Cognitively Difficult Phrases in Post-editing from English to Finnish: A Comparison of Two Analysis Methods

Tommy Nylund

Master's Thesis

Degree Programme in Multilingual Translation Studies, Department of English
School of Languages and Translation Studies
Faculty of Humanities
University of Turku

November 2023

Master's Thesis

Degree Programme in Multilingual Translation Studies, Department of English
Tommy Nylund
Cognitively Difficult Phrases in Post-editing from English to Finnish: A Comparison of Two Analysis Methods
Number of pages: 55 pages, 13 pages of appendices

Post-editing machine translations have become more common in the translation field in recent years. Studies in this field have examined the types errors post-editors have to correct, what sort of errors machine translations produce, and how useful machine translations are, but there is far less research on post-editing time and cognitive strain it produces.

In this paper, an analysis of cognitive strain in post-editing is presented where the method Choice network analysis is used together with the data processing tool Translog-II to explore, how cognitive difficulty manifests with five professional translators who post-edit the same English to Finnish text.

The different post-edits made by each participant of the same word or phrase in various sentences are examined with the Choice network analysis and the Translog-II tool measuring the time it took to post-edit these words or phrases, and then comparing these results with each other.

The results were mixed, where in 6 cases the two analysis methods showed similar cognitive difficulty and in 7 cases, they showed differing results of each other. These findings suggest that these two analysis methods might not be best suited to work by themselves but could yield better results when used together.

With Choice network analysis it is possible to analyse multiple possible post-edits with possible cognitive difficulty and with Translog-II it is possible to explore the time spent on the task and see the various alterations the post-editors make.

Key words: machine translation, post-editing, post-editing time, professional translators, choice network analysis, Translog-II, process analysis, cognitive strain

Table of contents

List of figures, tables, and abbreviations.

List of figures, tables, and abbreviations.

Abbreviations

CNA	Choice network analysis
MT	Machine translation
NMT	Neural machine translation
SMT	Statistical machine translation
RBMT	Rule-based machine translation
PE	Post-editing
SL	Source language
ST	Source text
TAP	Think-aloud protocol
TL	Target language
SKTL	The Finnish Association of Translators and Interpreters

1 Introduction

In the fast-paced world of language and technology, the field of translation has undergone a significant transformation. The advent of machine translation (MT) systems has brought about new challenges and opportunities, leading to the emergence of post-editing (PE) as a crucial process in translation studies. The ISO standard 18587 defines post-editing as “the editing and correction of machine translation output” (ISO/DIS 18587:2016).

PE is a crucial step in the translation process, where professional translators review and revise the machine-generated translations to produce a final, high-quality text. While the benefits of MT and PE are evident in terms of time and cost savings, it is essential to understand the cognitive demands of this task to optimise its efficiency and effectiveness (O'Brien, 2011).

As the demand for PE continues to grow, understanding the cognitive difficulty involved in this task becomes very important. Professional translators play a pivotal role in ensuring the quality of post-edited translations, making it essential to explore the cognitive processes and decision-making strategies employed during the PE process (Ketola, 2016).

In traditional translation practices, human translators have been relied upon to bridge communication gaps across different cultures. However, the introduction of MT systems has revolutionised the translation industry, offering rapid and automated translation services. Despite these advancements, MT output often requires human intervention to address linguistic nuances, contextual ambiguities, and cultural variations, prompting the need for PE (O'Brien, 2005).

PE is seen as a vital intermediary step, allowing for the refinement of MT output by human translators. While it promises efficiency, the cognitive demands of PE have raised questions about the decision-making processes that translators employ (Ketola, 2016). To maximise the potential of human-machine collaboration in translation, it is essential to visit the cognitive dimensions of PE, enabling us to identify factors that influence translators' choices (O'Brien, 2005).

The Choice Network Analysis (CNA) method, as proposed by Campbell (2000a) emerged as a promising method to investigate the cognitive aspects of PE. By identifying phrases where multiple translators made different changes to the MT output, the CNA method allows us to pinpoint segments that may pose cognitive difficulty for translators. These identified segments serve as focal points for understanding the decision-making strategies and potential

challenges faced by translators during PE (Ketola, 2016). CNA somewhat uses the cognitive psychology theory of problem-solving and decision-making, which is required during the translation process, for translators to evaluate various options before making a final choice (O'Brien, 2006). This approach allows researchers to study the decision-making patterns of post-editors and provides valuable insights into the cognitive difficulty associated with certain linguistic constructs.

Nonetheless, researchers have employed the use of Translog-II, a tool that records the keystrokes and time taken during the translation process, to gain a comprehensive understanding of the cognitive aspects of translation and PE (Schou, Dragsted, and Carl 2009) more than they have used CNA. Translog-II records the keystrokes and mouse movements of the translators during the PE task, offering valuable insights into their cognitive behaviour and attention allocation. By analysing the keylogging data, researchers can identify specific areas of the text that pose cognitive challenges to the translators, shedding light on the cognitive processes involved in PE (O'Brien, 2011).

Moreover, studies have shown that PE is influenced by various factors, such as the translator's level of experience, familiarity with the MT system, and the nature of the source text. These factors can significantly impact the cognitive difficulty of PE and can lead to variations in the PE process among different translators (O'Brien, 2006).

By investigating the cognitive difficulty of PE and understanding the decision-making processes of translators, we can develop effective training programs and tools to enhance the PE efficiency and quality of the final translations. Moreover, this research contributes to the broader field of translation studies, as it offers valuable insights into the evolving role of human translators in the age of automation and A.I. To address the cognitive complexity of PE, this thesis aims to investigate the cognitive difficulty encountered by translators during the PE process. Specifically, we focus on comparing the results obtained through CNA and Translog-II to explore the decision-making patterns and cognitive processes involved in PE.

Concurrently, Translog-II analysis will provide us with valuable information on the time taken to complete the PE of specific phrases and sentences. Longer PE times may indicate higher cognitive difficulty in comprehending or translating the source text, while shorter times may suggest that the PE process was relatively straightforward.

By triangulating the results from the CNA method and Translog-II, we can obtain a comprehensive understanding of the cognitive challenges faced by translators during PE. The combination of these two methods allows us to explore the relationship between cognitive difficulty and time taken to PE specific segments. Moreover, it enables us to discern whether certain decision-making patterns or cognitive processes contribute to efficient or cumbersome PE (O'Brien, 2005).

Furthermore, by analysing the CNA and Translog-II data in tandem, we can identify instances where the two methods diverge in their assessment of cognitive difficulty. Such discrepancies might offer valuable insights into the intricacies of PE and the factors that contribute to cognitive effort during this process. It is crucial to investigate why certain phrases show agreement between CNA and Translog-II in terms of cognitive difficulty, while others exhibit differences. Understanding these variations can shed light on the interplay between translators' decision-making and their PE behaviours.

In conclusion, this thesis investigates the cognitive difficulty associated with post-editing in machine translation, utilising the CNA method and comparing it with Translog-II in analysing the post-editing process. By exploring the decision-making process of translators and analysing their PE times, we gain valuable insights into the cognitive aspects of this essential translation step.

The thesis structure is as follows; in chapter 2 the focus is on background and related works that have been studied in regards of NMT, CNA and the cognitive strain of post-editing. Chapter 3 describes in detail the data collection, the participants, and the material. In chapter 4 the findings of this study are examined, analyses of the cognitively most difficult words/phrases with the CNA and Translog-II methods are made, and afterwards are the results. Chapter 5 is a discussion regarding the results of the previous chapter. These results will be examined by comparing them to previous research on the same topic. This thesis is concluded in chapter 6 by revisiting the objective and the results of this paper and suggesting possible future research in the translation fields. As a final note, chapters 1 Introduction and 6 Conclusion were written with the help of Chat-GPT 3.5 by using it to make the linguistic style of these chapters more appropriate for the thesis.

2 Background and Related Work

This section focuses on different background and related works that have been studied across the translation fields over the recent decades. Some of these studies have pioneered for translation, MT, or PE studies while some studies are extensions for these pioneered studies. Nonetheless, all of the research here has contributed to the translation field in general.

The subsections go as follows, first is the 2.1 NMT and the quality of NMTs are visited over the last two decades starting with Papineni et al. (2002) and their BLEU automated metrics, which can be argued to have launched a number of these studies. The second section 2.2 explores the CNA method which examines the cognitive difficulty of certain words or phrases in PE. Lastly, in section 2.3 PE is visited in greater detail and how cognitive strain affects translating and PE itself.

2.1 Neural machine translation quality

In the last two decades, the use of MT in general has seen a rapid increase. Currently, the most recent MT system, NMT, has gained a lot of attention in a rather short amount of time. It has been shown to outperform other systems, the statistical machine translation (SMT) and the rule-based machine translation (RBMT), regarding automatic scores and human evaluations. In some previous studies, automatic metrics such as BLEU (Papineni et al. 2002) has been used to compare MT output to translations made by humans. Various other studies have then shown, how the evaluation of translation quality has improved by using the automatic evaluation metrics (Junczys-Dowmunt, Dwojak, and Hoang 2016; Bahdanau, Cho, and Bengio 2016; Bojar et al. 2016). However, there is also research that suggests that these automatic metrics may not always be the best methods for evaluation. In their study, Shterinov et al. (2018) did a comparison of automatic evaluation and human evaluation, in which they noticed that the automatic scores underestimated the quality of NMT. While some of these studies show pros for NMT and some show cons like was mentioned above, other studies show mixed results such as Castilho et al. (2017). Yang et al. (2018) proposed two automatic evaluation metrics Otem and Utem to evaluate performance of NMT systems and human evaluation metrics, with mixed results for both metrics. Post-editing a NMT has also been reported to be faster than post-editing non NMT translations for some domain-specific texts, showing that NMT has great potential for future MT (Jia, Carl, and Wang 2019: 68).

Although NMT and MT in general are getting better every year (Way 2018) and in some special cases can surpass even human translators (Popel et al. 2020), some languages are, nonetheless, problematic. Languages that are morphologically diverse and rich like Finnish, are still proving to be difficult. In the last few years, NMT in Finnish has mainly seen use in academic settings (Östling et al. 2017; Mi, Xie, and Zhang 2020; Koponen, Salmi, and Nikulin 2019; Salmi and Koponen 2020), although its use is on the rise as Finnish was included in the eTranslation system of the European Commission (2018) as one of the first languages to utilise the new NMT system in 2017, and in fact the usage of MT post-editing has been on the rise (ELIS 2018). Translating to and from Finnish has not seen improvements in quality in a long time compared to some other languages. Some of these problematic characteristics in Finnish are linked to the rich inflectional morphology and the long compound words (Tiedemann, Ginter, and Kanerva 2015; Grönroos, Virpioja, and Kurimo 2017). Currently there are very few studies that examine NMT quality with Finnish as target language, but two studies were found that both show results where NMT was used to improve the quality of the translations and reduce errors (Toral and Sánchez-Cartagena 2017; Mi, Xie, and Zhang 2020).

2.2 Choice network analysis

The CNA is a method proposed by Stuart Campbell (2000a; 2000b) to study the translation process. In this method, all the different translations of the same word or phrase from different translators are compared together. This way it is presumed that, if a word or phrase has been translated the same way by multiple people it had required less cognitive processing. And on the other hand, when a word or phrase has multiple different translations it is cognitively challenging (Campbell 2000b). These various translations can be imagined to be different branches reaching out from a tree: more branches that form from each additional branch, more choices for the translator to decide and the more difficult the word or phrase is. In this way, every translation used forms a “net” which has all the alternatives in it from all the translators. In theory, this method should demonstrate the amount of possible translations each translator has at their disposal (Campbell 2000a, 215) but some studies are arguing against this. In their study O’Brien (2005: 46–47) suggests that there are some problematic parts in this model: e.g. does a large number of translations actually indicate cognitive difficulty or could it show individual creativity and secondly, can we safely assume that every translator has access to all

of these options? In their study, Ketola also surmised that a large number of translations could be attributed to translator's *creativity* rather than *difficulty* (Ketola 2018: 70–71).

According to Campbell (2000a: 32), comparing versions of target texts produced by different translators can provide information about choices made during the actual translation process. This then can be used to form hypotheses from individual texts and language pairs to test in other texts and language pairs. Even though this method does indeed show which word or phrase one translator ended up finally choosing as the final translation, it is not possible to know if the translator actually used other choices for the word or phrase before committing to the final result. Pavlović, (2009: 86) O'Brien (2006) and Ketola (2016: 80) all agree that it is not possible to get an answer with the CNA if the translator had other alternatives but ended up rejecting those in favour of the last entered. O'Brien also adds to the list the issue of previously translated or post-edited segments, which then can influence the future segments in some way, hence needing attention later on (O'Brien 2006: 178).

Although both O'Brien and Ketola list issues with the CNA, they still give it credit and explain how it is still valuable in translation research (O'Brien 2005; 2006; Ketola 2016; 2018). Even if CNA has some issues by itself, it can yield great results in triangulation with other methods such as Translog which has been used by O'Brien (2006); Koponen, Salmi, and Nikulin (2019). The CNA has been used to examine translators' cognitive processes during translations (Campbell 2000a; 2000b) and this also includes studies made in Finland. Besides Anne Ketola, Finnish researchers Salmi and Koponen (2020) have studied machine translations that had been post-edited by 33 translation students from Finland, where the researchers analysed if the CNA method can help with the process of collecting PE data. In their study, Salmi and Koponen compared the differences between single and longer segments using process data from Translog-II and results of post-editing. They examined alternatives for the post-edited phrases which had been translated by one or more of the MT engines, phrases that have repetition and if there was variation between these machine translators (ibid.). This study had as many as 33 participants in it where each participant had one text to edit which had been translated by three different MT engines to form nine different sentences. The order of these sentences translated by the MT engines were changed between each participant. The previously mentioned study used a larger participant group, Campbell also states that using CNA yields more conclusive results with a larger group to analyse (Campbell 2000a: 32). Ketola, on the other hand, makes a case for smaller groups, so that the researcher would not get overwhelmed by the vast number of possible answers and the large scale

figures or tables (Ketola 2018: 66). Because the analysis method of CNA does not inherently dictate how the analysed items are formed, they can range from single words to longer phrases (Ketola 2018: 67) or they can be analysed on various levels from single lexical choices to larger syntactical structures e.g. translating *passives* (Hale and Campbell 2002: 18). In this study, CNA and Translog-II are compared against each other to examine both of their results.

2.3 Post-editing output and cognitive strain

There are two primary reasons to use PE and they are to correct the MT, meaning that we assume that there are, or will be, errors in the translation and the second is to make translating faster, hence increasing productivity. As MT and PE in translating languages gets increasingly more common, so do the studies around this topic as well. It can easily be seen through the amount of studies on the topic that PE errors and the productivity of PE are, and have been, of great interest in translation domains (Krings 2001; O'Brien 2005; Temnikova 2010; Koponen et al. 2012; Nunes Vieira 2017; Koponen, Salmi, and Nikulin 2019). It can perhaps be argued that PE is the result of MT to check for translation errors. Hence, the idea of PE is for the post-editor to make changes, for the text that has been machine translated, they deem necessary to make the text coherent and grammatically correct. However, there are also studies which observed post-editors making unnecessary changes, such as de Almeida who examined post-editors making unnecessary "preferential" changes ranging from 16% to 25% (2013: 100), and Kopoonen's study gave results which showed 38% of light PE[1] done, was unnecessary (2017: 146). The light PE is of interest for my current study as well, as the participants were instructed to also light PE their texts.

The second interest in PE studies, which was already mentioned is productivity and the cognitive strain PE produces. In one of the earliest studies on this subject, Krings (2001) aimed to examine how much effort was required by traditional translating and by PE, while also arguing that the feasibility of PE should not be determined by comparing it to human translation and its processing time alone. In a later study, O'Brien (2011) has later added to this that PE productivity does not only mean the ratio of quantity and quality to time but also how the cognitive efforts have been spent on this i.e., the more effort has been used, the lower the productivity. PE can have a tremendous impact on productivity if the MT is well suited

[1] Light post-editing stands for making only minimal essential changes during post-editing according to TAUS (2010) and the International Standard ISO 18587 (2017).

for the task at hand, as it can reduce the keyboard time by up to 70% and pause time by 31% in some cases (Plitt and Masselot 2010: 14).

The cognitive strain of some of the post-edited words or phrases is of great importance for the current study, as we are later comparing these items through two separate methods. In their study, Temnikova (2010) proposed an update to the standard error classification system for MT by ranking the error categories according to 10 levels of cognitive effort by their expected cognitive difficulty. Temnikova also explored how some words can potentially be cognitively more challenging than others depending on the length of the machine translated sentence (ibid.). These cognitively difficult words are the ones we are also interested in the current study, although they will not be measured in the same way as Temnikova have done. In a later study, Koponen et al. (2012) examined PE time as a measure of cognitive effort and demonstrated that, although the amount of errors matters for the PE effort so does the type of error such as punctuation and word order errors. This way it was concluded that grammar was examined to be more difficult than lexis, for example.

Another common practice in translation and PE studies is the use of think-aloud protocols (TAP) and eye tracking. Nunes Vieira (2017) used both of these methods in their research to study the cognitive difficulty of PE but they, on the other hand, did not find a significant difference between lexical and grammatical cognitive effort. Although, it was hypothesised that due to the nature of the instructions for the study, where they were instructed to aim for high-quality, this may have triggered the post-editors to choose semantically closer lexical words over others, which is to be expected for higher quality translations (Nunes Vieira 2017). In the current study, no real hypothesis is made regarding the actual cognitive difficulty or errors made in the PE but rather, an examination is made if two different methods show similar cognitive difficulties for the same words or phrases in PE. But if any hypothesis is to be drawn: since the other method uses real-time data from pauses, keystrokes and mouse clicks while the other only shows the last translation, it could be argued that the real-time data could potentially show more cognitively difficult items.

Regarding post-editing and cognitive strain, one potential explanation for the differences could be the concept of "creative freedom" during PE, as highlighted by Ketola (2016). Translators may exercise their judgment and make various edits to explore alternative ways to convey the meaning of the source text. This aspect is particularly relevant when considering the nature of machine translation output, which might require significant adjustments to meet

specific translation preferences, style, or domain-specific requirements. Consequently, even when multiple translators make different changes to the same segment, it does not necessarily imply cognitive difficulty but rather reflects the translators' efforts to achieve a preferred translation output.

3 Materials and Methods

In this section, the present study is described in better detail with the materials and methods used. The sections are written as follows: Section 3.1 will go into detail how the data were collected. Section 3.2 explains how the subjects were gathered for the experiment and it also explains the participant's background. In section 3.3 the material used for this study is explained in more detail, what method was used to examine it with an example and finally the actual data analysis method is explained.

3.1 Data collection methods

The current study had one primary method to gather data which happens during the actual translation process. The primary data collection method to gather process data, Translog-II, is introduced first with a brief history of the program and afterwards the complementary background questionnaire is introduced. This data in question, was collected in November 2019 during a training session on post-editing for members of the SKTL.

3.1.1 Translog-II

In this study, a computer-based, data collecting, keystroke logger called Translog-II was used to gather data from the translators. Translog was originally created in 1995 for translation process research (TPR) by Arnt Lykke Jakobsen, to help study the writing process in translation. Since Translog's first debut, it has seen extensive use in translation process research starting from 1999 (Hansen 1999), where some of Translog's first empirical uses started to surface. Ten years after its release, Translog was still extensively used in translation studies, which is shown by the article Ten years of Translog (Schou, Dragsted, and Carl 2009) that states that Translog has been widely used in translation training and translation studies. Similar, up-to-date studies about the usage of Translog was not found. Since then, Translog has been updated to newer versions such as Translog 2006 and the current Translog-II that is also used in this paper.

At its core, Translog-II software is a keystroke logger that records both keyboard and mouse activity, without disturbing the translating process. It allows the researcher to examine the translation process in better detail than it would be possible with only the final text files of the translation. The software itself uses two applications to function which are Translog-II Supervisor and Translog-II User. The Translog-II Supervisor function is used to control the

actual project meaning it creates, modifies, replays, and analyses the projects. Translog-II User functions as the Translog-II test environment that the subjects can use, however the User function is not of importance for this study.

The Translog-II session was done by using the Translog-II Supervisor mode. As the files used in Translog-II can only be opened by the software itself, Translog-II was downloaded from the Center for Research and Innovation in Translation and Translation Technology (CRITT) from where it can freely be downloaded for academic purposes. To start examining the log files, they were opened in the Translog-II Supervisor mode by finding the correct translator's file to be analysed from the computer. Next, the Replay function opens a new window that allows the researcher to play, pause, and adjust the speed of the log file that has the translation. After this, from the Supervisor window, the *tools* dropdown menu is clicked from which *Linear view* is opened. This lets the researcher see the whole translation process from start to finish in a view, that shows every keystroke, mouse click, and pause done in the translation process (more on this in section 3.3.2).

3.1.2 Background questionnaire

Depending on what you are researching, questionnaires can be used as the main data gathering tool to sample a large pool of subjects, filter the sample pool to a smaller size, to gain a better view of the research topic in hand, or to gain background information of the subjects who are participating in your study. Regardless of the purpose of a questionnaire, it needs to be carefully designed to fit the task at hand and in some cases to keep the anonymity of the participants.

The purpose of the questionnaire for this study was to gather background information of the participants. The questionnaire was written in Finnish, and it was distributed on paper to the participants who then filled it in with a ballpen. The participants also signed a consent form to use the data collected from them for research purposes (Appendix 1). The filled questionnaires were later turned into pdf files and sent to me via email. The questionnaire included the following background questions: age, gender, native language, self-estimation of English and Finnish language skills if not native language, level of education, learned languages, work experience, and experience in PE machine translations.

3.2 Subject selection and participants

The data for this study was gathered in November 2019 in a monthly meeting of the local branch in Turku of the Finnish Association of Translators and Interpreters (SKTL) by Maarit Koponen and Leena Salmi. These meetings are held approximately every month, and this one was held in an IT-classroom at the University of Turku. Members of the SKTL will get an e-mail regarding these meetings and depending on the type of meeting, the members will need to sign up before hand for them. In this meeting, researchers Koponen and Salmi held a brief training lesson in machine translations and afterwards the participants were able to try machine translation and PE for themselves.

Five professional translators of various languages took part in the experiment, and they are introduced in this section based on their answers on the questionnaire. To keep their anonymity, the translators are referred here as Tr1, Tr2, Tr3, Tr4 and Tr5. The sampling of this group is primarily heterogeneous, but there are some factors which are homogeneous. For example, the group has one translator who belongs to the age group of 31 to 40, one who belongs to group 41 to 50, two in age group 51 to 60, and one who belongs to group 61 to 70. The subjects had also varied levels of experience in years for translating full time their chosen languages ranging from 7 years to 25 years. The homogeneous part comes primarily from their education and language skills, and all of the participants identify themselves as women. All the translators have a Master's degree in languages or translations, and everyone has experience in translating from French to Finnish. All five of the subjects have also studied at least one additional language in addition to their main language. Finally, regarding experience with machine translations, Tr5 was the only one who has had any experience with machine translations and had worked on assignments which used it.

Every participant's L1 was Finnish and all of them translate from or towards French. Table 1 below shows all working languages of the participants. It is worth noting, that when these translators were studying in the University of Turku, they graduated from Major subjects that are no longer available such as Romance philology or Translation studies. Three of the translators were studying translation for a specific language and two were studying philology/languages.

Table 1. Participants study subjects and translation directions.

	L1	Other working languages
Tr1	Finnish	Spanish
Tr2	Finnish	Italian
Tr3	Finnish	-
Tr4	Finnish	-
Tr5	Finnish	Swedish, Russian

Work experience saw more variation, which was to be expected based on the participant's age range. Table 2 shows that most of the translators have worked only full-time with Tr2 being the only one who has had experience working part-time. Most of the translators have worked at least 20 years with Tr2 maybe working as much, but the questionnaire did not specify if they have worked simultaneously or separately on their various languages. Tr5 being the one exception who has worked less than 10 years, and they were also the only one who had worked on occasional individual assignments for other languages, but the questionnaire did not specifically ask on how many assignments. Finally, it was not asked directly which languages the participants were translating to or from during their years as a professional translator.

Table 2. Participants experience working as translators.

	Experience translating English - Finnish		Experience translating to other languages		
	Full-time	**Part-time**	**Full-time**	**Part-time**	**Occasional individual assignments**
Tr1	-	-	25 years	-	-
Tr2	11 years	-	-	11 years	-
Tr3	20 years	-	25 years	-	-
Tr4	-	-	20 years	-	-
Tr5	7 years	-	-	-	Has made some

3.3 Study material

In the present study, the data consists of PE process data collected using Translog-II, as well as the post-edited MT sentences separated from the final text produced by each participant (Appendices 4 – 8). The next subsections will include an explanation of the actual data for this study in 3.3.1, an example of the primary data used from Translog-II to better explain what has been analysed for this study in 3.3.2, and the data analysis method used in 3.3.3.

3.3.1 Present study data

At the meeting, the five participants post-edited this translation during their training session at the University of Turku with the best of their knowledge without any aid from external sources such as internet or dictionaries. The source text was a 331-word, short English news article about an international smuggling network being shut down, which was translated to Finnish (256 words) using the European Commission's NMT system called eTranslation (2018). The text had 18 sentences that were later made into 18 segments to make the analysis easier in the excel tables. Because these final text files were TXT files and not Microsoft Word Document files, they had to be encoded by Microsoft Word into a readable format. This was done by double clicking the TXT files that opened an encoding window in Microsoft word, in which the default Windows encoder (Western European) was chosen. This was done, if these TXT files were needed at any point, although the Translog-II files had the same information already. This was done because it was easier to read the and compare the texts in Microsoft word than it was in notepad. The source text and the MT can be found in the appendices 2 and 3.

For the PE task, the participants were instructed to "light" post-edit the text as is defined in the draft International Standard ISO/DIS 18587:2016. The following instructions were given in Finnish to the participants:

- Make use of the raw machine translation as much as possible.
- Aim to produce a translation that conveys the correct meaning and follows correct Finnish grammar.
- Check that there is no extra information or missing information in the sentence.
- Fix sentence structure if the meaning is incorrect or unclear.
- Follow spelling and punctuation conventions

As of writing this thesis, the final ISO standard has later been updated and published again in 2017 with few modifications. Nonetheless, the instructions used in the experiment are still valid according to this ISO standard.

3.3.2 An example of Translog-II data

Translog-II is used to record several different types of information from translation sessions, as was stated in the previous subsection 3.1.1. It records all keyboard activity and some mouse activity, as well as inactivity that occurs between keyboard and mouse usage or different actions that happen outside of Translog-II. Translog-II uses two main components, Translog-II Supervisor and Translog-II User, but for this study only they Supervisor mode was used. The Supervisor mode can be used to view the recording as a video, as well as see a log file of the process.

The linear view shows the whole translation session from start to finish, with all keystrokes, mouse clicks, and inactivity that happened during this time. Figure 1 shows an example of a log file. The log shows, different keystrokes, mouse clicks, and various (in)activities are shown by different symbols. Keyboard activity is represented by black dots which represent spaces, triangles pointing left represent backspaces and the actual letters written are their corresponding keys on the keyboard. Cursor movement is shown with two consecutive triangles pointing downwards and upwards. Inactivity or pauses, are shown either as red dots marking 1 second pause times for each dot or red numbers showing the exact pause time.

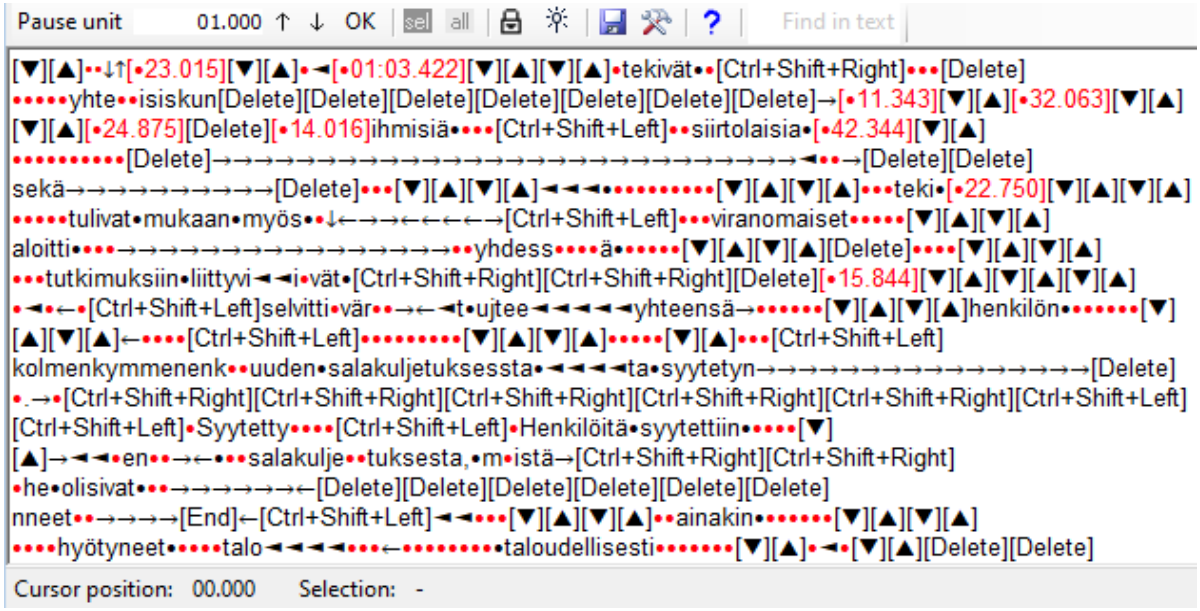

Figure 1. An example of Translog-II linear view data.

3.3.3 Identifying post-editing changes

To identify the PE changes made by the participants, a somewhat similar approach to Koponen, Salmi and Nikulin's study was used (Koponen, Salmi, and Nikulin 2019) where they used TER-plus with the basic H(TER) turned off to identify word-level changes edited by the subjects. However, the current study did not use the morphological analyser OMorFi and FinnPos morphological tagger toolkit as they did, as there were next to zero cases in the words and phrases analysed in this study that would have required lemmatising or stemming in larger scale. Hence, lemmatisation was checked manually to observe if there were any possible errors or homonymic words. In some situations, the participants made changes halfway during their post-editing and in these cases these changes were annotated separately.

The MT and PE texts were added to an Excel sheet manually one sentence at a time, by copying the sentences from the TXT files to their own Excel sheet per sentence, totalling 18 sheets. These sheets were arranged in a way where, the SL sentence being on the first row on each sheet, MT version on the second row and five PE versions on the next five rows. Every sentence was then divided into smaller segments if they contained three or more different post-edits made by the participants. This is an example:

Tr1 and Tr2 edited a word or phrase exactly the same way this counts as 1 edit for the purpose of this study, then Tr3 and Tr4 edited the previously mentioned word or phrase differently than Tr1 and Tr2 this counts as 2 different edits, and finally Tr5 edited the same word or phrase differently from the rest this now means we have 3 different edits which equals to three different post-edits regardless of the number of different post-editors.

Then segments were manually added to their own column according to their post-editors' row. Finally, 7 of these sentences were picked which contained a total of 13 segments for this study. Of these segments 7 were phrases with multiple words and 4 were single words. Following a somewhat similar approach (Koponen, Salmi, and Nikulin 2019: 72–73), specifically the 4th sentence was divided once more to smaller sub-segments because it contained several meaningful units, such as NPs and VPs based on their value for the current study. These segments were, once again, manually aligned with their annotations.

In some cases, the participant made changes that did not appear in the final edition, such as adding or changing the word halfway and later adding the translation. These cases were annotated separately according to their editing time as was mentioned above. The segments

were aligned according to their lemma being the same in the MT and PE, e.g., *another driver* (SL), *toinen kuljettaja* (MT), *yksi kuljettaja* (PE) where the words *another, toinen* and *yksi* were in the same column and the words *driver* and *kuljettaja* in another as they had the same lemma. Aligning was then continued by adding, again manually, the corresponding annotations such as insertions, deletions, and substitutions in the next column. No automatic annotation was used rather, a similar manual annotation that Koponen, Salmi and Nikulin used in their study (2019) was also used in the current research. In the current study, if a new word had been added to the PE which did not have any equivalent in the MT was this classified as *insertion,* and on the other end if a word had been deleted from the MT which did not have an equivalent was this counted as a *deletion*. Other cases that were noticed were situations where word form or order had been changed. Sometimes a word or phrase can be affected by multiple types of edits. These situations usually happened, when a word had its form changed and another word had been deleted or inserted. In these cases, annotations were marked with multiple actions (inserted / form changed, deleted / word changed, etc.).

These annotations are from Koponen, Salmi and Nikulin (ibid.) and they were used in the current study as well. The edits made by the participants were categorised with one or more of the following PE actions:

– unedited: no change;
– form changed: different morphological form;
– word changed: different lemma;
– deleted: word removed;
– inserted: word added;
– word order: position of a word changed (Koponen, Salmi, Nikulin, 2019)

In some cases, it is not possible to just analyse one or two words which are right next to each other, there are some cases where there are longer phrases in which the word order has changed drastically due to the Finnish morphology. As Ketola mentions in their study (2018: 67 – 68), in Finnish it's possible to write an entire prepositional phrase with one word. Based on this, in the current study there are some segments which will include larger units where multiple individual words are chained together. In practice, this means that it was necessary to take these situations case-by-case, while considering how the Finnish case inflections will be taken into account.

3.3.4 Identifying pauses

A text production unit is needed to define a moment when actual text producing is happening. This unit is "defined as a sequence of continuous typing separated by pauses of 1 s or longer" (Carl, Bangalore, and Schaeffer 2015: 35). As other researchers have used this definition in their studies to separate continuous units of activities and pauses in translation research (Carl, Bangalore, and Schaeffer 2015; Nunes Vieira 2017; Koponen, Salmi, and Nikulin 2019) it was then selected for this study as well. These units were measured with Translog-II and its process data for each participant. Because the present study is interested in PE and how it connects to specific parts of the text, further dividing was necessary. Due to the nature of the present study, unlike the text production units from Carl et al. (2015), a similar approach was taken with Koponen, Salmi and Nikulin (2019) and their division of text production units where cursor repositioning – mouse clicks, arrow keys and other function keys – was added to the activity of defining text production units.

Because this experiment did not use any TAPs or eye-tracking systems, it became clear quite fast that identifying segment boundaries and when they have been crossed would become problematic for measuring cognitive effort. Measuring technical effort would not be as difficult as this can be evaluated in a few different ways, such as when the participant stops making changes to their current segment and starts on a new segment (O'Brien 2006: 139), or by calculating text production and deleting keystrokes from sentences (Koponen, Salmi, and Nikulin 2019: 73). From a cognitive point of view, it becomes difficult to tell without any aid from other sources, when does the participant actually start processing a new segment. It was then decided that cursor movement would mark the changing of segment boundaries. A similar method has been used by O'Brien (2005): the boundary was seen as crossed, when the cursor was moved to a new segment after finishing the previous one and it did not move back before making any changes to the new one. Nonetheless, this is not an ideal method as it is still impossible to know if the participant actually stopped processing this segment or edit in their mind. However, after using Translog-II in this analysis it was noticed that when the subjects moved their cursor to a new segment to post-edit, there were only three cases where they went back to edit something else. In this light, it was decided that this approach for the segment boundary crossing would be used in this study.

It has been established, that post-editing a segment starts when the cursor crosses the boundary line but there are also pauses during and between words or phrases. To enhance the

analysis of cognitive effort in PE, we also measured the time between consecutive edits of the same text production units or segments. The idea is that if pauses are happening during PE of a specific segment, it is safe to assume that the PE is happening for that specific sentence (Koponen, Salmi, and Nikulin 2019). Both of these methods, cursor movement to cross segment boundaries and mid-segment pauses, were seen as exceedingly useful ways to approach this analysis of measuring cognitive difficulties of PE, when no other external aid had been used.

4 Results

The aim of this study is to examine if the cognitive effort demonstrated by the CNA method yields the same results as does the process analysis produced with Translog-II logs between the five professional translators, when they are post-editing a machine translation. The following sections will demonstrate in detail the findings of this study. In the first section 4.1, Translog-II logs will be analysed by examining the pauses and the post-edited words/phrases analysed in section 4.1. Secondly in section 4.2, an analysis of the cognitively most difficult words/phrases will be made, with the CNA method. Finally, in section 4.3 the results of the previous sections 4.1 and 4.2 will be compared. Due to the scope of this paper and the limitations that come with it, 11 total segments were considered for this study. From those 11 segments 7 had phrases with multiple words and 4 were single words, and these were analysed with the CNA for this study. These segments were taken from sentences 4, 7, 8, 13, 14, 15 and 16 (sentences can be found in the appendices 4 – 8 with these sentences numbered accordingly).

4.1 Translog-II Analysis

In the tables of this section, the first column shows the translator, second column the PE, and the third shows the time used for PE. In the third column calculation, each separate instance with a number in it means the participant has stopped any keyboard or mouse activity for a period of time until continuing again. For example: 5s+1s= 6s means the participant took a 5 second pause after moving their cursor at the start of the phrase/word they start to edit and then continued editing for a period of time until taking a 1 second pause until they finished the PE taking a total time of 6 seconds to PE the phrase.

In this paper, the speed of the post-edits is compared against the other participants and the three categories below reflect that. The cognitive difficulty of each PE belongs to one of three categories. As of writing this thesis it was difficult to find indications of a standard definition for 'short' or 'long' pauses in previous research. Therefore, a scale with 15 second increments between each category was made for this paper. If a phrase would belong to two different categories (or three) due to different PE times, the difficulty category would be the higher of those. Two examples: example 1, if a phrase has four participants who post-edited it in 15 seconds or less, and one participant took more than 16 seconds, it would be categorised as 2. Example 2, if two of the participants would belong to category 1 and the other three

participants in category 2 the difficulty would be 2. In a case where a phrase has post-edits from category 1, none in category 2, but has post-edits from category 3, and at least three of the participants' post-edits are from the lower category, the category is lowered to 2 as most participants have still post-edited within the time of category 1. The scale used in the process analysis of this study has three categories:

- PE time was 1 – 15 seconds for every participant, making it somewhat difficult = 1
- PE time was 16 – 30 seconds for every participant, making it difficult = 2
- PE time was over 31 seconds for every participant, making it very difficult = 3

The first sentence under scrutiny is 4, in which there are a number of phrases that have multiple different post-edits done to them, making it the most cognitively difficult sentence in the whole text. This sentence had its segments divided into smaller sub-segments for the purpose of this study, as these phrases would otherwise have been difficult to analyse. The exact segments are used as the point is to compare these two methods if they show the same cognitively stressful post-editions.

In the 4th sentence and its phrase *further looked into* machine translation *tarkastelivat edelleen* from each participant in table 3 below. The tables below, with columns regarding the post-edit and the time spent post-editing a phrase, have the final edit of each participant and do not include the whole post-editing process.

Table 3. Time spent post-editing the phrase *tarkastelivat edelleen*.

Post-editor	PE	PE time
Tr1	tarkastelivat sen jälkeen	**3s**
Tr2	tutkivat perusteellisesti	5s + 1s = **6s**
Tr3	tarkastelivat	**6s**
Tr4	seurasivat	4s + 4s + 3s = **11s**
Tr5	selvittivät	1s + 1s + 1s = **3s**

Tr1 can be seen to have started editing rather fast after moving their cursor at the beginning of the phrase as it only took 3s for them to start editing. Tr2 and Tr3 took few seconds longer to start their process and ended it with 6s. Tr4 took a little while longer and they also had few

short breaks. Tr5 only took 1 second after each stop to finish their PE. Every participant post-edited in less than 15 seconds, category 1.

Next is the second phrase of the 4th sentence *the activities of 36 people* machine translation *36 ihmisen toimia* in table 4.

Table 4. Time spent post-editing the phrase *36 ihmisen toimia.*

Post-editor	PE	PE time
Tr1	sellaisten 36 ihmisen toimia	**28s**
Tr2	36 syytetyn toimia	13s + 5s = **18s**
Tr3	36 ihmisen toimintaa	9s + 9s + 10s = **28s**
Tr4	36 ihmisen toimia	-
Tr5	yhteensä 36 salakuljetuksesta syytetyn henkilön toimia	**16s**

Tr1 took 28s to edit this phrase and they added one word to it, making this a somewhat lengthier edit. Tr2 took 13s to make a first edit until having a small pause, for a total of 18 seconds. Tr3 had three pauses with similar times; 9 - 10 seconds each for a total of 28s as was with Tr1. This phrase was left unedited by Tr4. Tr5 took 16s to PE the phrase which was edited the most out of these. Four participants post-edited between 16 to 30 seconds, category 2.

The third phrase of the 4th sentence is *accused of smuggling at least 580 migrants* MT *joita syytettiin salakuljetuksesta vähintään 580 siirtolaista.*

Table 5. Time spent post-editing the phrase *joita syytettiin salakuljetuksesta vähintään 580 siirtolaista.*

Post-editor	PE	PE time
Tr1	joita syytetään ainakin 580 siirtolaisen salakuljetuksesta	56s + 7s = **1m 3s**
Tr2	joissa oli salakuljetettu vähintään 580 siirtolaista	13s + 5s = **18s**
Tr3	Heitä syytettiin vähintään 580 siirtolaisen salakuljetuksesta	3s + 8s = **11s**
Tr4	joita syytettiin vähintään 580 siirtolaisen salakuljetuksesta	2s + 2s + 5s + 1s = **10s**
Tr5	Henkilöitä syytettiin ainakin 580 siirtolaisen salakuljetuksesta	9s + 5s + 3s + 2s = **19s**

Tr1 took 56s to start editing after moving the cursor at the start of the phrase. They later took a 7s pause to continue editing for a total PE time of 1 minute and 3 seconds making this one of the lengthiest PEs. Tr2 took a somewhat lengthier pause at the start for 13s while having a short pause during the phrase. Tr3 had two shorter pauses for 11s while Tr4 took a similar time to edit 10s, although they had a total of four pauses. Tr5 also had four pauses but edited this phrase for 19 seconds. Vastly different PE times for every participant and one post-editor with over 31 seconds, category 3.

The final phrase of the 4th sentence is *earning* MT *ja jotka ansaitsivat.*

Table 6. Time spent post-editing the phrase *ja jotka ansaitsivat*.

Post-editor	PE	PE time
Tr1	ja jotka ovat ansainneet	1s + 1s = **2s**
Tr2	ja ansaittu	**1s**
Tr3	josta he olivat ansainneet	7s + 1s = **8s**
Tr4	Syytetyt ansaitsivat	**3s**
Tr5	mistä he olisivat hyötyneet taloudellisesti	**4s**

This phrase saw short PE times all around as only Tr3 took over 5 seconds to PE. Tr1 had one additional pause after starting to PE, Tr2 stopped only for 1s and moved on, Tr4 took 3s and Tr5 4s and finished their PE after that. Every participant post-edited in less than 15 seconds, category 1.

In the 7th sentence, there is only one longer phrase which was post-edited by multiple people *with financial and logistical support provided by Eurojust,* MT *jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti.*

Table 7. Time spent post-editing the phrase *jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti.*

Post-editor	PE	PE time
Tr1	Eurojust, joka tukee taloudellisesti ja logistisesti	3s + 2s + 2s = **7s**
Tr2	jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti	-
Tr3	Eurojustin taloudellisella ja logistisella tuella	1s + 1s + 5s + 1s + 2s + 2s + 2s + 2s + 2s + 2s + 1s = **21s**
Tr4	jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti	-
Tr5	jota Eurojust ja Europol tukivat taloudellisesti ja logistisesti	**10s**

This longer phrase was left unedited by participants Tr2 and Tr4. Tr1 took two short 2s pauses after starting to edit for 7s total. Looking at Tr3's PE a curious pattern can be seen where they took a total of 9 very short pauses 1 – 2 seconds each and one 5s pause in the middle of the PE. When looking at the Translog log data it was noticed that Tr3 took these pauses after each word and sometimes during editing of a word. Their total PE time then was 21 seconds. Tr5 also shows a somewhat different result from the other two as they stopped for 10s after moving their cursor and then edited the whole phrase in one go. Three post-editors where one took 16 – 30 seconds, category 2.

Next is the 8th sentence which also has only one phrase which was accepted for this paper *the action day* MT *toimintapäivän.*

Table 8. Time spent post-editing the phrase *toimintapäivän.*

Post-editor	PE	PE time
Tr1	toimintapäivän	-
Tr2	pidätystoimien	3s + 2s + 5s + 2s + 43s + 3s = **58s**
Tr3	paljastusiskun	(1m 16s) + 1s + 10s + 7s = 18s **(1m 34s)**
Tr4	toimintapäivän	-
Tr5	iskun	3s + 1s = **4s**

This sentence left the participants somewhat divided on the PE needs as there are varying results here. This phrase was left unedited by Tr1 and Tr4, but participants Tr2 and Tr3 on the other hand took their time. Tr2 started with a 3s pause after the cursor and took multiple pauses during the edit from which one of the pauses was 43s. The Translog data showed that the translator edited the word *toimintapäivän* three times before keeping the final form, *pidätystoimien*, and during this time they also went to a previous phrase to edit this word which took few seconds extra time from them. The results from Tr3 are a little bit complicated. It was noticed during the analysis of their Translog data numerous times that they put their cursor at the end of the word they are about to edit. This phrase was one of those cases, hence it was deemed acceptable to add these PEs to the results. The PE time in the first brackets is the time they spent pausing after they moved the cursor to the end of the word they were about to edit, and the final brackets include the total time they spent editing with the exception included. Tr3 moved the cursor at the end of the word they were about to edit and paused for 1 minute and 16 seconds before editing it. The Translog data also shows that the rest of the pauses were used to correct typos in the word. And finally, Tr5 took a very short time to PE this phrase or in this case, word. It was also noticed later in the Translog file that Tr5 wrote their thoughts down and was later wondering if the *iskun* is the correct translation. Vastly different PE times with the participants who post-edited the phrase where two took over 31 seconds to post-edit, category 3.

Sentence 13 only has one phrase to post-edit *pieces of identification* MT *tunnistetietoja.*

Table 9. Time spent post-editing the phrase *tunnistetietoja*.

Post-editor	PE	PE time
Tr1	henkilötodistuksia	1m 23s + 1s = **1m 24s**
Tr2	henkilöasiakirjoja	3s + 3s = **6s**
Tr3	henkilöpapereita	5s + 1s + 11s = **17s**
Tr4	tunnistetietoja	-
Tr5	tunnistetietoja	-

This word saw varied results in its PE times. Tr1 took 1 minute and 23 seconds before they started editing the word while Tr2 took a 6s total to edit the word. The Translog data shows Tr3 starting to write something else, then they delete the word, take a 11 second pause, and then write down the final form for a total of 17s. Participants Tr4 and Tr5 left this unedited.

Vastly different PE times with the participants where one participant took over 31 seconds to post-edit, category 3.

The 14th sentence has two phrases which were cognitively stressful enough for this paper. First is *In related actions* MT *Toisiinsa liittyvissä oikeudenkäynneissä.*

Table 10. Time spent post-editing the phrase *Toisiinsa liittyvissä oikeudenkäynneissä*.

Post-editor	PE	PE time
Tr1	Tähän liittyvissä	6s + 2s = **8s**
Tr2	Asiaan liittyvissä	**15s**
Tr3	Asiaa koskevissa	4s + 7s + 1s + 1s + 6s + 6s + 1s + 6s + 2s = **34s**
Tr4	Toisiinsa liittyvissä	-
Tr5	iskuun liittyvissä	**3s**

Tr1 takes a little time to start editing and takes a one very short pause for a total PE time of 8s. Tr2 takes a somewhat longer time 15s to start editing. Tr3 takes 4s to start editing and after that takes two pauses during the edit to presumably think what to edit (a pause that lasts 7s and the second 6s pause). The rest of the pauses happen because they go back and forth to fix spelling errors, for a PE time of 34s. Tr4 did not post-edit this phrase and Tr5 took only 3s to start editing. One participant took more than 31 seconds to post-edit

The second phrase of the 14th sentence is *were sentenced* MT *tuomittiin.*

Table 11. Time spent post-editing the phrase *tuomittiin*.

Post-editor	PE	PE time
Tr1	tuomittiin	-
Tr2	annettiin tuomiot	**3s**
Tr3	langetettiin tuomioita	2s + 3s + 7s + 1s + 2s + 2s + 2s + 2s = **21s**
Tr4	tuomittiin	-
Tr5	annettiin ... tuomio	2s + 1s = **3s**

This second phrase was left unedited by Tr1 and Tr4. Only 3s was spent at the start of the word by Tr2. Tr3 on the other hand took more time to PE this phrase but spent less time at the start. From Translog it was noticed that Tr3 removed a word from this phrase they had added and then fixed numerous times some letters from the word *tuomittiin* until the word form was

changed to *tuomioita.* Tr5 is a somewhat special case in this phrase as they changed the word order of the whole sentence making this analysis a little more different from the rest. The word *tuomittiin* was changed to have a main verb *annettiin* which was moved to the beginning of the sentence, and then they changed the word *tuomittiin* from a verb to *tuomio* a noun. As these words make the phrase *tuomittiin* it was accepted for this analysis. From the first word Tr5 started to edit this phrase, *tuomittiin*, they moved to the start of the sentence to add the word *annettiin* and afterwards they took a 1s pause before they moved the cursor to the second edited word *tuomio* for a total PE time of 3s. One of the participants took 16 – 30 seconds to post-edit, category 2.

Second to last sentence is number 15 which also has two phrases of which the first one is *Another driver* MT *Toinen kuljettaja.*

Table 12. Time spent post-editing the phrase *Toinen kuljettaja*.

Post-editor	PE	PE time
Tr1	Erästä kuljettajaa	3s + 1s = **4s**
Tr2	Yksi kuljettaja	**1s**
Tr3	Yksi kuljettaja	6s + 1s = **7s**
Tr4	Yksi kuljettaja	3s + 4s = **7s**
Tr5	yhden kuljettajan	5s + 1s = **6s**

Every participant edited this phrase in similar fashion with similar PE times. To start with, Tr1 edited the whole sentence altogether with only one pause. They took 3s to start editing and one pause that lasted 1s. After moving their cursor to the word Tr2 is about to edit they take 1s before the PE of this phrase is done. Tr3 moves their cursor once again to the end of the word they are about to edit, and they delete it after 6s and afterwards they take a 1s pause. Tr4 took a total of 7s to PE the phrase from which 6s was used to start editing. Tr5 edited the whole sentence in one go in a total of 6s of which 5s was used to pause mid edit. Everyone edited in less than 15 seconds, category 1.

The second phrase of sentence number 15 is *is standing trial* MT *on seisova oikeudenkäynti.*

Table 13. Time spent post-editing the phrase *on seisova oikeudenkäynti*.

Post-editor	PE	PE time
Tr1	vastaan on vireillä oikeudenkäynti	-
Tr2	odottaa oikeudenkäyntiä	3s + 5s = **8s**
Tr3	odottaa oikeudenkäynti	1s + 2s + 3s + 1s + 1s = **8s**
Tr4	odottaa oikeudenkäynti	2s + 2s = **4s**
Tr5	on meneillään … oikeudenkäynti	-

In editing this phrase, the participants used two different approaches, namely that three of them edited the phrase in a same way and the two others edited the whole sentence in one go in a different fashion. Participants Tr1 and Tr5 both have their PE time written as none, because they post-edited the whole sentence without any pauses during this phrase, so the PE time of table 14 can be used here. Tr2 spent 3s before starting to edit the phrase and took a 5s pause during it for a little longer PE time of 8s. On the other hand, Tr3 took multiple short pauses during their PE still totalling 8s. Lastly, Tr4 edited the phrase very fast taking 2s to start and having one 2s pause during it. Everyone edited in less than 15 seconds, category 1.

The final sentence for this paper is the sentence number 16 with two phrases, of which the first one is *the main organizer* MT *pääjärjestäjän.*

Table 14. Time spent post-editing the phrase *pääjärjestäjän*.

Post-editor	PE	PE time
Tr1	pääjärjestäjän	-
Tr2	salakuljetusten pääjärjestäjän	**25s**
Tr3	pääjärjestäjän	-
Tr4	pääorganisaattorin	2s + 1s = **3s**
Tr5	salakuljetusringin johtajan	**18s**

This phrase saw few longer PE times and was also left unedited by Tr1 and Tr3. The longest PE time for this phrase was from Tr2 for 25s which they took in one go after moving the cursor. Tr4 took a total of 3s to edit with one mid phrase pause lasting 1s. Tr5 took a longer time to edit, 18s with no mid phrase pauses. Two of the participants took 16 – 30 seconds to post edit, category 2.

Finally, the last phrase of this paper is *on a German European Arrest Warrant* MT *saksalaisesta eurooppalaisesta pidätysmääräyksestä*.

Table 15. Time spent post-editing the phrase *saksalaisesta eurooppalaisesta pidätysmääräyksestä*.

Post-editor	PE	PE time
Tr1	sakasalaisen eurooppalaisen pidätysmääräyksen perusteella	3s + 5s + 3s = **11s**
Tr2	Saksassa vahvistetun eurooppalaisen pidätysmääräyksen nojalla	17s + 1s + 9s + 1s = **28s**
Tr3	saksalaisella eurooppalaisella pidätysmääräyksellä	12s + 1s + 1s + 11s + 1s + 1s = **27s**
Tr4	Saksan eurooppalaisen pidätysmääräyksen perusteella	14s + 1s + 2s = **17s**
Tr5	Saksan eurooppalaisen pidätysmääräyksen perusteella	1s + 1s + 1s = **3s**

This last phrase is a longer one with post-edits made by every participant. This phrase included multiple words because all the post-editors edited this phrase in a similar fashion. Meaning, they all had to edit their phrases the way they did because all the words had to have case endings that fit all these words in this phrase. The first PE made here by Tr1 took 11s with two pauses which took 5s and 3s respectively. The first pause taken after the start is because the post-editor goes back to edit the first word in the phrase *saksalaisesta* because they had spelled it incorrectly, although they still ended up spelling the word incorrectly to *sakasalaisen.* Tr2 takes the longest to PE this phrase for a total of 28s of which 17s is taken after moving the cursor to the start of the phrase. The 9s pause in the middle is taken because the post-editor first writes *saksalaisen* but then goes back to change it to *Saksassa*. Tr3 used slightly less time to PE this phrase than Tr2 with 27s. Similarly, to Tr2, Tr3 takes some time after moving the cursor to the beginning 12s and they also have one longer pause during this PE for 11s. Tr4 also took a longer pause at the beginning for 14s and then proceeded to have two very short breaks 1s and 2s before finishing this phrase. Tr5 on the other hand took only 3s to PE this phrase. They had three 1s pauses of which the second pause was used to change the first word they wrote from *saksalaisen* to *Saksan.* This phrase was particularly curious because it saw three of the participants come back a second time and edit the very first word

of the phrase *saksalaisesta* at some point of their PE. Three of the participants took 16 – 30 seconds to post edit, category 2.

4.2 Results of Choice Network Analysis

In this section, some of the same words/phrases that were analysed in section 4.1, are shown with three CNA figures and the rest are written analyses illustrating the PE times of the phrase/word in question for the various translators. As the CNA method in this study uses TXT files, this analysis will use only the text files that were recovered after the translators had post-edited their Translog-II files. As was mentioned earlier in section 3.3.1, these files were the final version taken from the Translog-II log files. After the files had been encoded to readable format, the comparison between the MT and the translator's post-editions could begin. The cognitive difficulty scale used for the CNA in this study has three categories:

- at least 3 participants post-edit making it somewhat difficult = 1
- 4 participants post-edit, difficult = 2
- every participant post-edit, very difficult = 3

The first phrase under scrutiny is the *further looked into* which has been machine translated into *tarkastelivat edelleen.* This is the first phrase that can be seen as cognitively difficult with the CAN making it category 3, as every participant has edited the phrase in a different way. The participants used 3 different approaches for post-editing: word changes, words deleted and a combination of both approaches where only one participant deleted one word. Figure 2 below illustrates the CNA method in a similar way as to how Campbell (2000a) and also Salmi and Koponen (2020) showcased it in their studies.

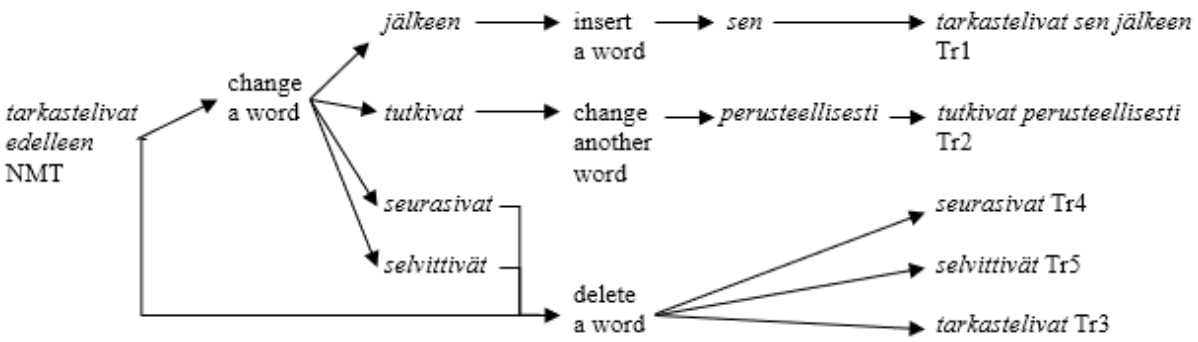

Figure 2. CNA network model for the 4th sentence's phrase *tarkastelivat edelleen.*

The word *tarkastelivat* is of more importance in this phrase than the adverb *edelleen* and this makes it the primary focus for the post-editors. Of the five post-editors two kept the original verb *tarkastelivat* while both of them still made changes to the phrase by inserting a new word to it and changing the other less important adverb (Tr1) and by deleting a word from the phrase (Tr3). This makes it so that four of five post-editors decided to change at least one word from the phrase. The other two post-editors decided to delete the other less important word entirely after changing one word (Tr4, Tr5) and the last person changed both of the words (Tr2.

Following the same sentence from the first example we have the second phrase which is a direct continuation of the first phrase in figure 3. The second phrase is *the activities of 36 people* which was machine translated to *36 ihmisen toimia*. This phrase was post-edited by four participants and one of the post-editors left it unedited (Tr4). This makes the phrase less difficult, category 2, as the previous one as it has still been post-edited by at least four participants. One post-editor only made changes to the form of one word leaving the other words intact (Tr3). Two post-editors inserted words to the original phrase and from these two, Tr1 inserted one word while Tr5 inserted three separate words, of which one was the verb of the phrase, changed a word, and also altered the word order of the whole phrase. Lastly, Tr2 changed the verb from of the phrase into *syytetyn*.

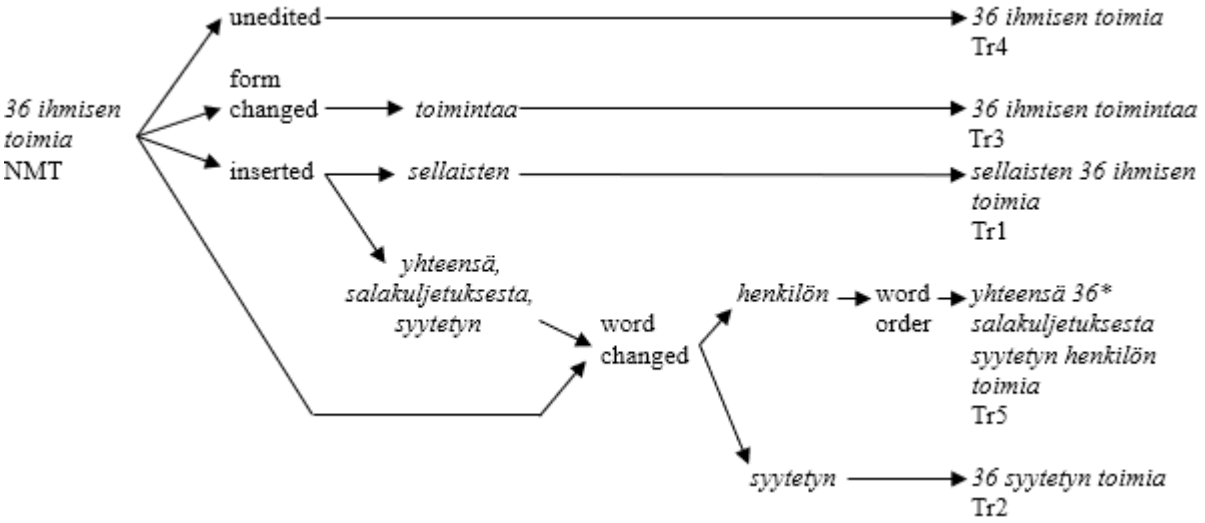

Figure 3. CNA network model for the 4th sentence's phrase *36 ihmisen toimia*.

In the figure 3 above, there is an asterisk (*) marked on the number *36* from Tr5. This is because in the PE text Tr5 actually wrote the number 36 (as in *thirty-six)* in Finnish which,

when written in Finnish (*kolmenkymmenenkuuden*), was too long to fit for the figure 3 CNA network model and as such was shortened to 36.

The next phrase, illustrated in the table 3 below, is again a direct continuation of the previous figure. This phrase is longer than the previous ones and has been cut to subsegments due to its structure, as the NMT translation differed somewhat of the one made by the post-editors and the word order was changed in a way that makes the analysis less optimal. This example is shown as a table instead of the traditional network model. This table model has been used in other studies to showcase CNA model in an alternative way which makes it easier to see the PEs at a glance, and it is also easier and less time consuming for the researcher to make than the traditional network model (O'Brien 2006; Salmi and Koponen 2020). This phrase has also been post-edited differently by every participant (category 3) making it cognitively difficult on CNA standards. Participants Tr3 and Tr5 cut the sentence in the middle by putting a full stop before the phrase *joita syytettiin* and starting with a new sentence. Every participant edited the phrase by changing the word order and changing the form of some words. Tr2 was the only one who kept the word *salakuljetettu* in somewhat same position with the NMT although, the form of the word was changed. The word *syytettiin* was moved to the previous phrase and a new word *oli* was added here. Every other participant kept the original words *salakuljetuksesta* and *siirtolaisen* and Tr1 and Tr5 also changed the word *vähintään.*

Table 16. CNA table model of a longer phrase in the 4th sentence.

ST4	**accused**	**of smuggling**	**at least**	**580 migrants**	
NMT	**joita syytettiin**	**salakuljetuksesta**	**vähintään**	**580 siirtolaista**	
Tr1	joita syytetään		ainakin	580 siirtolaisen	salakuljetuksesta
Tr2		joissa oli salakuljetettu	vähintään	580 siirtolaista	
Tr3	Heitä syytettiin		vähintään	580 siirtolaisen	salakuljetuksesta
Tr4	joita syytettiin		vähintään	580 siirtolaisen	salakuljetuksesta
Tr5	Henkilöitä syytettiin		ainakin	580 siirtolaisen	salakuljetuksesta

The table model, illustrated in table 16 above, makes it easier to quickly glance the differences of various PE made by different people. However, when the word order is changed this makes it somewhat more difficult to manage the table format as it might get stretch wide on paper.

The final difficult phrase (or word in the source language) of the 4th sentence is *earning,* machine translated to *ja jotka ansaitsivat.* This phrase was edited by everyone (category 3),

especially by Tr5 who reworked the whole phrase. Tr5 changed every word, inserted new words, and hence changed the word order. Other participants edited the word forms, primarily the word *ansaitsivat* (except Tr4) was altered, whereas Tr4 changed the word *jotka* to *syytetyt* while also deleting the word *jotka.*

Next, we move on to the 7th sentence, *with financial and logistical support provided by Eurojust* MT to *jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti*. This longer phrase has multiple words in it and was analysed as a single segment, as otherwise it would have been somewhat difficult to examine them due to the source language phrase and the NMT phrase being somewhat different from each other. Participants Tr2 and Tr4 left this segment unedited, while the rest made some changes making it category 1. Tr5 only changed the form of the word *tukevat* from its present tense to the past tense *tukivat*. Tr1 and Tr3 on the other hand changed this quite a bit. They both changed the word order of the whole phrase, deleted words, and changed the form of some of the words. Tr1 changed the word order so that they put a subordinate clause at the end of the sentence with the word *Eurojust* in it, and they removed the word *Europol* which is not seen in the ST and one of the *ja* words. *Tukevat* also had its from changed to *tukee.* Tr3 did similar alterations as they removed *Europol, jota,* and of the two *ja* words*,* changed the word order slightly, and also changed the form of every word in their post-edited phrase except *ja*.

In the 8th sentence there is again only one phrase that has multiple post-edits done to it, *the action day* machine translated as *toimintapäivän*. This has been post-edited by Tr2, Tr3, and Tr5 and left unedited by Tr1 and Tr4 (category 1). The three post-edits were all about changing the actual word where they removed the word *action* altogether. Tr2 changed it to *pidätystoimien* which is more in the realm of arresting action, whereas Tr3 and Tr5 changed the word in somewhat similar ways; Tr3 changed it to *paljastusiskun* and Tr5 shortened the word to *iskun.*

Like in the previous sentence, sentence 9 also only has one phrase that has been post-edited by three participants (category 1). The MT of the phrase is *pieces of identification* machine translation is *tunnistetietoja* which was left unedited by Tr4 and Tr5. The rest changed this one word to various different ones. Tr1 changed it to *henkilötodistuksia,* Tr2 to *henkilöasiakirjoja,* and Tr3 to *henkilöpapereita;* all these words are very similar as they are compound words that start with the word *henkilö-*.

Next, we move on to sentence 14 which has been cut to subsegments to make it easier to analyse as was also done in the 4th sentence. All the participants had changed various different parts of this phrase making this category 3, so it was decided to put these subsegments into the table form as was mentioned above due to the table form being easier for the researcher to make and it being easier to quickly see the differences. These phrases saw various modifications done to them: word changes, form changes, insertions, and word order changes, making this a cognitively difficult phrase.

Table 17. CNA table model of a longer phrase in the 14th sentence.

ST14	**In related actions,**	**four drivers**	**were sentenced**	**in Germany**			
NMT	**Toisiinsa liittyvissä oikeudenkäynneissä**			**Saksassa**	**tuomittiin**	**neljä kuljettajaa,**	
Tr1	Tähän liittyvissä oikeudenkäynneissä			Saksassa	tuomittiin	neljä kuljettajaa,	
Tr2	Asiaan liittyvissä oikeudenkäynneissä			Saksassa		neljälle kuljettajalle	annettiin tuomiot,
Tr3	Asiaa koskevissa oikeudenkäynneissä			Saksassa		neljälle kuljettajalle	langetettiin tuomioita,
Tr4	Toisiinsa liittyvissä oikeudenkäynneissä			Saksassa	tuomittiin	neljä kuljettajaa	rangaistuksiin,
Tr5	Saksassa annettiin iskuun liittyvissä oikeudenkäynneissä		tuomio			neljälle kuljettajalle.	

In the table 17 above, a quick a glance shows that Tr4 has kept the phrase almost unedited and Tr5 has made some significant changes to it. Next, it can be seen that each participant had kept the beginning of the sentence somewhat similar, mainly changing the pronoun, *Toisiinsa* with the exceptions of Tr4 who left the first subsegment unedited and Tr5 inserting new words and moving *Saksassa* at the very beginning. As Tr5 changed the word order of the whole sentence they were the only one to move *tuomio* (*sentence* to approximately the same place it would have been in the ST. Everyone else, kept *Saksassa* in its current place. The word *tuomittiin* was left here by Tr1 and Tr4. However, every participant left the *neljä kuljettajaa* here with some form changes done to it. Finally, Tr2, Tr3, and Tr4 moved various words at the end of the phrase. Tr2 inserted a new word *annettiin* with the word *tuomiot* which had its form changed, Tr3 did the same thing by inserting *langetettiin* and changing the form to *tuomioita,* whereas Tr4 just inserted the word *rangaistuksiin*.

Next is a short sentence 15, which has two phrases with multiple modifications and it's also worth noting that every participant post-edited this differently from the NMT. In fact, Tr2,

Tr3, and Tr4 post-edited the whole sentence identically. This makes the CNA result somewhat interesting in this sentence, as three of the participants came to the exact same conclusion still making this segment somewhat difficult cognitivelly. The first phrase is *Another driver* machine translation *Toinen kuljettaja,* which has been post-edited by every participant (category 3). Tr1 changed a word and form of these two words to *Erästä kuljettajaa,* while Tr5 changed a word, word order, and word form by moving this phrase towards the end of the sentence *yhden kuljettajan*. Tr2, Tr3, and Tr4 changed the word *Toinen* to *Yksi*. The other phrase of this sentence is a direct continuation of the first one *is standing trial* machine translation *on seisova oikeudenkäynti.* Although, in this paper there is no MT error analysis, here a clear error has happened where the NMT has *literally* translated the phrase which has then been post-edited by everyone (category 3). Tr1 has inserted the word *vastaan* at the beginning of the phrase and changed a word *vireillä*. As was said earlier, Tr5 changed the word order, but they also changed a word *meneillään*. Tr2, Tr3, and Tr4 deleted a word, changed a word *odottaa*, and changed word form *oikeudenkäyntiä.*

Finally, the 16th sentence has two phrases modified by multiple participants. The first phrase is *the main organiser* machine translation *pääjärjestäjän* which was left unedited by Tr1 and Tr3 (category 1). Tr2 inserted the attribute *salakuljetusten* before the main word *pääjärjestäjän.* Tr4 changed the word to *pääorganisaattorin* which is rather literal translation of the ST word. Tr5 on the other hand inserted *salakujetusringin* before the main word and then changed the main word to *johtajan.* The next phrase of this sentence was divided into subsegments to ease the analysis, although these phrases and words are directly linked to each other: *on a German European Arrest Warrant* machine translation is *saksalaisesta eurooppalaisesta pidätysmääräyksestä.* Every participant post-edited this phrase differently making this cognitively difficult (category 3). Every participant, except Tr3, inserted a word at the very end while everyone changed word forms of each word in this phrase. Tr1, Tr4, and Tr5 inserted the word *perusteella* at the end of the sentence to explain on behalf of what the arrests were made and Tr2 inserted the word *nojalla* for most likely the same reasons. Tr3 was the only who did not insert words to this phrase rather they just changed word forms to achieve similar results to the other post-editors.

4.3 Choice Network Analysis and Translog-II Analysis Comparison

With results from the both previous sections we can now compare the results of the two methods, did the CNA and Translog-II show the same cognitively difficult phrases/words in

the PE process of the five professional translators. The phrases are compared in the same order as they appeared in the previous sections 4.1 and 4.2. First the CNA result is given and then the Translog-II results are given for comparison. The difficulty scales had three categories: 1 being somewhat difficult, 2 being difficult, and three being very difficult. In CNA the cognitive difficulty was measured by counting how many participants post-edited the phrase with the minimum being three participants and with the process analysis tool Translog-II, the difficulty was measured by the time it took for the participants to post-edit the phrase.

Table 18. Sentence 4 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
4/1	5/5	3 – 11 sec.
4/2	4/5	16 – 28 sec.
4/3	5/5	10 – 63 sec.
4/4	5/5	1 – 8 sec.

From the 4th sentence the first phrase *tarkastelivat edelleen* was post-edited in some way by every participant making it cognitively very difficult by the CNA method which makes it difficulty category 3. The PE times ranged from 3s to 11s which actually makes this one of the shorter PE times of all the phrases analysed and in doing so, contradicts the CNA method being category 1. The second phrase *36 ihmisen toimia,* was post-edited by four participants making it difficult to PE, category 2. The PE times ranged from 16s to 28s which make it cognitively more difficult than the previous phrase, category 2. The third phrase *joita syytettiin salakuljetuksesta vähintään 580 siirtolaista* was post-edited by every participant, so its category 3. PE times give similar results as the time range was much wider from 10s to 1m 3s, although four of the participants edited the phrase in less than 20s, category 3. The final phrase of the 4th sentence was *ja jotka ansaitsivat* was post-edited by everyone making it very difficult in the eyes of CNA. The Translog-II on the other hand shows otherwise with PE times of 1s – 8s making this one of the shortest PE times of every phrase, category 1.

Table 19. Sentence 7 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
7/1	3/5	7 – 21 sec.

Next sentence was sentence number 7, which had only one phrase to analyse. The phrase *jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti* was post-edited by three people which was the minimum required for this paper, making this somewhat cognitively stressful with CAN, category 1. On the other hand, the PE times were 7s – 21s making this difficult, category 2.

Table 20. Sentence 8 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
8/1	3/5	4 – 94 sec.

Moving to sentence 8 which also has only one phrase to analyse *toimintapäivän*. This phrase was post-edited by three people as was the previous phrase making this category 1. This phrase was in fact one of the longest phrases to PE by the two participants who did PE this one, although one of the three post-edited this in 4s. The PE times were 4s – 1m 34s with the middle one taking 46s to PE showing that this was one of the more difficult phrases as the results ranged quite widely, category 3.

Table 21. Sentence 13 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
13/1	3/5	6 – 84 sec.

Next was the 13th sentence with one phrase *tunnistetietoja.* This was post-edited by three participants which would make this phrase also only somewhat cognitively difficult to PE, category 1. Translog-II shows some mixed results for these three as the PE times ranged from

6s to 1m 24s with the middle one taking 17s to PE. As one translator took more than 31s to PE the phrase it is category 3, which is very different from the CNA method.

Table 22. Sentence 14 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
14/1	4/5	3 – 34 sec.
14/2	3/5	3 – 21 sec.

The next comparison is for the 14th sentence which had two phrases to analyse in it. First one being *Toisiinsa liittyvissä oikeudenkäynneissä* which was post-edited by four of the five participants in this way making it cognitively difficult to PE by the CNA method, category 2. This phrase saw PE times ranging from 3s to 34s having two of the shortest PE times, but with the exception rule mentioned in 4.1, this belongs to category 2. This phrase can be said to be difficult with both methods due the PE times and the number of post-edits. The second phrase of this sentence was *tuomittiin.* This phrase was post-edited by three people which again makes it only somewhat difficult to PE in the eyes of CNA, category 1. The PE times were 3s and 21s, category 2.

Table 23. Sentence 15 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
15/1	3/5	1 – 7 sec.
15/2	5/5	4 – 8 sec.

Moving to the very short sentence numbered 15 which had two phrases to analyse as well. The first phrase was *Toinen kuljettaja* and this one was post-edited by every participant making it very difficult with the CNA method which would be category 3, but as three of the editors edited the phrase in the same way meaning it was, in fact only somewhat difficult, which ultimately is category 1. As CNA showed somewhat cognitively difficult results (category 1), Translog-II showed the same with PE times of 1s – 7s making this the shortest PE time in this paper meaning it was not cognitively difficult, category 1. The second phrase of this sentence was *on seisova oikeudenkäynti* which was post-edited again by everyone

making it very difficult. PE times on the other hand show a different story with times ranging from 4s to 8s, also making this phrase have the least amount of PE time difference between the participants, category 1. This short sentence showed vastly different results for these two methods.

Table 24. Sentence 16 number of post-editors and time used to post-edit.

Sentence / phrase	number of editors / CNA	time used / translation process (Translog-II)
16/1	3/5	3 – 25 sec.
16/2	5/5	3 – 28 sec.

Finally, the 16th sentence had two phrases to analyse. The first phrase *pääjärjestäjän* was post-edited by three people which makes this category 1. This phrase saw some variation with the PE times 3s – 25s with the middle one taking 18s making this difficult with the process analysis, category 2. The final phrase of this paper was *saksalaisesta eurooppalaisesta pidätysmääräyksestä*. This last phrase was post-edited by every participant which makes this category 3 with the CNA method. The PE times show different results with 3s – 28s. The longest PE time for this phrase was 28s and it still lies in the middle category of the PE times and in doing so, contradicts the CNA method in this phrase, category 2.

5 Discussion

The results of the present study, which are presented in section 4, will be discussed in better detail in this section. These results will be compared to previous research on cognitive difficulty and PE that were introduced in section 2 with related works and background framework. As the current study did not attempt to make any hypothesis regarding PE or the cognitive difficulty of it, but rather compare these two methods, the results will be compared to previous research on the topic.

In the present study, the CNA method was used to find phrases where more than three post-editors made different changes to the MT and these results were then compared to the editing processes of the participants to ascertain if they show the same cognitively difficult phrases. The results of this study are rather divided. The results showed that both of these methods exhibited the same results of cognitively difficult phrases on four cases and the other nine were somewhat, or very different from each other. For this research it would mean that the CNA method did not give the same cognitively difficult phrases as Translog-II did.

It has been established in previous studies that the CNA method might not yield the best possible results by itself but it can work really well with other methods or tools such as Translog (O'Brien 2005; 2006). And other studies suggest that having multiple different edits by translators does not always imply cognitive difficulty but can, in fact, also mean creative freedom (Ketola 2016). This study shows that the CNA method might not always equal cognitive difficulty, as the PE times shown in this paper imply. In this paper, on numerous instances the translators had multiple different post-edits done to a phrase for which the PE time was very short.

These differences between CNA and the translation process analysis in four of the cases in this study could imply what Ketola said in their research data (2016) that the translators use their creativity here instead of thinking what the "correct" translation could be. The short PE times, especially in tables 5, 8, 12, and 14, suggest that the participants did not have cognitive difficulties while post-editing these phrases. On the other hand, the aforementioned tables were post-edited by 4 to 5 participants which would mean cognitive difficulty with the CNA method.

The CNA method is used to determine if the translator had some or any cognitive difficulty while translating or post-editing text. In this paper, CNA showed variable results along with

the translation process analysis. By itself CNA would have showed perhaps somewhat misleading results on some of the phrases on the various segments examined in this paper. Although, this study did not examine every sentence in the text, as the scope was to examine those phrases which had three or more different post-edits done to them, it could be possible that the phrases not included in the analysis of this paper might have had better results along with the translation process analysis tool Translog-II. Four of the results showed that CNA and the translation process analysis are in an agreement with which post-edits could be cognitively difficult and which not. In these results, some post-edits were done fast and only three did any changes to them, and in some of these cases, every participant made changes and they also took considerably longer to edit these.

While CNA was useful in helping identify phrases which might be cognitively difficult, Translog-II was able to give more precise information e.g., how long it took to post-edit the phrases in question. It can, of course, be debated if PE time is more reliable than counting different post-edits made by various translators via CNA, but that was not the interest of this study. With Translog-II it was possible to see what the participants did during their PE work, even though this was not necessary for this paper. One of the participants wrote down some of their thoughts and questions for themselves on the document for later use, for example. Translog-II can also give insight on what the translator does while writing their post-edition as there are some occasions, even in this study, where it is recorded when a participant makes more than one edit to one word. This way, Translog-II can give more information on what word or phrase might be cognitively difficult as it takes longer to edit these specific words or phrases and it also shows words or phrases that have been edited multiple times, perhaps meaning that those segments were, in fact, cognitively difficult.

Although this study used a triangulation of CNA and the translation process analysis tool Translog-II, there are still some problems that remained for the research, after the analysis was done. This study only used Translog-II to gather data from the participants, there was no eye-tracking used so it is not possible to determine with 100% accuracy if the participants actually started or ended post-editing a certain phrase or segment. In this paper, it was determined that when the cursor moved to the start of the phrase to be edited and did not move immediately back, the line for post-editing that phrase was crossed. There is no way of knowing if the participant actually stopped thinking a previous edited phrase or not and neither was there a think-aloud protocol used although, one participant did write some of their thoughts on the document.

Another problematic factor was PE time to which there was no definite time measurement to decide what is a long PE time and what is not. I was not able to find a study that further expands on what is a suitable time to PE e.g., a phrase or a similar segment, which would serve this study. The PE times ranged from all the way from 1 second to 1 minute and 34 seconds. It was then determined that for this study, rough estimates of 15 second increments were to be used based on every participant's answers on all of the post-edits done and three categories were made according to the PE times.

As was mentioned earlier, it has been argued that post-editing a MT should ease the workload increasing productivity of translating (Plitt and Masselot 2010; O'Brien 2011). Considering this, it would behove us to believe that the less time spent post-editing the better this would be for productivity. As some of the lowest PE times were only a few seconds, it could hypothetically be said that those phrases were post-edited fast since this time is rather close to 0 (zero) seconds, and thus were, in fact cognitively not very difficult.

Although outside the scope of this paper, it could perhaps have been possible to theorise some results with the participants' experience in translations and the experience with PE times. The background information showed how long each participant had been translating professionally and the questionnaire sheet showed that four of the five participants had no experience with PE before this research. If a person has 20 years of experience in translation, can they still post-edit with relative ease and how easily did they understand the concept of light vs. heavy PE? If they did not follow the light PE rules, they might have done some unnecessary work which is not very productive or cognitively easy. On the other hand, if a translator has less than 10 years of experience in translation but has done multiple works by PE are they more, or less productive than someone who has more experience from the translation field in general? This would have needed more data to examine and as such, was not possible with the limitations of this study.

6 Conclusion

In this comprehensive study, we aimed to investigate the cognitive effort involved in post-editing machine translation and compare two methods, Choice Network Analysis and the translation process analysis tool Translog-II, to identify cognitively difficult phrases during the PE process in a somewhat similar fashion to O'Brien (2005) and Koponen and Salmi (2020). The comparison between these methods revealed both areas of agreement and discrepancies, shedding light on the complexities of the post-editing process and the factors influencing cognitive effort.

The study involved a group of five professional translators who participated in the PE process as part of the research investigation. These participants were gathered from the monthly SKTL meeting. Additionally, the participants' contributions help explain the cognitive complexities of translation, emphasizing the need for a nuanced approach to evaluating translation quality and difficulty.

We compared the outcomes of two distinct methods used to understand the cognitive intricacies in the PE process of skilled professional translators. The primary focus lies on the CNA and the translation process analysis as the chosen analytical tools for this investigation. Our aim was to discern any similarities and differences between these two methods in identifying phrases that pose cognitive challenges during the PE task, which, in turn, sheds light on the effectiveness and reliability of each method in comprehending the subtleties of translation cognitive effort.

As the analysis unfolds, we carefully scrutinised the results, presenting the cognitively difficult phrases pinpointed by both the CNA and Translog-II methods. In four of the cases, the two methods reached similar results, agreeing on the identification of genuinely challenging phrases, thereby somewhat validating the reliability of the CNA approach in certain instances. However, in the remaining cases, we encounter differing results that prompt a more in-depth examination of CNA's effectiveness in capturing cognitive difficulty, particularly when compared to the translation process analysis insights. These contrasting outcomes raise pertinent questions about the suitability of the CNA method under specific conditions and contexts. Consequently, we recognise the multifaceted nature of translation cognitive processes, acknowledging that no single method can claim unassailable supremacy

in assessing cognitive difficulty. But using both CNA and Translog-II bolsters the credibility of these two methods used in tandem with each other in assessing cognitive difficulty of PE.

Throughout this investigation, we noticed an interesting observation: certain phrases exhibit multiple post-edits despite short post-editing times. This phenomenon begs a profound question – do these multiple post-edits genuinely indicate cognitive difficulty (Campbell 2000a), or could they be manifestations of translators exercising their creative liberties during the PE task (Ketola 2016)? These insights underscore the need to consider factors that shape translators' decision-making and linguistic choices during PE.

Another intriguing facet of this study is the impact of variation in post-editing times among the skilled participants. This disparity in PE times contributed to diverse interpretations of cognitive difficulty. Some phrases, seemingly brief in post-editing duration, were deemed cognitively challenging by the CNA method, while the translation process analysis results were at odds with such classification. This revelation highlights the subjectivity inherent in defining translation difficulty and the idiosyncratic nature of cognitive engagement among the various translations and post-editions made by professional translators.

For instance, in the 4th sentence, the phrase *tarkastelivat edelleen* was post-edited by all participants, indicating high cognitive difficulty according to CNA. However, the PE times ranged from 3 to 11 seconds, contradicting CNA's assessment. Similarly in the 4th sentence, the phrase *ja jotka ansaitsivat* was post-edited by all participants, indicating high cognitive difficulty according to CNA. However, the translation process analysis revealed short post-editing times of 1 to 8 seconds for this phrase, showing disagreement with CNA's assessment.

Nonetheless, some cases did still comply with both methods such as the 4th sentence, with the phrase *joita syytettiin salakuljteuksesta vähintään 580 siirtolaista* was post-edited by every participant, denoting cognitive difficulty by CNA. The PE times ranged from 10 seconds to 1 minute and 3 seconds, aligning with CNA's evaluation. The 13th phrase *tunnistetietoja* was also post-edited by three participants making it somewhat cognitively difficult to PE. Translog-II showed varied results with PE times ranging from 6s to 1m 24s. Here only one translator took more than 20s to PE the phrase emphasising this being somewhat difficult with the CNA method.

On the subject of PE times, future research could perhaps focus on what is a short or long PE time and how does it factor in with the CNA. During the start of this study, the researcher did not find and relatable studies examining what is a short or long PE time. CNA is unable show PE times but with Translog-II it would be possible to gather this data, and then later compare this with CNA results.

In culmination, our comparative exploration of CNA and Translog-II in assessing cognitive difficulty throughout the PE process showcases profound insights into the intricate realm of translation cognition. While both methods provide valuable contributions to our understanding, their divergences and limitations prompt the realisation that the complexities of cognitive effort in translation can only be fully comprehended through a multi-faceted approach. Such an inclusive perspective entails the amalgamation of various analytical methodologies and a nuanced recognition of individual translator nuances, allowing for a more holistic comprehension of translation cognitive processes.

The present study examined and compared cognitive difficulty of professional translators' who post-edited a machine translated text from English to Finnish. There is very little research done on this topic regarding Finnish language and using professional Finnish translators in research. This study showed that professional translators, even those that have been translating for more than a decade, have either some cognitive difficulty post-editing or are taking their time using their creative freedom while translating or post-editing. It would benefit the translation field to research this topic more extensively in the future, for us to learn the intricacies of PE and the cognitive difficulty associated with it.

References

Almeida, Giselle de. 2013. 'Translating the Post-Editor: An Investigation of Post-Editing Changes and Correlations with Professional Experience across Two Romance Languages'. *De Almeida, Giselle (2013) Translating the Post-Editor: An Investigation of Post-Editing Changes and Correlations with Professional Experience across Two Romance Languages. PhD Thesis, Dublin City University.* Doctoral, Dublin City University. School of Applied Language and Intercultural Studies. https://doras.dcu.ie/17732/.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2016. 'Neural Machine Translation by Jointly Learning to Align and Translate'. https://doi.org/10.48550/arXiv.1409.0473. Accessed 2.2.2023.

Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, et al. 2016. 'Findings of the 2016 Conference on Machine Translation'. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, 131–98. Berlin, Germany: Association for Computational Linguistics. https://doi.org/10.18653/v1/W16-2301. Accessed 2.2.2023.

Campbell, Stuart. 2000a. 'Choice Network Analysis in Translation Research'. In *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*, 29–42. Manchester: St. Jerome.

———. 2000b. 'Critical Structures in the Evaluation of Translations from Arabic into English as a Second Language'. *The Translator* 6 (2): 211–29. https://doi.org/10.1080/13556509.2000.10799066. Accessed 19.9.2022.

Carl, Michael, Srinivas Bangalore, and Moritz Schaeffer. 2015. *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. 1st ed. 2016. New Frontiers in Translation Studies. Cham: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-20358-4. Accessed 27.2.2023.

Castilho, Sheila, Joss Moorkens, Federico Gaspari, Iacer Calixto, John Tinsley, and Andy Way. 2017. 'Is Neural Machine Translation the New State of the Art?' *The Prague Bulletin of Mathematical Linguistics* 108 (May): 109–20. https://doi.org/10.1515/pralin-2017-0013. Accessed 2.2.2023.

European Commission eTranslation. A Connecting Europe Facility website. n.d. Accessed 2 February 2023. https://ec.europa.eu/digital-building-blocks/wikis/digital-building-blocks/wikis/display/DIGITAL/Digital+Homepage. Accessed 2.2.2023.

European Language Industry Survey. n.d. Accessed 2 October 2023. https://elis-survey.org/wp-content/uploads/2022/02/2018-Language-Industry-Survey-Report.pdf

Grönroos, Stig-Arne, Sami Virpioja, and Mikko Kurimo. 2017. 'Extending Hybrid Word-Character Neural Machine Translation with Multi-Task Learning of Morphological Analysis'. In *Proceedings of the Second Conference on Machine Translation*, 296–302. Copenhagen, Denmark: Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-4727. Accessed 2.2.2023.

Hale, Sandra, and Stuart Campbell. 2002. 'The Interaction Between Text Difficulty and Translation Accuracy'. *Babel* 48 (1): 14–33. https://doi.org/10.1075/babel.48.1.02hal. Accessed 31.1.2023.

Hansen, Gyde. 1999. *Probing the Process in Translation: Methods and Results*. Samfundslitteratur. https://research.cbs.dk/en/publications/probing-the-process-in-translation-methods-and-results. Accessed 2.2.2023.

ISO/DIS 18587. 2017. 'Translation Services-Post-Editing of Machine Translation Output-Requirements. International Organization for Standardization'.

Jia, Yanfang, Michael Carl, and Xiangling Wang. 2019. ‘How Does the Post-Editing of Neural Machine Translation Compare with from-Scratch Translation? A Product and Process Study’. *The Journal of Specialised Translation*, January, 60–86. https://www.researchgate.net/publication/330901732_How_does_the_post-editing_of_neural_machine_translation_compare_with_from-scratch_translation_A_product_and_process_study. Accessed 1.2.2023.

Junczys-Dowmunt, Marcin, Tomasz Dwojak, and Hieu Hoang. 2016. ‘Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions’. https://doi.org/10.48550/arXiv.1610.01108. Accessed 2.2.2023.

Ketola, Anne. 2016. ‘An Illustrated Technical Text in Translation: Choice Network Analysis as a Tool for Depicting Word-Image Interaction’. *Trans-Kom* 9 (July): 79–97. Accessed 7.2.2023.

———. 2018. ‘Word-Image Interaction in Technical Translation: Students Translating an Illustrated Text’. https://doi.org/10.13140/RG.2.2.18766.15688. Accessed 6.2.2023.

Koponen, Maarit, Wilker Aziz, Luciana Ramos, and Lucia Specia. 2012. ‘Post-Editing Time as a Measure of Cognitive Effort’. In *Workshop on Post-Editing Technology and Practice*. San Diego, California, USA: Association for Machine Translation in the Americas. https://aclanthology.org/2012.amta-wptp.2. Accessed 20.9.2022.

Koponen, Maarit & Salmi, Leena. 2017. ‘Post-Editing Quality: Analysing the Correctness and Necessity of Post-Editor Corrections’. *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 16 (January). https://doi.org/10.52034/lanstts.v16i0.439. Accessed 19.9.2022.

Koponen, Maarit, Leena Salmi, and Markku Nikulin. 2019. ‘A Product and Process Analysis of Post-Editor Corrections on Neural, Statistical and Rule-Based Machine Translation Output’. *Machine Translation* 33 (1): 61–90. https://doi.org/10.1007/s10590-019-09228-7. Accessed 19.9.2022.

Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Process*. Translation Studies 5. Kent, OH: The Kent State University Press.

Mi, Chenggang, Lei Xie, and Yanning Zhang. 2020. ‘Improving Adversarial Neural Machine Translation for Morphologically Rich Language’. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4 (4): 417–26. https://doi.org/10.1109/TETCI.2019.2960546. Accessed 2.2.2023.

Nunes Vieira, Lucas. 2017. ‘Cognitive Effort and Different Task Foci in Post-Editing of Machine Translation: A Think-Aloud Study’. *Across Languages and Cultures* 18 (June): 79–105. https://doi.org/10.1556/084.2017.18.1.4. Accessed 20.9.2022.

O’Brien, Sharon. 2005. ‘Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability’. *Machine Translation* 19 (1): 37–58. https://doi.org/10.1007/s10590-005-2467-1. Accessed 19.9.2022.

———. 2006. ‘Machine-Translatability and Post-Editing Effort: An Empirical Study Using Translog and Choice Network Analysis’. https://www.semanticscholar.org/paper/Machine-translatability-and-post-editing-effort%3A-an-O%E2%80%99Brien/6ebc94f7da3f43d72d913a4d4085135378de94b7. Accessed 31.1.2023.

O’Brien, Sharon. 2011. ‘Towards Predicting Post-Editing Productivity’. *Machine Translation* 25 (September): 197–215. https://doi.org/10.1007/s10590-011-9096-7. Accessed 31.1.2023.

Östling, Robert, Yves Scherrer, Jörg Tiedemann, Gongbo Tang, and Tommi Nieminen. 2017. ‘The Helsinki Neural Machine Translation System’. https://doi.org/10.48550/arXiv.1708.05942. Accessed 2.2.2023.

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*, 311–18. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135. Accessed 1.2.2023.

Pavlović, N. 2009. 'More Ways to Explore the Translating Mind: Collaborative Translation Protocols.' *Copenhagen Studies in Language*. https://www.semanticscholar.org/paper/More-ways-to-explore-the-translating-mind%3A-Pavlovi%C4%87/b38534cd233494561896569e4cc9801239c9f793. Accessed 31.1.2023.

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 (January): 7–16. https://doi.org/10.2478/v10108-010-0010-x. Accessed 20.9.2022.

Popel, Martin, Marketa Tomkova, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar, and Zdeněk Žabokrtský. 2020. 'Transforming Machine Translation: A Deep Learning System Reaches News Translation Quality Comparable to Human Professionals'. *Nature Communications* 11 (1): 4381. https://doi.org/10.1038/s41467-020-18073-9. Accessed 20.9.2022.

Salmi, Leena, and Maarit Koponen. 2020. ''Valintaverkkoanalyysi Konekäännöksen Jälkieditoinnin Apuna' [Choice Network Analysis as an Aid for Investigating Machine Translation Post-Editing]'. *In MikaEL Kääntämisen Ja Tulkkauksen Tutkimuksen Symposiumin Verkkojulkaisu [Electronic Journal of the KäTu Symposium on Translation and Interpreting Studies]* 13 (May): 58–74. https://www.sktl.fi/liitto/seminaarit/mikael-verkkojulkaisu/vol/mikael-vol14-2021/. Accessed 19.9.2022.

Schou, L., Barbara Dragsted, and M. Carl. 2009. 'Ten Years of Translog'. *Copenhagen Studies in Language*. https://www.semanticscholar.org/paper/Ten-years-of-Translog-Schou-Dragsted/fe5a055cf27684f4a0439e6fecf235d3456e51de. Accessed 1.2.2023.

Shterionov, Dimitar, Riccardo Superbo, Pat Nagle, Laura Casanellas, Tony O'Dowd, and Andy Way. 2018. 'Human versus Automatic Quality Evaluation of NMT and PBSMT'. *Machine Translation* 32 (3): 217–35. https://doi.org/10.1007/s10590-018-9220-z. Accessed 1.2.2023.

TAUS. 2010. 'TAUS Transcreation Best Practices and Guidelines'. 2010. https://info.taus.net/taus-transcreation-best-practices-and-guidelines. Accessed 9.2.2023.

Temnikova, Irina. 2010. 'Cognitive Evaluation Approach for a Controlled Language Post-Editing Experiment'. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf. Accessed 20.9.2022.

Tiedemann, Jörg, Filip Ginter, and Jenna Kanerva. 2015. 'Morphological Segmentation and OPUS for Finnish-English Machine Translation'. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 177–83. Lisbon, Portugal: Association for Computational Linguistics. https://doi.org/10.18653/v1/W15-3021. Accessed 2.2.2023.

Toral, Antonio, and Víctor M. Sánchez-Cartagena. 2017. 'A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions'. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 1063–73. Valencia, Spain:

Association for Computational Linguistics. https://aclanthology.org/E17-1100. Accessed 6.2.2023.

Way, Andy. 2018. 'Quality Expectations of Machine Translation'. In *Translation Quality Assessment: From Principles to Practice*, edited by Joss Moorkens, Sheila Castilho, Federico Gaspari, and Stephen Doherty, 159–78. Machine Translation: Technologies and Applications. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_8. Accessed 20.9.2022.

Yang, Jing, Biao Zhang, Yue Qin, Xiangwen Zhang, Qian Lin, and Jinsong Su. 2018. 'Otem&Utem: Over- and Under-Translation Evaluation Metric for NMT'. Lecture Notes in Computer Science, 291–302. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-99495-6_25. Accessed 1.2.2023.

Appendix 1. Information form

Tunniste:

Tutkimus englanti-suomi-konekäännöksen jälkieditoinnista

marraskuu 2019

Turun yliopisto

Kieli- ja käännöstieteiden laitos

Maarit Koponen
maarit.koponen@helsinki.fi

Leena Salmi
leena.salmi@utu.fi

Editointiprosessitutkimus

Tässä tutkimuksessa tutkitaan englanti-suomi-konekäännöksen virheitä ja niiden korjaamista eli jälkieditointiprosessia. Koe koostuu tehtävästä, jossa koehenkilö editoi konekäännettyä tekstikatkelmaa tietokoneella. Kokeessa kerätään aineistoa seuraavasti:

1) Ennen jälkieditointia koehenkilö täyttää taustatietolomakkeen.
2) Jälkieditoinnin aikana tehtyjä korjauksia seurataan ja tallennetaan näppäilyntallennusohjelmalla.
3) Jälkieditoinnin jälkeen tallennetaan lopullinen käännösversio.
4) Jälkieditoinnin jälkeen koehenkilö saa halutessaan katsoa videon, jossa näkyy hänen editointiprosessinsa.

Aineistoa käytetään ainoastaan tutkimustarkoitukseen. Kaikki henkilötiedot (nimi, mahdollinen opiskelijanumero tai muu henkilöön yhdistettävä tieto) poistetaan tutkimusaineistosta ja säilytetään siitä erillään. Kaikki kerätty aineisto käsitellään ja säilytetään nimettömänä siten, ettei yksittäistä koehenkilöä voida tunnistaa. Koehenkilö voi halutessaan koska tahansa pyytää häneltä kerätyn aineiston poistamista ottamalla yhteyttä yllä nimettyyn tutkijaan.

Suostun että kokeessa syntyvää aineistoa saa käyttää edellä kuvattuun tarkoitukseen.

päiväys: _____

koehenkilön nimi: _____

Appendix 2. Source text

International migrant smuggling network dismantled 14 November 2019
Yesterday, the national authorities in Germany (Federal Police Department) and Romania (Directorate for Investigating Organised Crime and Terrorism and National Police) took part in a simultaneous action against an organised criminal group smuggling migrants from the Middle East. The German Federal Police initially carried out the investigation with the Romanian national authorities and were later joined by UK and Hungarian investigators. National authorities from Germany, Romania, Hungary, the UK and Turkey further looked into the activities of 36 people accused of smuggling at least 580 migrants, earning more than EUR 2 million for their services. The smugglers and their victims travelled from Iran, Iraq and Syria, via Turkey and Greece, to Romania where they were loaded into containers and transported on trucks via Hungary and the Czech Republic into Germany. This form of transport is potentially fatal for the migrants. A joint investigation team (JIT), with financial and logistical support provided by Eurojust, was formed in November 2018 among Germany, Romania, the UK and Hungary, with the participation of Eurojust and Europol. The JIT was instrumental in the success of the investigation and the action day. Two coordination meetings were held at Eurojust to prepare for the action day. Europol supported the actions by analysing data. During the action day, 78 police officers from Germany and 65 police officers from Romania searched 13 houses in Germany and Romania. Four people were detained in Romania and one person was detained in Greece. The police seized 14 cell phones, 2 computers, bank account documentation and various pieces of identification. In related actions, four drivers were sentenced in Germany to terms ranging from one year and eight months to five years and three months. Another driver is standing trial in Hungary. The main organiser was arrested on a German European Arrest Warrant in London earlier this year by German and UK police. His extradition is pending. At the same time, the Turkish Police arrested 14 suspects.

Appendix 3. Machine translation of the source text

Kansainvälinen siirtolaisten salakuljetusverkosto lakkautettu
Eilen Saksan (liittovaltion poliisiosasto) ja Romanian (järjestäytyneen rikollisuuden ja terrorismin tutkintaosasto ja kansallinen poliisi) viranomaiset osallistuivat samanaikaisiin toimiin Lähi-idästä muuttajia salakuljettavaa järjestäytynyttä rikollisryhmää vastaan. Saksan liittovaltion poliisi suoritti ensin tutkinnan Romanian kansallisten viranomaisten kanssa, ja myöhemmin siihen osallistuivat Yhdistyneen kuningaskunnan ja Unkarin tutkijat. Saksan, Romanian, Unkarin, Yhdistyneen kuningaskunnan ja Turkin kansalliset viranomaiset tarkastelivat edelleen 36 ihmisen toimia, joita syytettiin salakuljetuksesta vähintään 580 siirtolaista ja jotka ansaitsivat yli 2 miljoonaa euroa palveluistaan. Salakuljettajat ja heidän uhriensa matkustivat Iranista, Irakista ja Syyriasta Turkin ja Kreikan kautta Romaniaan, jossa heidät lastattiin kontteihin ja kuljetettiin kuorma-autoilla Unkarin ja Tšekin kautta Saksaan. Tämä liikennemuoto saattaa johtaa siirtolaisten kuolemaan. Marraskuussa 2018 Saksaan, Romaniaan, Yhdistyneeseen kuningaskuntaan ja Unkariin perustettiin yhteinen tutkintaryhmä, jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti. Yhteinen tutkintaryhmä oli ratkaisevassa asemassa tutkinnan ja toimintapäivän onnistumisen kannalta. Eurojustissa pidettiin kaksi koordinointikokousta toimintapäivän valmistelemiseksi. Europol tuki toimia analysoimalla tietoja. Toimintapäivän aikana 78 poliisia Saksasta ja 65 poliisia Romaniasta etsivät 13 taloa Saksasta ja Romaniasta. Romaniassa pidätettiin neljä henkilöä ja Kreikassa yksi henkilö. Poliisi takavarikoi 14 matkapuhelinta, 2 tietokonetta, pankkitiliasiakirjoja ja erilaisia tunnistetietoja. Toisiinsa liittyvissä oikeudenkäynneissä Saksassa tuomittiin neljä kuljettajaa, joiden kesto vaihteli vuodesta ja kahdeksasta kuukaudesta viiteen vuoteen ja kolmeen kuukauteen. Toinen kuljettaja on seisova oikeudenkäynti Unkarissa. Saksan ja Yhdistyneen kuningaskunnan poliisi pidätti pääjärjestäjän Lontoossa aiemmin tänä vuonna saksalaisesta eurooppalaisesta pidätysmääräyksestä. Hänen luovuttamisensa on kesken. Samaan aikaan Turkin poliisi pidätti 14 epäiltyä.

Appendix 4. Post-edited sentences of Tr1

Sentence 4

Saksan, Romanian, Unkarin, Yhdistyneen kuningaskunnan ja Turkin kansalliset viranomaiset tarkastelivat sen jälkeen sellaisten 36 ihmisen toimia, joita syytetään ainakin 580 siirtolaisen salakuljetuksesta ja jotka ovat ansainneet yli 2 miljoonaa euroa palveluistaan.

Sentence 7

Marraskuussa 2018 Saksan, Romanian, Yhdistyneen kuningaskunnan ja Unkarin kesken perustettiin yhteinen tutkintaryhmä, johon tulivat mukaan Eurojust, joka tukee taloudellisesti ja logistisesti, ja Europol.

Sentence 8

Yhteinen tutkintaryhmä oli ratkaisevassa asemassa tutkinnan ja toimintapäivän onnistumisen kannalta.

Sentence 13

Poliisi takavarikoi 14 matkapuhelinta, 2 tietokonetta, pankkitiliasiakirjoja ja eri henkilötodistuksia.

Sentence 14

Tähän liittyvissä oikeudenkäynneissä Saksassa tuomittiin neljä kuljettajaa, tuomioiden kesto vaihteli vuodesta ja kahdeksasta kuukaudesta viiteen vuoteen ja kolmeen kuukauteen.

Sentence 15

Erästä kuljettajaa vastaan on vireillä oikeudenkäynti Unkarissa.

Sentence 16

Saksan ja Yhdistyneen kuningaskunnan poliisi pidätti pääjärjestäjän Lontoossa aiemmin tänä vuonna sakasalaisen eurooppalaisen pidätysmääräyksen perusteella.

Appendix 5. Post-edited sentences of Tr2

Sentence 4

Saksan, Romanian, Unkarin, Yhdistyneen kuningaskunnan ja Turkin kansalliset viranomaiset tutkivat perusteellisesti 36 syytetyn toimia, joissa oli salakuljetettu vähintään 580 siirtolaista ja ansaittu yli 2 miljoonaa euroa.

Sentence 7

Marraskuussa 2018 Saksa, Romania, Yhdistynyt kuningaskunta ja Unkari perustivat yhteisen tutkintaryhmän (JIT), jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti.

Sentence 8

Yhteinen tutkintaryhmä oli ratkaisevassa asemassa tutkinnan ja pidätystoimien onnistumisen kannalta.

Sentence 13

Poliisi takavarikoi 14 matkapuhelinta, 2 tietokonetta, pankkitiliasiakirjoja ja erilaisia henkilöasiakirjoja.

Sentence 14

Asiaan liittyvissä oikeudenkäynneissä Saksassa neljälle kuljettajalle annettiin tuomiot, joiden kestot vaihtelivat vuodesta ja kahdeksasta kuukaudesta viiteen vuoteen ja kolmeen kuukauteen.

Sentence 15

Yksi kuljettaja odottaa oikeudenkäyntiä Unkarissa.

Sentence 16

Saksan ja Yhdistyneen kuningaskunnan poliisit pidättivät salakuljetusten pääjärjestäjän Lontoossa aiemmin tänä vuonna Saksassa vahvistetun eurooppalaisen pidätysmääräyksen nojalla.

Appendix 6. Post-edited sentences of Tr3

Sentence 4

Saksan, Romanian, Unkarin, Yhdistyneen kuningaskunnan ja Turkin kansalliset viranomaiset tarkastelivat 36 ihmisen toimintaa. Heitä syytettiin vähintään 580 siirtolaisen salakuljetuksesta, josta he olivat ansainneet yli 2 miljoonaa euroa.

Sentence 7

Marraskuussa 2018 Saksaan, Romaniaan, Yhdistyneeseen kuningaskuntaan ja Unkariin perustettiin Eurojustin ja ja Europolin kanssa yhteinen tutkintaryhmä (JIT) Eurojustin taloudellisella ja logistisella tuella.

Sentence 8

Yhteinen tutkintaryhmä oli ratkaisevassa asemassa tutkinnan ja paljastusiskun onnistumisen kannalta.

Sentence 13

Poliisi takavarikoi 14 matkapuhelinta, kaksi tietokonetta, pankkitilitietoja ja erilaisia henkilöpapereita.

Sentence 14

Asiaa koskevissa oikeudenkäynneissä Saksassa neljälle kuljettajalle langetettiin tuomioita, joiden kesto vaihteli vuodesta ja kahdeksasta kuukaudesta viiteen vuoteen ja kolmeen kuukauteen.

Sentence 15

Yksi kuljettaja odottaa oikeudenkäyntiä Unkarissa.

Sentence 16

Saksan ja Yhdistyneen kuningaskunnan poliisi pidätti pääjärjestäjän Lontoossa aiemmin tänä vuonna saksalaisella eurooppalaisella pidätysmääräyksellä.

Appendix 7. Post-edited sentences of Tr4

Sentence 4

Saksan, Romanian, Unkarin, Iso-Britannian ja Turkin kansalliset viranomaiset seurasivat 36 ihmisen toimia, joita syytettiin vähintään 580 siirtolaisen salakuljetuksesta. Syytetyt ansaitsivat yli 2 miljoonaa euroa palveluistaan.

Sentence 7

Marraskuussa 2018 Saksan, Romanian, Iso-Britannian ja Unkarin kesken perustettiin yhteinen tutkintaryhmä, jota Eurojust ja Europol tukevat taloudellisesti ja logistisesti.

Sentence 8

Yhteinen tutkintaryhmä oli ratkaisevassa asemassa tutkinnan ja toimintapäivän onnistumisen kannalta.

Sentence 13

Poliisi takavarikoi 14 matkapuhelinta, 2 tietokonetta, pankkitiliasiakirjoja ja erilaisia tunnistetietoja.

Sentence 14

Toisiinsa liittyvissä oikeudenkäynneissä Saksassa tuomittiin neljä kuljettajaa rangaistuksiin, joiden kesto vaihteli vuodesta ja kahdeksasta kuukaudesta viiteen vuoteen ja kolmeen kuukauteen.

Sentence 15

Yksi kuljettaja odottaa oikeudenkäyntiä Unkarissa.

Sentence 16

Saksan ja Iso-Britannian poliisi pidätti aiemmin tänä vuonna pääorganisaattorin Lontoossa Saksan eurooppalaisen pidätysmääräyksen perusteella.

Appendix 8. Post-edited sentences of Tr5

Sentence 4

Saksan, Romanian, Unkarin, Yhdistyneen kuningaskunnan ja Turkin kansalliset viranomaiset selvittivät yhteensä kolmenkymmenenkuuden salakuljetuksesta syytetyn henkilön toimia. Henkilöitä syytettiin ainakin 580 siirtolaisen salakuljetuksesta, mistä he olisivat hyötyneet taloudellisesti yli 2 miljoonaa euroa.

Sentence 7

Marraskuussa 2018 Saksan, Romanian, Yhdistyneen kuningaskunnan ja Unkarin virnomaiset perustivat yhteisen tutkintaryhmän, jota Eurojust ja Europol tukivat taloudellisesti ja logistisesti.

Sentence 8

Yhteinen tutkintaryhmä oli ratkaisevassa roolissa tutkinnan ja iskun onnistumisen kannalta.

Sentence 13

Poliisi takavarikoi 14 matkapuhelinta, kaksi tietokonetta, pankkitilitietoja ja erilaisia tunnistetietoja.

Sentence 14

Saksassa annettiin iskuun liittyvissä oikeudenkäynneissä tuomio neljälle kuljettajalle. Lyhin tuomioista oli 1 vuoden ja 8 kuukauden vankeusaika ja pisin 5 vuotta ja 3 kuukautta.

Sentence 15

Myös Unkarissa on meneillään yhden kuljettajan oikeudenkäynti.

Sentence 16

Saksan ja Yhdistyneen kuningaskunnan poliisi pidätti salakuljetusringin johtajan Lontoossa aiemmin tänä vuonna Saksan eurooppalaisen pidätysmääräyksen perusteella.

Appendix 9. Finnish summary

Johdanto

Tässä tutkimuksessa käsitellään jälkieditoinnin roolia kääntämisen alalla ja kääntämisen prosessiin liittyviä kognitiivisia haasteita. Tutkimuksessa käytetään kahta menetelmää, valintaverkkoanalyysiä (Choice Network Analysis) ja prosessianalyysityökalua Translog-II, näiden kognitiivisten haasteiden tutkimiseen. Valintaverkkoanalyysi -menetelmällä tunnistetaan ne kohdat, jotka aiheuttavat kognitiivisia vaikeuksia kääntäjille laskemalla, kuinka monta erilaista käännöstä on tehty, kun taas Translog-II-ohjelmalla kirjataan näppäinpainallukset ja käytetty aika kognitiivisten haasteiden ymmärtämiseksi. Tässä tutkimuksessa sivuutetaan myös jonkin verran jälkieditoinnin haasteellisuuteen vaikuttavia tekijöitä, kuten kääntäjän kokemus ja konekäännösjärjestelmien tuntemus.

Tämän tutkimuksen tavoitteena on selvittää, miten kognitiiviset haasteet ilmenevät kahdella eri menetelmällä: valintaverkkoanalyysillä ja Translog-II:lla. Tutkimuksessa vertaillaan valintaverkkoanalyysin ja Translog-II:n tuloksia päätöksentekomallien ja kognitiivisten prosessien tutkimiseksi jälkieditoinnin aikana ja sen jälkeen. Pidemmät jälkieditointiajat ja useat erilaiset jälkieditoinnit viittaavat korkeampaan haasteellisuuteen, ja lyhyemmät ajat sekä vähemmän erilaisia jälkieditointeja viittaavat suoraviivaisempaan prosessiin. Tutkimuksen tavoitteena on vertailla kahta menetelmää kognitiivisen haasteellisuuden perusteella sekä jälkieditointiin kuluvan ajan mukaan. Lisäksi käymme läpi näiden kahden menetelmän välisiä eroavaisuuksia ja niiden vaikutuksia.

Taustakirjallisuus

Tässä osiossa esitetään yleiskatsaus käännöstutkimuksen alan olennaisiin tutkimusaiheisiin ja tarjotaan perustietoa ja tutkimustaustaa myöhempää kognitiivisten haasteiden ja päätöksenteon tarkastelua varten jälkieditointiprosessissa. Käymme läpi neuroverkkokonekääntimen (Neural machine translation) ja sen vaikutusta käännösten laatuun, valintaverkkoanalyysin menetelmänä, jolla voidaan tutkia kognitiivista haastavuutta käännöstöissä, sekä jälkieditoinnin merkitystä ja sen vaikutuksia käännöstuloksiin ja kognitiiviseen rasitukseen.

Konekääntämisen, erityisesti neuroverkkokonekääntämisen, nopea kasvu on herättänyt paljon huomiota maailmalla. Neuroverkkokonekääntäminen on osoittanut ylivoimaista

suorituskykyä verrattuna muihin konekäännösjärjestelmiin. Keskustelua käydään kuitenkin jatkuvasti konekääntimien käännösten laadusta, ja tietyt kielet, kuten esimerkiksi suomen kieli, ovat edelleen haasteellisia neuroverkkokonekääntimelle. Viimeaikaisissa tutkimuksissa on tutkittu neuroverkkokonekäännösten laatua suomenkielisissä käännöksissä ja pyritty parantamaan käännösten laatua sekä vähentämään niiden virheitä ja näissä onkin jo todettu parannusta.

Käännösprosessin tutkimiseen käytetään Stuart Campbellin ehdottamaa valintaverkkoanalyysi -menetelmää. Valintaverkkoanalyysillä verrataan eri kääntäjien tekemiä saman sanan tai lausekkeiden useita eri käännöksiä kognitiivisen vaikeuden tunnistamiseksi. Valintaverkkoanalyysin tulosten tulkinnasta on kuitenkin käyty keskustelua, ja jotkut tutkijat kyseenalaistavat sen, osoittaako suuri määrä käännöksiä todella kognitiivista haastetta vai heijastaako se pikemminkin kääntäjän yksilöllistä luovuutta.

Jälkieditoinnilla on kaksi pääasiallista tarkoitusta: konekäännösten virheiden korjaaminen ja kääntäjien tuottavuuden lisääminen. Tutkimuksissa on keskitytty jälkieditointivirheisiin, jälkieditoinnin tuottavuuteen ja kognitiiviseen rasitukseen jälkieditointiprosessin aikana. Tutkijat ovat käyttäneet ääneenajattelututkimuksia (Think-aloud-protocols) ja katseenseurantaa (Eye-tracking) tutkiakseen kognitiivisia vaikeuksia jälkieditoinnissa. ”Luovan vapauden" käsite jälkieditoinnin aikana saattaa selittää kognitiivisen rasituksen vaihtelua, silloin kun kääntäjät käyttävät omaa harkintaansa saavuttaakseen haluamansa käännöstulokset.

Tutkimuksen toteutus

Tässä luvussa kuvataan kattavasti tässä tutkimuksessa käytetyt aineistot ja menetelmät. Aluksi kuvataan yksityiskohtaisesti aineistonkeruumenetelmiä, jotka toteutettiin pääasiassa Suomen kääntäjien ja tulkkien liiton (SKTL) jäsenille marraskuussa 2019 järjestetyn konekääntäjäkoulutustilaisuuden aikana. Tämän käytetyn tiedonkeruumenetelmän keskeisenä osana toimi Translog-II.

Translog-II on tietokonepohjainen näppäinpainallusten tallennusohjelma, joka on tässä tutkimuksessa ensisijainen tiedonkeruuväline. Tämä ohjelma tallentaa näppäimistö- ja hiiritoiminnot häiritsemättä käännösprosessia. Se tarjoaa tutkijoille yksityiskohtaisemman näkymän käännösprosessista, kun mitä voitaisiin saada pelkästään lopullisista käännetyistä tekstitiedostoista. Ohjelmassa on kaksi keskeistä sovellusta: Translog-II Supervisor, joka

ohjaa varsinaisia käännösprojekteja, ja Translog-II User, jota kääntäjät käyttävät, tosin tämä ei ollut tämän tutkimuksen kannalta olennaista. Translog-II-dataa tutkittiin siis perusteellisesti Translog-II Supervisor -tilassa.

Translog-II-tietojen lisäksi osallistujille annettiin täytettäväksi taustakyselylomake tietojen keräämiseksi. Lomakkeessa esitettiin kysymyksiä iästä, sukupuolesta, äidinkielestä, englannin ja suomen kielen taidosta (jos ne eivät ole äidinkieliä), koulutustasosta, opituista kielistä, työkokemuksesta sekä kokemuksesta konekäännösten jälkieditoinnissa. Kyselylomakkeet kirjoitettiin käsin suomeksi ja muutettiin myöhemmin PDF-tiedostoiksi sen jälkeen, kun osallistujat olivat täyttäneet ne.

Aineisto kerättiin viiden ammattikääntäjän ryhmältä. Nämä kääntäjät ovat tutkimuksessa nimettömiä ja heistä käytetään nimityksiä Tr1, Tr2, Tr3, Tr4 ja Tr5. Vaikka osallistujissa on eroavaisuutta iän perusteella, heillä on samankaltaisia piirteitä koulutuksen ja kielitaidon suhteen. Kaikki osallistujat ovat naisia ja heillä on maisterin tutkinto kielistä tai kääntämisestä, ja he ovat erikoistuneet pääasiassa kääntämiseen ranskasta suomeen. Heillä on myös kokemusta kääntämisestä suomesta ranskaan ja vähintään yhdestä muusta kielestä. Työkokemus vaihtelee odotetusti iän perusteella, ja suurin osa on työskennellyt kokopäiväisesti kääntäjinä vähintään 20 vuotta.

Tässä osiossa tutustutaan tutkimuksen aineistoon. Aineisto koostuu pääasiassa Translog-II:n avulla saaduista jälkieditointiprosessin tiedoista, kuten ajasta, sekä jälkieditoiduista lauseista. Konekääntämistä harjoiteltiin yhden tapaamisen aikana, jonka jälkeen kääntäjät kokeilivat jälkieditointia lyhyeen englanninkieliseen uutisartikkeliin kansainvälisen salakuljetusverkoston lopettamisesta, joka käännettiin suomeksi Euroopan komission neuroverkkokonekäännösjärjestelmän eTranslationin avulla. Tutkimuksessa korostetaan kansainvälisessä ISO/DIS 18587:2016 standardissa määriteltyä "kevyttä" jälkieditointia, johon kuuluu konekääntämisen maksimaalinen hyödyntäminen samalla kun varmistetaan lopullisen suomenkielisen käännöksen tarkkuus.

Lisäksi luvussa annetaan yksityiskohtainen esimerkki Translog-II-tietojen tallennusprosessista ja selvitetään, miten tämä ohjelma seuraa näppäinten painalluksia, hiiren napsautuksia ja taukoja käännösprosessin aikana. Nämä tallenteet auttavat ymmärtämään kääntäjän toimintaa jälkieditoinnin aikana.

Tämän tutkimuksen erikoinen piirre on laajempien tekstinpätkien analysointi, jossa käsitellään suomen kielen erityistä morfologiaa. Myös erityyppiset muutokset, kuten sanamuodon muutokset, sanojen korvaaminen, lisäykset, poistot ja sanajärjestysmuutokset, on luokiteltu.

Lopuksi luvussa selitetään menetelmä, jolla tauot tunnistetaan jälkieditoinnin aikana. Siinä määritellään tekstin tuottamisen yksiköt jatkuviksi kirjoitustapahtumiksi, joiden välissä on vähintään sekunnin pituisia taukoja. Tätä menetelmää täydennetään sisällyttämällä mukaan kursorin liikkeet, kuten hiiren napsautukset ja toimintonäppäimet. Peräkkäisten muokkausten välisiä aikavälejä tarkastellaan myös kognitiivisen haasteellisuuden mittaamiseksi. Viimeinen osio antaa kuvan siitä, miten tutkimuksessa kerättiin ja analysoitiin aineistoa, ja siten luodaan pohjaa myöhemmille tutkimustuloksille ja keskusteluille.

Tulokset

Tuloksissa vertaillaan kahta käytettyä menetelmää, valintaverkkoanalyysia ja Translog-II-analyysia, jotta voidaan tunnistaa kognitiivisesti haastavia kohtia ammattikääntäjien suorittamassa jälkieditointiprosessissa. Tulokset paljastavat eroavaisuuksia näiden kahden menetelmän välillä eri sanojen ja eri lausekkeiden osalta. Esimerkiksi neljännessä lauseessa valintaverkkoanalyysi luokitteli tietyt lauseet erittäin vaikeiksi, mutta Translog-II-analyysi osoitti näissä olevan lyhyet jälkieditointiajat, mikä viittaa vähäisempiin kognitiivisiin vaikeuksiin. Tutkimuksessa esitetään samanlaisia eroja ja myös samankaltaisuuksia muissa sanoissa ja lausekkeissa, mikä korostaa vaihtelua kognitiivisen haasteen arvioinnissa näiden kahden menetelmän avulla.

Pohdinta

Lopuksi syvennytään yksityiskohtaisempaan keskusteluun tutkimuksen tuloksista, joissa verrataan käytettyjä kahta menetelmää, valintaverkkoanalyysiä ja prosessianalyysi työkalu Translog-II, joita käytettiin kognitiivisesti haasteellisten lausekkeiden ja sanojen tunnistamiseen ammattikääntäjien suorittamissa jälkieditointiprosesseissa. Tutkimuksessa todetaan, että vaikka valintaverkkoanalyysin ja Translog-II:n välillä oli joitakin yhtäläisyyksiä, kognitiivisen haasteellisuuden arvioinnissa havaittiin useimmissa tapauksissa merkittäviä eroja. Tämä viittaa siihen, ettei valintaverkkoanalyysi ehkä anna johdonmukaista kuvaa kognitiivisesta haasteellisuudesta. Tämä pätee erityisesti tapauksissa, joissa lauseeseen tehdään useita jälkieditointeja, mutta jälkieditointiin käytetty aika on erittäin lyhyt. Tämä

saattaa viitata siihen, että kääntäjät käyttävät luovuuttaan pikemminkin kuin kamppailevat kognitiivisten haasteiden kanssa.

Tässä korostuu myös Translog-II:n antamat lisätiedot, erityisesti sen ymmärtämiseksi, kuinka kauan tiettyjen lausekkeiden jälkieditointi kestää ja mitä toimia kääntäjät käyttivät jälkieditoinnin aikana. Tutkimuksessa esitetään myös tiettyjä rajoituksia kuten esimerkiksi se, ettei käytetty katseenseurantaa, jonka avulla olisi voitu tarkemmin määrittää, milloin osallistujat aloittivat tai lopettivat jälkieditoinnin, eikä myöskään käytetty ääneenajattelututkimuksia. Tarkasteluna siis on, mikä on pitkä tai lyhyt jälkieditointiaika ja kuinka moni kääntäjä jälkieditoi saman kohdan eri tavalla, näin ollen tehden lausekkeesta tai sanasta kognitiivisesti haastavan. Käytettävä arviointi Translog-II:lla, käyttää 15 sekunnin aikavälejä aina puoleen minuuttiin asti ja siitä yli. Nämä jälkieditointiajat luokitellaan kolmeen ryhmään kuluneen ajan perusteella. Valintaverkkoanalyysillä kolme kategoriaa perustuu siihen, oliko jälkieditoijia kolme, neljä vai viisi. Tutkimuksessa pohditaan myös löyhästi kysymyksiä siitä, miten kääntäjien käännöskokemus ja perehtyneisyys jälkieditointiin mahdollisesti vaikuttavat jälkieditoinnin hyödyllisyyteen.

Kaiken kaikkiaan tässä osiossa korostetaan kognitiivisten haasteiden arvioinnin monimutkaisuutta ja vaihtelua jälkieditoinnin aikana ja lopuksi todetaan, että tarvitaan kattavampia tietoja jälkieditoinnin kognitiiviseen haasteellisuuden ja muiden tekijöiden vaikutuksen tutkimiseksi.