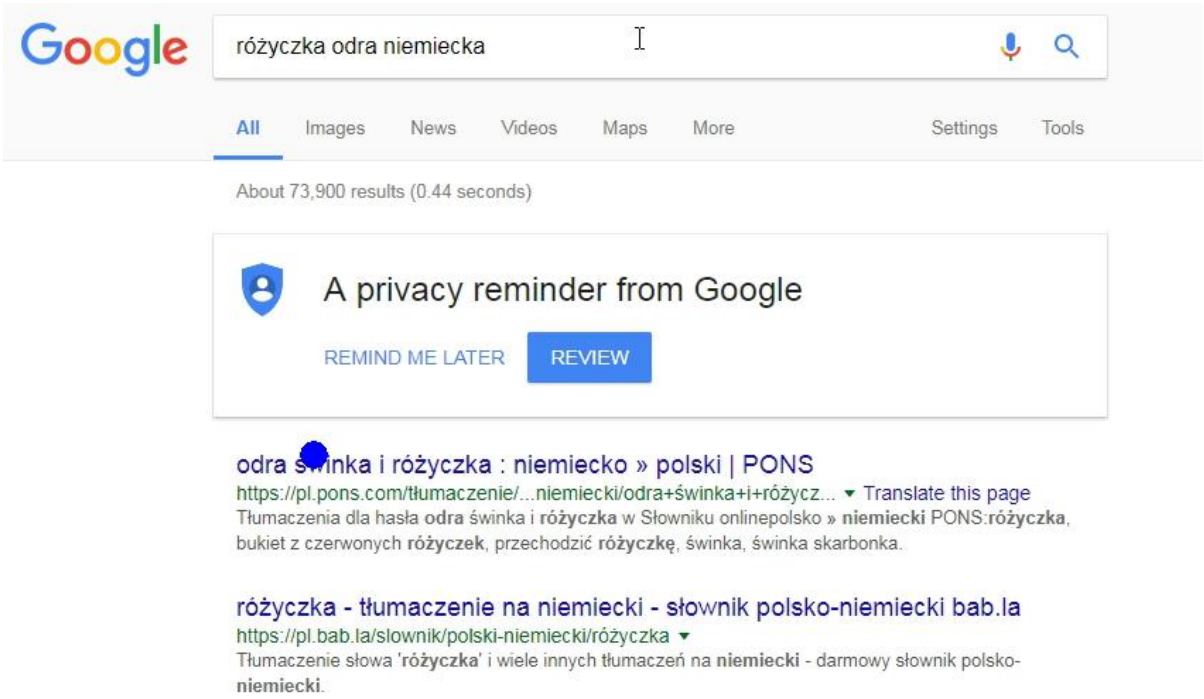


Fig. 26. Indirect research (*measles* and *mumps*) via search directly pertaining to another term (*rubella*)

One more very interesting research strategy was searching for a more general concept like *budowa* (anatomy) in Google Images in the context of headphones and ending up with a visual glossary of headphone parts. Such an instance of indirect research is provided in Fig. 27 below. The cursor is hovering directly under the Polish term for an ear pad: *nausznik.*

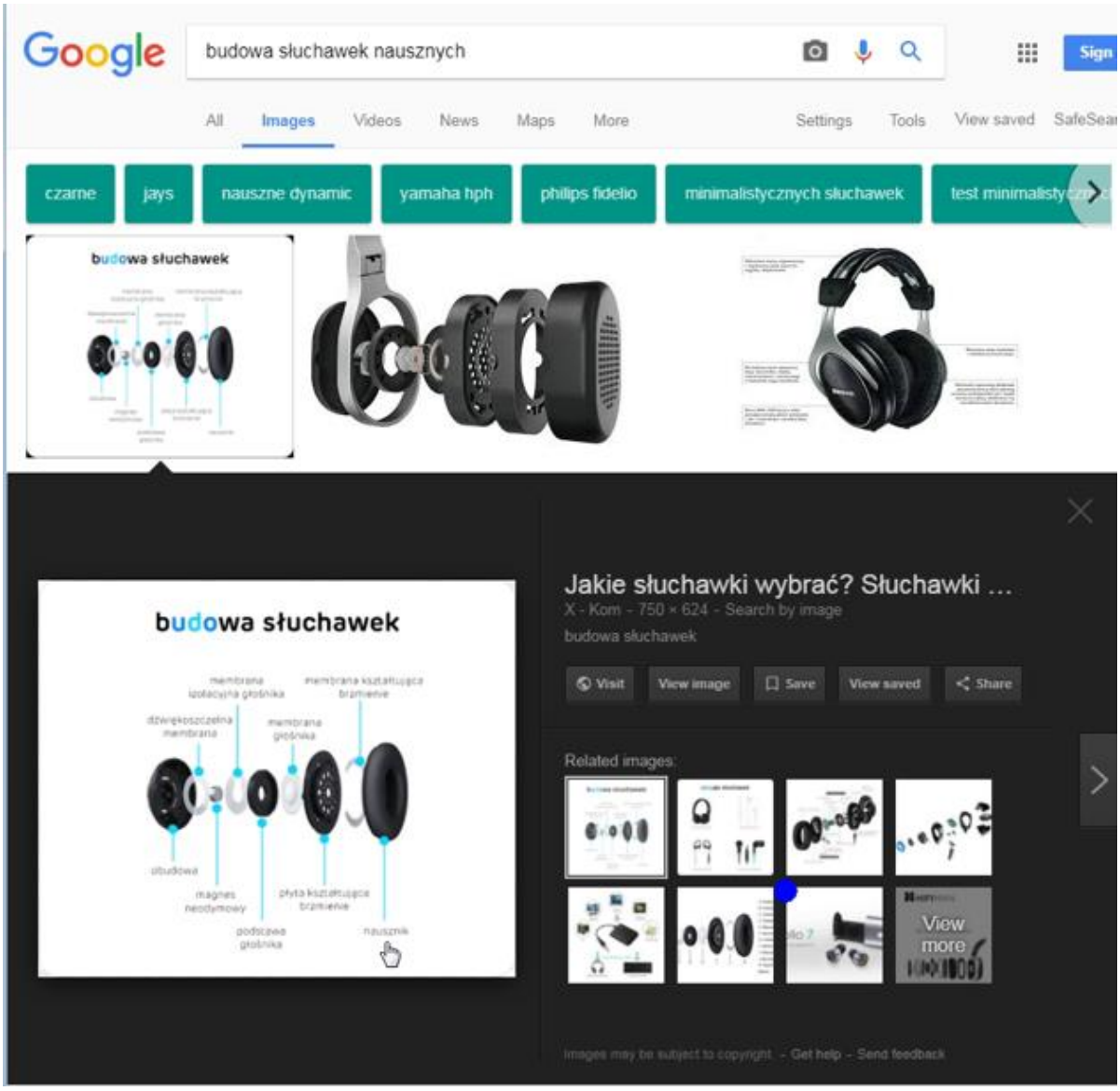


Fig. 27. Indirect research (*ear pads*) via general concept search

Finally, incomplete searches also were counted as instances of research. In the text WAT, G10 verified the phrasing *bransoletka zegarka* (=*watch bracelet*) only through typing a part of the phrase into the search bar and never actually hitting "search." The suggested search phrase(s) seemed enough proof for correct Polish phrasing (Fig. 28).

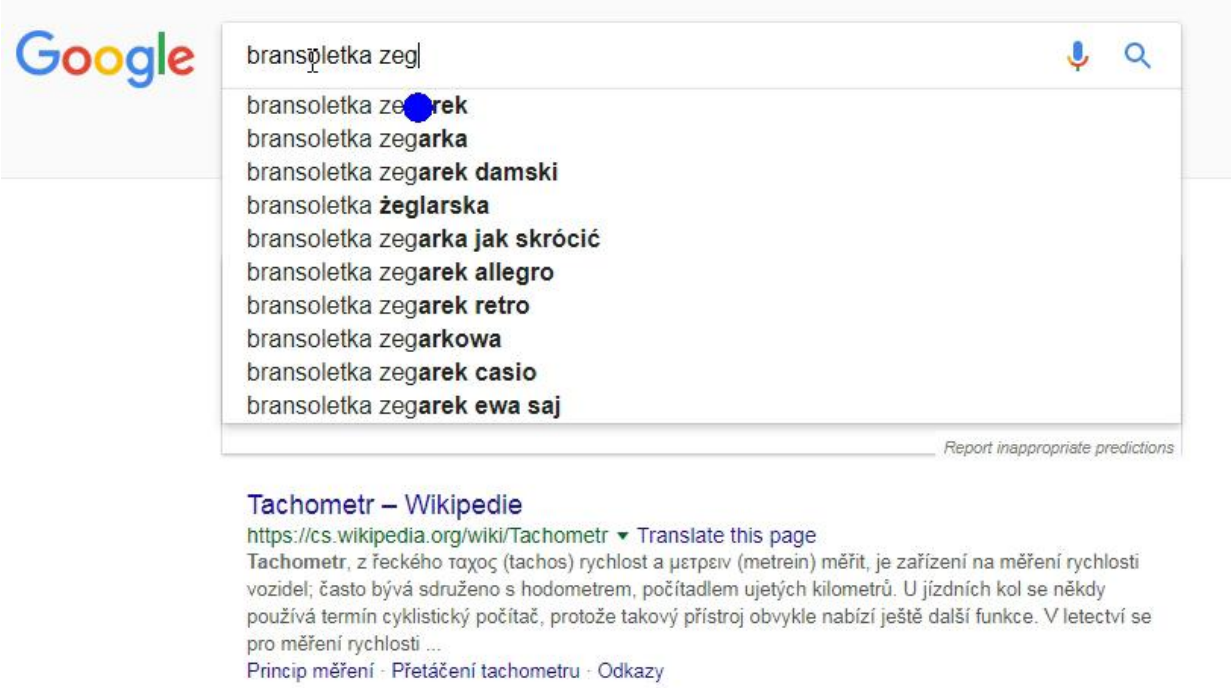


Fig. 28. Incomplete direct search via list of suggested search queries (*watch bracelet*)

After the analysis of product data via *accuracy* and *researched%*, the questionnaire data analysis was commenced.

4.10.3. Questionnaire data analysis

Finally, to supplement the process and product analyses, data collected via questionnaires was analysed regarding variables such as *attitude* and *perceived difficulty*. Open questions pertaining to the completed tasks and MT were then analysed as well.

Attitude towards MT here was operationalized as mean score from five statements from the pre-task questionnaire (see Appendix C: Pre-task questionnaire) that the participants were asked to rate on the visual analogue scale that translated into numerical values from -500 (strongly disagree) to 500 (strongly agree), -1 to 1 being the neutral attitude (in case the participant clicked the slider and changed their mind about sliding either way, the ±1 increment was there in a few cases). The question provided in the Appendix, but not on the list below ("*Internet is indispensable for translation*") was a distractor and was not counted into the analysis of MT attitude. The questions comprising the *attitude* variable were as follows:

(1) Machine translation is useful for everyday Internet browsing (e.g. shopping).

(2) Human translators will soon be replaced by machine translation.

(3) Machine translation cannot compete with human translation.

(4) Machine translation can speed up human translation.

(5) Machine translation is useful as a translation aid.

The values were hidden from participants, but each label translated into the ranges provided inTable 8. These labels themselves were not used in any analyses but are only provided here for reference to read and visualize the means and other descriptive statistics in the next paragraphs and tables. Values for questions (2) and (3) were then inverted to adjust the negative values to reflect the negative attitude and vice versa as it was for other questions.

Table 8. Attitude scale thresholds and their respective value ranges

Label	*Range*	
Strongly negative	-500	-251
Negative	-250	-2
Neutral	-1	1
Positive	2	250
Strongly positive	251	500

When it comes to the other questionnaire-based variable, *perceived difficulty*, it was calculated in the post-task questionnaire. In the analyses where both groups were treated as one, 79 data points were included (100%), with possible values ranging from 1 to 5 (Likert scale judgements, 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult), with *Mdn* = 3. The descriptive statistics with respect to individual texts by group were provided in Table 4, but more detailed descriptive statistics for *perceived difficulty* in both texts sorted by *group* and *task* are visible in Table 9 below.

Table 9. Descriptive statistics: *perceived difficulty* for *text* by *group* and *task*

Text	Group	Task	N	Median	Range	Minimum	Maximum	% of total N
Medical	T	TR	11	3.00	2	2	4	13.9%
		PE	11	3.00	2	2	4	13.9%
		Total	22	3.00	2	2	4	27.8%
	G	TR	8	3.00	3	1	4	10.1%
		PE	9	3.00	1	2	3	11.4%
		Total	17	3.00	3	1	4	21.5%
	Total	TR	19	3.00	3	1	4	24.1%
		PE	20	3.00	2	2	4	25.3%
		Total	39	3.00	3	1	4	49.4%
Technical	T	TR	11	3.00	1	3	4	13.9%

		PE	11	3.00	2	2	4	13.9%
		Total	22	3.00	2	2	4	27.8%
	G	TR	9	3.00	3	2	5	11.4%
		PE	9	3.00	3	2	5	11.4%
		Total	18	3.00	3	2	5	22.8%
	Total	TR	20	3.00	3	2	5	25.3%
		PE	20	3.00	3	2	5	25.3%
		Total	40	3.00	3	2	5	50.6%
Total	T	TR	22	3.00	2	2	4	27.8%
		PE	22	3.00	2	2	4	27.8%
		Total	44	3.00	2	2	4	55.7%
	G	TR	17	3.00	4	1	5	21.5%
		PE	18	3.00	3	2	5	22.8%
		Total	35	3.00	4	1	5	44.3%
	Total	TR	39	3.00	4	1	5	49.4%
		PE	40	3.00	3	2	5	50.6%
		Total	79	3.00	4	1	5	100.0%

Once the process, product, and questionnaire data analyses were completed, statistical analyses with reference to the seven hypotheses were conducted – which are described in the next section.

4.10.4. Statistical data analysis

The statistical tests for all hypotheses were conducted with IBM SPSS Statistics 25, 26, and 27. Graphs were also produced in SPSS. The specific tests for each hypothesis are provided in the results section below.

4.11. Results

The study tested the seven hypotheses introduced in Section 4.1.

4.11.1. H1: Both groups put more effort into information searching when translating than when post-editing

This hypothesis operationalised two dependent variables: *temporal effort* as time spent in applications (via Inputlog) and *cognitive effort* as average fixation duration for the same

researched groups (via eye-tracking). Both effort types were compared for the researched groups, G and T. The analysis involved examining potential differences between these two groups with respect to the type of task performed (PE, post-editing vs. TR – translation from scratch) and types of OR consulted. This hypothesis treated both groups, T and G, as one and tested for differences only on the level of task type with regards to cognitive and temporal effort (and technical effort in addition).

When it comes to the variables that had missing values, these could not be included in the model. As per the Shapiro Wilk test, the data were normally distributed. Thus, paired samples t-tests were calculated.

In terms of the eye-tracking data for the cognitive effort (*Text*, *Browser*, and *Total*), tests were run on a dataset with excluded data points (as per the data quality control parameters) and additionally with all data points included as well. Below are reported only the test results for the datasets with quality control applied (the variant with all data points included reached significance for the same pairs, but with lower p-values and effect sizes. In terms of Inputlog data for temporal effort (*WS_TT_Concordancer, WS_TT_BILINGUAL_DICTIONARIES,* and *WS_TT_Google_Search*), tests were run for both groups treated as one as well. Cohen's *d* was selected as the effect size measure for the t-test. The results of t-tests are summarized in Table 10 below.

Table 10. Paired sample t-test: Cognitive and temporal effort variables with respect to task type (PE – post-editing, TR – translation)

Variable	PE		TR		*t*	*p*	*d*	*df*
	M	*SD*	*M*	*SD*				
Text [ms]	257.94	31.90	272.83	40.36	-2.39	**0.028***	-0.53	19
Browser [ms]	289.23	49.33	299.05	29.89	-1.16	0.261	-0.26	19
Total [ms]	268.59	31.91	287.43	32.47	-4.07	**<0.001*****	-0.91	19
WS_TT_CONCORDANCER [s]	133.07	101.42	172.68	110.6	-2.15	**0.044***	-0.47	20
WS_TT_BILINGUAL_DICTIONARIES [s]	119.08	101.51	115.86	129.18	0.21	0.833	0.04	28
WS_TT_Google_Search [s]	190.68	153.65	288.15	219.53	-3.38	**0.002****	-0.56	36
WS_TT_LANGUAGE_REF [s]	45.62	10.77	63.18	47.17	-0.78	0.493	-0.39	3

The bar chart in Fig. 29 below illustrates the average fixation duration (eye-tracking data quality filter applied, $n = 20$).

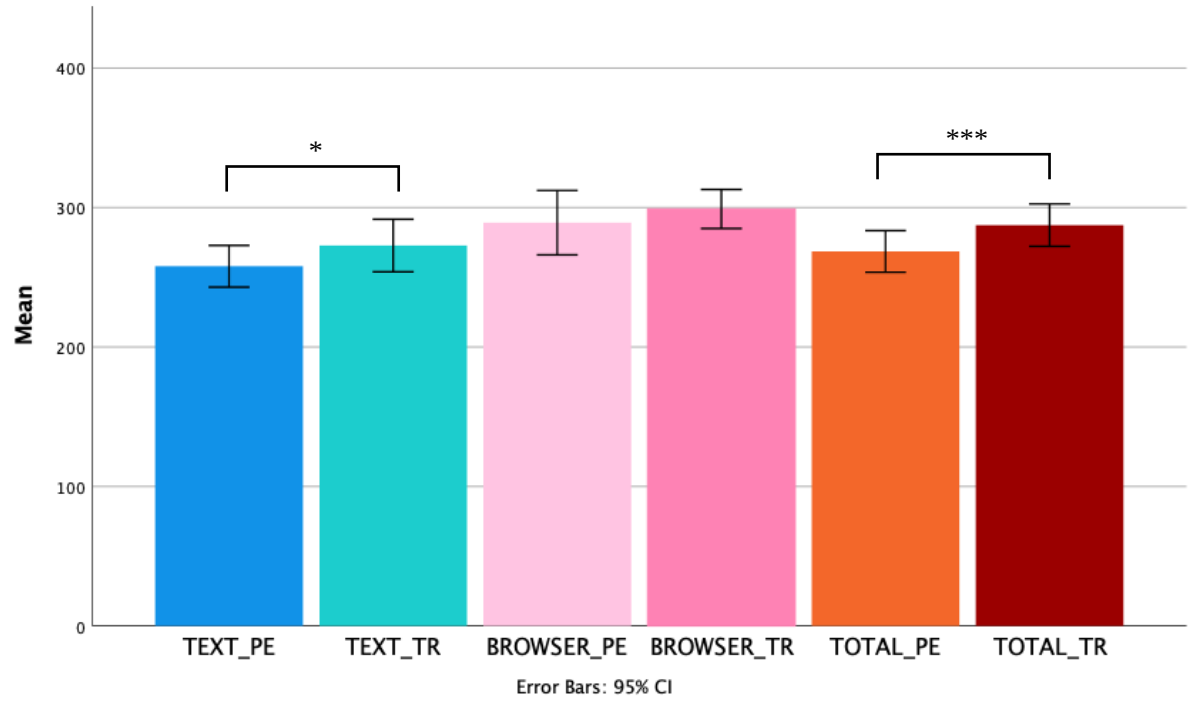


Fig. 29. Average fixation durations for cognitive effort variables Text, Browser, and Total by task (TR – translation, PE – post-editing)

Fig. 29 shows mean time spent in OR by task when both groups are treated as one for data points excluded pairwise as per the means analysed in the t-test. The below graph does not contain any error bars, for SD, see Table 10 above.

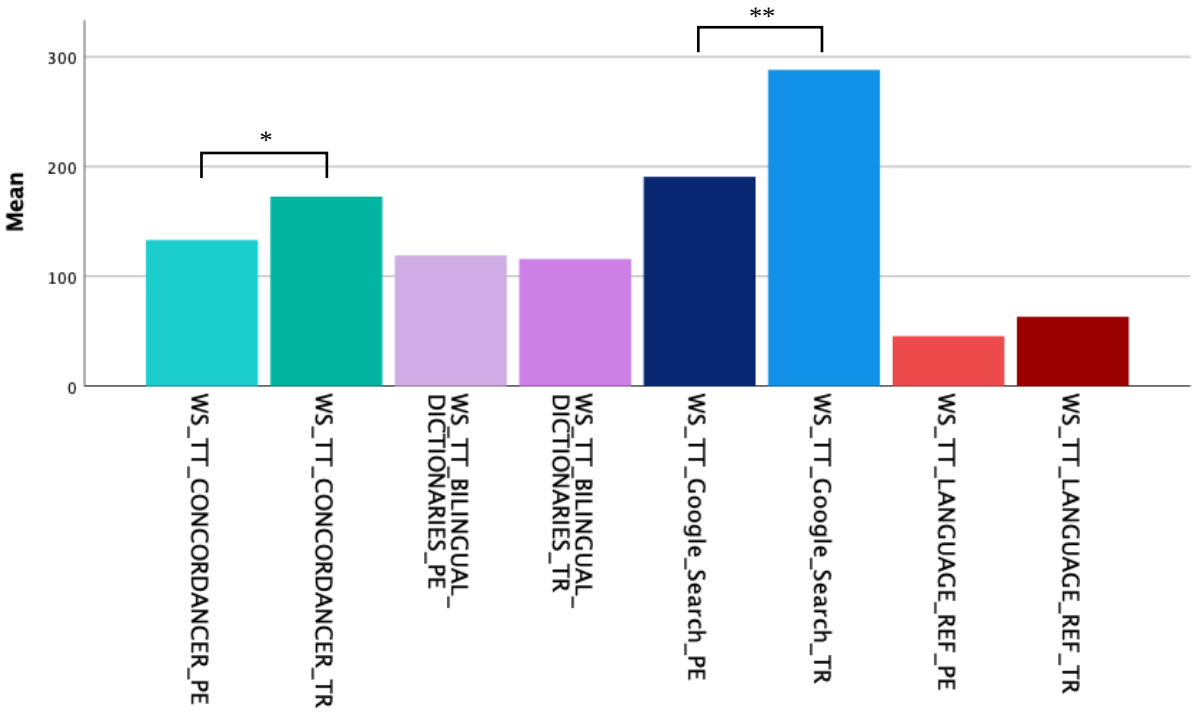


Fig. 30. Mean time spent in OR for temporal effort variables by task (TR – translation, PE – post-editing)

Four out of seven comparisons (*Text*, *Total*, *WS_TT_CONCORDANCER*, *WS_TT_Google_Search*) turned out to be statistically significant. Only for one of them (*Total*) the Cohen's *d* effect size coefficient proved to be large, whereas for other comparisons it was medium (*Total*, *WS_TT_CONCORDANCER*) and small (*WS_TT_CONCORDANCER*). For groups where the statistically significant differences were observed, the translation from scratch always had higher values for each variable than the post-editing task.

The remaining variables with the same number of complete data points (n = 39) were then analysed with the MANOVA, testing for significant differences between TR and PE. The effect size measure selected for the MANOVA was η_p^2. According to the Shapiro-Wilk W-test, the data were normally distributed and as per the Levene's test, the data also revealed homogeneity of variance. To check for significance in the differences regarding the fixation duration and time spent in ORs for respective tasks (TR – translation vs. PE – post-editing), the multivariate analysis of variance (MANOVA) has been conducted. Multivariate tests of the within-subjects main effect were made with the Wilk's Lambda test and the univariate tests were made with the Greenhouse-Geisser test.

The model proved to be statistically significant, $F(11, 28) = 37.38$; $p < 0.001$; $\eta_p^2 = 0.93$. This means that there are significant differences between TR and PE for the univariate tests between variables: *WS_TT_INPUTLOG_MS_Word, TT_Combined_OR, TT_All_EN_PL, WS_TT_Google, TT_EN_SUM, TT_PL_SUM, WS_T_totalTime, WS_totalKeypresses, WS_TT_REFERENCE, WS_TT_WIKIPEDIA, WS_TT_MONOLINGUAL, WS_TT_CORPORA*. The η_p^2 indicates a large effect size which means that the differences between tasks are very large. Comparisons for relevant dependent variables done with the Greenhouse-Geisser test are provided in Table 11 below.

Table 11. Greenhouse-Geisser test results for all variables in the MANOVA model (PE – post-editing, TR – translation)

Variable	PE		TR		*F*	*p*	η_p^2	Observed power
	M	*SD*	*M*	*SD*				
WS_TT_INPUTLOG_MS_Word [s]	604.04	337.60	790.95	221.51	11.29	**0.002****	0.23	0.906
TT_Combined_OR [s]	484.82	332.87	677.135	425.00	12.307	**0.001****	0.25	0.928
TT_All_EN_PL [s]	133.52	204.30	179.15	165.61	2.85	0.099	0.07	0.377
WS_TT_Google [s]	188.88	161.16	284.37	225.65	11.09	**0.002****	0.23	0.901
TT_EN_SUM [s]	40.24	81.81	40.55	52.93	0.00	0.977	0.00	0.050
TT_PL_SUM [s]	93.28	152.05	138.60	130.24	3.73	0.061	0.09	0.469
WS_T_totalTime [s]	1090.53	563.83	1469.75	522.89	15.25	**<0.001*****	0.29	0.967

WS_totalKeypresses	587.67	337.16	1539.92	299.13	185.74	**<0.001*** **	0.83	1.000
WS_TT_REFERENCE [s]	52.49	91.20	91.79	115.32	6.17	**0.018***	0.14	0.678
WS_TT_WIKIPEDIA [s]	55.80	85.30	50.60	79.40	0.10	0.752	0.00	0.061
WS_TT_MONOLINGUAL [s]	15.23	43.24	24.20	50.43	3.44	0.072	0.08	0.439
WS_TT_CORPORA [s]	3.94	17.89	0.89	5.56	1.01	0.321	0.03	0.165

Fig. 31 below illustrates the means from the table with significance levels.

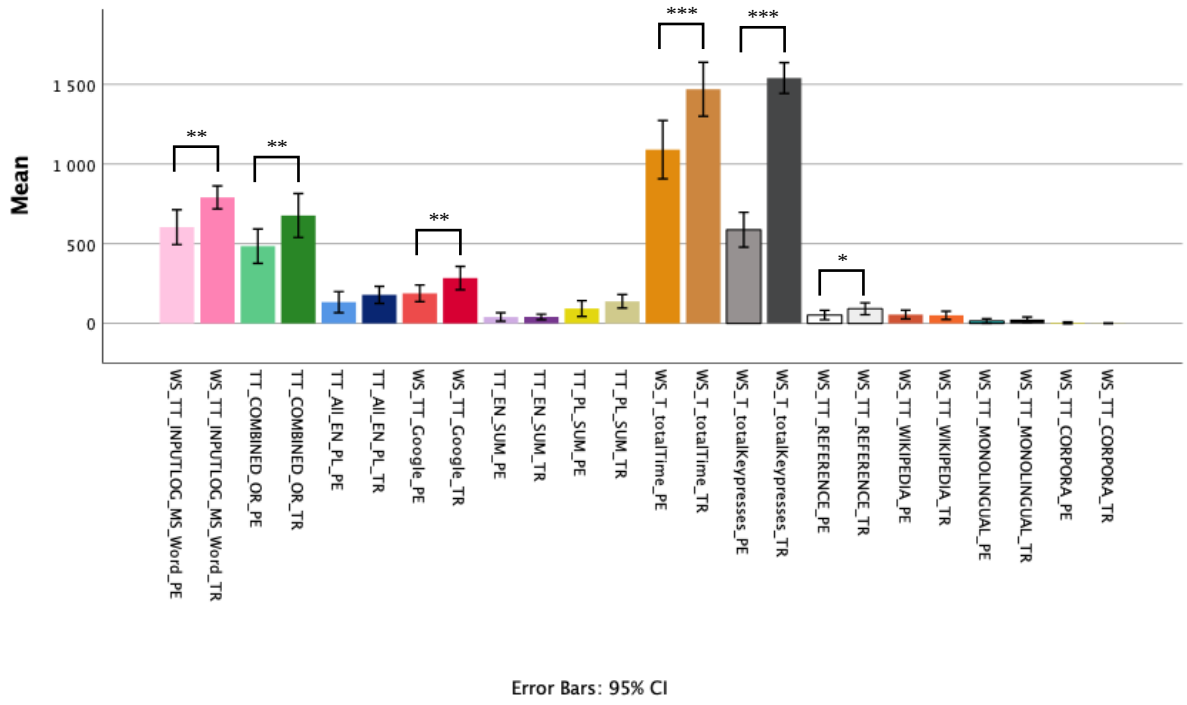


Fig. 31. Bar chart for temporal and technical effort means by task type (PE – post-editing, TR – translation) for *n* = 39

Six out of twelve tests (*WS_TT_INPUTLOG_MS_Word, TT_Combined_OR, WS_TT_Google, WS_T_totalTime, WS_totalKeypresses, WS_TT_REFERENCE*) turned out statistically significant with the effect size suggesting large differences between groups. Three tests (*TT_All_EN_PL, TT_PL_SUM, WS_TT_MONOLINGUAL*) approached statistical significance with clearly lower η_p^2 values. The remaining tests were not statistically significant and their low observed power is worth noting. Statistical power 1 - β (where β is the type II error probability) stands for the ability to reject false null hypothesis. Low values indicate the inability to reject the false null hypothesis, which suggests low reliability of nonsignificant tests.

Therefore, hypothesis 1 is partially confirmed, i.e. task type influences the amount of effort put into information searching and both groups put more effort into it when translating, albeit only in terms of browsing time, not cognitive effort – as indicated by non-significant results for the *Browser* variables comparison.

4.11.2. H2: Translation trainees put more effort into information searching than non-trainees in translation and post-editing

To test Hypothesis 2, independent samples t-tests, Mann-Whitney U test, and MANOVA were calculated in order to compare two tasks (TR and PE) and groups (T and G). For the variables included in the MANOVA, according to the Shapiro-Wilk test, the data were distributed normally and the Levene's test revealed the variance in the data to be homogenous.

When it comes to the remaining variables that had missing values, these could not be included in the model. For those variables where the data was normally distributed and the variance was homogenous, independent samples t-tests were calculated. For those variables with non-normal distributions, the non-parametric Mann-Whitney U test has been applied.

In terms of the eye-tracking data for the cognitive effort (*Text*, *Browser*, and *Total*), tests were run on a dataset with excluded data points (as per the data quality control parameters) and with all data points included as well. For both of these variants, tests were calculated for both tasks treated as one. None of the tests reached significance. The bar chart below in Fig. 32 illustrates the mean scores of mean fixation duration by group membership (n = 18 for G group and n = 22 for T group) when data quality filter was applied (some data points were excluded). The below Table 12 provides descriptive statistics for these variables by group.

Table 12. Cognitive effort: Descriptive statistics for Mean fixation duration by group for Text, Browser, and Total for both tasks treated as one (G – general, T – translation trainees)

Variable	Group	N	Mean [ms]	Std. Deviation
TEXT	T	22	263.97	38.09
	G	18	267.12	35.92
BROWSER	T	22	289.01	50.75
	G	18	300.41	22.75
TOTAL	T	22	278.20	34.56
	G	18	277.77	32.40

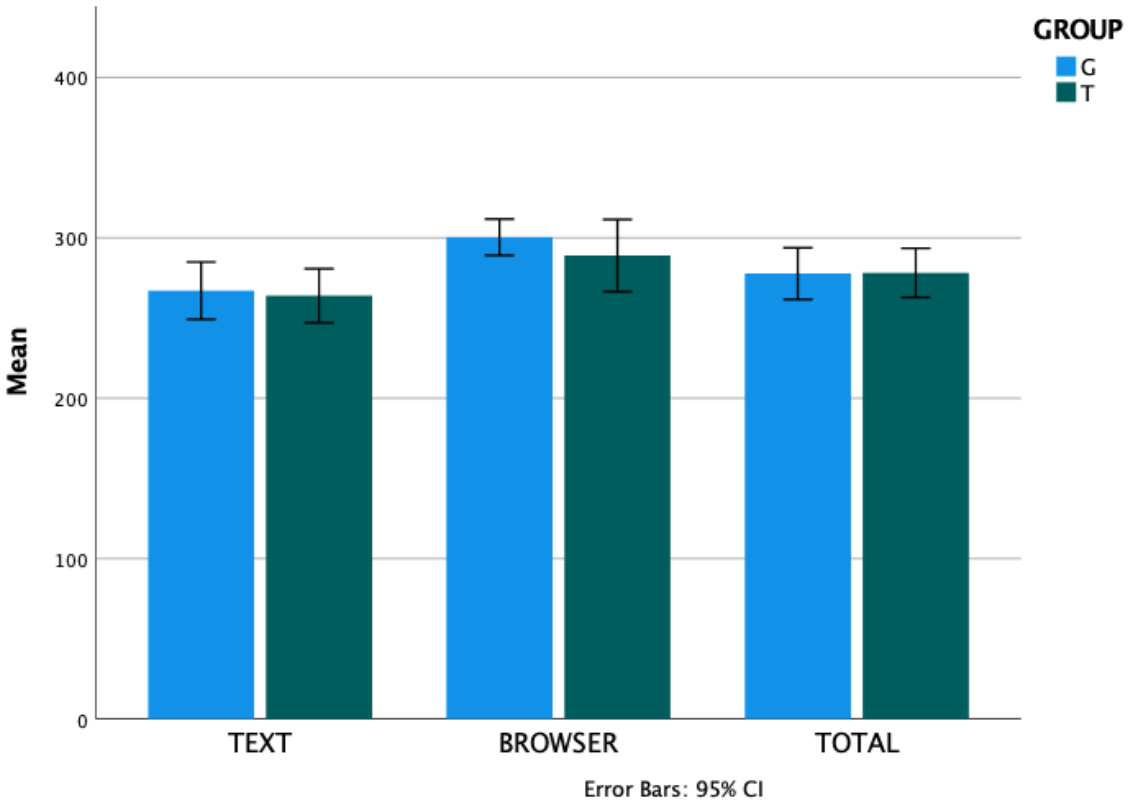


Fig. 32. Cognitive effort: Mean fixation duration by group for Text, Browser, and Total for both tasks treated as one (G – general, T – translation trainees)

In terms of Inputlog data for temporal effort (*WS_TT_Concordancer*, *WS_TT_BILINGUAL_DICTIONARIES*, *WS_TT_Google_Search*, and *WS_TT_LANGUAGE_REF*), the non-parametric tests were run for both tasks treated as one to test for differences between groups T and G. The only test to reach significance was for the time spent in language reference websites. Temporal effort in there in the T group (*Mdn* = 48.11 s) was higher than that in the G group (*Mdn* = 16.16 s). A Mann-Whitney U test indicated that this difference was statistically significant, $U(n_T = 6, n_G = 11) = 11.00$, $Z = -2.21$, $p = 0.027$. For this significant result, the effect size was large. Table 13 below provides the details of the Mann-Whitney U test, while Table 14 contains the descriptive statistics.

Table 13. Temporal effort: Descriptive statistics by group

Group	Variable	N	Mean	Median	Std. Deviation	Min.	Max.	Range
G	WS_TT_CONCORDANCER	17	128.10	92.21	131.91	11.78	471.34	459.56
	WS_TT_BILINGUAL_DICTIONARIES	28	94.19	78.86	66.87	4.77	242.24	237.46
	WS_TT_Google_Search	34	227.16	186.22	194.76	12.59	968.41	955.82
	WS_TT_LANGUAGE_REF	6	22.49	16.16	18.22	0.000	48.67	48.67
T	WS_TT_CONCORDANCER	36	140.42	118.37	100.08	1.25	328.26	327.01
	WS_TT_BILINGUAL_DICTIONARIES	38	126.16	87.56	134.58	15.77	615.38	599.61
	WS_TT_Google_Search	43	240.08	159.46	192.54	2.40	655.91	653.50
	WS_TT_LANGUAGE_REF	11	50.61	48.11	28.47	22.98	128.19	105.21

Table 14. Temporal effort: Mann-Whitney U test by group

Variable	*U*	*Z*	*p*	*R*
WS_TT_CONCORDANCER	261.000	-0.857	0.391	-0.15
WS_TT_BILINGUAL_DICTIONARIES	509.000	-0.298	0.765	-0.04
WS_TT_Google_Search	709.000	-0.226	0.821	-0.03
WS_TT_LANGUAGE_REF	11.000	-2.211	**0.027***	-0.67

The bar chart in Fig. 33 below illustrates the medians for the time spent in OR when both tasks are treated as one. The below graph does not contain any error bars, SD is provided in Table 13 above.

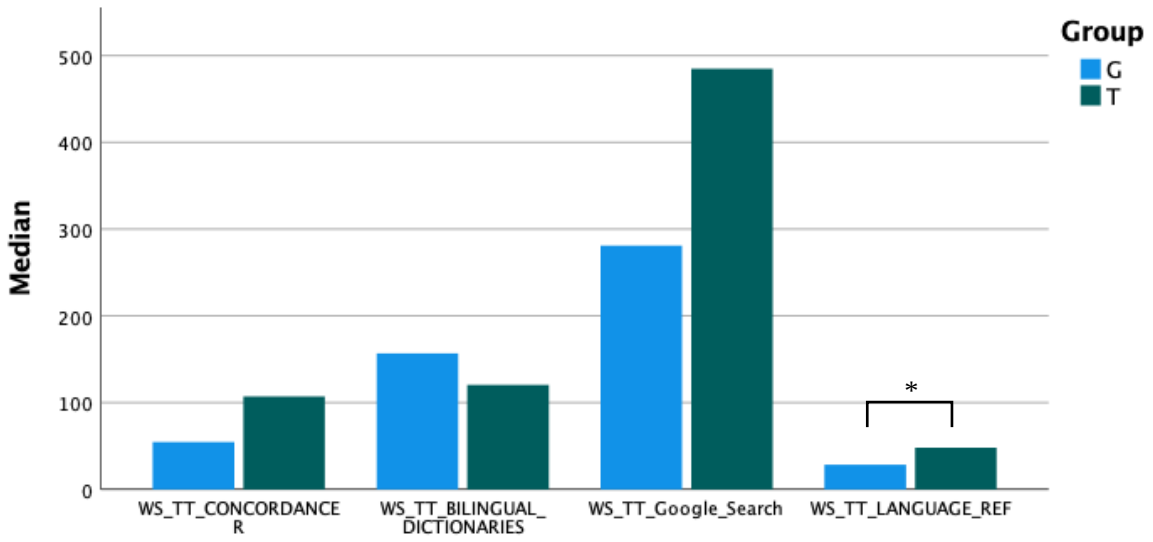


Fig. 33. Temporal effort: Median time spent in OR by group for both tasks treated as one (G – general, T – translation trainees)

A MANOVA was calculated to investigate the relationship between groups and tasks with the remaining variables, two main effects: within-subjects effect (task) and between-subjects effect (group). Multivariate tests of the within-subjects main effect were made with the Wilk's Lambda test and univariate tests were made with the Greenhouse-Geisser test.

The within-subjects main effect (task) turned out to be statistically significant, $F(11, 27) = 35.41$; $p < 0.001$; $\eta_p^2 = 0.94$. The between-subjects effect (group) was also statistically significant, $F(11, 27) = 2.36$; $p = 0.034$; $\eta_p^2 = 0.49$. However, the interaction between the group and task was non-significant, $F(11, 27) = 1.30$; $p = 0.279$; $\eta_p^2 = 0.35$. This means that the group variable does influence the multivariate construct created from the variables included in the analysis. It is interesting that according to the multivariate test the interaction between group and task did not reach significance, but did so for one variable in the univariate tests. These results are summarised in Table 15 below.

Table 15. Greenhouse-Geisser test results for the within-subjects main effect – task type (PE – post-editing, TR – translation) and for interaction between task type and group for temporal and technical effort

Effect	Effort	Variable	PE		TR		*F*	*p*	η_p^2	Observed power
			M	*SD*	*M*	*SD*				
TASK	Temporal effort [s]	WS_TT_INPUTLOG_MS_Word	604.04	337.60	790.95	221.51	10.85	**0.002****	0.23	0.894
		TT_Combined_OR	484.82	332.86	677.14	425.00	11.96	**0.001****	0.24	0.92
		TT_All_EN_PL	267.04	408.58	358.30	331.22	3.07	0.088	0.08	0.400
		WS_TT_Google	188.88	161.16	284.37	225.65	10.75	**0.002****	0.23	0.891
		TT_EN_SUM	40.24	81.81	40.55	52.93	0.01	0.913	0.00	0.051
		TT_PL_SUM	93.28	152.05	138.60	130.24	3.85	0.057	0.09	0.480
		WS_T_totalTime	1090.53	563.83	1469.75	522.89	14.74	**<0.001*****	0.29	0.962
		WS_TT_REFERENCE	52.49	91.20	91.79	115.32	6.52	**0.015***	0.15	0.701
		WS_TT_WIKIPEDIA	55.80	85.30	50.60	79.40	0.11	0.745	0.00	0.062
		WS_TT_MONOLINGUAL	15.23	43.24	24.20	50.43	5.70	**0.022***	0.13	0.643
		WS_TT_CORPORA	3.94	17.89	0.89	5.56	0.98	0.329	0.03	0.161
	Technical effort	WS_totalKeypresses	587.67	337.16	1539.92	299.13	177.99	**<0.001*****	0.83	1.000
TASK * Group	Temporal effort [s]	WS_TT_INPUTLOG_MS_Word	---	---	---	---	0.00	0.967	0.00	0.050
		TT_Combined_OR	---	---	---	---	0.03	0.857	0.00	0.054
		TT_All_EN_PL	---	---	---	---	0.48	0.494	0.01	0.103
		WS_TT_Google	---	---	---	---	0.02	0.883	0.00	0.052
		TT_EN_SUM	---	---	---	---	0.40	0.533	0.01	0.094
		TT_PL_SUM	---	---	---	---	0.25	0.618	0.01	0.078
		WS_T_totalTime	---	---	---	---	0.02	0.901	0.00	0.052
		WS_TT_REFERENCE	---	---	---	---	0.64	0.431	0.02	0.121
		WS_TT_WIKIPEDIA	---	---	---	---	0.02	0.899	0.00	0.052
		WS_TT_MONOLINGUAL	---	---	---	---	8.55	**0.006****	0.19	0.812
		WS_TT_CORPORA	---	---	---	---	0.00	0.963	0.00	0.050
	Technical effort	WS_totalKeypresses	---	---	---	---	0.00	0.976	0.00	0.050

For the within-subjects main effect, seven out of twelve Greenhouse-Geisser tests (*WS_TT_INPUTLOG_MS_Word*, *TT_Combined_OR*, *WS_TT_Google*, *WS_T_totalTime*, *WS_TT_REFERENCE*, *WS_TT_MONOLINGUAL*, and *WS_totalKeypresses*) were statistically significant. The η_p^2 coefficient indicates that only one of the effect sizes (*WS_TT_MONOLINGUAL*) is medium, but for all other variables the differences between tasks were large. The tests revealed the TR task to have higher values for dependent variables than the PE task.

When it comes to the interactions between the group and task, as illustrated in Table 15 above, only one of the tests (*WS_TT_MONOLINGUAL*) reached significance. The observed power is high for this interaction (0.81).

Table 16 below illustrates the results for the between-subjects effect tests. The only comparison to reach significance was for the variable *WS_TT_WIKIPEDIA*. The η_p^2 coefficient indicates that the effect size is only medium. The variable *WS_totalKeypresses* approached significance with a medium effect size. The values for the T group are higher for the dependent variable than the G group.

Table 16. Temporal and technical effort: Between-subjects effect of group (T vs. G)

Effort	Variable	T		G		*F*	*p*	η_p^2	Observed power
		M	*SD*	*M*	*SD*				
Temporal effort [s]	WS_TT_INPUTLOG_MS_Word	689.37	221.58	708.00	379.75	0.06	0.803	0.00	0.057
	TT_Combined_OR	653.37	446.51	487.28	285.67	2.35	0.133	0.06	0.321
	TT_All_EN_PL	378.03	448.43	228.09	219.98	2.02	0.164	0.05	0.282
	WS_TT_Google	240.52	193.45	231.58	212.42	0.03	0.876	0.00	0.053
	TT_EN_SUM	45.26	80.81	34.10	48.49	0.33	0.572	0.01	0.086
	TT_PL_SUM	143.76	166.22	79.94	94.86	2.79	0.103	0.07	0.369
	WS_T_totalTime	1344.33	603.14	1197.08	528.47	1.02	0.319	0.03	0.166
	WS_TT_REFERENCE	84.01	116.65	56.78	87.43	0.85	0.364	0.02	0.146
	WS_TT_WIKIPEDIA	72.96	96.45	27.62	48.35	5.23	**0.028***	0.12	0.606
	WS_TT_MONOLINGUAL	16.36	55.50	24.07	32.90	0.28	0.598	0.01	0.081
	WS_TT_CORPORA	3.04	15.69	1.61	9.36	0.22	0.639	0.01	0.075
Technical effort	WS_totalKeypresses	1122.34	589.19	988.03	554.02	3.41	0.073	0.08	0.436

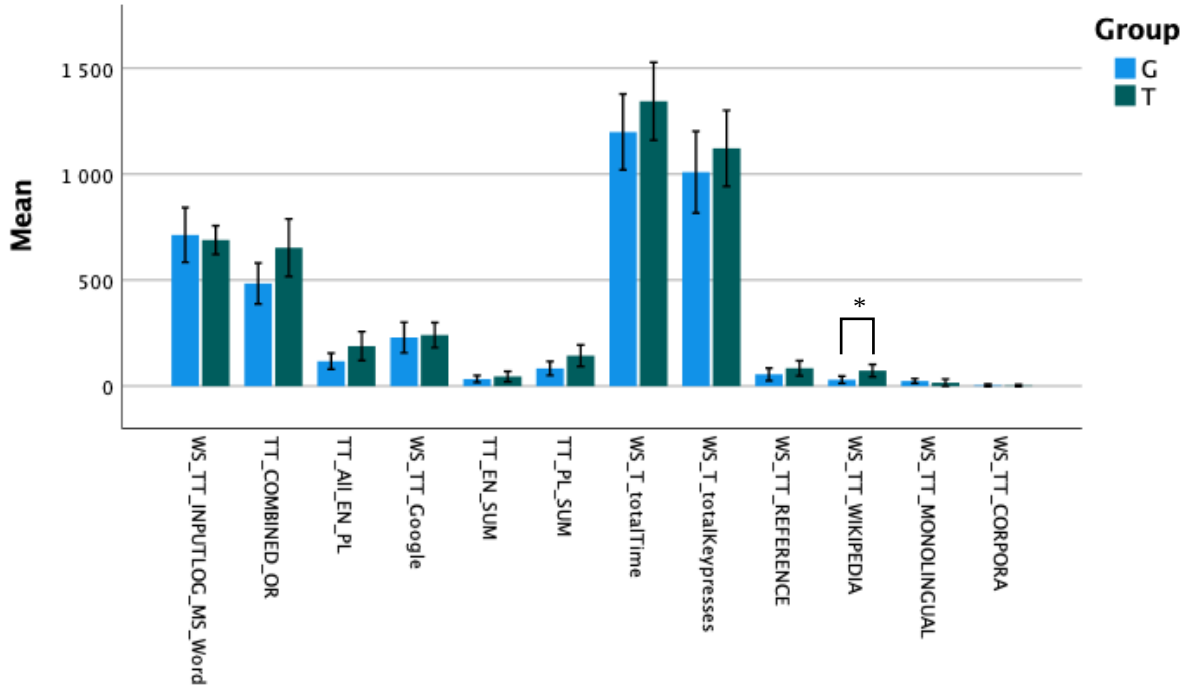


Fig. 34. Between-subjects main effect of group (G – general, T – translation trainees): mean temporal effort scores, *n* = 39 with 95% confidence intervals

According to both the test for the within-subjects effect and for the interaction test, the non-significant results usually feature much lower observed power than the significant ones, which increases the reliability of the ones which failed to reach significance.

Thus, Hypothesis 2 is only partially confirmed. While the group membership effect was found in the model, in univariate tests, the comparison for only one of the variables reached significance. As for the temporal effort and technical effort variables where the tests reached significance, the task type and group did not influence the dependent variables, except in the case of monolingual resources (*WS_Monolingual* in Table 15 and *WS_Wikipedia* in Table 16 along with *WS_TT_LANGUAGE_REF* in Table 14).

It can be concluded that, based on the conducted tests, translation trainees put more effort into information searching than non-trainees, but only for very specific types of resources. The effect was observed in three dependent variables, the effects for two of them are additionally illustrated in Fig. 35 and Fig. 36 below. The group membership effect can be seen in Fig. 35 below.

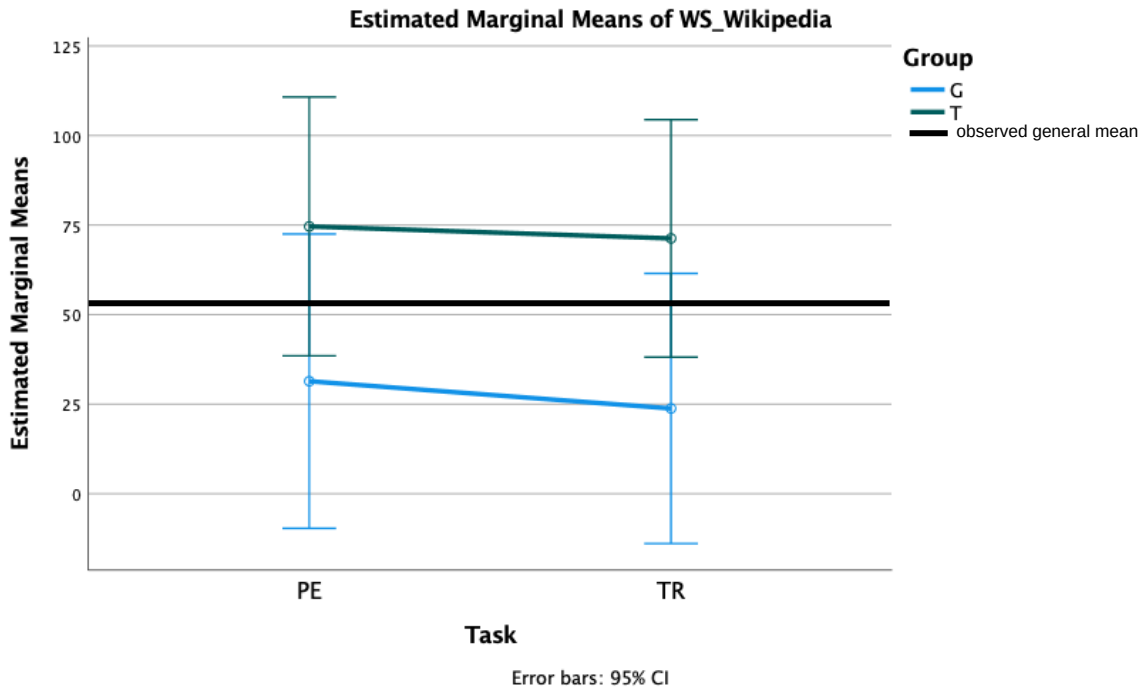


Fig. 35. Within-subjects and between-subjects effect for the dependent variable *WS_Wikipedia*. Estimated marginal means and 95% confidence intervals.

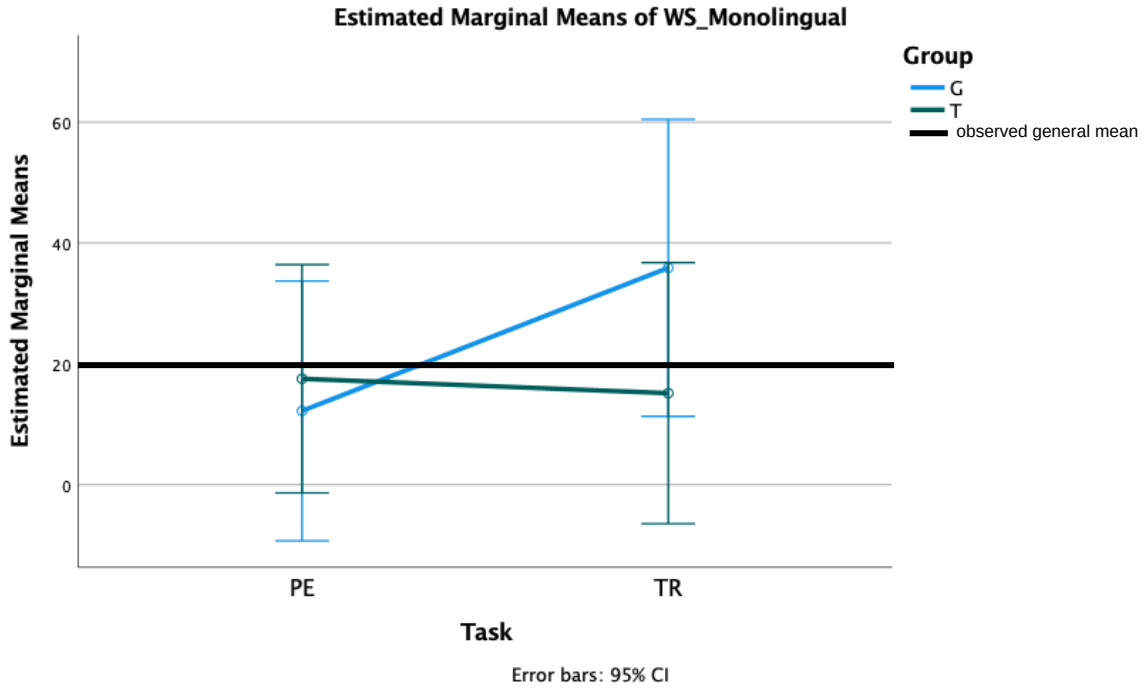


Fig. 36. Within-subjects and between-subjects effect for the dependent variable *WS_Monolingual*. Estimated marginal means and 95% confidence intervals.

As per Fig. 36 above, group differences in temporal effort for monolingual OR depend on task type. It was anticipated for both cognitive and temporal effort (the variable *Browser* along with the aggregate variable *TT_Combined_OR*) to be significantly higher in the T group

when compared to the G group, but a significant difference was observed only in temporal effort and not in the above mentioned aggregate variable of all OR combined. It could be concluded that the second hypothesis can be partially confirmed – the tasks and groups differ on the microscopic level of OR categories instead of the macro level of IS for all OR.

4.11.3. H3: The range of consulted resources is narrower when post-editing than when translating from scratch for both groups

For this hypothesis, the independent variables were the *Group* (two levels: T and G, like in the previous section) and *task* (two levels: TR and PE, like in the previous section). The dependent variable was *resource range* – range of consulted resources which was operationalized as the number of resource categories.

As per Table 17 below, the range was up to 10 resources per group or task, but the median oscillated around 4 and 5 per group or task. The difference between *resource range* means for both groups is small (5.49 for TR and 4.60 for PE). However, the potential significance of this difference is tested through the Generalised Linear Model (GLM) in the next sections.

Table 17. H3 Resource range: Descriptive statistics

GROUP	TASK	N	Mean	Median	Grouped Median	Std. Error of Mean	Minimum	Maximum	Range	Std. Deviation
G	TR	17	5.24	5.00	5.29	0.553	1	9	8	2.278
	PE	18	4.06	4.00	3.90	0.357	2	7	5	1.514
	Total	35	4.63	4.00	4.43	0.336	1	9	8	1.987
T	TR	22	5.68	5.00	5.50	0.433	2	10	8	2.033
	PE	22	5.05	5.00	4.86	0.499	0	8	8	2.340
	Total	44	5.36	5.00	5.21	0.330	0	10	10	2.190
Total	TR	39	5.49	5.00	5.40	0.340	1	10	9	2.126
	PE	40	4.60	4.00	4.31	0.324	0	8	8	2.048
	Total	79	5.04	5.00	4.83	0.239	0	10	10	2.121

The distribution of the counted number of resource categories is discrete and it is very often positively skewed (a value of 0 appears very often in the data set). According to the one-sample Kolmogorov-Smirnov (KS) test, the KS statistic shows that the variable *resource range* does not follow a normal distribution ($D(79) = 0.14$, $p = 0.001$). Moreover, in such situations

where the dependent variable is a count (the number of consulted resources), frequently a good approximation of the distribution of that variable is the Poisson distribution (the assumption of equal variances would be violated if we employed a normal distribution for a count variable). Therefore, to test whether the count data for the variable *resource range* come from a Poisson distribution a one-sample KS test was also employed and this variable is indeed Poisson distributed. The following paragraphs provide the results of the GLM for the *resource range* variable.

4.11.3.1. *Resource Range*: *Group* and *Task* as main effects, *Group*Task* as interaction

A GLM was constructed to test H3. The tested model included the *Intercept*, *group* and *task* (two main effects), as well as the Group*Task interaction effect. The Omnibus Test was conducted to compare the fitted model (here: (Intercept), *group*, *task*, *group***task*) against the intercept-only model. A likelihood ratio chi-square test evaluated whether the model that contains the predictors (the full set of predictors entered into the model, here: Intercept, *group*, *task*, and *group***task*) fits significantly better than a null model (the intercept-only model, without the predictors). The obtained Omnibus test result here is not significant (LR $\chi^2 = 5.57$, $p = 0.135$), hence it can be inferred that the model with predictors (added: GROUP, TASK, GROUP*TASK) does not show a significant improvement in fit over a null model (i.e. the intercept-only model).

In the model, *task* is not a significant predictor of the incidence rate for the number of resource categories (RR), $b = 0.255$, $SE = 0.1579$, $p = 0.106$). This indicates that there is no difference in predicted incidence rate between two types of tasks identified as TR and PE. Similarly, the regression coefficient (i.e. the regression slope) for *group* is non-significant ($b = 0.218$, $SE = 0.1507$, $p = 0.147$), indicating no difference in predicted incidence rate between participants identified as T and G. The regression coefficient for the interaction between *task* and *group* is also non-significant ($b = -1.137$, $SE = 0.2058$, $p = 0.505$).

4.11.3.2. *Resource Range*: *Group* and *Task* as main effects only

Another analysis without the *group***task* interaction as one of the predictors in the Poisson regression model was also conducted. In the tested model here, the Intercept was included in

the model along with the two main effects (*group* and *task*). The Omnibus Test compared the fitted model (here: (Intercept), *group*, *task*) against the intercept-only model. Here, the obtained Omnibus test result is non-significant (LR χ^2 = 5.12, p = 0.077), hence it can be inferred that the model with predictors (Intercept, *group*, *task*) does not show a significant improvement in fit over a null model (i.e. the intercept-only model).

In the model, *Task* is not a significant predictor of the incidence rate for the number of resources consulted (RR), b = 0.174, SE = 0.1005, p = 0.083). This indicates that there is no difference in predicted incidence rate between two types of tasks identified as TR and PE. Similarly, the regression coefficient (i.e. the regression slope) for *group* is non-significant (b = 0.145, SE = 0.1020, p = 0.156), indicating no difference in predicted incidence rate between participants identified as T and G.

4.11.3.3. *Resource Range* for two groups separately: *Task* as main effect

Additional analyses for the two groups separately have also been run. The results of the GLM (Poisson regression) for the variable *resource range* in the T and G groups separately show that the regression coefficient for Task does not reach statistical significance (b = 0.119, SE = 0.1304, p = 0.362 for T and b = 0.255, SE = 0.1579, p = 0.106 for T). Thus, there is no difference in predicted incidence rate between two types of tasks identified as TR and PE.

4.11.3.4. H3 results summary

A Poisson regression was run to predict the range of consulted resources based on translation training (T vs. G) and type of task performed (PE or TR). The results indicate that there is no effect of either *group* or *task* for the variable *resource range*, be it with or without Interaction. Thus, neither the T nor G group use a significantly narrower range of resources and neither of the groups do so in either of the *task* types (TR, PE). The hypothesis, therefore, is not confirmed – the range of consulted resources is not narrower when post-editing than when translating from scratch for both groups. There is no significant difference when it comes to the *resource range* variable for both groups.

4.11.4. H4: Translation trainees are more effective in Information Searching than non-trainees

The working definition of researching in the context of consulting resources for a given RU was established in Section 4.5. The effectiveness of IS here is operationalised via terminological accuracy for the selected RUs. The independent variable is *group* (two levels: T vs. G). The dependent variable is *accuracy*, a binary variable – 0 or 1. An additional correlational analysis on dependent variables *accuracy* and *researched unit %* was conducted in supplement of the between-group comparison. One research unit was excluded from analysis in the G group ("case" from the Technical text type, WAT text) because none of the participants in the group researched it. Additional analyses with the scenario including that unit were also conducted and are included below. The data proved to significantly deviate from a normal distribution (T group: $W = 0.693$, $p < 0.0001$; G group: $W = 0.793$, $p < 0.0001$), which is why nonparametric independent-samples Mann-Whitney U test was used to test the H4.

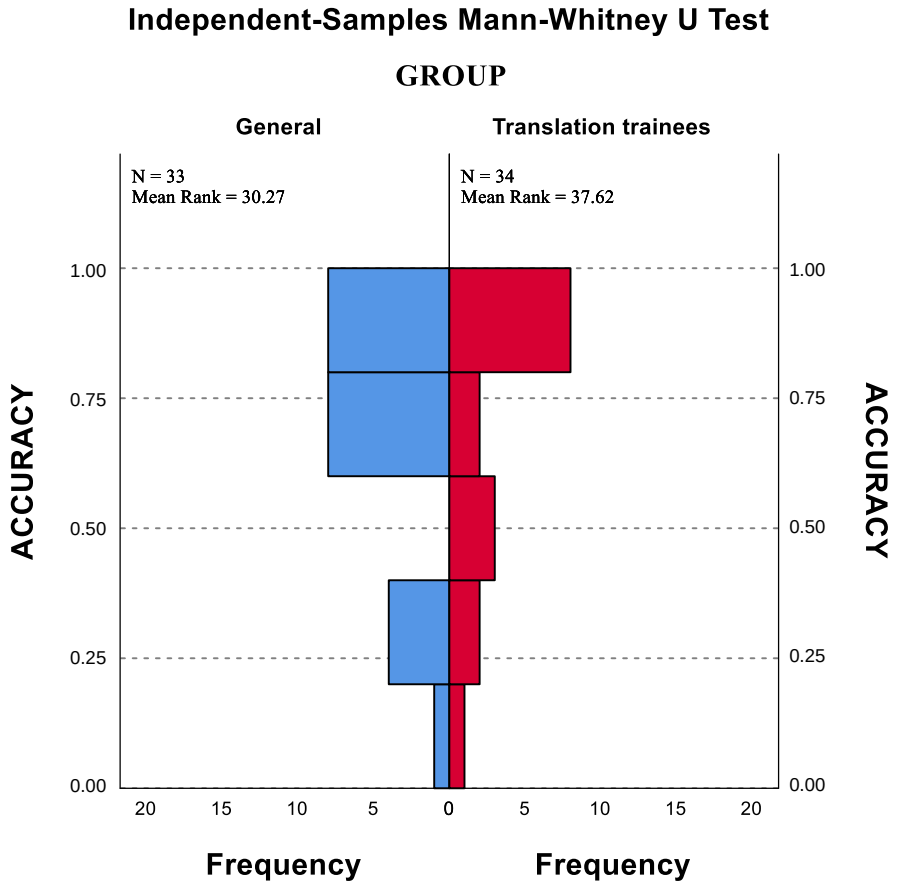


Fig. 37. H4: Independent-samples Mann-Whitney U Test

Fig. 37 shows the graphical interpretation of the analysis. The Mann-Whitney U statistic did not reach statistical significance ($U = 438.5$, $p = 0.106$). There are no statistically significant differences between T and G in their effectiveness (*accuracy* among all research units) in IS. The median *accuracy* was 1 and 0.89, respectively. Interestingly, when the same analysis was performed with the "case" research unit included – the Mann-Whitney U statistic for $n = 34$ also failed to reach significance ($U = 463$, $p = 0.138$). The medians did not change with this data point having been included.

Additionally, a correlational analysis was carried out between *accuracy* and the *researched unit (%)*. The value of the Shapiro-Wilk W-test calculated for *accuracy* is below 0.05 ($p < 0.0001$ for T and $p < 0.01$ for G). Thus, for each group separately (T and G) the nonparametric Spearman's correlation analyses were run to determine whether there is a statistically significant relationship between *researched unit (%)* variable and *accuracy*. No statistically significant correlation was found in either of the groups ($r_s = -0.167$, $p = 0.346$ for T and $r_s = 0.081$, $p = 0.653$ for G). Also, when the "case" research unit was included in both groups separately ($n = 34$), the Spearman's correlation failed to reach statistical significance as well ($r_s = -0.167$, $p = 0.346$ for T, $r_s = 0.016$, $p = 0.928$ for G).

The same analyses were conducted for both groups treated as one ($n = 67$), excluding one unit in the G group. The values of the Kolmogorov-Smirnov (K-S) test for the variable *accuracy* and for the variable *researched unit (%)* are below 0.05 ($p < 0.0001$ and $p < 0.01$, respectively). Therefore, Spearman's correlation was performed between the variable *accuracy* and the variable *researched unit (%)* and the analysis shows that the investigated correlation did not reach statistical significance ($r_s = 0.013$, $p = 0.920$). Finally, when the "case" research unit was included in the G group ($n = 68$), the Spearman's correlation also failed to reach significance ($r_s = -0.015$, $p = 0.905$).

To sum up, the hypothesis that translation trainees are more effective in IS than non-trainees cannot be confirmed as there is no significant difference between the groups in terms of accuracy of translated RUs. Furthermore, higher *accuracy* scores (per participant) do not correlate significantly with the percentage of researched RU (per person – per all four texts) – be it for the group of trainees, EFL students, or both groups together. This suggests that there may be no relationship between terminological accuracy and online consultations of given terms.

4.11.5. H5: Translation trainees' and non-trainees' attitude towards MT correlates with the percentage of time in online resources when post-editing

The percentage of time in ORs was calculated based on time spent in online resources (*TT_Combined_OR*) in relation to total task time (*WS_T_totalTime*) as per Inputlog data. Table 18 below provides the descriptive statistics for both correlated variables.

Table 18. Descriptive statistics for attitude and OR percentage of total time by group for PE task

		Statistic Group		
		G	T	Together
ATTITUDE	Mean	76.49	100.20	89.53
	Median	106.80	118.60	112.70
	Minimum	-119.60	-73.00	-119.60
	Maximum	230.80	196.20	230.80
	Range	350.40	269.20	350.40
	Std. Deviation	129.55	77.48	103.40
	N	18	22	40
PERCENTAGE_TIME_OR	Mean	40.74%	44.12%	42.60%
	Median	37.39%	44.17%	43.35%
	Minimum	15.61%	00.00%	00.00%
	Maximum	74.36%	75.48%	75.48%
	Range	58.75%	75.48%	75.48%
	Std. Deviation	16.92	16.60	16.62
	N	18	22	40

Each group separately as well as counted together on average displayed a positive attitude towards MT, the minimum scores never straying into the "strongly negative" range (below -250) and the maximum scores never into the "strongly positive" range (above 250). Interestingly, the range appeared smaller for trainees. For just the attitude scores between groups (n = 9 and n = 11 for G and T groups, respectively, data distributed normally), an independent samples t-test revealed no significant differences ($p > 0.05$) between the groups when it comes to their attitude scores.

Since the variables were not normally distributed ($D(18)=0.47$, $p < 0.05$ for attitude, group G) non-parametric correlation analyses were used (Spearman's correlation) to test the significance of the examined relationships. Two groups (T and G) were analysed separately as well as together, focusing on one task – post-editing. Correlations were made between *attitude* ratings (which were obtained before the participants saw any of the ST or MT) and *PERCENTAGE_TIME_OR*. No significant relationships were found in the T group or in the G

group. The variable *attitude* did not correlate significantly with this indicator of temporal effort. Fig. 38 below illustrates the values for both correlated variables.

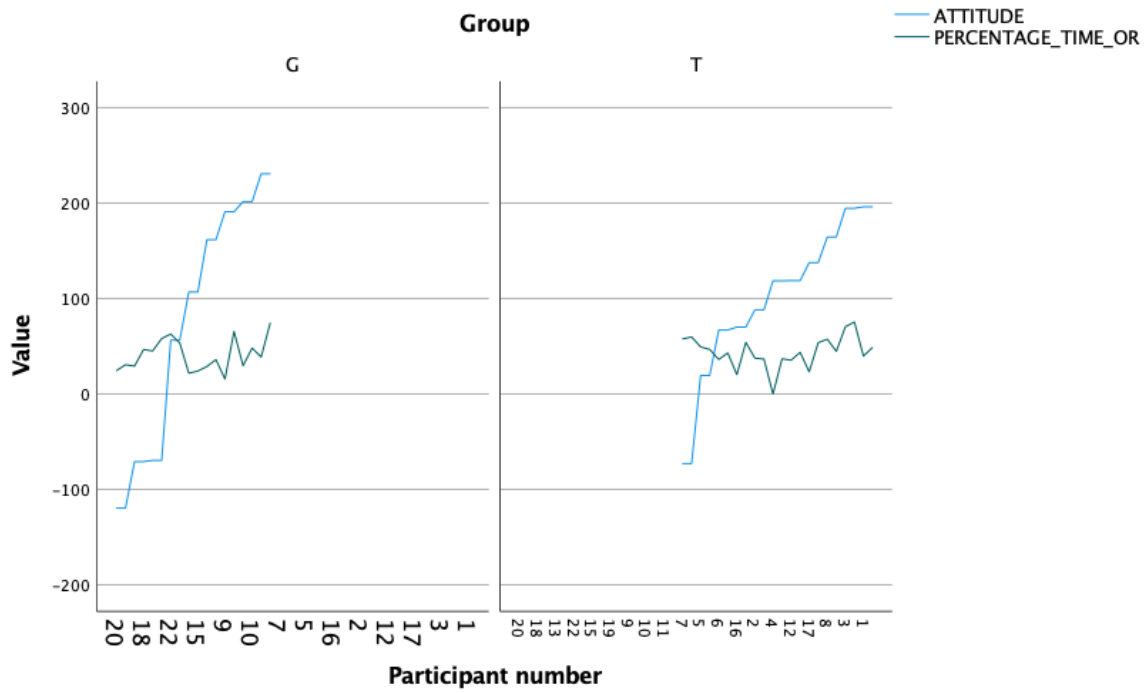


Fig. 38. Attitude and OR percentage of time values by group for PE task

As a follow-up to investigate other temporal effort variables in relation to attitude scores, *TT_Combined_OR* and *WS_T_totalTime* (as absolute values, not percentages) were also correlated with attitude both when both groups were treated as one and separately for the G and T groups. No significant correlations were found between these variables as well. Therefore, as attitude scores increase, the percentage of time in OR does not consistently increase or decrease with them, thus the hypothesis cannot be confirmed.

4.11.6. H6: There is a correlation between *perceived difficulty* and *temporal effort*

Since the variable *perceived difficulty* is ordinal, all conducted correlational analyses are nonparametric: in all cases Spearman's correlation coefficients (Spearman's rho) were calculated. In the case of all correlated variables, the analyses were conducted on raw data. In the analyses where both groups were treated as one, 79 data points were included (100%), with possible values ranging from 1 to 5 (Likert scale judgements, 1 = very easy, 2 = easy, 3 = medium, 4 = difficult, 5 = very difficult), with *Mdn* = 3. The descriptive statistics with respect

to the individual texts by group were provided in Table 3 in Section 4.8.1, but more detailed descriptive statistics for *perceived difficulty* in both texts sorted by *group* and *task* are visible in Table 9 in Section 4.10.3. All correlational analyses and significant results are reported in the subsequent sections.

4.11.6.1. Correlational analyses

Correlations were calculated between *perceived difficulty* and various temporal effort variables. A Spearman's correlation was run to determine whether there is a statistically significant relationship between temporal effort variables. The correlational analyses were calculated first on the aggregate categories (all resources combined, all English and Polish websites, all Google measures as well as MS Word window). Then, a more fine-grained analysis was run on subcategories (all English websites, all Polish websites, bilingual dictionaries). Subsequently, analyses on Google subcategories were also run (Google Search, Google images, Google Translate). Then, correlations were calculated on both English and Polish subcategories (reference websites, corpora, monolingual dictionaries, Wikipedia, and language reference websites). Finally, correlational analyses were run separately on each English and Polish OR subcategory (reference websites, corpora, monolingual dictionaries, Wikipedia, and language reference websites).

Tasks, groups, and texts treated as one:

- weak positive correlation for *TT_All_EN_PL* ($r_s = 0.24$, $n = 79$, $p < 0.05$);
- weak positive correlation for *TT_Combined_OR* ($r_s = 0.27$, $n = 79$, $p < 0.05$);
- weak positive correlation for *TT_EN_SUM* ($r_s = 0.33$, $n = 79$, $p < 0.01$);
- weak positive correlation for *WS_TT_CONCORDANCER* ($r_s = 0.31$, $n = 79$, $p < 0.01$);
- weak positive correlation for *WS_TT_REFERENCE_EN* ($r_s = 0.28$, $n = 79$, $p < 0.05$);
- weak positive correlation for *WS_TT_WIKIPEDIA_EN* ($r_s = 0.29$, $n = 79$, $p < 0.01$).

PE task, both groups and texts treated as one:

- weak-to-moderate positive correlation for *WS_TT_CONCORDANCER* ($r_s = 0.36$, $n = 40$, $p < 0.05$).

TR task, both groups and texts treated as one:

- moderate positive correlation for *TT_All_EN_PL* ($r_s = 0.42$, $n = 39$, $p < 0.01$);

- moderate positive correlation for *TT_EN_SUM* (r_s = 0.43, n = 39, p < 0.01);
- moderate positive correlation for *TT_PL_SUM* (r_s = 0.38, n = 39, p < 0.05);
- weak-to-moderate positive correlation for *WS_TT_Google_Translate* (r_s = 0.37, n = 39, p < 0.05);
- weak positive correlation for *WS_TT_REFERENCE* (r_s = 0.33, n = 39, p < 0.05);
- weak positive correlation for *WS_TT_WIKIPEDIA* (r_s = 0.32, n = 39, p < 0.05);
- weak-to-moderate positive correlation for *WS_TT_REFERENCE_EN* (r_s = 0.37, n = 39, p < 0.05);
- weak-to-moderate positive correlation for *WS_TT_WIKIPEDIA_EN* (r_s = 0.38, n = 39, p < 0.05).

T group, both tasks and texts treated as one:

- moderate positive correlation for *TT_EN_SUM* (r_s = 0.47, n = 44, p = 0.001);
- weak positive correlation for *WS_TT_CONCORDANCE* (r_s = 0.30, n = 44, p < 0.05);
- weak positive correlation for *WS_TT_REFERENCE_EN* (r_s = 0.31, n = 44, p < 0.05);
- weak positive correlation for *WS_TT_WIKIPEDIA_EN* (r_s = 0.33, n = 44, p < 0.05).

G group, both tasks and texts treated as one:

- weak-to-moderate positive correlation for *TT_All_EN_PL* (r_s = 0.34, n = 35, p < 0.05);
- moderate positive correlation between for *TT_Combined_OR* (r_s = 0.48, n = 35, p < 0.01);
- weak positive correlation for *TT_PL_SUM* (r_s = 0.35, n = 35, p < 0.05);
- weak positive correlation for *WS_TT_CONCORDANCE* (r_s = 0.35, n = 35 p < 0.05).

Informative-medical texts, both tasks and groups treated as one:

- weak-to-moderate positive correlation for *WS_TT_Google* (r_s = 0.38, n = 39, p < 0.05);
- weak-to-moderate positive correlation for *WS_TT_Google_Search* (r_s = 0.38, n = 39, p < 0.05);
- weak positive relationship for *WS_TT_REFERENCE* (r_s = 0.32, n = 39, p < 0.05);
- weak positive relationship for *WS_TT_REFERENCE_PL* (r_s = 0.32, n = 39, p < 0.05);

Operative-technical texts, both tasks and groups treated as one:

- weak positive correlation for *WS_TT_INPUTLOG_MS_Word* (r_s = 0.32, n = 40, p < 0.05);
- moderate positive correlation for *TT_All_EN_PL* (r_s = 0.47, n = 40, p < 0.01);
- moderate positive correlation for *TT_Combined_OR* (r_s = 0.53, n = 40, p < 0.001);

- marginally significant weak positive correlation for *TT_EN_SUM* (r_s = 0.31, n = 40, p = 0.05);
- moderate positive correlation for *TT_PL_SUM* (r_s = 0.38, n = 39, p < 0.05);
- moderate positive correlation for *WS_TT_CONCORDANCE* (r_s = 0.48, n = 40 p < 0.01);
- weak-to-moderate relationship for *WS_TT_Google_Translate* (r_s = 0.38, n = 40, p < 0.05);
- moderate positive relationship for *WS_TT_WIKIPEDIA* (r_s = 0.46, n = 40, p < 0.01);
- marginally significant weak positive correlation for *WS_TT_WIKIPEDIA_EN* (r_s = 0.31, n = 40, p = 0.052);
- moderate positive correlation for *WS_TT_WIKIPEDIA_PL* (r_s = 0.43, n = 40, p < 0.01);
- marginally significant weak positive correlation for *WS_TT_MONOLINGUAL_PL* (r_s = 0.35, n = 40, p < 0.05)

4.11.6.2. H6 summary

A number of temporal effort variables have been found to positively correlate with *perceived difficulty* values for all texts, tasks, and groups treated respectively as one – total time spent in all English and Polish websites combined, all OR combined, all English websites, concordancers, English reference websites, English *Wikipedia*. As perceived difficulty scores increased, so did temporal effort for all these categories. The effect size was weak for all above mentioned categories.

During the TR task, increased *perceived difficulty* correlated with total time spent in all English and Polish websites combined as well as English and Polish websites treated separately. The effect size was moderate for these categories. Increased perceived difficulty also weakly correlated with both *Wikipedias*, but for *Google Translate*, English reference websites, and English *Wikipedia* the correlation was weak-to-moderate. For the PE task, Concordancers were used more with increased *perceived difficulty* and the relationship was weak-to-moderate.

As the T group rated their *perceived difficulty* higher, they also spent more time in English websites (weak-to-moderate) as well as Concordancers, English reference websites, and English Wikipedia (all three weak correlations). As the G group rated their *perceived difficulty* higher, they spent more time in all OR combined (moderate) as well as in all English

and Polish resources combined (weak-to-moderate). Their temporal effort increased with difficulty ratings for Polish websites and Concordancers as well, albeit the effect size was weak.

As informative-medical texts were rated as more difficult, more time was spent in all Google measures as well as Google Search treated separately (weak-to-moderate). Furthermore, with higher difficulty ratings, more time was spent in English and Polish reference websites, Polish reference websites (both weak correlations). For operative-technical texts, higher ratings corresponded with more time spent in MS Word window (weak correlation). For all OR combined, English and Polish resources combined, Polish websites combined separately, concordancers, both *Wikipedias* and Polish *Wikipedia* separately – the effect size was moderate. For *Google Translate*, the correlation turned out to be weak-to-moderate. For English websites combined, English *Wikipedia* separately, and Polish monolingual dictionaries – the correlation only approached significance.

Below in Fig. 39 and Fig. 40, are presented the two strongest correlations (moderate) for the aggregate category of all OR combined (variable *TT_Combined_OR*).

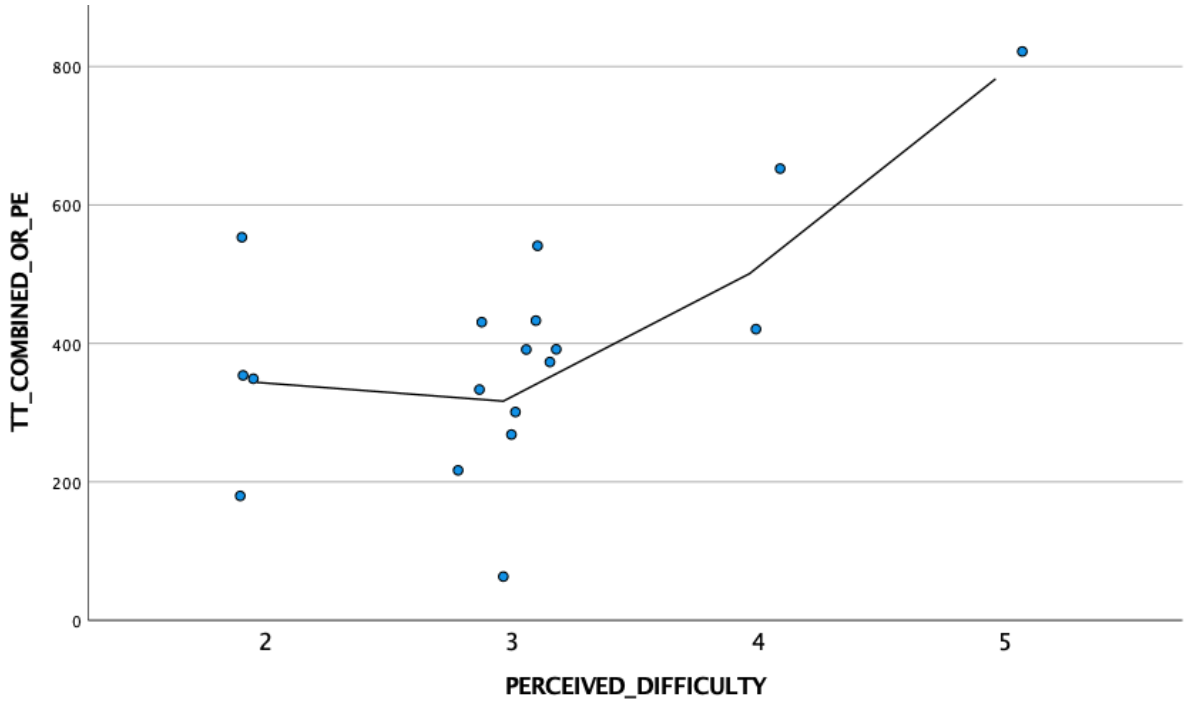


Fig. 39. Jitter plot for time in all OR by perceived difficulty for the G group

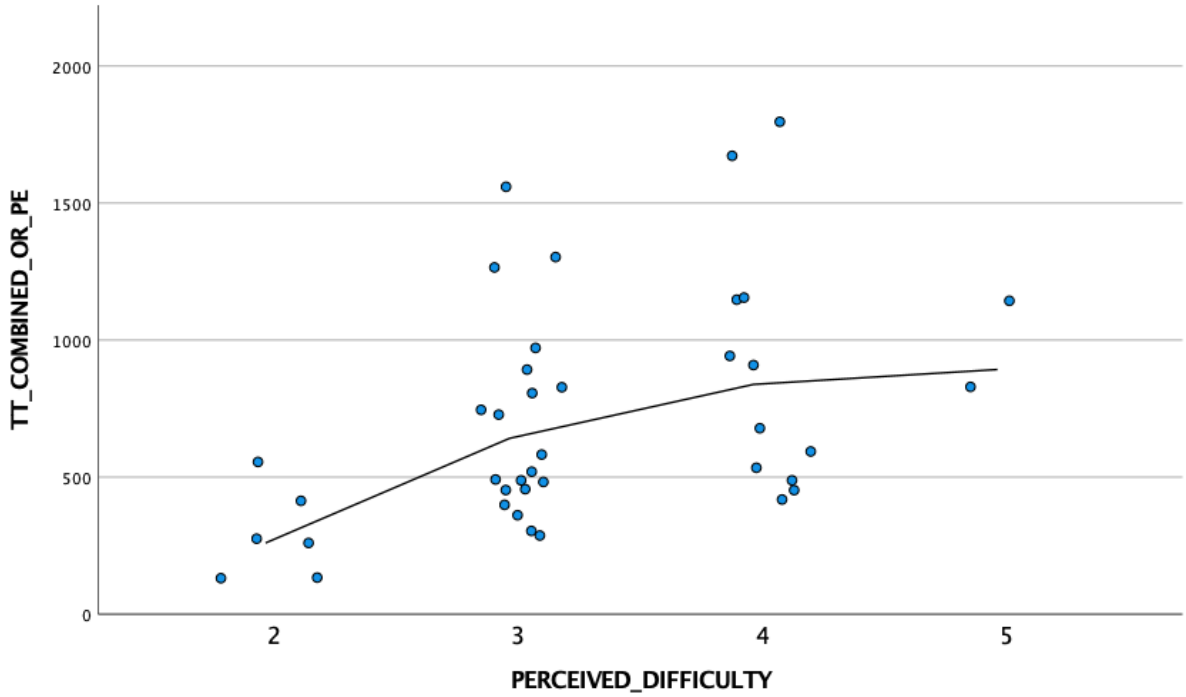


Fig. 40. Jitter plot for time in all OR by perceived difficulty for the operative-technical text

Therefore, in general, there is a relationship between students considering certain texts more difficult and their actual time spent in all OR combined or certain OR specifically. Thus, there exists a correlation between temporal effort and *perceived difficulty* – the hypothesis is partially confirmed.

4.11.7. H7: The range of consulted resources correlates with *perceived difficulty*

A Spearman's correlation was run to determine whether there is a statistically significant relationship between the *perceived difficulty* and the *resource range*. For the operative-technical texts, where both groups and tasks were respectively treated as one, there was a moderate positive relationship between the *perceived difficulty* and the *resource range* (r_s = 0.41, n = 40, p < 0.01). Thus, with *perceived difficulty* ratings increases for operative-technical texts, the range of consulted OR also increased. In conclusion, the hypothesis is only partially confirmed, as only for the operative-technical texts the increase in the *resource range* is significant.

4.11.8. Supplementary analysis: Conclusions from post-task questionnaire open questions

The final two questions in the post-task questionnaires were open and prompted the participants to reflect on their IS as affected (or not) by the presence of MT. The first question was: "How do you feel machine translation influenced your process of online researching (changed it completely, did not affect it at all, etc.)? Please provide a brief description." The second open question was meant to further prompt reflection on MT quality or other similar aspects: "Do you have any additional comments pertaining to MT or the task you just completed? Any insight will be greatly appreciated."

Out of 20 participants, twelve explicitly admitted that MT made the whole process faster/easier or that it was generally helpful (participants T1, T4, T5, T6, T9, T16, G10, G12, G18, G19, G20, G22). They admitted that with MT, their research "focused more on grammar, than on searching for vocabulary, which I suppose did save me some time" (G10) and that checking the already provided equivalent is easier and faster than looking for it without any prompt (T1, T5, T16). They did not need to "waste time on translating the easy parts" (G19).

Four participants expressed surprise at the quality of MT (T2, T5, T6, T16), their opinions ranging from "not that bad" to "really good". In general, six participants admitted explicitly that the MT quality was OK/good/very good (G20, T2, T5, T6, T16, T17). They wrote that it was easier to research minor things (T2) and that they "didn't have to search for all the terms" (T8). One participant admitted it made them spend less time checking things on the Internet (G13). They were suspicious of the high quality of MT, sometimes admitting that it made them verify whether the suggestions were correct or not (T17). Two people remarked on imperfections in MT (T4, G10) and one (T5) thought that certain text types should be translated from scratch ("it's better to translate an advertisement from scratch as the vocab in the text should be more vivd [sic]. I'm not sure whether MT could do this sufficiently.") It is not abundantly clear which text this referred to, but probably to one of the operative-technical texts which featured more of that persuasive language.

Three people (T7, G13, G15) insisted that MT did not affect their researching process. "If I had had to translate the same text from scratch, I suppose I would have followed the same pattern of proceedings. Since I don't trust MTs in general, I had to check translations of the terms anyway" is what one of the Trainees remarked (T7). The same person thought that translation from scratch would have been more "comfortable" for them – this referred to the WAT text, but when it came to the MMR text (medical), they expressed more enthusiasm

because they enjoy these types of texts in general. Thus, genre preference may be an interesting factor to investigate in future studies on attitude and other aspects of MT. Interestingly, two out of three participants mentioned here (T7, G13, and additionally G18 and G20) displayed a clearly negative attitude towards MT (their average score across five questions on their thoughts on MT were much below the neutral zero). However, what is also very intriguing, is that G18 and G20 initially (in the pre-task section on MT attitude) reported a negative attitude, but later on in the open questions (as reported in the two previous paragraphs) admitted that MT sped up the process (G18) and that it was quite good (G20).

Furthermore, the participants focused a great deal on verifying the correctness of the suggestions. As G15 wrote, "Sometimes MT is misleading", so a lot of participants explicitly remarked on their need to check the MT (T2, T3, T6, T7, T16, T17, G11, G20). Additionally, four people commented on the fact that MT prompted them to pursue certain solutions in the TT (T5, T7, G19) or how to interpret the ST (G15, G19). T7, the least enthusiastic about MT in general from all other participants, said "during the process of post-editing I wished I had been able to do it without looking at the MT text." A very interesting remark was provided by G19: "It was harder to approach the source text independently of the MT and to think of better ways to formulate sentences in Polish. In a way it was helpful, but in a way not that much, and I was all the time hesitating to what extend [sic] I can trust the MT" – they were more focused on style than correctness, albeit still thinking about potential errors suggested by MT.

Finally, there were two opinions that betrayed a certain degree of a misconception as to what MT's role actually is. T1 opined that "MT is very useful when translating specialized texts from a filed [sic, typo – should be field, OW] that is not necessarily well-known to the translator" which could potentially mean they think MT is a shortcut to dealing with complex texts. Since this sentence was not elaborated on by the author, it is difficult to judge whether they knew that MT suggestions could be dangerous traps when a field is unknown to the translator. Moreover, G19 said that as the text difficulty increases, MT is more helpful and the better it is to rely more on it, which could mean that it provides a certain degree of substitute for expertise.

4.12. Discussion

None of the hypotheses formulated to investigate IS in translation and post-editing were fully confirmed and out of all seven, four (H1, H2, H6, H7) were partially confirmed. Three (H3,

H4, H5) were not confirmed. In the subsequent sections all results are interpreted with reference to results of previous research in TPR and theoretical issues is TS. Limitations, future research avenues, and didactic considerations will also be considered.

4.12.1. Hypothesis 1

Both groups put more effort into information searching when translating than when post-editing

The first hypothesis tested the effect of task type on IS effort – temporal and cognitive. Hypothesis 1 has been partially confirmed: for the temporal effort indicators, i.e. all OR combined together, as well as other aggregate temporal effort variables: all English websites, Google measures, and all reference websites. The tests only approached significance for all English and Polish websites together and Polish websites separately, as well as in monolingual dictionaries.

For the tests where statistically significant differences were observed, TR always displayed higher values of the dependent variable than PE. In terms of eye-tracking measures, no significant differences were observed in reference to two tasks regarding the *Browser* variable, as opposed to the *Total* and *Text* variables. This could indicate that interaction with the text itself (the area of interest for both the ST and TT in the *Text* variable, i.e. the MS Word window) and the whole task (*Total*, the area of interest encompassing the entire screen) generated significantly more cognitive effort for the TR task than the PE task for both groups.

While cognitive processing is found to be more intense during OR than when looking at ST and TT (Hvelplund 2017; Whyatt et al. 2021), task type had no effect on the amount of cognitive effort in consulting OR. It could perhaps imply that the amount of cognitive processing involved in OR as compared to other areas of the screen reflects the complexity of this particular task, but is not influenced by task type. Conversely, cognitive processing involved in interacting with the ST and TT as well as for the entire screen throughout the task is indeed influenced by the task type – searching for information irrespective of task generates comparable cognitive effort when in the browser. The difference between the tasks might be visible before and/or after the IS even then.

In terms of temporal effort, more effort was put into the entire task of TR (*WS_totalTime*) and while engaging with the ST and TT in the MS Word application during the TR task, i.e. when reading ST/TT and producing TT. The fact that total task time for PE was significantly lower than for TR is in line with Daems et al. (2016) and previous studies, such as Plitt and Masselot (2010). Both, all OR combined as well as certain types of OR, generated significantly more effort for the TR task: all Google measures (*Search*, *Translate*, and *Images* counted together) as well as *Google Search* separately, concordancers, and reference websites. This selective effort increase for the TR tasks may indicate the more effortful nature of the particular goal of information searching happening in search engine consultation, reference resources, and concordancers. To further investigate this effect, a qualitative analysis of these searches and consultations would have to be conducted. The fact that for the significant comparisons, it was the TR mean that was always higher is also in line with Daems et al. (2016) who found a trend of more time spent in each resource when translating.

In addition to the indicators or temporal and cognitive effort, a technical effort variable (*WS_totalKeypresses*, the total number of keypresses) was included in the analyses. This was done to see whether the tasks differed in terms of the physical aspect of typing/editing/deleting as a whole. It turned out that the difference in terms of the total number of keypresses proved to be statistically significant with a large effect size, thus suggesting that participants were not inefficient with their editing/deleting/typing during PE and managed to use a substantial amount of MT instead of retranslating the TT. However, it ought to be stressed that this variable did not differentiate between windows – the scores and means were calculated based on the entire tasks, including what was typed and edited in the browser.

According to Densmer (2014), full post-editing could require more effort than translation from scratch. It was a probability here as well, considering that both groups were not experienced with post-editing. However, the results of this hypothesis confirm that as the tool was designed, MT not only results in decreased total task time, but also – which is paramount to this study – IS effort in OR. What is important is that translation experience for the participants was extremely limited or, for the EFL students – non-existent, which means that when provided with PE instructions, even translation trainees and non-translators have the potential to take advantage of MT as a translation aid in terms of temporal effort. It should be noted that translation quality was not verified beyond terminological accuracy for the selected rich points, so actual advantage in terms of final product quality remains to be researched.

When it comes to cognitive effort in IS, there was no significant difference in average fixation duration for OR between the tasks which suggests that processing visual input (textual, graphic, etc.) in consulting OR might be comparable for both task types. The other two cognitive effort variables differed significantly, which in turn suggests that cognitive processing is more intense on the global level of the task and on the level of interaction with the ST and TT. However, the more intense processing on the global level as well as on the level of ST/TT interaction cannot be regarded as independent from the effort involved in using OR – these are intertwined, albeit when considered alone the effort put into using the OR in the browser clearly is not significantly different between the tasks. As Hvelplund (2017) and Whyatt et al. (2021) concluded, the use of OR adds to the effort in the entire process. Different results were found for Hypothesis 2, which is discussed in the next section.

4.12.2. Hypothesis 2

Translation trainees put more effort into information searching than non-trainees in translation and post-editing

The second hypothesis posited that trainees would put more temporal and cognitive effort into IS in both tasks than EFL students would. The results show that this hypothesis can only be partially confirmed.

Similarly to Hypothesis 1, for cognitive effort there were no significant differences found in terms of average fixation durations between groups – not only for the IA in OR, but also on the text area in MS Word and the area of the entire screen.

When it comes to the temporal effort variables, the main effect of group membership was observed only for two variables – both *Wikipedias* treated as one and language reference websites. For the *Wikipedias* variable, trainees put more effort into using it (M = 72.96 s) than the EFL students (M = 27.62 s) and similarly for the language reference websites the median scores for the T group (Mdn = 48.11 s) was higher than that in the G group (Mdn = 16.16 s). The interaction of group and task was found to be significant only in the time spent in monolingual dictionaries i.e. the use of this type of OR is influenced by both the training and task type. In this case, the EFL students put more effort (M = 24.07 s) into using this type of resource than trainees (M = 16.36 s), when the tasks were treated as one. Thus, the hypothesis

can be confirmed only for those three types of resources, not on the global level of time spent in all OR or in terms of cognitive effort in the browser area. According to Kuznik (2017), spending a lot of time in OR and carrying out many searches is a characteristic of translators. It is interesting then that no significant difference was found in the data when it comes to time devoted to OR in general between the groups. In terms of the lack of group effect on the aggregate variable of all OR combined, it is in line with Daems et al. (2017: 257) who found there to be no significant difference between either groups or tasks in time spent in OR, as opposed to what the previous study established (less time in OR spent while post-editing than translating, according to Daems et al. 2016).

Kuznik (2017) also found that translators spent more time searching when translating into their L1 than into L2, while teachers spent less time when translating into L1. Conversely to that finding, neither of the groups spent significantly more time searching in Google (neither for all *Google* measures or *Google Search* separately) when translating. Like in a part of Kuznik's study, both of these tasks were done into L1 here.

Tirkkonen-Condit (1990) observed that non-professionals rely heavily on dictionary use which is indicative of treating translation as a lexical task. The fact that the majority of both groups apart from bilingual dictionaries used also reference works, *Wikipedia*, monolingual dictionaries, corpora, and language references means that both EFL students and translation trainees are aware of the need to understand the ST and TT beyond the linguistic layer. More than for half of all data points (n = 79; for groups and tasks treated as one) Polish reference websites and Polish *Wikipedia* were accessed (41 and 44 data points, respectively). 21 data points feature visits to Polish monolingual resources and for 16 data points Polish language reference websites were also accessed. This could be explained by the fact that even though the participants were Polish native speakers, there were still issues with production – probably caused by the specialised nature of STs. Another possible interpretation is that both groups are highly proficient in their L2 (English) and translating into their L1, i.e. their stronger language. Admittedly, 84% of all data points featured consultations of bilingual dictionaries and his tendency is already widely established (cf. Désilets et al. 2009; Gough 2017; Hvelplund 2017; Whyatt et al. 2021 among others). The IS behaviour of study participants does confirm what Tirkonnen-Condit (1990) posited, albeit there is much more nuance to the IS process – even among EFL students who have no training awareness of the necessity to go beyond the retrieval of equivalents. This means that both groups here appear to be aware of how language works and that terminology in translation requires extralinguistic knowledge.

4.12.3. Hypothesis 3

The range of consulted resources is narrower when post-editing than when translating from scratch for both groups.

The third hypothesis tested the relationship between the range of OR, task type, and group. It was expected for both groups to automatically assume that the ready-made translation draft for the PE task would not need extensive research. Their lack of post-editing experience was anticipated to drive them to that assumption, especially considering the fact that the post-editing brief did not include the original guidelines about making sure the terminology was correct. The hypothesis was not corroborated in the course of the statistical analysis – there was no effect of either group or task on the range of OR. There was no significant difference when it comes to the range of consulted OR categories for the trainees and EFL students.

This result was in line with Whyatt's (2018: 70) finding that transfer of training occurs for consulting OR in translation-like tasks (the project researched paraphrase as well as translation) – the study showed that the trainees turned to OR significantly more than language students even in a paraphrasing task. There was no difference in the number of visits to OR between the trainees and language students in the translation task, which is most likely indicative of a need for confirmation before making a decision in the TT. For the third group investigated in the study – translation professionals – this tendency was not observed. They used OR significantly less by half. Another study that found a similar lack of difference was Daems et al. (2016: 121–130), who established that the types of resources consulted in the course of both post-editing and translation from scratch are comparable (Daems et al. 2016: 121–130). The finding in the present study concerns the range of OR categories, while Daems et al. researched types of resources, but these two observations ought to be remarked on in relation to each other.

According to Pavlović (2007: 141), group profiles and individual preferences for certain types of OR play an important role. It appears that either the groups researched in this study were homogenous enough for there to be no difference in the range of OR consulted or they both transferred these behaviours from translation to post-editing in this regard. Another explanation based on Pavlović's (2007) conclusions is that the wealth of OR in English in contrast to materials in languages of limited diffusion plays a role in the selection process during IS. It turned out in her study that in L2 translation (Croatian into English), participants accessed

a wider selection of OR than just bilingual dictionaries (Pavlović 2007: 138f). The problem of OR scarcity in languages of limited diffusion was also noted by Gough (2017) and Ollala-Soler (2018). Had the participants translated and post-edited into English, there might have been a wider selection of OR categories, but it would have prolonged the recording session even more if another directionality had been added to the list of independent variables.

It is interesting that the present study's result is not in line with Kuznik (2017: 236) who established that translators accessed a higher number of different resources than teachers. Expertise appears to be a key factor in this difference as participants of that study were professional translators. The years of experience as opposed to the beginning of training most likely would have made a difference for the present study as well.

Gough (2017: 204) called the range of resources *the variety of types of resources*. The results obtained when testing Hypothesis 3 echo her results to some degree. In her study, there were also participants who displayed outlier behaviour and did not access any resources at all. Although her categorisation was slightly different from this study, the range of resources accessed by her participants was 0–12. Participants of this study accessed 0–10 resources, with values always slightly higher for the translation tasks on average, albeit this difference turned out non-significant.

This result might be again partially explained by the fact that trainees were still at the beginning of their translation training. At that point in their training (first semester of the first year of the MA programme) they had already received instructions pertaining to researching terminology and evaluating the reliability of OR. The lack of significant difference between the groups could mean that for the non-trainees and trainees alike the presence of MT output does not inherently alter their instinct to broaden their research horizon to a wider range of resources. Completing the tasks in a laboratory setting could have also impacted their need to search multiple resources or look for new ones – their performance most likely reflects their preferred choice of OR for similar tasks. The fact that the difference between the groups' attitude towards MT was non-significant might have implications for the lack of significant difference in the range of resources as well. Alternatively, their lack of experience with specialised terminology from these domains and text types might have similarly narrowed their range of resources. Perhaps a case study of the least and the greatest number of resource categories accessed could provide a more fine-grained insight into the patterns of types of OR used. This macroscopic perspective was bound to miss individual variations within respective OR categories – not only

in terms of quantity, but also quality of interaction (i.e. depth of search, level of engagement, website stickiness, etc.).

The level of prejudice towards MT could have comparably influenced the participants' interaction with OR. The fact that there was no significant difference in the range of OR for both tasks and groups is thus potentially indicative of caution in terms of trusting MT. This was reflected in the post-task questionnaire open questions. Participants commented that they felt compelled to check the suggestions proposed by MT. Apparently, researching the specialised terminology for PE also came with accessing a comparable range of OR.

Perhaps a task or group effect might have been observed if the dependent variable had been computed differently. For instance, instead of focusing on categories only, a more fine-grained approach with a focus on the number of separate websites accessed per a given category could have provided a more nuanced insight into potential differences.

4.12.4. Hypothesis 4

Translation trainees are more effective in information seeking than non-trainees.

In Hypothesis 4, product data was the focus of the analyses. It was expected that the trainees would be more effective in information searching than non-trainees, irrespective of the task or text type. The assumption of the group influencing the dependent variable of *accuracy* was based on the expected effect of translation training. The statistical tests found that there are not significant differences between the groups regarding their effectiveness.

A follow-up analysis was conducted to further examine the relationship between *accuracy* and the percentage of rich points researched by a given participant (*researched %*). No significant correlations were found in either of the groups separately or when they were treated as one. Hypothesis 4 cannot be confirmed as no statistically significant difference between the groups in terms of *accuracy* was found in the data.

To the best of the author's knowledge, no other study investigated effectiveness of information searching in this way when relating it to the percentage of researched rich points. Studies that address quality focus on accuracy in general (cf. Korpal 2017), its relationship with resource types (Daems et al. 2016 and Raído 2014) or with temporal/cognitive effort (cf.

Kuznik 2017; Whyatt et al. 2021). Examining the relationship with *accuracy* and percentage of researched rich points was a novel approach to interpreting this construct. Perhaps a different operationalisation of *accuracy* could have yielded significant results for the comparison and then also for the correlations with *researched %*. Effectiveness is an interesting construct and perhaps examining its relationship with effort and *resource range* would shed light on potential differences between groups or possibly also tasks in this study.

There are other studies that correlated the use of OR with the quality of the TT and employed quality as an indirect operationalisation of acceptability, such as Kuznik (2017) from the PACTE Group. Kuznik (2017: 226) found no relationship between the number of resources and acceptability scores in direct translation (into L1), but a significant relationship was found between the number of OR used by translators and teachers with medium or high acceptability scores. Participants in the high and low acceptability thresholds used more different resources than those who placed in the medium acceptability tier (Kuznik 2017: 226). Interestingly, translators with low acceptability scores used the greatest number of resources. For the relationship between the temporal measure (total time taken on searches) and acceptability scores, a tendency for translators with low acceptability scores was observed: they spent the most time searching (Kuznik 2017: 229). These tendencies were not tested for significance. Thus, perhaps indeed examining the relationship between the variable *resource range* could have provided significant results. Also, Kuznik's (2017) conclusions suggest that applying thresholds into the accuracy scores could have allowed for the correlations to turn significant in the analysis carried out in the present study.

Another study that examined the relationship between quality and a process variable was Whyatt et al.'s (2021). They found a significant negative correlation between the use of OR and quality of the TT (Whyatt et al. 2021: 11). Here the quality was measured with the time the proof-readers spent on improving the TT. The more time was spent in OR during translation, the less the proof-readers needed to make the final product publishable, albeit the correlation was slightly weaker for L1 translation (Whyatt et al. 2021: 11ff). The present study neither examined the product quality in terms of time needed to proof-read, nor analysed professional translators as an experimental group, so these results are unfortunately not comparable. But perhaps there could be a correlation between a temporal or cognitive effort variable and *accuracy* in the present study.

Finally, Pokorn et al. (2020) remarked that participants in their study mostly consulted terminological problems when allowed to access OR instead of focusing on the entire text for

comprehension purposes, for instance. It would be interesting to find out if participants of this study selected the acceptable terminological solutions before finding the equivalent online and more closely examine the difference between the researched and unresearched translation choices. Thus, a comparison for the variable *researched %* by group membership, task or even text type would be interesting. Some terms were researched by all participants (e.g. *keyhole surgery*) within a group, while others were only researched by some (e.g. *hesalite crystal glass*). This variable was only used in the correlational analysis with *accuracy*, and it is unknown whether the groups, tasks, and texts differed in terms of the percentage of terms researched.

To sum up, no significant differences between groups in terms of terminological accuracy were found and what is more – also no significant relationship between terminological accuracy and online consultations of given terms. Further investigations of effectiveness of IS would benefit from examining other variables involved in IS. In particular, analysing temporal and cognitive effort related to OR to see whether more time searching in L1 translation means lower translation product quality or even noticeably better terminological choices (cf. Kuznik 2017; Pokorn et al. 2020; Whyatt et al. 2021) for other language combinations, groups, and/or text types.

4.12.5. Hypothesis 5

Translation trainees' and non-trainees' attitude towards MT correlates with the percentage of time in online resources when post-editing.

This hypothesis examined the questionnaire and process data regarding post-editing only. It was assumed that a correlation would exist between attitude scores and percentage of time spent in OR during that task.

As one possible result, it was expected that an increasingly negative attitude (a decrease in attitude scores) would be accompanied with an increase in time spent in OR – participants would be more suspicious of suggestions provided by MT and thus spend more time double checking them. However, the deceptive fluency of NMT output could relax the vigilance of participants, which was another potential result. As Koehn (2017: 93) stated, NMT fluency oftentimes lulls the user into a false sense of security in terms of accuracy. This was also posited

in reference to Daems et al. (2017: 264) who suggested that students are less vigilant in IS during PE tasks – they assume correctness of MT suggestions. The situation was like the interactions with TM, MT, and OR as described by Pym (2013: 495): a special kind of risk management in relation to trust or mistrust in the data. Ultimately, the hypothesis was not confirmed – the results of the correlational analysis revealed no significant relationship between attitude and percentage of time spent in OR when post-editing.

Additionally, follow-up analyses were conducted to investigate potential relationships between total task time and attitude towards MT. No significant correlation has been found which is inconsistent with Guerberof Arenas' (2013: 83) finding about negative attitude towards MT potentially boosting productivity. However, the participants in Guerberof Arenas' (2013) study were all professional translators. Perhaps again only expertise can be the factor that allows translators to channel their negativity into a productivity boost. Another possible explanation is that none of the participants were used to post-editing NMT output. And similarly to what Guerberof-Arenas (2013) also concluded in her study: so much has changed since her experiment was conducted in 2013 that the relationships between SMT and attitudes back then might not hold up for the current state-of-the art technological advancements in NMT.

It is important to note that the attitude scores were never strongly negative, the mean values pointing towards weak positive attitudes (little above 100 out of 250 for both groups). It is then especially interesting that in the EFL students group the attitude was not significantly different from the trainees' attitude. The trainees had some brief introductory info in their courses why MT is used in professional translation at all which may have softened their attitude for this tool. But then also, the value range was wider among the EFL students. Since all participants fell into the category of digital natives (their mean age was 23), it is safe to assume that their experiences and attitudes towards technology and the Internet in particular as a source of information – which includes MT – was not confounded by generational differences that much.

The relationship between performance – especially for IS – and attitude towards MT is a complex one and no discernible pattern was found in the data. It was anticipated that there might be more nuance to how the groups differ in terms of attitude and how it could be related to the amount of time spent in OR. Perhaps a productivity effect and diminished use of OR requires professional experience with post-editing.

4.12.6. Hypothesis 6

There is a correlation between perceived difficulty and temporal effort

In Hypothesis 6, the questionnaire and process data were correlated. Correlational analyses were calculated between the variable *perceived difficulty* and time spent in OR in various categories – temporal effort. *Perceived difficulty* was treated here as a subjective indicator of effort and was expected to positively correlate with time spent in OR. The hypothesis was partially confirmed – the subjective difficulty judgements positively correlate with select temporal effort categories with respect to groups, tasks, and texts. As expected, all significant correlations were positive.

The results show that the correlation exists across groups, tasks, and texts especially for aggregate categories like all OR combined or English and Polish websites – either combined or treated separately, as well as *Wikipedias* – again, either combined or separately as English and Polish.

Interestingly, *Google* measures did not correlate across all groups, tasks, and texts – either together or treated separately. Positive correlations for *Google* measures appeared for the group of trainees, as well as for the translation task and informative-medical text type. For the operative-technical text type it was only *Google Translate* use that correlated significantly.

The strongest correlations were moderate and those were found within the TR task for aggregate categories of English and Polish websites combined (TR task and informative-medical texts) as well as treated separately (TR task, for informative-medical texts only for Polish websites). Other moderate correlations were found within for all OR combined (T group and informative-medical texts). For the informative-medical text, correlations of the same effect size were found to be for concordancers, both *Wikipedias*, as well as Polish *Wikipedia* separately. This means that depending on group membership, task, as well as text type, the sense of ST difficulty is significantly reflected in time spent in certain types of resources.

The fact that there were more significant relationships found for English language categories (9) than Polish language categories (5) is not surprising. Polish language categories were not consulted as much – Polish language blogs, websites, language reference websites, etc. While there are many missing values for Polish language categories, English counterparts were accessed much consistently by most participants. This could be interpreted that their need to understand the ST corresponded with their perception of difficulty of that text and thus

triggered more time in these types of OR. Also, while there was a significant difference between the groups in terms of time spent in Polish language reference websites (Hypothesis 2), the correspondence with difficulty clearly manifested for the English categories. This might be explained by the fact that for all participants English was their L2 and despite their high proficiency (cf. high mean LexTALE scores corresponding to the C1 threshold).

It is, however, quite surprising that there is no significant correlation for bilingual dictionaries and *perceived difficulty*. When it comes to the bilingual resources, it was only concordancers that correlated significantly for all independent variables treated as one and for the TR as well as PE task, T group and operative-technical text type. It appears that participants use bilingual dictionaries consistently, regardless of their perception of the text difficulty. However, as the participants perceived the texts to have been more difficult, they spent more time using concordancers such as *Linguee* or *Glosbe*. Perhaps then translators turn to concordancers for more specialised texts, while all participants regardless of text or task type go for the familiarity and ease of bilingual dictionaries. Whyatt et al. (2021: 12) found that translation direction interacts with text type and this interaction influences the IS process. There exists a complex relationship between these three aspects. Technical texts in the study turned out to require more consultations of bilingual resources in general (forums, dictionaries, corpora). In the light of this, the fact that *perceived difficulty* significantly correlates with the use of concordancers is not that surprising.

Moorkens et al. (2015) investigated the relationship between effort expectations and actual temporal effort measurements and found a moderate positive correlation. The *perceived difficulty* in the present study was measured post-task as a reflection after a completed translation and post-editing task. It was because human predictions of post-editing effort do not correlate strongly with actual time needed to post-edit, according to Moorkens et al. (2015: unpaginated). Thus, the decision to ask participants to provide their perception of text difficulty rather than prediction of effort was expected to better correlate as regards actual temporal effort scores. It needs emphasising that Moorkens et al. (2015) did not focus on effort in IS, but cognitive and temporal measures related to the whole task. Vieira (2016) reported subjective and objective measures of effort to be strongly correlated for translation professionals, but no OR were allowed in his study, so it is impossible to compare this conclusion in any way, unfortunately.

Herbig et al. (2019: 111f) found that subjective perception of effort is more connected with stress and exhaustion, which was one of the reasons that the wording in the questionnaire

of this study instead of using the word "effort" put the words "easy" and "difficult" to tie the rating to the texts as perceived post-tasks by the same participants who completed both translation and post-editing. This choice of wording was deliberate so as not to prompt the participants to think of the amount of time it took them to complete the task, but about the texts in relation to their general experience of task completion.

4.12.7. Hypothesis 7

The final hypothesis explored the relationship between the subjective perception of difficulty and the range of consulted OR categories. When groups, tasks, and texts were treated as one, the correlational analysis showed the relationship to be statistically non-significant ($p > 0.05$). For the TR and PE tasks (when both groups and texts respectively were treated as one), the analysis also showed the relationship to be statistically non-significant ($p > 0.05$) as well. The same was revealed for the T and G groups, where both tasks and texts were treated respectively as one. These non-significant results reveal that the *resource range* neither increases or decreases consistently when difficulty ratings change and this was observed when it comes to group membership and task type.

The only significant correlation was revealed to exist for the operative-technical texts, regardless of the task type or group membership. This means that as the perceived difficulty scores of a text increased, all participants accessed more types of resources. The relationship between text type and IS was already commented on with reference to *perceived difficulty* and time spent in OR for Hypothesis 6. In line with this correlation, the range of consulted OR increases with the *perceived difficulty* for the operative-technical text type. Whyatt et al.'s (2021: 13) information behaviour in bidirectional translation model showed that translation into L1 relies more on OR in the stage of meaning construction than on a translator's internal resources – more support is needed to deconstruct the ST when translating into L1. To interpret the relationship in this significant correlation with reference to Whyatt et al. (2021), it appears that regardless of task type or group, perceived difficulty of a given text type needs a wide range of OR because it triggered more intense cognitive uncertainty at the stage of meaning construction. Finally, it also should be borne in mind that, as stated previously, personal preference in the use of OR is likely to influence patterns of IS behaviour. As far as the conclusions from this study are concerned, it is impossible to establish how much of the range

of resources was affected by personal preferences. It would be interesting to see how aware of such preferences participants are when asked directly.

4.12.8. Open questionnaire questions discussion

In the open questions in the post-task questionnaires, participants expressed a variety of opinions and impressions after they completed the tasks. The majority of participants (60%, *n* = 20) admitted that MT made the entire task easier and/or faster. What is more, 20% wrote that they were surprised at the decent or good quality of MT. Admittedly, some of them also expressed their conviction that with the presence of good MT they didn't need to verify all the terms in the output, while others explicitly admitted to being suspicious of suggestions and checking them online. 40% of them explicitly said they needed to check what MT provided them with.

What also appeared in the questionnaire was musings on the relationship between creativity and MT. One person shared that they preferred translating the product description without the aid of MT. 20% of participants admitted to being prompted by MT to pursue a certain solution or interpretation of ST meaning. One of them wished they could not see the suggestions, apparently finding themselves unable to think of alternative solutions once presented with a particular option. Both of these sentiments echo what Cadwell et al. (2018: 312) found as well – that MT can both curb creativity as well as kickstart translation ideas. It is especially a belief shared by professionals who valued their creativity in translation and thus are wary of MT's capacity to speed up the translation process (Daems et al. 2017: 21).

Apart from affecting creativity, when it comes to MT and its influence on decision-making there is also the possibility of decreasing one's sensitivity to noticing errors. Interestingly, none of the participants commented on MT making them more or less tolerant of errors, as opposed to what Guerberof Arenas (2013: 78) noticed in her study: her participants admitted MT is helpful in noticing errors. Perhaps such conclusions would have appeared in the questionnaire answers if the texts were longer and/or participants were required to perform post-editing long-term.

The main goal of this open question was to gauge possible awareness of MT influencing IS behaviour. 15% of all the participants thought MT did not affect their online consultations. One of those three people who admitted that also wrote that they were generally wary of MT

suggestions, so perhaps this could have been connected with an increased awareness of MT capabilities.

In terms of comments regarding general attitudes towards MT, only 20% of the participants expressed an outright negative attitude towards MT with their scores much below zero. What is very intriguing, two EFL students were negatively disposed towards MT, but having actually used it for post-editing, their attitudes evolved. After the task they admitted that MT was good and sped up the process. This is in line with Tatsumi (2010: 185) who reported flexibility among participants' attitude towards post-editing. Similarly, a small-scale study conducted by this author also established this flexibility and openness with reference to using MT as a translation aid by translation students (Witczak 2016a).

Two people (one trainee and one EFL student) regarded MT as a shortcut to dealing with complex texts. Their opinions suggested that it appeared to them as a substitute for expertise thus allowing them to bypass years of training and experience and supposedly successfully deal with specialised texts. It was expected that some of the participants will fall prey to the deceptive fluency of NMT and comment on it in some manner in the questionnaire, but it is interesting that ultimately only two did so. As indicated by Cadwell et al. (2018: 315), frequent use of MT in professional translation leads to awareness of unreliability of MT suggestions when terminology is not verified (e.g. by the client), so it was anticipated that first time use of MT might lead to such convictions. It is, after all, an opinion voiced by some professional translators on social media – as Läubli and Orrego Carmona (2017) found – that untrained use of MT for translation would result in false sense of competence. Such decreased vigilance was not found to be reflected in any way in terms of temporal effort differences for the two groups – the even less experienced EFL students did not spend significantly less time in OR than trainees did. As Hirci (2013: 162) pointed out, trainees tend to trust OR when searching for information, so shifting the responsibility further to MT seems only natural, if only for some participants.

To sum up, the impressions gathered in the questionnaire also in general echo Guerberof Arenas' (2013: 88) finding that translators had a very open-minded and practical attitude towards MT and post-editing. It is important to emphasise that the participants were all students, not professionals. This is why they approached the new task (post-editing) with a clear mindset, devoid of experience-based biases or habits. According to Moorkens and O'Brien (2015), students are more eager than professional translators to engage with new tasks. Furthermore,

despite their generally positive attitude, these student participants showed enough caution to have limited trust towards MT and still used OR throughout the PE task.

4.13. Study limitations and future avenues

The main limitations of the study are the laboratory setting which lowered the ecological validity and the relatively small sample (20 participants). However, this is a perennial struggle in the TPR studies, when taxing tasks such as translation or post-editing are performed in a laboratory setting, thus resulting in long task times (e.g. Daems et al. 2016, 2017; Gough 2017; Hvelplund 2017; Hvelplund and Dragsted 2018; H. Kruger 2016; Massey and Ehrensberger-Dow 2014; Moorkens and O'Brien 2015; Ortiz-Boix and Matamala 2015; PACTE group 2017; Paradowska 2015; Pavlović 2007, 2014; Whyatt 2012, 2018; Whyatt et al. 2021). The gain of controlling the experimental environment and variables is that it limits as many possible confounding variables as possible – the performance of the experimental device (computer and software) or distractors (breaks, accessing other websites than OR needed for translation, interruptions, etc.). However, as it has been stated in Chapter 2, while the framework of Situated Translation emphasises the need for ecological validity, its tenets are not mutually exclusive with laboratory-controlled settings. This study could not have been conducted on participant's own computers or in their homes not only due to the need to remove as many confounds as possible, but mostly because of the eye-tracker's setup. Recruitment of target groups is another issue that is reported to hinder obtaining a sufficient number of eligible participants, especially whenever it is professionals rather than students. This study targeted a group of translation trainees at a specific time in their training, which immediately limited the pool of possible participants to less than 20 people, out of which some did not want to participate, thus resulting in a total of 11 trainees that not only agreed to participate but also completed both experimental sessions. The issue of a small sample was compensated for by averaging out either tasks or groups or treating either tasks or groups as one, which increased number of analysed data points.

Another limitation is the fact that familiarity with texts types and terminology was not investigated and it is unclear whether or not it could have influenced the results. Studies found that expertise level as well as familiarity with a given domain affect the choice of search terms (Hsieh-Yee 1993) and those cognitive resources combined with the retrieved content online

lead to successful retrieval of relevant information (Spink and Saracevic 1997). As Hvelplund and Dragsted (2018) established, genre familiarity leads to automated behaviours in the translation process.

One of future avenues for IS in the translation and post-editing process could include whether text types factor into potential differences between groups and tasks. The effect of text type was not tested in this thesis for some of the hypotheses and it would be interesting to calculate follow-up tests with this effect in mind. Also, no correlations were calculated between perceived difficulty and temporal effort per text type. Whyatt et al. (2021: 8) investigated correlations between directionality, text type, and OR. They found positive correlations between the number of searches and the number of pauses (both longer than 10 s and 5 s) – with comparable coefficients in both directions (EN<>PL) and for both text types. In Raído's (2014: 175) study, on average more time was spent translating from scratch the more specialised text among the two in the study. It would be valuable to examine that dimension as well.

Other future investigations could involve comparing preferences for certain types of OR across groups. Zapata (2016: 146) remarked that concordancers are preferred over dictionaries or term banks – since this is impossible to establish from temporal data alone whether such an OR is preferable among participants of this study, it would be an interesting question to ask in future studies. It was established that perceived difficulty correlates with the use of concordancers, but the fact that increased temporal effort is related to subjective judgements of text difficulty does not account for actual personal preference for one type of OR or another.

Furthermore, measuring and comparing trust with attitude and other variables (*effort*, *resource range*, etc.) might add nuance to the results because as Cadwell et al. (2018: 315) argued, lack of trust could result in refusal to adopt MT in the future. Examining possible correlations with trust measurements as well as attitude scores could establish interesting relationships between these variables. Another direction worth pursuing is re-examining the operationalisation of attitude with ratings of statements on trust in MT. Further analyses regarding potential suspicions regarding usefulness of MT could involve testing for differences between groups and/or tasks where attitude is an independent variable on the basis of thresholds (very negative, negative, neutral, positive, very positive). Perhaps the level of wariness might cause an increase or decrease in average fixation duration or time spent in OR.

Yet another possible direction of examining the data for the use of OR is product quality. Raído (2014: 144) established that higher translation quality is connected with the increase in total task time and effort, thus emphasising that IS is an important part of the whole task. Whyatt

et al. (2021: 11) established a negative correlation between time spent in OR by translators and time spent on proof-readers to make the text into a publishable quality. In this thesis, the product data was analysed in a very limited capacity, focusing on terminological accuracy instead of an extensive quality evaluation. This was mainly because process measures were of main interest in the analyses.

Furthermore, as only few statistically significant differences were found in the process data between the two groups, perhaps a quantitative analysis of search strategies would have revealed more pronounced differences. Examining search strategies with the focus on query formulation, search depth, and use of operators, among others, would be an interesting future avenue. These issues were briefly provided in Section 2.3 as background information to emphasise the complexity of IS as a process, but analysing them in depth was beyond the scope of this thesis, both quantitatively as well as qualitatively.

Finally, it would be interesting to see possible longitudinal effects of training on the use of OR as measured by variables used in this study: *effort*, attitude, *resource range*, etc. This study compared trainees and non-trainees while the former group was still at the very early stages, so a longitudinal approach could have shed light on the trajectory of their Translator Competence development with the emphasis on instrumental and technological components and their self-concept as translators. This would allow to test what Paradowska (2015) found: whether after completing a course, students turn to more OR (types) than at the beginning of their training. Alternatively, recording the trainees at a later stage of their training could have produced more robust differences with the control group. However, the later it would have been (both in terms of their studies and the academic year), the smaller the chances of both recruiting enough participants and all of them completing the recording sessions. Hence, it is important to consider potential pedagogical implications of the study results from the perspective of the training stage these students were at – these are provided in the next section.

4.14. General discussion and pedagogical implications

This thesis regarded the process of translation and post-editing through the lens of Situated Translation in an attempt to gain a more complete picture of these two tasks when performed by trainees and EFL students. The focus was on the interplay between the NMT and OR scaffolding the process. Presently, the translation profession cannot be decoupled from

computers and online resources, therefore the digital as well as information literacy have become an essential part of it. Digital artefacts in the translation process are also in constant flux and a part of the life-long learning process involving adapting to newer and improved tools. The fact that the study was conducted on translation trainees and EFL students provides insight into the effect of early training and how to potentially improve it with reference to IS and the use of MT.

In this study, participants probably relied on their intuition in both studied aspects: use of OR and benefitting from an MT output – however, training is about honing these intuitive behaviours into skills. To do so, experience-based learning, self-concept, self-reflection and collaborative learning can be employed (cf. Beeby et al. 2011; Kiraly 2013; Pym 2013). All to prepare for future technologized job market and introduce the trainees into it as confident and tech-savvy participants with a clear sense of self as translators. To echo Vieira's (2020) conclusions regarding the skills of translators in the face of impending automation: negativity towards MT involves more nuance than just plain dislike of the tech and fear of being replaced. The dislike of MT is often modulated by actual shortcomings of the tech in question along with business practices creating room for non-translating professions to tackle tasks such as post-editing or transcreation. While Vieira's (2020) comments were based on predictions and observations pertaining to the UK translation market, perhaps his conclusions could have important implications for translators of languages of limited diffusion as well. Namely, doomsday scenarios regarding automation ought to be regarded with caution and translators' competences transcending the linguistic aspect of the process should be emphasised to raise the awareness of what translation entails. This is tied with the issue of both wider social awareness and how business parties perceive the work of translators.

Another important aspect of this technology/business tandem is whether or not translators are confident enough with their knowledge of MT, i.e. how it works as well as what its capabilities and limitations are to use it in their individual workflows where appropriate and how to approach MT when requested directly to do so by clients. The fact that the participants in this study did mention that MT could substitute expertise suggests that explaining capabilities and limitations of (N)MT to students – especially translation trainees – is essential. But gauging their convictions and biases first would be beneficial to better establish discussion points. According to the questionnaire answers in this study, the participants' negative attitudes were flexible and subject to change based on new information – actual post-editing experience. The possibilities and limitations of MT could also be a point in a more general conversation on

educating wider audiences about what technology can do for translators and what it can – probably – never substitute. Furthermore, the issue of trust towards MT was extensively covered by Cadwell et al. (2018) and as it turned out to be connected to the socioeconomic reality of the workplace, the conclusion was that better technological awareness is key towards benefitting from what MT has to offer.

Combining the above with the results of the study in this thesis, the conclusion is apparent that intuitive transfer of skills from translation is a good start towards future proficiency in accommodating MT as it continues to develop – in line with the life-long learning philosophy. Incorporating post-editing into translation training has become a standard practice, either as a stand-alone course or a part of CAT classes. For instance, within the EMT framework (2017), post-editing is explicitly included on the level of strategic, methodological, and thematic competence and students are expected to know how to "[a]pply post-editing to MT output using the appropriate post-editing levels and techniques according to the quality and productivity objectives, and recognise the importance of data ownership and data security issues." Perhaps as the place of post-editing in the translator training programmes appears to be fixed, the way it is taught is better considered as Pym (2013: 497) suggested already in 2013, i.e. "we should be envisaging a general pedagogy, the main traits of which must start from the reasons why a specific course on TM/ MT may not be required." What it means is that standalone post-editing courses as suggested – a long time ago – by O'Brien (2002) may not necessarily be the best solution in translator training. However, a stand-alone course that would combine post-editing along with all of the elements proposed by O'Brien (programming, history of MT, controlled language) with teaching IS skills would be a different matter entirely. As Pym (2013: 497) pointed out, including post-editing tasks into each practical translation course would be difficult to achieve as a lot of teachers would see MT (and TM, which is also part of his argument) as a distractor from teaching translation from scratch. Therefore, designing a course that relies on text domain/genre unfamiliarity and targeting the development of research skills along with using MT as a tool/resource would be most beneficial, especially in illustrating the need to verify MT suggestions while making the most of what it has to offer – a draft of the TT.

When it comes to the IS skills, a detailed investigation of information behaviours has not been the aim of this project. Hence, conclusions regarding the qualitative aspect of IS cannot be drawn based on the analyses, in particular in terms of reliability of accessed OR and applied search strategies. However, this project can provide a meaningful contribution into the area of

IS and its place in the TC models by further emphasising how much effort the use of OR adds up to the entire process of translation and post-editing. The fact that almost no significant differences have been found between groups in terms of IS effort (both cognitive and temporal) suggests that early translation training combined with some degree of translation awareness is not enough to differentiate trainees from non-trainees.

Without the comparison with professionals it is not possible to establish whether the amount of effort the participants put into the tasks in IS was excessive or inefficient in any way or whether applying expert search strategies would have decreased the effort for the researched groups. Their accuracy scores varied which is indicative of room for improvement typical for novices and signals further investigation needed into the OR they consulted. But most importantly, the results can be used to further emphasise the importance of the amount of shadow work put into the translation and post-editing process. Gough (2019: 354) argued that translation-oriented IS can be regarded as a spectrum of behaviours, rather than a single recipe for finding certain types of information. OR are constantly changing and improving – while some are unlikely to perish from the Internet landscape in the near future (e.g. such popular ones like *Wikipedia* or concordancers), it is important to regard them as a dynamic system and account for diversity of information-searching behaviours among translators. According to Gough (2019: 354), "Many changes driven by the developments in information retrieval, AI and related areas will shape the information behaviour of future translators" which emphasises the importance of adaptability combined with awareness of the influence of MT on the process of IS.

Gough (2019: 349f) enumerates the four types of OR users (the Dictionary Enthusiast, the Mixed Type, the Parallel Text Fan, the MT Adopter) and patterns of research distribution pertaining to translation stages (e.g. the front-loaded research in which the translator does the bulk of research in the orientation phase). She suggests that self-reflection and self-awareness with reference to one's own research preferences might help with exploration of alternative ways to seek information (Gough 2019: 354). This could be achieved through screen-recorded and peer-reviewed translation exercises on each other's information behaviours, focusing on OR types, research direction and strategies, perhaps also supplementing them with TAPs. Such an exercise could be then discussed in class with demonstrations of other examples of IS strategies (Gough 2019: 354). The awareness of IS habits is, according to Gough (2019: 354), crucial in order to be able to efficiently adapt to emerging workflows (like

collaborative/concurrent translation modes when other translators are involved in the process for different parts of the workflow).

Gough's (2019) suggested IS-oriented exercise is a perfect starting point towards developing IS self-awareness. A follow-up exercise could focus on learning how to estimate and anticipate the amount of effort to be put into translation and post-editing based on IS rich points. Already aware of the patterns of IS behaviour at their disposal and their own preferences, trainees could start this exercise with a pre-translation/pre-post-editing task like the one suggested by Whyatt et al. (2021: 14). The pre-translation task in that study involved analysing a specialised text only focusing on rich points and consulting OR with only these rich points in mind. The exercise proposed here has two main goals: to calculate the time needed to research rich points and reflect on the types of OR accessed in relation to that time. Similarly to Whyatt et al. (2021), the exercise consists in researching the self-identified rich points, but would additionally be supplemented with screen-recordings of the entire process to calculate the time spent in OR and see the types of OR used. While students can employ different patterns of IS behaviour for this task, it requires them all to adopt a front-loaded research approach which happens in the orientation phase (Gough 2019: 351) as the task does not involve any translation or post-editing. Having recorded the IS process for a given text (be it for translation or post-editing), students then also peer-review their recordings and discuss their decisions as well as the time needed to complete the task. Depending on the task (post-editing vs. translation) and text type, different OR patterns and time would be expected. Also, in the case of languages of low diffusion – many of which are also languages of low OR availability – it would be useful to show students the disparity between OR availability for L2 translation into English and translation into their L1 and how that affects the time needed to consult OR/the types of OR accessed. This thesis explored L1 translation and post-editing only, but based on the findings by Whyatt et al. (2021), it would certainly be very beneficial to incorporate L2 translation into such an exercise as this type of translation is frequent among translators working with languages of low diffusion. Following the peer-review, a class discussion compares and contrasts the observations, conclusions, and doubts that the students had regarding their work and choices – all supplemented with feedback from the teacher. For the purpose of class discussion, students could prepare a short presentation of their own IS process with comments on their choices of rich points, OR types, and search strategies as well as conclusions regarding time spent on the task. Their peer reviewers could then briefly comment on their main discussion conclusions. Afterwards, other classmates comment on the presented strategies and ask questions. The

format of the presentation combined with the peer-review report allows the teacher to provide each student pair with individual feedback beforehand (for instance via an e-learning platform online feedback) as well as ensure structure of the class discussion. Finally, having received feedback from peers and the teacher in class, each student then proceeds with full translation or post-editing of the text. The complete target text could then be graded by the teacher with reference to how well the student integrated the discussed solutions from OR into the final product, combined with other grammar-, style-, and spelling-related grading criteria. The texts selected for different tasks could come from a variety of domains, depending on the aim of the course: either only specialised texts or a mixture of general purpose and specialised. Of course, it could also be the minimum of two texts intended for two tasks – translation and post-editing. The number of text types would determine the scope of the exercise and it could be adapted depending on time constraints and nature of the course. Using more than one text type for this exercise could allow recognising the different scope of OR types needed to satisfy the information needs of the ST. Pakkala-Weckström (2015: 164f) reported the genre of the ST to play a key role in the IS process – in the exercises her students completed, an EU text resulted in consulting mostly EU resources, while cooking recipes required an increase in the use of parallel texts.

The proposed pre-translation/pre-post-editing IS exercise could then include the following five stages:

1) Text translation/post-editing with screen recording and TAP to determine the preferred pattern of OR use (Gough's exercise described above);
2) Pre-translation and/or pre-post-editing task focused on researching self-identified rich points;
3) Peer-review of the recording in pairs: types of OR and time spent there as the focus;
4) Class discussion based on presentation of the peer-review exercise;
5) Individual translation and/or post-editing of the text(s) based on the discussion and teacher feedback.

Depending on the time constraints and group size, this exercise could be a short undertaking or a longer project. Before the first step, it would be beneficial to adopt Pakkala-Weckström's (2015: 155) approach to provide a background for data mining. This would allow the teacher to introduce the students to reliable OR, provide information about how to avoid suspicious sources, encourage to use parallel texts, and warn against being too sure about what does not require double-checking (Pakkala-Weckström 2015: 155). In line with Paradowska's

(2015) approach, it would also be most beneficial to instruct students on expert search techniques, including query formulation, as well as including/excluding phrases, among other things.

Besides the obvious purpose of this exercise to foster self-improvement and diversification of IS strategies and approaches, the exercise could also facilitate the development of skills regarding estimating time and price of individual assignments. Based on how well a student can estimate the time needed to complete the whole task in terms of producing the TT along with how much it takes to research individual rich points, students can be better prepared to negotiate rates and deadlines without running into exploitative assignments. Awareness of the degree to which MT can assist the IS process is also crucial in this regard. The experiential nature of such an exercise is in line with Kiraly's (2013) conception of emergent nature of translation competence. It is also in accordance with Piotrowska's (2015) argument that "the ultimate aim of translator education, ideally, is employability and preparing qualified graduates for the market" and this approach provides a seamless shift from the classroom into hypothetical scenarios echoing workplace conditions and challenges.

To sum up, the results and conclusions from this thesis could be regarded as a prelude to more studies on Human-Computer Interaction or, even more narrowly, Translator-Information Interaction in TPR with a holistic approach towards the process, i.e. treating it as a dance of agency with two of the main artefacts of modern translation: MT and OR.

Conclusion

This thesis explored information searching in the process of translation and post-editing and the results offer some insight into the use of OR in these task types among translation trainees in relation to EFL students. The study was designed with as much ecological validity as possible in a controlled laboratory experiment, relying of the tenets of Situated Translation (Risku 2010) which accounts for the consequences of the mind "leaking" into the environment – mainly technological in this context as seen in the process of IS and MT post-editing. The process is scaffolded by artefacts: OR and MT along with all other applications and devices (e.g. text processor) used in the process. The interaction of humans and technology was also considered in this thesis from the viewpoint of agency theory as suggested by Olohan (2011) after Pickering (2008): the translator's agency interacting with the material agency of artefacts used in the process and interconnected with their attitude towards the artefacts. The study is not without its limitations, but the results are an important contribution to the general understanding of Translator-Information Interaction from the viewpoint of translation novices and thus has valuable implications for translator training. Bearing in mind the limited scope of this project – the main focus being on effort, attitude, and accuracy – it is still a solid starting point towards developing experience-based exercises. To quote Pym (2013: 497): "In an ideal world, fully completed empirical research should tell us what we need to teach, and then we start teaching. In the real world, we have to teach right now, surrounded by technologies and pieces of knowledge that are all in flux." Therefore, surrounded by the constantly changing MT and OR landscape, even small-scale studies like this one can point towards more streamlined approaches to teaching.

Abstract

While translation has always required the ability to find information, currently this process has moved almost entirely into the digital realm. The universal revolution in translation, which happened many years ago, has transformed the profession into something resembling piloting an airplane because of the numerous tools to aid the process and help find information (Gouadec 2007: 263). Information mining and the use of other tools, such as machine translation, has become fixed points in translation curricula, but there remains a scarcity of research into both of these aspects when related to translation trainees.

In line with the translation process research paradigm, this thesis is an attempt to bridge this gap in research and to discuss information searching in the process of translation and post-editing. The aim of this project is to investigate translation trainees and EFL students as they interact with machine translation and online resources during translation and post-editing tasks for two text types (operative-technical and informative-medical, cf. Reiss 1976). The first objective of the thesis is to examine whether both groups put more effort into information searching when translating than when post-editing. Two indicators of effort have been used to test this hypothesis: time spent in applications (temporal effort) and average fixation duration (cognitive effort). The results show that the task type significantly influences the amount of temporal effort put into the use of online resources – both on the global level of all resource categories considered together and for some of them considered separately. No such effect has been found for the cognitive effort indicators. The second hypothesis in the study posits that translation trainees exert more temporal and cognitive effort in both translation and post-editing than EFL students. Again, the results show that this can only be partially confirmed. Significant differences exist only for temporal effort variables: the time spent on *Wikipedia* and language reference websites (like the Polish language advice centre, *Poradnia językowa PWN*). In both cases trainees spent more time consulting these resources. The interaction of the group and task effect was found in the use of monolingual dictionaries and it turns out that EFL students put more effort into consulting them. The third hypothesis focuses on the range of consulted online resources in relation to task type and group membership. Contrary to expectations, there is no effect of either group or task on the range of consulted resources. For the fourth hypothesis, accuracy in translating source text rich points is examined. Contrary to the expected group effect on accuracy scores, there is no statistically significant difference between the groups in

terms of how accurate they were. There is also no significant correlation between the accuracy of translations and the percentage of rich points (i.e. focal words or phrases) researched by a participant online. The fifth hypothesis concerns the relationship between the attitude towards machine translation and the percentage of time spent in online resources in relation to the whole task time during post-editing – the results show there is no statistically significant correlation between these variables, even for a follow-up correlational analysis between total task time and attitude scores. For the sixth hypothesis, an indicator of perceived effort is correlated with time spent in various online resource categories. The results reveal positive correlations with select temporal effort categories with reference to groups, tasks and texts as well as for each of these variables separately. For the last hypothesis, the correlation between the perceived effort indicator and the range of consulted online resources is examined. The results show a significant positive correlation only for one of the researched text types, i.e. a product description (operative-technical) – regardless of group membership or task type performed.

The results indicate that the relationship between effort, accuracy, and attitude in information searching during translation and post-editing is intensely nuanced. The findings of this study may be particularly valuable for translation trainers and translation process researchers. Although this project is limited in scope, it might provide a prelude into more extensive and focused studies of information searching in relation to translation training and translator competence development – and how machine translation influences the translation process as well. Examining the information searching process in translation students and incorporating self-reflection into translation pedagogy is likely to be beneficial for training more self-aware professionals, ready to commence the journey of life-long learning as translators.

Streszczenie

Tłumaczenie zawsze wymagało umiejętności poszukiwania informacji, jednak obecnie ten proces niemalże w całości ma miejsce w cyfrowej rzeczywistości. Powszechna rewolucja w tłumaczeniu – która miała miejsce już wiele lat temu – sprawiła, że zawód tłumacza przypomina teraz pilotowanie samolotu za pomocą licznych narzędzi wspomagających tłumaczenie i poszukiwanie informacji (Gouadec 2007: 263). Pozyskiwanie informacji i korzystanie z innych narzędzi, takich jak tłumaczenie maszynowe, to stały element programów nauczania tłumaczenia, jednak nadal niewiele jest badań dotyczących obu tych aspektów w odniesieniu do osób studiujących tłumaczenie.

Niniejsza rozprawa ma na celu zmniejszenie istniejącej luki w badaniach oraz zgłębienie poszukiwania informacji w procesie tłumaczenia i post-edycji zgodnie z nurtem badań nad procesem przekładu. Celem niniejszego projektu jest sprawdzenie, jak osoby studiujące tłumaczenie i filologię angielską korzystają z tłumaczenia maszynowego i źródeł internetowych podczas tłumaczenia oraz post-edycji dwóch typów tekstu (operatywno-techniczny i informacyjno-medyczny, por. Reiss 1976). Pierwszym z celów rozprawy jest sprawdzenie, czy obie grupy wkładają więcej wysiłku w poszukiwanie informacji, gdy tłumaczą od podstaw czy kiedy post-edytują tłumaczenie maszynowe. Do sprawdzenia tej hipotezy wybrano dwa wskaźniki wysiłku: czas spędzony w aplikacjach (wysiłek czasowy) oraz średnia długość fiksacji (wysiłek kognitywny). Wyniki pokazują, że rodzaj zadania istotnie wpływa na ilość wysiłku czasowego w korzystaniu ze źródeł internetowych – zarówno na poziomie globalnym wszystkich źródeł razem, jak i dla niektórych z nich osobno. Efektu typu zadania nie wykazały porównania dla wskaźników wysiłku kognitywnego. Druga hipoteza w niniejszej rozprawie zakłada, że studiujący tłumaczenie włożą więcej wysiłku czasowego i kognitywnego zarówno w tłumaczenie, jak i w post-edycję niż osoby z filologii angielskiej. Podobnie jak dla poprzedniej hipotezy, znaleziono jedynie częściowe potwierdzenie. Statystycznie istotne różnice wykazały tylko zmienne związane z wysiłkiem czasowym: czas spędzony na *Wikipedii* oraz stronach związanych z poradami językowymi (np. *Poradnia językowa PWN*). W obu przypadkach osoby ze specjalizacji tłumaczeniowej spędziły więcej czasu używając tych źródeł. Dla czasu spędzonego w słownikach jednojęzycznych wykazano interakcję grupy oraz typu zadania i okazało się, że osoby z filologii angielskiej wkładają więcej wysiłku w korzystanie z nich. Trzecia hipoteza skupiła się na zakresie wykorzystanych źródeł

internetowych w stosunku do typu zadania i przynależności do grupy. Wbrew oczekiwaniom nie wykazano ani efektu grupy, ani typu zadania na zakres źródeł internetowych. Czwarta hipoteza jest związana z poprawnością w tłumaczeniu wybranych słów lub fraz (tzw. *rich points*). Również wbrew oczekiwanemu wpływowi grupy na wyniki poprawności, analiza statystyczna nie wykazała istotnych statystycznie różnic między grupami w ramach poprawności tłumaczonych terminów. Nie wykazano również istotnej korelacji między poprawnością tłumaczeń a procentem sprawdzonych w Internecie wybranych słów lub fraz. Piąta hipoteza skupia się na związku stosunku do tłumaczenia maszynowego zadeklarowanym przez badanych z procentowo oszacowanym czasem spędzonym w źródłach internetowych w odniesieniu do całkowitego czasu post-edycji. Wyniki nie wykazały istnienia istotnych statystycznie korelacji, nawet po wykonaniu dodatkowych analiz na samym czasie trwania post-edycji i wynikach poprawności tłumaczenia. Dla szóstej hipotezy skorelowano wskaźnik subiektywnego postrzegania wysiłku z czasem spędzonym w różnych kategoriach źródeł internetowych. Wykazano istotne dodatnie korelacje w niektórych kategoriach, biorąc pod uwagę obie grupy, oba zadania i oba typy tekstów oraz dla każdej z tych zmiennych osobno. Ostatnia hipoteza jest związana z subiektywnie postrzeganym wysiłkiem oraz zakresem źródeł internetowych. Okazało się, że istnieje istotna dodatnia korelacja tylko dla jednego typu tekstów, tj. opisu produktu (typ operatywno-techniczny) – gdy grupy oraz typy zadania były analizowane wspólnie.

Wyniki niniejszej rozprawy wykazały, że relacje między wysiłkiem, poprawnością oraz stosunkiem do tłumaczenia maszynowego podczas poszukiwania informacji w procesie tłumaczenia i post-edycji są bardzo złożone. Zależności, które tu opisano mogą się okazać szczególnie przydatne w nauczaniu tłumaczenia oraz badaniach nad procesem przekładu. Mimo niewielkiej skali niniejszego projektu, może on stanowić preludium dla obszerniejszych i ściślej ukierunkowanych badań nad poszukiwaniem informacji w nauczeniu tłumaczenia i rozwijaniu kompetencji tłumaczy – oraz jak tłumaczenie maszynowe wpływa na proces tłumaczenia. Badanie procesu poszukiwania informacji u osób studiujących tłumaczenie i włączanie autorefleksji w dydaktykę przekładu może się okazać pomocne w kształceniu bardziej świadomych zawodowych tłumaczy oraz tłumaczek, przygotowanych do samorozwoju poprzez kształcenie ustawiczne.

References

Allen, Jeffrey H. 2003. “Post-editing”, in: Harold Somers (ed.), *Computers and Translation: A translator’s guide*. Benjamins Translation Library, Amsterdam: John Benjamins Publishing Company, 297–317. (https://benjamins.com/catalog/btl.35.19all).

Alonso, Elisa. 2015. “Analysing the use and perception of Wikipedia in the professional context of translation”, *JoSTrans: The Journal of Specialised Translation* 23: 89–116. (http://www.jostrans.org/issue23/art_alonso.php) (date of access: 3 Jan. 2020).

Alves, Fabio and Tânia Liparini Campos. 2009. “Translation technology in time: Investigating the impact of translation memory systems and time pressure on types of internal and external support”, in: Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees (eds.), *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen studies in language, Copenhagen: Samfundslitteratur, 191–218.

Aula, Anne. 2005. *Studying user strategies and characteristics for developing web search interfaces*. Tampere: Tampere University Press.

Austermühl, Frank. 2014. *Electronic tools for translators*. London; New York: Routledge.

Aziz, Wilker, Sheila Castilho and Specia, Lucia. 2012. “PET: a Tool for Post-editing and Assessing Machine Translation”, in: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*. Istanbul, Turkey: European Language Resources Association (ELRA), 3982-3987. (https://www.aclweb.org/anthology/L12-1000) (date of access; 3 Jan. 2020).

Bar-Hillel, Yehoshua. 1960. “The Present Status of Automatic Translation of Languages”, in: Franz L. Alt (ed.), *Advances in Computers*. Jerusalem, Israel: Elsevier, 91-163. (http://linkinghub.elsevier.com/retrieve/pii/S0065245808606075) (date of access: 31 May. 2018).

Marco Baroni, Pavel Rychlý, Adam Kilgarriff, Jan Pomikálek. 6-9. “WebBootCaT: a Web Tool for Instant Corpora”, in: Cristina Onesti Elisa Corino Carla Marello (ed.), *Proceedings of the 12th EURALEX International Congress*. Torino, Italy: Edizioni dell’Orso, 123–131.

Battelle, John. 2005. *The search: how Google and its rivals rewrote the rules of business and transformed our culture*. New York: Portfolio.

Beeby, Allison, Mónica Fernández, Olivia Fox, Amparo Albir, Inna Kozlova, Anna Kuznik, Wilhelm Neunzig, Patricia Rodríguez, Lupe Romero and Stefanie Wimmer. 2009. "Results of the validation of the PACTE translation competence model: Acceptability and decision making", *Across Languages and Cultures* 10, 2: 207–230.

Beeby, Allison, Mònica Fernández, Olivia Fox, A Hurtado Albir, Anna Kuznik, Wilhelm Neunzig, Patricia Rodríguez, Lupe Romero and Stefanie Wimmer. 2011. "Results of the validation of the PACTE Translation Competence model: Translation project and dynamic translation index", in: Sharon O'Brien (ed.), *Cognitive explorations of translation*. Continuum studies in translation, London; New York, NY: Continuum, 30–53.

Behnke, Maximiliana. 2017. *A small python script that extracts frequency of character bigrams in subtlex corpora.: fiqas/subtlex-bigram*. (https://github.com/fiqas/subtlex-bigram) (date of access: 12 Aug. 2019).

Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo and Marcello Federico. 2016. "Neural versus Phrase-Based Machine Translation Quality: a Case Study", in: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, 257-267. (https://www.aclweb.org/anthology/D16-1025) (date of access: 3 Jan. 2020).

Bermúdez, José Luis. 2014. *Cognitive science: an introduction to the science of the mind*. (Second edition.) Cambridge: Cambridge University Press.

Boitet, Christian, Herve Blanchon, Mark Seligman and Valerie Bellynck. 2010. "MT on and for the Web", in: IEEE, 1–10. (http://ieeexplore.ieee.org/document/5587865/) (date of access: 7 Feb. 2018).

Borja Albi, Anabel. 2008. "Corpora for translators in Spain. The CDJ-GITRAD corpus and the GENTT project", in: Gunilla M. Anderman and Margaret Rogers (eds.), *Incorporating corpora: the linguist and the translator*. Translating Europe, Clevedon; Buffalo: Multilingual Matters, 243–265.

Burack, Jacob A., Tamara Dawkins, Jillian Stewart, Heidi Flores, Grace Iarocci and Natalie Russo. 2012. "Chapter Six - 'The Mysterious Myth of Attention Deficit…' Revisited: A Discussion of How the Developmental Approach is Transforming the Understanding of Intellectual Disability", in: Robert M. Hodapp (ed.), *International Review of*

Research in Developmental Disabilities. International Review of Research in Developmental Disabilities, Academic Press, 147-177. (https://www.sciencedirect.com/science/article/pii/B9780123942845000061) (date of access: 22 Jan. 2021).

Cadwell, Patrick, Sharon O'Brien and Carlos S. C. Teixeira. 2018. "Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators", *Perspectives* 26, 3: 301-321.

Carl, Michael. 2012. "Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research", in: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey: European Language Resources Association (ELRA), 4108-4112. (http://www.lrec-conf.org/proceedings/lrec2012/pdf/614_Paper.pdf) (date of access: 5 May 2016).

Carl, Michael, Barbara Dragsted and Arnt Lykke Jakobsen. 2011. "A taxonomy of human translation styles", *Translation Journal* 16, 2. (http://translationjournal.net/journal/56taxonomy.htm) (date of access: 14 Oct. 2015).

Carl, Michael, Moritz Schaeffer and Srinivas Bangalore. 2016. "The CRITT Translation Process Research Database", in: Michael Carl, Srinivas Bangalore and Moritz Schaeffer (eds.), *New Directions in Empirical Translation Process Research*. Cham: Springer International Publishing, 13–54.

Carl, Michael and Moritz J. Schaeffer. 2017. "Models of the Translation Process", in: John W. Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 50–70.

Carpineto, Claudio and Giovanni Romano. 2012. "A Survey of Automatic Query Expansion in Information Retrieval", *ACM Computing Surveys* 44, 1: 1–50.

Chang, Chieh Ying. 2009. *Testing applicability of eye-tracking and fMRI to translation and interpreting studies: An investigation into directionality*. [Ph.D. Dissertation]. London: Imperial College London. (http://hdl.handle.net/10044/1/11409) (date of access: 8 Aug. 2019).

Chitu, Alex. 2007. "Google Switches to Its Own Translation System", *Google Operating System*. (https://googlesystem.blogspot.com/2007/10/google-translate-switches-to-googles.html) (date of access: 5 Feb. 2018).

Choi, Charles. 2002. “Multilingual Machines”, *Scientific American*. (https://www.scientificamerican.com/article/multilingual-machines/) (date of access: 5 Feb. 2018).

Clark, Herbert H. 1973. “The language-as-fixed-effect fallacy: A critique of language statistics in psychological research”, *Journal of verbal learning and verbal behavior* 12, 4: 335–359.

Common Sense Advisory. 2016. “Global Market Research Firm Common Sense Advisory Finds Post-edited Machine Translation (PEMT) Among Fastest-growing Segments of the Language Industry”, *CSA Research*. (https://csa-research.com/More/Media/Press-Releases/ArticleID/16/Global-Market-Research-Firm-Common-Sense-Advisory-Finds-Post-edited-Machine-Translation-PEMT-Among-Fastest-growing-Segments-of-the-Language-Industry) (date of access: 2 May 2018).

Cornellà, Alfons. 2000. “Cómo sobrevivir a la infoxicación”, in: *Trascripción de la conferencia del acto de entrega de títulos de los programas de Formación de Posgrado del año académico 1999-2000*. (https://www.academia.edu/34226090/C%C3%B3mo_sobrevivir_a_la_infoxicaci%C3%B3n) (date of access: 3 Jan. 2020).

Daems, Joke, Michael Carl, Sonia Vandepitte, Robert Hartsuiker and Lieve Macken. 2016. “The Effectiveness of Consulting External Resources During Translation and Post-editing of General Text Types”, in: Michael Carl, Srinivas Bangalore and Moritz Schaeffer (eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. Cham: Springer International Publishing, 111–133.

Daems, Joke, Sonia Vandepitte, Robert J. Hartsuiker and Lieve Macken. 2017. “Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators”, *Meta: Journal des traducteurs* 62, 2: 245.

Denkowski, Michael and Alon Lavie. 2012. “TransCenter: Web-based translation research suite”, in: *AMTA 2012 Workshop on Post-Editing Technology and Practice Demo Session*, 2012. (https://www.mjdenkowski.com/files/transcenter-amta2012.pdf) (date of access: 3 Jan. 2020).

Densmer, Lee. 2014. “Light and Full MT Post-Editing Explained” (https://info.moravia.com/blog/bid/353532/Light-and-Full-MT-Post-Editing-Explained) (date of access: 3 May 2018).

DePalma, Don. 2013. “Post-editing in practice”, *tcworld.info - translation and localization*. (http://www.tcworld.info/e-magazine/translation-and-localization/article/post-editing-in-practice/) (date of access: 3 May 2018).

Désilets, Alain. 2010. “Translation and Crowd Sourcing: Opportunity or Heresy?” (https://www.slideshare.net/alaindesilets/2010-06u-marylandcrowdsourcingworkshopv2010061610h44) (date of access: 5 Sep. 2018).

Désilets, Alain, Christiane Melançon, Geneviève Patenaude and Louise Brunette. 2009. “How translators use tools and resources to resolve translation problems: An ethnographic study” (http://mt-archive.info/MTS-2009-Desilets-2.pdf) (date of access: 5 Sep. 2018).

Doherty, Stephen. 2016. “Investigating cognitive processing in L1 and L2 multimedia viewers: Triangulating electroencephalography, eye tracking, and psychometric data in diverse samples”, in: *Tra&Co Symposium 2016*, 27–29 April 2016, Gemersheim, Germany.

Doherty, Stephen and Federico Gaspari. 2013. *Effective Post-Editing in Human & Machine Translation Workflows: Critical Knowledge & Techniques*. (Video of conference presentation slides with audio) (https://vimeo.com/81116667).

Domas White, Marilyn, Miriam Matteson and Eileen G. Abels. 2008. “Beyond dictionaries: Understanding information behavior of professional translators”, *Journal of Documentation* 64, 4: 576–601.

Donald, Merlin and Lars Andreassen. 2007. “Consciousness and Governance: From embodiment to enculturation – an interview”, *Cognitive Semiotics* 0: 68–83.

Dragsted, Barbara. 2006. “Segmentation in translation: Differences across levels of expertise and difficulty”, *Target* 17, 1: 49–70.

Dragsted, Barbara and Inge Gorm Hansen. 2008. “Comprehension and Production in Translation: A Pilot Study on Segmentation and the Coordination of Reading and Writing Processes”, in: Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees (eds.), *Looking at eyes: eye-tracking studies of reading and translation processing*. Copenhagen studies in language, Frederiksberg: Samfundslitteratur, 9–29.

Duchowski, Andrew. 2007. *Eye Tracking Methodology*. London: Springer London.

Duval, Cédric. 2012. “L’impact du Web en 4 questions”, *La Recherche* 467: 46–50.

Ehrensberger-Dow, Maureen and Sharon O’Brien. 2015. “Ergonomics of the Translation Workplace: Potential for Cognitive Friction”, *Translation Spaces* 4, 1: 98–118.

EMT Board. 2017. “EMT Competence Framework” (https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf) (date of access: 18 Feb. 2021).

EMT Expert Group. 2009. “Competences for professional translators, experts in multilingual and multimedia communication” (http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf) (date of access: 29 Apr. 2016).

Fidel, Raya. 2012. *Human information interaction: an ecological approach to information behavior*. Cambridge, Mass: MIT Press.

Flanagan, Marian and Tina Paulsen Christensen. 2014. “Testing post-editing guidelines: how translation trainees interpret them and how to tailor them for translator training purposes”, *The Interpreter and Translator Trainer* 8, 2: 257–275.

Flesch, Rudolph. 1948. “A new readability yardstick.”, *Journal of Applied Psychology* 32, 3: 221.

Furnas, G. W., T. K. Landauer, L. M. Gomez and S. T. Dumais. 1987. “The vocabulary problem in human-system communication”, *Communications of the ACM* 30, 11: 964–971.

GALA Global. 2016. TMX 1.4b Specification. (https://www.gala-global.org/tmx-14b) (date of access: 3 Jun. 2018).

Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves and Andy Way. 2014. “Perception vs reality: Measuring machine translation post-editing productivity”, in: *Third Workshop on Post-Editing Technology and Practice*, 60. (http://amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf) (date of access: 21 Nov. 2016).

Gerloff, Pamela Ann. 1988. “From French to English: A look at the translation process in students, bilinguals, and professional translators”. [Unpublished Ph.D. Thesis. Available on microfilm: University Microfilms International, Ann Arbor].

Gile, Daniel. 2004. “Integrated problem and decision reporting as a translator training tool”, *The Journal of Specialised Translation* 2, 2. (https://jostrans.org/issue02/art_gile.php) (date of access: 31 Aug. 2018).

Göpferich, Susanne. 2009. “Towards a model of translation competence and its acquisition: the longitudinal study TransComp”, in: Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees (eds.), *Behind the Mind: Methods, Models and Results in Translation Process Research*. Copenhagen studies in language, Copenhagen: Samfundslitteratur, 11–37.

Gouadec, Daniel. 2007. *Translation as a profession*. Amsterdam; Philadelphia: J. Benjamins Pub. Co.

Gough, Joanna. 2017. “The patterns of interaction between professional translators and online resources.” (http://epubs.surrey.ac.uk/813254/). [Unpublished Ph.D. dissertation, University of Surrey.].

Gough, Joanna. 2019. “Developing translation-oriented research competence: what can we learn from professional translators?”, *The Interpreter and Translator Trainer* 13, 3: 342–359.

Granka, Laura A., Thorsten Joachims and Geri Gay. 2004. “Eye-tracking analysis of user behavior in WWW search”, in: *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR ’04*. Sheffield, United Kingdom: ACM Press, 478. (http://portal.acm.org/citation.cfm?doid=1008992.1009079) (date of access: 16 Aug. 2018).

de Groot, Annette M. 1992. “Determinants of word translation.”, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18, 5: 1001–1018.

Guerberof Arenas, Ana Guerberof. 2013. “What do professional translators think about post-editing?”, *The Journal of Specialized Translation* 19. (http://www.jostrans.org/issue19/art_guerberof.php) (date of access: 20 Feb. 2017).

Gunning, Robert. 1952. *The technique of clear writing*. Toronto: McGraw-Hill.

Halverson, Sandra L. 2017. “Multimethod Approaches”, in: John W. Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 193–212.

Hansen, Gyde. 2003. “Controlling the process: Theoretical and methodological reflections on research into translation processes”, in: Fabio Alves (ed.), *Triangulating Translation: Perspectives in process oriented research*. Amsterdam: John Benjamins Publishing Company, 25–42.

Hansen-Schirra, Silvia. 2017. “EEG and Universal Language Processing in Translation”, in: John W. Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 232–247. (http://doi.wiley.com/10.1002/9781119241485.ch13) (date of access: 8 Aug. 2019).

Herbig, Nico, Santanu Pal, Mihaela Vela, Antonio Krüger and Josef van Genabith. 2019. "Multi-modal indicators for estimating perceived cognitive load in post-editing of machine translation", *Machine Translation* 33, 1–2: 91–115.

Hervet, Guillaume, Katherine Guérard, Sébastien Tremblay and Mohamed Saber Chtourou. 2011. "Is banner blindness genuine? Eye tracking internet text advertising", *Applied Cognitive Psychology* 25, 5: 708–716.

Hirci, Nataša. 2013. "Changing Trends in the Use of Translation Resources: The Case of Trainee Translators in Slovenia", *ELOPE: English Language Overseas Perspectives and Enquiries* 10, 2: 149–165.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka and Joost van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. United Kingdom: Oxford University Press.

Hsieh-Yee, Ingrid. 1993. "Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers", *Journal of the American Society for Information Science* 44, 3: 161–174.

Hu, Ke and Patrick Cadwell. 2016. "A Comparative Study of Post-editing Guidelines", in: *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, 34206-353. (https://www.aclweb.org/anthology/W16-3420) (date of access: 3 Jan. 2020).

Hurtado Albir, Amparo (ed.). 2017. *Researching translation competence by PACTE group*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Hutchins, Edwin. 1995. "How a cockpit remembers its speeds", *Cognitive science* 19, 3: 265–288.

Hutchins, Edwin. 2000. *Cognition in the wild*. (Nachdr.) Cambridge, Mass: MIT Press.

Hutchins, William John and Harold L. Somers. 1997. *An introduction to machine translation*. (2. printing.) London: Academic Press.

Hvelplund, Kristian Tangsgaard. 2011. *Allocation of Cognitive Resources in Translation: An Eye-tracking and Key-logging Study*. Denmark: Samfundslitteratur. [Ph.D. Thesis].

Hvelplund, Kristian Tangsgaard. 2014. "Eye tracking and the translation process: reflections on the analysis and interpretation of eye-tracking data", *Minding translation / Con la traducción en mente* 1: 201-223.

Hvelplund, Kristian Tangsgaard. 2016. “Cognitive efficiency in translation”, in: Ricardo Muñoz Martín (ed.), *Reembedding Translation Process Research*. Benjamins Translation Library, Amsterdam: John Benjamins Publishing Company, 149–170.

Hvelplund, Kristian Tangsgaard. 2017. “Translators’ Use of Digital Resources During Translation”, *HERMES - Journal of Language and Communication in Business* 2017, 56: 71–87.

Hvelplund, Kristian Tangsgaard and Barbara Dragsted. 2018. “Genre familiarity and translation processing: Differences and similarities between literary and LSP translators”, in: Isabel Lacruz and Riitta Jääskeläinen (eds.), *Innovation and Expansion in Translation Process Research*. American Translators Association Scholarly Monograph Series, Amsterdam: John Benjamins Publishing Company, 55–76.

Illich, Ivan. 1981. *Shadow work*. Boston; London: M. Boyars.

Inhoff, Albrecht Werner and Keith Rayner. 1986. “Parafoveal word processing during eye fixations in reading: Effects of word frequency”, *Perception & Psychophysics* 40, 6: 431–439.

Inside Google Translate. 2010. *Internet Archive Wayback Machine*. (https://web.archive.org/web/20100822193710/http://translate.google.com:80/about/intl/en_ALL) (date of access: 3 Jan. 2020).

ISO. 2015. “ISO 17100:2015: Translation services – Requirements for translation services” (http://www.iso.org/iso/catalogue_detail.htm?csnumber=59149) (date of access: 3 May 2018).

Jääskeläinen, Riitta. 1989. “Translation assignment in professional vs. non-professional translation: A think-aloud protocol study”, in: Candace Séguinot (ed.), *The translation process*. H.G. Publications, 87–98.

Jääskeläinen, Riitta. 1990. “Features of successful translation processes: A think-aloud protocol study”. University of Joensuu, Savonlinna School of Translation Studies. [Unpublished Ph.D. Thesis.].

Jääskeläinen, Riitta. 2010. “Are all professionals experts?: Definitions of expertise and reinterpretation of research evidence in process studies”, in: Gregory M. Shreve and Erik Angelone (eds.), *Translation and Cognition*. American Translators Association Scholarly Monograph Series, Amsterdam: John Benjamins Publishing Company, 213–227.

Jakobsen, Arnt Lykke. 1999. “Logging target text production with Translog”, *Copenhagen studies in language* 24: 9–20.

Jakobsen, Arnt Lykke. 2014. “The Development and Current State of Translation Process Research”, in: Elke Brems, Reine Meylaerts and Luc van Doorslaer (eds.), *The Known Unknowns of Translation Studies*. Benjamins Current Topics, Amsterdam: John Benjamins Publishing Company, 65–88.

Jakobsen, Arnt Lykke. 2017. “Translation Process Research”, in: John W. Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 19–49.

Jansen, Bernard J. and Udo Pooch. 2001. “A review of Web searching studies and a framework for future research”, *Journal of the American Society for Information Science and Technology* 52, 3: 235–246.

Jansen, Bernard J., Amanda Spink, Chris Blakely and Sherry Koshman. 2007. “Defining a session on Web search engines”, *Journal of the American Society for Information Science and Technology* 58, 6: 862–871.

Jansen, Bernard J., Amanda Spink and Tefko Saracevic. 2000. “Real life, real users, and real needs: a study and analysis of user queries on the web”, *Information Processing & Management* 36, 2: 207–227.

Jarodzka, H. and S. Brand-Gruwel. 2017. “Tracking the reading eye: towards a model of real-world reading: Towards a model of real-world reading”, *Journal of Computer Assisted Learning* 33, 3: 193–201.

Jensen, Astrid. 1999. “Time pressure in translation”, in: Gyde Hansen (ed.), *Probing the process in translation: Methods and results*. Copenhagen studies in language, Frederiksberg: Samfundslitteratur, 103–119. (http://research.cbs.dk/en/publications/time-pressure-in-translation(9fd81a50-8ebb-11db-a124-000ea68e967b).html) (date of access: 3 May 2018).

Johnson, Melvin, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda B. Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes and Jeffrey Dean. 2016. “Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation”, *CoRR* abs/1611.04558. (http://arxiv.org/abs/1611.04558).

Johnson, Peter. 1992. *Human Computer Interaction: Psychology, task analysis and software engineering.* Maidenhead: McGraw-Hill.

Junczys-Dowmunt, Marcin, Tomasz Dwojak and Hieu Hoang. 2016. “Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions”. (https://workshop2016.iwslt.org/downloads/IWSLT_2016_paper_4.pdf) (date of access: 3 Jan. 2020).

Just, Marcel A and Patricia A Carpenter. 1980. “A theory of reading: from eye fixations to comprehension.”, *Psychological review* 87, 4: 329.

Just, Marcel Adam and Patricia A Carpenter. 1976. “Eye fixations and cognitive processes”, *Cognitive Psychology* 8, 4: 441–480.

Kenny, Dorothy. 2017. “Of Meaning, Materiality and Magic: Conceptualising Machine Translation in the 21st Century” [Keynote lecture 3 at the Translation [And Interpreting] In Transition 3, 14 July 2017, Belgium: Gent].

Kincaid, J. Peter, Robert P. Fishburne Jr, Richard L. Rogers and Brad S. Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. DTIC Document. (http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA006655) (date of access: 25 Oct. 2015).

Kiraly, Don. 1995. *Pathways to translation: pedagogy and process*. Kent, Ohio: Kent State University Press.

Kiraly, Don. 2013. “Towards a view of translator competence as an emergent phenomenon: Thinking outside the box (es) in translator education”, in: Don Kiraly, Silvia Hansen-Schirra and Karin Maksymski (eds.), *New prospects and perspectives for educating language mediators*. Translations-Wissenschaft, Tübingen: Gunter Narr Verlag, 197–224.

Kiraly, Don. 2014. “From Assumptions about Knowing and Learning to Praxis in Translator Education”, *inTRAlinea Special Issue: Challenges in Translation Pedagogy*. (http://www.intralinea.org/specials/article/2100) (date of access: 8 Apr. 2021).

Kirsh, David. 1995. “The intelligent use of space”, *Artificial Intelligence* 73, 1-2: 31–68.

Kirsh, David. 2000. “A Few Thoughts on Cognitive Overload”, *Intellectica* 1, 30: 19–51.

Kit, Chunyu and Jian-Yun Nie. 2015. “Information Retrieval and Text Mining”, in: Sin-wai Chan (ed.), *Routledge encyclopedia of translation technology*. London; New York: Routledge, 494–535.

Koehn, Philipp. 2017. “Neural Machine Translation”, *CoRR* abs/1709.07809. (http://arxiv.org/abs/1709.07809).

Kool, Wouter, Joseph T McGuire, Zev B Rosen and Matthew M Botvinick. 2010. “Decision making and the avoidance of cognitive demand.”, *Journal of Experimental Psychology: General* 139, 4: 665.

Koponen, Maarit, Wilker Aziz, Luciana Ramos and Lucia Specia. 2012. “Post-editing time as a measure of cognitive effort”, *Proceedings of WPTP*. (http://157.56.13.76/AMTA2012Files/html/13/13_paper.pdf) (date of access: 5 May 2016).

Korpal, Paweł. 2017. *Linguistic and psychological indicators of stress in simultaneous interpreting*. Poznań: Wydawnictwo Naukowe UAM.

Koskinen, Kaisa. 2010. “Agency and causality: Towards explaining by mechanisms in translation studies”, in: Ewald Reuter (ed.), *Translators’ agency*. Tampere: Tampere University Press, 165–187.

Krings, Hans Peter. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes.* Kent, Ohio: The Kent State University Press.

Kruger, Haidee. 2016. “What’s happening when nothing’s happening? Combining eyetracking and keylogging to explore cognitive processing during pauses in translation production”, *Across Languages and Cultures* 17, 1: 25–52.

Kruger, Jan-Louis. 2016. “Investigating the cognitive processing of subtitles: EEG, eye tracking, and self-report data”. [Presentation at Tra&Co Symposium 2016, 27–29 April 2016, Gemersheim, Germany].

Krüger, Ralph. 2016. “Contextualising Computer-Assisted Translation Tools and Modelling Their Usability”, *Trans-Kom-Journal of Translation and Technical Communication Research* 9, 1: 114–148.

Kuznik, Anna. 2017. “Use of Instrumental Resources”, in: Amparo Hurtado Albir (ed.), *Researching Translation Competence by PACTE Group*. Benjamins Translation Library, Amsterdam: John Benjamins Publishing Company, 219–241.

Lacruz, Isabel. 2017. “Cognitive Effort in Translation, Editing, and Post-editing”, in: John W Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 386–401. (10.1002/9781119241485.ch21).

Lacruz, Isabel. 2018. “An experimental investigation of stages of processing in post-editing”, in: Isabel Lacruz and Riitta Jääskeläinen (eds.), *Innovation and Expansion in*

Translation Process Research. American Translators Association Scholarly Monograph Series, Amsterdam: John Benjamins Publishing Company, 217–240.

Lacruz, Isabel, Michael Denkowski and Alon Lavie. 2014. "Cognitive Demand and Cognitive Effort in Post-Editing", in: *11th Conference of the Association for Machine Translation in the Americas – Third Workshop on Post-EditingTechnology and Practice*. Vancouver BC, Canada. (https://kilthub.cmu.edu/articles/journal_contribution/Cognitive_Demand_and_Cognitive_Effort_in_Post-Editing/6473261) (date of access: 19 Nov. 2015).

Lacruz, Isabel and Riitta Jääskeläinen (eds.). 2018. *Innovation and Expansion in Translation Process Research*. Amsterdam: John Benjamins Publishing Company.

Lacruz, Isabel and Gregory M. Shreve. 2014. "Pauses and Cognitive Effort in Post-editing", in: Laura Winther Balling and Michael Carl (eds.), *Post-editing of machine translation: Processes and applications*. Cambridge Scholars Publishing, 246–272. (date of access: 23 Jan. 2017).

Lacruz, Isabel, Gregory M. Shreve and Erik Angelone. 2012. "Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study", in: *AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012)*. San Diego, USA: Association for Machine Translation in the Americas (AMTA), 21–30.

Lagoudaki, Elina. 2008. "The value of machine translation for the professional translator", in: *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas*. Citeseer, 262–269.

Latour, Bruno. 2005. *Reassembling the social: An introduction to actor-network-theory*. Oxford university press.

Läubli, Samuel. 2017. "3 Reasons Why Neural Machine Translation Is a Breakthrough" [Presentation at SlatorCon, Zurich].

Läubli, Samuel. 2017. "3 Reasons Why Neural Machine Translation is a Breakthrough", *Slator*. (https://slator.com/technology/3-reasons-why-neural-machine-translation-is-a-breakthrough/) (date of access: 3 Jan. 2020).

Läubli, Samuel and David Orrego-Carmona. 2017. "When Google Translate is better than Some Human Colleagues, those People are no longer Colleagues", *Proceedings of the 39th Conference Translating and the Computer* 59-69.

Lee, Young-Jin. 2003. “Effective Web Searching for Open-Ended Questions: The Effects of Visualization and Data Mining Technology”. [Unpublished Ph.D. Thesis, University of Illinois].

Lee, Young-Jin. 2005. “VisSearch: A collaborative Web searching environment”, *Computers & Education* 44, 4: 423–439.

Leijten, Mariëlle and Luuk Van Waes. 2013. “Keystroke logging in writing research using Inputlog to analyze and visualize writing processes”, *Written Communication* 30, 3: 358–392.

Leijten, Mariëlle and Luuk Van Waes. 2019. “Inputlog website”, *Inputlog | A keystroke logging application*. (http://www.inputlog.net/) (date of access: 12 Aug. 2019).

Leijten, Mariëlle, Luuk Van Waes and Eric Van Horenbeeck. 2019. “Help documentation (in progress)”. [Inputlog PDF manual].

Lemhöfer, Kristin and Mirjam Broersma. 2012. “Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English”, *Behavior Research Methods* 44, 2: 325–343.

Leontiev, Aleksei Nikolaevich. 1977. “Activity and consciousness”, in: *Philosophy in the USSR, Problems of Dialectical Materialism*. Moscow: Progress Publishers, 180-202. (https://www.marxists.org/archive/leontev/works/1977/leon1977.htm) (date of access: 2 Jun. 2018).

Lommel, Arle R. and Donald A. DePalma. 2016. *Post-editing Goes Mainstream. How LSPs Use MT to Meet Client Demands*. Common Sense Advisory, Inc. [Common Sense Advisory Report].

Mandera, Paweł, Emmanuel Keuleers, Zofia Wodniecka and Marc Brysbaert. 2015. “Subtlex-pl: subtitle-based word frequency estimates for Polish”, *Behavior Research Methods* 47, 2: 471–483.

Massey, Gary and Maureen Ehrensberger-Dow. 2011a. “Investigating information literacy: A growing priority in translation studies”, *Across Languages and Cultures* 12, 2: 193–211.

Massey, Gary and Maureen Ehrensberger-Dow. 2011b. “Technical and Instrumental Competence in the Translator’s Workplace: Using Process Research to Identify Educational and Ergonomic Needs”, *ILCEA [Online]* 14: 1–14. (https://doi.org/10.4000/ilcea.1060) (date of access: 3 Jan. 2020).

Massey, Gary and Maureen Ehrensberger-Dow. 2014. “Looking beyond the text: the usefulness of translation process data”, in: Dagmar Knorr, Carmen Heine and Jan Engberg (eds.), *Methods in Writing Process Research*. Peter Lang D, 81–98.

Massey, Gary and Peter Jud. 2015. “Teaching audiovisual translation with products and processes: Subtitling as a case in point”, *Accessing audiovisual translation* 99–116.

Matacic, Catherine. 2016. “Google’s new translation software is powered by brainlike artificial intelligence”, *Science | AAAS*. (http://www.sciencemag.org/news/2016/09/google-s-new-translation-software-powered-brainlike-artificial-intelligence) (date of access: 7 Feb. 2018).

MateCat. 2018. *Machine translation and post editing in the language industry*. (https://www.youtube.com/watch?v=ESqfo3Jtpsw&feature=youtu.be) (date of access: 1 Jun. 2018). [Webinar].

Meijer, Siety. 1993. “Attitudes towards machine translation”, *Language International* 5, 6: 11–13.

Mellinger, Christopher D. and Thomas A. Hanson. 2018. “Order effects in the translation process”, *Translation, Cognition & Behavior* 1, 1: 1–20.

Mesa-Lao, Bartolomé. 2013a. “Eye-tracking Post-editing Behaviour in an Interactive Translation Prediction Environment”, in: *Book of Abstracts: 17th European Conference on Eye Movemement, 11-16 August 2013*. Lund University.

Mesa-Lao, Bartolomé. 2013b. “Introduction to post-editing – The CasMaCat GUI”. (http://bridge.cbs.dk/projects/seecat/material/hand-out_post-editing_bmesa-lao.pdf) (document not available any more).

Moorkens, Joss. 2018. “Eye tracking as a measure of cognitive effort for post-editing of machine translation”, in: Callum Walker and Federico M. Federici (eds.), *Eye Tracking and Multidisciplinary Studies on Translation*. Benjamins Translation Library, Amsterdam: John Benjamins Publishing Company, 55–70.

Moorkens, Joss and Sharon O’Brien. 2015. “Post-Editing Evaluations: Trade-offs between Novice and Professional Participants”, in: İlknur El-Kahlout, Mehmed Özkan, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez, Fred Hollywood and Andy Way (eds.), *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*. Antalya, Turkey: European Association for Machine Translation, 75-81. (https://www.aclweb.org/anthology/2015.eamt-1.0) (date of access: 3 Jan. 2020).

Moorkens, Joss, Sharon O'Brien, Igor A. L. da Silva, Norma B. de Lima Fonseca and Fabio Alves. 2015. "Correlations of perceived post-editing effort with measurements of actual effort", *Machine Translation* 29: 267–284.

Muñoz Martín, Ricardo. 2010. "On paradigms and cognitive translatology", in: Gregory M. Shreve and Erik Angelone (eds.), *Translation and Cognition*. American Translators Association Scholarly Monograph Series, Amsterdam: John Benjamins Publishing Company, 169–187.

Muñoz Martín, Ricardo. 2012. "Just a matter of scope", *Translation Spaces* 1: 169–188.

Muñoz Martín, Ricardo. 2016. "Of minds and men–computers and translators", *Poznan Studies in Contemporary Linguistics* 52, 2: 351–381.

Mutta, Maarit, Sanna Pelttari, Leena Salmi, Aline Chevalier and Marjut Johansson. 2014. "Digital Literacy in Academic Language Learning Contexts: Developing Information-seeking Competence", in: Janel Pettes Guikema and Lawrence Williams (eds.), *Digital literacies in foreign and second language education*. Calico Monograph Series, Texas: Texas State University: Computer Assisted Language Instruction Consortium (CALICO), 227-244.

Nicholas, David, Paul Huntington, Peter Williams and Tom Dobrowolski. 2006. "The Digital Information Consumer", in: Amanda Spink and Charles Cole (eds.), *New Directions in Human Information Behavior*. Berlin/Heidelberg: Springer-Verlag, 203–228.

Nord, Christiane. 2005. *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam; New York, NY: Rodopi.

O'Brien, Sharon. 2002. "Teaching post-editing: a proposal for course content", in: *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*. Manchester, England: European Association for Machine Translation, 99–106.

O'Brien, Sharon. 2005. "Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability", *Machine Translation* 19, 1: 37–58.

O'Brien, Sharon. 2006. "Pauses as indicators of cognitive effort in post-editing machine translation output", *Across Languages and Cultures* 7, 1: 1–21.

O'Brien, Sharon. 2007. "Eye-tracking and translation memory matches", *Perspectives* 14, 3: 185–205.

O'Brien, Sharon. 2010. "Introduction to Post-Editing: Who, What, How and Where to Next?", in: *Proceedings of the Ninth Conference of the Association for Machine Translation in the Americas*. Denver, Colorado.

O'Brien, Sharon. 2012. "Translation as human–computer interaction", *Translation Spaces* 1, 1: 101–122.

O'Brien, Sharon. 2017. "Machine Translation and Cognition", in: John W. Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 311–331.

Och, Franz. 2006. "Statistical machine translation live", *Research Blog*. (https://research.googleblog.com/2006/04/statistical-machine-translation-live.html) (date of access: 5 Feb. 2018).

Och, Franz Josef. 2005. "Statistical machine translation: Foundations and recent advances", *Proceedings of MT Summit X: the Tenth Machine Translation Summit: 12–16 September 2005, Phuket, Thailand*.

OED Online. 2018. "machine learning, n." (http://www.oed.com) (date of access: 8 Feb. 2018).

OED Online. 2019. "'readability, n.'." (http://www.oed.com) (date of access: 16 Aug. 2019).

Office Support. 2017. "Test your document's readability" (https://support.office.com/en-us/article/Test-your-document-s-readability-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2) (date of access: 20 Mar. 2017).

Olohan, Maeve. 2011. "Translators and translation technology: The dance of agency", *Translation Studies* 4, 3: 342–357.

Ortiz-Boix, Carla and Anna Matamala. 2015. "Quality assessment of post-edited versus translated wildlife documentary films", in: *4th Workshop on Post-editing Technology and Practice (WPTP4)*. (https://ddd.uab.cat/record/142792) (date of access: 26 Jan. 2017).

Paas, Fred G. 1992. "Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach.", *Journal of Educational Psychology* 84, 4: 429–434.

Paas, Fred, Juhani E. Tuovinen, Huib Tabbers and Pascal W. M. Van Gerven. 2003. "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory", *Educational Psychologist* 38, 1: 63–71.

PACTE Group. 2005. "Investigating Translation Competence: Conceptual and Methodological Issues", *Meta: Journal des traducteurs* 50, 2: 609.

PACTE Group. 2017. *Researching Translation Competence by PACTE Group*. Amsterdam: John Benjamins Publishing Company.

Pakkala-Weckström, Mari. 2015. "Student's Data Mining Skills in Second-Year Undergraduate Translation", *Current Trends in Translation Teaching and Learning E* 2: 139–170.

Papineni, Kishore, Salim Roukos, Todd Ward and Wei-jing Zhu. 2002. "BLEU: a Method for Automatic Evaluation of Machine Translation", in: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Philadelphia, 311–318.

Paradowska, Urszula. 2015. "Expert web searching skills for translators – a multiple-case study", in: Mikołaj Deckert and Paulina Pietrzak (eds.), *Constructing Translation Competence*. Łódź Studies in Language, Frankfurt am Main; New York: Peter Lang Edition, 227–244. (https://www.peterlang.com/view/9783653969429/xhtml/chapter013.xhtml) (date of access: 24 Jan. 2018).

Pastor, Verónica and Amparo Alcina. 2009. "Search techniques in corpora for the training of translators", in: *Proceedings of the Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning*. Association for Computational Linguistics, 13–20.

Pavlović, Nataša. 2007. "Directionality in collaborative translation processes". [Unpublished Ph. D. Thesis. Universitat Rovira i Virgili.].

Pavlović, Nataša and Kristian Jensen. 2009. "Eye tracking translation directionality", in: A. Pym and A. Perekrestenko (eds.), *Translation research projects 2*. Tarragona: Intercultural Studies Group, 101–119.

Pavlović, Tanja. 2014. "Exploring directionality in translation studies", *Explorations in English Language and Linguistics* 1, 2: 149–165.

Peters, Isabella and Paul Becker. 2009. *Folksonomies: indexing and retrieval in Web 2.0*. Berlin: De Gruyter/Saur.

Pickering, Andrew. 1993. "The Mangle of Practice: Agency and Emergence in the Sociology of Science", *American Journal of Sociology* 99, 3: 559–589.

Pickering, Andrew. 2008. "New ontologies", in: Andrew Pickering and Keith Guzik (eds.), *The mangle in practice: science, society, and becoming*. Science and cultural theory, Durham: Duke University Press, 1–14.

Pickering, Andrew. 2010. *The Mangle of Practice: Time, Agency, and Science.* Chicago: University of Chicago Press.

Piolat, Annie, Jean-Yves Roussey, Thierry Olive and Murielle Amada. 2004. "Processing time and cognitive effort in revision: effects of error type and of working memory capacity", in: *Revision Cognitive and Instructional Processes*. Springer, 21–38.

Piotrowska, Maria. 2015. "Revisiting the Translator Competence in the 21st Century", in: Paulina Pietrzak and Mikołaj Deckert (eds.), *Constructing translation competence*. Lódź studies in language, Frankfurt am Main; New York: Peter Lang Edition, 13–27. (https://www.peterlang.com/view/9783653969429/xhtml/chapter001.xhtml) (date of access: 24 Jan. 2018).

Plitt, Mirko and François Masselot. 2010. "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context", *The Prague Bulletin of Mathematical Linguistics* 93, 1. (http://www.degruyter.com/view/j/pralin.2010.93.issue--1/v10108-010-0010-x/v10108-010-0010-x.xml) (date of access: 14 Oct. 2015).

Pokorn, Nike K., Jason Blake, Donald Reindl and Agnes Pisanski Peterlin. 2020. "The influence of directionality on the quality of translation output in educational settings", *The Interpreter and Translator Trainer* 14, 1: 58–78.

Pym, Anthony. 2011. "What technology does to translating", *Translation & Interpreting* 3, 1: 1–9.

Pym, Anthony. 2013. "Translation Skill-sets in a Machine-translation Age", *Meta: Journal des traducteursMeta:/Translators' Journal* 58, 3: 487–503.

Radach, Ralph and Alan Kennedy. 2013. "Eye movements in reading: Some theoretical context", *Quarterly Journal of Experimental Psychology* 66, 3: 429–452.

Raído, Vanessa Enríquez. 2014. *Translation and web searching*. New York; London: Routledge, Taylor & Francis Group.

Rayner, Keith. 1998. "Eye movements in reading and information processing: 20 years of research.", *Psychological bulletin* 124, 3: 372.

Rayner, Keith. 2009. "Eye movements and attention in reading, scene perception, and visual search", *Quarterly Journal of Experimental Psychology* 62, 8: 1457–1506.

Rayner, Keith and Alexander Pollatsek. 1989. *The psychology of reading*. Englewood Cliffs, N.J: Prentice Hall.

Reiss, Katharina. 1976. *Texttyp und Übersetzungsmethode: der operative Text*. Kronberg: Scriptor Verlag.

Rele, Rachana S. and Andrew T. Duchowski. 2005. "Using Eye Tracking to Evaluate Alternative Search Results Interfaces", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 49, 15: 1459–1463.

Richman, Dan. 1995. "Translation software picks up steam as demand booms", *Computerworld* 29, 43: 81–84.

Risku, Hanna. 2010. "A cognitive scientific view on technical communication and translation: Do embodiment and situatedness really make a difference?", *Target* 22, 1: 94–111.

Risku, Hanna and Florian Windhager. 2013. "Extended translation: A sociocognitive research agenda", *Target. International Journal of Translation Studies* 25, 1: 33–45.

Risku, Hanna, Florian Windhager and Matthias Apfelthaler. 2013. "A dynamic network model of translatorial cognition and action", *Translation Spaces* 2: 151–182.

Roberts, Maxwell J. and George Erdos. 1993. "Strategy Selection and Metacognition", *Educational Psychology* 13, 3-4: 259–266.

Rowlands, Ian, David Nicholas, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, Richard Withey, Hamid R. Jamali, Tom Dobrowolski and Carol Tenopir. 2008. "The Google generation: the information behaviour of the researcher of the future", *Aslib Proceedings* 60, 4: 290–310.

Sales Salvador, Dora. 2006. "Documentation as ethics in postcolonial translation", *Translation Journal* 10, 1.

Schaeffer, Moritz and Michael Carl. 2013. "Shared representations and the translation process: A recursive model", *Translation and Interpreting Studies* 8, 2: 169–190.

Schaeffer, Moritz and Michael Carl. 2017. "A Minimal Cognitive Model for Translating and Post-editing", in: Sadao Kurohashi and Pascale Fung (eds.), *Proceedings of MT Summit XVI*. Japan: International Association for Machine Translation, 144–155.

Schuster, Mike, Melvin Johnson and Nikhil Thorat. 2016. "Zero-Shot Translation with Google's Multilingual Neural Machine Translation System", *Research Blog*. (https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html) (date of access: 7 Feb. 2018).

Séguinot, Candace. 1991. "A study of student translation strategies", in: Sonja Tirkkonen-Condit (ed.), *Empirical research in translation and intercultural studies*. Tubingen, Germany: Guntere Narr Verlag, 79–88.

Séguinot, Candace. 2000. “Knowledge, expertise, and theory in translation”, in: Andrew Chesterman, Natividad Gallardo San Salvador and Yves Gambier (eds.), *Translation in Context: Selected papers from the EST Congress, Granada 1998*. Benjamins Translation Library, Amsterdam: John Benjamins Publishing Company, 87–104.

Shreve, Gregory M. and Isabel Lacruz. 2017. “Aspects of a Cognitive Model of Translation”, in: John W. Schwieter and Aline Ferreira (eds.), *The Handbook of Translation and Cognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 127–143.

Simard, Tanya. 2013. “Analyse comparative de la couverture et de l’acceptabilité des solutions d’un dictionnaire bilingue spécialisé, d’une banque de données terminologiques et d’un concordancier en ligne: application au domaine de la traduction médicale”. [Unpublished Ph.D. Thesis].

Sjørup, Annette Camilla. 2013. *Cognitive effort in metaphor translation: an eye-tracking and key-logging study*. Frederiksberg: Copenhagen Business School (CBS).

Smallwood, Jonathan and Jonathan W Schooler. 2006. “The restless mind.”, *Psychological bulletin* 132, 6: 946.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla and John Makhoul. 2006. “A study of translation edit rate with targeted human annotation”, in: *Proceedings of association for machine translation in the Americas*, 223–231. (http://mt-archive.info/AMTA-2006-Snover.pdf) (date of access: 14 Oct. 2015).

Specia, Lucia, Marco Turchi, Nicola Cancedda, Marc Dymetman and Nello Cristianini. 2009. “Estimating the sentence-level quality of machine translation systems”, in: *13th Conference of the European Association for Machine Translation*, 28–37.

Spink, Amanda, Howard Greisdorf and Judy Bateman. 1998. “From highly relevant to not relevant: examining different regions of relevance”, *Information Processing & Management* 34, 5: 599–621.

Spink, Amanda and Tefko Saracevic. 1997. “Interaction in information retrieval: Selection and effectiveness of search terms”, *Journal of the American Society for Information Science* 48, 8: 741–761.

SR Research. 2018. “How does eye tracking work?”, *SR Research*. (https://www.sr-research.com/eye-tracking-blog/background/how-does-eye-tracking-work/) (date of access: 4 Jan. 2020).

Stefaniak, Karolina. 2020. “Evaluating the usefulness of neural machine translation for the Polish translators in the European Commission”, *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* 263–269.

Strohner, Hans. 1995. *Kognitive systeme: eine einführung in die kognitionswissenschaft*. Darmstadt: Westdeutscher Verlag.

Sutcliffe, Alistair and Mark Ennis. 1998. “Towards a cognitive theory of information retrieval”, *Interacting with Computers* 10, 3: 321–351.

Sweller, John. 1988. “Cognitive Load During Problem Solving: Effects on Learning”, *Cognitive Science* 12, 2: 257–285.

SYSTRAN. [n.d.]. *What is Machine Translation? Rule Based vs. Statistical*. (http://www.systransoft.com/systran/translation-technology/what-is-machine-translation/) (date of access: 5 Feb. 2018).

Tatsumi, Midori. 2010. “Post-editing machine translated text in a commercial setting: Observation and statistical analysis”. [Unpublished Ph.D. Thesis.].

TAUS. 2010. *Post-editing in Practice. A TAUS Report*. (https://www.taus.net/think-tank/reports/postedit-reports/postediting-in-practice) (date of access: 9 Apr. 2018).

TAUS. 2016. “TAUS Machine Translation Post-editing Guidelines” (http://info.taus.net/mt-post-editing-guidelines) (date of access: 3 May 2018).

Tirkkonen-Condit, Sonja. 1990. “Professional vs. Non-Professional Translation: A Think-Aloud Protocol Study”, in: M.A.K. Halliday, John Gibbons and Howard Nicholas (eds.), *Learning, Keeping and Using Language Selected papers from the Eighth World Congress of Applied Linguistics, Sydney, 16–21 August 1987*. Amsterdam: John Benjamins Publishing Company, 381–384.

Tra&Co Team. 2016. “Translation performance metrics: From empirical data to translation process modeling”. [Presentation at Tra&Co Symposium 2016, 27–29 April, Germersheim: Germany.].

Turovsky, Barak. 2016. “Found in translation: More accurate, fluent sentences in Google Translate”, *Google*. (https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/) (date of access: 7 Feb. 2018).

Vakkari, Pertti. 1999. “Task complexity, problem structure and information actions”, *Information Processing & Management* 35, 6: 819–837.

Van Waes, Luuk, Mariëlle Leijten, Tom Pauwaert and Eric Van Horenbeeck. 2019. “A Multilingual Copy Task: Measuring Typing and Motor Skills in Writing with Inputlog”, *Journal of Open Research Software* 7: 30.

Vieira, Lucas Nunes. 2016. “How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data”, *Machine Translation* 30, 1-2: 41–62.

Vieira, Lucas Nunes. 2020. “Automation anxiety and translators”, *Translation Studies* 13, 1: 1–21.

White, Marilyn Domas and Mirja Iivonen. 2001. “Questions as a factor in Web search strategy”, *Information Processing & Management* 37, 5: 721–740.

Whyatt, Bogusława. 2012. *Translation as a human skill: from predisposition to expertise*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Whyatt, Bogusława. 2018. “Testing Indicators of Translation Expertise in an Intralingual Task”, *HERMES - Journal of Language and Communication in Business* 57: 53–78.

Whyatt, Bogusława, Katarzyna Stachowiak and Marta Kajzer-Wietrzny. 2016. “Similar and different: cognitive rhythm and effort in translation and paraphrasing”, *Poznan Studies in Contemporary Linguistics* 52, 2: 176–208.

Whyatt, Bogusława, Olga Witczak and Ewa Tomczak. 2021. “Information behaviour in bidirectional translators: focus on online resources”, *The Interpreter and Translator Trainer*, 1–18.

Wilson, T.D. 1999. “Models in information behaviour research”, *Journal of Documentation* 55, 3: 249–270.

Winograd, Terry. 1972. “Understanding natural language”, *Cognitive Psychology* 3, 1: 1–191.

Witczak, Olga. 2016a. “Incorporating post-editing into a computer-assisted translation course. A study of student attitudes.”, *Journal of Translator Education and Translation Studies* 1, 1: 35–55.

Witczak, Olga. 2016b. “Tłumacze kontra maszyny, czyli o tłumaczeniu wspomaganym komputerowo”, in: *Tłumacz – praktyczne aspekty zawodu*. Filologia Angielska, Poznań: Wydawnictwo Naukowe UAM, 203–234.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang,

Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes and Jeffrey Dean. 2016. “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation”, *CoRR* abs/1609.08144. (http://arxiv.org/abs/1609.08144).

Yang, Jin and Elke D. Lange. 1998. “SYSTRAN on AltaVista A User Study on Real-Time Machine Translation on the Internet”, in: David Farwell, Laurie Gerber and Eduard Hovy (eds.), *Machine Translation and the Information Soup*. Berlin, Heidelberg: Springer Berlin Heidelberg, 275-285. (http://link.springer.com/10.1007/3-540-49478-2_25) (date of access: 5 Feb. 2018).

Yang, Liu and Zhang Min. 2015. “Statistical Machine Translation”, in: Sin-wai Chan (ed.), *Routlege Encyclopedia of Translation Technology*. London; New York: Routlege, 201–212.

Zapata, Julián. 2016. “Investigating Translator-Information Interaction: A Case Study on the Use of the Prototype Biconcordancer Tool Integrated in CASMACAT”, in: *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. New Frontiers in Translation Studies, Springer, 135–152.

Zimowska, Sylwia. 2017. “Tłumacz Google lepiej zrozumie język polski”, *CHIP*. (https://www.chip.pl/2017/03/tlumacz-google-lepiej-zrozumie-jezyk-polski/) (date of access: 7 Feb. 2018).

Appendix A: Source texts and MT output

WAT ST: Omega Moonwatch text

We are very proud to stock the Omega Speedmaster Professional Moon Watch which used to be a part of all six Apollo lunar landings. This incredibly well engineered timepiece has a stainless steel case which houses a perfectly proportioned black face with tachymeter and black chronograph sub dials with white detailing protected by hesalite crystal glass. The case is paired with a matching stainless steel bracelet. A manual movement ensures extremely accurate timekeeping, making this a very serious contender for one of the finest watches currently available on the market.

WAT MT: Omega Moonwatch MT output

Z dumą możemy pochwalić się Omega Speedmaster Professional Moon Watch, który był częścią wszystkich sześciu lądowań księżycowych Apollo. Ten niezwykle dobrze zaprojektowany zegarek ma obudowę ze stali nierdzewnej, w której mieści się idealnie proporcjonalna czarna twarz z tachymetrem i czarnymi podrzędnymi tarczami chronografu z białymi detalami chronionymi kryształowym szkłem hesalitowym. Obudowa jest połączona z pasującą bransoletką ze stali nierdzewnej. Ruch ręczny zapewnia niezwykle dokładny pomiar czasu, co sprawia, że jest to bardzo poważny pretendent do jednego z najlepszych zegarków dostępnych obecnie na rynku.

MMR ST: MMR vaccine text

MMR is a safe and effective combined vaccine that protects against three separate illnesses – measles, mumps and rubella (German measles) – in a single injection and the full course of MMR vaccination requires two doses. Measles, mumps and rubella are highly infectious conditions that can have serious, and potentially fatal, complications, including meningitis, swelling of the brain (encephalitis) and deafness. They can also lead to complications in pregnancy that affect the unborn baby, and can lead to miscarriage. Since the

MMR vaccine was introduced in 1988, it's rare for children in the UK to develop these serious conditions.

MMR MT: MMR vaccine MT output

MMR jest bezpieczną i skuteczną kombinowaną szczepionką chroniącą przed trzema odrębnymi chorobami - odrą, świnką i różyczką (odrę) w pojedynczym wstrzyknięciu, a pełne szczepienie MMR wymaga dwóch dawek. Odra, śwince i różyczce są wysoce zakaźnymi chorobami, które mogą mieć poważne i potencjalnie śmiertelne powikłania, w tym zapalenie opon mózgowych, obrzęk mózgu (zapalenie mózgu) i głuchota. Mogą również prowadzić do powikłań w ciąży, które mają wpływ na nienarodzone dziecko i mogą prowadzić do poronienia. Ponieważ szczepionka MMR została wprowadzona w 1988 r., Rzadko zdarza się, że dzieci w Wielkiej Brytanii rozwinęły te poważne choroby.

APP ST: Appendicitis text

In most cases of appendicitis, the appendix needs to be surgically removed as soon as possible. Removal of the appendix, known as an appendectomy or appendicectomy, is one of the most common operations in the UK and its success rate is excellent. The operation is most commonly performed as keyhole surgery (laparoscopy), which involves making several small cuts in your abdomen, through which special surgical instruments are inserted. Open surgery, where a larger, single cut is made in the abdomen, is usually carried out if the appendix has burst or access is more difficult.

HEA ST: Headphones text

The DT770 Pro Limited Edition by Beyer Dynamic are closed backed headphones for professional studio applications. Featuring an 80 Ohm drivers you can also use the DT770 Pro with your smartphones, iPods, iPads and Hi-fi equipment. The single-sided cable makes the handling of the headphone easy, the low weight diaphragms produce an excellent sound reproduction and the soft comfortable ear pads ensure your ears will not suffer with fatigue

after long periods of use. The DT770 Pro is a closed dynamic headphone which has been designed for critical music and sound monitoring.

Appendix B: Language History Questionnaire

Administered via printed copy

Language History Questionnaire (LCL custom version)

Participant #:	
Age:	
Sex (circle):	M / F

What is your country of residence? ....

What is your country of origin? ....

(1) Indicate your native language(s) and any other languages you have studied or learned, the age at which you started using each language in terms of listening, speaking, reading, and writing, and the total number of years you have spent using each language:

Language↓	Listening	Speaking	Reading	Writing	Years of Use*

*You may have learned a language, stopped using it, and then started using it again. Please give the total number of years.

(2) Rate your current ability in terms of listening, speaking, reading, and writing in each of the languages you have studied or learned. Please rate according to the following scale (circle the number in the table):

Very poor	Poor	Limited	Functional	Good	Very good	Native-like
1	2	3	4	5	6	7

Language ↓	Listening	Speaking	Reading	Writing
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

Appendix C: Pre-task questionnaire

Administered via 1ka.si online questionnaire

Q1 - Participant code

Q2 - Year:

- ○ 1 MA Translation
- ○ 2 MA Translation
- ○ 1 MA General
- ○ 2 MA General
- ○ 3BA General

BLOCK (1)
Q3 - Have your ever post-edited machine translation output before?

- ○ Yes
- ○ No

Q4 - What did/do you post-edit machine translation output for?
You can select more than one answer

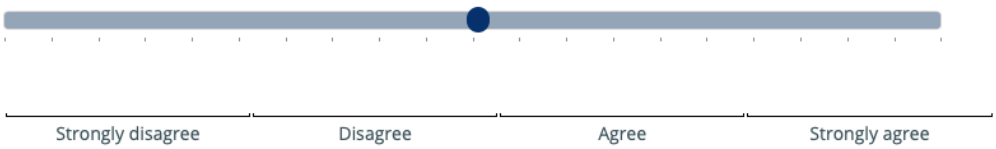
- ☐ Translation assignment from a client
- ☐ Homework assignment
- ☐ Other:

Q5 - How often do you use machine translation for your own needs?

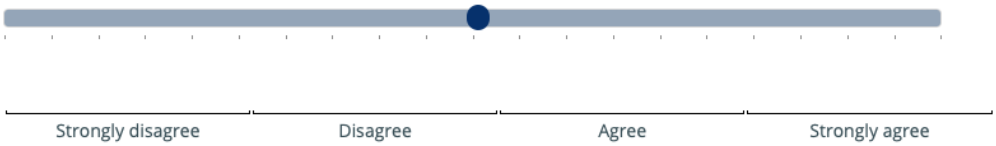
- ○ Not at all
- ○ Rarely (less frequently than once a month)
- ○ Up to a few times a month
- ○ Weekly
- ○ Daily

Please indicate how much you agree with the following statements:

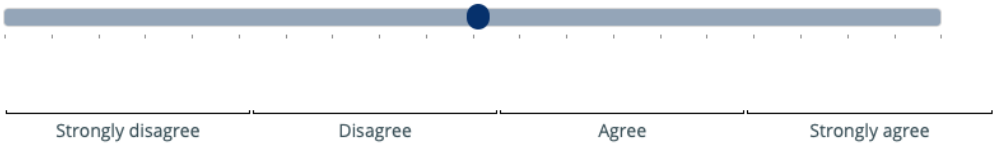
Q6 - Machine translation is useful for everyday Internet browsing (e.g. shopping).
The middle of the slider stands for neutral/I don't know.



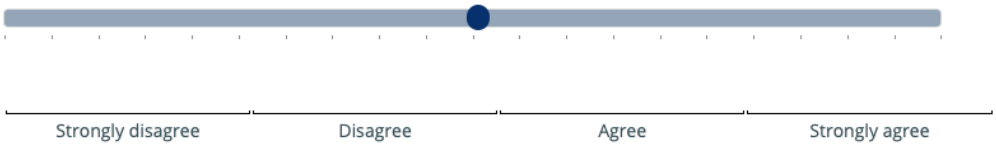
Q7 - Human translators will soon be replaced by machine translation.
The middle of the slider stands for neutral/I don't know.



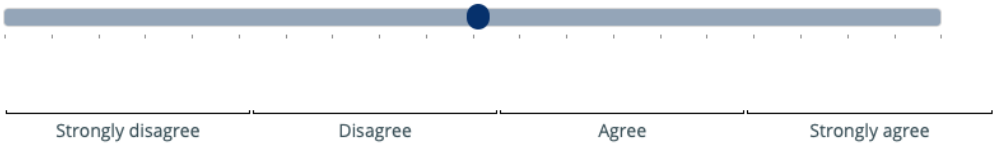
Q8 - Internet is indispensable for translation.
The middle of the slider stands for neutral/I don't know.



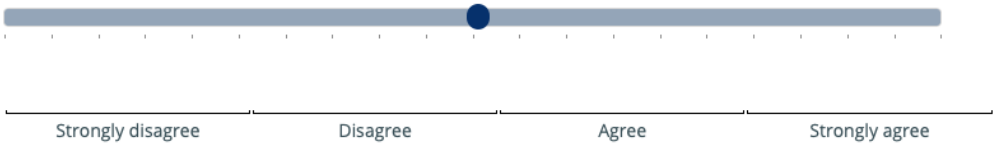
Q9 - Machine translation cannot compete with human translation.
The middle of the slider stands for neutral/I don't know.



Q10 - Machine translation can speed up human translation.
The middle of the slider stands for neutral/I don't know.



Q11 - Machine translation is useful as a translation aid.
The middle of the slider stands for neutral/I don't know.



Appendix D: Post-task questionnaire (WAT+APP)

Administered via 1ka.si online questionnaire

Q1 - Participant code

Q2 - Please rate the difficulty of the source text (Omega Moon Watch):

- ◯ Very easy
- ◯ Easy
- ◯ Medium
- ◯ Difficult
- ◯ Very difficult

Q3 - Please rate the difficulty of the source text (Appendicitis):

- ◯ Very easy
- ◯ Easy
- ◯ Medium
- ◯ Difficult
- ◯ Very difficult

Q4 - How do you feel machine translation influenced your process of online researching (changed it completely, did not affect it at all, etc.)? Please provide a brief description.

Q5 - Do you have any additional comments pertaining to MT or the task you just completed? Any insight will be greatly appreciated.

Appendix E: Post-task questionnaire (HEA+MMR)

Administered via 1ka.si online questionnaire

Q1 - Participant code

Q2 - Please rate the difficulty of the source text (Headphones):

- ◯ Very easy
- ◯ Easy
- ◯ Medium
- ◯ Difficult
- ◯ Very difficult

Q3 - Please rate the difficulty of the source text (MMR vaccination):

- ◯ Very easy
- ◯ Easy
- ◯ Medium
- ◯ Difficult
- ◯ Very difficult

Q4 - How do you feel machine translation influenced your process of online researching (changed it completely, did not affect it at all, etc.)? Please provide a brief description.

Q5 - Do you have any additional comments pertaining to MT or the task you just completed? Any insight will be greatly appreciated.

Appendix F: General instructions for participants

INSTRUKCJA DLA UCZESTNIKA BADANIA

Dziękuję serdecznie za udział w badaniu.
Celem eksperymentu jest analiza tłumaczenia i post-edycji tłumaczenia maszynowego.
Post-edycja polega na korekcie tłumaczenia wygenerowanego maszynowo w taki sposób, aby tekst docelowy nadawał się do publikacji.
Eksperyment polega na wykonaniu tłumaczenia 2 tekstów i post-edycji 2 tekstów oraz składa się z dwóch części, które odbywają się w różne dni. Eksperyment polega na:

1. Wypełnieniu kwestionariusza poprzedzającego tłumaczenie i post-edycję;
2. Wykonaniu testu pisania na klawiaturze;
3. Przetłumaczeniu jednego tekstu i post-edycji innego tekstu (każdy poniżej 100 słów);
4. Wypełnieniu kwestionariusza na temat tekstów;
5. Wypełnieniu testu LexTALE oraz kwestionariusza LHQ.

Druga część eksperymentu będzie polegała na wykonaniu czynności 3 i 4 dla innych tekstów.

Odpowiedzi w kwestionariuszach oraz dane z eksperymentu zostaną zakodowane.

Eksperyment będzie rejestrowany okulografem i programami komputerowymi, służącymi do zapisu aktywności na ekranie (Inputlog i Morae Recorder). Nie wiąże się to z nagrywaniem wizerunku w żadnej formie.

Przed tłumaczeniem każdego tekstu okulograf zostanie skalibrowany (przyzwyczajony do Twojego wzroku). **Po kalibracji proszę o wykonanie zadania w naturalnej i wygodnej pozycji, ale bez gwałtownych ruchów głową. Proszę również o nieobracanie się na krześle.**

W czasie tłumaczenia ekran będzie podzielony na 2 części:

- po lewej: tekst źródłowy i miejsce do wprowadzenia tekstu docelowego;
- po prawej: przeglądarka internetowa Google Chrome.

Proszę o nieprzesuwanie żadnego z wyżej wymienionych okien, niezmienianie ich rozmiaru ani nieminimalizowanie ich.

Nie edytuj też tekstu źródłowego i nie zmieniaj jego formatowania (nie klikaj na „wstążkę" programu Word).

Po zakończeniu tłumaczenia/post-edycji **wpisz na końcu dokumentu literkę "f" (bez dodatkowych enterów)** i zawołaj „Gotowe!". Po wpisaniu "f" niczego więcej nie klikaj, nie zapisuj ani nie zamykaj.

Jeśli masz jakiekolwiek pytania do osoby prowadzącej eksperyment w trakcie wykonywania jednej z części badania, zawołaj prowadzącą, ale nie wstawaj.

Translation:

INSTRUCTION FOR THE PARTICIPANT

Thank you for participating in my study.
The aim of this experiment is the analysis of translation from scratch and post-editing of machine translation.
Post-editing is the correction of machine translation in such a way that the target text can be published.
The experiment consists in translating two texts and post-editing two texts. It is divided into two parts which have to take place over two different days. The experiment involves:

1. Filling in the pre-task questionnaire which precedes the translation and post-editing part;
2. Completing the Copy Task;
3. Translating one text and post-editing another text (each below 100 words);
4. Taking the LexTALE test and filling in an LHQ questionnaire.

Part two of this study involves completing steps 3 and 4 for different texts.

Answers in questionnaires and data from the experiment will be anonymised with codes.

The procedure will be recorded with an eye-tracker and computer software capturing activity on screen (Inputlog and Morae Recorder). This does not involve recording of your face in any form.

Before commencing translation of each text, the eye-tracker will be calibrated (familiarised with your eyes). **After the calibration, please complete the task in a comfortable position, but don't make any sudden head movements. Also, please do not move around on the chair.**

During the tasks, the screen will be divided into two parts:

- On the left: source text and space to type the target text;
- On the right: Google Chrome Internet browser.

Please do not move any of the above windows, change their size, or minimise them.

Don't edit the source text or change its formatting (don't click on the "ribbon" in MS Word).

When you finish the translation/post-editing **type the letter "f" at the end of the document (without any additional line breaks)** and say to the researcher: "Done!". After typing "f", do not click on anything, do not save the file, do not close any windows.

If you have any questions to the researcher during one of the task parts, call the researcher, but don't stand up.

Appendix G: Informed consent form

Zgoda na udział w badaniu naukowym do projektu doktorskiego

Osoba odpowiedzialna: mgr Olga Witczak, Wydział Anglistyki,
Uniwersytet im. Adama Mickiewicza w Poznaniu
Kontakt: owitczak@wa.amu.edu.pl
Opiekun naukowy: dr hab. Bogusława Whyatt, prof. UAM

Oświadczam, że zaznajomiłem/-am się i zrozumiałem/-am informację dla osoby badanej. Wyrażam dobrowolną i świadomą zgodę na udział w badaniu. Jestem również świadomy/-a faktu, iż w każdej chwili mogę odstąpić od udziału w badaniu.
Wyrażam zgodę na przetwarzanie moich danych osobowych uzyskanych w trakcie eksperymentu, zgodnie z ustawą z dnia 29 sierpnia 1997 r. o ochronie danych osobowych, *lecz wyłącznie* w celach naukowych.
Niniejszy dokument, potwierdzający zgodę na udział w badaniach będzie przechowywany zgodnie z zasadami przechowywania dokumentacji poufnej.

.....
Imię i nazwisko badanego

....
Podpis badanego data

Oświadczam, że osoba badana zapoznała się z informacją dla uczestnika badania, a dane uzyskane podczas eksperymentu będą przechowywane oraz przetwarzane zgodnie z ustawą z dnia 29 sierpnia 1997 r. o ochronie danych osobowych.

....
Podpis przeprowadzającego badanie data

Appendix H: Translation and post-editing briefs

(1) HEA translation

INSTRUKCJA: TŁUMACZENIE

Biuro tłumaczeń przysłało zlecenie na polską wersję fragmentu opisu produktu, który później zostanie opublikowany w Internecie. Przetłumacz fragment na polski.

Translation:

INSTRUCTION: TRANSLATION

Translation agency hired you to translate a product description into Polish. The text will be later published online. Translate the text fragment into Polish.

(2) APP translation

INSTRUKCJA: TŁUMACZENIE

Biuro tłumaczeń przysłało zlecenie na polską wersję fragmentu tekstu medycznego, który później zostanie opublikowany w Internecie. Przetłumacz fragment na polski.

Translation:

INSTRUCTION: TRANSLATION

Translation agency hired you to translate a medical text into Polish. The text will be later published online. Translate the text fragment into Polish.

(3) WAT post-editing

INSTRUKCJA: POST-EDYCJA

Biuro tłumaczeń przysłało zlecenie na polską wersję opisu produktu, który później zostanie opublikowany w Internecie. Dokonaj post-edycji przetłumaczonego maszynowo fragmentu.
Wskazówki dotyczące post-edycji tłumaczenia maszynowego:

- Postaraj się, aby tłumaczenie było poprawne pod względem gramatycznym, syntaktycznym i semantycznym.
- Upewnij się, że żadne informacje nie zostały dodane ani pominięte.

- Skoryguj wszelkie obraźliwe, niestosowne lub kulturowo nieodpowiednie fragmenty.
- Wykorzystaj jak najwięcej tekstu wygenerowanego maszynowo.
- Przestrzegaj podstawowych zasad ortografii, interpunkcji i dzielenia wyrazów.

Translation:

INSTRUCTION: POST-EDITING

Translation agency hired you to post-edit a medical text into Polish. The text will be later published online. Post-edit the machine translation of the fragment into Polish.

Here are tips on how to post-edit machine translation:

- Aim for grammatically, syntactically and semantically correct translation.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling, punctuation and hyphenation apply.

(4) MMR post-editing

INSTRUKCJA: POST-EDYCJA

Biuro tłumaczeń przysłało zlecenie na polską wersję fragmentu tekstu medycznego, który później zostanie opublikowany w Internecie. Dokonaj post-edycji przetłumaczonego maszynowo fragmentu.

Wskazówki dotyczące post-edycji tłumaczenia maszynowego:

- Postaraj się, aby tłumaczenie było poprawne pod względem gramatycznym, syntaktycznym i semantycznym.
- Upewnij się, że żadne informacje nie zostały dodane ani pominięte.
- Skoryguj wszelkie obraźliwe, niestosowne lub kulturowo nieodpowiednie fragmenty.
- Wykorzystaj jak najwięcej tekstu wygenerowanego maszynowo.
- Przestrzegaj podstawowych zasad ortografii, interpunkcji i dzielenia wyrazów.

Translation:

INSTRUCTION: POST-EDITING

Translation agency hired you to post-edit a product description into Polish. The text will be later published online. Post-edit the machine translation of the fragment into Polish.

Here are tips on how to post-edit machine translation:

- Aim for grammatically, syntactically and semantically correct translation.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling, punctuation and hyphenation apply