University of Stavanger

FACULTY OF ARTS AND EDUCATION

MASTER'S THESIS

Programme of study: MA in English and Literacy Studies	Spring semester, 2021 Open
Author: Kristian Fellowes Haukelid	…………………………………………… (Author's signature)
Supervisor: Rebecca Anne Charboneau Stuvland	
Thesis title: Google Translate: A helpful aide? A mixed method sequential explanatory study on the usage and effects of Google Translate in three Norwegian EFL classes	
Keywords: EFL, Machine Translation, Google Translate, ICT, CALL	No. of pages: 115 + attachment/other: 39 Stavanger, 11th May 2021

ABSTRACT

This thesis investigates how learners in Norway use Google Translate to aid them in EFL writing as well as how the usage of Google Translate affects the quality of the texts they write. A mixed method study was used in three Norwegian EFL classes situated in the same school along with learner stimulated recall interviews and teacher interviews. The aim of the study was to determine how effective Google Translate is at helping Norwegian learners at different performance levels and to see how dependant these learners are on Google Translate to help them in EFL writing.

Two writing sessions were conducted to gather data on learners' usage of Google Translate and determine whether the translation tool had any effect on the quality of their written products. The first writing session consisted of learners using dictionaries to help them translate from Norwegian to English, and in total, 33 learners participated in this writing session. In the second writing session, learners were encouraged to use Google Translate as they would normally to help in EFL writing, and 33 learners participated in this writing session, 22 of whom used Google Translate. By comparing vocabulary, syntax errors, subject-verbal concord errors, essay length, and spelling mistakes between the texts from the two writing sessions of learners who used Google Translate in the second writing session, comments could be made on the effectiveness Google Translate has on the quality of learners' texts. A separate analysis was conducted of how Google Translate was being used by learners in the second writing session, i.e. how many times Google Translate was used to translate words, phrases, sentences, entire texts or for other purposes.

Based on the analysis comparing two sets of 22 learners' screen recorded videos and the learners' writing along with information from stimulated recall interviews and teacher interviews, it seems that learners' current usage of Google Translate does not affect the quality of their written product in a positive nor negative way, except for syntax related errors which lowered for all learner groups from the first to the second writing session. Google Translate mostly serves as a quicker alternative to dictionaries as it provides translations at a much fast speed compared to dictionaries. Further, learners mostly use Google Translate to aid in the translation of words and phrases, rarely using it to translate whole sentences or longer texts. However, there is a disparity between lower performing learners, average performing learners, and higher performing learners in how much they use the tool.

Furthermore, there is a lack of training given to learners and teachers on how to use Google Translate as an efficient translation tool. Both learners and teachers that participated in the study reported a lack of training received on proper usage of the tool and reviewing 31 screen recorded videos from the second writing session, it was clear that learners lacked knowledge of the many capabilities the tool has to offer.

There has previously been concern amongst teachers that Google Translate hinders learners in learning English as the translation tool produces incorrect output or that learners use the tool to translate large amounts of text. However, statements from teacher interviews reveal that these teachers seem to have become more acceptant of the translator being used in their class. Previous studies also reveal that Google Translate has reached the point where it has the capability of providing output equivalent to the minimum level of accuracy required for university entrance, thereby providing output better than what most learners in primary school could produce themselves (Mundt & Groves, 2015; Stapleton & Leung, 2019).

Finally, based on the analysis of 64 screen recordings, the author of this thesis argues that Google Docs and other word processors should be a greater concern for teachers than Google Translate and other tools that aid in translation. This is due to a substantial number of learners who participated in the study being heavily dependent on the grammatical and spelling correction tools that the software provides, making it difficult for teachers to assess whether learners know various grammatical rules and possess the ability to apply these rules in written texts.

ACKNOWLEDGEMENTS

I would first like to extend my most sincere gratitude to my supervisor Rebecca for all the help that she has given me throughout the writing of this thesis. From answering countless questions that I have had throughout the year to going above and beyond in providing detailed feedback on my thesis drafts, your feedback and guidance have been invaluable. It has been an utmost pleasure having you as both my lecturer and supervisor, and I appreciate immensely everything that you have done for me. From the bottom of my heart, thank you.

I would also like to express my appreciation and gratitude to my mother and father who have been incredibly supportive of me throughout my life, always encouraging me to do the best that I can. The love and support you show every day mean the world to me, and I could not have asked for better parents.

A special and heartfelt thank you goes to my brother Espen whose help has been invaluable during my time in school and university. Taking time out of your busy schedule to help me with proofreading exams, giving feedback, supplying me with free books, and always being so supportive of me has not gone unnoticed, and I thank you from the bottom of my heart. I truly could not have asked for a better brother.

Lastly, I would like to thank all my friends (extra gratitude to maths wizard Ben) that I have made during the past five years at university. There are too many of you to name, but I have enjoyed spending time with each and every one of you, both inside and outside of lectures. Sharing struggles, great memories, and everything in between, thank you all for making my life so much more wonderful, and I hope to share many more great memories with you all in the future!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AP	Average performing
CEFR	Common European Framework of Reference
CALL	Computer-Assisted Language Learning
EFL	English as a Foreign Language
FL	Foreign Language
GT	Google Translate
GNMT	Google Neural Machine Translation
HP	Higher performing
ICT	Information and Communications Technology
L1	First Language
L2	Second Language/Target Language
LP	Lower performing
MT	Machine Translation
NMT	Neural Machine Translation
NSD	Norwegian Centre for Research Data
OT	Online Translation/Online Translator
SMT	Statistical Machine Translation
UI	User Interface

1. INTRODUCTION

This thesis presents a sequential explanatory mixed method study that explores the questions of how Google Translate (GT) is used by Norwegian primary school learners in the English as a foreign language (EFL) classroom and how the use of GT affects the quality of learners' texts. The thesis focused on three year 7 classes situated in the same school with learners aged between 12 and 13. The introduction chapter covers the motivation for writing the current thesis, then it covers the aims and scope of the thesis, after which a brief section on the relevance of the thesis is covered, and lastly, a more detailed outline of the chapters of this thesis is given.

In this day and age, the importance of being technologically competent cannot be overstated. Technology is used and found in every career field, every city, and every corner of the world. To perform well in today's society, individuals are required to possess some degree of digital competency. Technological literacy is so vital that the Norwegian Directorate for Education and Training places digital skills at the same level of importance as reading, writing, numeracy, and oral skills; the five of which combined called the five basic skills (Kunnskapsdepartementet, 2017b, p. 12). Furthermore, in 2017 the Ministry of Education and Training published a strategy for digitalisation for the school years of 2017-2021 that, amongst other things, sets out to ensure that learners develop the digital skills required to participate in society and to succeed in private life, education, and work. In order to succeed at these goals, schools need to effectively use the possibilities that digital technologies bring with them to have a meaningful impact on learners' learning outcomes (Kunnskapsdepartementet, 2017a; Søby, 2018, pp. 4-5).

With the widespread availability of technological tools on the Internet that learners have at their disposal, it becomes only natural to take a more extensive look at how these tools are being used in school settings and how they might promote or obstruct learning. Google Translate (GT), an online translation (OT) tool used by millions of people every day to translate documents, web pages, conversations, and much more has also had the interest of language learners ever since its release in 2006 (Turovsky, 2016b). Used to save time, as a dictionary or out of laziness, GT has many applications to speed up the process of translating words, phrases, or texts from one language to another. This study looks further into GT's limitations and strengths in order to make it more accessible for teachers to aid their learners

in using GT as a supplementary tool for their language learning rather than a tool that hinders their language learning processes.

1.1. BACKGROUND

Having such an appreciation for technology, particularly information and communications technology (ICT) tools and their widespread applications, I found it was only natural to incorporate my interest into this thesis. As I also just completed my teacher education programme last year, looking at learners' usage of ICT tools became an ideal topic of choice for a thesis.

The discussion around using GT to aid in foreign language (FL) learning was constant during my own years as a learner in primary and secondary school. Some teachers found any use of GT to be considered cheating, whilst others accepted its use as a bilingual dictionary. The one point they all seemed to agree on was the fact that GT should not be used as a translator for sentences or whole texts for that matter. Their perceived views of GT were that its outputs would be severely lacking and contain a significant number of errors. These views might have been well grounded when GT was released in 2006, but significant technological advancements have been made since then. Conversing with some of the teachers that I have encountered during school practice these last years, many of them still hold a sceptical and negative view surrounding learners' usage of GT to aid in written English.

Especially in lower secondary and upper secondary school, the dominant focus of learners in FL learning is achieving good grades, or in other words on the end results rather than the process itself of learning an FL. Support of this statement is given by Imsen (2010, pp. 135-136) who discusses that learners only think of achieving good grades rather than anything else after they start lower secondary school. A survey from Ekholm, Lander, and Wernersson (1977, as cited in Imsen, 2010, p. 135) also shows support for this statement. In a survey conducted with learners in years 4 through 9 in Swedish schools, which closely resemble Norwegian schools, regarding what they thought was the most important thing for them in school, the number one answer was achieving good grades, whereas learning for the sake of learning was number four. I am guilty of this in both lower and upper secondary school. Learning German and Nynorsk was difficult, and although I attempted to achieve good grades by studying properly, the ease of use and availability of Google Translate and other translation programs meant that whenever an opportunity presented itself, translation

programs were used instead of writing sentences myself. Although Google Translate had a lot of drawbacks back then, I knew that the output that it gave would be better than what I could produce myself. Teachers therefore need to find ways of incorporating such tools into the FL classroom. Disregarding and banning the use of these tools instead of teaching learners proper use of them only incites negative consequences which is where a lot of the inspiration for this thesis originates from. Having hampered my own language learning experience by using translation tools to translate whole texts, I wish to investigate how learners today make use of GT in English and how it affects the quality of their texts. Investigating the usage of GT may have implications for my future teaching career, where I can guide learners to make appropriate use of said tools.

1.2. AIMS

The research aims for this thesis are twofold. Firstly, the author of this thesis aims to determine how EFL learners use GT in a regular school setting so that their current behaviours and attitudes towards the translation tool can be laid out and provide context for how to change these behaviours and attitudes for the better. Secondly, the thesis aims to address how learners' current use of translation tools, GT specifically, affects the quality of the texts they write in English. Changes in vocabulary, syntax errors, subject-verbal concord, essay length, and spelling mistakes were compared between two writing sessions to determine the effects of GT. Subject-verbal concord in particular is a type of error that Norwegian learners struggle with which is why it was included in this thesis (Nygaard, 2019, pp. 8-9).

The following two research questions were therefore formulated for this thesis:

- How is Google Translate used by Norwegian primary school learners when they write in English?
- How does the use of Google Translate affect the quality of Norwegian learners' texts when they are writing in English?

A combination of qualitative and quantitative approaches was used to gather the required information needed to answer the two research questions. By combining observations through screen recording software for both a quantitative and qualitative approach, and stimulated recall interviews from three learners in each class for a qualitative approach, a mixed method approach of collecting data was used.

1.3. RELEVANCE

This research is relevant and important because similar research is severely lacking in a Norwegian context seeing as the existence of research on the effects GT has on EFL learners in Norwegian classrooms is absent all together. Only one thesis that delved into the usage and perceptions of GT in Norwegian classrooms could be found (Aksnes, 2018, see section 2.6.).

There were three main rationales behind why Google Translate was chosen as the machine translation (MT) tool that learners would use in this study. The first reason was that most similar research conducted earlier employed the use of GT, meaning that there already exists some empirical data on how GT affects the writing of language learners unlike other MT tools where the data is even more scarce. The second reason was that it seems that language learners in Norway use GT as their MT tool of choice when translating between languages, evidenced by personal observations in teaching practice and personal experiences as a learner and student as well as GT being one of the most prominent MT tools in the world. This means that most Norwegian language learners are already familiar with GT to some extent and know the basics of GT's user interface (UI) in order to translate between two languages. Adding on to this point, the author of this thesis is also most familiar with this MT tool compared to others, making it easier with which to conduct research. The third reason as to why GT was used for this study was that it is completely free with no account creation required to use the tool. This made it easier for the author of this thesis as no time had to be spent on guiding learners through creating accounts, which in turn reduced the probability that technical issues would arise when they were writing their essays.

Some comparative studies have been conducted internationally on how Google Translate affects student writing. However, there are two aspects that separate most of them from the current study. Firstly, most studies conducted have been on secondary school learners and university students (Giannetti, 2016; Mundt & Groves, 2016; Bahri & Mahadi, 2016; Lee, 2020). Little attention on the use of GT has been given to primary school learners. Secondly, a lot of these earlier studies and articles have become less credible as GT's method of translating from one language to another was greatly altered in November 2016, years or months after several of these studies were published (Niño, 2009; O'Neill, 2012; Mundt & Groves, 2015; Turovsky, 2016a). Furthermore, most of these studies and other studies looking at GT in school settings employed different methods of collecting data compared to the current study.

This thesis aims to further the knowledge around the use of machine translation in foreign language learning, especially in Norway where research on this topic is close to non-existent. It might also alter the perceived views that a lot of teachers around the world have regarding GT and its usefulness. After Google's overhaul of how Google Translate operates, studies and research conducted that are based on these new algorithms are lacking, making this thesis of interest to anyone who teaches foreign languages. The author of this thesis believes that the results of this study can aid in filling gaps of knowledge within research that were created after GT changed from a statistical machine translation system to a neural machine translation system (see subsection 2.5.3.). Taking into consideration the extent that GT is being used by learners every day, it is important to study this machine translation tool in great detail to ensure that learners gain better awareness of its strengths and weaknesses.

1.4. THESIS OUTLINE

Chapter 2 details the theoretical background and relevant background information that are covered in this thesis. Furthermore, five previous studies conducted related to MT and FL learning are discussed to shed light on where current research stands in terms of using MT for educational purposes.

Chapter 3, methodology, provides insight into the various methods of how data was collected. Chapter 3 further covers a pilot study that was done prior to the real study, the reliability and validity of the current study, and lastly, ethical considerations that had to be in order prior to the study taking place.

Chapter 4, results, presents the data from the first writing session, the second writing session, the comparison of learners' two written texts between the two writing sessions, and data related to learners' dependency on Google Docs' correction features. This is followed by learner stimulated recall interviews and teacher interviews.

In chapter 5, discussion, the results focused on learners' usage of Google is discussed, followed by how Google Translate affects the quality of Norwegian learners' English texts, and how learners seem to be heavily dependent on the correction tools provided by Google Docs.

Finally, chapter 6 concludes the thesis by summarising major findings, details the limitations and delimitations of the study, discusses implications of future teaching as well as suggestions for future research that should be examined, and concludes with a final statement.

2. THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1. INTRODUCTION

This chapter outlines the various theoretical aspects and important background knowledge that are considered relevant for the thesis. ICT and its usage in Norway, the process of learning L2 knowledge through writing, CALL, machine translation, Google Translate, and pedagogical uses of machine translation for FL learning are all addressed in this chapter. The end of this chapter presents earlier studies that have been conducted internationally on Google Translate and its impact in the language classroom, as similar studies conducted in Norway are non-existent.

2.2. ICT AND ITS USE FOR EDUCATIONAL PURPOSES

2.2.1. DIGITAL SKILLS IN THE *LK20* CURRICULUM

The *LK20* curriculum is the newest school reform in Norway that started its rollout in 2020, with the aim of being completely rolled out in all school years by the 2022/2023 school year (Kunnskapsdepartementet, 2017b). In primary school as well as in years 8 and 9 in lower secondary school, the transition to the new curriculum has been completed as of writing this thesis, entirely replacing the old *LK06* curriculum.

The framework for basic skills was also revised in 2017 to reflect the transition over to the *LK20* reform which defines digital skills as:

> Digital skills include being able to obtain and process information, be creative and inventive with digital resources, and being able to communicate and cooperate with others in digital environments. That includes being able to use digital resources appropriately and justifiably to solve practical tasks. Digital skills also involve being able to develop digital discernment by acquiring knowledge and good strategies of internet usage[1] (Utdanningsdirektoratet, 2017).

Digitals skills are divided into five sub-skills, being the ability to use and understand, the ability to find and process, the ability to produce and edit, the ability to communicate and cooperate, and lastly, the ability to apply digital discernment. Within each skill set are five levels that learners are placed in determined by their current progress in each category. Using Google Translate efficiently and well requires learners to master these digital skills,

---

[1] Translated from Norwegian

particularly the ability to find and process as well as the ability to apply digital discernment as learners need to be able to consider whether the output received by GT is legitimate. Learners must also be taught to be aware of copyright related to translating whole texts from one language to another using translation tools. For this thesis, the framework is used for reference purposes to gauge learners' mastery of the sub-skills that are applicable to the study.

2.2.2. ICT USAGE AND EFFECTS IN NORWEGIAN EDUCATION

The terms ICT and ICT tools in this thesis primarily refer to computers (phones included), both hardware and software as well as the Internet. Kalnina and Kangro (as cited in Isisag, 2012) talk highly of ICT and its development over the years, stating that ICT no longer serves to be a simple additional means in teaching and learning, rather it has become an indispensable part of the modern and contemporary learning environment. ICT tools such as the Internet, laptops, smartphones, e-mail, and social media enable people to communicate, share ideas, and collaborate with one another, increasing the necessity of learning other languages. As they have become a vital part of everyday life, the use of these tools should also be incorporated into the classroom, especially in the FL classroom, as learners will encounter and utilise these tools on a daily basis later in life.

Houcine (2011) mentions more positive impacts that ICT can bring with it in FL learning. Among these impacts, she comments that teaching material can be more easily adapted to learners' needs, feedback can be delivered more quickly to learners, the use of multimedia can make use of all the basic skills at once and lectures have the possibility to become more interesting and engaging. Effective use of ICT can also promote independent learning, more effective collaborations, and stimulate learners' motivation (Houcine, 2011). Padurean and Margan (2009, p. 100) group the advantages of ICT usage into four categories: Capacity to control presentation, novelty and creativity, feedback, and adaptability. Capacity to control presentation entails the difference between books and computers, whereas books have a fixed presentation, computers can combine multiple elements such as text, video, and sound. Novelty and creativity concern how teachers can use different teaching materials every lesson, unlike teaching with textbooks where much of the material is the same. Feedback entails that learners receive feedback from the ICT tools through error correction. Some programs also provide appropriate advice on how to fix errors. Lastly, computer programs and tools are easier to adapt to learners at all levels compared to a textbook. Teachers can

adapt their learning materials a lot smoother and easier to encompass all learners which is also required by schools in Norway through section 1-3 of the Education Act (1998).

Learners typically experience ICT tools and devices as platforms for communicating, sharing hobbies, and playing games rather than as tools for language learning or other types of learning (OECD, 2015, p. 32). Finding ways of incorporating these tools and devices in the classroom might therefore improve learners' learning outcomes quicker due to them already being familiar with the tools, consequently reducing the time that teachers must spend on teaching learners the basics of how said tools and devices work. However, a lot of ICT tools bring with them the possibility of becoming distractions, with many learners surveying that they at times become distracted by ICTs and perform non-academic uses of the tools that they have at their disposal (Kunnskapsdepartementet, 2017a, p. 7). Teachers and schools therefore need to ensure that learners are trained and disciplined in using hardware and software for academic purposes that contribute to their learning outcomes, instead of hampering their development. Schools in Norway have the ability to block access to certain webpages for example in order to prevent distractions from happening. However, not only would it require a lot of effort and time to block these sites, but learners also usually find ways of circumventing these blocks regardless.

Within some areas, studies have been conducted that suggest that the use of ICT can have negative effects on learners' learning outcomes (Kunnskapsdepartementet, 2017a, p. 8). One should therefore not assume that the use of ICT automatically improves learning. Good implementation and usage are prerequisites for it to be successful and provide learners with superior learning outcomes. The impact that ICT can provide is highly dependent on the teachers' engagement of the various tools that they have seeing as technology can amplify great teaching, but great technology cannot compensate for poor teaching (Houcine, 2011; OECD, 2015).

In 2012, PISA results showed that more than 99% of Norwegian learners had access to at least one computer at home and that over 99% had access to home internet (OECD, 2015). As these numbers were from eight years ago, one can reasonably presume that these numbers are now nearing 100%. Additionally, in a lot of municipalities almost every learner in primary and secondary school has access to his or her own ICT tool. This usually comes in the form of a Chromebook or an iPad depending on the what the municipality has invested in. In other municipalities though, around 40% of secondary schools have one ICT device per learner (Kunnskapsdepartementet, 2017a, p. 8). In a newer report from 2019, 155 schools and

their learners from various municipalities in Norway participated in a survey answering to what extent they had access to a computer at school (Fjørtoft, Thun & Buvik, 2019, p. 24). Based on the results, in over half the country's municipalities all learners had access to their own ICT tool and based on current trends this number will only further increase. The percentage of 1:1 learner-device ratio also increases in the later school years. The report by OECD (2015) also revealed that Norway has one of the greatest integrations of ICT in schools out of any country that participated in the survey.

Yet from a survey conducted in 2013 by the European Commission (cited in Gudmundsdottir et al., 2014, p. 4), it was concluded that the pedagogical use of ICT in Norwegian schools was mediocre. An explanation as to why this is the case could be due to teachers not being trained in using ICT effectively. In a survey done of teacher graduates in the period of 2011-2012, 76% of respondents answered that there was little correspondence between the ICT given in their teacher education programmes and the expectations that were imposed on them regarding the use of ICT in the teaching profession (Gudmundsdottir et al., 2014, p. 20). A report from NIFU in 2013 concluded much the same that training in ICT in teacher education is fragmented and poorly anchored (Gudmundsdottir et al., 2014, p. 9). Kunnskapsdepartementet (2017a, p. 9) also came to that conclusion, stating that the competence level of teachers in ICT is varied and that newly educated teachers believe that their ICT training during university was not sufficient.

To combat this, the government committed 90 million NOK in 2017 to strengthen the professional digital competence in teacher education programs as well as offering voluntary online post-graduate courses that aim to improve teachers' pedagogical use of ICT and bolster their professional digital competence (Kunnskapsdepartementet, 2017a, p. 23; Søby, 2018, p. 18). By providing ICT related courses in the teacher education programme that are worth 60 credits, the government also aims to have these qualified ICT teachers spread their competence regarding good pedagogical teaching of ICT to other schools and to offer support and guidance to those who need it.

In 2020, all subjects in years 1 through 9 were given new curricula as part of a major restructuring process of the education reform in order to provide learners with a better quality of learning throughout their educational courses, starting from nursery school all the way through higher education (Meld. St. 28 (2015-2016), p. 5). In the English subject, this new curriculum provided teachers with more flexibility in how they plan their lessons, being able

to influence as well as having more authority on how learners are supposed to achieve the curriculum's stated competency aims.

One of the competence aims that the new English curriculum brought with it is quite relevant for the thesis. It is one that learners are expected to master after finishing year 7 in primary school: "use digital resources and different dictionaries in language learning, text creation and interaction" (Utdanningsdirektoratet, 2020). Arguments can be made that Google Translate fits into this competence aim considering that the aim does not specify if the various dictionaries are supposed to be electronic or physical. The fact that it also groups digital resources together with these various dictionaries, GT fits nicely in between both categories. Therefore, strictly speaking, GT and other MT tools are considered a vital part of the English curriculum in Norwegian primary schools, providing further support that teachers should incorporate these ICT tools in their teaching.

2.3. WRITING IN L2

2.3.1. WRITER-BASED PROSE AND READER-BASED PROSE

Writing is essentially a method of communication between a writer and a reader, with the writer and reader sometimes being the same person (Drew & Sørheim, 2016, pp. 89-90). In order for writing to be understandable to both the writer and reader, the text has to be clear and concise. The function of writing varies greatly, with context and target audience being some of the factors behind why writing takes place. Flowers (1979) distinguishes between two types of writing, writer-based and reader-based. Writer-based writing's target audience is the writer themself, meaning that the texts produced might not make sense to an outside audience as the writer shows little mindfulness to them. On the other hand, reader-based writing shows a deliberate attempt to communicate one's thoughts to an audience. Although writers should always write in a reader-based prose when writing for someone other than themselves, it is especially important that the subjects in this study keep a reader-based prose in mind whilst writing as Google Translate might struggle with comprehending writing that is not clear and concise, thereby creating a less than ideal output. The importance of writing in a reader-based prose was not explicitly addressed to the learners in the study, but the author of this thesis was told by the teachers of the respective English classes that learners had been taught to write clear and concise when translating in GT.

Writing as a skill is a major component of this thesis as the output that machine translation tools provide is heavily contingent on the input that they receive. Although Google Translate, for instance, can compensate for a misspelt word or incorrect sentence structures at times, for these tools to provide the best output possible, proper input must be given. Writing as a basic skill in the *LK20* curriculum is not language specific, rather it applies to writing in any language, but for the purpose of this thesis, English and Norwegian are the main focus.

Writing as a skill is defined by the Framework for basic skills as:

> To write entails being able to express oneself in an understandable and appropriate way regarding various topics and the ability to communicate with others. Writing is also a tool to develop one's thoughts and learning. To be able to write in an understandable and appropriate way, different sub-skills must be developed and interrelated. This involves being able to plan, design and edit texts that are adapted to the contents and the purposes of the writing taking place[2]

(Utdanningsdirektoratet, 2017).

Similar to digital skills, writing as a skill is divided into sub-skills. In this case, there are four instead of five, being the ability to plan and edit, the ability to design, the ability to communicate, and the ability to reflect and assess. Each sub-skill has five levels where learners are placed according to their current capabilities in the various sub-skills. As learners improve and show more proficiency within the sub-skills, their levels increase with the aim of achieving level 5 in all sub-skills.

All subjects taught in Norwegian schools incorporate the five basic skills. However, teachers do have the ability to place a larger emphasis on some skills if they are more prevalent in the subject being taught. Having the five basic skills incorporated into all subjects means that learners are constantly learning how to use these skills, digital and writing skills especially. This can be beneficial in that teachers can, to some extent, expect that learners will retain knowledge of how to write in different genres or how to use digital resources appropriately as some of it is repeated to them in all subjects. EFL teachers can therefore focus more of their attention on the language teaching aspect of the subject instead of spending a lot of time on areas that shift the focus away from language learning.

---

[2] Translated from Norwegian

2.3.3. LANGUAGE TRANSFER FROM WRITING IN NORWEGIAN TO ENGLISH

When learning a second language, unless the learner is a simultaneous bilingual, the use of the first language will be either subconsciously or consciously applied in the learning process (Drew & Sørheim, 2016, p. 17). This use of language transfer from their first language (L1) to L2 may both be of help or a hinderance, dependent on the similarity between the two languages (Koda, 2007). Between the Norwegian and English languages, some areas will benefit from the use of language transfer, such as vocabulary and syntax due to the two languages both being Germanic. It is vital though that learners are taught that although some aspects might benefit from language transfer, they should never rely on the fact that it will work every time (Drew & Sørheim, 2016, p. 17). Comparing the vocabulary between the two languages, one can see that they have multiple cognates between them, but there are also many false friends that learners need to be aware of when writing. Many English words can also be difficult for Norwegian learners to spell correctly as English has a deep orthography, meaning that the language is very inconsistent in its reliability of correspondence between print and speech (Schmalz et al., 2015). An example of this can be seen in the different pronunciation between similarly spelt words such as *through, though, cough, tough*, and *thorough*.

Related to language transfer is the interlanguage hypothesis. Interlanguage can be seen as step before language acquisition, which is never completed but is always in development (Aljumah, 2020). In other words, it can be seen as the transitional period between a learner's L1 and the L2 which he/she is learning. As the learner has not become completely capable in the L2 nor his/her own language, structures from both languages may be displayed when writing in the L2 (Aljumah, 2020). The interlanguage is different for each individual and it is based on each learner's current comprehension of the L2. These structures are formed from using various learning approaches such as simplification and language transfer.

While learners typically possess a vocabulary of several thousand words and the ability to grasp grammatical structures in Norwegian as they begin primary school, once they start writing in English things become more difficult. Not only do learners have to learn how to write, they must also learn English at the same time (Hyland, 2003, p. 34). Due to these setbacks, language learner texts written in English are usually less effective than texts written by native speakers of English. The texts are usually shorter, have less cohesiveness, are less fluent, and contain more errors (Purves, 1988, as cited in Hyland, 2003, p. 34). This could be

one of the reasonings behind why, especially EFL learners at the beginning and novice stages, gravitate towards Google Translate and other machine translation tools as these tools may provide a superior English output than what they can produce themselves. What learners fail to realise, however, is that error production is a major part of language learning as errors enable teachers to react to the observable thought process of the learners and give appropriate feedback to the various errors being made (Brown, 2000, p. 66).

2.3.4. GRAMMAR-TRANSLATION METHOD

The grammar-translation method is a method of teaching L2 that was used extensively in the 20th century, with the focus being on teaching learners how to read and appreciate foreign language literature. By inspecting the grammar of the L2 that the learners were studying, there was a hope that they would also become more familiar with the grammar of their L1 which in turn would cause them to both read and write better in their L1 (Larsen-Freeman & Anderson, 2011). This method of language teaching focuses on reading and writing, as when the approach was conceived, it was used in the teaching of Latin and Greek, languages that were dead, so to actually speak them was not worth considering (Drew & Sørheim, 2016, p. 23).

In a Norwegian context, this method of teaching involves learning vocabulary lists by heart as well as doing grammar exercises and translating texts and passages to and from English. The teacher usually speaks Norwegian with some exceptions where English is used instead such as in asking comprehension questions related to texts that were read earlier. Producing written texts in English and translating texts between Norwegian and English are standard activities in class (Drew & Sørheim, 2016, p. 23). As new school reforms were introduced and altered the curriculum, this method is now employed less in schools, as other language teaching methods that focus more on communication are more widely used, such as the communicative method.

Regarding positive aspects in relation to the grammar-translation method, Hell (2009, as cited in Mart, 2013) speaks of several. He considers the use of the grammar-translation method as positive due to how it enables learners to enrich their vocabularies, increase the number of figures of speech they can use and develop their ability of interpretation. By studying the best writers, Hell argues that it allows learners to produce similarly good texts due to how translation forces them to notice details that escape the attention of other readers. Furthermore, Stern (1992, as cited in Mart, 2013) emphasises that translation is important in language learning. Translation is a useful resource that enables

learners to see the similarities and differences between their L1 and their L2 which supports improvement in understanding the language system. By comparing the L2 to their L1, they may use the L2 more effectively.

On the other hand, the grammar-translation method does have its disadvantages. Firstly, this approach focuses heavily on reading and writing, with little regards to use of oral language (Drew & Sørheim, 2016, p. 23). With the *LK20* reform and earlier reforms, a lot of emphasis was given to the use of oral skills in English, making this method for language learning less beneficial. That is not to say that it cannot be used at all, rather that it should not be used exclusively. Another disadvantage to the grammar-translation method is the fact that learners are not active participants in the classroom (Eisa, 2020, p. 385; Larsen-Freeman & Anderson, 2011). Although the old approach of grammar-translation is seldom used for language teaching in Norway, modified versions of it have been introduced into the FL classroom as more modern methods were used in new textbooks and teaching materials (Drew & Sørheim, 2016, p. 23). Further, Chang (2011) argues that the old grammar-translation method is not suited for younger learners as its original intended use was for those who were already highly educated with knowledge of classical grammar. In addition, he argues that this method is inappropriate to use in group teaching in the classroom as it is primarily a self-study method. However, arguments can be made that the grammar-translation method can be used in the classroom for language teaching, although not exclusively. As translation tools and teaching materials that aid learners have progressed significantly since the 20th century, it has become easier to translate from one language to another and thus provides learners with a better activity to engage in.

MT tools have the ability to replace some of the key roles that the teacher is usually tasked with doing as well as changing some of the core fundamentals that the grammar-translation method possesses (Tsai, 2020). The use of MT tools enables learners to become active participants in the classroom rather than passive on the condition that they have their own device. In this context, active is defined as the learners themselves finding translations of words or phrases rather than the teacher providing the information for them. This shifts the interaction from between the teacher and the learners, to between the learners and their device. Each of them can use Google Translate or other MT tools to discover similarities and differences between two languages. Furthermore, the use of MT enables instant feedback to the learners when they are writing, freeing up some of the workload for the teacher. By giving learners MT tools and guiding them in how to use the tools efficiently, they may acquire a

larger vocabulary, learn figures of speech and learn grammatical structures in L2 due to the instant feedback that they receive (Niño, 2009; Alhaisoni & Alhaysony, 2017; Tsai, 2020). However, all of this is dependent on that the MT output is accurate, as inaccurate MT output will only promote incorrect vocabulary and blur the similarities and differences between their L1 and the L2 that they are studying. As will be seen in a later section, although Google Translate is far from being a perfect translation tool, its quality of output can be very high if certain conditions are met.

The grammar-translation method considers learners who are able to translate from one language to another as successful language learners (Lars-Freeman & Anderson, 2011). Although the conventional method of grammar-translation did not take machine translation (MT) into consideration as such tools did not exist at that time, learners today can technically be considered successful language learners with the use of Google Translate (GT) or other MT tools. According to this line of argument, if learners are able to produce a near perfect output with the assistance of GT or other translation tools, the author of this thesis argues that they can be considered successful language learners. This is because the language learners are showing that they possess the ability to spot differences and similarities between their first language and their target language, which is also one of the learning aims of the *LK20* curriculum that learners are expected to master after year 7: "explore and talk about some linguistic similarities between English and other languages that he or she is familiar with and use this in his or her language learning" (Utdanningsdirektoratet, 2020). They also show proficiency of the grammar rules in the target language as they must approve of the translator's output, thus proving themselves to be successful language learners according to the grammar-translation method. However, discussions must be made whether the learners can be considered successful language learners if they decide to copy and paste the output without processing any of the information presented to them.

2.3.5. WRITING AS A PROCESS VS WRITING AS A PRODUCT

In the context of writing, there are several approaches that teachers can choose from in teaching learners the writing skill, all of which have their strengths and weaknesses. In this thesis, two of the more well known and most popular methods of approaching writing will be discussed and how they can be applied in the FL classroom together with ICT tools, namely the process approach and the product approach.

The process-based approach of teaching writing looks at writing as a process in which multiple activities such as pre-writing, drafting, revising, and editing are involved in a reflective way (Osanloo & Kolahi, 2016, p. 87). Further, Emig, Meyers and Raimes (1977; 1983; 1986; as cited in Gomez, 1996, p. 210) advocate for process-based writing by arguing that writing is a tool for learning and self-discovery, not just a means to demonstrate learning. Learners are expected to participate actively and work during the whole process of the written work instead of only looking at the finished product. In Norway, process-based writing has been done in both L1 lessons and L2, although it seems that process writing has been more widely used in L1 lessons (Drew & Sørheim, 2016, p. 97). Drew and Sørheim further argue that process writing should be used more in L2 lessons than what has been done earlier as learners need just as much, if not more help in writing in their second language than in their first language.

The pre-writing stage focuses on finding ideas that one can write about, which can include oral discussions, brainstorming sessions, and reading texts (Drew & Sørheim, 2016, p. 98). In the first draft stage, learners find ways of organising and presenting these ideas so that they become simpler for the reader to understand. The focus of the first draft is content rather than language, as focusing too much on the language at this stage might take away valuable time from the actual completion of the text. The revision stage focuses on giving learners constructive feedback on their current progress. Assuming that machine translation supplies learners with a perfect one-to-one output of their input, it has the opportunity to replace the teacher's role of giving feedback related to grammar and possibly content in the future when learners input phrases and sentences as MT becomes even smarter and more features are added. However, at present, machine translation is not intelligent (or unintelligent) enough to give a perfect output of the input that it has been given, meaning that teachers still need to assess the language aspect of texts that are written by learners and not let them rely solely on machine translation. In the editing stage, focus is given to the formal components of the text such as grammar, vocabulary, spelling, and punctuation (Drew & Sørheim, 2016, p. 101). At this stage, a different type of ICT tool than machine translation, such as a grammar checker can prove useful in aiding learners with feedback regarding sentence structure, grammar, spelling, and punctuation.

However, it is very important to note that none of these ICT tools have perfect accuracy as of now, meaning that teachers and learners themselves need to be critical and assess the output that they are being given. Some of the pedagogical ways that ICT tools can

be implemented into the classroom which will be discussed later share a lot of the same processes that writing as a process has. Especially the editing stage in the process-based approach is quite similar to one pedagogical use of implementing machine translation tools into FL learning, namely post-editing.

The other approach to writing that is of interest to this thesis is writing as a product. Product-based writing supplements the grammar-translation method because they both share similar aspects in how they give premade texts for learners to study (Larsen-Freeman & Anderson, 2011; Osanloo & Kolahi, 2016). The product-based approach focuses on the product of the text, where learners are given model texts written by an adequate writer to read and mimic the qualities of what makes that text good in order to become good writers themselves. Supporters of this approach believe that learners can learn how to become good writers and produce little error when they are given a well-made text to study before they start writing (Nunan, 1999; Adams, 2006; as cited in Osanloo & Kolahi, 2016, p. 87). Learners are also given writing exercises that reinforce language structures that they have learnt through the imitation of grammatical patterns, much the same as in the grammar-translation method (Osanlo & Kolahi, 2016, p. 87). In this approach, the focus is on the product that is being written rather than how learners should approach the various stages of writing. Steele (n.d.) details four stages of a product approach to writing. The first one is reading model texts and highlighting features of the genre that is being read. The second stage is controlled practice of the highlighted features read earlier, usually in isolation. The third stage is the organisation of ideas where the organisation is more important than the ideas themselves and the last stage is where the learners individually choose a writing task that incorporates the skills, structures, and vocabulary that was highlighted earlier to produce a product.

A product-based approach to writing can also be accompanied by ICT tools to help learners in the FL classroom when writing. Although not perfect, certain grammar checker tools on the Internet can prove useful to learners at the last stage of this approach in assessing the various components of the text that they have produced such as vocabulary and sentence structure. Just as in the previous approach, it is important to note that these tools are not perfectly accurate, so some caution has to be taken by both learners and teachers. One such program is Grammarly[3], which guides learners on whether there is anything to improve in the text inputted, and if so, it highlights various words, phrases, and sentences that should be

---

[3] Grammarly.com to see more. User creation is required to enable specific highlights on how to improve the text inputted. Some features are locked behind a premium fee.

edited. If users copy parts of their texts into this program, feedback is given on aspects such as correctness i.e., syntax errors, spelling mistakes, and sentence structure. It can also make recommendations to the user regarding clarity, engagement, and delivery. The user can also customise how the program considers the input, whether the audience of the text is the general population or people with expertise within the field that is being written about. It can also be customised according to formality, tone that the text should emit, the intent of the text and more. By researching and analysing these types of ICT tools, teachers can find ways of integrating them into FL lessons and transfer some of the workload, for instance giving feedback to learners, over to machines. It is this author's belief that this way of approaching ICT tools in the classroom should, as of now, only be done by teachers who are confident and knowledgeable of how these tools work.

2.4. CALL (COMPUTER-ASSISTED LANGUAGE LEARNING)

Computer-Assisted Language Learning (CALL) can be defined as any process in which a learner uses a computer and, as a result, improves his or her language (Beatty, 2010, p. 7). This definition is quite broad, however, the field of research within CALL is rather fragmented and its use covers a broad range of activities. It still proves as a useful definition for this thesis, and it relates largely with the earlier subsection on ICT.

CALL is used regularly in schools and other education centres where learners are exposed to language learning. Although the focus of this thesis is EFL, CALL can be used to aid in the learning of any foreign language. In Norwegian schools, there is sufficient access to ICT tools and equipment to use CALL in language teaching. Especially in the upper years of primary school, it is not unusual for the learner to ICT device ratio to be 1:1 (Fjørtoft, Thun & Buvik, 2019, p. 24). A majority of primary and lower secondary classrooms in Norway are fitted with smartboards and schools typically have computer rooms where learners can take advantage of CALL tools to promote learning (Fjørtoft, Thun & Buvik, 2019, p. 56). Rather, the limiting factor on how much and well CALL is used for language learning is the teacher. The quality of CALL to stimulate language learning is therefore very dependent on how much teachers are educated in its use and the knowledge they have surrounding the quality of the various tools within CALL that exist.

Languagenut[4] and Duolingo[5] are two examples of CALL applications that can be used to improve language learning. Especially Duolingo is popular in school sectors and for private use to promote language learning for learners at various stages. This tool provides language learning opportunities for learners just starting a new language at Common European Framework of Reference (CEFR) level A1 all the way up to bilingual or native learners at CEFR level C2, the highest level of proficiency, meaning that it can easily be adapted by teachers to provide challenges to each individual learner in their class. With the technological advancements during the past ten years or so in the smartphone industry, CALL tools are now more readily available than ever. Duolingo in particular is used by millions of language learners to either learn new languages or improve upon old FL knowledge due to its accessibility, lack of cost, and effectiveness (Jiang, Rollinson & Blanco, 2020). The portability and accessibility of this CALL tool is useful for language learners in schools where they are able to easily continue from where they left off in the FL classroom by using the same login credentials on the app on their smartphones or on a laptop at home, as progress is saved across devices. Some studies have been conducted on the effect Duolingo has on language learning, with the one from Loewen et al. (2019) in particular indicating a moderate positive correlation between the amount of time spent on Duolingo and learning gains.

CALL therefore creates opportunities for learners where they are able to study by themselves and even go back to practice previous parts of a language where they struggle, independent of teachers. However, the autonomy that learners are given by these tools can at times be troublesome in programs that follow a lock-step scope and sequence as they provide learners with only limited opportunities to organise their own learning or tailor it to their needs (Beatty, 2010, pp. 11-12).

Padurean and Margan (2009, p. 98) discuss how CALL has evolved into supplying instructional materials to learners in FL classrooms and how activities that involve CALL supply learners with a wide range of tasks to create engagement. The use of computers as instruments for understanding language using spelling and grammar checkers as well as other editing programs is also brought up. Google Translate, Grammarly, and a multitude of other machine translation tools, as well as grammar checkers, can therefore be categorised as CALL tools as they have the possibility to assist in language learning if used appropriately.

---

[4] https://www.languagenut.com
[5] https://www.duolingo.com

Meskill and Pennington (1996; 1996; as cited in Farzi, 2016, pp. 49-50) offer further support for CALL, stating that computer tools and software facilitate the teaching and learning of L2 writing. By using CALL tools in FL classes, L2 writers are offered various advantages over normal writing in the form of easy access to saving, organising and editing documents, access to helpful tools such as grammar checkers and spell checkers, and other software that can enhance the quality of their L2 writing. Pennington (1996, as cited in Farzi, 2016, p. 50) delves deeper into the advantages of CALL by categorising said advantages into five areas: Quality of written work, quantity of writing, writing process, revision behaviour, and effective social outcomes. She concludes that L2 learners writing with computers end up with higher holistic and analytic ratings of compositions, produce longer texts, edit their texts more often, and experience a better attitude towards L2 writing. It would seem that with a good implementation of CALL into language classrooms, learners have the potential to perform better, both qualitatively and quantitatively, compared to traditional language teaching, furthering the argument that a more substantial ICT education for teachers should be implemented to promote higher quality learning for learners.

Somers (2003, pp. 325-326) provides input regarding teaching L2 learners proper use of machine translation tools, suggesting that due to the extent that translation is used in L2 learning, MT and translation software should be a part of the curriculum for language learners. He goes on to mention that other researchers go beyond this and suggest that translation software can be used to strengthen different aspects of language learning, thereby calling machine translation a CALL tool. Although the amount of research conducted on the use of Google Translate and other machine translation tools in schools is limited, and some of it outdated, studies such as the one from Lee (2020) and the one by Bahri and Mahadi (2016) are supportive of GT functioning as an effective CALL tool in writing in the FL classroom. However, Lee ends his study by saying that in order for GT to function as a beneficial CALL tool, teachers have to be aware of its limitations and provide proper guidance to learners.

2.5. MACHINE TRANSLATION

2.5.1. INTRODUCTION

In Douglas Adams' The Hitchhiker's Guide to the Galaxy, a small yellow fish named the Babel fish was all one needed in one's ear to be capable of understanding any language that was spoken as the fish would perform real-time translations. When the series started in 1985,

Adams probably did not envision that such a thing would ever exist, but as of the writing of this thesis, technology is inching ever closer to it becoming a reality. In 2017, Google launched its first version of the Google Pixel Buds, a pair of earphones with one of its prominent features being real-time translation between dozens of languages. Although the feature was not impeccable, it served as a proof of concept of what is to come of machine translation in the near future. An updated version of the Google Pixel Buds was released in April 2020, which improved the real-time translation even further, making it even more instantaneous. The Pixel Buds are an exceptional example of the real-life application that MT can have, with Google Translate working as the brain behind its real-time translations. Looking at the past to see how MT became what it is today will prove useful in understanding where the future of MT software might go next.

As noted at the end of the last subsection, Google Translate is one example of a CALL tool that can aid in language learning. It is one of many various MT tools available for the public to use in translating words, texts or phrases from one language to another, but what defines an MT system? Microsoft (2020) defines them as applications or online services that use machine-learning technologies to translate any amount of text from a source language into a target language. In order to better understand the context behind the role MT can play in language learning and its limitations, a brief history of it will first be addressed as well as covering some of the algorithms that MT software has used earlier and use today.

2.5.2. A BRIEF HISTORY ON THE EARLIEST MACHINE TRANSLATION SYSTEMS

Thierry Poibeau (2017) details the history of machine translation (MT) systems, from their emergence up until recently. Within MT history, there have currently been four generations of approaches to MT, and they will be examined briefly. In order from oldest to newest, they are the rule-based MT systems, the example-based translation systems, the statistical machine translation (SMT) systems, and the neural machine translation (NMT) systems. When MT systems were first developed, three different systems to translate between two languages were the starting point for the development of future translation systems. These three systems were direct translation systems, transfer systems, and interlingua systems, all of which properly kickstarted the era of MT in the 1950s, partly due to the newly developed computer. At this point in time, MT systems were not meant for public usage as computers were mostly reserved for academic, research, and military purposes.

Briefly summarised, direct translation systems correspond to word-for-word translations. This framework for translation does not analyse the source text in depth and in its simplest form, it can use bilingual dictionaries to translate from the source language to the target language with some tweaks to morphology and syntax (Poibeau, 2017). With the vast differences in syntax and other grammatical areas between languages, this strategy for translation does not prove particularly useful. However, it can at times output understandable translations if the two languages are considered close to each other (Poibeau, 2017).

Transfer systems on the other hand, divide sentences into chunks or linguistic units, as the structure of sentences are too variable to translate as a whole. The systems then translate these linguistic units via specific rules that are implemented, such as how adjectives in French are usually placed after the noun, while in English they are usually placed before the noun (Poibeau, 2017). More complex rules can also be implemented in how these linguistic units are to be translated, for example at the semantic level in order to determine the appropriate meaning of a word concerning the chunk it belonged to, for instance, if the word bark relates to the verb that dogs do versus the noun that one can find on a tree. When this system for translating between languages first emerged, it was difficult to implement properly as it was impossible to predict all the contexts in which words and linguistic units could be used and therefore not feasible to implement manually (Poibeau, 2017).

The last system out of the three that emerged in the 1950s, is the interlingua system. This system shares some aspects with transfer systems but differs in that it employs the use of an interlingua (Poibeau, 2017). In transfer systems, translation always concerns two different languages, therefore needing adaptation for every new language couple that is going to be translated (Poibeau, 2017). The interlingua system, however, first translates the content from the source language into an interlingua, a representation that is language-independent and attempts to maintain the meaning, that is to say, the characteristics, syntactic structures, and other grammatical rules from the source language. This interlingua can then be used to translate into any other language by employing a "generation module", a module that looks at the rules and structures that need to be applied first in order to produce proper sentences in the target language into which one wishes to translate (Poibeau, 2017).

In modern times, especially once SMTs and later on NMTs were introduced, the use and benefits of rule-based machine translation systems were outweighed by the other two and could not compete in terms of the quality of translations, making rule-based systems near obsolete. However, SMTs and NMTs require massive amounts of source material and

computing power to provide translations of good quality, while rule-based machine translation can operate on smaller servers. Some online translators (OT), such as Apertium[6], still operate with rule-based systems for its translations with various degrees of success. However, it is important to note that this OT does not seem to be heavily supported any longer, meaning users should be cautious in its use (Forcada et al., 2011, p. 128). Interestingly, Apertium seems to be the only web-based OT that can translate between Bokmål and Nynorsk, making it useful for Norwegian learners, even with its shortcomings.

Example-based machine translation was introduced in the 1980s in Japan as an answer to the difficulties in translating between Japanese or other Asian languages and English due to the languages being so vastly different that rule-based translation systems usually could not provide adequate translations (Poibeau, 2017). With the increasing complexity of rule-based translation systems, these systems also became increasingly difficult to maintain, furthering the need to explore new methods of approaching MT. Put simply, example-based MT looks at earlier translations made between the source language and the target language in existing bilingual corpora that one wishes to translate between, to see if similar translations have been made before (Way & Gough, 2005, p. 298). If fragments of the written sentence match with earlier translations, it will imitate them. It then swaps out any unknown words that do not match in the fragment by looking up the corresponding word in a bilingual dictionary (Poibeau, 2017). This method of translating between two languages increases in quality the more previous translations there are to search through. Although this method of translating is not really used anymore by itself, aspects of it are at times used by SMTs as a hybrid-system to produce higher quality translations compared to what they are able to perform by themselves (Groves & Way, 2006, p. 301).

2.5.3. MACHINE TRANSLATION SYSTEMS IN MODERN TIMES

In the mid-2000s and early 2010s, SMTs became the norm for most OTs and MTs such as Microsoft Translator, Google Translate, and Yandex Translate, however, research into creating an SMT system already began in the 1980s and 1990s (Och, 2006; Yandex, n.d.; Poibeau, 2017; Microsoft, 2020). SMTs employ advanced statistical analysis to estimate the best possible translation for a word given the context of surrounding words (Microsoft, 2020). This system performs these translations by applying algorithms to vast amounts of corpora

---

[6] https://apertium.org

made up of similar texts in different languages in attempts to understand the patterns between the languages (Poibeau, 2017). In other words, SMT is purely data-driven, it does not rely upon rules of the source language nor the target language. It analyses the extensive number of corpora that it has access to and finds a corresponding word or phrase based on probability and statistics. Within SMT, there are three different methods for approaching translations: Word-based SMT, phrase-based SMT, and syntax-based SMT.

In its infancy, SMT operated with a word-based method of translation where it would use advanced algorithms to determine the correct translation for each word (Poibeau, 2017). As is implied by its name, this method looks at sentences at word-level to produce word-for-word translations, although it will at times produce more words than inputted due to one word in the source language corresponding to several words in the target language. As Poibeau (2017) describes, going for a word-for-word translation approach is generally considered a poor solution because too many basic errors arise. This is due to there being too many variables and too much contextual information in each sentence for word-based approaches to produce good quality translations.

Phrase-based SMT is the most widely used method of translation out of the three SMT systems and differs from word-based SMT in that instead of focusing solely on each word, a phrase-based SMT learns to translate words along with phrases (Zens, Och & Ney, 2002, p. 22; Systran, 2016). In simple terms, this method for translation segments sentences into words and phrases, and then it translates each phrase and word based on statistics and probabilities derived from looking at large amounts of corpora into what it thinks is the correct translation. It then merges the phrases and words to compose the finished sentence in the target language. In theory, this enables translations with a higher quality due to being able to take context into consideration to a greater degree than the previous method (Poibeau, 2017). Up until the transition over to NMTs, phrase-based SMT was the best method for approaching translation as its output accuracy was shown to be the highest (Koehn, Och & Marcu, 2003).

To briefly explain syntax-based SMT, it operates by furthering the work of phrase-based SMT by breaking down the phrases of the source language into syntactic units such as verb phrases, noun phrases, pronouns, and more. It then reorders them to fit the target language, thereby improving word alignment over phrase-based SMT (Williams et al., 2016). There was a time when it seemed that syntax-based SMT would become the successor to phrase-based SMT just as phrase-based SMT was the successor to word-based SMT

(Williams et al., 2016). However, before it managed to surpass phrase-based SMT to the extent where it was able to produce translations of significantly higher quality, NMTs were introduced and became the next generation of machine translation systems.

Most research conducted on the use of machine translation tools for educational purposes was done during the timeframe when the phrase-based statistical machine translation system was most widely in use. As the major OTs, such as the ones from Microsoft and Google[7] do not operate solely with this MT system any longer, it is important for new research to be conducted on the topic as earlier results and conclusions can be misleading in the present-day.

Today, most translation systems, whether they are free OTs or MT software designed for businesses, operate using NMT. Some of the companies that have created their own NMTs include Google, Microsoft, Amazon, Facebook, and Linguee. In very simple terms, NMT has a two step-process to translate from the source language into the target language. The first process, encoding, analyses vast amounts of training data and translates the source text into an interlingua, a language which only a computer can comprehend and is reminiscent of the rule-based interlingua machine translation system (Poibeau, 2017). This differs from phrase-based SMTs which typically use English as the interlingua. The second process, decoding, automatically produces a translation from the source text into the target language based on the data that the encoding process created (Poibeau, 2017).

The advantage of using an NMT system is that it translates sentences at word-level, phrase-level, and sentence-level in order to better gain awareness of the context in sentences so that it can improve the quality of translations (Systran, 2016; Turovsky, 2016a; Poibeau, 2017). Furthermore, NMT systems learn over time, meaning that the more they are used, the smarter they become, thereby improving the translation quality of languages the more they are used (Turovsky, 2016a). While a human has to design the NMT system in regard to training regime and network architecture, most of the learning process is done by the system itself, meaning that rules of syntax and other linguistic features do not have to be inputted manually as the system will learn it on its own if the data given is of proper quality (Goldberg, 2017, p. 3).

---

[7] Latin is, at the time of writing, still translated by Google Translate with the use of phrase-based machine translation, a method within Statistical Machine Translation. https://cloud.google.com/translate/docs/languages

Before reporting on empirical data of how machine translation affects writing production in a school setting, teachers' and learners' perceptions of machine translation will be discussed briefly. Niño (2009, pp. 248-250) conducted a survey on learners' perceptions on the use of MT in FL learning and reported that a majority of learners surveyed reported a favourable and positive attitude towards the use of MT tools for FL learning, especially when they combined it with post-editing where they were able to put their language knowledge to use. On the other hand, teachers reported a less favourable view towards machine translation tools in the classroom. Teachers cited low quality output, the required training in using MT tools, and text type constraints, i.e. MT tools not accurately translating poems as well as factual texts, as factors for showing scepticism towards these tools (Niño, 2009, pp. 250-252). The teachers surveyed did however hope to incorporate MT tools into FL learning in the future once the output was of higher quality. As eleven years have passed since those surveys were conducted, MT tools have become a lot better at providing more accurate translations, meaning that teachers' perceived views of these tools might have changed drastically. Clifford, Merschel, and Munne (2013) had similar results from their survey, showing a large discrepancy between learners' and teachers' views on the usefulness of machine translation tools in language learning. Whilst learners found MT tools to be useful as dictionaries and as tools to double-check their work, teachers were more sceptical of the positive impact that MT tools can provide in language learning.

A recent study from Stapleton and Leung (2019, p. 25) seems to indicate that teachers' attitudes towards machine translation tools are changing. Although all twelve teachers interviewed agreed that MT tools should not be used by learners to translate passages or sentences from L1 to L2 that learners themselves had written, nine of the twelve teachers were not against MT tools as learning tools to help learners translate at word-level or phrase-level. Some teachers commented on MT tools as being powerful education tools and appreciated it when MT tools were used to strengthen learners' language skills. Many teachers from the study did, however, show some confliction in the use of MT tools in the sense that they believed that these tools were only beneficial if used correctly, with some of them suggesting that schools and teachers should provide guidance to learners so that these tools are used properly for learning.

Earlier data on the use of machine translation tools for educational purposes mostly reports positive trends on the effect they have on the written production of learners in FL

classrooms, however, some studies seem to indicate a neutral outcome from using MT tools. As written above, Stapleton and Leung (2019) conducted a study where they had twelve teachers mark texts created by year 6 learners in primary school. The texts were written in Chinese by the learners themselves before the learners translated them from Chinese to English. The researchers also translated some of the texts from Chinese to English using Google Translate, however, they did not tell the teachers that some of the texts they were marking were translated by a machine translation tool. A majority of the teachers commented on the overall quality of the texts translated by GT as ranging from typical to very good and impressive (Stapleton & Leung, 2019, p. 24). Overall, grammar was seen to be of average or above-average level compared to what the teachers were used to, while teachers had widely differing views regarding the range and quality of the vocabulary seen in the texts. Some teachers experienced that GT had given the texts advanced vocabulary, while other teachers experienced that the texts only contained basic words and lacked variety. However, the teachers' different experiences might be explained by the variance in the input vocabulary. In terms of comprehensibility, none of the teachers had any issues of understanding the texts given to them.

A study by Fredholm (2014) was conducted to investigate the effects of using machine translation tools in Spanish FL learning, more specifically morph syntactic and lexical-pragmatic accuracy. The study was conducted on Swedish learners aged between 17 and 18, and it was done by comparing the written product between two groups, one with free access to the Internet and the spelling and grammar checker of Microsoft Word, while the other could only use printed dictionaries without internet access, nor did they have access to the spelling and grammar checker of Microsoft Word. Most learners in the first group employed the use of MT tools, specifically Google Translate to help in the writing process. The results from the study were overall neutral as it seemed that the use of MT tools compared to writing with no aid other than a dictionary, did not affect writing performance negatively nor positively. It is worth mentioning that in this study, 84 essays were received by the first group with access to the Internet, compared to 28 from the group with no internet access, meaning that the conclusion reached by Fredholm (2014) can be argued to be uncertain due to the vast difference in the amount of data that was collected between the two groups.

To briefly summarise a few more studies, one study by Garcia and Pena (2011) was conducted to look at how MT tools affect language learning for beginners, specifically

English students learning Spanish. The results from that study indicated that language learning beginners not only wrote more, but also better when using MT tools to aid in writing in Spanish compared to writing directly in Spanish. However, writing directly in Spanish seemed to require more effort, therefore creating more engagement with the task. A different study by Garcia (2010) one year earlier, concluded a similar outcome, learners at the beginner stage benefit by using MT tools to communicate more and better compared to not using them. On the other hand, in this study as well, learners seemed to learn more when writing directly in L2 without the use of MT tools due to them engaging more with the task. Garcia (2010) reached this conclusion by studying editing interventions of the participants that showed engagement levels which indirectly showed learning taking place based on the number of successful and unsuccessful edits.

Giannetti (2016) and Garcia (2010) also both provide empirical data, showing that learners write more and with a more advanced vocabulary when using MT compared to that of directly writing in an L2. However, research is still scarce on whether learners actually retain the more advanced vocabulary they show in essay writing for use in future writing. Both O'Neill (2012) and Al-Tuwayrish (2016) conclude their studies by saying that the field of MT for educational purposes is nascent and a lot more research needs to be conducted before proper conclusions can be drawn on the effects and roles MT tools have in language learning.

A few more noteworthy points surrounding the use of MT tools for educational purposes is that online translation tools used to struggle heavily to maintain a high level of accuracy when translating between English and languages whose syntactic structures vary from English (Giannetti, 2016, p. 18; Aiken & Balan, 2011, as cited in Tsai, 2020). This can still hold true, but as the algorithm behind how MT tools translate languages has changed, this claim might not be as accurate today. Online translators also seem to provide the highest quality output when translating technical, scientific, and academic texts due to their lack of figurative language, something that MT has been known to struggle with at times (Shih, 2016, as cited in Giannetti, 2016, pp. 18-19). The last point to address is the positive impact OTs have when used as dictionaries as they provide near-instant speed translations that can result in improved comprehension as less time is spent away from the text. Just as print dictionaries can help promote vocabulary development for language learners, OTs and online dictionaries can do the same with the advantage of these ICT tools being that they can take less time away from the texts (Jim & Deifell, 2013, as cited in Giannetti, 2016, pp. 26-27; Giannetti 2016).

2.6. GOOGLE TRANSLATE

Google Translate (GT) is a free online translation tool released by Google in April 2006 (Och, 2006). It is used by millions of people every day to translate between all kinds of languages and for different purposes (Turovsky, 2016b). Not only can GT translate text inputted by a keyboard, in a lot of languages it also has the ability to translate uploaded documents, handwritten texts, spoken conversations, images, and perform real-time translations with a user's phone by using the phone's camera (Google Translate, n.d.). Google Translate is also embedded into many ICT tools, such as the web browser Google Chrome which uses GT to translate whole web-pages into a language of the user's choice. GT is also used in hardware that utilises Google Assistant, an artificial intelligence-powered virtual assistant; for example, most phones running on the Android operating system, Google's Pixel Buds or Google's smart speakers, Google Nest. In the first ten years of its release, it operated by using a phrase-based statistical machine translation approach, before starting a transition over to a neural machine translation approach in 2016 (Turovsky, 2016a). The OT currently supports the translation of 109 languages, and at the time of writing, GT now uses an NMT approach to translate between 108 of its supported languages, with all the various methods of translating named above using NMT. Latin is currently the only language that GT still uses an SMT approach in translating to or from other languages (Google Translate, n.d.).

As explained, Google Translate used to operate with Google's own phrase-based SMT system but changed to using Google's neural machine translation (GNMT) in November 2016 (Schuster et al., 2016). GNMT translates between languages with comparable algorithms to other NMT systems, but as all NMT systems are created by different companies, the output they provide will vary greatly due to the tiniest differences in either input or how the various algorithms function. To provide an example of how different NMT outputs can be, the Norwegian question written in a Stavanger dialect, "Ka kan eg gjør for deg?" was translated in GT and Microsoft Translator into English to compare how they both handle translations that can be deemed as problematic. GT provided the output *What can I do for you?* which is the correct translation, while Microsoft Translator outputted *Ka can eg do for you?*, only translating the words that were spelt identically in both a Stavanger dialect and Bokmål. This is just one example of how vastly different the intelligence between OT and MT translators using different NMT systems can be.

Manually training the GNMT system in being able to perform translations for all of its 108 languages that it currently supports would be an almost impossible task due to the

sheer number of combinations of language pairs that one can translate between. One of the ways GNMT handles translations between all its language pairs is by using what Google calls *Zero-Shot Translation* (Schuster et al., 2016). Simplified, *Zero-Shot Translation* allows the GNMT system to perform translations between language pairs that it has never been taught to translate by instead comparing translations of other languages that the system has received training in. For example, the three following languages can be considered: Norwegian, English, and Finnish. The GNMT system receives training in translating between Norwegian and English, and between Finnish and English. By comparing and looking at the translations the system made in the other two language pairs, it is able to produce reasonable translations between Norwegian and Finnish (Schuster et al., 2016). Although this algorithm creates opportunities in providing more options for languages to translate between faster than training the system manually, it could also help to explain to some extent why GT occasionally creates weird and widely inaccurate translations between some language pairs.

The Google Translate user interface (UI) (see Figure 1) has a simplistic design in an attempt to make it as intuitive as possible for people to use the OT tool without creating confusion. The left side of the UI is dedicated to user input. In the left text box, users can input their own words, phrases or texts, or they can choose to upload documents that GT will translate. GT attempts to automatically detect the language that the user is inputting but users have the choice to manually select the language which they are inputting if GT picks the wrong input language. In the right text box, Google Translate performs instant translations of what the user is inputting. Users can choose which language they would like the output to be by selecting any of the 109 supported languages.
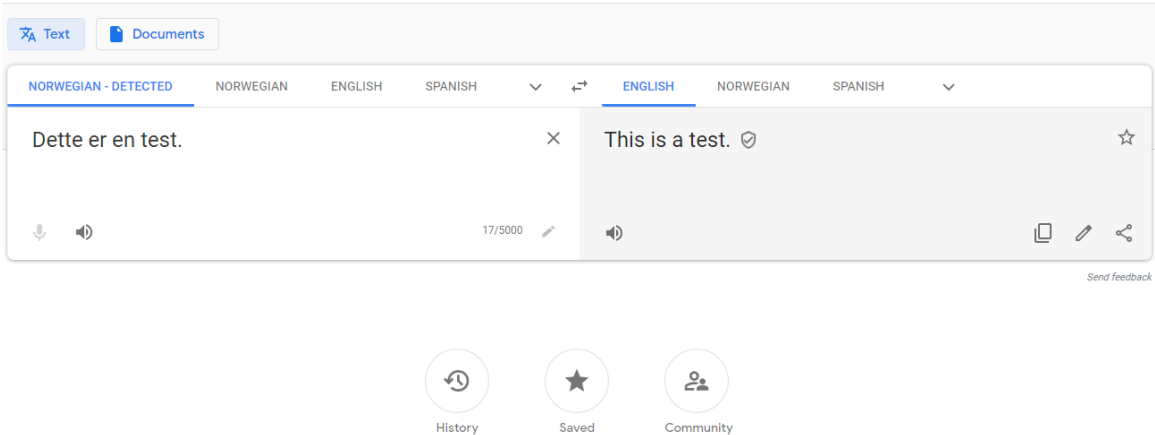
Figure 1: User interface (UI) of Google Translate

Users can dictate their input by clicking the microphone button in the input field, however, the accuracy of dictation is worse than manually typing, and especially in Norwegian, it struggles with dialects that pronounce words differently than how they are written in Bokmål. Looking at it from a different perspective, the dictation function of GT can prove very useful for language learners who, for example, have learning disorders such as dyslexia or learners who have reduced motor skills. Both the input box and the output box provide the ability to listen to the pronunciation of the text that is displayed, either in the source language or the output language by clicking the speaker button. This can prove useful for language learners in two ways, one being that they can learn pronunciations of new and unfamiliar words. The second reason being that it can help them determine whether a word they wrote is correct based on pronunciation if they are familiar with how it is supposed to be pronounced.

Google Translate also functions as a dictionary as well as being able to produce synonyms to the words that users input. As of writing, GT is unable to provide definitions of Norwegian words, although it does provide the functionality in English which is still a useful function for Norwegian language learners who wish to learn new and unfamiliar words. However, GT can provide synonyms in Norwegian for any word that is typed in, and if longer texts are inputted, users can highlight a specific word to see its synonyms, if any are available. Even though the aforementioned translator Apertium is, as far as the author of this thesis could find, the only OT that supports translations between Bokmål and Nynorsk, GT seems to have the capability of translating from both Bokmål and Nynorsk into other languages. This is useful as not all learners in Norway have Bokmål as their main language. Although GT can handle both Bokmål and Nynorsk input, only Bokmål can be outputted when translating to Norwegian.

Another important feature of GT and one that helps the MT tool in achieving higher quality translation outputs is user translations and the Google Translate community behind it. When translating from one language to another, a shield icon with a check mark within it will sometimes appear next to the output language (see Figure 1 and Figure 3). When this shield appears, it means that the Google Translate community has gone over this specific translation earlier and flagged it as being accurate. In order to prevent people from flagging incorrect translations as being correct, Google has a peer review process in place where other users voluntarily check other users' translations to see whether or not they are correct. This system is not perfect, and one should not trust every single translation with a shield next to it as being

31

correct, but it has helped the OT in achieving better translations for certain phrases or unusual words.

Addressing some earlier studies that discuss the positives and negatives of Google Translate, much has changed since those studies were conducted. For instance, Santoso (2010, as cited in Chandra & Yuyun, 2018) and Medvedev (2016) state that GT is unable to translate longer sentences correctly. Although, GT can produce errors, no matter the length of the sentence inputted, these statements seem to have lost some credibility. In order to test this claim, the author of this thesis tested GT by writing 15 self-composed sentences of lengths between 40-50 words, translating from Norwegian to English. When sentences were tested, the output never produced any errors. Despite the small sample size, this gives an indication that this claim may be questionable.

In addition, Giannetti (2016, p. 29) remarks in her research that OTs are unable to produce proper translations to words that are misspelt in English when translating to other languages. If GT is provided with words that are misspelt, even multiple words in the same sentence, it is still able to produce the translation that the user intended to write (see Figure 2). This works with other languages besides English; misspelt Norwegian words can still be translated correctly into their corresponding English word, but it varies based on the amount of training GT has had on those particular sets of misspellings.
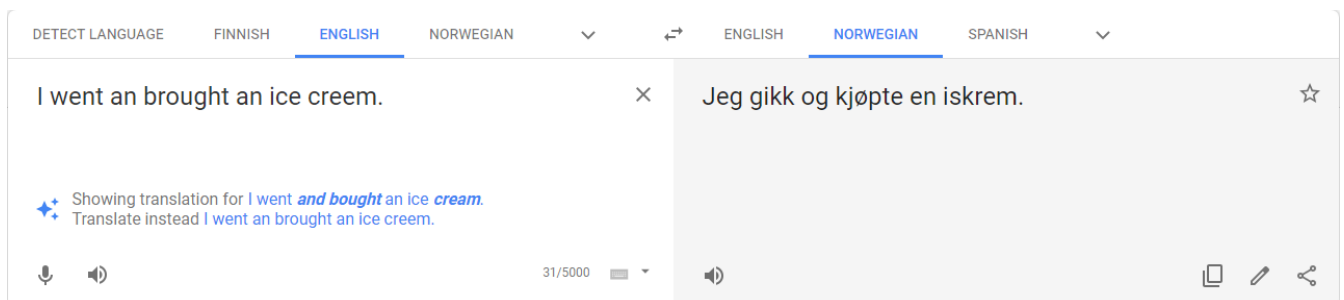
Figure 2: Example of GT successfully translating a sentence with multiple misspelt words

Santoso (2010, as cited in Octaviani & Jakarta, 2018) states that OT tools are incapable of handling idioms or figurative language. This can be investigated by inputting various idioms and other figurative phrases into GT to see what the output is. The author of this thesis inputted various Norwegian idioms and other figurative phrases into GT to determine whether it successfully translated them or not. Two popular Norwegian idioms, "prikken over i-en" and "gå rundt grøten" were inputted into GT, both of which were correctly translated into the English equivalents *finishing touch* and *beat around the bush,* respectively. By clicking the

output that GT produces, users can see and select other outputs to determine which equivalent corresponds best to the input. In this case, *beat around the bush* (see Figure 3) was one of the offered translations to "gå rundt grøten". It is unsure why *beat around the bush* was not the default output shown for the translation as it would be the closest translation that retained its figurative meaning. However, the default output still provided a correct explanation of the input. Other Norwegian idioms such as "gikk bort" to describe someone who has passed away and "hva er i veien" to ask what is wrong were also translated correctly by Google Translate, even when they were used in complete sentences. Some idioms, on the other hand, were only translated correctly by GT when there was a specific word order. An example of this was the phrase "her er det noe muffens" which translated into English as *here is something muffens*. However, changing the word order to "det er noe muffens her" resulted in the correct translation of *there is something wrong here*. This signifies that Google Translate is capable of handling numerous idioms but does struggle at times.
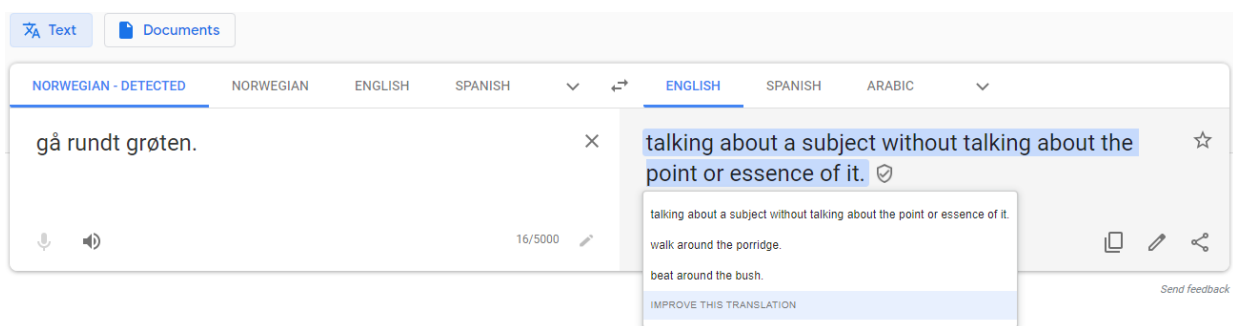
Figure 3: Example of GT translating a Norwegian expression into English

Addressing the perception of GT in a Norwegian school context, Aksnes (2018) conducted a study in two Norwegian upper secondary schools on the attitudes to the use of GT amongst teachers and learners by interviewing four teachers and four learners. Results from her study showed a mix of positive and negative attitudes towards GT. Some learners favoured the tool because of its capability of translating between a multitude of languages and because it is very quick and easy to use compared to dictionaries. Other learners and teachers were more sceptical of the output that GT produced, citing that the translation tool's grammar output was inaccurate and unreliable. Teachers were also torn on the idea of using GT as a tool in a language learning context, some willing to see its usefulness, whilst other teachers rejected its use completely, seeing no usefulness of the tool for any purposes.

As discussed previously, one of the strengths of Google Translate is the community verification process in which community members can suggest edits or peer review other translation contributions. Idioms and figurative language are examples of formulaic language

that benefit greatly from this as it helps GT to translate phrases correctly more quickly compared to only relying on its own training. This is due to the large amounts of idioms that exist in every language which would take a great deal of time for a few researchers to manually train the system in handling.

Google Translate is one of the most widely used OTs in the world, translating more than 100 billion words a day (Turovsky, 2016b), and it is also the OT most familiar to the author of this thesis which is one of the main rationales as to why it was employed in the study. GT has also evolved a lot since its creation in 2006 to handle more advanced input as well as more languages to translate between. It also has downfalls related to translating between certain languages and certain types of texts. However, it is important to reiterate that with the GNMT system, words, phrases, and sentences that are incorrectly translated by the translator at this moment in time have a high probability of being translated successfully in the future as the system learns and adapts every day.

2.7. THREE PEDAGOGICAL METHODS OF IMPLEMENTING MT IN FL LEARNING

2.7.1. INTRODUCTION

Online translation tools are being used increasingly by learners despite their flaws and errors in translations. As Bueno (1992, as cited in Farzi, 2016, p. 70) states in his report: "To think that students would invest hours translating with their pencil in one hand and their dictionary in the other, when the computer can do it in a few minutes, is at best wishful thinking." Schools and teachers need to adapt to the technological advancements that are emerging and learn to find ways of embracing them rather than prohibiting them from educational institutions. When learners decide to make use of MT tools despite their flaws in translations, teachers have no reason not to find ways of employing them to their advantage as well (Bueno, 1992, as cited in Farzi, 2016, p. 71).

As *LK20* constitutes the use of digital resources in language learning as one of its learning outcomes for learners after finishing year 7 in the English subject, it is a beginning, but it is the author of this thesis' belief that an even bigger emphasis on CALL tools should be incorporated into the curriculum. In the case of machine translation tools, pedagogical methods of implementing these tools have been studied by others, and positive effects have been reported by researchers on the effects of MT in FL learning. In this thesis, three methods

that approach pedagogical uses of MT will be discussed: Pre-editing, post-editing, and using machine translation as *a bad model*.

2.7.2. PRE-EDITING

One pedagogical use of MT in the classroom for aiding language learning is the use of pre-editing. Pre-editing is the modification of the source text in order to obtain an appropriate output from the MT tool (Niño, 2009, p. 243). By observing irregularities in the output, and then modifying the input to receive better output, Shei (2002) argues that this method boosts learners' learning both in the cognitive and affective domains as well as training the learners in using MT tools.

The learners in Shei's (2002) study first modified the input of a text in their mother tongue (Chinese), observed the output (English) and noted down the limitations and changes done by the MT tool. They would then think of what parts of the input to modify in order to provide an output of higher quality and repeat the process until a satisfactory result was achieved or until they felt that the MT tool could not improve upon its output any further. The learners would then do the opposite, use an English text as the source language to translate into Chinese as the target language, this time pre-editing the English text. Shei (2002) points to two reasons behind doing this, first one being to let learners experience how MT can help them in their L2 learning. Secondly, he wanted to let them experiment with English words and structures, replacing words and phrases written by a native English speaker with their own English to evaluate the quality of the Chinese output. Shei (2002) concludes his study by saying that the learners were full of zeal when pre-editing texts and argues that pre-editing is one of the most exciting and effective ways of learning to use MT, learning to translate, and learning languages.

At the time of writing, the intelligence and capabilities of MT tools have increased tremendously compared to when Shei conducted his study in 2002. This does not mean that pre-editing texts cannot be a meaningful learning process for learners who are attempting to learn new languages. Rather, it means that learners should explore the use of pre-editing to discover the new limitations of MT. The author of this thesis believes that teachers should be aware that some prerequisites need to be in order before employing pre-editing of texts into the FL classroom. Firstly, teachers need to consider what texts they want their learners to pre-edit, if the source language should be in their mother tongue and the target text be the FL they are trying to learn or the opposite. The genre and difficulty of the text also needs to be

considered to ensure that all learners can receive an adequate learning outcome from the exercise. Furthermore, teachers need to consider the prior knowledge that the learners have of the FL they are learning before implementing pre-editing. If learners do not master aspects of the FL such as sentence structures or grammar, they will not be able to recognise irregularities and errors between the FL and their mother tongue, thereby not receiving a satisfactory learning outcome.

2.7.3. POST-EDITING

Another method that can be used for pedagogical uses of MT in the FL classroom is post-editing. Niño (2009, p. 243) defines post-editing as the correction of raw MT output into an acceptable text for a particular purpose. Another definition of post-editing is given by Allen (2003, p. 297) who defines it as editing, modifying and/or correcting pre-translated text that has been processed by an MT system from a source language into a target language.

In one of her earlier studies, Niño (2008) argues that the use of post-editing as an activity in the FL classroom is useful for advanced learners as their post-edited output had fewer errors compared to that of unedited MT translations and learners' own translated outputs. She suggests that if teachers wish to employ post-editing tasks in the FL classroom, they should control the input text in such a way that the output that the MT produces is relevant and appropriate for the learners' levels. French (1991, as cited in Niño, 2008, p. 33) believes that the use of MT post-editing is useful for learners as it challenges them to be self-critical towards their written productions and, consequently, it prompts them to towards the creation of cohesive and coherent texts.

In a study conducted by Kliffer (2005), FL learners were asked to post-edit three texts that had been translated from French to English by an MT tool. In his study, learners at all stages, lower, average, and higher showed improved results and fewer errors in their post-edit texts compared to that of a translation program. The learners were also asked to give an evaluation of the exercise that they had participated in where they would answer three questions. One question asked learners if they believed that the exercise was useful via a 5-point scale system. A majority of participants reacted positively to the exercise with one positive comment being that they felt the exercise made them focus on challenging errors such as idioms and figurative language. Some of the participants that were classified as higher performing felt that one negative aspect of the exercise was the plethora of trivial errors

produced by the MT, while some of the lower performing learners felt that editing these errors was helpful in acquiring FL knowledge.

Overall, Kliffer (2005) concludes that the exercise was more useful for lower performing learners who valued post-editing more than translating the text from scratch, but it did show to have a positive effect for learners at all levels. In terms of who benefits the most from this type of MT exercise, this study comes to a different conclusion to that of Niño (2008) who argues that advanced learners benefitted the most from post-editing. Some reasons as to why this is, could be due to the use of different MT tools or the different language pairs that the learners were asked to post-edit. Both studies did reach the same conclusion, however, that FL learners post-editing a text resulted in fewer errors than what an MT tool provided. Garcia (2011, p. 227) reaches a conclusion that is in agreement with both Kliffer and Niño, where post-editing seems to be beneficial to both lower and higher performing learners, although the difference in post-editing vs translating directly from the source text was more noticeable in lower performing learners, especially when they post-edited in their first language.

In an updated study, Kliffer (2008) discusses more positive attributes of using post-editing as a pedagogical method of using MT in FL learning. Kliffer remarks that learners experience post-editing as a confidence-building exercise, especially lower and average performing learners. Furthermore, he observed that post-editing made learners more aware of the importance of a holistic approach of interpreting and translating a source text, rather than translating word-for-word. Belam (2003) adds more arguments as to why post-editing should be included in MT lessons. The amount of post-editing required by learners is a good measure of translation quality by MT, making learners become more aware of the limitations that MT tools have. She further adds that learners have to study the text in detail to post-edit effectively, thereby learning new vocabulary and expressions, new grammatical points and stylistic futures, giving post-editing exercises positive benefits to learners' language learning.

The use of post-editing in the FL classroom enables learners to not only save time by not having to translate a text from scratch, which is a time-consuming process, time that could be spent on refining the learner's language or improve the text's contents, but it also offers opportunities to raise language awareness and make use of language skills by finding errors in the MT output and correcting them. Combining the use of MT tools, the use of post-editing and testing their output in grammar checkers such as Grammarly, enables learner autonomy as they can independently receive some feedback without the need of a teacher. Given the results

of the studies touched upon above, post-editing seems to be an effective and meaningful method to introduce to FL learners when using MT tools as it can positively impact their learning outcome.

2.7.4. USING MACHINE TRANSLATION AS A BAD MODEL

There is a very limited amount of research conducted on this particular method, but Somers (2003) and Niño (2009) discuss the use of machine translation as *a bad model* as one method in which learners can facilitate language learning, both referring to it with said name. Somers (2003, p. 327) characterises MT as *a bad model* as using MT software's weaknesses and mistakes to bring out subtle aspects of language differences or to reinforce learners' appreciation of both L1 and L2 grammar and style. Using MT as *a bad model* is partly incorporated into the other two pedagogical methods of approaching MT in the classroom as the learners either pre-edit the input or post-edit the output in this method. Furthermore, Anderson (1995) provides insight into how the use of this method can be beneficial for learners of an FL. When learners manually enter sentences one by one from a suitable text provided to them, note the results given, and then use native-speaker intuition to identify and correct errors, language learning is facilitated. He explains that for translations from an FL into learners' L1, this can be a useful exercise because the poorer-quality translations are typically too close to the lexical and syntactic structure of the source language, meaning that the exercise can reinforce students' understanding of differences between the languages by showing them bad translations into their own language.

If the use of this method is to be used in the FL classroom, one has to be careful of how to approach it as using this method with translations into the second language carries with it the danger of reinforcing or introducing incorrect language habits to the language learners (Somers, 2003, p. 327). Somers furthers this point by saying that learners tend to have a natural "respect" for printed words, and that there is a tendency for them to believe that the system is an authority on the target language, believing that anything it produces must be correct. Teachers therefore need to disabuse the learners of this mindset. Richmond's (1994, as cited in Somers, 2003, pp. 327-328) solution to disabuse learners of this mindset is by providing model translations where learners are provided with an adequate premade translation in the target language and the source text that they have to translate into the target language using MT. The learners take note of the errors made by the MT software and are then asked to pre-edit the text until similar output is given to that of the premade translation.

His pedagogic reasoning as to why this works is because it causes learners to focus on the differences between the source language and the target language.

As seen from the positive results of earlier studies, using MT as *a bad model* can be an appropriate method to reinforce language learners' learning outcome. However, this type of exercise, and the other two discussed above, require a great deal of planning from the teacher before it can be implemented. Finding lesson plans for these exercises in different languages might not prove helpful as the output and weaknesses of MT tools vary between every language pair. Further, as these exercises require a higher cognitive level than some other activities, they are not recommended for learners who struggle in either L1, L2, or both. The teacher needs to be prepared and have specific texts that he/she knows will have flaws and mistakes in them. Another challenge that these three methods bring with them is that earlier studies and lesson plans created for this kind of exercise may constantly change and be ineffective. Due to the constant evolutions and improvements in machine translation, mistakes and errors made by them today might be corrected soon. That is not to say that teachers should be discouraged from pursuing these types of exercises, as when done correctly, they have the potential to reinforce language learning and provide learners with knowledge of the limitations and weaknesses of MT. This can help in preventing them from becoming overly dependent on these tools and instead see them as supplementary tools.

2.8. REVIEW OF SIMILAR STUDIES AND CONTRIBUTIONS

2.8.1. INTRODUCTION

In this section, an overview of similar studies and contributions will be addressed. Some criteria had to be established so that the empirical data presented, as of the time of writing, is relevant for this study. Google Translate started changing its method of translating between languages from an SMT to an NMT in November 2016. Studies that were conducted prior to this date are therefore not relevant for this study as the data collected in those studies are outdated and can be misleading. As a result, there is a severe lack of relevant studies that look at GT and its use in the FL classroom. In this subsection, the five studies that were conducted after the algorithm changed and that look at GT in a FL classroom are examined. Although these studies had different methods of collecting data than the current study, they still prove useful in providing background context, and they further the implications and applications of how GT can alter teaching practices in the FL classroom.

In their study, Ducar and Schocket (2018, p. 780) aim to advance the theoretical discussion surrounding machine translation in the FL classroom, help teachers understand what MT can and cannot do, and help them equip their learners in using the technology in an educationally and interculturally respectful manner. In their study, they discuss the strengths of Google Translate, some of them being its versatility and pervasiveness. Another aspect they consider is that teachers typically work with learners who are at the beginning- and intermediate-level where the FL is basic enough that GT can translate their input with relative ease. Ducar and Schocket (2018, p. 783) present two arguments for GT's intellectuality. Firstly, they argue that the OT can correct spelling mistakes inputted in the source language. Secondly, the OT is now becoming better at translating idioms, especially high frequency ones due to the Google shield, the human verification process feature that Google implemented into GT, which they comment still cannot be trusted every time it appears in the translation output.

They further discuss the previous limitations and weaknesses of GT that had largely been fixed at the time the study was written and issues that still plagued the OT. Some of the former issues that GT had largely overcome were proper nouns being translated into the target language, literal translations, misspelt words not being translated, and difficulties with less-common idioms (Ducar & Schocket, 2018, p. 784). Issues that were still prevalent at the time the study was written were grammatical inaccuracies, register (formal vs. informal varieties) and cultural expectations, and pragmatic breakdown, e.g. issues related to context, connotation, and denotation. Although Google Translate is becoming increasingly powerful, it still cannot provide adequate translations to some of the finer human language aspects such as cultural norms, dialectal variations, and pragmatics. The researchers also mention that in the near future as MT technology continues to improve, identifying translation errors will become increasingly difficult to notice for language teachers. Instead of teachers noticing errors that MT tools previously made, it will soon become the subtle successes of these tools that signal that they have been used by language learners if plagiarism is of relevance for teachers.

In their discussion section, Ducar and Schocket (2018, pp. 787-793) offer insight into the pedagogical implications that GT brings with it in the foreign language classroom and the importance of teachers instructing learners of how to use translation tools in a responsible way that promotes, rather than circumvents learning. Learners need to feel motivated in learning a foreign language and see its use as a 21st-century skill to persuade them from

becoming overly dependent on the use of OT software. Teachers therefore have a responsibility of helping learners in wanting to communicate and autonomously seek to progress their proficiency in a language rather than seeing it as simply completing an assigned task.

Furthermore, they advocate that teachers should implement translation exercises that detail the pitfalls and drawbacks of OT software to make learners aware of the subtleties that are involved in conveying one's intended message from one language to another. Lastly, they conclude their study by calling out the severe lack of research into the effects MT has on learning despite its prevalent use in the FL classroom, which this research will hopefully address.

2.8.3. THE USE OF GOOGLE TRANSLATE IN EFL ESSAY WRITING

Chandra and Yuyun (2018) conducted a case study on eight Indonesian university students learning English as an FL. The study dealt with how these students used GT in EFL essay writing, more specifically, for which writing purposes learners consulted GT.

The methodology that Chandra and Yuyun (2018, p. 231) employed consisted of a qualitative approach where two methods of collecting data were used, observations through screen recordings and interviews. The students took part in a writing session where they were prompted to write a narrative essay in which their screens were recorded for further analysis. Only one writing session was conducted in the study, stating time restrictions as their reasoning behind only one writing session, and the students were then interviewed by the researchers afterwards to gain further knowledge of their use of GT.

When analysing the screen recordings, the researchers categorised the way students used GT into five categories: Word, phrase, sentence, grammar, and spelling. The results from their study showed that students mainly used GT for three reasons when writing essays: vocabulary, grammar, and spelling, in which vocabulary was divided into three further levels: word-level, phrase-level, and sentence-level (Chandra & Yuyun, 2018, pp. 231-232). Based on their results, students mainly used GT as a dictionary. Phrases were translated second most of total translations, followed by sentences, spelling, and finally the tool was used the least to check grammatical aspects. The researchers provide reasoning as to why the students' use of GT to check for spelling was relatively low. This was mainly due to word-processing programs such as Microsoft Word and Google Docs having the embedded ability to proofread

the text that was written, providing learners with suggestions for proper spelling of words that were written incorrectly.

Based on their results, Chandra and Yuyun (2018, p. 236) concluded that their students used Google Translate as a supportive tool in their language learning, not as a replacement tool which they were overly dependent on. It is unsure if the students in the study had been given prior training in how to use Google Translate, but they stated in their interviews that they found the use of GT helpful in essay writing, especially when using it for word-level translations. However, they did not think that Google Translate would prove useful if longer sentences were translated as they believed that the output would be of poor quality. Mainly, Google Translate was used as an alternative to a dictionary as it provided results in real time unlike a dictionary which took a longer time to provide translations.

2.8.4. ASSESSING THE ACCURACY AND TEACHERS' IMPRESSIONS OF GOOGLE TRANSLATE: A STUDY OF PRIMARY L2 WRITERS IN HONG KONG

Stapleton and Leung (2019) conducted a study on Chinese learners where the learners were given two writing tasks to carry out, one in English without the use of GT, and one in their native language which would be translated into English via GT. Teachers were then given the scripts written and asked to grade them while being unaware that GT had been used, after which interviews were conducted with the teachers regarding their impressions.

The learners who participated in the study were primary 6 learners who were 11 or 12 years old. The study took place in 2018, and learners were asked to write a composition based on the prompt "Is half-day school a good idea?" for 60 minutes in English, which was an FL for them. Several days later, the researchers then asked the learners to write a composition in Chinese, their mother tongue, based on the same prompt as was given to them earlier. The English composition they had written a few days earlier was also given back to them which they could refer to when writing.

After both writing tasks were completed, Stapleton and Leung (2019, p. 20) had collected 22 Chinese compositions and 26 English compositions to analyse. Before the Chinese compositions were translated into English via Google Translate, the researchers deemed it necessary to perform data cleaning on the scripts such that Google Translate would handle the texts better. The Chinese scripts were then translated into English via GT and randomly interspersed with the learners' own English compositions for teachers to grade and

comment on. Twelve teachers were then recruited to grade all the writing compositions and to participate in interviews. The researchers also gave the teachers a rubric to grade the scripts based on three criteria: grammar, vocabulary, and comprehensibility, grading each individual criterion on a scale from A-D, with A being the highest. The teachers were also instructed to ignore content and organisation in the learner written texts as those elements are not within MT's capabilities.

Results from the study showed that teachers graded scripts translated by Google Translate higher than what learners themselves could produce directly in English (Stapleton & Leung, 2019, p. 23). Based on analysis, the GT scripts were considered significantly better than non-GT scripts in grammar, and the GT scripts were considered better in vocabulary to some extent. The analysis from t-test results showed that in comprehensibility, non-GT scripts scored insignificantly higher, meaning that both scripts seemed to be equally comprehensible to teachers.

Summarised, a majority of teachers in the study held the belief that GT could be used as a tool, but they had concerns that GT might negatively affect learners' language learning. A suggestion was given by the teachers that training on the correct use of GT should be provided in schools so that learners could benefit from it.

Stapleton and Leung's (2019, p. 29) conclusion is that GT did at times generate more formal and sophisticated language than the learners managed to produce own their own due to the large corpora of texts that GT has been trained on. The researchers further discuss the interesting disruptive changes that are happening to language learning and teaching due to the rapidly advancing artificial intelligence, with MT being one example. Presently, pedagogy lags behind the improvements that are happening to technology as teachers struggle to learn how to best use and manage the new tools that appear. With the constant improvements in MT, the researchers argue that teachers of foreign languages need to quickly develop a broader realisation abut technologies that are likely to have a significant impact on teaching and learning in L2 contexts. As it is highly unlikely that learners will not take advantage of the ICT tools that exist, teachers and schools need to rethink L2 pedagogy to accommodate and adapt for technology as pedagogical tools.

Lee (2020) conducted a study on Korean learners majoring in English at university. The objective of her study was to study the effects MT has on learners when they translate their L1 writing into L2 without the aid of MT, and then correct their L2 writing with the use of an MT for comparison. As little research has been conducted on the use of MT as a learning tool in FL learning, Lee (2020, p. 158) states that it is important to investigate how MT can facilitate language learning based on empirical data and acknowledge potential benefits, as well as drawbacks on the use of MT in education.

In Lee's (2020, p. 161) study, 34 university students participated, all majoring in English at a Korean university where data was collected over the course of six weeks. The students were given a task to write about the texting language of today's young people. Five steps had to be followed when performing the task, first watching a TED video related to the topic, writing a one-page paper about the topic mentioned in their L1, translating their scripts into English without the aid of MT, translating their source language script into English solely using MT, and finally editing their initial English translation by comparing it with the MT version that was generated. In the data analysis, Lee first analysed the students' first script, the one written in their L1 to check for overall quality. Content and meaning were especially paid attention to as they were supposed to remain the same in both the students' scripts and the MT scripts.

Overall, most students generated acceptable writing in their L1. In their initial L2 scripts, a mean score of 3,76 was given based on a 6-point scale. This mean score increased significantly in the final version of the scripts to 4,59. Furthermore, the average number of errors produced in the students' scripts decreased significantly between the students' initial scripts, with the mean number of errors being 21,94, and their final scripts, with a mean number of errors totalling 13,64. In her data analysis, Lee (2020, p. 164) also concludes that students did not simply adopt the MT translation when they were revising their scripts.

Lee's (2020, pp. 171-172) conclusion was that there were indications that MT improved vocabulary, grammar, and expressions, which resulted in writing quality improvements. The use of MT also positively influenced student writing strategies during revision, and she remarked that MT could be an effective supplementary tool to language learning under appropriate conditions. However, for MT to be beneficial for student learning, teachers must understand the role of MT in language learning and provide students with

adequate guidelines on its use, detailing both its strengths and weaknesses to students, along with methods of using it as a language resource.

2.8.6. CHINESE STUDENTS' PERCEPTIONS OF USING GOOGLE TRANSLATE AS A TRANSLINGUAL CALL TOOL IN EFL WRITING

Tsai (2020) conducted a study on EFL students at a Taiwanese university where he aimed to investigate the effectiveness of using Google Translate as a translingual CALL tool in EFL writing. One of the main objectives of the study was to investigate how non-English major students, who make up a majority of EFL students, can use GT as a translingual CALL tool to improve writing performance.

The methodology used by Tsai (2020, p. 6) is similar to how Lee (2020) conducted her study. Students were first asked to write an extemporaneous reflective essay based on a 5-minute passage from a movie in their mother tongue (Chinese) for 30 minutes, then compose a similar text in English for 30 minutes, then submit their Chinese essay to GT, and finally revise their self-written English text by referring to the text outputted by GT. To aid in assessment, two online assessment freeware programs were used, Vocabprofiler and 1Checker. Vocabprofiler was used to determine content improvement and vocabulary in the students' texts by comparing their self-written English text, the text produced by GT, and their revised English text. 1Checker was used to determine surface-level mistakes such as spelling and grammar.

Results from Tsai's (2020, p. 8) study showed that for both English major students and non-English major students, the GT texts outperformed their first self-written texts in terms of vocabulary. The GT texts also contained fewer mistakes and achieved a writing style that was more advanced compared to their first self-written texts. In their revised self-written texts, both English majors and non-English majors performed better than their first texts, however, the improvement rates were greater for non-English majors, most likely due to English majors being more knowledgeable of the English language. Interestingly, non-English majors expressed significantly less information and fewer ideas than English majors in the first writing session when they were asked to write directly in English. However, they were able to deliver thoughts and information equivalent to the level of English major students in the revised texts after they consulted the GT texts.

Tsai (2020, pp. 19-20) concludes his study by expressing that results seem to indicate that GT can help EFL students enhance their English writing performance if done correctly, creating opportunities for more enriched content, more advanced vocabulary and fewer spelling and grammatical errors, especially for those with a lower English proficiency. He further argues that although the use of GT makes translating quicker, more convenient, and possibly more accurate for L2 learners, they still have to improve their own English proficiency to be able to comprehend what is being translated and make revisions as needed. EFL teachers also need a thorough understanding of Google Translate and other CALL tools to develop language learning models that fit the digital era.

2.8.7. CONNECTING THE LITERATURE REVIEW STUDIES TOGETHER

Ducar and Schocket's (2018) discussion on the importance of teachers instructing learners how to use translation tools in a responsible way that promotes learning relates to the discussion from Stapleton and Leung's (2019) study. They conclude that teachers and schools need to rethink L2 pedagogy to accommodate and adapt to technology as pedagogical tools seeing as it is highly unlikely that learners will not take advantage of the ICT tools available at their disposal. Both these studies along with the studies from Lee (2020) and Tsai (2020) conclude the importance of teachers adapting to the technology that surrounds them and incorporating it into the FL classroom in meaningful ways that promote learning. Furthermore, results from Stapleton and Leung's (2019) study of how GT is able to produce texts of higher quality than learners can create themselves also agree with the results from both Tsai (2020) and Lee (2020), both studies reaching the conclusion that GT can aid learners in creating texts of a higher quality, that is to say fewer errors, improved vocabulary, and expressions compared to learners writing directly in their L2. Lastly, Chandra and Yuyun (2018) conclude their study by stating that based on their results, learners used GT as a supportive tool in their language learning, typically translating at word-level. These findings are in line with Lee's (2020) remark that MT could be an effective supplementary tool to language learning under appropriate conditions.

3. METHODOLOGY

3.1. INTRODUCTION

In the following chapter, the methodology and materials used to collect data are explained in detail. To answer the research questions that this thesis presents, an explanatory sequential mixed method design was used. This mixed method study was comprised of a quantitative study using a quasi-experimental design and a qualitative study using a case study design.

Firstly, descriptions of qualitative, quantitative, and mixed method approaches are given followed by the design process of the current study with an explanation of what makes this thesis a mixed method study. Detailed explanations of the methodologies used to collect data, i.e. screen recordings, learner stimulated recall interviews, teacher interviews, and the study participants are then presented. Afterwards, the pilot study that was conducted prior to the actual research and how it affected the methodology used to collect data is discussed. Lastly, this chapter details the validity and reliability of the thesis, and all ethical considerations that had to be in order before data collection could be conducted.

3.2. QUALITATIVE, QUANTITATIVE AND MIXED METHOD RESEARCH

Holliday (2015) describes that the aim of qualitative research is to search for the richest possible data. In this approach, the ideas and presence of the researcher can be influential in what the data looks like and the way in which it is interpreted. As this is the case, the outcome of the research will always be influenced by the researcher's beliefs. Holliday (2015) further comments that the basic aim of qualitative research is to get to the bottom of what is going on in all aspects of social behaviour. Zoltan Dörnyei (2007, p. 24) details qualitative research as data collection procedures that result primarily in open-ended, non-numerical data that is analysed primarily with the use of non-statistical methods. Interviews are one such example of data collection that fall within qualitative research as, especially for this thesis, the answers given by learners and teachers are not numerical data, nor are they analysed by statistical methods. Another characteristic of qualitative research that distinguishes itself from quantitative research is that it typically uses a small sample size (Dörnyei, 2007, p. 38). Although qualitative research has its advantages in terms of collecting data, it is not without flaws. Firstly, qualitative research is often criticised for its small sample size of participants. Examining cases may be very helpful in providing insights into a phenomenon, but the

conditions or insights that are examined may not apply broadly to others (Dörnyei, 2015, p. 41). Another weakness related to qualitative research is that it can be quite time consuming and labour-intensive to process, which explains why qualitative research often uses small sample sizes (Dörnyei, 2015, p. 42).

In contrast to qualitative research, quantitative research emphasises control of variables to the extent that the researcher's influence is minimised (Holliday, 2015). Quantitative research typically involves numbers, quantification, and statistics to address a research problem, with a large sample size usually involved (Phakiti, 2015). There are three main characteristics associated with quantitative research. The first one is that quantitative research is centred around the use of numbers. The second one is that that it looks at variables rather than cases, meaning quantitative research is more interested in common features of groups of people than individuals. Lastly, the third one is that it focuses on standardised procedures to assess objective reality, eliminating individual-based subjectivity from the various phases of the research process (Dörnyei, 2015, pp. 32-33). One downside of using quantitative research to collect data is that its exploratory capacity is rather limited, meaning that quantitative methods typically delve deeper to cover the reasons for particular observations or dynamics underlying examined situations (Dörnyei, 2015, p. 35).

To combat the weaknesses of the two aforementioned methods, both were used to supplement each other and strengthen the data collected, which is also referred to as a mixed method approach of collecting data. Mixed method research can be defined as a combination of qualitative and quantitative methods within a single research project (Dörnyei, 2015, p. 44). Further, Dörnyei (2015, p. 45) states that qualitative and quantitative principles can be combined in the data analysis stage by either *quantifying* or *qualitising* the data that has been collected. Riazi (2016, p. 189) discusses similar traits to mixed method research, stating that it is possible to mix methods from the two research approaches, quantitative and qualitative, at different stages in the process of research and thereby providing a better understanding of the research problems. Dörnyei (2015, pp. 45-46) also provides some strengths and weaknesses related to using mixed method research in studies. The main positive attribute to this type of method as already mentioned, is that it increases the strengths while eliminating the weaknesses of both quantitative and qualitative methods. Mixed method research also provides the potential to produce evidence for the validity of research outcomes through corroboration and convergence of findings. This can be done by triangulating the results that are collected to further assist the researcher in confirming or denying whether one type of data

collected is useful by comparing it with another one. One potential downfall of using mixed method research to collect data is whether the researcher that is conducting the research is adequately trained in handling both qualitative and quantitative data. This is especially something to consider for this thesis as the researcher needs to be careful in determining which data proves useful to further an argument and which data might hinder the potential of reaching adequate conclusions.

3.3. THE DESIGN PROCESS OF THE STUDY

In this section, the methods used in the study to answer the research questions are detailed. The two research questions are:

- How is Google Translate used by Norwegian primary school learners when they write in English?
- How does the use of Google Translate affect the quality of Norwegian learners' texts when they are writing in English?

This thesis employed the use of a mixed method research design, specifically a sequential explanatory method design. A sequential explanatory method design is characterised by the collection and analysis of quantitative data followed by the collection and analysis of qualitative data (Creswell, 2003, p. 245). This thesis employed the use of qualitative methods, in the form of learner stimulated recall interviews and teacher interviews, and one method that was used for both qualitative and quantitative purposes, this being observation through screen recordings.

Presented in Figure 4 is the study procedure used to collect and analyse data, totalling six steps. In step 1, three learners from each class that had given their consent to participate in all aspects of the study (see Appendix 2A) were randomly selected by the lead teacher of each corresponding class, for a total of nine learners, to participate in stimulated recall interviews (SRI). These chosen learners were also observed in real-time by the author of this thesis during both writing sessions such that the SRIs would contain questions that could elicit responses from learners of actions they had done and could remember doing during the writing sessions.
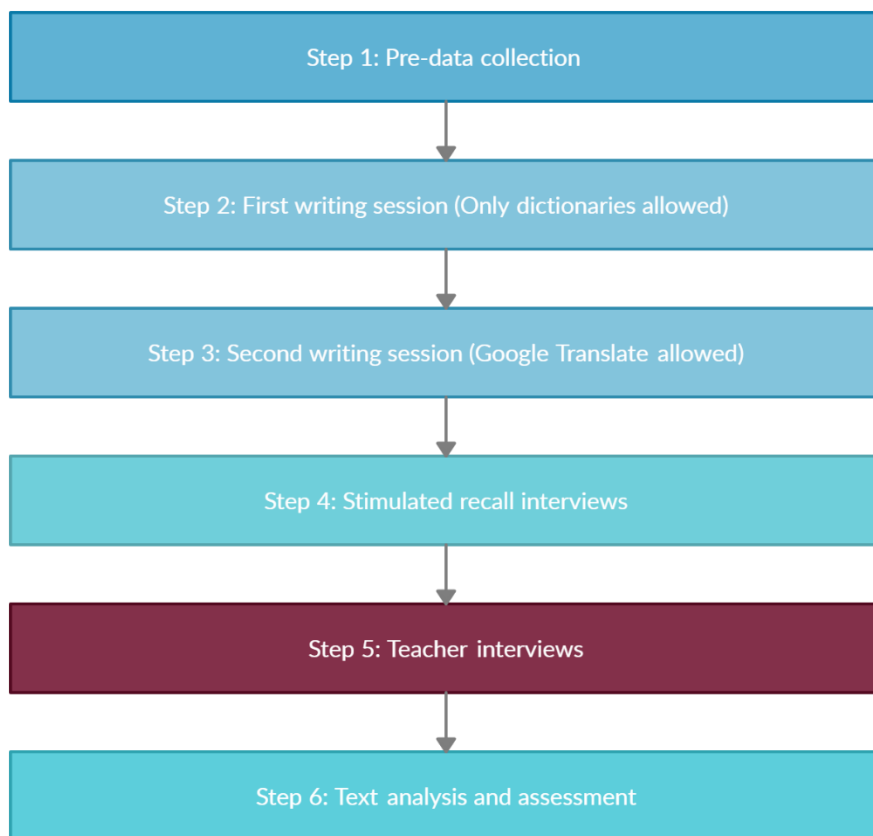
Figure 4: Study procedure

In step 2 of the study procedure, learners were tasked to write about the topic *My favourite place to be* for 30 minutes in Google Docs. In this writing session learners were not allowed to make use of Google Translate or other translation tools that had the capability of translating at phrase-level or above, only tools that would translate at word-level were permitted. Following the conclusion of the first writing session, 36 texts and 33 screen recorded videos were received and downloaded onto an encrypted hard drive for future analysis.

In step 3, the second writing session, learners were given the topic *My second favourite place to be* to write about for 30 minutes in Google Docs. Contrary to the previous writing session, in this writing session learners were encouraged to make use of Google Translate in the same manner as they would have normally. After the second writing session had finished, 33 texts and 31 screen recorded videos were downloaded onto an encrypted hard drive for later analysis.

Step 4 took place following the conclusion of step 3 where the randomly chosen learners from each class partook in audio-recorded SRIs. Based on real-life observations of these learners in the first and second writing sessions, questions were asked to elicit

information from them of their actions and thought processes behind those actions to provide supplementary data to the screen recorded videos (see Appendix 3B or Appendix 3D).

In step 5, audio-recorded teacher interviews were conducted to elicit information from two of the lead teachers of the participating classes on their perceived use of Google Translate (see Appendix 3A or Appendix 3C). They were also used as supplementary data to confirm or reject statements made by learners during their SRIs and provide context to help answer why some learners performed certain actions during the writing sessions.

Lastly, in step 6, all data collected in the previous steps was analysed and assessed to determine answers to the two thesis questions. There were a total six analyses conducted to provide answers to the thesis questions, and the participating learners had to pass various requirements in order to be included in the different analyses (see section 3.4.5.).

To address the first research question, an explanatory case study was used. A case study is defined by Riazi (2016, pp. 25-26) as a study where the researcher concentrates on a single case. In other words, a case study focuses on one specific subject and examines it in great detail. Casanave (2015) describes case studies as an approach in which the researcher's interest is in an in-depth investigation of the particular rather than the general.

In this study, extensive analysis was conducted on learners' GT usage when writing using stimulated recall interviews and screen recorded videos from the second writing session to provide empirical data for the specific subject investigated. With the use of screen recorded videos, the author of this thesis was able to acquire an extensive look into how each individual learner, with a total of 22 learners, used GT to assist them in their essay writing. This included if they used GT to assist them at word-level, phrase-level, sentence-level, or other uses when translating from their L1 into English.

To address the second research question, a quasi-experimental study with a within-group design was used to collect relevant data. Experimental studies mainly employ quantitative research where the researcher attempts to find cause-and-effect relationships between variables (Riazi, 2016, p. 112). There are two types of variables, independent variables and dependent variables. Independent variables can be thought of as the cause to something changing or the variable believed to influence other variables, while dependent variables are considered the effects that are caused by independent variables (Phakiti & Paltridge, 2015).

In the study, the independent variable was Google Translate, whilst the dependent variables were the different aspects of writing that were analysed which in turn contributed to

an impression of either a positive, neutral, or negative effect GT has on writing. For this study, the written texts that learners produced acted as quantitative data for comparing vocabulary, syntax errors, subject-verbal concord, essay length, and spelling mistakes between learners' texts to determine the effects that Google Translate has on learners' written products. The screen recorded videos acted as qualitative data as they were used to determine whether learners made use of translation tools in the first writing session or if other translation tools than GT were used in the second writing session.

The purpose of conducting the study in three different classes was to provide more data to analyse to address possible issues related to reliability and validity. In total, only two school days were spent collecting data as it was possible to conduct the study in all three classes during the same day as well as interviewing the teachers in quick succession.

3.4. SCREEN RECORDINGS OF WRITING

3.4.1. SCREEN RECORDINGS: AIMS AND RESEARCH COMPONENTS

Observation is a widely used method for data collection, and it is typically associated with studies using a qualitative research design (Curdt-Christiansen, 2020, p. 336). Observation offers the researcher the opportunity to gather first-hand, "live" data from naturally occurring social situations, and it has the potential to provide more authentic data than other methods (Cohen et al., 2018, p. 542). Furthermore, Robson (2002, as cited in Cohen et al., 2018, p. 542) remarks that what people do may differ from what they say they do, and observation provides a reality check. The learners that participated in this study may not have been able to recall how they use Google Translate in FL lessons, or they might not have felt comfortable providing the truth of how they use it if asked about it. Hence, observation, specifically observation through screen recording was used as a data source as it provided the researcher with data that occurred naturally in situations and it avoided issues with selective or faulty memory due to a time gap between the act of observation and the recording of the event (Cohen et al., 2018, p. 542).

Although observation is mainly associated with studies using a qualitative research design, for this thesis observation was used for both qualitative and quantitative purposes through the screen recordings. It was used for qualitative purposes when determining how learners used Google Translate, both live observation of the learners who were participating

in the SRIs and screen recorded videos of the other learners, while it was used for quantitative purposes to help analyse and compare the learners' texts.

3.4.2. SCREEN RECORDINGS SAMPLE

The current study was conducted in three different year 7 classes situated at the same school, and in total, 37 learners gave their consent to participate in the study (see Table 1).

Table 1: Number of learner participants for each writing session in each class and number of documents and screen recordings received

Class	Learners in total	Learners who consented to participate in the study	Documents received from the first writing session	Documents received from the second writing session	Screen recordings received from the first writing session	Screen recordings received from the second writing session
Class A	22	9	9	8	7	7
Class B	19	15	14	14	13	13
Class C	23	13	13	11	13	11
Total	64	37	36	33	33	31
Notes from each class						
Class A	One participant was absent from the second writing session. One participant failed to record their screen in the first writing session. One participant did not consent to their screen being recorded but did approve for their text to be analysed.					
Class B	One participant did not consent to their screen being recorded but did approve for their text to be analysed. One participant was absent from both writing sessions.					
Class C	Two participants were absent from the second writing session.					

For this study, individuals were not randomly assigned to participate as there already existed a non-random context that the participants were situated in, i.e. intact classes. Instead, this type of experimental study, where participants are not chosen randomly and not all variables could be completely controlled by the researcher, is called quasi-experimental research (Gass, 2015). Rogers and Révész (2020, p. 134) state that quasi-experimental studies do not require a control group, something which was not used in this study. Instead, a comparison group was used, which in this study was the same group of learners. This is known as within-group design, and it is used when data is collected from the same group of participants twice to check whether the independent variable, i.e. GT, had any effect on learners' writing performance (Riazi, 2016, p. 272).

To gain a better understanding of the extent GT affects learners' writing, the teachers provided information regarding the learners' language background as well as their levels in English. This information was only received regarding learners who had agreed to share their language background. By receiving the learners' language backgrounds and seeing whether they spoke English at home, it could provide context as to why they performed as they did. However, out of all the participants who consented to sharing their language background, none of them were native speakers of English, but some did have a native language other than Norwegian.

The teachers provided information about each individual learner's level in English so that they could be categorised into three groups: lower performing (LP), average performing (AP), and higher performing (HP). This was done to determine the impact that GT has on the different levels, i.e. whether LP learners gain more from using GT compared to HP learners. The learners' levels were based on previous assignments that they had handed in and the overall competence level that teachers believed they mastered. In terms of performance levels in the English subject, in class A, there were three participants categorised as LP, four participants categorised as AP, and two participants categorised as HP. In class B, four participants were categorised as LP, five participants were categorised as AP, and six participants were categorised as HP. Finally, in class C, there were two participants categorised as LP, six participants categorised as AP, and five participants categorised as HP.

3.4.3. SCREEN RECORDINGS INSTRUMENT

The free screen recording extension, *Screen Recorder* by Erich Behrens, available in the Google Chrome web store was used to record learners' screens during both writing sessions. Neither microphone audio nor screen audio were recorded as they were irrelevant for this study. Furthermore, the learners themselves were never recorded in any form during the writing sessions.

3.4.4. SCREEN RECORDINGS PROCEDURE

The study consisted of two separate writing sessions consisting of 30 minutes each where learners wrote their essays in Google Docs (see Figure 4). In the first writing session, learners were not permitted to use Google Translate nor any other type of translation software that could translate segments of text. Instead, they had to use dictionaries, electronic or physical if

they needed help to define or translate words. For the second writing session, learners were encouraged to use Google Translate in a similar manner to how they typically would use it in any other writing situation.

For each writing session, one 60-minute session was used. The first 20-30 minutes of each lesson was allocated to instructing learners on how to download the accompanying screen recording software that would be used and on the details of what was going to happen during each writing session. This was done using two presentation slides, one for each writing session that had slightly differing information. For the data collected to be reliable, the topics that learners were going to write about had to be similar, in order to have a fair comparison of the product they wrote. For the first writing session, learners were asked to write about the topic *My favourite place to be,* while in the second writing session they were asked to write about the topic *My second most favourite place to be.* The second half of each 60-minute lesson was allocated to the actual writing session. After 30 minutes had passed in each writing session, that is to say after the lessons were finished, the learners were asked to leave the classroom. The author of this thesis together with the teacher of each class then proceeded to download the texts and accompanying screen recorded videos onto an encrypted hard drive to be analysed at a later point.

3.4.5. SCREEN RECORDING DATA ANALYSIS AND TECHNIQUES

From the data analysed from the screen recorded videos and the texts that learners handed in after each writing session, they could be included in up to six analyses depending on what they handed in and whether they were present in the different writing sessions:

1. An analysis relating to learners' use of dictionaries if they were present in the first writing session and handed in a screen recorded video.
2. An analysis of the quality of the learners' texts from the first writing session if they were present in both writing sessions, handed in screen recorded videos from both writing sessions, and made use of Google Translate during the second writing session.
3. An analysis of how learners used Google Translate to help write in English if they handed in a screen recorded video in the second writing session and made use of Google Translate in some manner.
4. An analysis of the quality of the learners' texts from the second writing session if they were present in both writing sessions, handed in screen recorded videos from both writing sessions, and made use of Google Translate during the second writing session.

5. An analysis comparing the quality of the texts handed in by learners from the first writing session to the second writing session.

6. An analysis looking at how learners used two of Google Docs' features: grammar checker and spell checker if they were present during any of the two sessions and handed in a screen recorded video along with their text.

Based on these criteria, 33 learners were included in the first analysis, 22 learners were part of the second, fourth and fifth analyses, 22 learners were included in the third analysis, and 34 learners were included in the sixth analysis. As learners made use of Google Docs in both writing sessions, there were a total of 64 screen recorded videos included as part of the sixth analysis. One learner was present in both writing sessions and did make use of Google Translate in the second writing session, however, he failed to record his screen during the first writing session and was therefore omitted from the second, fourth, and fifth analyses.

From the pilot study, it was noted that a few learners managed to not save their screen recordings. There was a probability that it might occur in the proper study as well, and if it were the case, learners would still be asked to hand in their documents. However, due to not being able to confirm whether GT was used or not during the first writing lesson, only learners who were present during both writing sessions and that handed in accompanying screen recordings with both their texts were included in the analysis relating to the second thesis question. In total, five texts were submitted that did not have a corresponding screen recording, three coming from the first writing session and two coming from the second writing session. Two learners accounted for two of the screen recordings in both writing sessions.

To speed up the analysis of over 40 texts written by participants, the ICT tool Duolingo CEFR checker was used to compare differences in vocabulary between the first and second writing session. As the author of this thesis had to analyse more than 30 hours' worth of screen recorded videos and over 40 texts by himself, it was deemed appropriate to make use of applicable ICT tools wherever possible to alleviate some of the workload.

The CEFR checker by Duolingo is a tool where users input text and the tool will analyse the text to determine which CEFR level the text is suitable for, ranging from A1 being the simplest, to C being the most complex. There are a total of six CEFR levels that can be broadly categorised into three levels: Basic user (A1 and A2), independent user (B1 and B2), and proficient user (C1 and C2), however, the Duolingo tool does not differentiate between the two C-levels (Council of Europe, 2018, p. 34). In the Duolingo tool, the text is broken

down into word-level, and each word is assigned a CEFR level which can be identified by colour, as well as a generated list where each word is shown its equivalent CEFR level (see Figure 5). The CEFR checker estimates the CEFR level at which a learner might comprehend each word in a text and assigns an appropriate CEFR level to each word. The tool has estimates for hundreds of thousands of words in English (McDowell & Settles, 2019). It achieves estimates of CEFR levels through the use of machine learning and training the system on thousands of hand-annotated CEFR labelled words (McDowell & Settles, 2019). Additionally, the CEFR checker labels words that it has not been trained on with a ?-label. This ?-label has no effect on the overall predicted CEFR level of the text, but it does affect the overall percentage of predicted words for each label which is something that has to be adjusted for when comparing vocabulary between learners' texts.
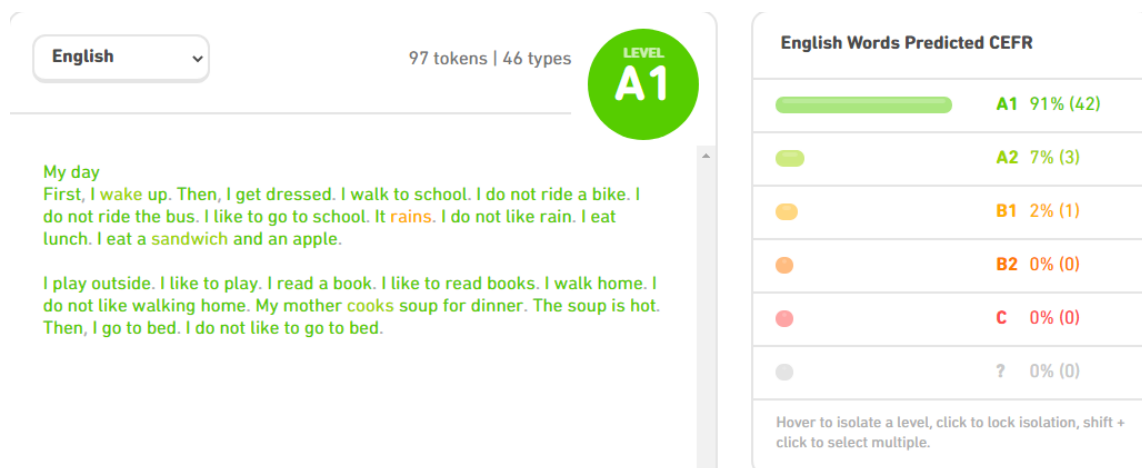
Figure 5: The estimated CEFR level of the text inputted in this instance is A1

Rigorous testing was carried out by the author of this thesis to determine the accuracy of the CEFR checker before it was decided to be used as a tool to compare the vocabulary between learners' texts. 5000 words CEFR labelled from A1 to C1 by the Oxford Learner's Dictionary[8] were inputted into the CEFR checker to examine whether the words would be labelled the same by the ICT tool. Although the two differed slightly between their labelling, it was usually only one CEFR level separating them, and the majority of words were labelled identically. Additionally, texts and passages CEFR-labelled by Lingua[9] were inputted to the CEFR checker to further determine its accuracy. Most texts and passages inputted were given the same CEFR level by the ICT tool as Lingua had labelled them while a few texts had a discrepancy of one CEFR level between the CEFR checker and the texts labelled by Lingua.

---

[8] https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000
[9] https://lingua.com/english/reading/

Moore (2020, p.10) points out that the CEFR contains descriptors of vocabulary development, which describes the progress in terms of control of various types of vocabulary as learners move from A1 to C2. However, specific vocabulary lists are not found in the CEFR, only generic terms referring to learners having a "sufficient" or "good" range of vocabulary enabling them to achieve various communicative functions. The Oxford Learner's Dictionary word list used to determine the accuracy of the Duolingo CEFR checker is therefore not absolute as it is merely intended as a guide, but it can still prove as a principled basis for learners' vocabulary CEFR level.

In addition, it is important to differentiate between receptive and productive vocabulary related to the CEFR levels. Some word lists base their CEFR labelled words in regard to the level at which learners are likely to recognise a word receptively, while other CEFR labelled word lists relate theirs to when learners are typically starting to use the word productively (Moore, 2020, p. 11). Learners' productive vocabulary size is thought to be smaller than the learner's receptive vocabulary size (Milton & Alexiou, 2020, p. 21). The CEFR checker by Duolingo bases its machine learning CEFR labelled word list on learners' receptive vocabulary, i.e. reading or listening level. Although this study focuses on comparing learners' productive vocabulary, the Duolingo CEFR checker can still prove useful. Considering that the CEFR checker was used to compare two texts composed by the same writer, the argument for its use can be made as the baseline was equal for every participant as they were only measured against themselves. However, the author of this thesis did not trust the tool blindly. Comparisons of learners' texts were also done manually to reduce the probability of the overall CEFR levels being completely incorrect. For instance, if a word was labelled C by the CEFR checker but the author of the thesis believed the word did not fit the given label, a manual review process was conducted by looking at what the Oxford Learner's Dictionary labelled the word in question.

Support for measuring vocabulary against the CEFR is given by Milton and Alexiou (2009). They express that as learners become better and more skilled in an FL, they tend to know more words. The vocabulary size scores that emerged from their study suggest that certain levels for vocabulary knowledge are associated with performance at each CEFR level (Milton & Alexiou, 2009, p. 208). Furthermore, Milton (2009, p. 191) concludes with empirical data that vocabulary size can be connected to the CEFR in a plausible way. With formal examinations, such as the Cambridge examinations being placed within the CEFR

framework, Milton (2009, p. 172) raises the implication of a direct link between vocabulary knowledge and CEFR levels.

Mean CEFR-levels for the different groupings were calculated by classifying each CEFR-level to a number, A1 equalling 1, A2 equalling 2, and so forth up to C equalling 5. The higher the number, the greater the CEFR-level. A mean CEFR-level was then given by taking the mean score received from a group of learners and rounding to the nearest whole number, e.g. a score of 2,7 would be rounded to 3 which correlates to CEFR-level B1. The author of this thesis argues that this approach of creating mean CEFR-levels is valid due to estimations made by Cambridge Assessment which show that each CEFR-level from A1 to C1 is proportionate to the previous one (Cambridge English Language Assessment, 2015).

The analysis process of the data obtained from screen recordings and texts, both in the first writing session and the second writing session, was conducted in six steps. Firstly, in analyses 2, 4, and 5, each learner that participated in both writing sessions, handed in a screen recording and text in both writing sessions, and made use of Google Translate during the second writing session had their texts and screen recorded videos analysed. Data from the two writing sessions relating to vocabulary, syntax errors, subject-verbal concord errors, spelling mistakes, and essay length were calculated and inputted into two separate tables (see Table 2 and Table 6). The data obtained from the two writing sessions was then compared against each other to determine whether the three learner groups' usage of GT and the learners' overall mean usage of GT had any statistical significance in the different categories using paired t-testing (see Table 7). 22 learners were included in the three analyses mentioned above. In analysis 1, learners' screen recorded videos were reviewed to determine how they used dictionaries to help in EFL writing, and in this analysis only learners that handed in a screen recorded video along with their text in the first writing session were included. In this analysis, 21 learners were included as they made use of a dictionary during the first writing session. In analysis 3, learners' screen recorded videos were examined to determine how they used Google Translate to help in EFL writing. 22 learners were included in this analysis as they handed in an accompanying screen recording with their text in the second writing session and made use of Google Translate in some manner. Lastly, analysis 6 was conducted by reviewing learners' screen recorded videos from both writing sessions to determine how Google Docs' features helped them correct mistakes as well as determining to which extent learners at the three different performance groups were dependent on Google Docs' features to help them create a text of a higher quality, meaning a text containing fewer errors. In total,

34 learners were included as part of this analysis, and as screen recorded videos from both writing sessions were analysed, a total of 64 videos were examined.

In order to ensure that learners' texts were analysed and compared as objectively as possible, a set of guidelines were created to aid when performing analytics (see Appendix 4). The guidelines were divided into five sections, one for each category that the author of this thesis set out to compare between a learner's two texts. Essay length was compared by inputting the two texts that learners wrote into a word processing program and comparing the overall number of words. Vocabulary was compared by performing data cleansing on misspelt words, then inputting the texts into the CEFR checker by Duolingo and comparing the overall CEFR level of each text as well as the predicted CEFR levels of all words. Syntax errors, subject-verbal concord errors, and spelling mistakes were analysed similarly. A scoring system consisting of 0 points and 1 point was created, where each sentence with one or more errors in syntax or subject-verbal concord would result in 0 points, whilst awarding 1 point for sentences with no mistakes related to those two categories. To create fair comparisons for the last two types of errors mentioned due to learners writing longer texts in one of the writing sessions, the number of correct sentences was divided by total number of sentences to create *Minimum of x per sentence*, x representing either syntax errors or subject-verbal concord errors. For spelling mistakes, the number of words misspelt was divided by the total number of words in the text to achieve a fair comparison.

In order to determine whether there were any statistically significant changes that occurred between the first and second writing session in the five different categories, two-tailed paired t-tests were used. According to Dörnyei (2007, p. 215), paired t-tests are used when the same group of participants are measured more than once, in this case the learners' results from the two writing sessions were compared against each other. Further, their usage is to determine whether the difference between two sets of means is big enough to reach statistical significance.

Prior to using paired t-testing, the mean scores from learners' performance groups in the various categories were checked to see if they followed a normal distribution. This was due to the small sample size used in this study as the normality test ensures more reliable and valid results from the t-tests (Riazi, 2016). The Kolmogorov-Smirnov Test of Normality was used to check whether the means calculated from each learner group and the overall mean of all learners followed a normal distribution. The HP learners' data could not be used to determine a normal distribution as there were too few data points, however, their data was

included to determine whether there was a normal distribution in the overall mean categories. Their data was still used in t-tests to determine statistical significance, but these results are not as reliable as the other groups. After running the various groups' data and the overall means through Kolmogorov-Smirnov Test of Normality, a majority of data was determined to have a normal distribution. The data sets determined by the normality test to not possess a normal distribution consisted of: overall subject-verbal concord errors in both writing sessions, subject-verbal concord errors for AP learners in both writing sessions, subject-verbal concord errors for LP learners in the second writing session, overall CEFR-levels in both writing sessions, and CEFR-levels for AP learners in the first writing session. T-tests were nonetheless performed on these data sets, however, discussions related to the results of the t-tests take into consideration that the results from the above-mentioned data sets are not as reliable nor valid as the other data sets.

As noted, two-tailed paired t-tests were used to determine statistical significance. The reason for doing two-tailed t-tests rather than one-tailed t-tests was to determine whether there were any statistically significant differences between the two writing sessions, both positive and negative. In the t-tests, the α-levels 0,05 and 0,01 were used to determine whether results were statistically significant, and then whether statistically significant results were even more statistically significant.

3.5. STIMULATED RECALL INTERVIEW

3.5.1. STIMULATED RECALL INTERVIEW AIMS AND RESEARCH COMPONENTS

Stimulated recall is defined by Sanchez and Grimshaw (2020, p. 312) as an introspective research method that has mostly been used to serve one of two purposes: Either to recall the elicitation of interactive thought processes or to recall events to facilitate a discussion of the factors used to influence them. In stimulated recall, the use of video or audio recordings are used to help participants recall certain events. Interviews typically follow one of three formats: Unstructured interviews, semi-structured interviews and structured interviews (Riazi, 2016, p. 291). In this study, stimulated recall interviews were used to elicit learners' thoughts and thought processes of how they were using dictionaries and Google Translate to aid them in their writing.

3.5.2. STIMULATED RECALL INTERVIEW SAMPLE

Prior to the first writing session, three learners from each class were randomly chosen to participate in stimulated recall interviews after they had completed the second writing session. These learners were placed together, and away from the rest of the class so that the author of this thesis could observe how these learners without distracting the other learners by moving back and forth in the classroom. The randomly chosen participants were asked questions based on how they used dictionaries or Google Translate to assist them in writing. The plan was initially to conduct two stimulated recall interviews for each class, i.e. one for each writing session. However, whilst some of the SRI participants did use a dictionary to help them in the first writing session, the author of this thesis did not witness any events from any of the learners that called for a stimulated recall interview right after the writing sessions were completed.

3.5.3. STIMULATED RECALL INTERVIEW INSTRUMENT

In this study, the stimulated recall interviews were semi-structured, meaning that some pre-prepared questions were used for probing and follow-up questions where participants could provide more detailed information about aspects of a particular response (Riazi, 2016, p. 162). Some of the questions that were asked to all the participants and were not based on events that transpired during the writing sessions were (See also Appendix 3B or Appendix 3D):

- What do you usually use Google Translate for?
- Is it easier for you to write when you can use Google Translate?
- Why do you use Google Translate?

3.5.4. STIMULATED RECALL INTERVIEW PROCEDURE

For this study, the stimulated recall interviews took place immediately following the second writing sessions, which is a typical approach for stimulated recall (see Figure 4). During the writing sessions, non-participative observation was used so as to not distract them from their writing as time was limited. Notes of interesting events that transpired were written down so that participants could be asked for clarification and their thought processes in the stimulated recall interviews right after the second writing sessions were completed. The observed participants were given the choice to have the interviews conducted in either Norwegian or

English, all of whom chose to have it in Norwegian. The stimulated recall interviews were recorded with an audio-recorder as to not take away too much time from the interviews to write down replies from learners and to not miss any crucial details that the participants might provide.

3.5.5. STIMULATED RECALL INTERVIEW DATA ANALYSIS AND TECHNIQUES

The audio-recorded interviews of learners were listened to multiple times to aid in analysis, were partly transcribed, and questions and answers were given timestamps as it would take too much time transcribing all the interviews by hand. If something deemed very interesting was said by any of the participants, that was also given its own timestamp with a short description of what was said.

3.6. TEACHER INTERVIEW

3.6.1. TEACHER INTERVIEW AIMS AND RESEARCH COMPONENTS

Interviews are one of the most common types of data collection in qualitative research and are typically classified differently dependent on the degree of structure involved (Riazi, 2016, p. 161). As noted in the previous subsection, these classifications are unstructured, semi-structured, and structured. Structured interviews are designed with little flexibility in mind and are usually conducted with standardised formats (Riazi, 2016, p. 161).

Teacher interviews were used in order to gain background knowledge of why and how learners use Google Translate in class based on their previous experiences with their teachers. The answers that teachers provided during the interviews may help in explaining why learners performed certain actions during the writing sessions as well as providing more insight into how teachers perceive Google Translate as a language learning tool.

3.6.2. TEACHER INTERVIEW SAMPLE

The lead teachers for two of the participating classes were interviewed. The teacher of Class B had been teaching for three years, while the teacher of Class C had been teaching for two years.

3.6.3. TEACHER INTERVIEW INSTRUMENT

For the teacher interviews, a structured interview approach was used. Questions that were asked during the interviews dealt with the teachers' perceptions and experiences of GT in class, whether the learners had had any training in how to use Google Translate, and positive and negative aspects regarding learners' use of GT in class (see Appendix 3A or Appendix 3C).

3.6.4. TEACHER INTERVIEW PROCEDURE

The teachers were given the choice whether their interview should be conducted in Norwegian or English, both of whom preferred to have it in Norwegian. Both teacher interviews were recorded using an audio-recorder as to make the analysis process easier and achieve accurate testaments of what teachers said during the interviews rather than what the author of this thesis would be able to write down by hand.

3.6.5. TEACHER INTERVIEW DATA ANALYSIS AND TECHNIQUES

The audio-recorded interviews of teachers were listened to multiple times to aid in analysis, partly transcribed, and questions along with answers that correlated directly to the questions were given descriptive timestamps as it would take too much time transcribing both interviews by hand.

3.7. PILOT STUDY

A pilot study was conducted prior to the proper study to determine if it were necessary to implement any changes and to make sure that all technical aspects worked properly. The pilot study was conducted at a different school to the one of the proper study, and it was conducted on learners in year 6. The pilot school had divided its learners into groups based on their English level, one for LP learners, one for AP learners, and one for HP learners. As the author of this thesis assumed that learners at every level could write something on the topic when given GT to assist them, it was decided that for the pilot study, the first writing session would be conducted where learners were not permitted to use GT, only dictionaries. The author of this thesis also decided to conduct the pilot study on the HP learners in order to determine whether the topic was too challenging. There were concerns regarding whether the topic that

learners were going to write about was a topic that they could produce more than one sentence on. It was reasoned that if the HP learners struggled to write anything substantial on the topic, the topic should change for the proper study. As LP learners might struggle to write a lot on any topic when only given a dictionary, it proved more useful to see what the HP learners were capable of writing on the topic.

The pilot study writing session took place over a 60-minute lesson, with the first 30 minutes being split between instructing learners how to download and use the screen recording program, the topic that they were going to write about, and preparing the documents and video files for download. The second 30 minutes of the lesson were given to the actual writing session. Some issues did arise during the pilot study writing session, mostly technical issues where the screen recording program would stop recording, one learner whose Chromebooks turned off due to low battery, and two learners who struggled to think of anything to write. Some learners also struggled to prepare their files for download due to confusion of where the saved video file's location was on their Chromebooks.

Overall, the pilot study was considered a success as it went better than expected. Learners wrote considerably more in 30 minutes than expected, but it is important to note that these were HP learners. Additionally, the author of this thesis did not download any of the files, nor did he test the recording equipment for interviews as his application from the Norwegian Centre for Research Data (NSD) had not yet been approved at that point, meaning no personal data could be collected. Some adjustments and further tests were made after the pilot study to decrease the risk of something not going according to plan in the actual study.

Firstly, some learners claimed that their screen recording program stopped working on its own which was worrisome. After talking with these learners, it seemed more likely that user errors had happened as one of them had updated his Chromebook by accident, thereby rebooting the whole system and closing the recording software. Others had accidentally stopped and closed the screen recording program when trying to minimise the window. It was also noted that a lot of learners had clicked on the program shortcut multiple times before the writing session started, thereby opening multiple instances of the program. If any of the learners' screen recording program did crash, this might have been the reason behind it. To combat this, notes were put in the presentation slides to remind learners not to open the program more than once. Nevertheless, after the pilot study was completed, the author of this thesis conducted multiple tests to determine if the screen recording program could crash on its own. Multiple one-hour long recordings were done on three separate devices to see if the

program ever crashed, but no crashes or errors could be reproduced. In case the program were to crash during the actual writing sessions, the author of this thesis added a few notes to the instruction part of the session to instruct learners to restart the screen recording program.

Secondly, to provide support for learners who struggled with writing, a few sentence starters were added to both presentation slides used for instructing learners on the screen recording software and the topic which they would write about. This could prove especially useful for LP learners who usually spend a long time getting started with writing compared to HP learners. Another adjustment made for the actual study was to have the teacher publish a copy of the slides that were used in the introduction to the learners on their learning management system once the writing sessions started so that the learners could revisit earlier slides if they forgot any instructions that had been talked about earlier.

Thirdly, it became apparent that during the pilot study, learners were not familiar with electronic dictionaries as they did not know the difference between a dictionary and a translator, nor did they know the name of any electronic dictionaries. To support learners during the first writing session in the proper study, hyperlinks to three Norwegian-English dictionaries were added to the presentation slides for them to use. The pilot study also brought to light another notable issue. When using Google to search for dictionaries or words e.g. "hund engelsk ordbok" (*dog English dictionary*) or "hund på engelsk" (*dog in English*), Google Translate is embedded so deeply into Google searches that the translator shows the translation of the specified words and phrases in English as the first Google result. For the proper study, the author of this thesis therefore decided that learners must have the dictionary websites up and running prior to starting the writing session as GT is not permitted in any form during the first writing session.

Finally, as a lot of time was spent on learners preparing their video files and documents for download. Contemplations were made whether to reduce the time limit in the writing sessions from 30 minutes to 20 minutes to keep the study within the scheduled 60-minute lesson. Discussions were had with the teachers and the thesis supervisor, and it was determined to keep the 30 minutes for writing, especially to accommodate for learners who needed a lot of time to produce a script. To preserve the 30 minutes for writing, the teachers and the author of this thesis reached an agreement. Once learners completed their writing session, they would either be given a new assignment to perform outside the classroom if they were finished with the essay before the 30 minutes had elapsed, or they would leave the classroom once 30 minutes had passed. The teachers and the author of this thesis would then

stop the screen recordings and prepare the assignments for download on behalf of the learners. This also resulted in another consequence. By having the learners not touch their Chromebooks once they finished writing their texts, the probability of user errors, such as video files not being saved, lessened as the author of this thesis and the teachers were more experienced with the ICT tools used than the learners.

3.8. RELIABILITY AND VALIDITY

For research purposes, the concept of reliability has two applications: The reliability of the instruments of data collection and the reliability of the research reported (Riazi, 2016, p. 271). If the instruments for data collection are inconsistent, then the data will be unreliable, making it difficult to draw meaningful conclusions from the data and analysis. It is therefore vital that for the study to be reliable and valid, the data collection instruments used in the study need to be consistent and stable in eliciting data from the participants (Riazi, 2016, p. 271). Another way to refer to reliability is the whole research process, whether the data collection, analysis, and interpretation of data are reliable or not (Riazi, 2016, p. 271). In research papers, detailed explanations of the research process should be provided so that readers can judge for themselves the plausibility of the conclusions drawn in the studies related to the information the obtain from the reported procedures for data collection and analysis.

In research, the term validity is often associated with traditional quantitative research, however the term can be used in qualitative research as well, often with the alternative term trustworthiness (Phakiti & Paltridge, 2015, p. 31). In essence, validity is the extent in which one can trust the research findings, or in other words what researchers claim as knowledge and understanding of a research problem. Riazi (2016, p. 341) gives another definition to the term, stating that validity is the best possible approximation of the truth stated in the form of an interpretation or interference put forth by evaluators or researchers. He further goes on to say that the more comprehensible and stronger the evidence is, the more valid the interpretation of the evidence will be, which in turn furthers the validity of whichever conclusion one reaches. To provide validity to a study, reliability will be a prerequisite as readers of research papers must be able to see the details of the process of how the researcher came to his/her interpretations and conclusions (Riazi, 2016, p. 341).

As touched upon earlier, a pilot study was conducted before the proper study to expose any flaws in the methodology and to find ways to improve the overall reliability. Furthermore, some of the design choices of the research methodology were made specifically

to strengthen the reliability and validity of the study. For instance, the decision to record the learners' screens while they were writing was two-fold. Firstly, it was necessary to analyse their writing processes through screen recordings to answer the first and second research question. Secondly, having users record their screens while writing also enabled the author of this thesis to see afterwards whether any learners used language translation programs during the first writing session even when they were explicitly told not to. This prevented any falsified results from being mixed with results of learners who adhered to the given guidelines.

A consideration was made whether to delete the learners' first written text from their Chromebook after they had been downloaded onto an encrypted hard drive. This was due to the concern that some learners would refer to their first written text in the second writing session as the topics that they wrote about were quite similar. However, as screen recordings had been made, it would be simple to determine if any learners had done this and in if so, to what extent they copied their previous work. Ultimately, the decision was made not to delete the learners' first written text to address the concern that the author of this thesis' hard drive might corrupt, or files might be mistakenly deleted, thereby losing all data and not being able to retrieve it again if the learners had deleted their texts.

Another concern that could raise questions about the reliability and validity of this study was the number of schools asked to participate. It would have been more beneficial and could have strengthened the validity of the study more if at least one more class at a different school participated in the study. However, as this thesis was written during the COVID-19 pandemic, it was deemed appropriate to only conduct research at one school. There were, however, no indications that showed that the three classes used in the study did not resemble other EFL classes in Norway.

For the results to seem valid and reliable, there had to be enough participants who consented to being a part of this study. Out of 64 learners asked to participate in the study, 37 gave their consent (see Table 1). It was also important that there were enough learners present when the study was conducted, and that these learners were present during both writing sessions so that their texts could be compared. After both writing sessions were completed, four participants were not present during both sessions, meaning that they could not be included in the analysis of how GT affected their written production skills. Furthermore, four participants were not present during the second writing session, meaning these learners' habits of how they used GT in essay writing could not be analysed.

As this thesis explored how learners' usage of GT affected their written texts, it was important that as many learners as possible used GT to aid in their writing during the second writing session. However, forcing them to make use of the tool would have caused invalid results as not all learners typically make use of the tool to help in writing. This is another reason for asking three classes to partake in the study to combat the numbers of learners who would be excluded from the analysis due to them not using GT. Reviewing the 31 screen recordings collected from the second writing session, only 22 learners used GT in some capacity to aid in writing their texts: five LP, 13 AP, and four HP. Due to the exceedingly low number of LP and HP learners who were included in the paired t-test analysis of the quality of their written texts, results relating to these two groups to provide answers to the second thesis question may be unreliable. This unreliability is furthered by the normality tests that could not be calculated for HP learners' data groups and the results of some normality tests which showed some data groups to not be normally distributed. However, the other options would have been to include learners who made no use of Google Translate as part of the analysis process or force learners to use GT during the second writing session, but this would have resulted in invalid data. These unreliability variables were taken into consideration when discussing the data in relation to the second thesis question.

In this study, learners were only given 30 minutes for each writing session, and the study was only conducted in three classes. This was due to time pressure and the unfeasibility of one person analysing all the collected data within an acceptable time frame. As such, the results and conclusions that come from this study should only act as indications. However, as the conclusions of earlier research conducted on this topic point towards positive effects of using GT, the reliability and validity on this aspect of the study can be strengthened as well.

The vocabulary analysis tool, Duolingo CEFR checker could raise concerns regarding reliability. As this ICT tool partly uses machine learning to label various words to the different CEFR levels, it is not completely reliable. However, as the Duolingo CEFR checker is trained using words labelled by international researchers within vocabulary development in conjunction with the vigorous amount of testing that the author of this thesis has done to determine its accuracy and that the tool was only used to compare each learner's written product against themselves, its reliability was deemed acceptable.

Lastly, to ensure that learners' texts were compared as objectively as possible, guidelines were created to make the process of analysis and comparisons faster, more efficient and more reliable (see Appendix 4). The author of this thesis recruited help from one of the

teachers who had participated in the study to determine whether the guidelines used for analysis were reliable. Five randomly chosen texts created by learners after the first writing session were analysed by the author of this thesis and the teacher to determine the interrater reliability of the guidelines provided in the form of percentage agreement (Albano, 2016). Albano (2016) describes percentage agreement as the simplest measure of interrater agreement and that it is calculated as the number of times a set of ratings agree, divided by the total number of sets of ratings, multiplied by one hundred. The main strength of percentage agreement is its simplicity of use, while its major drawback is that it does not account for chance agreements between raters. Although the drawback of this interrater agreement measure was important to keep in mind, the probability of the two raters reaching the same conclusion and thereby agreeing for the various rating categories purely by coincidence was modest at best. This was because the raters were comparing numbers that could fluctuate a lot where the odds of reaching the same number would be quite small. 25 ratings were done in total to determine the guidelines' reliability, five for each learner. In total, 23 out of 25, or 92% of ratings were in agreement between the two raters (see Appendix 5), thereby making the guidelines seem reliable when analysing the learners' texts in terms of objectivity.

3.9. ETHICAL CONSIDERATIONS

Prior to the research being conducted and data being collected, ethical considerations had to be made as they are critical to all research (Phakiti & Paltridge, 2015, p. 31). Firstly, an application had to be sent in to NSD for approval before the author of this thesis could be allowed to conduct his research. After the application was approved (see Appendix 1), participant information sheets were sent out to the teachers, learners, and learners' parents who were going to participate in the study. These information sheets contained information regarding the aims of the study, the participants' rights and how their data was being collected and stored (see Appendix 2A and Appendix 2B). All participants were also made aware via the information sheets that all data collected would be destroyed prior to the project being finished. The information sheets also contained a consent form where teachers had to consent that they would be willing to participate in an audio-recorded interview, and the parents of the learners consented to their child participating in the study. An important note was that although the parents of the learners were the ones who had to consent on behalf of their child, the learners themselves could still refuse to participate with no reasoning needed to be given. Furthermore, all participants were promised that their information would be made

anonymous. For this study, this meant that learners were given numbers instead of names when referred to, and the name of the school where the study took place was not included.

Besides the ethical considerations accounted for above, performing research related to applied linguistics typically carries a low-risk, and the participants in this study were never exposed to any threatening issues (Riazi, 2016, p. 106). However, the author of this thesis does acknowledge that using screen recording to collect empirical data, although not obtrusive, can still be highly invasive of the learners' personal lives. Nevertheless, its use was determined as being acceptable for two reasons. Firstly, participants had been made aware that it would be used in the participant information sheet, and secondly, since the Chromebooks that learners possess are school property, the IT department already has some access to what learners are doing on their Chromebooks. Obtaining approval from NSD signified that the proposed project complied with all ethical considerations and that it is compliant with the European Parliament's General Data Protection Regulation.

4. RESULTS

4.1. INTRODUCTION

This chapter presents the results that were gathered from the two writing sessions along with summaries of the stimulated recall interviews that were conducted with learner participants as well as summaries of the teacher interviews. Data collected from the learners' first writing session is presented first, how they made use of dictionaries during the writing session and the analysis of their texts. This is followed by the results from the second writing session, first presenting how learners used Google Translate to aid them in writing, then the analysis of their texts. The subsequent section compares the texts from two writing sessions to determine differences between learners using a dictionary and Google Translate to aid in EFL writing. Data relating to learners' dependency on Google Docs' correction tools is then presented. Lastly, summaries of the learner stimulated recall interviews are given, followed by summaries of the teacher interviews.

4.2. FIRST WRITING SESSION USING DICTIONARIES

This section presents data collected from the first writing session where learners were not permitted to use Google Translate nor any other tools that allowed for multiple word translations to help them. Instead, they had to rely on electronic dictionaries if they needed assistance in writing their texts. Notable events and interesting information that were observed from the screen recordings are also presented here. After confirming that each learner did not make use of Google Translate in this specific writing session, their data was inputted, and calculations were made of their scores.

In total there were 36 participants in the first writing session but only 22 of them were included in the analysis of the quality of their written texts as these learners were also present and made use of GT in the second writing session as well as handing in a screen recording for both writing sessions. The other learners were not included in this analysis, but data of how these learners used a dictionary to aid in writing were included in a separate analysis (see section 4.2.1). The 22 learners who were part of the first analysis were grouped based on their performance level: five LP, 13 AP, and four HP.

In the first writing session, dictionary usage was mixed between learners. Some learners did not opt to use a dictionary at all, others used it once or twice, and some learners used it to translate multiple words to help them write their essay. Out of the 33 participants who were part of this analysis from the first writing session, 21 of them attempted to use a dictionary in some way. Five LP learners and 11 AP learners used a dictionary or attempted to use one, some using it once or twice, other using it a lot. On the other hand, only five HP learners used a dictionary, and each of these learners only made use of it once or twice.

Each learner who made use of a dictionary to help them in writing used the same dictionary, Dinordbok[10]. Learners were not told that they had to use this specific dictionary, but it was shown as one example of several that learners could make use of in the writing session. Based on screen recordings, the author of this thesis started to wonder whether learners had received any training in using dictionaries, as a majority of learners that did make use of one seemed to either struggle with using it or input an incorrect word into their text after using the dictionary. For instance, a lot of learners inputted various Norwegian verbs into the dictionary and then copied one of the results the dictionary outputted into their document without taking conjugation into consideration. Some of these learners also inputted conjugated verbs into the dictionary which yielded incorrect outputs. The Norwegian words "løper" and "hopper" are two such examples that were inputted by two different learners where both learners used the words to represent the present tense of the verbs "løpe" and "hoppe". However, as the dictionary does not take conjugation into consideration when translating, it instead translated the words as their noun counterparts which resulted in the wrong outputs *runner* and *jumper* instead of *run(s)* and *jump(s)*, respectively. This also happened when nouns were inputted. It appears learners assume that the result outputted from the dictionary is the correct noun needed for their texts as they do not consider whether the noun they receive and input into their text needs to be singular or plural. The author of this thesis was further led to believe that the learners had not received training in using dictionaries as a large number of learners attempted to use the dictionary identically to how they used Google Translate, in that they inputted phrases and multiple words into the dictionary with the expectation of receiving an output. "Helt nye" (*brand new*), "hele huset"

---

[10] https://www.dinordbok.no/norsk-engelsk/

(*the whole house*), and "maser på meg" (*nagging me*) are three examples of phrases that were inputted into the dictionary, none of which the dictionary could provide an output for.

Furthermore, the dictionary that learners used in this writing session did not handle misspellings very well nor did it provide outputs for most phrases, resulting in learners not managing to express what they intended to write, instead deleting the sentence that they had started (see Figure 6). In other instances, a lot of learners seemed to be overwhelmed by the large number of translated suggestions given by the dictionary, also resulting in them deleting their sentences.
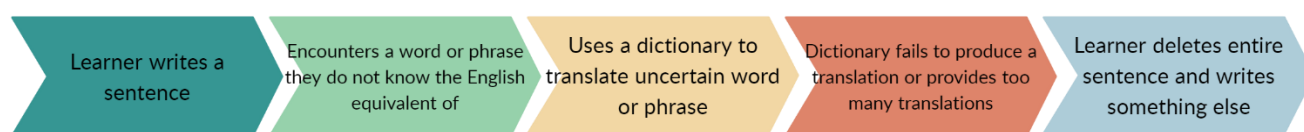
Figure 6: Visualisation of the process of learners' failed dictionary usage

One example of this came from a learner who was writing the sentence *I'm afraid if it's going to happen something with them maybe…* and was unsure how to translate "overkjørt" (*run over*). She inputted the incorrect spelling "over kjørt" into the dictionary instead of the correct spelling. The dictionary was unable to provide an output for the word and it did not provide a suggestion of the correct spelling which resulted in the learner deleting the whole sentence that she had started and writing a sentence detailing her bed instead. In another instance, one learner was unsure how to translate "dykke" (in this context, *scuba dive*) in her sentence *I put on a lot of equipment before I go…*, so she used the dictionary to translate the word. The dictionary provided thirteen translations of the inputted word, but the sheer number of suggestions seemed to confuse the learner even further, resulting in her deleting her half-written sentence and writing about a different activity that she enjoyed doing in Greece instead.

The 22 learners who were included in this analysis were grouped based on their performance level: five LP, 13 AP, and four HP.

Table 2: Analysis of the texts written in the first writing session

Mean category: Means for:	Number of participants analysed	Minimum of one syntax error per sentence*	Minimum of one subject-verbal concord error per sentence*	Spelling mistakes based on percentage of words written*	CEFR-level	Essay length in words
Lower performing learners	5	0,552	0,033	0,048	2,2 ~ A2	57
Average performing learners	13	0,326	0,022	0,018	2,31 ~ A2	152
Higher performing learners	4	0,133	0,025	0,007	2,25 ~ A2	197
Overall mean	22	0,342	0,025	0,023	2,27 ~ A2	138

*Lower number means fewer errors

With regard to mean errors that learners' texts had after the learners had completed their first writing session, results were mixed. As shown in Table 2, there is a mostly even decline in errors from one performance level to the next in regard to syntax errors and spelling mistakes as well as an increase in essay length.

However, there were two exceptions where learners' performance levels did not result in an increase of *quality* in their texts, subject-verbal concord errors and the overall CEFR-levels. As discussed in the methodology section, the author of this thesis did not blindly trust the CEFR-level produced by the CEFR checker. However, after manually reviewing each text written by the learners, data cleansing misspelt words that were given a higher CEFR-level than they were supposed to or did not receive a CEFR-label due to being misspelt and comparing some uncertain labelling of words with the labelling that the Oxford Learner's Dictionaries gave the same words, the results confirm that AP learners scored slightly higher overall in CEFR-level than HP learners. This result can possibly be explained by the low number of HP learners included in the analysis. On the other hand, the slight difference between the two groups related to mean subject-verbal concord errors per sentence is more difficult to explain, but again, it may have been due to the large number difference of learners between the two groups.

4.3. SECOND WRITING SESSION USING GOOGLE TRANSLATE

Relevant data collected and notable events that transpired in the second writing session are presented in this section. In the second writing session, learners were permitted to use Google Translate to aid them however they wished. Prior to the writing session starting, it was emphasised to the learners that they were encouraged to use the translation tool the same way that they would regularly use it to support them as to not force the tool onto them in ways they would never make use of in other contexts. Some of the learners did not make use of any translation tools to help in writing during this writing session. It also quickly became clear that they had more experience of using GT based on how accustomed they were with the user interface of the translation tool. Learners also spent more time away from their texts when they used the online dictionaries compared to Google Translate as the dictionaries took longer to perform translations of words compared to GT's near instant translations.

A total of 33 learners participated in this writing session, but two learners did not give consent to their screens being recorded but did allow for their documents to be downloaded and analysed. As their screen recordings could not be analysed, they were omitted from both analyses that were conducted from the second writing session, both for what purposes they used GT to help in English writing as well as how GT affected the quality of their written texts.

4.3.1. HOW LEARNERS USED GOOGLE TRANSLATE TO AID IN EFL WRITING

Out of the 31 learners who participated in the second writing session and had given consent to record their screens, 23 of them used Google Translate to some capacity: six LP learners, 13 AP learners, and four HP learners. The remaining eight learners who had no help from GT whilst they were writing their texts comprised of seven HP learners and one LP learner. The finished product that the LP learner produced consisted of only three sentences and contained a simplistic vocabulary which might provide reasoning as to why he did not need help from a translation tool. The seven HP learners who did not receive any help from translation tools in this writing session, also never made use of a dictionary in the first writing session.

Table 3 shows how the different groupings based on performance level in English used Google Translate in the second writing session. Whenever a learner used GT, notes were taken of what was translated, how GT handled the translation, and for what purpose GT was used to help:

- Word-level to aid in the translation of a single word
- Phrase-level to help in translating a phrase or a few words of a sentence
- Sentence-level to translate a full sentence
- Whole-text-level to translate the entire text from L1 to L2
- Double-checking tool to determine whether the word that the learner had already written in Google Docs was correct

In total, Google Translate was used to aid in translation at word-level 63 times, phrase-level 35 times, sentence-level 15 times, whole-text-level once, and once as a double-checking tool. Learners seem to be hesitant to use GT to translate anything more than phrases. Learners mostly use GT to translate single words from Norwegian to English, sometimes translating at phrase-level to offer the translation tool more context to produce a correct output of words that are dependent on surrounding words, e.g. verb tense, singular or plural nouns, or words with multiple meanings.

Table 3: Learners' usage of Google Translate

Translation type: Learner group:	Word-level	Phrase-level	Sentence-level	Whole-text-level	Double-check that a word was correct
Lower performing learners	15	7	3	1	0
Average performing learners	36	18	9	0	1
Higher performing learners	12	10	3	0	0
Total usage	63	35	15	1	1

As there was a disproportionate number of learners in the three performance groups that used Google Translate, means were created that accounted for said number difference to produce results which were not skewed (see Table 4).

Table 4: Mean number of times a learner used Google Translate to aid in writing

Translation type: Learner group:	Word-level	Phrase-level	Sentence-level	Whole-text-level	Double-checking tool
Lower performing learner	2,50	1,17	0,50	0,17	0,00
Average performing learner	2,77	1,38	0,69	0,00	0,08
Higher performing learner	3,00	2,50	0,75	0,00	0,00
Overall mean	2,74	1,52	0,65	0,04	0,04

Results show that, on average, AP learners used Google Translate to translate at word-level, phrase-level, and sentence-level more compared to the LP learners, even when the difference in the number of learners in each group was accounted for. One logical reason for this is that the different learner groups' mean essay length differed considerably. To compensate for this, further calculations were made to account for the mean essay length of each performance group in relation to the number of times they used GT (see Table 5). Results from Table 4 also show that HP learners use GT more than the two other performance groups, however, one has to take into consideration that out of 11 HP learners who participated in the second writing session, only four used GT in some manner. This indicates that HP learners show more independence than the other two performance groups as a majority of them made no use of the tool at all.

Table 5: Mean number of words written by a learner prior to each use of GT

Translation type: Learner group:	Word-level	Phrase-level	Sentence-level	Whole-text-level	Double-checking tool
Lower performing learner	18	39	92	276	N/A
Average performing learner	46	91	182	N/A	1638
Higher performing learner	47	56	188	N/A	N/A
Overall mean	37	62	154	92	546

Table 5 is read as the mean number of words a learner within a performance group was able to write independently before each time GT was consulted to help at a specified level. The higher the mean word count for each GT use was, the higher level of independence the learner

showed. When mean essay length is taken into consideration, the data shows that AP learners consulted GT for word-level translations for every 43 words written, whilst LP learners consulted GT for every 18 words written to translate at word-level. This means that AP learners showed more independence and used GT less than LP learners, not only at word-level translations, but also phrase-level and sentence-level. AP learners and HP learners used GT for word-level translation approximately as much as one another when essay length was taken into consideration. On the other hand, AP learners used GT for phrase-level translations less than HP learners, but HP learners used the tool less for sentence-level translations.

In terms of how effectively learners were able to use Google Translate, data collected from the screen recordings show mixed results. Some learners were able to use GT effectively, providing the translation tool with multiple words as to give the program context to work with when providing an output, whilst others seemed to lack knowledge of some of the more *advanced* capabilities of the translation tool which resulted in them not being provided the output they were hoping for.

For instance, one LP learner attempted to translate the Norwegian word "hytten" (*cottage* or *cabin*) into English, but misspelt the word thrice, first as "huten", then "huta", and finally as "huyten". Unfortunately for the learner, Google Translate did not understand which word the learner was attempting to write and therefore did not provide a suggestion for the correct spelling of the word in Norwegian. The learner therefore decided to change the topic of their text. It appears this learner had not been taught that GT operates better when given context surrounding the word. The author of this thesis provided the following sentence into GT to determine whether the translator would be able to interpret the word correctly when given more context to work with: "Jeg elsker å være på huyten", which GT promptly asked whether the word "huyten" was meant to be "hytten", the correct spelling. Confirming that this was the case, GT corrected the Norwegian misspelling and provided the output *I love being at the cabin*. This learner was not the only one who struggled with receiving a correct output from GT when inputting misspelt words. In other instances, learners inputted misspelt words and phrases into GT where the translation tool would provide suggestions of correct spellings, however, it would not provide the correct output automatically until the learner pressed on the suggestion. Some learners seemed to be unaware that pressing the suggestion would result in the correct output as they would instead delete their input and write synonyms of their original intended word.

Learners also seemed to be unaware of another feature that GT provides, namely that it can provide multiple possible outputs. One learner inputted "tursti" into GT which the tool automatically detected as a Norwegian word, but the default output shown was the same word as the Norwegian input. The learner therefore decided to alter their input which resulted in a different translation altogether. Not a single learner that participated clicked on the output that GT provided to check whether there were any other suggestions (synonyms or words of similar meaning) that the OT had to offer. The author of this thesis checked whether GT had any other suggestions for what "tursti" could be in English, and it did have a suggestion for the word *trail* which the learner could have used if they had clicked on the output. The learner also could have provided GT with more context by writing a phrase or sentence that contained the word, as GT would have provided *hiking trail* as an output for "tursti" if more context had been given. Another learner inputted the Norwegian word "boller" into GT which GT outputted as *bowls*, not an incorrect translation but the learner was looking for the translation related to the food item. Just as the previously mentioned learner, this learner altered their input to create a completely different translation than what they had originally intended to write. The author of this thesis again tested whether GT had other suggestions to the input offered, and the translation tool did offer *buns* as an alternate output which would have been the correct output for the learner's sentence. The learner could have also provided the translation tool with more context to receive a translation of a specified word, for example by writing "jeg liker å spise boller" which GT would have outputted as *I like to eat buns*, the correct translation.

When the learners used Google Translate, there was a common trend amongst them where if GT provided them with an output that was seemingly unknown to them or had advanced vocabulary, they would either disregard the output or alter the input to force the translation tool to offer them different output that was more familiar to them even if the original output was correct. One example of this was from a learner who inputted "av og til" into GT and received the output *occasionally*. Instead of using this output, which was correct, the learner altered the input to "noen ganger" and received the output *sometimes* which the learner opted to use instead. However, despite learners altering the input to force GT into providing a different output of simpler vocabulary or changing the input because GT failed to produce an output due to spelling mistakes or other reasons, learners still seemed to be able to produce something in English related to what they wanted to express in Norwegian in their essays when receiving help from GT. When learners struggled with translation using dictionaries, a lot of them deleted their sentences due to a word or phrase proving difficult to

translate, and then continued writing their texts without a replacement sentence to accommodate for the original deleted sentence. On the other hand, when learners were able to use GT to help them translate, this trend disappeared. Although some learners did alter their original sentences that they intended to write when GT struggled to translate certain words or phrases, their sentences were never entirely deleted, instead being altered to keep as much of the meaning as the original intended sentences did.

Another observation made was that only a handful of learners copied and pasted the output received by Google Translate into their documents. The majority of learners instead opted to either drag the output from GT over to their document or manually input the output received. It is unsure whether this was due to learners not knowing how to copy and paste text or if there was a different reasoning. It is interesting that not a single learner used GT's *copy translation* button which copies the output to the clipboard which can easily be pasted anywhere. In contrast, learners manually inputting the translation given by GT also brought with it two consequences. Firstly, it used up a lot of time as learners continuously alternated between the two tabs to write a portion of the translation into their document before going back to GT to check how to further write the translation. Secondly, a lot of learners misspelt the words and phrases that GT outputted when they manually wrote them in their documents, meaning either Google Docs had to correct their mistakes, or their spelling mistakes were so severe that the word processor had no alternate suggestions which resulted in the words remaining incorrect even when the words had red underlines.

Further, only one learner, classified as LP, out of 31, translated their whole text from Norwegian to English through Google Translate. This learner's Norwegian input was riddled with spelling mistakes (seven in one sentence), but GT was still able to produce a near perfect output.

The final notable observation made regarding learners' usage of Google Translate was that GT helped a lot of learners, especially LP ones, in producing correct outputs or offering suggestions for correct inputs when words, phrases, and sentences were misspelt. Misspelt Norwegian words and phrases for which GT produced correct outputs included: "trenings appareter" (treningsapparater, *exercise equipment*), "slapeav" (slappe av, *relax*), "definetivt" (definitivt, *definitely*), "någen ganger spiler vi" (noen ganger spiller vi, *sometimes we play*), "jallefall" (i hvert fall, *at least*), and "egntli" (egentlig, *actually*).

The 22 learners who were included in this analysis were grouped based on their performance level: five LP, 13 AP, and four HP.

Table 6: Analysis of the texts written in the second writing session

Mean category: / Means for:	Number of participants analysed	Minimum of one syntax error per sentence*	Minimum of one subject-verbal concord error per sentence*	Spelling mistakes based on percentage of words written *	CEFR-level	Essay length in words
Lower performing learners	5	0,470	0,000	0,049	2,00 ~ A2	46
Average performing learners	13	0,196	0,042	0,011	2,15 ~ A2	126
Higher performing learners	4	0,082	0,085	0,005	2,25 ~ A2	141
Overall mean	22	0,237	0,041	0,019	2,14 ~ A2	111

*Lower number means fewer errors

When learners used Google Translate to aid in their writing, the number of errors they produced varied greatly amongst the different performance groups (Table 6). LP learners had the highest mean of errors related to syntax errors and spelling mistakes in the second writing session. However, none of the LP learners' texts contained any subject-verbal concord errors. A possible explanation to this, with empirical evidence from the screen recorded videos, was that these learners created shorter sentences with less advanced vocabulary compared to the other two performance groups. This may also explain why HP learners' texts contained a higher mean subject-verbal concord errors than AP learners. With the exception of subject-verbal concord errors, the *quality* [11] of the performance groups' texts increased in all other categories from one performance level to the next.

4.4. COMPARISON OF THE LEARNERS' OWN TEXTS

In total, 22 learners' own two texts were compared against each other to measure changes in vocabulary, syntax error, subject-verbal concord, spelling mistakes, and essay length as these learners participated in both writing sessions, made use of Google Translate in some manner

---

[11] Improved text quality refers to a decrease to syntax errors, subject-verbal concord errors, and spelling mistakes, and an increase to CEFR-level and essay length.

during the second writing session, and handed in screen recorded videos along with their texts in both writing sessions.

Table 7: Mean difference of learners' written texts between the two writing sessions

Mean category: / Learner group:	Number of participants analysed	Minimum of one syntax error per sentence	Minimum of one subject-verbal concord error per sentence	Minimum of one subject-verbal concord error per sentence	Spelling mistakes based on percentage of words written	CEFR-level	Essay length in words
Lower performing learners	5	-0,082	-0,033	0,001	-0,20	-10
Average performing learners	13	-0,131**	0,020	-0,007*	-0,15	-25
Higher performing learners	4	-0,051	0,060	-0,002	0,00	-56
Overall mean difference	22	-0,105**	0,015	-0,004	-0,14	-28*

*$p<0,05$ and **$p<0,01$. Significant difference in specified category between the first and second writing session.

Presented in Table 7, a negative mean in relation to the categories indicates a positive change from the first writing session to the second writing session as it means that the number of errors dropped for those three categories: minimum of one syntax error per sentence, minimum of one subject-verbal concord error per sentence, and spelling mistakes based on percentage of words written. Inversely, a positive mean in the same categories indicates an increase in number of errors from the first writing session to the second writing session. In the other two categories: CEFR-level and essay length in words, a negative mean indicates a drop in performance from the first writing session to the second writing session. Furthermore, means shown with * or ** indicate that there were significant differences between the first and second writing session.

Firstly, the mean number of syntax errors per sentence reduced for all learner groups from the first to the second writing session. Furthermore, for AP learners this decrease in number of errors was shown to have a positive significant difference. The overall mean difference from the first to the second writing session also saw a positive significant difference, which indicates that Google Translate may have had a positive impact on learners' texts when it comes to syntax errors.

On the other hand, there was a slight decrease in number of subject-verbal concord errors from the first to the second writing session for LP learners, and an increase for both AP and HP learners. None of these results were calculated to be of statistical significance,

however, as some of the data groups related to this category of errors did not pass the normality test which made the t-test not as reliable, there is uncertainty to how accurate these findings are. As learners typically used Google Translate for word-level and phrase-level translations rather than sentence-level translations, this might also be an explanation as to why no statistically significant changes occurred in this category between the two writing sessions.

In regard to spelling mistakes based on percentage of words written, AP learners saw a statistically significant difference for the better from the first to the second writing session. LP learners saw a negligible increase of errors in the same category in the second writing session compared to the first writing session, whilst HP learners' texts had a negligible decrease of spelling mistakes produced.

The CEFR-levels of LP and AP learners' texts had a small mean decrease, but it was neither a statistically significant change nor did it change the learners' mean CEFR-level. However, as with subject-verbal concord errors, some data groups relating to CEFR-levels did not pass the normality test, meaning that the results of the t-tests for AP learners and overall mean difference might be unreliable. HP learners saw no change at all in CEFR-level between the two writing sessions.

Lastly, all performance groups saw a decrease in essay length from the first to the second writing session, but none of these decreases were calculated to be of statistical significance. However, the overall mean decrease in essay length was shown to have a statistical significance. These decreases in essay length can most likely be attributed to the provided writing topics and not Google Translate. This possibility was further strengthened by statements made by learners in the stimulated recall interviews.

4.5. LEARNERS' DEPENDENCY ON GOOGLE DOCS' CORRECTION TOOLS

While screen recordings were being analysed, it became increasingly noticeable that Google Translate was not the primary tool that helped learners achieve the quality of texts that they did. Instead, three features of the word processor Google Docs that all learners used in the two writing sessions had a considerable impact on the quality of the texts: auto-correct, spell check, and grammar check. These three features helped learners at every skill level to correct errors, from simple spelling mistakes, correcting agreement between singular/plural nouns, and subject-verbal agreement to rearranging whole sentences that initially made no sense. For instance, one learner wrote *home maid straberry gam* into Google Docs, and the correction

tools offered to correct the inputted phrase into the correct phrase *homemade strawberry jam*. The same learner also wrote *help me always* and *I have also* which Google Docs offered to correct into *always helps me* and *I also have* respectively.

All three performance-based groups were divided into three new subgroupings based on how much dependency they had on Google Docs correcting their mistakes (Table 8): little dependency, some dependency, and high dependency. Learners who received help from Google Docs' features in every sentence they wrote were grouped as highly dependent. Learners who had help from Google Docs' features in most sentences written, that is to say every second or third sentence, were grouped as having some dependency. Finally, learners who, on average, received aid from Google Docs' features less than every third sentence were classified as having little dependency. Learners who did not permit their screen recordings to be downloaded were not included in this analysis since there was no record of how Google Docs helped them correct their mistakes. Both writing sessions were included in this analysis but were separated to see whether the use of Google Translate had any effect on learners' dependency of Google Docs' features. Furthermore, only two features of Google Docs were analysed: the spell checker and grammar checker. Spelling mistakes that were auto-corrected by Google Docs were not included in this analysis as they only comprised a fraction of overall errors. However, two such examples are when one learner incorrectly wrote *shoping* which was automatically corrected into *shopping*, and when one learner incorrectly wrote *swater* which was corrected automatically into *sweater*. In Google Docs and most other word processors, underlines highlighted in red are spelling mistakes, whilst underlines highlighted in blue are grammar suggestions.

Table 8: Learners' dependency of Google Docs to correct their errors

Learner group:	Feature:	First writing session (7 lower performing, 13 average performing, and 13 higher performing learners)	Spell check	Grammar check	Second writing session (7 lower performing, 13 average performing, and 11 higher performing learners)	Spelling check	Grammar check
Number of lower performing learners with *x* dependency	Little dependency		1	5		1	2
	Some dependency		0	1		3	2
	High dependency		6	1		3	3
Number of average performing learners with *x* dependency	Little dependency		3	1		3	1
	Some dependency		4	4		4	6
	High dependency		6	8		6	6
Number of higher performing learners with *x* dependency	Little dependency		7	7		7	5
	Some dependency		6	5		3	5
	High dependency		0	1		1	1

In the first writing session, there was an almost even split amongst HP learners having either little dependency on the features of Google Docs and having some dependency. In the second writing session, there were two fewer HP learners that participated, but once again the vast majority of HP learners were categorised as having either little or some dependency on Google Docs' correction features.

The results differ for AP learners, wherein both writing sessions, AP learners with a high dependency of Google Docs' spell check feature and grammar check feature made up the largest group. Looking at the screen recorded videos, AP learners often misspelt outputs they received from the dictionary and Google Translate, only for Google Docs to then correct their misspellings. For instance, one learner received the output *definitely* from GT and manually wrote the word in her document. She misspelt the word as "definitly" and Google Docs' spell checker feature flagged the word as incorrect and offered the correct spelling. Just as for HP learners, the dependency on Google Docs' correction tools did not change for AP learners between the two writing sessions, meaning Google Translate did not affect their dependency on Google Docs for spelling mistakes nor grammar suggestions. Although AP learners

received a lot of help from Google Docs, the fact that Google Docs was able to produce corrections of their mistakes shows that these learners still managed to produce words and sentences that one could still make sense of. Had their texts been riddled with severe mistakes, both spelling mistakes and grammatical mistakes, Google Docs might have not been able to produce correction suggestions at all, as was the case for multiple LP learners in the first writing session.

4.6. LEARNER STIMULATED RECALL INTERVIEWS

4.6.1. CLASS A

Three learners from class A had been randomly chosen by their teacher to participate in the stimulated recall interview prior to the first writing session, and they were interviewed shortly after the second writing session had concluded. Two learners (henceforth known as learner A1 and A2) were categorised by their teacher as AP, while the third learner (henceforth known as A3) was categorised as LP.

It was observed during the first writing session that learners seemed to be unaware of how a dictionary functioned which is why the first question the three learners were asked was whether they knew how to use a dictionary. The three learners all replied that they had no experience in using a dictionary, only relying on Google Translate to help in translation.

This led to the question of how the three learners used Google Translate to aid them in translating from Norwegian to English. Learners A2 and A3 reported that they typically used the OT for word-level translations, which their screen recordings also confirmed. Learner A1 answered that he seldom used GT at all, and if he did, it was mostly used as a confirmation tool to check whether a word was written correctly. However, observing him during the second writing session revealed that he used Google Translate thrice, but at word-level, not as a double-checking tool. When asked about this discrepancy, he was unable to provide a specific reasoning. When prompted as to whether the learners used GT for sentence-level translations, the three learners did admit to sometimes using the tool for phrase-level translations, but very rarely would they input more than phrases into GT. This was due to being told by their teacher that GT should not be used for translating larger pieces of texts.

The learners were then asked why they used Google Translate instead of other translation tools, to which they all replied that GT was the only translation tool that they knew

of. Learner A2 said that when she wished to translate something from her L1 to L2, she would typically use Google and search for *translate*. Based on the screen recordings from the second writing session, all learners who performed search queries used Google as their search engine of choice. Learner A2 is not the only learner who made the search query *translate* in Google during the writing session. In fact, multiple learners did, and as they all used Google to perform their search, GT was the first result to appear, not only the URL that brings the learners to the Google Translate site, but GT is also embedded in Google searches where learners can directly make use of the tool (see Figure 7). It is therefore understandable that learners might be unaware of other translation tools as out of seven results shown on the first page of results when searching for *translate*, four of them are related to GT.
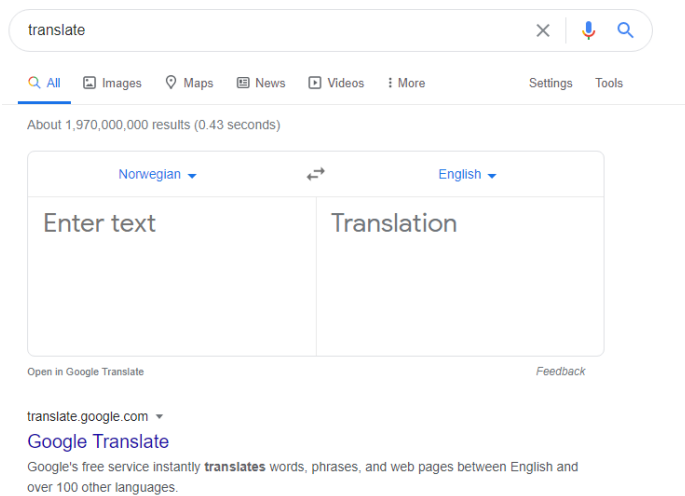
Figure 7: Google Translate embedded in the Google search engine

Learner A1 commented that he was unsure how to properly use GT. The three learners reported that they had not received any training at all in using GT. The only guideline they had received from their teacher was that they should only use it for word-level translations. Although learner A1 was unsure of how to properly use GT, he still preferred its usage over a dictionary as the tool gave him the opportunity to translate more than a single word at a time if he should ever need it. He also commented that he trusted GT more than a dictionary because the dictionary produced a list of multiple synonyms in English based on his Norwegian input and that choosing the correct word needed for his sentence was difficult at times.

From observing the three learners over the two writing sessions, learner A1 wrote significantly less in the second writing session compared to the first once, whilst learners A2 and A3 had a significant percentage increase in essay length in the second writing session. The learners were asked a hypothetical question on whether they believed they would write

more when given the choice between using a dictionary or Google Translate. Learners A1, A2, and A3 all agreed that the usage of GT would enable them to write more in an essay compared to a dictionary. Based on his reply, learner A1 was then asked as to why he wrote significantly less in the second writing session where he had access to Google Translate compared to the writing session where he used a dictionary. The learner replied that he simply ran out of ideas of what to write about regarding the topic given to him, but he would most likely have written more if he used GT compared to a dictionary if there was only one topic to write about instead of two. His reasoning for this was firstly because he felt he was more familiar with GT compared to a dictionary. Secondly, he stated that if he were to struggle with writing a sentence in English, GT could help him to a greater extent. This was because the tool could translate phrases and sentences from his L1 into English compared to a dictionary only translating at word-level, which became difficult to use when struggling with parts of a sentence and not just a single word.

Lastly, learner A3 commented that she wished GT had the function of showing her synonyms of the input she gave the translator, something the dictionary she used in the first writing session did. The other two learners were asked if they knew that GT already had this function where it shows synonyms of the input they write as well as other definitions of the words they input, e.g. showing both *bowls* and *buns* as suggestions for the Norwegian input "bolle", if they click on the output that GT produces. None of the learners knew of this function which further supports their prior statement that they had not received any training in using GT.

4.6.2. CLASS B

In class B, the three learners that had been randomly chosen to participate in the stimulated recall interview following the second writing session were categorised by their teacher as LP (henceforth known as learner B1), AP (henceforth known as learner B2), and HP (henceforth known as learner B3).

Based on screen recordings and observing the three learners in the first writing session, they were first asked if they knew how to use a dictionary and if they had received any training in using one. None of the three learners had received any training in how to use one, but both learner B2 and B3 believed that they did well in the first writing session using a dictionary considering the lack of training. Learner B1 said that she struggled a lot in the first writing session as she did not understand how to use a dictionary, which the screen recording

confirmed. The screen recording showed learner B1 using the dictionary a lot to try and aid in translation of words, but that she was rarely able to receive an output that she was content with or she mistook the output that the dictionary gave as automatically being correct. For instance, the learner inputted "i dag" (*today*) into the dictionary, but the dictionary provided a long list of English synonyms with *today* being the third to last option. The learner chose the first word from the list that was familiar to her which was *day* and inputted that into her text. The learner also inputted "sted" (*place)* into the dictionary but ended up using the first result presented to her, *berth*. As a third example, the learner inputted "hall" (*gymnasium* or *sports hall*) into the dictionary and received the output *antechamber*. However, this time the learner opted not to use the output given.

Learner B2 said she enjoyed the simplicity of GT in that it presented only one output instead of the multiple suggestions that the dictionary gave her. Learner B3 agreed with learner B2's comment, saying that there were times where the dictionary's output of multiple words was too confusing, finding the simplicity of GT a lot better. Both also preferred being able to write more than just words into GT compared to the dictionary as they sometimes lacked the vocabulary to compose phrases or longer pieces of texts. However, learner B2 also commented that the synonym function of the dictionary was a nice feature at times as it helped her essay not to become repetitive with the same words used repeatedly. This learner, along with learners B1 and B3 were asked based on learner B2's comment if they knew that GT also had a feature that would show synonyms or other words the input could be translated into, to which they all replied no, reinforcing the idea that learners have not received a substantial amount of training in using the OT.

When asked for what they typically used GT, learner B2 and B3 commented that they mostly use the tool for word-level translations, sometimes phrase-level if there are multiple unknown words that need translation. Looking at the screen recordings, learner B2 only made use of GT for word-level translations whilst learner B3 made no use of GT at all. Learner B3 was asked why he did not feel the need to use GT during the second writing session, and his reply was that he did not encounter any words during the writing session that he needed help to translate. Learner B1 replied that she never uses GT for whole-text translation, but that she does use it for word-level, phrase-level, and sentence-level translations depending on how much she struggles with certain translations. Her statement agrees with her usage of GT in the writing session, as based on her screen recording, learner B1 used GT for those three purposes. She was further questioned on how she felt GT helped

her writing in the second writing session, to which she did not have a specific reply other than she felt that the tool made it easier for her to communicate what she wished to express in Norwegian into English. The three learners were then questioned on why they use GT instead of other translation tools, to which learner B3 replied that he did not know of any other translation tools beside GT. Learner B2 said that she sometimes used IntoWords instead of GT, but due to its unresponsiveness and more sluggish experience, GT was the preferred translation tool. Learner B1 agreed with learner B2 but said that she never used IntoWords for the reasons mentioned by learner B2.

Questioned on whether they find it easier to write when given the choice between GT or a dictionary to help them, all three agreed that GT was the superior tool to make writing easier. Learner B2 further commented that she felt that GT made her a more independent writer as she did not have to ask the teacher for help or anyone else as much as if she did not have the tool at her disposal. Regarding the length of their texts, learner B1 commented that she simply had more to write about in the second writing sessions and thus the text was substantially longer, not that GT was more helpful. In the second writing session, she used GT to aid in word-level translations four times, phrase-level translations four times, and sentence-level translations three times. The analysis of her two texts also shows that she performed better during the second writing session compared to the first one as her spelling mistake percentage and syntax error percentage decreased. For the other two learners whose texts were about the same length, learner B3 replied that he could not think of anything more to write about in the two writing sessions, and that it was not due to the fault of the dictionary or GT, which is confirmed by his screen recordings as he did not make use of either tool in the writing sessions. Learner B2 agreed with Learner B3 that it was simply a matter of not coming up with more ideas to write about.

4.6.3. CLASS C

From class C, one learner that had been randomly chosen to participate in the interview was categorised by his teacher as AP (henceforth known as learner C1), while the other two learners that had been chosen randomly were categorised as HP (henceforth known as learner C2 and C3).

It was observed that in the first writing session learners in this class as well seemed to struggle a lot with using a dictionary. They were asked if they knew how to use one, both electronic and physical, and if they had received any training on how to use dictionaries. To

both questions, learners C1, C2, and C3 all replied that they had no experience nor had they received any training in how to use one. During the first writing session, learner C1 made use of a dictionary once to aid in the translation of the word "bane" (in this context, *pitch*), but the dictionary failed to provide the output that the learner was looking for and instead opted to use the word *court*. Learner C2 used it a few times, but she failed to conjugate the output that the dictionary provided, resulting in errors that Google Docs highlighted for the learner to fix. Learner C3 made no use of the dictionary during the first writing session. Learners C1 and C2's reasoning as to why they used the dictionary was because they could not think of a suitable translation of the words they wanted to write in English themselves.

The learners were then asked to name the translation tools that they were aware of. All learners named Google Translate as their first answer, but only learner C3 could name another tool, IntoWords. When asked if any of them used IntoWords since they all had it installed on their Chromebooks, all replied that they did not. Learner C2 expressed that she did not understand the tool as it was clunky and not user friendly.

Regarding for what purposes the learners use Google Translate, learner C2 used it mostly at word-level, sometimes using it at sentence-level if necessary. Learner C1 mostly used it at word-level and phrase-level but did make use of it at sentence-level when he struggled to come up with a good sentence translation himself. The learner also mentioned that he had dyslexia and that he used GT as a tool to help him when writing in English because he often struggled with spelling English words correctly. The learner struggled less with spelling Norwegian words correctly, so it was easier for him to input the Norwegian word into GT and receive a correctly spelt output to use in his text. Learner C3 said that she rarely used GT at all, but if she did, she used it solely at word-level as a dictionary. She further commented that she typically used GT as a tool to hear the pronunciation of words that were unknown to her by using the listen function.

All three learners agreed that they believe they find it easier to write when using GT compared to using a dictionary. Learner C2 argued that using GT to translate takes less time compared to dictionaries and that the times she needed help with translating more than individual words, GT could provide that support while dictionaries could not.

When observing the three learners over the two writing sessions, learner C1 and C2 wrote less in the second writing session compared to the first one. A hypothetical question was therefore given to the learners where they would only be having one writing session and if they had the choice between GT and a dictionary, did they believe that they would write

more when given a dictionary or would they write more if Google Translate was an option to use. Learner C1 was unsure and could not provide a definite answer, while learner C2 expressed that she would probably write more if given GT due to its speed and ease of translation. Learner C3 said that she would most likely write the same amount since she barely makes use of any translation tool, be it a dictionary or GT. As results from the two writing sessions showed that learner C1 and C2 wrote less in the second writing session even when given GT to assist them and based on the responses from the previous question, the author of this thesis probed as to why this might have been the case. Learner C2 and C1 expressed that they wrote as much as they did because the two topics were too similar and that they simply did not have anything more to write about than they did.

In the last section of the interview, the learners could freely express any thoughts or experiences that they had with dictionaries, GT, other translation tools, or the writing sessions themselves. Just as in the previous two interviews, the only dislike they had with GT was that it only provided them with one output for any word that they inputted. This confirmed that learners did not know that GT provided synonyms or other words that the input could be if the output was clicked on. The three learners also commented that they had not received a lot of training in using GT, only how to use it at word-level.

4.7. TEACHER INTERVIEWS

4.7.1. TEACHER OF CLASS B

When asked about the experiences that the teacher had had with learners using Google Translate in class, his reply was that when learners used the OT the way he wanted them to use it, that is to say, as a dictionary, he had positive experiences with the tool. The teacher attributed this positive experience of Google Translate mostly to learners categorised as AP or HP because they were more often able to use the output given by Google Translate in contexts other than what the output displayed, e.g. the word "løper" being translated by GT as *runs* but the learner wants to use it in first person singular, therefore conjugating it to *run*. The teacher had less favourable experiences with the tool being used by LP learners, stating that they more often translated large amounts of text at once from Norwegian to English. Since a proportion of the learners that were categorised as LP in English were also categorised as LP in Norwegian, their Norwegian input in GT would often be riddled with spelling mistakes, which caused GT to output sentences that made little to no sense in English. The teacher also

noted that he at times had negative experiences with learners at all performance levels using GT out of laziness.

The teacher reported that learners made use of GT a lot for various purposes such as a dictionary or translating phrases or sentences. The teacher also commented that he believed that learners associated GT synonymously with a dictionary as the only experience they had with translation tools was GT, further stating that learners would probably be unable to name a translation tool other than GT.

Further, the teacher was asked if any other translation tools were used by learners. The teacher responded that learners were encouraged to make use of IntoWords[12], a tool that provides learners with reading and writing support. However, both teachers and learners had negative experiences with the tool, citing frequent freezes and crashes as well as buggy behaviour as the main reasonings as to why learners preferred to use Google Translate over IntoWords. On the other hand, teachers had been given some training on how to teach learners how to use IntoWords properly, however, due to the negative experiences with it, they instead encouraged learners to consult GT for word-level translations.

Finally, the teacher reported that the learners belonging to their class had received some training on proper usage of GT, but they were unsure as to how much of it they had retained. The learners had received training on how to use GT as a dictionary, the pitfalls of GT, and that GT provides a more accurate output when given parts of a sentence compared to a single word. The teacher also reported that he had not received any form of training in Google Translate, only being aware of the capabilities the tool provides that he had taught himself.

4.7.2. TEACHER OF CLASS C

Teacher C was first asked about the experiences he had with learners using Google Translate in class. His reply was that GT is used when necessary, having positive experiences when the tool is used as a dictionary. However, the teacher also expressed a drawback of GT related to learners not always understanding that the output given by GT cannot always be copied and pasted as is, requiring conjugation or other changes depending on the context the word will be placed in. He further commented that HP learners typically handle the output given by GT better than LP learners as LP learners not only use GT more than other learners, but that they

---

[12] https://vitec-mv.com/no/produkter/intowords/ for more information

also copy and paste the output given into their documents without any alterations of the output. The teacher also commented that he does not typically experience learners using GT out of laziness, but LP learners using it out of necessity as their vocabulary and mastery of the English language is so low. Learners who have a better understanding and mastery of the language mostly use it at word-level or phrase-level to translate words they are uncertain of.

According to the teacher, IntoWords was used by some learners, almost exclusively LP learners, but that GT was the main translation tool used by every learner in his class. Learners had also received a very small amount of training on how to use GT effectively, but that this training had taken place over a year prior to the interview, with the teacher expressing doubts as to how much knowledge from the training the learners had retained. The training that learners had received consisted of how to use GT at word-level, that is to say, as a dictionary, and giving GT more context to work with to translate a specific word. Teacher C, just as teacher B, had not received any official training on using Google Translate as a CALL tool, instead being reliant on self-taught skills.

Finally, when asked to comment on positive and negative aspects that GT provides learners, the teacher answered that in his experience, GT offers benefits when learners already possess some degree of mastery of the language that they wish to translate into, something that a lot of learners in their class currently do not possess. When learners already possess some level of mastery of the language they translate into, the teacher argued that the learners have a higher probability of noticing outputs that contain errors or outputs that make little to no sense in the context they are being used. The teacher did not specify a CEFR-level that learners needed to have before GT would be beneficial for them, but he did state that the tool was not beneficial for most of his LP learners. This was because, although the tool might help them achieve a text with fewer errors than they would have achieved on their own, they would not actually learn any vocabulary or grammar from using the tool as they do would not know when GT produces correct or incorrect output. The teacher also commented that the ease of use and simplicity of the program are two big benefits that the tool has to offer compared to other tools used to help in translation. Related to the negative aspects of Google Translate, the teacher replied that certain learners have a habit of blindly trusting the output given by GT, even when it is completely incorrect as they do not possess the language skills to notice that the word/phrase/sentence is wrong.

5. DISCUSSION

5.1. INTRODUCTION

The following chapter discusses the results in the previous chapter to provide answers to the two research questions. Firstly, the chapter discusses how learners currently make use of Google Translate to aid them in EFL writing. Secondly, the chapter discusses whether GT positively or negatively affects learners' written skills. In the final section of this chapter, the issue related to learners' use of Google Docs' correction tools will be discussed.

5.2. HOW IS GOOGLE TRANSLATE USED BY NORWEGIAN PRIMARY SCHOOL LEARNERS WHEN THEY WRITE IN ENGLISH?

The first research question concerned how Norwegian learners' currently use Google Translate to aid them in translating between Norwegian and English whilst they are writing. In order to provide an answer to this question, an explanatory case study was used which entailed the analysis of 31 screen recordings of the learners' second writing session. Learner stimulated recall interviews also played a role in obtaining relevant information regarding learners' habits of using Google Translate. The teacher interviews also provided insight into how learners had been taught to use the translation tool.

Firstly, examining 31 screen recordings revealed that not all learners need Google Translate to aid in EFL writing, but of the 22 learners who did use GT in some manner, it was used to aid them in their writing a total of 115 times, where over half of the total usage stemmed from learners translating at word-level. These results agree with the outcomes of the study conducted by Aksnes (2018) and the study conducted by Jolley and Maimone (2015), whose conclusions were that learners typically used GT mostly for word-level translations. Furthermore, a lot of teachers' earlier concerns regarding the usage of GT in FL lessons came from learners using the tool to translate paragraphs or complete texts (Jolley & Maimone, 2015; Stapleton & Leung, 2019). The outcome from this study, however, reveals that the utmost few learners use the translation tool in ways that some teachers would deem unethical.

As shown in Table 5 and taking into consideration that a majority of HP learners did not need the tool to aid in EFL writing, HP learners, on average, used GT less to help in every category compared to the other two groups with the exception of phrase-level translations. HP learners might possess more knowledge of the translation tool than the other two groups,

providing GT with context to increase the probability of a correct output which might explain their higher usage of GT for phrase-level translations. It is not surprising to see that HP learners use GT to aid in translation less than the other two groups as their performance level should reflect a higher degree of mastery in the English language. These results also correspond to a report that Niño (2009, p. 246) wrote, which stated that HP learners use GT as a phrase or sentence dictionary to a lesser extent than LP learners and AP learners. However, her argument as to why this is, is that HP learners are aware of the many grammatical inaccuracies that GT possesses. The author of this thesis argues against Niño's statement for two reasons. The first being that GT has progressed a lot since 2009, the year her study was published, making her claims of GT's weaknesses outdated, and secondly, HP learners possess a more advanced vocabulary and a better mastery of the English language, thereby not requiring the same amount of assistance from GT to aid in English writing as the other two groups.

Further, the one learner who translated his whole text from Norwegian to English had multiple spelling mistakes in his L1 input. However, GT was still able to produce a near perfect output of his input. This somewhat contradicts what the teacher of class B said in his interview when he stated that GT would produce sentences that made little to no sense when given input riddled with mistakes. It is completely true that this learner's input was riddled with spelling mistakes, but GT was still able to produce a near perfect output. This is not to say that GT is able to correct spelling mistakes or other grammatical errors in all cases, but it is apparent that the tool performs better than teachers previously thought. These results also contradict Giannetti's (2016) original statement that Google Translate and other translation tools were unable to properly translate misspelt words.

As for gauging learners' mastery of applicable sub-skills in relation to the framework of digital skills and their usage of translation tools, some learners seem to possess more competence than others. For instance, many learners from the study showed a low degree of mastery related to dictionary usage when referenced against the sub-skill *find and process* (Utdanningsdirektoratet, 2017). Many learners struggled to make any use of the dictionary during the first writing session, inputting misspelt words or long phrases, and were unable to find suitable translations from the list of words produced by the dictionary. Further, all learners who did make use of a dictionary to aid in writing, only consulted one dictionary even though they were told that they had the opportunity to use multiple. A high degree of mastery in *find and process* entails that learners, amongst other things, are able to critically

assess information from different digital sources. This is difficult to fully assess as learners were only encouraged to use GT for the second writing session, nonetheless, parts of it are still relevant. Some learners directly copied the output that GT produced with no hesitation into their texts, thereby having Google Docs suggest the correct output afterwards. Other learners, on the other hand, seemed to be more critical of the tool, altering their input at times to see how that would affect the output given to them, and were able to manipulate the output given by GT to fit the context of their texts.

5.3. How does the use of Google Translate affect Norwegian learners' English written production skills?

The second research question concerned how Norwegian learners' current usage of Google Translate affected the quality of their written work, either negatively or positively. A within-group quasi-experimental study comparing vocabulary, spelling mistakes, subject-verbal concord, syntax errors, and essay length between learners' two texts, one without the use of GT and one with the use of GT, was used to provide an answer to the question. Only learners that were present during both writing sessions, made use of GT in some manner during the second writing session, and handed in screen recorded videos in both writing sessions were included in the analysis. Learner stimulated recall interviews also provided insight into how learners think whilst they are writing, and the teacher interviews provided background information as to why learners use GT as they do.

Firstly, looking at syntax errors, there was a mean reduction of errors for all performance-based groups from the first writing session to the second. Whether the mean decrease of essay length in each performance group from the first to the second writing session was the main cause for the decrease in syntax errors or if it was the use of Google Translate is difficult to comment on. Even if there is some uncertainty that GT was the main factor behind the decrease in syntax errors, Gianetti's (2016) study obtained similar results in that learners produced writing with fewer syntactic errors. More research would be useful to provide more empirical evidence on the matter. However, as all performance groups had a decrease in syntax errors from the first writing session to the second one and that the t-test determined that AP learners' reduction in syntax errors was statistically significant, GT might have played some part in the results, despite the translation tool mostly being used for word-level translations.

Subject-verbal concord errors had a negligible difference between the first and writing session (see section 4.4.). A mean error increase of such a small margin, indicates that GT has a neutral effect on learners' written production skills related to this category, neither hindering them nor helping them. Reviewing screen recorded videos and texts from both writing sessions, learners' texts did not contain many sentences with subject-verbal concord errors. This is in contrast to Nygaard's (2019) findings that L1 Norwegian L2 English learners struggle with this. However, given the relatively small sample size, this needs further study. Additionally, Google Docs can correct mistakes related to subject-verbal concord errors, thus the completed scripts would not be representative of what they had originally written. This is addressed further in section 5.4.

A paired t-test on the AP learners' texts found that there was a statistically significant difference in the mean number of spelling mistakes. However, for the other two performance groups, there was no statistical difference in the mean number of spelling mistakes. Reviewing the screen recorded videos, it is unsure as to why AP learners' texts had a significant decrease in spelling mistakes in the second writing session. It does seem, though, that Google Translate did not play a large role in helping learners reduce their number of spelling mistakes as Google Docs handled most errors related to that category (see section 4.5. and section 5.4).

From the first to the second writing session, mean CEFR-levels dropped for LP and AP learners whilst they stayed exactly the same for HP learners. One possibility behind these results is the two topics given to learners might have not interested them, supported by the learners' comments during the SRIs, hindering them in using their vocabulary to their full potential, which then lead to scripts consisting of vocabulary with fewer words considered advanced. If this study were to be run again multiple times in all three classes with a variety of topics to write about, the CEFR-level means might have changed.

The mean essay length of all performance groups did have a decline from the first writing session to the second. Based on comments from learners during the stimulated recall interviews, the reasoning for this most likely stems from the topic given to them to write about, not the fact that they used Google Translate instead of a dictionary to aid in EFL writing.

To conclude this section, results from the current study, SRIs and teacher interviews included, indicate that Google Translate does not have a large impact on the quality of learners' written texts, neither positive nor negative. The one exception was the category

number sentences with minimum one syntax error, which did see a decrease from the first to the second writing session for all performance groups and a statistically significant decrease for AP learners. These results align considerably with Fredholm's (2014, p. 100) results which showed that the use of GT did not affect writing performance in any decisive way, neither improving nor giving worse results than if the tool had not been used. Although GT has the possibility to provide an output which reaches university acceptance levels if given input of similar quality, most learners aged 10-11 have not reached a vocabulary level in their L1 to provide GT with such input. Norwegian learners also seem to take seriously the suggestions made by their teachers of not using GT to translate large amounts of texts as it can lead to a reduced learning outcome (Stapleton & Leung, 2019). If learners were to translate their whole texts from Norwegian to English using Google Translate, it would have been interesting to see how the results would have looked compared to how they did in this study. Based on previous research, the learners' texts might have performed better had they translated their whole texts from Norwegian to English using GT, seeing as those were the results that Stapleton and Leung (2019) achieved in their study. These were Chinese university students though, not Norwegian primary school learners, and the comparison might be unfair as some of the learners in the current study still struggle to write accurately in their L1. Nevertheless, although the quality of their texts might have been superior had they translated their whole texts from Norwegian to English, their learning outcome would most likely have decreased significantly, which is not beneficial when learning a language. Connecting this with the grammar-translation method, learners might have experienced a positive learning outcome from translating their whole texts from L1 to L2 had they worked further with the texts and compared similarities and differences between the two texts related to language structure and vocabulary (Stern, 1992, as cited in Mart, 2013). However, as the current study looked at the effects GT has on the quality of learners' texts, and not the effects GT has on learners' learning outcomes, it is difficult to comment on the topic. Additionally, comparing individual words and phrases between languages should be familiar to the learners, but to be critical to translations between two languages of whole texts may be beyond the abilities of year 7 learners.

5.4. DISCUSSION ON LEARNERS' DEPENDENCY ON GOOGLE DOCS' CORRECTION TOOLS

Once it became increasingly clear that learners seemed to be more dependent on the correction tools provided by Google Docs than Google Translate, another analysis was

undertaken to document how dependent the different performance groups were on these correction tools and document how they used these correction tools to aid the quality of their texts. Although the existence of these correction tools is, in the author of this thesis' opinion, incredibly useful and helpful with writing more efficiently as well as improving the form and meaning of texts, for assessment purposes, their use is more controversial as it hinders teachers in knowing what their learners are capable of writing themselves and what knowledge they have of the language they are learning. For instance, if learners have just been taught how to conjugate the verb *to be,* and they worked with this newly acquired knowledge in Google Docs, how would a teacher be able to determine which learners truly understood how to conjugate the verb and which learners did not when Google Docs corrects incorrect conjugations of *to be*?

Five out of seven LP learners were classified as having little dependency on Google Docs related to the grammar checker in the first writing session where they were only allowed dictionaries. At first glance, these results might seem out of place as LP learners typically struggle with correct spelling and sentence structure. However, the reason as to why they had such little dependency on Google Docs' grammar checker feature was due to a lot of their texts making little to no sense as there were misspellings that were impossible to derive meaning from as well as several syntax errors within the same sentences. This meant that Google Docs had no suggestions to give as it could not derive the meaning of the content it analysed. On the other hand, once LP learners were given access to Google Translate in the second writing session, the dependency on Google Docs' grammar checker feature increased. This might possibly be due to the additional help the learners received from GT which meant that the sentences made enough sense for Google Docs to derive meaning from them and offer suggestions to correct these mistakes. The use of GT also seemed to help some LP learners make fewer spelling mistakes as the number of learners with a high dependency on the spelling mistake feature in the first writing sessions halved in the second writing session.

These findings are also in line with Garcia and Pena (2011) and Garcia (2010) whose studies reached the conclusion that LP learners benefitted from using GT as they communicated more and better compared to not using them. On the other hand, it is not surprising to see that most HP learners were not categorised as having a high dependency on Google Docs, as they are expected to master many aspects of the language, correct spelling of words and grammatical structure included.

Another observation made from the screen recordings related to Google Docs was that multiple learners at every performance level did not write the first-person singular pronoun *I* with a capitalised letter, instead relying on Google Docs to correct it for them. The two teachers were asked in their interview about this phenomenon, whether learners had not been previously taught to capitalise *I* despite being in year 7 of primary school. Both teachers were surprised to hear this as learners had been taught every year since they started learning English that they needed to capitalise the *I*. It could be argued that learners either do it out of laziness as they expect Google Docs to correct it for them or they have forgotten the rule. When asked about this occurrence in their SRIs, none of the learners could think of a specific reasoning as to why they failed to capitalise the *I* when using the pronoun. Interestingly, Google Docs did not correct this error automatically. Instead, the mistake was flagged by the grammar checker, which learners then had to manually accept, perhaps multiple times if they had written *I* without the capitalisation multiple times throughout their essays. A possible explanation as to why many learners failed to capitalise the *I*, is the interlanguage hypothesis (Aljumah, 2020). Learners might be used to the capitalisation system of the Norwegian language wherein only proper nouns and the beginning of a sentence are capitalised, whilst English has capitalisation for a few more categories, e.g. the pronoun *I*, days and months, and adjectives related to regions such as *a Norwegian man*. Whilst they are writing in English, they may be using the language rule that they are most familiar with, in this case the Norwegian, and therefore they decided to write *I* with a lowercase letter. Another possible explanation as to why this happened might be due to learners being used to their other devices auto capitalising the pronoun, an explanation which is supported by a study conducted by Wood (2014, p. 26).

It also seemed that with the exception of one learner's screen recording, learners are not aware of how the autocomplete function of Google Docs (Smart Compose) functions. Smart Compose is a feature of Google Docs which uses machine learning to offer further suggestions to what one is currently typing. Most learners were at least once offered a suggestion to autocomplete a phrase or sentence that they were currently typing, however only one learner accepted the suggestion. The learners in the stimulated recall interviews were asked if they had any knowledge regarding the Smart Compose feature, to which they all replied that they saw the faded suggestions in their Google Docs, but they did not know what it meant or how it worked. Although it is offered as a way to write documents faster and more easily, as learners had not received training on what it was or how to use it, it seemed to be more of a hindrance to them. This was seen in some of the learners' screen recordings as they

would instinctively start typing the suggestion they were offered by Smart Compose, even if it were wrong, before deleting it and writing what they were actually intending to write. However, overall, Smart Compose did offer learners more correct suggestions than incorrect ones, which could have saved learners time if they knew how to use the feature properly.

Learners using Google Docs to write their texts also helps explain why there was such a small percentage of errors made related to subject-verbal concord in both writing sessions and why the usage of GT did not help reduce the number of errors related to this category. Google Docs corrected countless errors amongst learners at all performance levels related to subject-verbal concord, underlining incorrect usage of, for example, was/were in blue for learners to approve the suggestion made by the word processor. It is believed that there would have been a large increase in errors related to this category if learners had to write by hand or if these Google Docs features were turned off prior to the writing sessions.

The spell checker and grammar checker features of Google Docs may not be a big concern when they are merely used a few times, that is to say when learners only make a few mistakes. However, it is when learners make multiple errors in every sentence they write that the concern starts to arise from an assessment point of view. Learners may produce a product that is not telling of what they are actually capable of writing themselves with no help. Reviewing the 64 screen recordings, it is also worrisome that most learners approve the suggestions made by Google Docs without any hesitation, almost as if they are doing it unconsciously. This is concerning because Google Docs is not perfect and not all the suggestions the word processor makes are necessary nor correct. Learners also do not learn the correct spelling nor, for example, the correct subject-verbal concord if they approve these suggestions without first studying the suggestions given. Learners' lack of critical assessment of the suggestions that Google Docs provides shows a low degree of mastery related to the two sub-skills found in the framework for digital skills: *use and understand*, and *find and process* (Utdanningsdirektoratet, 2017). Firstly, because they are not critically assessing their work processes, and secondly, because they are not critically assessing the information given to them by Google Docs. This ties in with Brown's (2000) statement that error production is a major part of language learning. Rather than teachers reacting to the observable thought process of the learners and giving appropriate feedback to the various errors being made, Google Docs' correction features fill some of the teacher's role by highlighting various errors made in learners' texts. However, for these correction features to prove useful at increasing learners' language knowledge, learners need to pay more attention to the suggestions being

made by Google Docs and become better at critically assessing these suggestions. This ties in with the Noticing Hypothesis which claims that L2 learners need to notice relevant material in the linguistic data afforded by the environment in order to learn aspects of the L2 (Ortega, 2009, p. 63). The more L2 learners notice, the more they learn, and in the context of the current study, learners need to notice the suggestions provided by Google Docs in order to achieve a greater learning outcome. Since it was not the primary focus of this thesis, no direct empirical data from the study can back up these claims. However, a study by Rimbar (2017) reported data showing that spell checkers help in eliminating surface errors, but that they have little influence on correcting the same errors on a cognitive level. A different study by Lin, Liu, and Paas (2017) concluded that spell checkers do help learners on a cognitive level to fix corrections, but it is heavily dependent on the effort spent on searching for the correct words. In both the current study and Rimbar's (2017), learners typically spent little to no effort searching for the correct word when a word was flagged as being spelt incorrect, instead accepting the first suggestion provided without checking that it was actually correct. Additionally, although most learners made use of these features with no hesitation, some learners, for unknown reasons, ignored a lot of the suggestions that were provided, mainly suggestions related to grammar. Unfortunately, it is unknown why this was the case. The focus of the live observations and SRIs was on learners' usage of GT. The analysis of learners' dependency on Google Docs became of interest after the SRIs had concluded; as such, insight into learners' thought processes surrounding the topic was not given, but it was nevertheless an interesting observation.

It felt somewhat ironic writing the current section related to the dependency that learners have on Google Docs, seeing as the author of this thesis made use of the exact same features multiple times throughout the writing of this thesis, often without considering the suggestions being offered, merely right-clicking and accepting the suggestions given. Without the auto-correction feature, grammar check feature, and spell check feature, this thesis would have contained substantially more errors when reading through the drafts of it compared to what it did with the help of these features. It is therefore understandable why most learners typically approve of the suggestions being made by their word processor of choice without first reading the suggestion as they seem to have such a large amount of trust in the correctness of the suggestions. This is supported by Wood (2014) who conducted a study on the iPhone's auto-correct feature. One of the questions Wood asked the participants in the study was how confident they felt in changes made by the auto-correct feature when it was used for different purposes. Results were mixed, but for most purposes, a majority of

participants were completely confident in the changes made by auto-correct and many of them did not realise the necessity of proofreading their work. However, it is vital for both the author of this thesis and learners that suggestions are reviewed before approval as these features do not always provide the correct suggestion or a necessary suggestion.

5.5. IMPLICATIONS FOR FUTURE TEACHING

The findings of this study have implications for future teaching. Firstly, based on the results of learners' current usage of Google Translate, it is clear that the translation tool has little effect on the quality of the product learners create, and it is evident that they have received little training on the many features that the tool provides. This was an unexpected outcome, as the opposite was anticipated, that GT would have a large impact on the quality of learner's texts and that they would have had a better understanding of the tool they use regularly to help in translating between Norwegian and English. The implications these results have for future teaching are that learners first and foremost need to be given more training on how they use Google Translate. As learners make use of the tool regularly in primary school, all the way up through university and later in life as well, there needs to be a larger emphasis given to how the tool functions and the features it has to offer. Jolley and Maimone's (2015, p. 195) study agrees with the suggestion, as the majority of both learners and teachers in their study believed that it would be helpful and beneficial to spend more time teaching learners strategies for maximising their effectiveness of using translation tools such as GT.

A bigger emphasis on teaching learners ethical and effective usage of Google Translate should be given as the tool is not going away from the FL classroom, which is supported by multiple other studies stating that teachers need to adapt to the technologies surrounding them and find ways of implementing them appropriately and pedagogically in the classroom (McCarthy, 2004; O'Neill, 2012; Stapleton & Leung, 2019; Tsai, 2020). However, all of this is dependent on the digital competence that the teacher possesses. The biggest hurdle behind the incorporation of pedagogical use of ICT is a lack of digital competence amongst teachers, which is why not only learners, but teachers as well need to be taught ethical and effective usage of ICT tools (Kunnskapsdepartementet, 2017a). Both Kelentrić et al. (2017) and Giannetti (2016) argue that teachers need to be provided with training to develop the competence required to integrate ICT tools into their pedagogical work.

Seeing that Norwegian learners' current usage of Google Translate is largely based on using it as a dictionary, translating at word-level and phrase-level, and not translating

larger pieces of texts at once, there is currently not a big concern that the translation tool is hindering learners in actively using the language knowledge they received in the FL classroom, at least in relation to the English language. Rather, it seems as though learners have taken the input they have received from their teachers seriously. They do not use the tools in unethical ways such as translating whole passages or texts, meaning that Google Translate is mostly used as a supplementary tool to language learning. However, for future teaching, the emphasis on reminding learners not to use the tool in unethical ways cannot go away, especially as translation tools become smarter every day, possibly increasing the temptations of using GT and other translation tools in ways that make learners not actively engage with the language which they are supposed to learn.

Some schools have decided to ban all usage of Google Translate. For instance, the teacher at the pilot study school mentioned that the lower secondary school in the area had decided to ban all usage of Google Translate. It seems as though some upper secondary schools are starting to do the same. This trend is both odd and concerning. It is odd because there seems to be a gradual shift amongst teachers' attitudes towards Google Translate for the better (Stapleton & Leung, 2019). It is also odd as the results from the current study shows that learners primarily only use Google Translate as a dictionary, translating at word-level and phrase-level. This is also in agreement with Jin and Deifell's (2013) study, with their additional comment that GT offers the advantage of instant translations, meaning learners can spend less time away from their texts to translate words.

The banning of Google Translate is concerning for two reasons. Firstly, if teachers are concerned with learners translating larger parts of their texts from their L1 to English, then banning the usage of GT and other translation tools is not the answer because learners already have the ability to translate their whole text from most languages to English using solely built-in features found in their word processor, such as Google Docs or Word. Secondly, if teachers are concerned that GT provides learners with a vocabulary and grammar which is not correspondent of what they can write by themselves, then the author of this thesis argues that this line of thought is hypocritical. Based on the results of this study, Norwegian learners' current usage of Google Translate has barely any effect on their written products. However, teachers feel it is acceptable for learners to write their essays using a word processor, which based on the results of this thesis, does appear to affect the quality of their work by providing corrections to spelling mistakes and suggestions related to incorrect grammar use. If learners are writing their essays writing in Google Docs, they are also allowed to make use of Smart

Compose, which can essentially do some of the writing for them. This feature uses machine learning, and it therefore has the possibility to become more prevalent in learners' writing as it may be able to deduce what the learners are writing about and provide suggestions more frequently. Furthermore, there is no current research which shows that Google Translate has a negative effect on the results of learners' writing, but there is a multitude of research suggesting that despite banning the use of such technology, learners will nevertheless make use of it. Instead, teachers have to learn to adapt to the technology surrounding them and implement appropriate measures for its use in the classroom (McCarthy, 2004; O'Neill, 2012; Stapleton & Leung, 2019; Tsai, 2020).

The use of Google Translate and other translators also provides learners with ample opportunities to understand the contents of whatever they are reading in an FL that they do not master. This means that learners are no longer hindered by a language in order to gain knowledge of a new subject matter through that language. This furthers the argument as to why Google Translate and other translation tools should not be banned from schools. They provide learners with enriched opportunities to not only learn new words and phrases of foreign languages, but also acquire the knowledge that was previously unavailable to them.

The pedagogical methods of integrating machine translation into the FL classroom are methods which will be useful when teaching learners about the benefits and drawbacks of machine translation in relation to FL learning. Teaching learners post-editing can be done in such a way that encourages them to be critical of the output they receive from GT, and to always confirm that the output received is correct by analysing it manually. That is not to say that the output they receive will be incorrect, on the contrary, with the current way learners are using Google Translate, most outputs they receive will most likely be correct, but it teaches them critical thinking, and not to blindly trust all suggestions given. This also relates to the grammar-translation method where it can be argued that the output given by Google Translate and the correction suggestions from Google Docs can function as both *the best writers* which learners can study and learn from if the suggestions given are correct (Hell, 2009, as cited in Mart, 2013). On the other hand, if learners notice incorrect or unnecessary outputs and correction from GT and Google Docs, it reinforces their appreciation of both L1 and L2 grammar and style (Somers, 2003). Lastly, by using machine translation as *a bad model*, it teaches learners the weaknesses and faults of Google Translate to enhance their critical thinking even further, and it shows them that translation tools can and will create incorrect outputs, which is why they should be used critically. Teaching learners about ethical

107

and effective usage of Google Translate is also supported by Giannetti (2016), who states that teaching learners strategic uses of translation tools equips them with both the tools to aid in developing their foreign language literacy and with linguistic skills to aid in their comprehension. Additional ICT tools and MT tools that may be of interest to pursue in future teaching can be found in Appendix 6.

Looking towards the future, how might translation tools alter the way in which FLs are taught in schools? There are no concrete answers to that question, but considering the rampant advancements being made to machine learning and technology as a whole, the current system of how learners are taught FLs might be overhauled in the future. Mundt and Groves (2015) ask a hypothetical question of why a learner would want to go to the effort and expense of learning an FL if they are able to produce an acceptable L2 text from their own L1 writing instantly with the use of machine translation? After all, Google Translate and other translation tools do have their weaknesses and might never achieve a one hundred percent perfection rate of translations, but they continually improve and inch closer to that goal every day. Even if they were to *only* achieve a success rate of ninety percent, why would learners feel the need to learn the written aspects of a language when they have a near perfect translation tool readily available in their pockets which offers translations of almost every language in the world? In a different study, Mundt and Groves (2016) offer another hypothetical question of what the future of FL learning might look like, asking if a freely available and effective technology is able to allow learners to bypass the very difficult and time-consuming process of learning a language, would it not be difficult to make the case that learners should ignore the technology and rather return to the classroom for slow and strenuous studying?

Although it is difficult to comment on definite changes that will be made to FL teaching in the future, one thing is certain, machine translation tools are here to stay, which is why it is worth considering how future curriculums might put a larger emphasis on content rather than form. As translation tools might be able to handle everything related to form, which is more or less objective, when translating from L1 to L2, content, on the other hand, is subjective, which is impossible for translation tools or other ICT tools to write. Furthermore, both Stapleton and Leung (2019), and Farzi (2016) show support of the changing landscape of L2 learning that machine translation brings with it. If the future were to ever move towards FL learning being redundant due to translation tools providing near perfect translations, the

focus would then become on teaching learners to be fluent in their L1 so that they can fully benefit from what translation tools have to offer.

On the topic of the future of machine translation, a testament of how quickly GT learns to adjust for people's mistakes, one learner that participated in the study inputted "slappeav", a misspelling of the Norwegian words "slappe av" into GT during the second writing session. The translator suggested that the learner may have intended to spell "slappe av", but she did not produce a correct output on her own. When the author of this thesis translated the same misspelling into GT a few weeks later, the translator suddenly produced the correct output *relax* instantly, whilst also still offering the suggestion to whether the words "slappe av" was the intended input. This is merely one example of how the translation tool continuously evolves to become smarter and handle not only misspellings, but also new combination of phrases that it might not have heard of. For example, GT does not, as of the writing of this thesis, offer the correct translation of "turvei" as its first output when translating from Norwegian to English. The default output it produces is "turvei", and the user has to click on the output to receive *walking trail*, the correct output. However, with enough training, there is a large probability that the default output that GT produces will be *walking trail* in the future.

6. CONCLUSION

6.1. SUMMARISING AND MAJOR FINDINGS

This thesis investigated how Norwegian learners use Google Translate to aid them in EFL writing along with the effects that their current usage of GT has on the quality of their written texts. There is currently a lack of research conducted related to the current topic internationally, and no similar studies have been conducted in Norway. This thesis has attempted to pave the way for future studies to be conducted on the topic of GT in Norway as the tool is prevalent in most classrooms in the country, which is why it is vital to understand its strengths, weaknesses, and the impact it has on learners.

A mixed methods sequential explanatory study was conducted to answer the two research questions created for this thesis, combining a quantitative study, in the form of a within-group quasi-experimental design, with a qualitative study, in the form of an explanatory case study design. The explanatory case study was employed to supply answers to the first research question of how Norwegian learners currently use GT in their EFL writing. This qualitative study collected data by studying learners' screen recorded videos while writing and through stimulated recall interviews. The quasi-experiment, on the other hand, was used to provide answers to the question of how learners' current usage of GT affects the quality of their written texts. This quantitative study collected data through learners' screen recorded videos while writing, learner stimulated recall interviews, and teacher interviews.

Three classes of year 7 learners situated within the same school participated in the study along with two of the classes' lead teachers. The study consisted of two writing sessions where learners were given similar topics in order to stay as consistent as possible. In the first writing session, learners were given the topic *My favourite place to be* to write about, and they had no access to translation tools capable of translating at more than word-level, instead having to rely on electronic dictionaries. In the second writing session, learners were given the topic *My second favourite place to be* to write about, and they were encouraged to use GT the same way that they would typically use the tool to help them in EFL writing. In total, 37 learners agreed to participate in the study, however, due to illnesses and other issues, 36 documents and 33 screen recorded videos were received by learners after the first writing session, whilst 33 documents and 31 screen recorded videos were received after the second

writing session. Further, three learners from each class participated in stimulated recall interviews after the second writing sessions had concluded, followed by interviews of the two teachers who had consented to partake in the study.

For the data analyses, learners were grouped based on their performance level, i.e. lower performing, average performing, and higher performing. This was done to determine how learners at different levels use the tool to aid them, and whether the use of GT affects the learners' texts differently. The screen recorded videos were firstly used to determine learners' usage of GT, which was accomplished by counting how many times it was utilised by learners to translate at word-level, phrase-level, sentence-level, longer pieces of texts, or for other purposes, along with analysing how learners utilised the tool if GT did not produce the output that they expected. The screen recorded videos along with the learners' completed texts from the two writing sessions were then assessed and compared against each other to determine changes in vocabulary, syntax errors, subject-verbal concord, essay length, and spelling mistakes to determine whether GT had any effect on the quality of their texts.

Firstly, results indicate that learners mostly use Google Translate for word-level translations and phrase-level translations when translating from Norwegian to English. These findings are also supported by statements that both learners and teachers gave in their interviews, learners stating that they typically only use the tool for word-level translations, whilst teachers stating that they encourage learners to mostly use GT as a dictionary. These results are also in line with previous studies (Jin & Deifell, 2013; Chandra & Yuyun, 2018).

Furthermore, there is a discrepancy between learners at different performance levels related to when they use GT to help translate a piece of text. After the variance in the number of learners in the different performance-based groupings was accounted for, results show that LP learners use the tool the most out of the three performance-based groups for translations done at word-level, phrase-level, and sentence-level. Only one LP learner out of all the participants that had a corresponding screen recording to their texts used GT to translate their entire text from Norwegian to English. AP learners used GT the second most to translate at word-level and sentence-level, whilst HP learners made use of GT the least out of the three groups for all categories except phrase-level translations. However, only four out of eleven HP learners made use of the translation tool during the second writing session which must be taken into consideration when interpreting the data.

After thoroughly comparing participating learners' two texts against one another, the first one where they used an electronic dictionary against the second one where they used GT

and receiving valuable data from SRIs and teacher interviews, results indicate that GT has a negligible effect on the quality of learners' written products in relation to the categories studied in this thesis. The one exception was syntax errors where the overall mean and AP learners' texts had significant improvements when using GT, whilst the other two groups had some improvements. It is difficult to comment on whether GT was the reason for the decrease in errors related to syntax errors as there was no other empirical data to reference, but more research would aid in strengthening or weakening these results. Further, these results are mostly in agreement with Fredholm's (2014) study which reached a nearly identical conclusion as the current study, stating that MT has an overall neutral effect compared to writing with no aids other than a dictionary as it did not affect the writing performance of learners negatively nor positively. Vocabulary and essay length did suffer slightly from the first to the second writing session for all performance groups with the exception of vocabulary for HP learners which saw no improvement nor deterioration. However, based on the SRIs, these decreases stemmed from the topic that learners were given to write about and were not the fault of using GT to assist in writing. The texts of AP learners also saw a significant improvement in the spelling mistakes category, however, the author of this thesis believes that the impact of spelling mistakes came mostly from Google Docs' correction tools, not GT. Lastly, none of the learners' texts nor the overall mean saw a statistically significant change related to subject-verbal concord errors from the first to the second writing session.

The biggest surprise to come out of this study was the realisation of seeing how dependent learners at all performance levels are on the correction tools offered by word processors, in this case Google Docs. With the combined help of Google Docs' auto-correct feature, spell checker, and grammar checker, learners were able to produce texts of a higher quality than they would have been on their own. These three features corrected everything from learners' spelling mistakes to rearranging whole sentences that were incorrectly written, and constantly corrected subject-verbal concord errors that learners made in the writing of both their texts. As teachers have previously been concerned by learners' usage of GT, thinking that the tool replaces language learning rather than supplementing it, the author of this thesis argues that teachers should instead be more concerned by how word processors are replacing language learning. Based on the screen recordings, learners instinctively accept any suggestion and error correction that the word processor offers without first studying the suggestions offered nor double checking that the suggestions are correct.

6.2. LIMITATIONS AND DELIMITATIONS OF THE STUDY

Initially, the plan was to conduct research at two different schools as to have more impartial results. By conducting research at different schools, one is able to better determine if there are any variables that might affect the results. However, due to this study being conducted during the COVID-19 pandemic, it was decided that research would only be conducted at one school. This was due to safety concerns as the author of this thesis did not feel comfortable travelling more than necessary, and minimising exposure to other people wherever possible was in everyone's best interest. To compensate for only conducting research at one school, it was decided to conduct the research in three different classes that had different English teachers. Another consideration was that the learners in the three different classes did not share the same English teacher in previous years. These considerations were put in place to create a comparable scenario of conducting research at different schools.

Another potential limitation was learners not charging their Chromebooks before the writing sessions took place. To address this, there was good communication to the students and an exception to the normal school policy was made for the duration of the study for learners to charge their computers at school.

A delimitation put forth by the author of this thesis was the number of participants asked to participate in the study. As this was a master thesis and taking into consideration the vast amounts of data that needed to be analysed to answer the two research questions, it was decided that three classes of EFL learners would be acceptable. Having one individual analyse over 40 texts and over 30 hours' worth of screen recorded videos, comparing them against each other, and entering all the data into spreadsheets requires a lot of time and effort, and as such, involving more EFL learners in the study was not feasible. It can also be considered a potential risk as there was only one person analysing the complete data set.

Two weaknesses related to qualitative research that affected this study in particular, were the small number of participants for the teacher interviews and the learner stimulated recall interviews, in this case only two teachers and nine learners. The other weakness was the difficulty of assessing which findings were of general importance and which were simply idiosyncratic to a particular case (Dörnyei, 2007, p. 42).

Furthermore, in all three stimulated recall interviews (SRIs), the participants were often unsure or did not remember why they performed the actions that they did, even when the interviews were conducted as close to the end of the writing sessions as possible. It was

also difficult at times to elicit information from the learners of specific events that transpired during the writing sessions as the learners did not usually think through their actions whilst they performed them, as told by the learners themselves in the SRIs. To add to this, many learners in the SRIs also expressed that they wrote as much as they did in the two writing sessions because of the similarity of the two topics given. This is something to consider in relation to the data comparing text length as there is a possibility that learners would write more or less had the topics differed more. These comments are an argument that both of the writing sessions should have been piloted.

It is also necessary to assert that as this study was only conducted three times, it is difficult to state with certainty how learners use Google Translate to aid them with translation between their L1 and English, but this thesis provides good indications. Furthermore, this study only focused on how learners made use of the translator during writing sessions. Therefore, there is a potential probability that learners do not make use of the translation tool the same way in other settings such as for homework purposes, take-home assignments, or for other school related purposes that do not entail them being observed by others. A majority of the learners that participated in the SRIs stated that they mostly used GT the same way for other purposes as they did in the writing session, but with no empirical evidence to back up those statements, a conclusion cannot be reached on that matter. Although results shown in various tables accounted for the different group sizes based on performance level, it is not unreasonable to believe that these results might have had a different outcome had there been more learners in the study and an equal number of learners in each performance group.

6.3. SUGGESTIONS FOR FUTURE RESEARCH

There are many topics related to the current study that would be both interesting and highly relevant to pursue for future research, some of which will be mentioned here. As research related to the current study is scarce, a lot of it predating 2016 which is when Google Translate's algorithms were completely overhauled, and even non-existent in a Norwegian context, more research related to the topic of GT in the EFL classroom would be invaluable, both on how learners use the translator and the effects it has on their EFL writing. As there already exists some studies internationally on the usage and effects of GT, the main focus of future research should be to fill the knowledge gaps of how the tools are used in a Norwegian context, if possible. Similar studies that extend over a longer period of time would also be useful to pursue to obtain a better grasp of whether GT helps learners obtain new vocabulary.

Further, based on comments by learners in the stimulated recall interviews, they believed that they would have written more with the use of Google Translate compared to a dictionary if they had been only given one topic. Thus, it would be interesting to undertake a quasi-experimental study with a between-subjects design to determine the effects GT has on learners' EFL texts.

One of the teachers interviewed commented that many learners who struggle in written English also struggle in written Norwegian. This sparked the idea of pursuing research related to how learners with dyslexia or learners who just struggle with the written part of English and Norwegian would cope if instead of writing their input into the translation tool, they instead made use of the *translate by voice* feature of Google Translate where learners can use their microphone to say the words, phrases, and sentences that they wish to translate from Norwegian to English.

It needs to be stated that the results collected in the current study should only be interpreted in relation to translation of text between Norwegian and English. Learners are typically far more competent in English than they are in other foreign languages that are being taught at lower secondary school such as Spanish, French, and German. Thus, the probability of learners using GT and other translation tools differently in other foreign languages compared to English is high. Therefore, future research should also pursue how learners use Google Translate to aid in the writing of foreign languages other than English and how the translator affects their written texts.

6.4. FINAL STATEMENT

This thesis has presented valuable empirical data to a topic for which there is no prior research, at least within a Norwegian context. This study indicates that there is a need for more research on GT and other MT tools in relation to the FL classroom. To conclude, the author of this thesis would like to advocate the idea that The Norwegian Directorate for Education and schools should invest more time and resources into CALL. Additionally, a larger part of the curriculum related to language learning should focus on the implementation of teaching L2 learners the use of machine translation tools and other software that can strengthen their language learning.

7. BIBLIOGRAPHY

Aksnes, V. (2018). *Google Translate: Friend or Foe? An exploration of the use and attitudes to the use of Google Translate among teachers and pupils in two Norwegian upper secondary schools* [Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway]. NTNU Open. https://ntnuopen.ntnu-xmlui/bitstream/handle/11250/2568637/Masteroppgave%2c%20Vilde%20Evensen%20Aksnes.pdf?sequence=1&isAllowed=y

Albano, A. D. (2016). Chapter 6: Interrater Reliability. *Introduction to Educational and Psychological Measurement Course Notes*. University of Nebraska-Lincoln. https://cehs01.unl.edu/aalbano/intromeasurement/mainch7.html#x9-970006

Alhaisoni, E., & Alhaysony, M. (2017). An Investigation of Saudi EFL University Students' Attitudes towards the Use of Google Translate. *International Journal of English Language Education, 5*(1), 72-82. https://doi.org/10.5296/ijele.v5i1.10696

Aljumah, F. H. (2020). Second Language Acquisition: A Framework and Historical Background on Its Research. *English Language Teaching*, *13*(8), 200-207. https://doi.org/10.5539/elt.v13n8p200

Allen, J. (2003). Post-editing. In Somers, H. L. (Ed.), *Computers and translation: A translator's guide* (pp. 297-317). John Benjamins Pub. Co.

Al-Tuwayrish, K. (2016). An Evaluative Study of Machine Translation in the EFL Scenario of Saudi Arabia. *Advances in Language and Literary Studies, 7*(1), 5-10. http://www.journals.aiac.org.au/index.php/alls/article/view/1950/1759

Anderson, D. D. (1995) Machine translation as a tool in second language learning. *CALICO Journal, 13*(1), 68-97. https://journals.equinoxpub.com/CALICO/article/view/23406/19411

Bahri, H., & Mahadi, T. (2016). Google Translate as a Supplementary Tool for Learning Malay: A Case Study at Universiti Sains Malaysia. *Advances in Language and Literary Studies*, *7*(3), 161-167. http://www.journals.aiac.org.au/index.php/alls/article/view/2305/2014

Beatty, K. (2010). *Teaching and Researching Computer-Assisted Language Learning*. (2nd ed.). Pearson Longman.

Belam, J. (2003). "Buying up to falling down": a deductive approach to teaching post-editing. *MT Summit IX Workshop on Teaching Translation Technologies and Tools*, 1-10. http://mt-archive.info/MTS-2003-Belam.pdf

Brown, H. D. (2000). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. (2nd ed.). Pearson Longman.

Cambridge English Language Assessment. (2015). *The Cambridge English Scale*. University of Cambridge. https://www.cambridgeenglish.org/images/167506-cambridge-english-scale-factsheet.pdf

Casanave, C. P. (2015). Case Studies. In Phakiti, A. & Paltridge, B. (Eds.), *Research Methods in Applied Linguistics: A Practical Resource*. (pp. 119-136). Bloomsbury Publishing Plc.

Chandra, S., & Yuyun, I. (2018). The Use of Google Translate in EFL Essay Writing. *LLT Journal: A Journal on Language and Language Teaching, 21*(2), 228-238. https://doi.org/10.24071/llt.2018.210212

Chang, S.-C. (2011). A Contrastive Study of Grammar Translation Method and Communicative Approach in Teaching English Grammar. *English Language Teaching*, *4*(2), 13-24. https://doi.org/10.5539/elt.v4n2p13

Chen, H.-J. H., & Yang, T.-Y. Y. (2013) The impact of adventure video games on foreign language learning and the perceptions of learners. *Interactive Learning Environments*, *21*(2), 129-141. https://doi.org/10.1080/10494820.2012.705851

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the Landscape: What is the Role of Machine Translation in Language Learning?. *@tic. revista d'innovació educative*, *10*, 108-121. https://doi.org/10.7203/attic.10.2228

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.). Routledge.

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Council of Europe.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Sage.

Curdt-Christiansen, X. L. (2020). Observations and field notes: Recording lived experiences. In McKinley, J. & Rose, H. (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 336-347). Routledge.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics Quantitative, Qualitative, and Mixed Methodologies*. Oxford University Press.

Drew, I., & Sørheim, B. (2016). *English Teaching Strategies Methods for English teachers of 10 to 16-year-olds*. (3rd ed.). Det Norske Samlaget.

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*, *51*, 779-795. https://doi.org/10.1111/flan.12366

Eisa, S. A. H. (2020). "The Pros and Cons of the Grammar Translation Method on the Performance of Saudi EFL Learners" (A case study of the third class in Saudi secondary schools in Northern boarders). *The Arab Journal of Academic Publishing, 17*, 381-392. https://www.ajsp.net/research/The%20Pros%20and%20Cons.pdf

Farzi, R. (2016). *Taming Translation Technology for L2 Writing: Documenting the Use of Free Online Translation Tools by ESL Students in a Writing Course* [Doctoral thesis, The University of Ottawa, Ottawa, Canada]. uO Research. https://doi.org/10.20381/ruor-5717

Fjørtoft, S. W., Thun, S., & Buvik, M. P. (2019). *Monitor 2019 - En deskriptiv kartlegging av digital tilstand i norske skoler og barnehager*. SINTEF Digital. http://hdl.handle.net/11250/2626335

Flower, L. (1979). Writer-Based Prose: A Cognitive Basis for Problems in Writing. *College English*, *41*(1), 19-37. https://doi.org/10.2307/376357

Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., Sánchez-Martínez, F., Ramírez-Sánchez, G., & Tyers, F. M. (2011). Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation, 25*(2), 127-144. https://doi.org/10.1007/s10590-011-9090-0

Fredholm, K. (2014). Effects of online translation on morphosyntactic and lexical-pragmatic accuracy in essay writing in Spanish as a foreign language. *CALL Design: Principles and Practice - Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands,* 96-101. https://doi.org/10.14705/rpnet.2014.000201

Garcia, I. (2010). Can Machine Translation Help the Language Learner? *International Conference "ICT for Language Learning" 3rd edition.* PIXEL. https://conference.pixel-online.net/conferences/ICT4LL2010/common/download/Proceedings_pdf/TRAD02-Garcia.pdf

Garcia, I. (2011). Translating by post-editing: Is it the way forward? *Machine Translation, 25*(3), 217-237. https://doi.org/10.1007/s10590-011-9115-8

Garcia, I., & Pena M. I. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, *24*(5), 471-487. https://doi.org/10.1080/09588221.2011.582687

Gass, S. (2015). Experimental Research. In Phakiti, A. & Paltridge, B. (Eds.), *Research Methods in Applied Linguistics: A Practical Resource*. (Chapter 6). Bloomsbury Publishing Plc.

Giannetti, T. R. (2016). *Google Translate as a Resource for Writing*. [Master's thesis, St. John Fisher College, New York, USA]. Fisher Digital Publications. https://fisherpub.sjfc.edu/cgi/viewcontent.cgi?article=1358&context=education_ETD_masters

Goldberg, Y. (2017). *Neural Networks Methods for Natural Language Processing*. Morgan & Claypool Publishers.

Golonka, E. M., Bowles, A. R, Frank, V. M, Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their

effectiveness. *Computer Assisted Language Learning*, *27*(1), 70-105.
https://doi.org/10.1080/09588221.2012.700315

Gomez, R. J. (1996). Process versus Product Writing with Limited English Proficient
Students. *Bilingual Research Journal, 20*(2), 209-33.
https://ncela.ed.gov/files/rcd/BE021579/Process_Versus_Product.pdf

Google Translate. (n.d.). *Translate*. https://translate.google.com/intl/en/about/languages/

Groves, D., & Way, A. (2006). Hybrid data-driven models of machine translation. *Machine
Translation*, *19*, 301–323. https://doi.org/10.1007/s10590-006-9015-5

Gudmundsdottir, G. B., Loftsgarden, M., & Ottestad, G. (2014). *Newly qualified teachers
Professional digital competence and experiences with ICT in teacher education*.
https://www.udir.no/contentassets/fb484acf74dc4eb3bea0f16bf9995641/nulrapport_en
glish.pdf

Holliday, A. (2015). Qualitative Research and Analysis. In Phakiti, A. & Paltridge, B. (Eds.),
Research Methods in Applied Linguistics: A Practical Resource. (Chapter 3).
Bloomsbury Publishing Plc.

Houcine, S. (2011). The effects of ICT on learning/teaching in a foreign language.
International Conference "ICT for Language Learning" 4th edition. PIXEL.
https://conference.pixel-
online.net/conferences/ICT4LL2011/common/download/Paper_pdf/IBL69-437-FP-
Houcine-ICT4LL2011.pdf

Hyland, K. (2003). *Second Language Writing*. Cambridge University Press.

Imsen, G. (2010). *Elevens verden Innføring i pedagogisk psykologi* (4th ed.).
Universitetsforlaget.

Isisag, K. U. (2012). The positive effects of integrating ICT in foreign language teaching.
International Conference "ICT for Language Learning" 5th edition. PIXEL.
https://conference.pixel-
online.net/conferences/ICT4LL2012/common/download/Paper_pdf/235-IBT107-FP-
Isisag-ICT2012.pdf

Jiang, X., Rollinson, J., & Blanco, C. (2020, August 19). How well does Duolingo work? *Duolingo Blog*. https://blog.duolingo.com/how-well-does-duolingo-work/

Jin, L., & Deifell, E. (2013). Foreign Language Learners' Use and Perception of Online Dictionaries: A Survey Study. *Journal of Online Learning and Teaching*, *9*(4), 515-533. https://jolt.merlot.org/vol9no4/jin_1213.pdf

Jolley, J. R., & Maimone, L. (2015). Free online machine translation: use and perceptions by Spanish students and instructors. In Moeller, A. J. (Ed.), *Learn Languages, Explore Cultures, Transform Lives. Selected Papers from the 2015 Central States Conference on the Teaching of Foreign Languages* (pp. 181-200). Robert M. Terry.

Kelentrić, M., Helland, K., & Arstorp, A.-T. (2017). *Professional Digital Competence Framework for Teachers.* The Norwegian Centre for ICT in Education. https://www.udir.no/globalassets/filer/in-english/pfdk_framework_en_low2.pdf

Kliffer, M. D. (2005). An experiment in MT post-editing by a class of intermediate/advanced French majors. *Proceedings EAMT 10th Annual Conference*, 160-165. http://www.mt-archive.info/EAMT-2005-Kliffer.pdf

Kliffer, M. D. (2008). Post-editing machine translation as an FSL exercise. *Porta Linguarum*, *9*, 53–67. http://www.ugr.es/~portalin/articulos/PL_numero9/4%20Michel%20D.pdf

Klimova, B., & Kacet, J. (2017). Efficacy of Computer Games on Language Learning. *The Turkish Online Journal of Educational Technology*, *16*(4), 19-26. https://files.eric.ed.gov/fulltext/EJ1160637.pdf

Koda, K. (2007). Reading and Language Learning: Crosslinguistic Constraints on Second Language Reading Development. *Language Learning, 57,* 1-44. https://doi.org/10.1111/0023-8333.101997010-i1

Koehn P., Och, F. J., & Marcu, D. (2003). Statistical Phrase-Based Translation. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 127-133. https://www.aclweb.org/anthology/N03-1017

Kunnskapsdepartementet. (2017a). *Framtid, fornyelse og digitalisering Digitaliseringsstragtegi for grunnopplæringen 2017-2021.* https://www.regjeringen.no/contentassets/dc02a65c18a7464db394766247e5f5fc/kd_framtid_fornyelse_digitalisering_nett.pdf

Kunnskapsdepartementet. (2017b). *Overordnet del - verdier og prinsipper for grunnopplæringen.* https://www.regjeringen.no/contentassets/53d21ea2bc3a4202b86b83cfe82da93e/overordnet-del---verdier-og-prinsipper-for-grunnopplaringen.pdf

Lars-Freeman, D., & Anderson, M. (2011). *Techniques & Principles in Language Teaching.* Oxford University Press.

Lee, S.-M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, *33*(3), 157-175. https://doi.org/10.1080/09588221.2018.1553186

Lin, P.-H., Liu, T.-C., & Paas, F. (2017). Effects of Spell Checkers on English as a Second Language Students' Incidental Spelling Learning: A Cognitive Load Perspective. *Reading and Writing: An Interdisciplinary Journal*, *30*(7), 1501–1525. https://doi.org/10.1007/s11145-017-9734-4

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL: The Journal of EUROCALL, 31*(3), 293-311. https://doi.org/10.1017/S0958344019000065

Mart, C. T. (2013). The Grammar-Translation Method and the Use of Translation to Facilitate Learning in ESL Classes. *Journal of Advances in English Language Teaching 2013*, *1*(4), 103-105. http://european-science.com/jaelt/article/view/281/pdf#

McCarthy, B. (2004). Does Online Machine Translation Spell the End of Take-Home Translation Assignments?. *CALL-EJ Online*, *6*(1). http://callej.org/journal/6-1/mccarthy.html

McDowell, B., & Settles, B. (2019, September 23). The Duolingo CEFR Checker: An AI tool for adapting learning content. *Duolingo Blog.* https://blog.duolingo.com/the-duolingo-cefr-checker-an-ai-tool-for-adapting-learning-content/

McGarr, O., & McDonagh, A. (2019). Digital Competence in Teacher Education. *Output 1 of the Erasmus+ funded Developing Student Teachers' Digital Competence (DICTE) project*. https://dicte.oslomet.no/wp-content/uploads/2019/03/DICTE-Digital-Competence-in-Teacher-Ed.-literature-review.pdf

Medvedev, G. (2016). Google translate in teaching English. *Journal of Teaching English for Specific and Academic Purposes*, *4*(1), 181-193. http://espeap.junis.ni.ac.rs/index.php/espeap/article/view/318/221

Meld. St. 28 (2015-2016). *Fag – Fordypning – Forståelse — En fornyelse av Kunnskapsløftet*. Kunnskapsdepartementet. https://www.regjeringen.no/no/dokumenter/meld.-st.-28-20152016/id2483955/?

Microsoft. (2020). *Machine Translation.* https://www.microsoft.com/en-us/translator/business/machine-translation/

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. In Richards B., Daller M. H., Malvern D. D., Meara P., Milton J. & Treffers-Daller, J. (Eds.), *Vocabulary Studies in First and Second Language Acquisition* (pp. 194-211). Palgrave Macmillan. https://doi.org/10.1057/9780230242258_12

Milton, J., & Alexiou, T. (2020). Vocabulary Size Assessment: Assessing the Vocabulary Needs of Learners in Relation to Their CEFR Goals. In Dodigovic, M. & Agustín-Llach, M. P. (Eds.), *Vocabulary in Curriculum Planning: Needs, Strategies and Tools* (pp. 9-27). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-48663-1_2

Moore, J. (2020). *The Oxford 3000 and Oxford 5000: The Most Important Words to Learn in English*. Oxford University Press.

Mundt, K., & Groves, M. (2015). Friend or Foe? Google Translate in Language for Academic Purposes. *English for Specific Purposes*, *37*, 112-121. https://doi.org/10.1016/j.esp.2014.09.001

Mundt, K., & Groves, M. (2016). A double-edged sword: the merits and the policy implications of Google Translate in higher education. *European Journal of Higher Education*, *6*(4), 387-401. https://doi.org/10.1080/21568235.2016.1172248

Niño, A. (2008). Evaluating the use of machine translation postediting in the foreign language class. *Computer Assisted Language Learning*, *21*(1), 29-49. https://doi.org/10.1080/09588220701865482

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, *21*(2), 241-258. https://doi.org/10.1017/S0958344009000172

Nygaard, M. O. (2019). *The effects of grammar instruction when learning L2 English subject-verb agreement: An investigation of L1 Norwegian learners' acquisition of L2 English* [Master's thesis, The Arctic University of Norway, Tromsø, Norway]. UiT Munin. https://munin.uit.no/bitstream/handle/10037/16101/thesis.pdf?sequence=2&isAllowed=y

Och, F. J. (2006, April 28). Statistical machine translation live. *Google AI Blog*. https://ai.googleblog.com/2006/04/statistical-machine-translation-live.html

OECD. (2015). *Students, Computers and Learning: Making the Connection*. PISA & OECD Publishing. https://doi.org/10.1787/9789264239555-en

O'Neill, E. (2012). *The effect of online translators on L2 writing in French* [Doctoral thesis, University of Illinois, Urbana, USA]. IDEALS. http://hdl.handle.net/2142/34317

Ortega, L. (2014). *Understanding second language acquisition*. Routledge.

Osanloo, N., & Kolahi, S. (2016). The Comparative Effect of Direct and Indirect Corrective Feedback in Process-Based Vs. Product-Based Writing Instruction on EFL Learners' Writing Performance. *Journal of Applied Linguistics*, *9*, 84-103. https://pdfs.semanticscholar.org/b75d/39babc7b4d7ee7b2869de2fa19b99106ba1d.pdf

Padurean, A., & Margan, M. (2009). Foreign Language Teaching Via ICT. *Revista de Informatică Socială*, *7*(12), 97-101.

https://www.researchgate.net/publication/242555379_Foreign_Language_Teaching_Via_ICT

Phakiti, A. (2015). Quantitative Research and Analysis. In Phakiti, A. & Paltridge, B. (Eds.), *Research Methods in Applied Linguistics: A Practical Resource.* (Chapter 2). Bloomsbury Publishing Plc.

Phakiti, A., & Paltridge, B. (2015). Approaches and Methods in Applied Linguistics Research. In Phakiti, A. & Paltridge, B. (Eds.), *Research Methods in Applied Linguistics: A Practical Resource* (pp. 5-25). Bloomsbury Publishing Plc.

Poibeau, T. (2017). *Machine Translation*. The MIT Press.

Riazi, A. M. (2016). *The Routledge Encyclopedia of Research Methods in Applied Linguistics: Quantitative, qualitative, and mixed-methods research.* Routledge.

Rimbar, H. (2017). The Influence of Spell-Checkers on Students'ability to Generate Repairs of Spelling Errors. *Journal of Nusantara Studies (JONUS)*, *2*(1), 1-12. https://doi.org/10.24200/jonus.vol2iss1pp1-12

Rogers, J., & Révész, A. (2020). Experimental and quasi-experimental designs. In McKinley, J. & Rose, H. (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 133-143). Routledge.

Sanchez, H. S., & Grimshaw, T. (2020). Stimulated recall. In McKinley, J. & Rose, H. (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 312-323). Routledge.

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychon Bull Rev, 22*, 1614-1629. https://doi.org/10.3758/s13423-015-0835-2

Schuster, M., Johnson, M., & Thorat, N. (2016, November 22). Zero-Shot Translation with Google's Multilingual Neural Machine Translation System. *Google AI Blog*. https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Shei, C.-C. (2002). Teaching MT through pre-editing: Three case studies. *Proceedings of the 6th EAMT Workshop Teaching Machine Translation*, 89-98. http://www.mt-archive.info/00/EAMT-2002-Shei.pdf

Somers, H. L. (2003). Machine translation in the classroom. In Somers, H. L. (Ed.), *Computers and translation: A translator's guide* (pp. 319-340). John Benjamins Pub. Co.

Stapleton, P., & Leung, K. K. B. (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes, 56,* 18-34. https://doi.org/10.1016/j.esp.2019.07.001

Steele, V. (n.d.). *Product and process writing: A comparison*. TeachingEnglish. https://www.teachingenglish.org.uk/article/product-process-writing-a-comparison

Systran. (2016, October 17). How does Neural Machine Translation work? *SYSTRAN Blog*. https://blog.systransoft.com/how-does-neural-machine-translation-work/

Søby, M. (2018). *Norway Country Report on ICT in Education*. http://www.eun.org/documents/411753/839549/Country+Report+Norway+2018_v2.pdf/e8c32816-d56e-4080-8154-d2f6ca6f9961

The Education Act. (1998). *Act relating to Primary and Secondary Education and Training* (LOV-1998-07-17-61). https://lovdata.no/dokument/NLE/lov/1998-07-17-61

Tsai, S.-C. (2020). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1799412

Turovsky, B. (2016a, November 15). Found in translation: More accurate, fluent sentences in Google Translate. *Google Translate Blog*. https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/

Turovsky, B. (2016b, April 28). Ten years of Google Translate. *Google Translate Blog*. https://www.blog.google/products/translate/ten-years-of-google-translate/

Utdanningsdirektoratet. (2017). *Rammeverk for grunnleggende ferdigheter*.
https://www.udir.no/laring-og-trivsel/rammeverk/rammeverk-for-grunnleggende-
ferdigheter/2.1-digitale-ferdigheter/

Utdanningsdirektoratet. (2020). *Curriculum in English* (ENG01-04).
https://www.udir.no/lk20/eng01-04?lang=eng

Way, A., & Gough, N. (2005). Comparing example-based and statistical machine
translation. *Natural Language Engineering, 11*(3), 295-309.
https://doi.org/10.1017/S1351324905003888

Williams, P., Sennrich, R., Post, M., & Koehn, P. (2016). *Syntax-based Statistical Machine
Translation*. Morgan & Claypool Publishers.

Wood, N. (2014). *Autocorrect Awareness: Categorizing Autocorrect Changes and Measuring
Authorial Perceptions* [Undergraduate honours thesis] Florida State University
Libraries. https://diginole.lib.fsu.edu/islandora/object/fsu:204751/datastream/PDF/

Yandex. (n.d.) *Machine Translation*. https://yandex.com/company/technologies/translation/

Zens, R., Och, F. J., & Ney, H. (2002). Phrase-Based Statistical Machine Translation. *Annual
Conference on Artificial Intelligence*, 18-32. https://doi.org/10.1007/3-540-45751-8_2

LIST OF WEB RESOURCES REFERENCED

Apertium - https://apertium.org/

Bad News - https://www.getbadnews.com/

Dinordbok - https://www.dinordbok.no/norsk-engelsk/

Duolingo - https://www.duolingo.com/

Duolingo CEFR checker - https://cefr.duolingo.com/

Factitious- http://factitious.augamestudio.com/

Google Translate - https://translate.google.com/

Grammarly - https://www.grammarly.com/

IntoWords - https://vitec-mv.com/no/produkter/intowords/

Keep Talking and Nobody Explodes - https://keeptalkinggame.com/

Languagenut - https://www.languagenut.com/

Lingua - https://lingua.com/english/reading/

Microsoft Bing Translator - https://www.bing.com/translator

Oxford 3000 and 5000 wordlist -

https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000

Yandex Translate - https://translate.yandex.com/

APPENDICES

APPENDIX 1 - NSD APPROVAL

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

A comparative study on the effects of using Google Translate in Norwegian EFL classrooms.

Referansenummer

184782

Registrert

08.09.2020 av Kristian Fellowes Haukelid - kf.haukelid@stud.uis.no

Behandlingsansvarlig institusjon

Universitetet i Stavanger / Fakultet for utdanningsvitenskap og humaniora / Institutt for kultur- og språkvitenskap

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Rebecca Anne Charboneau Stuvland, rebecca.a.stuvland@uis.no, tlf: 51831577

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Kristian Fellowes Haukelid, Kristian.fellowes@gmail.com, tlf: 90646487

Prosjektperiode

24.08.2020 - 20.06.2021

Status

28.10.2020 - Vurdert

Vurdering (1)

28.10.2020 - Vurdert

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet 28.10.2020 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:
https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 20.06.2021.

LOVLIG GRUNNLAG UTVALG 1
Prosjektet vil innhente samtykke fra foresatte til behandlingen av personopplysninger om barna/elevene. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som foresatte kan trekke tilbake. Barna/elevene vil også samtykke til deltakelse.

Lovlig grunnlag for behandlingen vil dermed være foresattes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

LOVLIG GRUNNLAG UTVALG 2
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Utvalg 2 består av lærere. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

• lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
• formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
• dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
• lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen som de registrerte og foresatte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert/foresatt tar kontakt om sine/barnets rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og eventuelt rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

130

Lykke til med prosjektet!

Kontaktperson hos NSD: Maren Urheim
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

131

Vil du delta i forskningsprosjektet

«Effektene av å bruke Google Translate i engelskfaget i norsk skole»?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å analysere hvordan elever i norsk skole bruker Google Translate som et verktøy og hvordan bruken av Google Translate i engelskfaget påvirker elevenes skriveevner. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål
Jeg ønsker å gjøre forskningsarbeid til masteroppgaven min om hvordan elever på mellomtrinnet bruker verktøyet Google Translate når de skriver i engelskfaget, og videre hvordan bruken av dette verktøyet påvirker deres skriveevne, om det har en positivt eller negativ virkning på produktet de skriver. Det er gjort svært lite forskning om hvordan Google Translate påvirker elevers skriveevner så jeg håper at forskningen min kan behjelpe bruken av verktøyet for fremtidig undervisning, både for meg selv og andre lærere.

Prosjektet vil finne sted over to ulike engelsktimer på to ulike dager. I prosjektet vil elevene skrive to korte stiler om et tema som er forhåndlaget av meg. Den første gangen de skal skrive vil de ikke få tilgang på Google Translate, men de kan bruke elektroniske eller fysiske ordbøker. Den andre gangen dere skal skrive en stil vil bruken av Google Translate være tillatt/oppmuntret, men ikke påkrevd. Informasjonen som blir hentet inn blir videoopptak av skjermene deres under skriveprosessene samt det ferdige dokumentet som de produserer. Med skjermopptak menes at et program på Chromebooken deres vil ta opp det som foregår på skjermen deres, webkamera og filming av elevene vil ikke skje. I tillegg ønsker jeg å intervjue to-tre tilfeldig valgte elever mens de skriver for å finne ut deres tanker mens de er i skriveprosessen og bruker verktøyet. Jeg ønsker også å få vite barnet ditt sin språklige bakgrunn fra læreren hans/hennes ettersom det kan hjelpe meg med å få en bedre forståelse av hvilken grad Google Translate har en påvirkning på skrieevnen til barnet ditt. For eksempel om barnet ditt bare bruker engelsk hjemme, vil det mest sannsynlig ha en påvirkning på nivået ditt barn behersker i engelsk som da kan forklare hvorfor barnet ditt ikke har en stor forskjell i skriveevne mellom de to skriveøktene.

Hvem er ansvarlig for forskningsprosjektet?
Universitetet i Stavanger er ansvarlig for prosjektet.

Hvorfor får barnet ditt spørsmål om å delta?
Utvalget ble trukket via at jeg tok kontakt med skolen som eleven din går på, og spurte om det var noen klasser på mellomtrinnet hvor jeg kunne gjennomføre et forskningsprosjekt som ikke ville ta mange timer ut av den ordinære undervisningen. Jeg fikk beskjed om at læreren til ditt barn på 7. trinn var mer enn villig til å stille opp med klassen sin. Det er læreren til din elev som har sendt ut dette informasjonsskrivet, jeg har ingen kontaktopplysninger til hverken elever eller foresatte.

Hva innebærer det for ditt barn å delta?
Hvis du velger å la ditt barn delta i prosjektet, innebærer det at barnet vil gjennomføre to 30 minutters skriveøkter hvor opptak av skjermene deres vil bli gjort. Disse skjermopptakene vil bli lastet ned på en kryptert harddisk, sammen med stilene som de skriver. Under selve skriveøktene vil jeg ikke notere

ned noe elevene foretar seg ved Chromebookene sine ettersom skjermopptakene samler inn den dataen som trengs. Dersom barnet også blir trukket ut til å bli intervjuet mens han/hun skriver, vil lydopptak bli gjort i tillegg som også blir lagret kryptert på en harddisk. Spørsmål som vil være på intervjuet inkluderer «Er det lettere for deg å skrive når du får lov å bruke Google Translate?» og «Hvorfor brukte du heller en ordbok i stedet for Google Translate for dette ordet?». Spørsmålene kan ha litt variasjon avhengig av hva eleven gjør når han/hun skriver, men det er kun slike type spørsmål som vil bil stilt.

Du/dere som forelder kan gjerne få innsyn i intervjuguiden på forhånd hvis det skulle være et behov for det, i så fall ta kontakt med meg. Ingen av intervjuspørsmålene vil på noen måte være inngripende, det vil kun bli stilt spørsmål i henhold til hvordan de bruker Google Translate som et verktøy.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Det samme gjelder også for barnet ditt. Dersom han/hun ikke ønsker å delta, har han/hun lov til å si nei når som helst i prosjektet, til og med selv om du som forelder har samtykket til at han/hun kan delta i forskningen. Barnet ditt kan trekke tilbake sin samtykke uten å oppgi grunn, og hans/hennes personopplysninger vil bli slettet fra prosjektet.

Å bli med eller ikke i forskningsprosjektet vil ikke påvirke ditt/deres forhold til skolen/lærer på noen måte. Det samme gjelder også for barnet deres.

Dersom du ikke ønsker at ditt barn deltar i forskningsarbeidet, vil barnet ditt fortsatt kunne delta i skriveøktene for å få trening i engelsk, men jeg vil da ikke innhente noe informasjon eller data om hva barnet ditt gjør. Filmopptak og intervju vil IKKE bli gjort i så fall. Læreren vil også ha noen engelskoppgaver tilgjengelige mens skriveøktene pågår hvis det skulle være behov for dem.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Resultatene av skjermopptakene og intervjuene vil bli anonymisert slik at din anonymitet vil være beholdt i det ferdige resultatet av masteroppgaven.

- Navn og personopplysninger vil også bli erstatt med en kode som lagres på en egen navneliste som blir lagret på en kryptert harddisk som adskilles fra øvrige data

- Resultatene av intervjuet og skjermopptakene vil kun bli benyttet til bruk av masteroppgaven.

- Tilgang til transkripsjonen av intervjuene samt resultatene av skjermopptakene vil være limitert til meg og min veileder.

- Endringer av betingelsene ovenfor vil kun skje dersom ny samtykke blir gitt

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?
Opplysningene anonymiseres mens dataen behandles og når prosjektet avsluttes/oppgaven er godkjent, som vil være juni 2021. Innen prosjektslutt vil all data som skjermopptak, lydopptak på intervju og dokumentene som elevene skrev bli destruert.

Dine rettigheter

Så lenge ditt barn kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om ditt barn, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om ditt barn,
- å få slettet personopplysninger om ditt barn, og
- å sende klage til Datatilsynet om behandlingen av ditt barns personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Universitetet i Stavanger har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:
- Universitet i Stavanger ved Rebecca Anne Charboneau Stuvland, rebecca.a.stuvland@uis.no, 51 83 15 77
- Kristian Fellowes Haukelid, Kristian.Fellowes@gmail.com, 906 46 487
- Vårt personvernombud: personvernombud@uis.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:
- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Rebecca Anne Charboneau Stuvland   Kristian Fellowes Haukelid
(veileder)   (student)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Effektene av å bruke Google Translate i engelskfaget i norsk skole?*, og har fått anledning til å stille spørsmål. Jeg samtykker til at mitt barn kan/sin:

- ☐ Delta i forskningsprosjektet hvor skjermopptak vil bli gjort
- ☐ Språklige bakgrunn kan bli gitt til forsker via læreren
- ☐ Stille til intervju

Jeg samtykker til at mitt barns opplysninger behandles frem til prosjektet er avsluttet.

--   --
(Foresattes samtykke/signatur på vegne av sitt barn, dato)   (Barnets navn, dato)

Vil du delta i forskningsprosjektet

Effektene av å bruke Google Translate i engelskfaget i norsk skole?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å analysere hvordan elever i norsk skole bruker Google Translate som et verktøy og hvordan bruken av Google Translate i engelskfaget påvirker elevenes skriveevner. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Jeg ønsker å gjøre forskningsarbeid til masteroppgaven min om hvordan elever på mellomtrinnet bruker verktøyet Google Translate når de skriver i engelskfaget, og videre hvordan bruken av dette verktøyet påvirker deres skriveevne, om det har en positivt eller negativ virkning på produktet de skriver. Det er gjort svært lite forskning om hvordan Google Translate påvirker elevers skriveevner så jeg håper at forskningen min kan behjelpe bruken av verktøyet for fremtidig undervisning, både for meg selv og andre lærere. Samtidig ønsker jeg å stille deg som lærer noen spørsmål angående bruken av Google Translate i klasserommet og dine holdninger rundt det.

Prosjektet vil finne sted over to ulike engelsktimer på to ulike dager. I prosjektet vil elevene skrive to korte stiler om et tema som er forhåndlaget av meg. Den første gangen de skal skrive vil de ikke få tilgang på Google Translate, men de kan bruke elektroniske eller fysiske ordbøker. Den andre gangen dere skal skrive en stil vil bruken av Google Translate være tillatt/oppmuntret, men ikke påkrevd. Informasjonen som blir hentet inn blir videoopptak av skjermene deres under skriveprosessene samt det ferdige dokumentet som de produserer. Med skjermopptak menes at et program på Chromebooken deres vil ta opp det som foregår på skjermen deres, webkamera og filming av elevene vil ikke skje. I tillegg ønsker jeg å intervjue to-tre tilfeldig valgte elever mens de skriver for å finne ut deres tanker mens de er i skriveprosessen og bruker verktøyet.

Hvem er ansvarlig for forskningsprosjektet?

Universitetet i Stavanger er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Utvalget ble trukket via at jeg tok kontakt med skolen som du er lærer på går på, og spurte om det var noen klasser på mellomtrinnet hvor jeg kunne gjennomføre et forskningsprosjekt som ikke ville ta mange timer ut av den vanlige undervisningen. Jeg fikk beskjed om at du var mer enn villig til å stille opp med klassen sin.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at elevene dine vil gjennomføre to 30 minutters skriveøkter hvor opptak av skjermene deres vil bli gjort. Disse skjermopptakene vil bli lastet ned på en kryptert harddisk, sammen med stilene som de skriver. Under selve skriveøktene vil jeg ikke notere ned noe elevene foretar seg ved Chromebookene sine

ettersom skjermopptakene samler inn den dataen som trengs. Noen elever blir også trukket ut til å bli intervjuet mens han/hun skriver, hvor lydopptak vil bli gjort i tillegg som også blir lagret kryptert på en harddisk. Spørsmål som vil være på intervjuet inkluderer «Er det lettere for deg å skrive når du får lov å bruke Google Translate?» og «Hvorfor brukte du heller en ordbok i stedet for Google Translate for dette ordet?». Spørsmålene kan ha litt variasjon avhengig av hva eleven gjør når han/hun skriver, men det er kun slike type spørsmål som vil bil stilt.

Jeg ønsker også å stille deg som lærer noen spørsmål rundt bruken av Google Translate i ditt klasserom. Dette er gjort slik at jeg kan få en bedre forståelse for hvorfor elever gjør slik som de gjør når de tar i bruk verktøy som Google Translate, og hvordan holdningene til lærere rundt bruken av maskinoversettere har utviklet seg gjennom årene og deres tanker rundt dem.

Du som lærer kan gjerne få innsyn i intervjuguiden på forhånd hvis det skulle være et behov for det, i så fall ta kontakt med meg. Ingen av intervjuspørsmålene vil på noen måte være inngripende, det vil kun bli stilt spørsmål i henhold til bruken av Google Translate.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Resultatene av intervjuet vil bli anonymisert slik at din anonymitet vil være beholdt i det ferdige resultatet av masteroppgaven.

- Tilgang til transkripsjonen av intervjuet vil være limitert til meg og min veileder.

- Resultatene av intervjuet vil kun bli benyttet til bruk av masteroppgaven.

- Endringer av betingelsene ovenfor vil kun skje dersom ny samtykke blir gitt.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Opplysningene anonymiseres mens dataen behandles og når prosjektet avsluttes/oppgaven er godkjent, som vil være juni 2021. Innen prosjektslutt vil all data som skjermopptak, lydopptak på intervju og dokumentene som elevene skrev bli destruert.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Universitetet i Stavanger har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:
- Universitet i Stavanger ved Rebecca Anne Charboneau Stuvland, rebecca.a.stuvland@uis.no, 51 83 15 77
- Kristian Fellowes Haukelid, Kristian.Fellowes@gmail.com, 906 46 487
- Vårt personvernombud: personvernombud@uis.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:
- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Rebecca Anne Charboneau Stuvland   Kristian Fellowes Haukelid

(veileder)   (student)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Effektene av å bruke Google Translate i engelskfaget i norsk skole?*, og har fått anledning til å stille spørsmål. Jeg samtykker til:

☐ At jeg stiller opp til intervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

----

(Signert av prosjektdeltaker, dato)

APPENDIX 3A - TEACHER INTERVIEW QUESTIONS IN ENGLISH

Which experiences do you have of using Google Translate in the classroom? (Positive, negative, other?)

Is Google Translate used in the class?

      If yes, how?

      If no, why not?

Do you use other translation programs instead of Google Translate, if so, why?

Have the learners received any form of training on how to use Google Translate?

Have you as a teacher received any form of training on how to use Google Translate?

Do you see any positive aspects of using Google Translate?

Do you see any negative aspects of using Google Translate?

Why do you think learners decide to use Google Translate or other translation programs?

Do you know any of the strengths and/or weaknesses that Google Translate or other translation programs have?

Examples of questions that may arise if the learners perform certain actions while they are writing their essays and decide to use/not use Google Translate:

Why did you paste the results that you received from Google Translate into your document without editing it afterwards?

Why did you decide to not use the result that you received from Google Translate?

Why did you edit the result you received from Google Translate in your document?

Why did you rather use a dictionary instead of Google Translate for this word?

Would you rather use Google Translate for this word if you were given the opportunity instead of using a dictionary?

Why did you choose to use a synonym of the word that Google Translate gave you instead of the original?

Is it easier for you to write when you are allowed to use Google Translate?

Is it easier for you to write when you use a dictionary rather than Google Translate?

Why do you use Google Translate?

What do you usually use Google Translate for?

Do you know how to use a dictionary?

Hvilke erfaringer har dere i klassen med Google Translate? (Postive, negative, andre?)

Blir Google Translate brukt i klassen?

      Dersom ja, hvordan?

      Eventuelt nei, hvorfor ikke?

Bruker dere andre oversettingsprogrammer istedenfor Google Translate, hvis så, hvorfor?

Har elevene fått opplæring i hvordan å bruke Google Translate?

Har du som lærer fått noen opplæring i å bruke Google Translate?

Kan du se for deg noen positive sider ved bruk av Google Translate?

Kan du seg for deg noen negative sider ved bruk av Google Translate?

Hvorfor tror du at elever bruker Google Translate eller andre oversettingsprogrammer?

Kjenner du til styrkene og/eller svakhetene til Google Translate eller andre oversettingsprogram?

Eksempler på spørsmål som kan dukke opp dersom elevene gjør visse handlinger mens de skriver stilene sine og bruker/ikke bruker Google Translate:

Hvorfor limte du resultatet du fikk fra Google Translate rett inn i dokumentet ditt uten å redigere?

Hvorfor bestemte du deg for å ikke bruke det resultatet du fikk fra Google Translate?

Hvorfor redigerte du svaret som du fikk fra Google Translate i dokumentet ditt?

Hvorfor brukte du heller en ordbok i stedet for Google Translate for dette ordet?

Ville du heller brukt Google Translate for dette ordet om du fikk sjansen istedenfor en ordbok?

Hvorfor valgte du synonymet for det ordet istedenfor det originale ordet som Google Translate gav deg?

Er det lettere for deg å skrive når du får lov å bruke Google Translate?

Er det lettere for deg å skrive når du heller bruker ordbok ovenfor Google Translate?

Hvorfor bruker du Google Translate?

Hva pleier du å bruke Google Translate til?

Klarer du å bruke en ordbok?

Vocabulary:

- Before determining vocabulary of each text, perform data cleansing on words that are misspelt but still understandable. If words inputted are spelt completely wrong and meaning cannot be derived from them, remove them from the text. Correct these words to their correct spelling, then input the text into the CEFR checker by Duolingo to determine the overall vocabulary of the learner's two texts. Compare the overall CEFR level of two texts and the predicted CEFR levels of words to determine if the vocabulary of learners' texts increased, decreased or stayed the same between the two texts.

Syntax errors:

- Create a point scoring system that has 0 points and 1 point. For each sentence in the learners' text that has no syntax errors, award 1 point. For each sentence that has one or more syntax errors, award 0 points. Add up the points to calculate the sum of syntax errors. A syntax error is defined for this guideline as any error that violates the rules, principles and processes that govern the structuring of a sentence. As these learners are quite young, an exception will be made for not placing commas as long as it does not massively hinder the flow of the sentence.

Subject-verbal concord:

- Create a point scoring system that has 0 points and 1 point. For each sentence in the learners' text that has no subject-verbal concord errors, award 1 point. For each sentence that has one or more subject-verbal concord errors, award 0 points. Add up the points to calculate the sum of subject-verbal concord errors. For this guideline, a subject-verbal concord error is defined as any mismatch between the subject and the verb.

Spelling mistakes:

- Count every word that is misspelt in the text. A word is considered misspelt if the word is either spelt incorrectly or if a different word (spelt correctly but wrong context) was used, e.g. *too apples* instead of *two apples*. Words that are supposed to have capital lettering, e.g. *I* and proper nouns, but learners have failed to capitalise will not count as spelling mistakes for this analysis. Words that use the apostrophe incorrectly will be treated as a misspelt word. Furthermore, a word that counts as a subject-verbal concord error will not count as a spelling mistake error.

Essay length (Word count):

- Input each text into a word processing program such as word to determine essay length of both texts. Calculate the increase or decrease in essay length between the two texts in percentages.

Creating a fair comparison for subject-verbal concord, syntax errors, and spelling mistakes:

- As an increase in text length will most likely lead to an increase in subject-verbal concord errors, syntax errors, and spelling mistakes, calculations need to be made for these three categories to have a fair comparison between the two texts that learners write. Look for and count all errors for both texts as normal, but also make a comparison by dividing the number of correct sentences on total number of sentences for subject-verbal concord and syntax errors. For spelling mistakes, divide the final tally of misspelt words on the total number of words for each texts.

APPENDIX 5 - PERCENT AGREEMENT BETWEEN TWO RATERS

If both raters reached the same conclusion for a specified category, 1 was given for agreement. 0 was given if the two raters did not reach an agreement in a specified category.

In total, 23/25 or 92% of ratings were in agreement.

	Learner 1			Learner 2		
	Rater 1	Rater 2	Agreement	Rater 1	Rater 2	Agreement
Vocabulary Rating after data cleansing	A2	A2	1	B1	B1	1
Number of syntax error sentences	4	4	1	0	0	1
Number of subject-verbal concord error sentences	1	1	1	0	0	1
Number of spelling mistakes	6	6	1	1	1	1
Essay length count in words	71	71	1	281	281	1
Agreement percentage:			100% agreement			100% agreement

	Learner 3			Learner 4		
	Rater 1	Rater 2	Agreement	Rater 1	Rater 2	Agreement
Vocabulary Rating after data cleansing	A2	A2	1	A1	A1	1
Number of syntax error sentences	1	2	0	1	1	1
Number of subject-verbal concord error sentences	0	0	1	1	1	1
Number of spelling mistakes	1	1	1	1	1	1
Essay length count in words	190	190	1	109	109	1
Agreement percentage:			80% agreement			100% agreement

	Learner 5		
	Rater 1	Rater 2	Agreement
Vocabulary Rating after data cleansing	A2	A2	1
Number of syntax error sentences	2	1	0
Number of subject-verbal concord error sentences	0	0	1
Number of spelling mistakes	2	2	1
Essay length count in words	159	159	1
Agreement percentage:			80% agreement

APPENDIX 6 - IMPLICATIONS FOR FUTURE TEACHING (ADDITIONAL ICT TOOLS EVALUATED)

Multiple CALL tools and other ICT tools that can be used as CALL tools have been looked at that might be of interest to pursue in future teaching. Duolingo in particular is a CALL tool that is used to learn foreign languages in fun and interactive ways. As it can be used both in a web browser and on phones through their app, there are definitely ways to make use of it in the FL classroom. The study by Loewen et al. (2019) concludes that there is a positive correlation between the amount of time spent on Duolingo and learning gains, meaning that the tool can serve a purpose in the FL classroom, albeit not as the sole method for learning a language. For learners who are struggling in English, Duolingo offers English language courses, but AP learners and HP learners might not benefit from using the tool as much. As the percentage of learners with 1:1 device ratio is quite high in Norway, with current trends pointing towards these percentages becoming even higher in the future, more and more FL education is seemingly becoming digitalised (Fjørtoft, Thun & Buvik, 2019, p. 24). Thus, it is important to be aware of relevant ICT tools that can be incorporated into lesson plans.

Microsoft Translator, a competitor to Google Translate, has a highly useful feature which Google Translate currently does not support that can be used in classes with learners of many language backgrounds. This feature, called Conversation[13], allows for real-time translation between the language that the host wishes to use and whichever language the participants wish to use. The conversation only shows written text for all participants, but the participants can speak and have the translator produce a speech-to-text translation of what was said as well as having the option to enable text-to-speech for the language that the participant chose before entering the conversation. For example, the teacher has a group of three learners, all of whom speak a different language and struggle to understand both Norwegian and English, the two languages the teacher is familiar with. With this feature, all three learners can participate in the teacher's conversation with a code and choose which language they wish to see the conversation in. This is individual for each learner, meaning that in this example, there are four different languages being typed (or spoken if the language is supported), but all participants only see the language which they selected to join the conversation with. However, a great addition to this feature is that all participants can choose to see the text that

---

[13] https://translator.microsoft.com/

was sent in its original language as well, meaning they can learn words, phrases, and sentences whilst conversing with each other (see Figure 8).
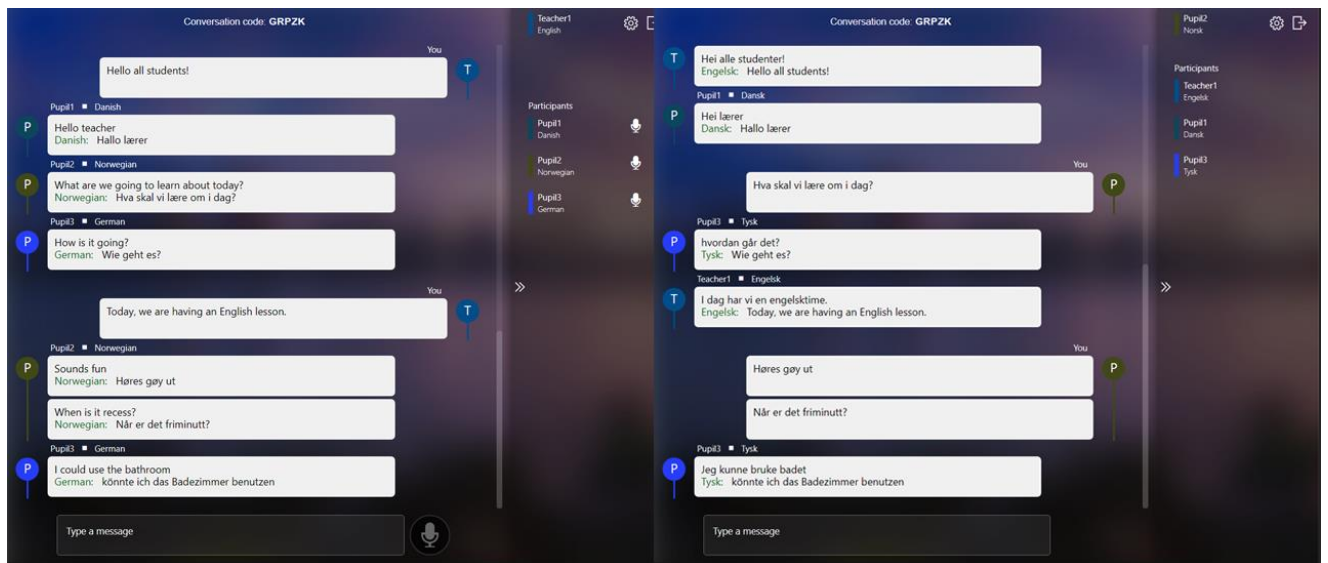

Figure 8: An example of a conversation of four people who chose different languages prior to joining the conversation. Left picture is what the conversation looks like for a person who chose English as their language while the right picture is what the conversation looks like for a person who chose Norwegian as their language.

Other ICT tools not initially thought of as CALL tools, such as creating wikis or blogs, can promote language learning if done correctly and are definitely worthy options to consider integrating into the FL classroom (Golonka et al., 2014). Video games are also a great source of teaching subjects in schools, some of which can be used in English or other FL subjects. Multiple studies show that video games can provide learners with input to enhance their English listening, reading, vocabulary skills and learning motivation (Chen & Yang, 2013; Klimova & Kacet, 2017). The author of this thesis is definitely considering some games that can be used to promote FL oral fluency, such as Keep Talking and Nobody Explodes[14], or games that can be used in relation to important topics, e.g. the ever-growing importance of being critical of news where games such as Bad News[15] or Factitious[16] can be used.

---

[14] https://keeptalkinggame.com/
[15] https://www.getbadnews.com/
[16] http://factitious.augamestudio.com/