

Miriam Marie Ringrose

ChatGPT and differentiated reading materials

The possibilities and limitations of AI-generated graded readers for an ESL context

Master's thesis in Language Studies with Teacher Education

Supervisor: Nicole Louise Busby

Co-supervisor: Annjo Klungervik Greenall

May 2024

Miriam Marie Ringrose

ChatGPT and differentiated reading materials

The possibilities and limitations of AI-generated graded readers for an ESL context

Master's thesis in Language Studies with Teacher Education
Supervisor: Nicole Louise Busby
Co-supervisor: Annjo Klungervik Greenall
May 2024

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature



Norwegian University of
Science and Technology

Abstract

In 2023, the use of AI chatbots, especially ChatGPT, witnessed a notable surge, sparking debate on the application of such tools in various domains, including education. One such application of AI is to use its generative capabilities to rewrite texts to be more suitable for students learning English as a second language. To investigate how ChatGPT solves this task, this study compared two types of simplified versions of literary works, referred to as graded readers, namely human-made graded readers and graded readers generated by ChatGPT 3.5. The focus of this study was the vocabulary choices of the two types of graded readers, and the comparison was based on an analysis of lexical strategies and word frequency. The results of this study suggest that ChatGPT managed to simplify the text, by means of using several lexical strategies and including a greater proportion of high frequency words. The findings also show differences between human-made and AI-generated graded readers in the ways that lexical items, such as low frequency words, were omitted and changed. Based on the lexical strategy results, it was hypothesised that these differences in strategy use may indicate an overall difference between how ChatGPT and humans adapt and shorten text. In addition, the findings indicate that characteristics of the literary works, such as the proportion of low-frequency words, seem to have an impact on the extent to which ChatGPT simplifies the vocabulary to the desired level. Overall, this study indicates that ChatGPT 3.5 could be a useful resource for teachers to simplify text, but it may require adjustments to accommodate its distinct simplification style.

Sammendrag

I 2023 var det en kraftig fremvekst i bruken av KI-baserte språkmodeller, som ChatGPT. Som følge av dette har det oppstått debatt rundt anvendelsen av disse verktøyene på forskjellige arenaer, inkludert i utdanning. En mulig anvendelse av KI i skolen er å bruke det til å omskrive tekster slik at de blir bedre egnet for elever som lærer engelsk som andrespråk. For å undersøke hvordan ChatGPT løser slike oppgaver, ble to typer forenklinger av litterære verk (*graded readers*) sammenlignet. Dette var menneskeskaptede forenklete versjoner og KI-genererte forenklete versjoner fra ChatGPT 3.5. De leksikalske endringene i de forenklete variantene ble så analysert. Analysene inkluderer de leksikalske oversettelsesstrategiene som ble benyttet og ordfrekvensfordelingen. Funnene i denne studien tyder på at ChatGPT klarte å forenkle teksten ved å bruke leksikalske strategier og øke den totale ordfrekvensen. Videre viste analysene forskjeller mellom ChatGPT og menneskers oversettelser i måten ord, særlig lavfrekvente ord, ble fjernet og endret. Basert på resultatene argumenteres det for at de ulike resultatene kan tilsi at ChatGPT 3.5 og mennesker har forskjellige metoder for å redusere tekstlengde. I tillegg virker kvaliteter ved de litterære verkene, som høy forekomst av lavfrekvente ord, å være av betydning for hvor store likhetstrekk man kan finne mellom menneskegenererte og KI-genererte forenklinger. Alt i alt framstår ChatGPT 3.5 som en nyttig ressurs for å forenkle tekst, men lærere burde være klare over noen særtrekk ved forenklingsstilen som kan påvirke tekstens kvalitet.

Acknowledgements

Thank you to my supervisors, Nicole Louise Busby and Annjo Klungervik Greenall. Your feedback has been extremely valuable, and I always looked forward to our supervision sessions. I really appreciate the time and effort you have put in to helping me throughout this project. Future students will be lucky to have you both as their supervisors.

I also want to thank my wonderful family and friends, and a special thanks to my mum, Priscilla, and my boyfriend, Gunnar, for listening to me, helping, and comforting me throughout my master's. Thank you for reading through my master's and helping with the final touches.

A final thanks to my friends at school who have kept me going on many long days. A special thanks to Tilde, Amanda, Hanna, Daniel, Nicolai, Anna-Marie, and Marthe for making the process of writing a masters fun.

Thank You!

Trondheim, 15th of May 2024

Miriam Marie Ringrose

Table of contents

List of tables and figures	5
List of abbreviations	6
1 Introduction.....	7
2 Theoretical background	11
2.1 Second language acquisition and comprehensible input	11
2.2 Graded readers as translations	12
2.3 Artificially generated graded readers	15
3 Methods.....	18
3.1 The choice of naturalistic language generation model	18
3.2 The material	19
3.3 Method of analysis for translation strategies	25
3.4 Method of analysis for frequency distribution.....	30
4 Results.....	32
4.1 Results of the analysis of translation strategies	32
4.2 Results of the analysis of word frequency distribution.....	38
5 Discussion.....	40
5.1 Translation strategies in human-made graded readers.....	40
5.2 Comparison of translation strategies in AI-generated and human-made graded reader	43
5.3 Discussion of frequency distribution	45
5.4 Implications.....	47
5.5 Limitations	48
6 Conclusion	50
References.....	52
Appendixes	59

List of tables and figures

List of tables

1. Prompt development table23
2. Sample of table containing source text and target text comparison of coupled pairs 29
3. Sample of table containing source text and target text comparison of strategies 30
4. Sample of the table from the analysis of Text A showing the different use of omission 34
5. Frequency distribution table for source and target texts 38

List of figures

1. Source text and target text relationship 18
2. Distribution of translation strategies in human-made (HM) and AI-generated graded reader.32

List of abbreviations

AI:	Artificial intelligence
BNC:	British National Corpus
COCA:	Corpus of Contemporary American English
EFL:	English as a Foreign Language
ESL:	English as Second Language
HM:	Human-made
L2:	Second language
NLG:	Natural language generation
PC-to-word:	Phrase/clause-to-word
ST:	Source text
TS:	Text-simplification
TT:	Target text
Udir:	The Norwegian Directorate for Education and Training

1 Introduction

The aim of this study is to compare AI-generated and human-made differentiated reading material, focusing on lexical choices. In this thesis, by analysing lexical strategies and word frequency, the vocabulary choices of ChatGPT 3.5 are compared to the vocabulary choices of humans. The aim is to investigate whether ChatGPT can be used to simplify pre-existing text to be suitable for an English as a second language (ESL) context.

A one-size-fits-all approach to teaching reading is not appropriate in today's diverse reading classrooms (Firmender et al., 2013). In Norway, by 10th grade, students are expected to be able to read, reflect and interpret “various types of texts” including “English-language fiction” (The Norwegian Directorate for Education and Training (Udir), 2019, p. 9). Despite these aims, the results from national tests in Norway indicate that many students have not achieved this level (Udir, 2023). Descriptions of second language proficiency levels by the Council of Europe (2024), indicate that students’ English reading levels at lower secondary school range from A1 to B2 level (Udir, 2023). The English classrooms in Norway therefore consist of a diverse student body with various proficiency levels (Udir, 2023).

Through the internet and widespread accessibility of English medias, students have the opportunity to immerse themselves in English language spheres in their spare time. The classroom is not necessarily the primary language learning arena for all students, and students’ extracurricular involvement with English can have a major impact on their proficiency level (De Wilde et al., 2020). On top of this, students may have learning disabilities and have various language and socioeconomic backgrounds that also play into their proficiency levels (Lervåg et al., 2019). Reading levels may then vary, and this calls for differentiated material to meet these diverse needs (Murty et al., 2021).

Choosing reading materials that are at an appropriate level for learners has been shown to be beneficial for language acquisition as well as motivation for reading (Murty, et al., 2021). Reading texts at an appropriate level helps with several language skills: including word learning, reading fluency, and language comprehension (Chiang, 2015; Pulido, 2007; Stanovich, 1980). Material of this sort can support reading comprehension and in turn help develop the network of related language skills.

Reading comprehension is a flexible and ongoing cognitive process which is dependent on negotiation between the reader and the text (Woolley, 2011). This means that text selection

plays a vital role for reading comprehension. The complexity of the content, structure, syntax and vocabulary of the text all have an impact on the readers ability to comprehend the material (Woolley, 2011). For vocabulary in particular, studies have established a link between reading comprehension and the vocabulary of the text, indicating that texts consisting of difficult words are harder to comprehend (Cain, 2007; Gersten et al., 2001; Oakhill & Cain, 2007; Vellutino et al., 2007; Pulido, 2007). Within this research, one established determiner for reading comprehension is that the vocabulary should mostly consist of words that are known to the reader (Cobb, 2007). This suggested number of known words for reading comprehension is high. Research suggests that the readers need to understand at least 98% of words in the text to enable adequate comprehension (Nation & Deweerdt, 2001; Hu & Nation, 2000). Words that are more common in a language, referred to as high frequency words, tend to be learnt first since language learners are more often exposed to them (Nation, 2022). Based on this, a strategy to increase the number of known words is to use a lot of high frequency words in differentiated reading material to ensure that the text consist of known words to the reader.

As explained by Cobb (2007), to measure the frequency of words, corpus material has been studied and words have been organized into groups of 1000 based on their frequency. In this system, the 1000 most frequent word families in English are categorized as level 1 (K-1) and viewed as very frequent and most likely to be known. Following this, differentiated material can be adjusted to the student based on the number of high frequency words, for example by including some of English's most frequent words to ease reading comprehension for ESL students.

Although differentiated reading materials such as newsletters, stories, and textbooks are already used in ESL classrooms, Murty et al. (2021) argue that there is still need for more differentiation in the English as a second language classroom. This is supported by Klein (2008), who suggests that teachers struggle to address the individual reading needs of their students throughout the year. A possible explanation for this may be that teachers lack access to enough differentiated material to meet the needs of their diverse classrooms. This could be due to the lack of time teachers have to search for these texts, or the lack of accessibility and funding, as many differentiated texts are behind a paywall. In addition to this, not all resources have been adapted for different reading levels, meaning that if teachers want to use this undifferentiated material, they would be left with no choice but to give the undifferentiated reading material to their students or adapt the material to fit the students' needs themselves.

This calls into question whether there is an easier way for teachers to get differentiated material into the classroom.

A possible solution to this problem is to use artificial intelligence (AI) to create differentiated reading material. Natural language generation tools, such as ChatGPT, have the capability to adapt text much faster than humans. In addition, differentiated texts are often simplified versions of original literary works, and ChatGPT has been found to be successful at text simplification tasks (Jeblick, et al., 2023). A previous analysis of ChatGPT found that the language model can generate human-like text in terms of “accuracy, relevance, depth, breadth, logic, persuasiveness, and originality” (Liu, et al., 2023, p. 10; Sunsjak, 2022). If ChatGPT can take an original undifferentiated text as input and create versions suited for the ESL classroom, this would give teachers access to free differentiated material of any text they want to use in their teaching.

To investigate whether ChatGPT is suitable for this task, the output from ChatGPT 3.5 when asked to create a simplified version of a literary work will be analysed. More specifically, ChatGPT will be asked to create differentiated material of pre-existing literary works for the ESL classroom. The literary works that will be used as source material are Text A) *A Long Way Home* by Saroo Brierley (2013) and Text B) *Brick Lane* by Monica Ali (2003). The imagined target group for these texts are ESL 10th grade students in Norway reading at a A2/B1 level. The AI output created with this group in mind will be compared to corresponding sections in a human-made differentiated version, typically referred to as a graded reader, of the same literary work. The human-made graded reader versions of Text A and B are from Penguin readers and are retold by Parmar (2022) and Khan (2023). Accordingly, the material used in the analysis consists of a total of six texts: excerpts from two literary works (source text A and B, see Appendix B), pre-existing human-made graded reader versions of the excerpts (see Appendix C) and the artificially generated graded reader version of the excerpts (see Appendix D).

According to Simensen (1990), changing the vocabulary of a text is the most common way of adapting a text into a graded reader. Due to the central role of vocabulary in creating a graded reader, as well as its role in reading comprehension, the analysis is focused on comparing the vocabulary choices of the texts. The analysis is divided into two parts that focus on different aspects of the vocabulary. Part one consists of an analysis of the lexical translation strategies and part two consists of a word frequency analysis.

Using methods from translation studies, including Toury's (2012) coupled-pair method, the first part of the analysis consists of analysing the lexical translation strategies used to create the human-made and AI-generated graded readers. Lexical strategies concern the translation of words, so an analysis of lexical strategies consists of an investigation of how specific lexical items have been translated (Hill-Madsen, 2014). The lexical strategies are identified by finding the corresponding units, referred to as coupled pairs, between the original excerpts and the human and AI differentiated versions. By analysing the coupled pairs, the focus of this part of the analysis is categorising which translation strategies have been used to create the human and AI simplified texts. The five strategies that will be investigated are *demetaphorisation* (removing figurative language), *explicitation* (making implicit lexical information explicit), *phrase/clause-to-word simplification* (replacing a phrase or clause with a single word), *omission of lower frequency words* (removing uncommon words), and finally *use of higher frequency words* (replacing uncommon words, lower frequency, with more common words, higher frequency). One of the strategies will receive extra focus in the second part of the analysis, namely the high frequency word strategy. To analyse the word frequency further, the second part of the analysis consists of an investigation of the word frequency distributions of the texts.

The analysis of lexical strategies and frequency distributions will provide the foundation to compare the human-made and AI-generated translations, in order to gain a better understanding of how AI changes the lexicon of a text compared to humans. As vocabulary is central in reading comprehension, the lexical build-up of a text is vital for its success as a graded reader. A general aim in this study is then to investigate whether there are considerable differences in vocabulary choices between human-made and AI-generated graded readers, and possible limitations with using AI-generated graded readers in the ESL classroom. The current study investigates how ChatGPT 3.5 performs compared to translators of pre-existing human-made graded readers when asked to create differentiated material for an A2/B1 level. The specific research questions that will be investigated are:

Q1: To what extent are the selected translations strategies used in human-made and AI-generated graded reader versions of Text A and Text B and how do the results compare in terms of prevalence?

Q2: What are the relative word frequency distributions of the human-made and AI-generated versions of Text A and B and how do they compare?

2 Theoretical background

2.1 Second language acquisition and comprehensible input

Reading is an easy way of getting exposed to a language and can function as a pathway for language learning. One undoubtable factor in language learning is language exposure, often denoted as *input* in linguistics. As articulated by Krashen (1992), this input, needs to be *comprehensible*. By comprehensible Krashen (1992) means that the input must not surpass the language learner's grasp, adhering to Krashen's $i+1$ hypothesis. In this model, 'i' signifies the learner's current level of competence, and '+1' represents the elements that slightly exceed the learner's current level (Krashen, 1985). According to this hypothesis, the ideal input, should be slightly more challenging than the learner's current language level.

The concept of comprehensible input has played an important role in second language (L2) didactics and research, as teachers and scholars have highlighted the benefits of giving students reading materials that are adapted to their level. The texts need to be at a level where students can make connections between “the new and the familiar”, so that reading can lead them towards higher proficiency (Tomlinson, 2003, p. 19). Furthermore, there are a range of benefits in providing texts at a suitable level for L2 students, including comprehension and retention of new vocabulary (Chiang, 2015; Pulido, 2007). According to Grabe (2009), reading experience causes an increase in reading fluency and processing speed, and this is because the more experience one has with reading the more automatic several complex processes become. This is in line with Stanovich (1980), who ties increased comprehension to increased reading fluency. According to Stanovich, with low reading comprehension cognitive resources will be allocated to word recognition and other lower-level processes rather than to overall comprehension. So, providing students with comprehensible input in the form of reading material slightly more advanced than their current level, will allow them to allocate more cognitive resources to overall comprehension, creating an environment where new words can be learned from the context.

The question remains what this “suitable” reading material consists of. Since 98% of the words in a text should be known to the reader in order to facilitate reading comprehension, Nation and Wang Ming-Tzu (1999) argue that type of input is suitable for L2 reading. This density of known words allows the reader to learn words from the context without external help (Nation & Deweerd, 2001). This is supported by Krashen (1992), who argues that new words are

learned from contextual cues, and reading material of this sort will be useful for acquiring language. Based on this, Nation and Deweerdt (2001) argue that simplified texts have clear value in language education, and that this material should consist of around 98% familiar words to the reader.

One strategy to create texts with words that are likely to be known to the reader, is to include high frequency words. The prevalence of a word in a language is a good indicator for how likely it is for an individual to know it (Nation, 2022). It is hard to imagine someone knowing some English without knowing the most frequent words in English: *the, of, and, a, and to*. The most frequent word families, up to 3000, will cover a large part of the vocabulary of a text, while words that are in the 6000-9000 range of most frequent word families will occur “infrequently” and words over 9000 will occur “very infrequently” (Nation, 2022, p. 29). Since high frequency words are more likely to be known, a text written to consist of mainly known words for an ESL context should aim to use a large amount of high frequency words, depending on the level of the student. Findings by Nation (2006) suggest that for a student to be able to read various types of texts, they will need to know 4000-word families to know 95% of the words in the text and 9000-word families to know 98%. Graded readers, on the other hand, tend not to proceed systematically past the 3000-word families level (Cobb, 2007). If students do not know 9000-word families in English, this 98% threshold will not be met when reading undifferentiated texts, and this may hinder the full reading comprehension, fluency, and vocabulary learning benefits that one can gain from a text consisting of mainly known words. Using texts that contain a greater proportion of high frequency words means that students are more likely to know the words which in turn creates a text at a more suitable reading level.

2.2 Graded readers as translations

One way of providing language learners with comprehensible input is by giving them graded readers. Graded readers can be books specifically written in simple language, or they can be original works that have been rewritten for this target audience. In the literature on graded readers, several scholars have conceived graded reader versions of literary works as intralingual translations (Tovar, 2019; Zethsen 2009; Hedge, 1985; Simensen, 1990). Intralingual translations are translations occurring within a language, so they include changes in complexity, style or register, instead of changes from one language to another (Hill-Madsen, 2019). The term originated from Jakobsen (1959) who presented a typology of translation. In

this well-known translation typology, intralingual translations are seen as one of three types, alongside translations between languages (interlingual translation) and translation from verbal signs to non-verbal signs (intersemiotic translation) (Jakobsen, 1959; Hill-Madsen, 2014). The concept of intralingual translations has been expanded on in recent years and various studies have used translation methods and theory to analyse and discuss different types of intralingual translations. Some examples include Hill-Madsen (2014), who analysed several forms of expert-to-layman translations, Zethsen (2018), who analysed expert-to-layman translations within medicine, Brems (2018) who analysed translations between regional varieties of Dutch, and finally Hansen-Schirra (2020) who looked into translation into a simplified form of language developed for those with disabilities. Taking the above research into account, the main point is that graded readers can be categorised as translations, and methods from translation studies can therefore be used to analyse graded readers.

The inclusion of intralingual translations within the term *translation* is nevertheless debated, as many adopt a stricter and more narrow definition of the term, only including changes from one language to another, so called interlingual translations (Newmark, 1981; Tymoczko, 2005; Mossop, 2016). However, in line with Tovar (2019), Schmid (2008) and Zethsen (2009), a broader definition of translation will be adopted in this study, which includes intralingual translations.

Expanding upon Jakobsen's typology (1959), Gottlieb (2008) identified five types of intralingual translations. These are diachronic (modernising text), dialectal (translations between dialects), transliteration (between writing systems), diamesic (between speech and writing), and finally diaphasic, which involves changing the register of the text to meet a new target audience within the same language system (Hill-Madsen, 2019). Based on these descriptions graded readers can be seen as belonging to this latter subcategory of diaphasic translations. This is because graded readers involve adapting the text to a new audience within the same language. For graded readers, this new audience is typically either native-speaking children or language learners. Following this terminology, graded readers that are rewritten versions of other literary works will be seen as diaphasic translations in this study.

2.2.1 Translation studies on diaphasic translations and graded readers

Using this broader definition of translation, scholars have utilised translation terminology and methodology to label and identify the characteristics of intralingual translations, including diaphasic translations. Zethsen (2009) analysed five different intralingual translations of the Danish bible including a diaphasic translation created for children. Based on her analysis, she identified translation strategies that occur in intralingual translations. Zethsen found that the translational strategies used for intralingual translations only differ from interlingual translations in matter of “degree and motivation rather than kind” (p. 809). This means that the shifts identified in the analysis of the five versions of the Danish Bible were not unique to intralingual translation and can be found in interlingual translation. Based on this, one can argue that the use of translation strategies from interlingual translation methodology can be used in the analysis of intralingual translations.

In addition to this general conclusion about strategies in intralingual translations, Zethsen (2009) identified some strategies that more specifically centre around “the level of background knowledge and ability of comprehension of the target group” (p. 803). These strategies are relevant for graded readers as meeting the comprehension level of the target group matches the *skopos* (the motivation) for creating a graded reader (Velmeer, 1989). According to Zethsen, some strategies that centre around this *skopos* are omission, explicitation and simplification.

Building on Zethsen, Hill-Madsen (2014) aims to “fill in “the empirical gap” (p. 85) specifically regarding the lexical strategies found in intralingual translations between registers. Through analysing diaphasic translations of medical information simplified for patients, Hill-Madsen identified translational strategies including explicitation, and several forms of lexical simplification, such as *decrease-in-formality* and *decrease-in-technicality*. According to Hill-Madsen, these forms of lexical simplification involve using easier vocabulary to meet the level of knowledge of the target group. These two articles together highlight some important translational strategies that are relevant for diaphasic translations and centre around the comprehension level and background knowledge of the reader.

Other scholars have used translation terminology and methodology to label and identify the characteristics of graded readers, including Tovar (2019), and Simensen (1990). Basing himself on Hedge (1985), Tovar (2019) identified domains of control used to create a graded reader version of *And Then There Were None* by Agatha Christie. The domains identified were lexical control, informational control and syntactic control. Put simply, these domains are ways in

which graded readers are qualitatively different from texts written for native speaking adults. It will then be expected that creating a graded reader involves changing the text syntactically, informationally, and lexically. According to Simensen (1990), lexical control is the most common form of control in the creation of graded readers. Lexical control involves changes from more difficult words and phrases to more common words, which are more likely to be known by the target group (Tovar, 2019). However, as explained by Tovar, lexical control can also involve the removal of figurative phrases to more transparent vocabulary items, and removal of nonstandard forms of words, for example dialectal variation. Tovar (2019) also identified the above-mentioned strategies of omission and explicitation in his analysis of the graded reader. His findings indicate that explicitation is a strategy used in graded readers, albeit a rare strategy, and highlights omission as the most “commonly used technique in abridgments” (p. 61). He states that a norm in the creation of graded readers is to omit “to the point that the size of the TT may be reduced by more than one half” (p. 65). In his analysis he found that the graded readers word count was cut down 65% from the original.

Overall, based on the studies discussed we can make some assumptions about the characteristics of graded readers as intralingual translations: 1) the strategies used to create graded readers are the same as for other forms of translation, 2) the strategies used may, however, vary in degree and motivation of use from interlingual translations, 3) the strategies one can expect to find in diaphasic translations such as graded readers are: omission, explicitation, and lexical simplification, 4) lexical simplification in graded readers can involve changes from low to high frequency, and the removal of figurative language and non-standard language.

2.3 Artificially generated graded readers

Traditionally, graded readers are made by humans; however, the recent developments of natural language generation models may open the door for artificially generated graded readers. According to Dong et al. (2022), Natural Language Generation (NLG) is the process of generating natural language text, based on specific communicative goals. Dong et al. account for the various types of NLGs and state that one type of NLG is text-to-text generation, in which a text is rewritten into a different style. Within text-to-text generation, is a subcategory called text simplification (TS), a process in which the complexity of the text is reduced in both vocabulary and sentence structure while the core meaning is preserved (Sheang & Saggion,

2021). TS can be used to create diaphasic intralingual translations since it can be used to adapt the text to a new audience within the same language. NLGs can therefore be used to create texts which share characteristics of graded readers, such as a simplified vocabulary.

Today there are numerous NLG programs that can be used to simplify language, one of which is the readily available model ChatGPT 3.5. ChatGPT produces text in response to a *prompt*, a written task, by predicting based on a large amount of text what words are likely to be used in certain contexts (OpenAI, 2024). It then predicts which words would work in the context of its response and in relation to the prompt it has been given. Based on this, the same prompt can yield different outputs (Jiao, et al., 2023). According to Jiao et al., this has to be taken into consideration when working with ChatGPT. They suggest repeating the generation several times for more reliable results and to a better idea of ChatGPT's full capability. Another more well-known limitation with text generated from ChatGPT is its accuracy. According to Borji (2023), ChatGPT has demonstrated factual errors and inconsistency in reasoning. Nevertheless, according to Moons and Van Bulck (2023), ChatGPT is a powerful tool which has a broad application and has been trained to "offer services like language translation, summarisation, and text completion" (p. 124).

There is some research investigating whether the highly available NLG tool ChatGPT, can be used to simplify text to an easier register. According to Lui et al. (2023), the research on text generation points to ChatGPT being successful at text simplification, by means of reducing the length of the text and the complexity of the content. A study by Moons and Van Bulck (2023) investigated ChatGPT's ability to simplify medical information for children. They asked ChatGPT to simplify patient information for 6th-graders. Their findings indicate that the higher the reading level was to begin with, the more ChatGPT managed to reduce the difficulty. However, ChatGPT did not manage to reach the desired proficiency of a 6th grade-reading level for any of the texts. Moons and Van Bulck hypothesise that this may be due to ChatGPT being designed to have a reading level comparable to a high-school student (Moons & Van Bulck, 2023). Another study looking at ChatGPT's ability to simplify medical information for children found that ChatGPT improved readability and reading ease (Haver, et al., 2024). In addition, they found that ChatGPT's responses were deemed clinically appropriate in 92% of cases. However, Haver et al. did find incorrect information in the generated texts and stated that it is therefore necessary for medical professionals to check texts before they are made available to patients.

Furthermore, researchers have investigated ChatGPT's utility in an English as a foreign language (EFL) context. A study by Young and Shishido (2023a) investigated ChatGPT's ability to simplify text for an English as a foreign language context. They found that the text generated was less complex than the original. They also investigated the way in which the text was simplified. They found that ChatGPT shortened the text by reducing the number of sentences, and by reducing the number of words per sentence, even though it was not asked to shorten the text. This can be paralleled to findings from Tovar (2019), where omission was identified as a tendency in human-made graded readers. Finally, in another study, Young and Shishido (2023b) analysed the output of ChatGPT when asked "to make a dialogue to help EFL students to practice their English" (p. 69). Their findings suggest that ChatGPT's output was comprehensible for EFL students and that the material generated was suitable for students with an A2 proficiency in English. Comparing these findings to Moons and Van Bulck (2023), this raises the question of whether ChatGPT can produce text at an A2 level but cannot simplify text to this level.

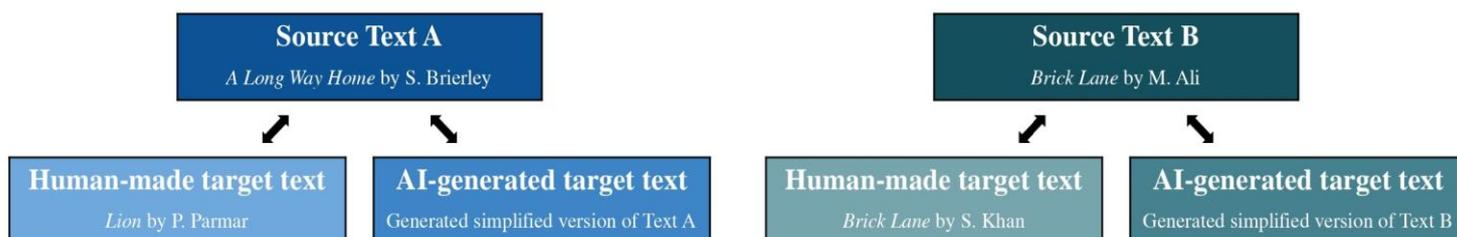
The above-mentioned studies compared the AI generated text to the original text or other artificially generated texts. There is a lack of research comparing ChatGPT's ability to simplify text up against human-made simplified text, especially within the ESL context and the creation of graded readers. The current study aims to see if ChatGPT shows significant differences to human-made graded readers when asked to create a simplified version of an original work fit for a A2/B1 proficiency level. To do this the vocabulary choices of the generated text will be compared to the vocabulary choices of prewritten graded readers by two metrics: lexical strategies and word frequency profiles (see section 3).

3 Methods

The current study consists of a comparative analysis of the translation strategies and word frequency distributions in human-made graded readers versus AI-generated graded readers. There are two original works of literature, Text A and B, with two graded reader versions each (human-made and AI-generated). In line with standard terminology in Translation Studies, the original works will be referred to as the *source texts* and the graded reader versions will be referred to as *target texts* throughout the thesis (Munday, 2016). As illustrated in Figure 1 below, the excerpts from each of the STs will be analysed with the corresponding excerpt from one of the graded reader versions. The STs and the corresponding TTs will be mapped onto each other to identify the extent to which the set of pre-selected strategies are used in the human-made and AI-generated versions. In addition, the different texts will also be compared based on their word frequency distribution. The excerpts from the STs and the TTs will be analysed using a word frequency distribution analysis tool, called VocabProfile (Cobb, Compleat Web VP, 2024). This will give an overview of the amount of high and low frequency words found in the human-made and AI-generated differentiated texts.

Figure 1

Source text and target text relationship



Note: This figure shows the relationship between the versions of Text A (the same is true for Text B).

3.1 The choice of naturalistic language generation model

There are many naturalistic language generation models which could have been used for this study. However, as an underlying focus of this thesis is to investigate ways in which AI tools can be helpful for teachers, it is more relevant to choose a model that is widely used and is easily accessible for all. Because ChatGPT 3.5 is free, readily available, and widely used, it

was selected as the AI model for this study (Meyer, et al., 2023). It should be noted that the results may not be entirely transferable to other AI models and newer models of ChatGPT, for example ChatGPT 4.

3.2 The material

The process of selecting the material to analyse for this study began with finding suitable human-made graded readers. Following this, an excerpt from the graded reader was selected, and the corresponding excerpt from the ST was identified. The ST excerpt was then used as input alongside a prompt to generate the AI graded reader. This section describes the selection criteria and considerations that were taken during the process of selecting and generating the material of analysis.

There were two main selection criteria for the choice of texts, namely that the human-made graded readers had to be suitable for the Norwegian ESL classroom, and that the human-made graded reader should not be accessible to ChatGPT (discussed in 3.2.1). Furthermore, to aid the process of comparison, another selection criterion was that the excerpts that were selected from the human-made graded reader had to have a similar structure as the ST (see section 3.2.2). In the final section, 3.2.3, the process of generating the AI graded readers will be discussed.

3.2.1 Criteria for the selection of texts

The human-made graded readers that were selected for this study had an interest age and reading age fit for the ESL classroom at a lower secondary school level in Norway. In the beginning of the process of selecting text, the first selection criterion was that the reading material had to be suitable for the study's target group. Murty et al. (2021) highlights the importance of the distinction between reading age and interest age for the ESL context. According to Murty et al., reading age is related to the complexity of the text, while interest age is related to the maturity level, and these are not necessarily the same. Many graded readers are written for a native speaking context and may therefore have the same reading age and interest age. For example, a graded reader may be created for students at a third-grade level in a native speaking context. Based on this target group, the text complexity and the maturity of the content are both adapted with third-graders in mind. This will not work well for many

students in an ESL context, as the students may be much older than their reading age in English. Separating reading age and interest age allows students to read at their maturity level while still receiving differentiated material (Murty et al., 2021).

Based on this consideration, the human-made graded readers that were selected had to have an interest age and reading age fit for the ESL classroom at a secondary level in Norway. In the process of selecting the human-made graded readers, many books were excluded due to this criterion, mainly graded readers for native-speakers and for younger age groups. Based on recommendations by Cobb (2007), a website called Penguin Readers with books specifically made for the ESL context was found (Penguin Readers, 2024a).

After finding the penguin reader website, the next step was to select books at appropriate level from their selection of books (Penguin Readers, 2024a). A search filter was used to sort the text from the highest to the lowest reading age, and the books were checked for their interest age and the year they were produced (discussed below). The two graded readers that were selected were at an A2+ and B1+ reading level and were written for students above the age of 12 (Penguin Readers, 2024b; Penguin Readers, 2024c). The age restriction of 12+ is the lower limit, and the books explore complex themes such as religion, arranged marriage and poverty. Based on this, they were considered sufficient for the maturity and interest age of students at Norwegian lower secondary school. The reading levels were also deemed as relevant for a lower secondary school level of English. Even though the reading levels of the books are A2+ and B1+, which is lower than the expected competence of a 10th grade student indicated by the competence aims, this level of difficulty may be useful for many of the students in the diverse ESL classroom (see introduction). The students who find it challenging to read, are also arguably those who are most in need of differentiated material, as students at an expected level will have the course book, and those at a higher level can often use material written for native speakers. In addition, one needs to consider that the competence aims are *aims*, and that it is not the case that all students achieve all aspects of the competence aims. Based on this, the combined interest age and reading level of the two penguin reader books were deemed relevant for the Norwegian lower secondary school ESL context.

At this stage of the selection process another selection criterion emerged that impacted the choice of the human-made graded reader. Through researching the process of text generation and finding out that ChatGPT has access to a large amount of data from the internet and third parties, it became clear that this may impact the validity of the study (OpenAI, 2024). It is

difficult to track down and identify what ChatGPT has used to produce text. If ChatGPT has access to a graded reader version of the ST, it may reproduce elements of the graded reader text rather than generating its own version. Using this data, would then not surely show ChatGPT 3.5's true ability to create a graded reader, and would rather reflects the AI's ability to copy text.

To avoid the possibility that ChatGPT had access to a graded reader version of the ST, another selection criterion was added, namely that the human-made graded readers that were selected for this study had to be written after January 2022. ChatGPT 3.5 has a knowledge cut-off date of January 2022. This means that without add-ons or providing ChatGPT with extra information, the natural language generation tool does not have access to data produced after this date (Rice et al., 2024). Since the AI would naturally get access to the STs, as it is a part of the information provided to the chatbot, the publishing date of the STs did not matter. However, to ensure no level of reproduction of existing graded readers, the graded readers used in the analysis had to be written after January 2022. In addition, there could not be other versions of the graded reader books that were written before this date. This selection criteria excluded many classic books such as *Mrs Dalloway* (1925) and *Moby Dick* (1851), which have several graded reader versions, some of which were written before January 2022. Many of these books had higher reading levels but were nonetheless excluded. After the books were filtered by reading age, interest age and finally by the ChatGPT's cut-off date, the books that were selected were graded reader versions of *A Long Way Home* by Saroo Brierley and *Brick Lane* by Monica Ali.

3.2.2 Criteria for the selection of excerpts

After choosing the two human-made graded readers, the next step was to choose the specific excerpts from the ST and human-made target text for the analysis. The aim was to find an excerpt that could potentially be used as an excerpt in class, and which had a natural start and end. In addition, to ensure some level of comparability, another criterion was that the excerpts from Text A and B should be around the same length. Based on these criteria, a few potential excerpts were identified. However, after trying to generate an AI version of the excerpts and doing a sample analysis, a final selection criterion was added, namely that the information structure of the human-made graded reader needed to be similar to the ST.

The ST and its corresponding graded reader version had to have similar information structure to be able to compare and identify the related segments. Some sections in the graded reader version had many structural changes, for example, moving a sentence or a paragraph earlier or later in the book. This made it difficult to know whether the segments in the target text were translations of the segments in the ST, meaning that it was difficult to identify the coupled pairs (see section 3.3.1 for further explanation on coupled pairs). On top of this, these structural changes made it difficult to compare the human-made version to the AI-generated one. This was due to ChatGPT not having the opportunity to apply such structural changes across sections. In this project, ChatGPT 3.5 only had access to the section of the ST it was asked to translate intralingually, and could therefore not make structural changes outside of this section. Based on these structural considerations, the STs and corresponding human-made graded readers were carefully matched to find excerpts that had similar information structure. The final excerpts that were selected can be found in Appendix C. Both excerpts could be used as material for in-class reading and had a word count of around 750 (765 for Text A and 743 for Text B). In addition, the translations of the excerpts in the human-made graded reader versions had a similar information structure.

3.2.3 Generating the AI graded readers

When the final excerpts from the ST and human-made graded reader were selected, the next task was to generate the AI graded readers. It should be noted that due to the exploratory nature of this study, and the lack of existing set protocols, the process of prompt development was based on a process of trial and error, where the goal was generating an AI graded reader similar to the human-made graded reader. This process concerned developing a prompt that would yield a graded reader that was similar to the human-made graded reader in terms of length and word frequency distribution. In addition, the content of the AI-generated graded reader needed to make sense and reflect the content of the ST. According to Jiao et al. (2023), “The style of prompts may affect the quality of translation outputs” of ChatGPT (p. 1), so throughout the project, numerous prompts were tested on both books to reach the final prompt (see Table 1). To give an overview of the process of prompt development, two of the temporary prompts will be presented first, followed by the final prompt (see “Prompt 3” in Table 1). The development of the final prompt was done by a process of testing different inputs and outputs, until arriving at the output which was the most similar to the human-made graded reader in terms of length

and word frequency. After each generation, the output was analysed, and the text was put into VocabProfile. If the output was a lot shorter than the graded reader version, had too many low frequency words, or departed too much from the ST in content, it was discarded, and a new prompt was tested.

Table 1

Prompt Development Table

Prompt 1	Rewrite this text so that it is adapted for a B1 level of English. Make the new text approximately 90 words.
Prompt 2	Rewrite this text so that it is adapted for a B1 level of English. Good adaptations are characterized by removing the following without disrupting the content too much: complex words, figurative language, metaphors, similes, non-standard words and phrases, anaphora, canephora, exophora, ellipsis, difficult construction. Make the new text somewhere between 80 and 100 words. Simplify lexically, syntactically and remove irrelevant information. It is okey to remove whole sentences if they are not important to the plot. Be creative and reformulate the text so it sounds natural.
Prompt 3	<p>Main prompt:</p> <ul style="list-style-type: none"> Keep as much of the content as possible but rewrite this text to fit for a A2/B1 level of English. Change difficult words to simple and frequent words without changing the content. The text should be a bit shorter than the original. <p>Adjustment prompt</p> <ul style="list-style-type: none"> you use a few difficult words. With minimal changes, please replace the difficult words with easier words or remove them. For example: courtesy <p>Adjustment prompt :</p> <ul style="list-style-type: none"> Still too difficult. With minimal changes, try to make the text only consist of high frequency words.

“Prompt 1” in Table 1 was one of the prompts from the beginning of the project. An aim for the generated text was that it should be similar in length and be suitable for the target group. Based on this, a word count was included in the prompt in the hope of generating a text of a similar length and enabling a more direct comparison.¹ In addition, the reading level was included to ensure that the generated text was suitable for the target audience.

¹ Note that at this point in the project, shorter sections from the two STs were used to test out these prompts, which explains the lower word count in the first two prompts.

Following the generation from “Prompt 1”, there were two main factors that called for prompt development. The first output did not include many typical features of graded readers, such as explicitation and removal of figurative language. As a result, the prompt was changed to include some the characteristics of graded readers found by Tovar (2019). As can be seen in Table 1, this was done by adding a list of characteristics of ‘good’ adaptations to the prompt. The second major change from Prompt 1 to Prompt 2 was based on ChatGPT being too conservative in the earlier versions. In the result from the first prompt, ChatGPT made minimal changes. Finally, the three last sentences were added to give ChatGPT approval to be creative, remove irrelevant information, and even remove whole sentences (see Table 1). Nonetheless, Prompt 2, despite all of its detail, did not produce the wanted output, and was not selected. The main reasons for this were that the output generated from this prompt contained too many complex low frequency words and departed too much from the content of the ST. A hypothesis at this point was that the prompt included too many tasks and did not focus enough on simplifying the vocabulary of the text. Based on these issues, the prompt development process continued, and the final result was Prompt 3.

There were three major changes from Prompt 2 to Prompt 3. Firstly, the focus of the task was narrower. Instead of focusing on all of the characteristics of a graded reader, as described by Tovar (2019), Prompt 3 simply focused on the vocabulary. The focus on vocabulary was based on lexical control being the most common form of control in creating a graded reader according to Simensen (1990) and on the fact that a simplified vocabulary is linked to improved reading comprehension (see introduction). Changing the vocabulary was emphasised through asking ChatGPT to change difficult words to simple and frequent words, followed by asking again for it to replace difficult words with easier words or to remove them. The second major change was that prompt 3 had three parts. Instead of giving all the information in the initial prompt, by using a shorter initial prompt followed by two adjusting prompts, the AI-generated version was slightly adjusted, and then adjusted once more. After each of these three prompts, the number of low frequency words was reduced, so this was deemed a successful way of getting closer to the desired frequency distribution. Additionally, difficulty was changed from B1 to B1/A2, based on ChatGPT generally not taking the simplification far enough.

The third noteworthy change was the removal of the word limit. According to Cheng and Yu (2023), ChatGPT performed “unexpectedly poor[ly]” at a relatively simple counting task (p. 190). This was consistent with the output generated from the earlier prompts, where the text

generated by the AI rarely met the word requirements specified in the prompts. Since the word count did not seem to have an effect, it was not included in Prompt 3, the final prompt.

3.3 Method of analysis for translation strategies

The source texts (STs) and target texts (TTs) were compared to identify the extent to which selected translation strategies were used in AI-generated and human-made translations. Due to the scope of this thesis, the strategies that were accounted for were limited to the domain of lexical control. This domain was chosen based on the clear link between vocabulary and reading comprehension (see introduction), and the importance of lexical control in creating a graded reader (Simensen, 1990). The strategies chosen for the analysis were omission, explicitation, phrase/sub-clause-to-word simplification, demetaphorisation, and use of higher frequency words. These strategies were selected based on previous studies on diphasic translations and graded readers as translations, as well as, being informed by the data (See section 2.2.2). Of these strategies, two were taken from Zethsen (2009) and Hill-Madsen (2014), namely omission and explicitation, and two other strategies were based on descriptions in Tovar (2009) of lexical control, namely use of higher frequency words and demetaphorisation. The final strategy, PC-to-word simplification, emerged from analysing the data. This strategy arose from seeing the need for a category that describes the lexical simplification found when phrases and sub-clauses are translated as one word.

One of the strategies, use of higher frequency words, was investigated further in the second part of the analysis, the word frequency distribution analysis. Here the frequency analysis tool VocabProfile was used to get a more detailed image of the use of low- and high-frequency words in the various texts. The further emphasis on *this* strategy was in part based on the existence of a tool to investigate the strategy in more depth. However, this choice is also supported by the widespread agreement that text comprehension is dependent on the number of known words, often high frequency words, in a text (Cobb, 2007).

3.3.1 Coupled-pairs method

The research method of this analysis follows Gideon Toury's (2012) 'coupled-pairs method'. This is a well-known method from Translation Studies which consists of a comparative analysis of the ST and TT at a micro-level (Hill-Madsen, 2014). In this method, the ST and TTs are

mapped onto each other to identify related segments which are then analysed in terms of strategy use (Munday, 2016). These related segments are referred to as coupled pairs. Each coupled pair then consists of a segment from the ST which is being *replaced*, and a related segment from the TT, which is *replacing* the ST segment.

In the first part of the analysis, the STs were compared to the TTs to identify shifts. The comparison will be used to identify the extent of use of selected translation strategies in the human-made graded reader versus the AI-generated graded reader. According to Munday (2016), the term *strategy* in translation studies typically refers to “the overall orientation of the translated text”, meanwhile the term *procedure* is used for the “specific techniques used at a given point in a text” (p. 24). However, both Hill-Madsen (2014) and Zethsen (2009) use the term *strategy* in this narrower sense. For example, they both use the term to describe the use of explicitation which is identified as a translation procedure in Munday (2016). However, they both do specify that what they really are analysing is “micro-level translation strategies” (Hill-Madsen, 2014, p. 90; Zethsen, 2009, p. 795). Following the terminology in these articles, this thesis will use the term *strategy* meaning “micro-level translation strategies”, and will use this term to describe the specific approaches that are used at particular points in a text.

To operationalise the concept of the coupled pair, in this analysis, coupled pairs will be identified by the boundaries of the strategy: where a strategy starts and where it ends. Taking the example of explicitation, the unit of comparison in the ST and in the TT will be similar in meaning, the difference being the use of explicitation. In the ST, the unit of comparison will be what has been made more explicit: what element within the ST has been expanded on. Meanwhile, in the TT the corresponding unit will be the ST meaning plus the expanded meaning. For example, the word *bag* in the ST can become *suitcase*, or the name *Hamid* in the ST can become *Hamid, a friend* in the TT. The boundaries of the coupled pair are dependent on where the strategy, in this case explicitation, starts and ends. Since for example, *Hamid, a friend* is making the meaning of *Hamid* more explicit, they are a coupled pair.

3.3.1.1 The translation strategies

There are five translation strategies that were investigated as a part of this study. This section provides an explanation of how each of the terms have been used. The first strategy, explicitation, is used when implicit information in the ST is made explicit in the TT (Munday, 2016). According to Munday, explicitation can take many forms, including making implicit

grammatical coding, such as gender, explicit or explaining cultural items. In this thesis, the label explicitation is used to describe cases in which the meaning of a vocabulary item has been made more explicit. For example, cases in which indexical items such as *her* and *there*, are rendered more explicit by simply directly stating the referent, for example *Nazneen* and *Kawanda*. Some other examples involve abstract items such as *poverty* being rendered more concrete, translated as *homeless people*.

The label phrase/sub-clause-to-word simplification (henceforth PC-to-word simplification) is used to describe cases in which phrases and subordinate clauses in the ST have been translated into single words. Examples of this second strategy include the phrase *outskirts of the town* translated as *town*, or the clause *found myself growing* translated as *grew*. The label demetaphorisation, is used for cases in which figurative language is replaced by non-figurative language, and also often involves PC-to-word simplification. Cases where PC-to-word simplification and demetaphorisation occur together are categorised as demetaphorisation. Examples of demetaphorisation include *press of people* translated to *crowds*, *struck by* translated to *surprised* and then *flattened* translated to *devastated*.

For the last two strategies, omission and use of higher frequency word, the tool VocabProfile was used to identify the frequency level of the words in the coupled pairs (Cobb, Compleat Web VP, 2024). The frequency levels were then used to identify whether the translation could be labelled as omission or use of higher frequency word, by means of identifying when lower frequency items were omitted, or when lower frequency words were translated as higher frequency words. In translation studies, omission is typically used for any unit in the ST that is not in the TT (Munday, 2016). This is consistent with the use of the strategy in Zethsen (2009) and Tovar (2019). However, in the current study, the label omission has solely been used when a word that is of lower frequency has been omitted. This is due to the focus of this study being the lexical changes involved in creating graded readers that are suitable for the ESL classroom. Omission is then most relevant in the cases where lower frequency items are removed, as these are more likely to be unknown and perceived as difficult to the target group (see section 2.1).

The strategies of omission and use of higher frequency word were identified by inserting the ST and TTs into VocabProfile, to pinpoint changes of word frequency from the ST to the TT. VocabProfile takes text as input and gives a frequency distribution table as output. The tool grades words on a frequency scale from 1-25 based on data from the BNC COCA corpus. The corpus consists of the British National Corpus (BNC) and the Corpus of Contemporary

American English (COCA). Level one includes the 1000 most frequent families found in the corpus, and thereafter each level increases incrementally with 1000. Based on this the levels are marked with *K*, meaning thousand, so that K-1 contains the 1000 most common word families and K-25 contains the 25000 most common word families. The tool uses word families because this is a known indicator of vocabulary size. According to Bauer and Nation (1993), a word family consists of a “base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately” (p. 253). While encountering new words can require a lot of cognitive effort, encountering other family members of a known word family “requires little or no extra effort” (p. 253). If a learner knows the word *develop*, deciphering the meaning of the word *developer* then requires less effort than deciphering an unrelated word, for example *procrastinator*. The concept of word families is therefore central in calculating a vocabulary knowledge of a learner and the vocabulary size of a text. The VocabProfile tool then provides data which is useful in analysing the lexical make-up of a graded reader.

Based on the data generated from VocabProfile, all words that were omitted that had a frequency level higher than the first level (K-1), were categorised as omission. For example, the following words in the ST were omitted in both TTs and were categorised as omission: *exorbitant* (K-10 level word) and *established* (K-2 level word). For the final strategy, all coupled pairs that changed frequency level, from lower frequency to a higher frequency, were categorised as the strategy use of higher frequency word. Examples include *organized* translated as *got* and *official* translated as *man*.

3.3.1.2 Method of comparison for translation strategies

After the coupled pairs and translation strategies in the human-made and AI-generated graded readers were identified, they were added into a spreadsheet so that the different translations of the STs could be compared (see Appendix E). The data from this spreadsheet formed the basis for the analysis of translation strategies. The spreadsheet consisted of a total of 7 columns coding the coupled pairs and translation strategies. The most important columns and codes will be explained in this section.

To arrive at the comparative data found in the results section, columns for showing the coupled pairs were established in the spreadsheet. To show the difference between the human-made and AI-generated translations, these were placed side by side. For each ST unit, the

corresponding unit from the human-made translation (HM TT in the table) and the AI-generated translation (AI TT in the table) could be found in the subsequent columns. If the TT unit was omitted then the cell was marked with the symbol “-“. Table 2 below shows that the word *established* was omitted in the human-made graded reader and the AI-generated graded reader, and that the word *official* was translated to *man* in the human-made graded reader and translated to *airport staff* in the AI-generated version.

Table 2

Sample of table containing source and target comparison of coupled pairs

ST	HM TT	AI TT
established	-	-
official	man	airport staff
missing	lost	missing

The following columns, “Strategies HM” and “Strategies AI”, were added to the spreadsheet to show the strategies used for the coupled pairs in human-made and AI-generated translations (see Table 3). Here, one of the five strategies was added into the row of the corresponding coupled pair, and the cell was left blank in cases where the corresponding ST-TT pair did not display use of one of the five strategies, for example with *missing* (see “Strategies AI”). In addition, columns were added to show the number of times that the human-made TT and the AI-generated TT used the same strategies in regards to the same unit in the ST (“Strategies (Same/different)”).

Table 3

Sample of table containing source text and target texts comparison of strategies

ST	HM TT	AI TT	Strategies HM	Strategies AI	Strategies (Same/different)
missing	lost	missing	Higher frequency word		Different
there	Khandwa	Khandwa	Explicitation	Explicitation	Same
much dirtier	dirty	dirtier	PC-to-word	PC-to-word	Same

For each ST unit, one of the five strategies could be identified in both translations, one translation, or neither translation. The first two cases are included in the results data. If none of the five strategies was identified for either of the translations, then these translations were not included in the results spreadsheet. This means that each row in the spreadsheet includes an ST unit that has been replaced using one of the five strategies in either both or one of the TTs. The ST unit “missing”, in the Table 3 above, is an example of an ST unit for which a strategy was only identified in one of the TTs. The strategy of use of higher frequency word was identified in the human-made translation, while none of the five strategies were identified for the AI translation. This is shown in the strategy columns, in which a strategy, in this case use of higher frequency word, is included in the cell corresponding to the human-made translation but not for the AI translation (see “Strategies human-made” and “Strategies AI” in Table 3). The strategy column for the AI translation is then left blank as none of the five strategies were identified in relation to the ST unit.

3.4 Method of analysis for frequency distribution

To investigate the distribution of higher and lower frequency words, each section, namely excerpts from the STs, the two human-made graded readers, and the two AI-generated graded

readers, were all put into VocabProfile (see section 3.3.1.1 for a description of how the VocabProfile tool works). The website gives several options for which corpus the data generation should be based on. The option chosen for this project was the Nation's BNC-COCA frequency list (Nation, 2012). This list is often used for L2 research, and to set vocabulary goals for L2 learners (Skjelde, 2015). In addition to showing the distribution of the frequency values, the VocabProfile shows the specific words that belong to the different categories (1-25) and shows for which frequency level in which 95% and 98% of the words of the text belong. All of these three visualisations of the frequency values were used in the analysis of the texts.

4 Results

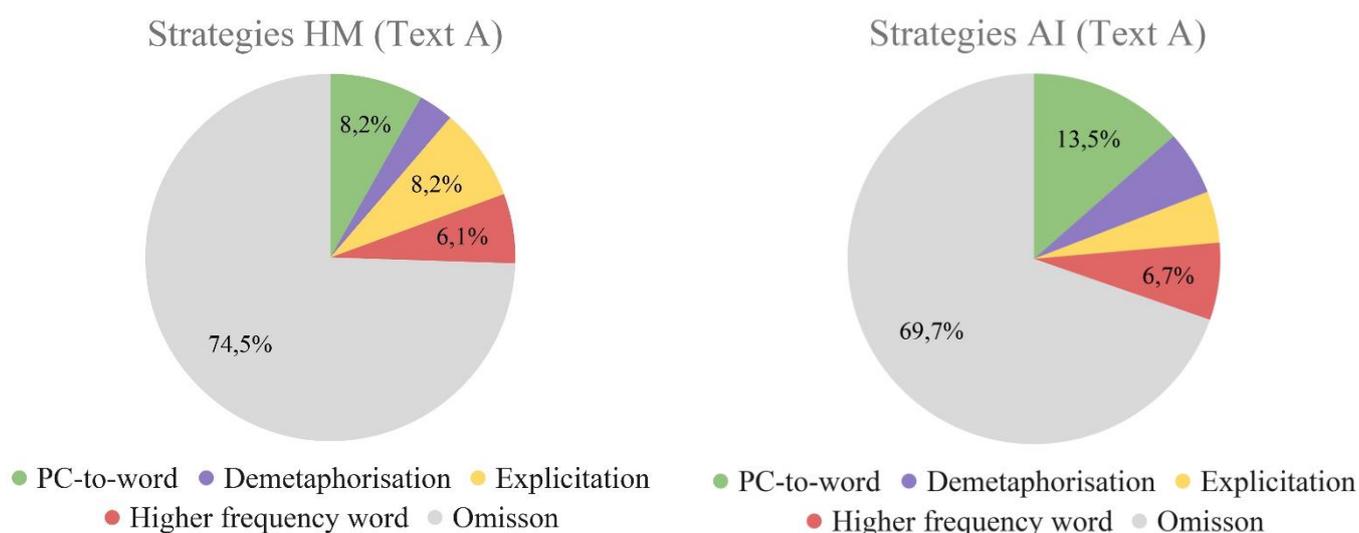
4.1 Results of the analysis of translation strategies

The first research question of this thesis relates to the use of translation strategies in the graded readers. Specifically, the question concerns to what the extent selected translation strategies are used in the excerpts of human-made and AI-generated graded reader versions of Text A and Text B, as well as how these results compare. The selected translation strategies are omission, PC-to-word simplification, demetaphorisation, explicitation and use of higher frequency word. This section will discuss each strategy separately before summarising the overall strategy use in the adapted texts.

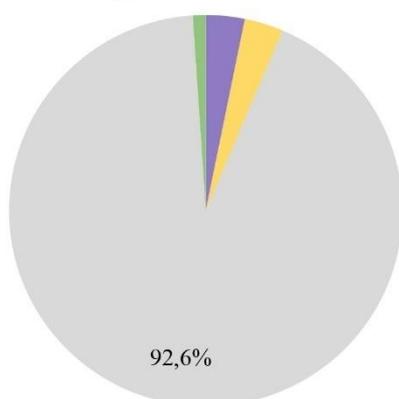
Figure 2 below show the results for the first research question and shows the occurrences of all five translation strategies for each of the four TTs. Note that these results only pertain to the five strategies investigated in this thesis, so formulations such as “using a strategy” refers to using one of these five strategies.

Figure 2

Distribution of translation strategies in human-made (HM) and AI-generated graded readers

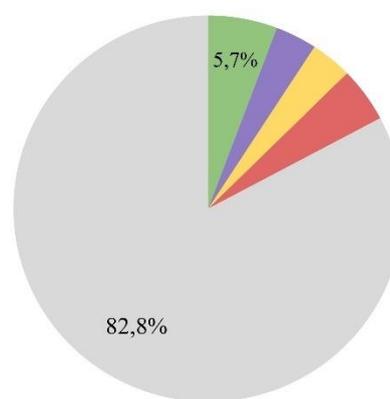


Strategies HM (Text B)



- Demetaphorisation ● Explication ● Omission
- PC-to-word

Strategies AI (Text B)



- PC-to-word ● Demetaphorisation ● Explication
- Higher frequency word ● Omission

4.1.1 Omission

As we can see in Figure 2, omission was the most used strategy for all four translations. A pattern that emerged regarding the use of omission was that human-made versions used more omission than the AI counterparts. This can be seen in Figure 2 above, in which the strategy was used more in the human-made versions, seen on the left, than in the AI-generated ones, seen on the right. The strategy of omission was most prevalent in the human-made version of Text B, where it was found in 92.6 % of the coupled pairs. Omission was also prevalent in the AI translation of the same text and was found in 82.8% of the coupled pairs identified. For translations of Text A, the strategy was somewhat less prevalent, found in 74.5% of coupled pairs of the human-made translation, and 69.7% of the AI-generated translation. This means that the translations of Text B both used more omission than the translations of Text A. In addition, these numbers show that there were greater differences in use of omission between the human-made and AI-generated versions of Text B (a difference of 9.8% for Text B compared to 4.8% for Text A).

As stated above, there was a notable difference between the human-made and AI-generated translations concerning the use of omission. These differences concerning omission will now be discussed in more detail. There were longer stretches of consecutive use of omission in the human-made translations compared to the AI-generated translations (see Table 4 below). For Text A the average stretch of omission was 8.1 while for the AI-generated version it was 4.1. This means that the strategy was on average used around 8 times before another strategy was

used in the human-made version. In the AI-generated version, on the other hand, the strategy was only used around 4 times in a row before another strategy was used. A similar pattern emerged with the translations of Text B, in which the average stretch of omission for the human-made version was 17.4, while for the AI-generated version it was 8.0. In the human-made translation, omission was then used around 14 times in a row before another strategy was used, while for the AI-generated version, omission was used around 8 times in a row. This means that the average span of consecutive use of omission was around double in the human-made translations compared to the AI-generated translations. The trend was then that the AI-generated translation alternated more between omission and other strategies, creating shorter stretches of repeated use of omission, while the human-made translation alternated less between omission and the other four strategies, creating long stretches of omission.

Table 4

Sample of the table from the analysis of Text A showing the different uses of omission

Strategies HM	Strategies AI
Higher frequency word	Omission
Explicitation	Omission
PC-to-word	Omission
Omission	Demetaphorisation
Omission	Omission
Omission	PC-to-word
Omission	Omission
Omission	Higher frequency word
Omission	Omission
Omission	Demetaphorisation
Omission	Omission
Omission	PC-to-word
Higher frequency word	Omission
PC-to-word	Omission

Note: This sample shows the longer stretches of omission in the human-made (HM) graded reader compared to the AI-generated graded reader. Omission occurs the same number of times in the human-made and AI-generated translations, but the distribution pattern is different.

A similarity between human-made and AI-generated translations was that both translations regularly used the strategy of omission to translate the same items in the ST. By putting the

translations of the STs side by side, the strategies of the coupled pairs were coded so that if the same strategy had been used to translate the same ST unit in both translations, it would be marked as ‘Same Strategy’ (see section 3.3.1.2). Out of all of the times that the AI-generated and human-made translations used the same strategy for the same ST unit, 93.9% of the instances were omission. Omission was, for example, used to translate the word *scavenging* in both human-made and AI-generated translations of Text A. Despite this similarity, the comparison between human-made and AI-generated translations in regards to omission reveals a discrepancy, most notably the frequency and consecutive use of the omission strategy, being used more frequently and more consecutively in human-made translations.

4.1.2 Phrase/sub-clause-to-word simplification

As can be seen in Figure 2, PC-to-word simplification was the second most used strategy in all translations, with the exception of the human-made translation of Text B. For PC-to-word simplification, a pattern concerning human-made versus AI-generated translations was also identified. In contrast to the strategy of omission, which was more prevalent in human-made translations, PC-to-word simplification was more frequent in the AI-generated translations than in the human-made translations of the same texts. In the translations of Text A, the strategy of PC-to-word simplification was identified in 13.5% of the coupled pairs from the AI-generated translation, and identified in 8.2% of the coupled pairs from the human-made translation. For translations of Text B, PC-to-word simplification was identified in 5.7% of the coupled pairs in the AI-generated translation compared to 1.1% in the human-made translation. For Text B, there were in addition no instances in which both translations used PC-to-word simplification for the same ST unit. For translations of Text A, on the other hand, there were two instances of same strategy use. Both translations of Text A used PC-to-word simplification to translate the ST unit *much dirtier* (translated as *dirty* in the human-made translation and as *dirtier* in the AI translation) and *the outskirts of the town* (translated as *town* in both translations).

4.1.3 Demetaphorisation

As can be seen in Figure 2, demetaphorisation was one of the lesser used strategies in all translations. One pattern that arose concerning demetaphorisation was that the strategy was slightly more prevalent in the AI-generated translations than in the human-made translations.

For Text A, demetaphorisation was identified in 3.1% of coupled pairs in the human-made translation, while identified in 5.6% of coupled pairs in the AI-generated translation. This difference was smaller for translations of Text B. For Text B, demetaphorisation was identified in 3.2% of coupled pairs in the human-made translation and 3.4% in the AI translation. This means that demetaphorisation was more prevalent in the AI-generated version of Text A, and slightly more prevalent in the AI-generated version of Text B. These numbers also show that there was no pattern of prevalence in regard to which ST had translations with more demetaphorisation ($3.2 > 3.0$ but $5.6 > 3.4$), and that the differences between the AI and human-made translations were greater for translations of Text A than for Text B (Text A ($5.6 - 3.0$) > Text B ($3.4 - 3.2$)). Finally, both the translations of Text A and Text B had instances of “same strategy” use for demetaphorisation. There was one instance in Text A and two in Text B. An example is *rush of adrenaline* from Text A which was translated as *exciting* in the human-made translation and *excited* in the AI-generated translation.

4.1.4 Explication

As can be seen in Figure 2, explication was also one of the lesser used strategies, with the exception of its use in the human-made translation of Text A, where it was used in 8.2% of coupled pairs. For explication, there was no clear pattern when comparing the human-made and AI-generated translations, as explication was used *more* in the human-made translation than in the AI translation of Text A (8.2% compared to 4.5%), while used *less* in the human-made compared to the AI translation of Text B (3.2% compared to 3.4%). The strategy was used more frequently in the translations of Text A, with 8.2% for the human-made translation and 3.2% for the AI-generated translation, than in Text B, with 4.5% for the human-made translation and 3.4% for the AI-generated translation. In addition, these numbers show that for explication there were greater differences between the human-made and AI-generated translations of Text A, than for Text B. Furthermore, there were three instances of same strategy use for explication, two of which were found in the translations of Text B. Some examples include *there* being translated as *Khandwa* in both translations of Text A and *her* being translated into Nazneen's in both translations of Text B.

4.1.5 Use of higher frequency word

As can be seen in Figure 2, use of higher frequency word was the only strategy that did not occur in all translations. In addition, for use of higher frequency word, the strategy was slightly more prevalent in AI-generated translations. For translations of Text A, the strategy was identified 6 times in both translations, equating to similar percentages, with 6.1% in the human-made translation and 6.7% in the AI-generated translations. For Text B, the strategy was not identified in the human-made translation but was identified in 4.6% of coupled pairs for the AI-generated translation. As these numbers show, use of higher frequency word was more prevalent in the translations of Text A, than for Text B, while there were greater differences between the translations of Text B than the translations of Text A regarding the prevalence of the strategy. Finally, for Text A, there were in addition two cases where the strategy was used for the same ST unit. An example of this was the translation of *organized* as *got* in both human-made and AI-generated translations.

4.1.6 Summary of translation strategy results

When comparing human-made and AI-generated translations, there were two strategies that had clear patterns of prevalence. Omission was more prevalent in human-made translations, while PC-to-word simplification was more prevalent in the AI-generated translations. Demetaphorisation and use of higher frequency word were also more prevalent in AI-generated translations taken together. However, both strategies had a small difference for one of the STs, less than 1%, so the pattern of prevalence (AI>HM) is weaker. For explicitation, no prevalence pattern was identified for the human-made versus AI-generated translation distinction. There were also similarities between human-made and AI-generated translations, most notably 131 cases of same strategy use compared to 83 instances of different strategy use across both STs.

In comparing translations of Text A and B, other patterns emerged concerning strategy prevalence. One of the five strategies was more prevalent in the two translations of Text B, namely omission. PC-to-word simplification, explicitation and use of higher frequency word were all more prevalent in the two translations of Text A. There was no pattern for demetaphorisation. Additionally, as can be seen in Figure 2, a difference between the translations of the two STs, was that the five translation strategies were more evenly distributed for translations of Text A, than for Text B, where omission dominated to a greater extent.

Finally, for a majority of strategies the difference between human-made and AI-generated translations was greater between translations of Text B than for Text A.

4.2 Results of the analysis of word frequency distribution

The second research question that this study aimed to investigate was how the word frequency distribution compared between the human-made and AI-generated adaptations of Texts A and B. In order to determine what kind of vocabulary size would be needed to attain the 95% and 98% coverage levels needed for reading comprehension, the vocabulary profile of the text was analysed. Table 5 shows a simplified version of the frequency distribution tables found in Appendix F. This table gives an overview of the word frequency distribution through showing the 95% and 98% coverage thresholds. These coverage values, marked in green, show the highest frequency value (K-1, K-2, etc.) for which 95% and 98% of the vocabulary in each text is covered (see section 2.2 for the significance of these thresholds, see 3.3.1.1 for descriptions of VocabProfile). This means that for both translations of Text A, 95% of the words were at a K-2 level or lower, meaning that 95% of the words occurring in these translations were of the 2000 most frequent word families in the corpus. For further information, the frequency distribution tables can be found in Appendix F.

Table 5

Frequency distribution tables for source texts and target texts

Text A			Text B		
ST	HM	AI	ST	HM	AI
K-1	K-1	K-1	K-1	K-1	K-1
K-2	K-2	K-2	K-2	K-2	K-2
K-3	K-3	K-3	K-3	K-3	K-3
K-4	K-4	K-4	K-4	K-4	K-4
K-5	K-5	K-5	K-5	K-5	K-5
K-6	K-6	K-6	K-6	K-6	K-6
K-7	K-7	K-7	K-7	K-7	K-7

Coverage 95	Coverage 98
-------------	-------------

Note: This is a simplified version of frequency distribution tables for source texts (ST), human-made translations (human-made) and AI-generated translations (AI).

There were three main findings related to word frequency. The first finding was that the proportion of high frequency words increased in both human-made and AI-generated graded readers compared to the source texts. The second finding was that human-made translations diverged further from the ST, containing more high frequency words than the AI-generated counterparts. Finally, the third finding was that this difference was greater between AI-generated and human-made translations of Text B than it was for the translations of Text A.

As indicated by the 95% coverage levels, Text B had a lower frequency vocabulary than Text A. This can be seen in Table 5 above, where the 95% coverage level is at K-2 for Text A, while at a K-4 level for Text B. Altogether, the word frequency distributions of the AI-generated translations are closer to the STs than the human-made translations. This can be seen in the table above, in regard to translations of Text B, but is missing in this simplified table in regard to translations of Text A, where the differences are more subtle. The human-made translation of Text A, did however, consist of a higher percentage of the most frequent words, K-1 words, with 92.4% compared to 86.4% for the AI translation, indicating a slightly higher proportion of the most frequent words in the human-made translation compared to the AI-generated translation (see Appendix F).

The 95 and 98 coverage values also indicate that there are greater differences between the AI-generated and human-made translations of Text B than the translations of Text A. The AI-generated and human-made translations of Text A have the same 95% and 98% coverage lines, at K-2 and K-3 respectively. This means that they are similar in their word frequency distribution profiles. Meanwhile, for Text B, the coverage 95% line for the human-made translation was at K-1 and for the AI-generated translation it was at K-2, indicating a greater difference.

5 Discussion

The aim of the study is to investigate whether ChatGPT 3.5 can be used to make differentiated reading material for the English as a second language classroom, and whether these AI generated texts differ from human-made graded readers. The study's focus is the vocabulary choices involved in creating graded readers, as this is described as a central aspect in reading comprehension and differentiated reading material (see introduction). To investigate these vocabulary choices, two lexical aspects of the human-made and AI-generated graded readers were analysed, namely lexical translation strategies and word frequency distributions. Looking at both translations of Text A and Text B, the results of this study were that all of the selected strategies were found in both human-made and AI-generated graded readers. However, the results indicate some differences between human-made and AI-generated graded readers. For example, omission was used more in the human-made graded readers while PC-to-word simplification was used more in AI-generated graded readers. Furthermore, the word frequency distribution analysis results indicate that the human-made graded readers overall used a higher proportion of high frequency words than the AI-generated graded readers, even though ChatGPT was asked numerous times to use vocabulary items that are easy, simple and of high frequency (see 3.2.3)

This chapter consists of a discussion of the results, as well as the implications and limitations of this study. In the first three sections, 5.1-5.3, the results will be discussed. In the final two sections, the implications of the study (section 5.4), and the limitations of the study (section 5.5) will be discussed.

5.1 Translation strategies in human-made graded readers

The results show that four of the selected strategies, namely omission, PC-to-word simplification, explicitation and demetaphorisation, were found in both human-made translations. The final strategy, use of higher frequency word, was only identified in the human-made translation of Text A and not identified in the human-made translation of Text B (see 4.1.5). The results regarding strategies used in the human-made graded readers were in part consistent with a prior analysis of a graded reader by Tovar (2019), as well as analyses of diaphasic translations by Hill-Madsen (2014) and Zethsen (2009).

The most used strategy in both human-made translations was omission (see 4.1.1). The abundance of this strategy is consistent with findings from Tovar (2019), in which he states that a norm for creating graded readers is to cut all peripheral information. In addition, the long stretches of repeated use of omission in the human-made graded readers of this study is also consistent with Tovar's findings, as he states omission of full sentences and paragraphs is the most commonly used technique for graded readers (see 2.2.2). If a sentence containing five lower frequency words was removed from the ST, this would cause the categorisation of five consecutive instances of omission in the analysis. With this in mind, the long stretches of repeated use of omission may be caused by the omission of several sentences. This is reflective of the human-made translations analysed in this thesis, in which, for instance a whole section of Text A describing the main characters journey through a town to a railway was removed from the human-made translation (see Appendix C and E). The consecutive use of omission found in this study is then consistent with Tovar's findings on the norms of graded readers, in which graded readers tend to involve the removal of full sentences and paragraphs. In addition, Tovar (2019) states that another norm of graded reader translations is to remove unfamiliar words and non-standard language. Based on this, the high amount of omission of lower frequency words found in the human-made graded readers is again consistent with Tovar's findings.

PC-to-word simplification was not mentioned as a strategy for graded readers by Tovar (2019), however the results of this study indicate that this strategy can be used in human-made graded readers (see 4.1.2). The use of PC-to-word simplification is in part in line with Hill-Madsen (2014), who highlights examples in diphasic translations in which phrases are translated into words, for example the phrase *excipients and active substance* translated to the word *ingredients*. In addition, Zethsen (2009) describes lexical simplification as a strategy for diphasic translations. As PC-to-word simplification can be viewed as a form of lexical simplification, the results of this study concerning PC-to-word simplification are partly in line with Zethsen's descriptions as well.

Concerning the use of explicitation, the occurrence of the strategy in the human-made graded reader is consistent from studies on diphasic translations by both Tovar (2019) and Hill-Madsen (2014). Although Hill-Madsen (2014) mentions explicitation as a core strategy for expert-to-layman diaphasic translations, Tovar's (2019) findings indicate that it is one of the rarer strategies in graded readers. The use of the strategy is therefore consistent with both Hill-Madsen (2014) and Tovar (2019), however, the frequency of use, especially in the graded

reader version of Text A, is more consistent with Hill-Madsen's findings (see 4.1.3). In this translation, explicitation, alongside PC-to-word simplification, was the second most prevalent strategy, used in 8.1% of coupled pairs. The use of demetaphorisation in both human-made translations is also consistent with Tovar (2019) who identified that figures of speech such as metaphors and similes tended to be omitted in the graded reader that he analysed (see 4.1.4).

However, the lack of instances of use of higher frequency words in the human-made graded reader of Text B is inconsistent with Tovar (2019). Tovar identified changes of this sort, from unknown to known words, as a norm of graded reader translations (see section 2.2). The material used for analysis from the graded reader version of Text B was, however, a lot shorter than Tovar's material, which was three chapters long. The length of the section may have impacted the presence of this strategy. By analysing a longer section of the human-made graded reader version of Text B, the strategy may have been identified. This is supported by the word frequency distribution results, where there was a clear reduction in the use of low frequency words from the ST to the human-made graded reader (see 4.2). This in turn points to a use of higher frequency words in the TT, even though it was not identified in the coupled pairs analysed.

Based on this study's findings and prior research, one can expect graded readers to contain the strategies that were analysed in this study, although the prevalence of these strategies may vary. In particular, this study and prior research on graded readers point to omission as the most prevalent strategy in graded readers. Based on this, teachers should be aware that ST information, in particular peripheral information such as descriptive sections, may be omitted in graded readers. These cuts have a purpose and have benefits for readability, but may impact the content of the text and this should be kept in mind when using graded readers in the classroom. Furthermore, the use of the translation strategies may impact the writing style, for example by removing figurative language. If an ST is selected to be used as reading material based on its use of figurative language, a graded reader version may not work as well for the same purpose.

5.2 Comparison of translation strategies in AI-generated and human-made graded readers

There were some notable similarities between the human-made translations and the AI-generated translations. Namely, all strategies were identified at some point in both human-made and AI-generated translations and omission was the most used strategy in all translations. In addition, ChatGPT often used the same strategy as the humans when translating specific ST units. Across both STs, there were 131 instances in which ChatGPT used the same strategy as the human translators and 83 instances where it used a different strategy to the human translators (see 4.1.6). All together this indicates that ChatGPT 3.5 has the capacity to simplify language with some similarities to humans.

The results also indicate some differences between human-made and AI-generated translations. The main differences identified in AI-generated translations compared to the human-made translations concerned the use of omission and PC-to-word-simplification (see section 4.1.6). Omission was more prevalent in human-made translations while phrase-to-word simplification was more prevalent in AI-generated translations. In addition, in regard to omission, the AI-generated translations alternated more between omission and other strategies, while the human translation alternated less between omission and the other four strategies, causing long stretches of omission (see 4.1.1). This difference is exemplified by the two alternative translations of a section from Text A. This section, describing the main character's journey through a town to a railway, was removed from the human-made translation (see Appendix C). For this section, omission was the only strategy identified (identified 23 times) in the human-made translation. In the AI-generated translation, on the other hand, each strategy was identified in this section, and omission was identified in between the use of the other strategies (see rows 88-114 in Appendix E).

Prior findings on human-made graded readers and use of ChatGPT for creating simplified text, may shed light on these differences. As explained in section 5.1, in line with Tovar (2019), the long stretches of omission identified in the human-made graded reader, may be due to the translator choosing to omit full sentences and sections. In light of this, the more frequent alternation between omission and other strategies found in the AI translations may indicate that ChatGPT is less likely to omit full sentences and sections. Since the human-made and AI-generated translations are similar in length, the question remains of what alternative strategy ChatGPT relies on to shorten the text to the same length.

The answer may be explained by findings by Young and Shishido (2023a) on the omission tendencies of ChatGPT when creating reading material for the EFL context. Young and Shishido (2023a) found that reading material produced by ChatGPT for an EFL setting, had fewer words per sentence than the original (see 2.3). To attain fewer words per sentence one can either replace several words with fewer words, for example through PC-to-word simplification, or one can simply omit words within the sentence. This tendency of ChatGPT may then explain the higher use of PC-to-word simplification and the shorter stretches of repeated use of omission in the AI-generated graded readers analysed in this study. In context of the fewer words per sentence finding, the higher use of PC-to-word simplification may be a means of reducing words per sentence through replacing several words with one. Additionally, omission may be a means of reducing words per sentence by removing words within the sentence. In light of Tovar (2019), Young and Shishido (2023a), and this study's results, a possible difference between human-made and AI-generated graded readers, may be that human-made graded readers remove larger chunks of text, while AI-generated graded readers tend to shorten the text by shortening sentences.

However, these differences might have been impacted by the final prompt. In the final prompt ChatGPT was asked to keep as much of the original content as possible. Based on this prompt, it is feasible that ChatGPT solved the task of keeping as much as possible by shortening sentences rather than removing them. Further research on ChatGPT's strategies for shortening text is therefore needed to know whether this is a real difference between the human and AI text simplification process.

Furthermore, the ST seemed to have an impact on the use of translation strategies. As described in the results section, patterns were identified between the ST and the prevalence of different strategies. Omission was more prevalent in both human-made and AI-generated translations of Text B and PC-to-word simplification, explicitation and use of higher frequency word were more prevalent in translations of Text A. In addition, the strategies seemed to be more evenly distributed in translations of Text A than in translations of Text B (see Figure 2). These results indicate that the ST had an impact on the translation strategies used in the TTs. In addition, there were greater differences between the human-made and AI-generated translations of Text B, compared to Text A. This may indicate that the ST has an impact on whether the AI can produce a human-like graded reader, by means of using the same strategies as humans. The section of Text B that was used as material for the AI translation consisted of more lower frequency words than the section from Text A and is therefore arguably more difficult (see

4.2). A possible explanation for the greater difference between the strategy use in the human-made and AI-generated translations of Text B, may then be that a more difficult ST is more challenging for ChatGPT to simplify (a position that is supported by the frequency distribution results, discussed in 5.3).

Based on the possible differences between human-made and AI-generated graded readers discussed in this section, it may be fruitful for teachers to keep in mind that ChatGPT seems to shorten texts differently from human translators. The findings suggests that ChatGPT shortens the text by shortening sentences, while human translators remove sentences and sections of the text deemed peripheral. If we assume that ChatGPT does not understand the text like a human translator, the tool will then not know what is core and what is peripheral content. Therefore, ChatGPT may not know what is important to keep and what can be removed. A possible issue that may arise when generating graded readers with ChatGPT, is then that peripheral sentences and sections may be given as much weight as core sentences and sections from the ST. A solution to this problem could be to highlight the most important parts of the text, ask ChatGPT to not shorten the respective sections too much, and potentially even generate a graded reader version of the most important sections separately to ensure more control over the translation of the section.

Furthermore, it may be useful for teachers to keep in mind that the ST may impact the translation strategies chosen by ChatGPT. The findings of this study suggest that the strategies used by ChatGPT seem to be dependent on the nature of the source text. The source text may then impact the way in which ChatGPT translates and therefore the graded reader it produces. It may be useful for teachers to test out different STs as input to find an ST that can successfully be translated into effective differentiated material for L2 students.

5.3 Discussion of frequency distribution

In terms of word frequency, the analysis showed that the human-made graded readers contained a greater proportion of higher frequency words than the original texts, which is in line with previous research (Nation & Wang Ming-Tzu, 1999; Cobb, 2007). The analysis also showed that AI-generated graded readers contained a greater proportion of high frequency words compared to the STs (see 4.2). However, this result may have been impacted by the final prompt, in which ChatGPT was specifically asked to use high frequency words. Nevertheless,

these results are consistent with Haver et al. (2024) who found that ChatGPT improved readability and reading ease when asked to simplify medical information for children, and Young and Shishido (2023a) who found that reading material generated by ChatGPT for the EFL classroom was less complex than the original. Overall, results and prior research indicate that both human translators and ChatGPT increase the proportion of higher frequency words when asked to simplify text.

For the AI-generated translation of Text A, the reduction of lower frequency words was similar to the human-made graded reader version (see 4.2). This AI-generated translation had the same 95% and 98% coverage results as the human-made translation. The results, therefore, indicate a similar ability to increase the word frequency of the vocabulary of the text between humans and ChatGPT when asked to simplify text. This marks a shift from Moons and Van Bulck (2023), who found that ChatGPT did not manage to simplify text to a 6th grade reading level, which is equivalent to an A2 reading level (Young & Shishido, 2023b). However, they investigated ChatGPT's ability to simplify medical information and not fiction, which may explain the different result. In addition, there were *some* differences between the frequency distributions of the texts. For example, the human-made version of text A had more K-1 words than the AI-generated version (see 4.2 and Appendix F). Overall, the difference in the human-made and AI-generated translation of Text A were so small that they may be due to chance. Further analysis of a larger amount of text is necessary to know whether ChatGPT has the ability to simplify text to an A2/B1 reading level.

For translations of Text B, on the other hand, the AI-generated graded reader did not reduce the amount of low frequency words as much as the human-made graded reader of the same ST. The result for the AI-generated translation of Text B is consistent with Moons and Van Bulck (2023), who found that ChatGPT did not reach the desired reading level when asked to simplify medical text for children. They argue that this finding was not surprising since ChatGPT is designed to have a reading level comparable to a native English-speaking high school student, indicating that ChatGPT can only simplify text to a certain level. Furthermore, their findings indicated that the complexity of the original text had an impact on the complexity of ChatGPT's output. A higher ST reading level was associated with a greater improvement in readability in the TT. This is consistent with the results from the frequency distribution analysis in this study, in which the changes from ST to the AI-generated graded reader of Text B were greater than for Text A (see 4.2). Nevertheless, the AI-generated translation of Text A was closer to the human-made graded reader. This is reflective of Moons and Van Bulcks (2023) results as well.

They found that the simplified version of the more complex ST (17th grade level) had greater improvements in readability, but the simplified version of the less complex ST (11th grade level) was closer to the desired reading level (6th grade level). In light of prior research, the results from the frequency distribution analysis may indicate that ChatGPT can decrease the amount of low frequency words of a text, and that the complexity of the text that it simplifies impacts the frequency distribution of the output. A simpler ST may make it easier for ChatGPT to reach the desired complexity, as the text is already quite simple, while more complicated STs may allow ChatGPT to simplify to a greater extent, since there is a greater gap between the ST's difficulty and the desired difficulty. Further analysis of a larger amount of material is nevertheless necessary to determine the accuracy of this claim.

In line with the translation strategy results, the frequency results indicate that the ST seems to have an impact on ChatGPT's output. Once again, it may be useful for teachers to be aware of this when producing differentiated material with ChatGPT. Potentially, the differences between Text A and Text B may explain the word frequency results of their respective AI-generated translations. Text B consisted of a larger proportion of low frequency words and more figurative language, which may explain the frequency result. In line with Moons and VanBulck (2023), there may be a threshold for how far ChatGPT can simplify text. A possibility is then that difficulty of the ST impacts the success of the AI translation. Based on this, when generating graded readers, teachers may have to avoid more complicated STs consisting of many low frequency words and figurative language, as ChatGPT may not be able to simplify these texts to the desired level.

5.4 Implications

The results from this study and previous findings suggest that ChatGPT 3.5 has the capability to simplify text and change the vocabulary of the text from lower frequency to higher frequency. In addition, the similarities in the use of translations strategies, may indicate that ChatGPT simplifies in a somewhat similar way to humans in terms of what and when specific strategies are used (see 5.2). Based on this, the tool could be useful in an ESL context to create differentiated reading material. The tool can be used in the same way it was used in this study or new applications could be tested. A possibility would be to test ChatGPT's ability to create differentiated texts that specifically introduced new vocabulary items at an ideal rate. As mentioned in the introduction, to learn from the context, 95-98% of words in the text need to

be known (Nation & Deweerdt, 2001), but in addition to this, for word learning to occur, a word needs to be encountered 6-10 times (Cobb, 2007). This means that a beneficial way of learning words from reading is to introduce the same unknown words multiple times within a text of mainly known words. Based on this, ChatGPT could be asked to make high frequency texts that introduce the same lower frequency words repeatedly.

Despite these possible applications, there are several possible problems with using ChatGPT for generating reading material. The first possible limitation of using ChatGPT to create graded readers, is the way it shortens the text. Instead of removing sections deemed as peripheral in the way that human-made graded readers do, ChatGPT seems to shorten the text more evenly, omitting words within sentences and replacing phrases and clauses with words (Tovar, 2019; Young & Shishido, 2023). Shortening this way may cause important information from the ST to be given as much weight as peripheral information, which can cause problems for the storyline. The second problem concerns the word frequency results. In line with previous findings, the results from this study indicate that ChatGPT may not reach the desired reading level when asked to simplify text, especially STs consisting of low frequency words. Finally, there were additional issues identified in the AI-generated texts that go beyond the scope of the research questions. These include information inconsistent with the ST, issues with shortening the text too much, and unnatural formulations. In line with the recommendations of Haver et al. (2024), these problems call for proofreading by professionals, in this case teachers, before the texts are provided to students.

5.5 Limitations

A limitation of this study is its scope. The material analysed may not be representative of human-made graded readers or of the abilities of ChatGPT 3.5 to generate graded readers. There can be great variability in the way that texts are translated. This difference can even be seen in the results of the two human-made translations that were analysed (see 4.1 and 4.2). Analysing excerpts from two graded readers is then not sufficient to gain generalisable knowledge on the translation strategies or on the word frequency distribution patterns of graded readers. The same can be said for the AI-generated material used in this thesis. Considering all the possible prompts that could have been used and that one prompt can generate different outputs, there is a huge amount of variability in the output from ChatGPT (see 3.2.4). In addition, the choice of the excerpt source text may have impacted the output as well. The

material analysed is then not sufficient to generalise on ChatGPT 3.5's ability to generate graded readers. To account for this variability, further research at a larger scale is needed. Comparing several graded readers, choosing longer excerpts, and testing out the results of several prompts would provide better evidence of the differences between human-made and AI-generated graded readers.

6 Conclusion

The goal of this study was to investigate whether ChatGPT 3.5 can be used to make graded readers for ESL students, and whether these AI generated texts differ from human-made graded readers. As indicated by the similarities in the results between human-made and AI-generated translations, ChatGPT appears to have the potential to be used to create graded readers. Nevertheless, there are notable differences between the vocabulary choices of ChatGPT compared to humans, which teachers need to be aware of.

A similarity was that all the selected translations strategies were found in both human-made and AI-generated graded readers. Another similarity was that omission was the most common strategy in all four translations analysed. The main differences between human-made and AI-generated translations pertained to the prevalence of PC-to-word simplification, which was more common in AI-generated translations, and the use and prevalence of omission, which was more prevalent in human-made translations. As discussed in section 5.2 and 5.4, these findings may indicate differences between the way that humans and ChatGPT 3.5 shorten text when asked to create a graded reader.

Compared to the STs, the word frequency distribution of human-made translations shows a decrease in low frequency words, and so does the frequency distribution of the AI-generated translations. For the ST that contained more low frequency words, Text B, the human-made translations reduced the number of low frequency words more than the AI-generated translation, while for Text A, the differences were almost non-existent. The results for translations of Text A mark a possible shift from prior research that concluded that ChatGPT did not reach the desired reading level. However, the results for Text B were consistent with prior research on ChatGPT's capabilities and may indicate the role of the ST in ChatGPT's success as a diaphasic translator.

Although these findings are not sufficient in themselves to generalise on the full capabilities and shortcomings of AI-generated graded readers, they provide a snapshot of its potential. To gain proper insight into how ChatGPT changes the vocabulary of a text in comparison to humans, further research is necessary. For future research, the number of texts and the length of texts analysed needs to be increased to account for the variability in the graded readers, especially the generated ones (see 5.4). In addition, testing out several prompts will be necessary to gain better insight into ChatGPT's capabilities. Furthermore, other aspects of AI simplification should be investigated such as syntax and content. Finally, the effect of these

texts on students learning needs to be investigated. The question remains as to whether the graded readers produced by ChatGPT have the same learning benefits as human-made graded readers.

The findings of this study suggest that ChatGPT 3.5 effectively simplifies text by employing the same lexical strategies as humans, for example by increasing the proportion of high frequency words. However, the findings also indicate that there are some significant differences between AI-generated and human-made graded readers. ChatGPT 3.5 seems to not deal as effectively with simplifying more complex texts and the tool seems to shorten the text more evenly rather than removing peripheral information. Consequently, while ChatGPT 3.5 can serve as a valuable tool for educators who wish to use the tool to generate graded reader material, they may need to make adjustments to AI-generated graded readers in order to compensate for ChatGPT 3.5's distinct simplification style.

References

- Ali, M. (2003). *Brick lane*. London: Black Swan Edition.
- Baker, M. E. (2001). Shifts of translation. In M. Baker, *Routledge Encyclopedia of Translation* (pp. 226-231). London: Routledge.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279.
- Borji, A. (2023). A categorical archive of chatgpt failures. *arXiv*.
- Brems, E. (2018). Separated by the same language: Intralingual translation between Dutch and Dutch. *Perspectives*, 26(4), 509-525.
- Brierley, S. (2013). *A long way home*. Viking Australia Publisher.
- Cain, K. (2007). Deriving word meanings from context: Does explanation facilitate contextual. *Journal of Research in Reading* 30(4), 347–359.
- Cheng, V., & Yu, Z. (2023). Analyzing ChatGPT's mathematical deficiencies: Insights and contributions. *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing* (pp. 188-193). ROCLING 2023.
- Chiang, M.-H. (2015). Effects of varying text difficulty levels on second language (L2) reading attitudes and reading comprehension. *Journal of Research in Reading*, pp. 1-21.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology* 11(3), 38-63.
- Cobb, T. (2024, January). *Compleat Web VP*. Retrieved from Lextutor: <https://www.lexutor.ca/vp/comp/>
- Council of Europe. (2024). *Self-assessment grid - Table 2 (CEFR 3.3): Common Reference levels*. Retrieved from Council of Europe: <https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale>

- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure: How do word-related variables and proficiency influence receptive vocabulary learning? *Language Learning*, 70(2), 349-381.
- Dong, C., Li, Y., Gong, H., Chen, M., Li, J., Shen, Y., & Yang, M. (2022, December). Survey of natural language generation. *CM Computing Surveys*, pp. 1-38.
- Firmender, J. M., Reis, S. M., & Sweeny, S. M. (2013). Reading comprehension and fluency levels ranges across diverse classrooms: The need for differentiated reading instruction and content. *Gifted Child Quarterly*, 57(1), pp. 3-14.
- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research* 71, 279–320.
- Gottlieb, H. (2008). Multidimensional translation. In A. Schjoldager, *Understanding Translation* (pp. 39-65). Aarhus: Academica.
- Gottlieb, H. (2017). Semiotics and translation. In K. Malmkjær, *The Routledge Handbook of Translation Studies and Linguistics* (pp. 45-63). London: Routledge.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Hansen-schirra, S., Bisang, W., Nagels, A., Gutermuth, S., Fuchs, J., Borghardt, L., . . . Sommer, J. (2020). Intralingual translation into easy language – Or how to reduce cognitive processing costs. In S. Hansen-Schirra, & C. Maaß, *Easy Language Research: Text and User Perspectives* (pp. 197-226). Berlin: Frank & Timme.
- Haver, H. L., Gupta, A. K., Ambinder, E. B., Bahl, M., Oluyemi, E. T., Jeudy, J., & Yi, P. H. (2024). Evaluating the use of ChatGPT to accurately simplify patient-centered information about breast cancer prevention and screening. *Radiology* 6(2), e230086.
- Hedge, T. (1985). *Using readers in language teaching*. London: Macmillan.
- Hill-Madsen, A. (2014). Lexical strategies in intralingual translations between registers. *Journal of Language and Communication in Buisness*, 54, pp. 85-15.
- Hill-Madsen, A. (2019). The heterogeneity of intralingual translations. *Meta: Translators' Journal*, pp. 537-560.

- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13, 1*, pp. 403-430.
- Jakobsen, R. (1959). On linguistic aspects of translations. In L. Venuti, *The Translation Studies Reader* (pp. 113-118). London: Routledge.
- Jeblick, K., Schachtner, B., Dexl, J., Mittermeier, A., Stüber, A. T., Topalis, J., . . . Ingris, M (2023). ChatGPT makes medicine easy to swallow: An exploratory case study on simplified radiology reports. *European Radiology, 1-9*.
- Jiao, W., Wang, Wenxuan, Huang, J.-t., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv*.
- Khan, S. (2023). *Brick lane*. Dublin: Penguin Books Ltd.
- Klein, A. F. (2008). *Providing differentiated reading instruction to meet the individual needs of students*. Retrieved from Learning A-Z: https://www.raz-plus.com/updates/reading_az_white_paper.pdf
- Krashen, S. (1992). The input hypothesis: An update. In J. E. Alatis, *Linguistics and language pedagogy: The state of the art* (pp. 409-431) Washington D.C.: Georgetown University Press.
- Krashen, S. D. (1985). The input hypothesis. In S. D. Krashen, *The input hypothesis. Issues and implications* (pp. 1-32). London: Longman.
- Lervåg, A., Dolean, D., Tincas, I., & Melby-Lervåg, M. (2019). Socioeconomic background, nonverbal IQ and school absence affects the development of vocabulary and reading comprehension in children living in severe poverty. *Developmental Science, 22(5)*, 1-15.
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., . . . Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology, 100017*.
- Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P. C., . . . Moore, J. H. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining, 16(1)*, 20.

- Ministry of Education and Research. (2019). Læreplan i engelsk (ENG01-04) [Curriculum for English]. Established as regulations. The National curriculum for the Knowledge Promotion 2020. Retrieved from <https://data.udir.no/k106/v201906/laereplaner-lk20/ENG01-04.pdf?lang=eng>
- Moons, P., & Van Bulck, L. (2023). Using ChatGPT and Google Bard to improve the readability of written patient information: A proof of concept. *European Journal of Cardiovascular Nursing*, 1-5.
- Mossop, B. (2016). 'Intralingual translation': A desirable concept? *Across Languages and Cultures*, 17(1), 1-24.
- Munday, J. (2016). Main issues of translation studies. In J. Munday, *Introducing translation studies: Theories and applications* (pp. 7-28). London: Routledge.
- Murty, L., Sinar, B., Williams, G., Pande-Rolfesen, M. S., Normann, A., & Vicary, T. (2021). Reading for everyone. In G. Williams, & A. Normann, *Literature for the English classroom: Theory into practice* (pp. 169-196). Bergen: Fagbokforlaget.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I. S. P. (2022). The goals of vocabulary learning . In I. S. Nation, *Learning vocabulary in another language* (pp. 8-48). Cambridge: University Printing House.
- Nation, I. S. P. (2012, September 17). *The BNC/COCA word family lists*. Retrieved from Victoria University of Wellington: https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/Information-on-the-BNC_COCA-word-family-lists.pdf
- Nation, I. S. P., & Deweerdt, J. P. (2001, December). A defence of simplification. *Prospect*, 16 (3), pp. 55-65.
- Nation, I. S. P., & Wand Ming-Tzu, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, 12 (2), 155-181.
- Newmark, P. (1981). *Approaches to translation*. London: Pergamon Press.

- Oakhill, J., & Cain, K. (2007). Introduction to comprehension development. In K. Cain, & J. Oakhill, *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 3-40). London: The Guilford Press.
- OpenAI. (2024, March 11). *How ChatGPT and our language models are developed*. Retrieved from OpenAI: <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed>
- Parmar, P. (2022). *Lion*. Dublin: Penguin Books Ltd.
- Penguin Readers. (2024a). *Explore books*. Retrieved from Penguin Readers: <https://www.penguinreaders.co.uk/books/>
- Penguin Readers. (2024b). *Brick lane*. Retrieved from Penguin Readers: <https://www.penguinreaders.co.uk/ladybird-books/brick-lane/>
- Penguin Readers. (2024c). *Lion*. Retrieved from Penguin Readers: <https://www.penguinreaders.co.uk/ladybird-books/lion/>
- Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28, pp. 66-86.
- Rice, S., Crouse, S. R., Winter, S. R., & Rice, C. (2024). The advantages and limitations of using ChatGPT to enhance technological research. *Technology in Society*, 76, 102426.
- Schmid, B. (2008). A duck in rabbit's clothing: Integrating intralingual translation. In M. Kaiser-Cooke, *Das entrprinzip: Translation aus neuen perspektiven* (pp. 19-80). Bern: Peter Lang.
- Sheang, K., & Saggion, H. (2021). Controllable sentence simplification with a unified text-to-text transfer. *International conference on natural language generation* (pp. 341-352). Aberdeen: Association for Computational Linguistics.
- Simensen, A. M. (1990). *Using graded readers: Differentiated teaching in English*. Oslo: Informasjonssenteret for språkundervisning.
- Skjelde, K. (2015). *Academic vocabulary: Unleashed potential? A corpus study of English course materials for advanced Norwegian learners of English*. Department of Foreign Languages: University of Bergen.

- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly* 16(1), pp. 32-71.
- Sunsjak, T. (2022). ChatGPT: The end of online exam integrity. *ArXiv: 2212.09292*, 1-21.
- The Norwegian Directorate for Education and Training. (2019). *English subject curriculum (ENG01-04)*. [Circular]. The Government. Retrieved from The Norwegian Directorate for Education and Training: <https://data.udir.no/kl06/v201906/laereplaner-1k20/ENG01-04.pdf?lang=eng>
- The Norwegian Directorate for Education and Training. (2022). *Mestringsbeskrivelser for nasjonale prøver i engelsk [Descriptions mastery levels for national tests in English]*. Retrieved from The Norwegian Directorate for Education and Training: <https://www.udir.no/eksamen-og-prover/prover/nasjonale-prover/mestringsbeskrivelser-og-hva-provene-maler/engelsk-mestringsbeskrivelse/#a127084>
- The Norwegian Directorate for Education and Training. (2023). *Nasjonale prøver 2023: Resultater nasjonale prøver ungdomstrinnet*. Retrieved from The Norwegian Directorate for Education and Training: <https://www.udir.no/tall-og-forskning/statistikk/statistikk-grunnskole/analyser/2023/analyse-av-nasjonale-prover-for-ungdomstrinnet-2023/>
- Tomlinson, B. (2003). Materials evaluation. In B. Tomlinson, *Developing materials for language teaching* (pp. 15-36). London and New York: Continuum.
- Toury, G. (2012). *Descriptive translation studies - and beyond. Revised edition*. Amsterdam: John Benjamins.
- Tovar, M. M. (2019). (A)bridging the Gap - A study of the norms and laws in the intralingual translation of the novel *And Then There Were None* by Agatha Christie. *Revista de Lenguas para Fines Especificos* 26(1), pp. 51-68.
- Tymoczko, M. (2005). Trajectories of research in translation studies. *Meta* 50 (4), pp. 1082-1097.
- Vellutino, F. R., Tunmer, W. W., Jaccard, J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading* 11(1), 3-32.

- Velmeer, H. (1989). *Skopos and translation*. Heidelberg: Universitat.
- Woolley, G. (2011). Reading comprehension. In G. Woolley, *Reading comprehension: assisting children with learning difficulties* (pp. 15-33). New York: Springer.
- Young, J. C., & Shishido, M. (2023a). Evaluation of the potential usage of ChatGPT for providing easier reading materials for EFL students. *EdMedia + Innovate Learning* (pp. 155-162). Vienna: Association for the Advancement of Computing in Education (AACE).
- Young, J. C., & Shishido, M. (2023b). Investigating OpenAI's ChatGPT potentials in generating Chatbot's dialogue for English as a foreign language learning. *International Journal of Advanced Computer Science and Applications*, 14 (6), 65-72.
- Zethsen, K. K. (2009). Intralingual translation: An attempt at description. *Meta* 4, pp. 795-812.
- Zethsen, K. K. (2018). Access is not the same as understanding: Why intralingual translation is crucial in a world of information overload. *Across Languages and Cultures*, 19(1), 79-98.

Appendixes

Table of appendixes

A. The didactic implications of this study.....	60
B. Expert from source text	61
C. Expert from human-made graded reader.....	65
D. Prompt and material generated from ChatGPT	66
E. Translations strategy analysis.....	70
F. Word frequency distribution tables	81

Appendix A: Relevance for the teaching profession

The aim of this study was to investigate the use of ChatGPT 3.5 for creating differentiated reading materials for the ESL classroom. As this aim relates directly to the ESL context, there are several aspects of this study that are relevant for the teaching profession. One clear aspect of relevance is the use of AI in schools. Through the investigation of the vocabulary choices of simplified text generated by ChatGPT, this study has built on previous research to investigate how these texts differ from texts written by humans, and the possible applications of this AI tool. Gaining knowledge on this will guide us in implementing the tool in schools in a way that is beneficial, rather than detrimental, for learning.

The results of this study indicate that ChatGPT 3.5 has the ability to simplify text. The application of AI in creating suitable reading material for all students in the Norwegian ESL classroom, can then be beneficial, as AI is efficient at simplifying text. However, the results point to some possible characteristics of text simplified by ChatGPT 3.5 that teacher should be aware of. For instance, the results indicate that some *possible* tendencies are that it may not consistently simplify to the desired reading level, may shorten sentences to a greater extent and removes sentences to a lesser extent than humans, and as a result, may give peripheral information as much weight as core information. These characteristics may cause issues when using the texts in the classroom, however knowing these tendencies of ChatGPT may limit the effect of the issues. In addition, knowing of these ChatGPT tendencies may help teachers in the difficult task of identifying when their students have cheated using this AI model.

This thesis has provided me with valuable experience and knowledge relevant for my future work as a teacher. From this project, I have developed a deeper understanding of central aspects for L2 learning, such as comprehensible input, the characteristics of suitable reading materials, and how to aid students in developing reading proficiency. This knowledge will help me implement differentiated instruction for all students concerning reading and teaching in general. In addition, I have gained insights of the possible applications of AI in schools, and issues that may come with this application. I have experienced the value of experimenting with AI to produce learning materials for the ESL classroom, and have become familiar with some possible characteristics of AI to be wary of. On the basis of this project, I feel more confident in tackling the likely implementation of AI in schools in the future.

Appendix B: Expert from source text

Excerpt from *A Long Way Home* (Text A) by Saroo Brierly (2013, pp. 111-112)

I landed on February 11, 2012, in the city of Indore, the biggest city in Madhya Pradesh, my feet touched the ground in India for the first time since I left as a child. In the pre-dawn dark, I felt a rush of adrenaline as the magnitude of this journey hit me.

India didn't exactly welcome me back. My first experiences firmly established me as a stranger—I might have come "home," but this was a country foreign to me. My bag was missing from the luggage claim carousel. When I tried to ask an airport official where it might be, he replied in what I think was Hindi and I didn't understand a word. The official soon went to fetch someone who spoke some English. It seems a little thing to not speak the language, but it carried extra weight for a man making an emotional journey home after years of being lost. It was like being lost all over again, unable to understand what anyone said or to make people understand me.

I made a mess of negotiating my way out of several insistent but exorbitant offers of a taxi to the hotel where I was staying overnight before going on to Khandwa, and I eventually found the courtesy bus. The sun blazed into life as the bus pulled out of the airport, and I got my first look at the pressing confusion of twenty-first-century India. At first much of it looked like the India I had known a quarter of a century earlier. I saw black wild pigs scavenging in side streets, recognizable trees on street corners, and the familiar press of people everywhere. The poverty was still evident, but I was quickly struck by how much dirtier everything appeared from how I remembered it. People were relieving themselves on the roadside and there was rubbish strewn everywhere—I didn't remember the same things from my own neighborhood, but maybe I'd become accustomed to the clean, open spaces of Hobart.

When I stepped off the bus at the hotel, the unrelenting noise of heavy traffic and the strong smell of sulfur, from drains and sewerage, hit my senses. I realized that after such a long time, Khandwa would probably seem different, too. I was exhausted and decided to try to rest a bit before trying to find a driver to take me to Khandwa. After a fitful few hours' sleep, I organized a car and driver who said he would take me there the next day.

Khandwa was two hours away, and I paid half of what the drivers at the airport had quoted me for the few kilometers' ride to the hotel. But perhaps you paid extra for safety: my short, skinny driver took to the roads like a maniac (even by the famously carefree standards of India), which added another shot of adrenaline to my overloaded system. The road from

Indore runs through hills and valleys, but I noticed little of the scenery. We stopped occasionally for a chai and a cigarette and I found myself growing more and more anxious about what awaited me in Khandwa. The death-defying trip couldn't go quickly enough.

Under a hot sun in clear skies, we approached the outskirts of town, I didn't recognize the place at all, which gave me an instant chill. The area had a dusty gray industrial look that I wasn't familiar with, Suddenly, I decided to go straight to the railway station before the hotel—I was past dragging things out, and that would be the quickest and easiest way to discover whether what I had worked out on my laptop back at home in Tasmania was right. We changed direction.

The roads were narrow and traffic slowed to a crawl—it was Sunday and people were out and about everywhere. When I was little, there had been more horses and carts than auto rickshaws, but now the streets. were Clogged with cars and motorbikes.

My mobile phone had a GPS service, which would have laid out a street map for me, but my battery was low and I wanted my memory to be jolted into service So I directed the driver to the best of my memory, and, sure enough, we found the station where I expected it to be. My spirits lifted.

The station looked a little different from how I remembered it, but I found that I instantly had my bearings—from this point, I could find my way to anywhere in Khandwa, I knew where I was . . . and I wasn't far from home.

Excerpt from *Brick Lane* (Text B) by Monica Ali (2003, pp. 4-5)

As Nazneen grew she heard many times this story of How You Were Left To Your Fate. It was because of her mother's wise decision that Nazneen lived to become the wide-faced, watchful girl that she was. Fighting against one's Fate can weaken the blood. Sometimes, or perhaps most times, it can be fatal. Not once did Nazneen question the logic of the story of How You Were Left To Your Fate. Indeed, she was grateful for her mother's quiet courage, her tearful stoicism that was almost daily in evidence. Hamid said – he always looked away as he spoke – your mother is naturally a saint. She comes from a family of saints. So when Rupban advised her to be still in her heart and mind, to accept the Grace of God, to treat life with the same indifference with which it would treat her, Nazneen listened closely with her large head tilted back and her cheeks slack with equanimity.

She was a comically solemn child. 'How is my precious? Still glad you came back to life?' said Mumtaz when she had not seen her for a couple of days. 'I have no complaints or regrets to tell you,' said Nazneen. 'I tell everything to God.'

What could not be changed must be borne. And since nothing could be changed, everything had to be borne. This principle ruled her life. It was mantra, fettle and challenge. So that when, at the age of thirty-four, after she had been given three children and had one taken away, when she had a futile husband and had been fated a young and demanding lover, when for the first time she could not wait for the future to be revealed but had to make it for herself, she was as startled by her own agency as an infant who waves a clenched fist and strikes itself upon the eye.

Her sister Hasina, born only three days after Banesa passed away (one hundred and twenty years old then and for evermore), listened to no one. At the age of sixteen when her beauty was becoming almost unbearable to own or even to look at, she eloped to Khulna with the nephew of the saw-mill owner. Hamid ground his teeth and an axe besides. For sixteen hot days and cool nights he sat between the two lemon trees that marked the entrance to the compound. For that time his only occupation was throwing stones at the piebald dogs that scavenged in the dump just beyond, and cursing his whore-pig daughter whose head would be severed the moment she came crawling back. Those nights, Nazneen lay awake listening to the rattling of the corrugated tin roof, starting at the owl calls that no longer sounded like owls but more like a girl felled by an axe on the back of her neck. Hasina did not come. Hamid went back to supervising the laborers in the paddy fields. But for a couple of thrashings given on only the slightest of provocation, you would not know he had lost a daughter.

Soon after, when her father asked if she would like to see a photograph of the man she would marry the following month, Nazneen shook her head and replied, 'Abba, it is good that you have chosen my husband. I hope I can be a good wife, like Amma.' But as she turned to go she noticed, without meaning to, where her father put the photograph. She just happened to see it. These things happen. She carried the image around in her mind as she walked beneath the banyans with her cousins. The man she would marry was old. At least forty years old. He had a face like a frog. They would marry and he would take her back to England with him. She looked across the fields, glittering green and gold in the brief evening light. In the distance a hawk circled and fell like a stone, came up again and flew against the sky until it shrank to nothing. There was a hut in the middle of the paddy. It looked wrong: embarrassed, sliding down at one side, trying to hide. The tornado that had flattened half the neighboring village

had selected this hut to be saved, but had relocated it. In the village they were still burying their dead and looking for bodies. Dark spots moved through the far fields. Men, doing whatever they could in this world.

Appendix C: Expert from human-made graded reader

Excerpt from graded reader version of *A Long Way Home* (Brierly, 2013) retold by Parakash Parmar (2022, p. 52)

On 11 February 2012, I arrived at the airport in the city of Indore. I got off the aeroplane and my feet touched the ground in India for the first time in twenty-five years. It was very exciting to be back, but after all this time, I felt like a foreign visitor in the country.

My suitcase was lost at the airport. When I tried to ask someone about it, he replied in Hindi and I did not understand a single word. The man found an English-speaking person, but I still felt like a lost boy again, trying to find someone to help me.

On the bus journey to the hotel, I noticed that lots of things were the same as twenty-five years ago. There were still animals in the street, homeless people and large crowds. But I was surprised at how dirty everything looked, which was so different from Hobart.

After a few hours' sleep in the hotel, I got a taxi to Khandwa. The journey took two hours, and I smoked a few cigarettes during it to try to calm my feelings. When we reached the town, I did not recognize it at all, and I immediately felt worried. I asked the driver to go straight to the railway station. When we arrived there, I knew where I was. I was not far from home.

Excerpt from *Brick Lane* (Text B) by Monica Ali (2003)

As Nazneen grew, she often heard the story of How You Were Left To Your Fate. She never questioned accepting fate. She saw the example in her mother, who suffered without complaining every day until the day she died.

Nazneen's sister, Hasina, was very different. She listened to no one. At the age of 16, when she was already almost too beautiful, she ran away with the nephew of a factory owner. Hamid was extremely angry. For weeks, he sat waiting for her, ready to cut off her head. But Hasina did not come back.

Soon after, Nazneen's father asked if she would like to see a photograph of the man she was going to marry next month. Nazneen shook her head and replied, "Abba, it is good that you have chosen my husband. I hope I can be a good wife, like Amma was." But she noticed where her father put the photo, and later she looked at it. The man was at least 40 years old. He was ugly. And he was going to take her to live in England with him.

Appendix D: Material from ChatGPT

AI-generated version of excerpt from *A Long Way Home* (Brierly, 2013)

On February 11, 2012, I reached the city of Indore, the big city in Madhya Pradesh. It was my first time back in India since I left as a child. In the early morning, I felt excited about this journey. However, India didn't welcome me warmly. My bag was missing, and when I tried to ask for help, the airport staff spoke in Hindi, which I didn't understand.

I struggled to find a cheap taxi but eventually found the free bus to my hotel. As the bus left the airport, I saw the busy confusion of modern India. It seemed dirtier than I remembered, with people relieving themselves on the roadside and rubbish everywhere.

After some restless sleep, I got a driver to take me to Khandwa the next day. The journey was two hours, and though it cost less than airport taxis, the ride was intense. The roads twisted through hills, and I grew anxious about what awaited me in Khandwa.

Approaching the town, everything looked different, and I went straight to the railway station. The area had an unfamiliar industrial look. We changed direction through crowded streets filled with cars and motorbikes. My phone's battery was low, so I relied on memory to find the station. When we arrived, it looked slightly different, but I instantly knew where I was. I wasn't far from home.

AI-generated version of excerpt from *Brick Lane* (Ali, 2003)

Nazneen grew up hearing a story about her fate. Her mom's smart choice saved her life. Nazneen never questioned this story; she was grateful for her mom's bravery. Hamid, a friend, praised her mom, calling her like a saint from a saintly family.

Nazneen believed in accepting life and not fighting fate. She faced everything calmly, even when Mumtaz asked about her well-being. Nazneen shared everything with God and had no complaints.

Nazneen's sister, Hasina, was different. She eloped at sixteen, causing trouble for her family. Despite Hamid's anger, Hasina didn't return, and life went on for Nazneen's family. At thirty-four, Nazneen had to shape her future.

Later, when Nazneen's dad showed a photo of her future husband, she didn't object. She hoped to be a good wife like her mom. However, she noticed the man was older with a unique face. They would marry, and he would take her to England.

While walking with her cousins, Nazneen carried the image, observing fields and a circling hawk. In the distance, a hut stood, saved by a tornado that devastated a nearby village. Life continued for villagers, with dark spots moving through fields as men worked on various tasks.

Prompt and outputs from ChatGPT for AI version of Text A

<p>Prompt 1</p>	<p>Keep as much as possible of the content, but rewrite this text to fit for a A2/B1 level of English. Change difficult words to simple and frequent words without changing the content. The text should be a bit shorter than the original.</p> <p>*inserted Text A*</p>
	<p>Meeting My Mother</p> <p>On February 11, 2012, I arrived in the city of Indore, the largest city in Madhya Pradesh. It was my first time back in India since I left as a child. In the early morning darkness, I felt excited about this journey. However, India didn't welcome me warmly. My bag was missing, and when I tried to ask for help, the airport staff spoke in Hindi, which I didn't understand.</p> <p>I struggled to navigate through offers for an expensive taxi but eventually found the courtesy bus to my hotel. As the bus left the airport, I saw the bustling confusion of twenty-first-century India. It seemed dirtier than I remembered, with people relieving themselves on the roadside and rubbish everywhere.</p> <p>After some restless sleep, I arranged for a driver to take me to Khandwa the next day. The journey was two hours, and though it cost less than airport taxis, the ride was intense. The roads twisted through hills, and I grew anxious about what awaited me in Khandwa.</p> <p>Approaching the town, everything looked different, and I decided to go straight to the railway station. The area had an unfamiliar industrial appearance. We changed direction through crowded streets filled with cars and motorbikes. My phone's battery was low, so I relied on memory to find the station. When we arrived, it looked slightly different, but I instantly knew where I was. I wasn't far from home.</p>
<p>Adjusting prompt</p>	<p>you use a few difficult words. With minimal changes, please replace the difficult words with easier words or remove them. For example: courtesy</p>
	<p>10. Meeting My Mother</p> <p>On February 11, 2012, I arrived in the city of Indore, the biggest city in Madhya Pradesh. It was my first time back in India since I left as a child. In the early morning darkness, I felt excited about this journey. However, India didn't welcome me warmly. My bag was missing, and when I tried to ask for help, the airport staff spoke in Hindi, which I didn't understand.</p> <p>I struggled to navigate through offers for an expensive taxi but eventually found the free bus to my hotel. As the bus left the airport, I saw the busy confusion of twenty-first-century India. It seemed dirtier than I remembered, with people relieving themselves on the roadside and rubbish everywhere.</p> <p>After some restless sleep, I arranged for a driver to take me to Khandwa the next day. The journey was two hours, and though it cost less than airport taxis, the ride was intense. The roads twisted through hills, and I grew anxious about what awaited me in Khandwa.</p> <p>Approaching the town, everything looked different, and I decided to go straight to the railway station. The area had an unfamiliar industrial appearance. We changed direction through crowded streets filled with cars and motorbikes. My phone's battery was low, so I relied on memory to find the station. When we arrived, it looked slightly different, but I instantly knew where I was. I wasn't far from home.</p>

Adjusting prompt	with minimal changes, make try to make the text only consist of high frequency words.
Selected output	<p>Meeting Mom</p> <p>On February 11, 2012, I reached the city of Indore, the big city in Madhya Pradesh. It was my first time back in India since I left as a child. In the early morning, I felt excited about this journey. However, India didn't welcome me warmly. My bag was missing, and when I tried to ask for help, the airport staff spoke in Hindi, which I didn't understand.</p> <p>I struggled to find a cheap taxi but eventually found the free bus to my hotel. As the bus left the airport, I saw the busy confusion of modern India. It seemed dirtier than I remembered, with people relieving themselves on the roadside and rubbish everywhere.</p> <p>After some restless sleep, I got a driver to take me to Khandwa the next day. The journey was two hours, and though it cost less than airport taxis, the ride was intense. The roads twisted through hills, and I grew anxious about what awaited me in Khandwa.</p> <p>Approaching the town, everything looked different, and I went straight to the railway station. The area had an unfamiliar industrial look. We changed direction through crowded streets filled with cars and motorbikes. My phone's battery was low, so I relied on memory to find the station. When we arrived, it looked slightly different, but I instantly knew where I was. I wasn't far from home.</p>

Prompt and outputs from ChatGPT for AI version of Text B

Prompt 1	<p>Keep as much as possible of the content, but rewrite this text to fit for a A2/B1 level of English. Change difficult words to simple and frequent words without changing the content. The text should be a bit shorter than the original.</p> <p>*inserted Text B*</p>
	<p>As Nazneen grew up, she often heard the story of how she was left to fate. Her mother's wise choice saved her life. Nazneen never questioned this story and was thankful for her mother's courage. Hamid, a friend, praised her mother, saying she was like a saint from a saintly family.</p> <p>Nazneen believed in accepting life as it came and not fighting against fate. She faced everything with equanimity, even when Mumtaz asked about her well-being. Nazneen replied that she shared everything with God and had no complaints.</p> <p>Nazneen's sister, Hasina, was quite different. She eloped at sixteen, causing distress to her family. Despite Hamid's anger, Hasina didn't return, and life moved on for Nazneen's family. Nazneen, at the age of thirty-four, found herself in a situation where she had to shape her own future.</p> <p>Later, when Nazneen's father showed her a photo of her future husband, she didn't object, hoping to be a good wife like her mother. However, she couldn't help but notice the man was older with a frog-like face. They would marry, and he would take her to England.</p>

	While walking with her cousins, Nazneen carried the image in her mind, observing the fields and a circling hawk. In the distance, a hut stood, saved by a tornado that had devastated a nearby village. Life went on for the villagers, with dark spots moving through the fields as men worked on various tasks.
Adjusting prompt	you use a few difficult words. With minimal changes, please replace the difficult words with easier words or remove them. For example: equanimity
	<p>10. Meeting My Mother</p> <p>On February 11, 2012, I arrived in the city of Indore, the biggest city in Madhya Pradesh. It was my first time back in India since I left as a child. In the early morning darkness, I felt excited about this journey. However, India didn't welcome me warmly. My bag was missing, and when I tried to ask for help, the airport staff spoke in Hindi, which I didn't understand.</p> <p>I struggled to navigate through offers for an expensive taxi but eventually found the free bus to my hotel. As the bus left the airport, I saw the busy confusion of twenty-first-century India. It seemed dirtier than I remembered, with people relieving themselves on the roadside and rubbish everywhere.</p> <p>After some restless sleep, I arranged for a driver to take me to Khandwa the next day. The journey was two hours, and though it cost less than airport taxis, the ride was intense. The roads twisted through hills, and I grew anxious about what awaited me in Khandwa.</p> <p>Approaching the town, everything looked different, and I decided to go straight to the railway station. The area had an unfamiliar industrial appearance. We changed direction through crowded streets filled with cars and motorbikes. My phone's battery was low, so I relied on memory to find the station. When we arrived, it looked slightly different, but I instantly knew where I was. I wasn't far from home.</p>
Adjusting prompt	with minimal changes, make try to make the text only consist of high frequency words.
Selected output	<p>Meeting Mom</p> <p>On February 11, 2012, I reached the city of Indore, the big city in Madhya Pradesh. It was my first time back in India since I left as a child. In the early morning, I felt excited about this journey. However, India didn't welcome me warmly. My bag was missing, and when I tried to ask for help, the airport staff spoke in Hindi, which I didn't understand.</p> <p>I struggled to find a cheap taxi but eventually found the free bus to my hotel. As the bus left the airport, I saw the busy confusion of modern India. It seemed dirtier than I remembered, with people relieving themselves on the roadside and rubbish everywhere.</p> <p>After some restless sleep, I got a driver to take me to Khandwa the next day. The journey was two hours, and though it cost less than airport taxis, the ride was intense. The roads twisted through hills, and I grew anxious about what awaited me in Khandwa.</p> <p>Approaching the town, everything looked different, and I went straight to the railway station. The area had an unfamiliar industrial look. We changed direction through crowded streets filled with cars and motorbikes. My phone's battery was low, so I relied on memory to find the station. When we arrived, it looked slightly different, but I instantly knew where I was. I wasn't far from home.²</p>

² OpenAI. (2024b). *ChatGPT* (Jan) [Large language model]. <https://chat.openai.com/share/3bee1af4-0c46-45be-ad9a-5406b09a6c70>

OpenAI. (2024c). *ChatGPT* (Jan) [Large language model]. <https://chat.openai.com/share/7c7c5540-52ad-4ba1-8bb1-3c1e1dea13429>

Appendix E: Translation strategy analysis

This appendix shows the table from the analysis of translation strategies. For descriptions of the columns and symbols in the table see section 3.3.1.2. The first table shows the coupled pairs and strategies identified in the translations of Text A. The table for Text B can be found on page 76.

The analysis of translations strategies in human-made and AI-generated translations of Text A

ST	HM	AI	Strategies HM	Strategies AI	Strategies (Same/different)
the pre dawn dark	-	early mourning	Omission	Demetaphorisation	Different
a rush of adrenaline	exciting	excited	Demetaphorisation	Demetaphorisation	Same
the magnitude of this journey	-	journey	Omission	PC-to-word	Different
welcome	-	welcome	Omission		Different
firmly	-	-	Omission	Omission	Same
established	-	-	Omission	Omission	Same
stranger	foreign visitor	stranger	Explicitation		Different
foreign	foreign	-		Omission	Different
bag	suitcase	bag	Explicitation		Different
missing	lost	missing	Higher frequency word		Different
the luggage claim curousel	airport	-	PC-to-word	Omission	
an airport official	someone	airport staff	PC-to-word		Different
he	he	airport staff		Explicitation	Different
a word	a single word		Explicitation		Different
the official	the man	airport staff	Higher frequency word	Higher frequency word	Same

fetch (went to fetch)	found	fetch	Higher frequency word		Different
someone	a person	someone	Explicitation		
who spoke some English	English-speaking		PC-to-word		Different
language	-	-	Omission	Omission	
emotional	-	-	Omission	Omission	
journey	-	-	Omission	Omission	
made a mess of		struggled		Demetaphorisation	Different
negotiating	-	-	Omission	Omission	Same
insistent	-	-	Omission	Omission	Same
exorbitant	-	-	Omission	Omission	Same
taxi	-	taxi	Omission		Different
hotel where I was staying overnight	-	hotel	Omission	PC-to-word	Different
eventually	-	eventually	Omission		Different
courtesy	-	free	Omission	Higher frequency word	Different
blazed	-	-	Omission	Omission	Same
pressing	-	busy	Omission	Demetaphorisation	Different
confusion	-	confusion	Omission		Different
black wild pigs	animals	-	PC-to-word	Omission	Different
scavenging	-	-	Omission	Omission	Same
side streets	street	-	PC-to-word	Omission	Different
recognizable	-	-	Omission	Omission	Same
familiar	-	-	Omission	Omission	Same

press of people	crowds	-	Demetaphorisation	Omission	Different
poverty	homeless people	-	Explicitation	Omission	Different
evident	-	-	Omission	Omission	Same
quickly struck by	surprised	-	Demetaphorisation	Omission	Different
much dirtier	dirty	dirtier	PC-to-word	PC-to-word	Same
revealing	-	revealing	Omission		Different
rubbish strewn	-	rubbish	Omission	PC-to-word	Different
accostemed	-	-	Omission	Omission	Same
unrelenting	-	-	Omission	Omission	Same
traffic	-	-	Omission	Omission	Same
sulfer	-	-	Omission	Omission	Same
drains	-	-	Omission	Omission	Same
sewerage	-	-	Omission	Omission	Same
senses	-	-	Omission	Omission	Same
exhausted	-	-	Omission	Omission	Same
fitful	-	restless	Omission	Higher frequency word	Different
sleep	sleep in the hotel	sleep	Explicitation		Different
organized	got	got	Higher frequency word	Higher frequency word	Same
car	taxi	driver	Explicitation		Different
there	Khandwa	Khandwa	Explicitation	Explicitation	Same
airport	-	airport	Omission		Different
quoted	-	-	Omission	Omission	Same

kilometers	-	-	Omission	Omission	Same
hotel	-	-	Omission	Omission	Same
skinny	-	-	Omission	Omission	Same
maniac	-	-	Omission	Omission	Same
famously	-	-	Omission	Omission	Same
standards	-	-	Omission	Omission	Same
adrenaline	-	-	Omission	Omission	Same
overloaded	-	-	Omission	Omission	Same
road from Indore		roads		PC-to-word	Different
runs through		twisted through		Explicitation	Different
valleys	-	-	Omission	Omission	Same
scenery	-	-	Omission	Omission	Same
occasionally	-	-	Omission	Omission	Same
chai	-	-	Omission	Omission	Same
cigarettes	cigarettes	-		Omission	Different
found myself growing		grew		PC-to-word	Different
anxious	-	anxious	Omission		Different
awaited	-	awaited	Omission		Different
defying	-	-	Omission	Omission	Same
the outskirts of the town	town	town	PC-to-word	PC-to-word	Same
the place	it	-	PC-to-word	Omission	Different
instant	immediately	-	Higher frequency word	Omission	Different

chill	felt worried	-	Higher frequency word	Omission	Different
dusty	-	-	Omission	Omission	Same
industrial	-	industrial	Omission		Different
wasn't familiar	-	unfamiliar	Omission	PC-to-word	Different
decided to go		went		PC-to-word	Different
hotel	-	-	Omission	Omission	Same
dragging	-	-	Omission	Omission	Same
laptop	-	-	Omission	Omission	Same
direction	-		Omission		Different
narrow	-	-	Omission	Omission	Same
traffic	-	-	Omission	Omission	Same
cawl	-	-	Omission	Omission	Same
people were out and about everywhere		crowded		Demetaphorisation	Different
carts	-	-	Omission	Omission	Same
auto rickshaws	-	-	Omission	Omission	Same
clogged	-	filled	Omission	Higher frequency word	Different
motorbikes	-	-	Omission	Omission	Same
mobile	-	phone	Omission	Higher frequency word	Different
map	-	-	Omission	Omission	Same
battery	-	My phone's battery	Omission	Explicitation	Different
memory	-	-	Omission	Omission	Same
jolted	-	-	Omission	Omission	Same

directed	-	-	Omission	Omission	Same
memory	-	-	Omission	Omission	Same
spirits	-	-	Omission	Omission	Same
The station		it		PC-to-word	Different
a little		slightly		PC-to-word	Different
found that I instantly	-	instantly	Omission	PC-to-word	Different

The Analysis of translations strategies in human-made and AI-generated translations of Text A

Source text	HM	AI	Strategies HM	Strategies AI	Strategies (Same/different)
many times	often		Phrase to word		Different
wise	-	smart	Omission	Higher frequency word	Different
decision	-	choice	Omission	Higher frequency word	Different
fate	-	-	Omission	Omission	Same
weaken	-	-	Omission	Omission	Same
fetal	-	-	Omission	Omission	Same
not once		never		PC-to-word	Different
logic	-	-	Omission	Omission	Same
the story of How You Were Left To Your Fate		this story		PC-to-word	Different
grateful	-	grateful	Omission		Different
courage	-	bravery	Omission	Higher frequency word	Different

tearful	-	-	Omission	Omission	Same
stoicism	-	-	Omission	Omission	Same
evidence	-	-	Omission	Omission	Same
Hamid		Hamid, a friend		Explicitation	Different
saint	-	saint	Omission		Different
family of saints	-	saintly family	Omission		Different
advised	-	-	Omission	Omission	Same
grace	-	-	Omission	Omission	Same
indifference	-	-	Omission	Omission	Same
tilted	-	-	Omission	Omission	Same
cheeks	-	-	Omission	Omission	Same
slack	-	-	Omission	Omission	Same
equanimity	-	-	Omission	Omission	Same
comically	-	-	Omission	Omission	Same
solemn	-	-	Omission	Omission	Same
precious	-	-	Omission	Omission	Same
complaints	-	-	Omission	Omission	Same
regrets	-	-	Omission	Omission	Same
principle	-	-	Omission	Omission	Same
mantra	-	-	Omission	Omission	Same
fettle	-	-	Omission	Omission	Same

challange	-	-	Omission	Omission	Same
futile	-	-	Omission	Omission	Same
fated	-	-	Omission	Omission	Same
demanding	-	-	Omission	Omission	Same
furutre	-	-	Omission	Omission	Same
revealed	-	-	Omission	Omission	Same
startled	-	-	Omission	Omission	Same
agency	-	-	Omission	Omission	Same
infant	-	-	Omission	Omission	Same
clenched	-	-	Omission	Omission	Same
fist	-	-	Omission	Omission	Same
Her	Nazneen's	Nazneen's	Explicitation	Explicitation	Same
evermore	-	-	Omission	Omission	Same
her beauty was becoming	she was		Demetaphorisation		Different
eloped	-	eloped	Omission		Different
nephew	nephew	-		Omission	Different
saw-mill	factory	-		Omission	Different
sixteen hot days and cool nights	weeks		Phrase to word		Different
ground his teeth	was extremely angry	Hamid's anger	Demetaphorisation	Demetaphorisation	Same
axe	-	-	Omission	Omission	Same
lemon	-	-	Omission	Omission	Same

entrance	-	-	Omission	Omission	Same
compound	-	-	Omission	Omission	Same
occupation	-	-	Omission	Omission	Same
scavenged	-	-	Omission	Omission	Same
dump	-	-	Omission	Omission	Same
cursing	-	-	Omission	Omission	Same
whore	-	-	Omission	Omission	Same
pig	-	-	Omission	Omission	Same
svered	-	-	Omission	Omission	Same
crawling	-	-	Omission	Omission	Same
awake	-	-	Omission	Omission	Same
rattling	-	-	Omission	Omission	Same
corrugated	-	-	Omission	Omission	Same
tin	-	-	Omission	Omission	Same
roof	-	-	Omission	Omission	Same
owl	-	-	Omission	Omission	Same
owls	-	-	Omission	Omission	Same
felled	-	-	Omission	Omission	Same
come back	come back	return		PC-to-word	Different
supervising	-	-	Omission	Omission	Same
labourers	-	-	Omission	Omission	Same

paddy	-	-	Omission	Omission	Same
thrashings	-	-	Omission	Omission	Same
provocation	-	-	Omission	Omission	Same
Soon after	Soon after	Later		PC-to-word	Different
her father	Nazneen's father	Nazneen's dad	Explicitation	Explicitation	Same
Amma	Amma	mom		Higher frequency word	Different
of the man she would marry	of the man she was going to marry	husband		PC-to-word	Different
a face like a frog	ugly	a unique face.	Demetaphorisation	Demetaphorisation	Same
take her	take her to live	take her	Explicitation		
image	-	image	Omission		Different
banyans	-	-	Omission	Omission	Same
cousins	-	cousins	Omission		Different
frog	-	-	Omission	Omission	Same
glittering	-	-	Omission	Omission	Same
brief	-	-	Omission	Omission	Same
distance	-	distance	Omission		Different
hawk	-	hawk	Omission		Different
circled	-	circiling	Omission		Different
shrank	-	-	Omission	Omission	Same
hut	-	hut	Omission		Different

paddy	-	-	Omission	Omission	Same
embarrassed	-	-	Omission	Omission	Same
sliding	-		Omission		Different
tornado	-	tornado	Omission		Different
flattened		devastated		Demetaphorisation	
village	-	village	Omission		Different
selected	-	-	Omission	Omission	Same
hut	-	-	Omission	Omission	Same
relocated	-	-	Omission	Omission	Same
village	-	-	Omission	Omission	Same
burying	-		Omission		Different

Appendix F: Word frequency distribution tables

This appendix contains the table for the word frequency distribution results. There are six tables for each of the six texts. In cases where there are more than three consecutive rows that were empty, the rows were compacted into one row (e.g. K-15...K-25). For further descriptions of these metrics see section 3.4 and 4.2.

The word “families” column in the table shows in more detail the differences between the proportion of high, mid and low frequency words in the various texts (low= 0-3000, mid= 3000-6000, low= above 6000). Here we can see the differences between the human-made and AI-generated graded reader versions of Text A.

Source text A							
Freq. Level	Families	Families (%)	Types	Types (%)	Tokens	Tokens (%)	Cumul. Token (%)
K-1	235	72.8	267	74.79	684	87.0	87.0
K-2	51	15.8	53	14.85	64	8.1	95.1
Coverage 95							
K-3	9	2.8	9	2.52	9	1.1	96.2
K-4	9	2.8	9	2.52	9	1.1	97.3
K-5	1	0.3	1	0.28	1	0.1	97.4
K-6	7	2.2	7	1.96	7	0.9	98.3
Coverage 98							
K-7	3	0.9	3	0.84	3	0.4	98.7
K-8	2	0.6	2	0.56	2	0.3	99.0
K-9	1	0.3	1	0.28	2	0.3	99.3
K-10	2	0.6	2	0.56	2	0.3	99.6
K-11							
K-12	1	0.3	1	0.28	1	0.1	99.7
K-13	1	0.3	1	0.28	1	0.1	99.8
K-14	1	0.3	1	0.28	1	0.1	99.9
K-15...K-25							
Off-List	??	0.00	0	0.00	0	0.00	
Total (unrounded)	323	357 (100)	357	≈100.00	786		

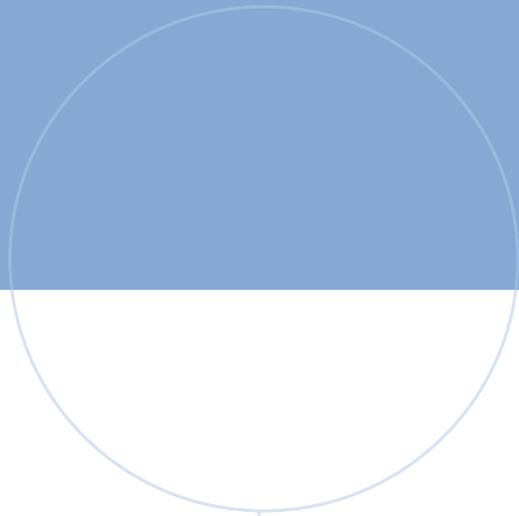
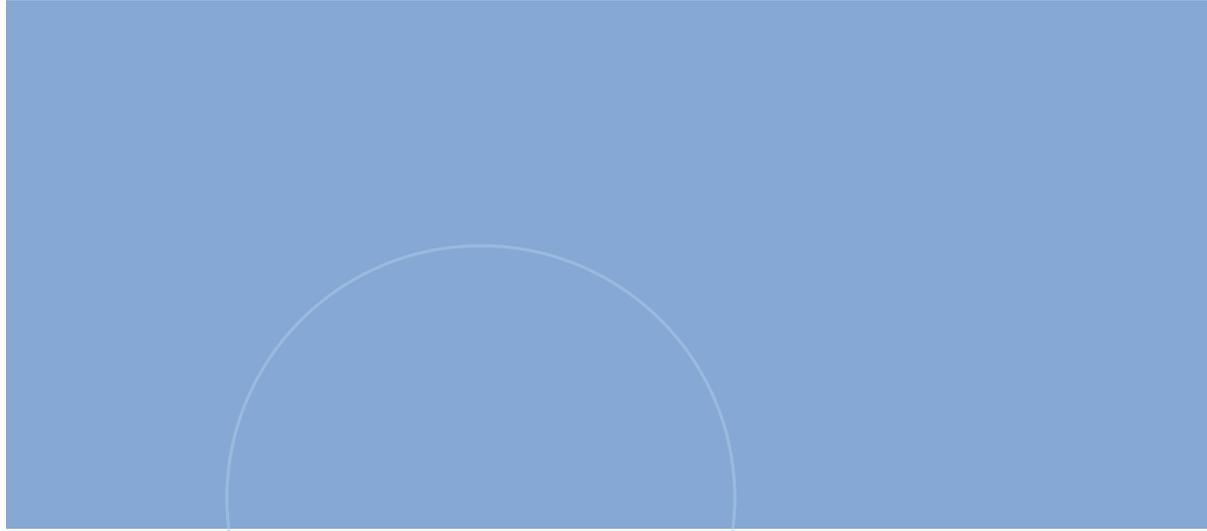
Human-made graded reader version of Text A							
req. Level	Families	Families (%)	Types	Types (%)	Tokens	Tokens (%)	Cumul. Token (%)
K-1	111	88.8	122	89.05	220	92.4	92.4
K-2	13	10.4	13	9.49	16	6.7	99.1
Coverage 95							
K-3	1	0.8	1	0.73	1	0.4	99.5
Coverage 98							
K-4...K-25							
Off-List	??	1 (0.73)	1	0.42	99.92		
Total (unrounded)	125	137 (100)	137	≈100.00	237		

AI-generated graded reader version of Text A							
Freq. Level	Families	Families (%)	Types	Types (%)	Tokens	Tokens (%)	Cumul. Token (%)
K-1	109	79.6	116	80.00	204	86.4	86.4
K-2	25	18.2	26	17.93	29	12.3	98.7
Coverage 95							
K-3	2	1.5	2	1.38	2	0.8	99.5
Coverage 98							
K-4	1	0.7	1	0.69	1	0.4	99.9
K-5...K-25							
Off-List	??	0.00	0	0.00	0	0.00	
Total (unrounded)	137	145 (100)	145	≈100.00	236		

Source text B							
Freq. Level	Families	Families (%)	Types	Types (%)	Tokens	Tokens (%)	Cumul. Token (%)
K-1	226	72.9	264	75.00	653	87.4	87.4
K-2	31	10.0	32	09.09	33	4.4	91.8
K-3	15	4.8	16	4.55	18	2.4	94.2
K-4	14	4.5	15	4.26	17	2.3	96.5
Coverage 95							
K-5	7	2.3	7	1.99	7	0.9	97.4
K-6	3	1.0	3	0.85	3	0.4	97.8
Coverage 98							
K-7	1	0.3	1	0.28	1	0.1	97.9
K-8	2	0.6	2	0.57	2	0.3	98.2
K-9	2	0.6	2	0.57	2	0.3	98.5
K-10	2	0.6	2	0.57	3	0.4	98.9
K-11							
K-12	1	0.3	1	0.28	1	0.1	99.0
K-13	1	0.3	1	0.28	1	0.1	99.1
K-14							
K-15	2	0.6	2	0.57	2	0.3	99.4
K-16							
K-17							
K-18	1	0.3	1	0.28	1	0.1	99.5
K-19	1	0.3	1	0.28	1	0.1	99.6
K-20							
K-21							
K-22	1	0.3	1	0.28	1	0.1	99.7
K-23...K-25							
Off-List		1 (0.28)	1	0.13	99.83		
Total (unrounded)	310	352 (100)	352	≈100.00	746		

Human-made graded reader version of Text B							
Freq. Level	Families	Families (%)	Types	Types (%)	Tokens	Tokens (%)	Cumul. Token (%)
K-1	101	92.7	110	93.22	175	95.1	95.1
Coverage 95							
K-2	4	3.7	4	3.39	4	2.2	97.3
K-3	2	1.8	2	1.69	3	1.6	98.9
Coverage 98							
K-4							
K-5	1	0.9	1	0.85	1	0.5	99.4
K-6...K-21							
K-22	1	0.9	1	0.85	1	0.5	99.9
K-23...K-25							
Off-List	??	0.00	0	0.00	0	0.00	
Total (unrounded)	109	118 (100)	118	≈100.00	184		

AI-generated graded reader version of Text B							
Freq. Level	Families	Families (%)	Types	Types (%)	Tokens	Tokens (%)	Cumul. Token (%)
K-1	91	77.8	96	77.42	172	85.1	85.1
K-2	14	12.0	16	12.90	17	8.4	93.5
K-3	8	6.8	8	6.45	9	4.5	98.0
Coverage 95							
K-4	1	0.9	1	0.81	1	0.5	98.5
Coverage 98							
K-5	1	0.9	1	0.81	1	0.5	99.0
K-6							
K-7	1	0.9	1	0.81	1	0.5	99.5
K-8...K-12							
K-13	1	0.9	1	0.81	1	0.5	100.0
K-14...K-25							
Off-List	??	0.00	0	0.00	0	0.00	
Total (unrounded)	117	124 (100)	124	≈100.00	202		



 **NTNU**

Norwegian University of
Science and Technology