Jekyll & Hyde: Looking Behind '*de Gevel*'

The Potential Aid of Machine Translation for the Beginning Musical Translator

Maaike de Rijk

S1838172

MA Thesis

Linguistics: Translation in Theory and Practice

Leiden University

Supervisor: Dr. A.G. Dorst

Second reader: Dr.T. Reus

5 May 2021

Abstract

This master thesis attempts to portray the potential benefits of using an MT in the translation of expressive texts, specifically that of the novel *The Strange Case of Dr Jekyll and Mr Hyde*, as well as the dialogue and song lyrics of the musical *Jekyll & Hyde*. Three excerpts per genre were translated using Google Translate, and the output was assessed and post-edited with Moorkens' (2018) method for assessing MT output. This method was adapted for a more specific assessment of style in addition to content by using the four categories of language by Leech & Short (2007), namely lexical categories, grammatical categories, figures of speech, and cohesion and context. Furthermore, the error typology by Farrús et al (2012) was used instead of Moorkens' own typology, in order to achieve a more specific view on errors made by the MT. From these results, it can be concluded that all three genres show significant quality of their output. Secondly, most errors were either lexical or syntactic errors. It can also be concluded that while the editing effort of song lyrics is significantly higher than that of prose or dialogue, it could still be considered a large aid for beginning translators within the field of musical translation, as it creates a first, literal draft on which can be easier built upon than translating from scratch.

Keywords: Machine Translation, NMT, Musical Translation, Song Translation

Contents

Introduction

The translation of foreign musicals is a common strategy in Dutch musical theatre. Starting in 1960 with the translation of *My Fair Lady*, which was highly popular with the Dutch audience, large producing companies in the Netherlands like Stage Entertainment and V&V Entertainment currently translate many of its musicals, mainly from English, but also German or French (Busscher, 2018). From 2010 to 2019, 27 of the 41 musicals produced by Stage Entertainment Nederland were translated musicals, as well as two different iterations of *Musicals in Concert*, which contained translated songs as well as songs from original Dutch musicals (StageEntertainment, 2020).

As a field of study, musicals have been the subject of academic research less frequently than other forms of stage entertainment. A significant number of researchers and critics consider the musical to be only entertainment, and not high culture (Croonen, 2011), while other instances of performance such as ballet and the opera are considered to be more impactful and more deserving of the label 'art'. Not unlike the practice of translation, which is seen as a less expressive practice than the writing process. Annie M.G. Schmidt, when asked to translate a musical, said *"Zolang ik zelf nog zoveel ideeën heb, vertaal ik niets!"* [As long as I have ideas of my own, I will not translate anything!] (Linders-Nouwens, 1999). While over half of the musicals put on stage are in fact translations, critical reviews of musicals rarely discuss the quality of the translation, unless the translation is found to be of poor quality.

This disregard of many musicals' translated nature is striking, given that the translation of expressive texts is a considerably difficult task. Not only must the translator convey the content correctly, but the form and style of the text play an important role in supporting its function as entertainment and must be understandable and natural to properly entertain and interest the reader. The song lyrics in musicals create additional challenges. In order to create a singable translation, the translator should respect the pre-existing melody of

the original song and attempt to follow its rhythm as much as possible (Low, 2008). This creates difficulties for the translator, who must be aware of the rhythm and melody and take this into account while translating. Musical translation on its own, but even expanding to song translation in general, has remained as one of the more underdeveloped areas of translation research. Available research does discuss the difficulties of song translation (Buhler, 2017), the assessment of existing song translations (Smola, 2011; Russel, 2018) and even some attempts towards creating an assessment approach specifically designed for song translation (Low, 2005). Nevertheless, there is still very little research that specifically goes towards the practice of translating for singing.

Another field of research touched on in this research is the use of machine translation in the translation process of expressive texts, specifically in song translation. A survey by Torres Dominguez has shown that, of the 500 vendors and translation companies surveyed, over one third were actively using MT in their translation practice, half of which believed that MT cuts down their working time and increases their productivity (2012). However, MT is not often for expressive texts, mainly because the belief that "the resulting literary style would be atrocious and fuller of 'howlers' and false values than the worst that any human translator produces" (Holmström, 1951). Holmström's reason for this statement was that a machine would only translate what was on a page, while a human translator would make their own choice regarding what is best for the text, what the target audience would prefer and what style they themselves are partial to. To reiterate this point, Holmström further states:

> "translation is an art; something which at every step involves personal choice between uncodifiable alternatives; not merely direct substitutions of equated sets of symbols but choices of values dependent for their soundness on the whole antecedent education and personality of the translator" (1951).

Research by several scholars, such as the works of Jones and Irvine (2013) and Constantine (2019) has established that, depending on the complexity of the syntax of the literary text, machine translation could indeed function as an aid in the translation of literature. The general consensus is that the more straightforward the text is, the more likely the end-result is to be grammatically correct. With this information, it could be assumed that machine translation could equally be used in the translation of musicals, especially within the translation of spoken text. As the spoken dialogue of musicals needs to be memorized by actors, as well as understood by the audience in one hearing, these are often following relatively easy to understand syntax, which, according to previous research, should result in accurate translations. Moreover, the output produced by an MT specifically from song lyrics, could function as a first draft in which the text is translated literally, after which a human translator can edit this output to fit the melody.

In this master thesis, I will assess the MT output of Google Translate for three excerpts from the novel *The Strange Case of dr. Jekyll and mr. Hyde* (Stevenson, 1886), as well as three song lyrics and three dialogue scenes from the musical *Jekyll & Hyde* (2013). First, I will assess the MT output based on Moorkens (2018) method for assessing MT output, which was adapted for a more specific assessment on style as well as content by using the four categories of language (Leech & Short, 2007), and the error typology by Farrús et al (2012). Secondly, I will post-edit all of the machine-translated fragments and discuss particular translation problems. In doing so, I will address the following three research questions:

RQ1: Is the quality of MT output of song lyrics and dialogue comparable to that of written prose and what are the main difficulties encountered by machine translation in these three areas?

RQ2: Is the post-editing effort of MT output for song lyrics and dialogue comparable to written prose, and is the post-editing effort of MT output less demanding translating from scratch?

RQ3: In what ways can machine translation function as an aid for (the beginning) translator in musical translation?

1. Literature Review

1.1 Musical Translation

1.1.1 A Brief History of Musical Theatre

Modern musical theatre for the West emerged in the 19th century, mainly through the efforts of Gilbert and Sullivan in England, and Harrigan and Hart in the United States. These countries were mostly seeking a lighter, easier to digest musical theatre form that could co-exist with the opera (Scholten, 2004). In a period of thirty years, these countries saw the creation of many different genres of theatre, including vaudeville, the operetta and burlesque. Scholten (2004, p. 4) distinguishes these new genres from other pre-existing types of theatre by, among others, their use of both sung and spoken text, dance, colourful costumes and backdrops. According to Brouwer (1984), *The Beggar's Opera,* a 1720's satirical ballad opera, functions as a starting point from which the modern musical was conceived. From this point onwards, theatre plays started integrating different genres and different techniques in theatre, eventually resulting in the first officially recognized musical *Showboat* (1927). From this point onwards, the musical continued to grow into common practice in the United States, but it took until 1960 for a musical to be first performed in the Netherlands (Busscher, 2018).

In 1960, Dutch audiences saw both the first original Dutch musical performed on a professional scale, *Alle Wegen Gaan Naar Amsterdam*, as well as the first translated musical, *My Fair Lady*. While this first translated musical was wildly popular and was performed for

over two years with a total of 702 performances (Brouwer, 1984, p. 74), American musicals were no guarantee for success: the foreign musicals that were produced in the years that followed did not appeal to the audiences as had been expected and did not receive many positive reviews by critics or the general audience. This was not due to their translation: according to Brouwer, the Dutch audience differed in their preferences of entertainment compared to the English and American. As a result, a successful musical abroad did not guarantee that a Dutch audience would like the content as well (1984, p. 79). During the early 1960's, audiences preferred the Dutch musical comedies over translations, because the originals were more unassuming and less theatrical than the foreign musicals. When the concept of the musical became more commonplace within the Netherlands, this attitude shifted, and translated musicals were visited much more frequently. From the 1980's onward, most of the musicals offered by the big producing companies are translations of Broadway hits, and musicals originally written in Dutch are not necessarily visited more often than foreign musicals (Brouwer, 1984).

1.1.2. The Limitations of Musical Translation

For this thesis, I will use Reiss' term 'expressive texts' as an umbrella term for the three genres used, which are prose, dialogue and song lyrics. This type includes all texts that portray "the communication of artistically organized content" (Reiss, 1981, p. 124). This term involves all types of creative writing, as well as all texts that do not specifically aim to be informative or operative. Although most texts are a combination of two of the text types, for example, a brochure is both informative and operative, while a commercial for a clothing brand can be both expressive and operative, this term still encompasses the general idea of expressive texts that do not have any particular aim outside of entertainment. Therefore, I will use the term 'expressive texts' to describe all types of texts that have entertaining or artistic values as a main function, mostly referring back to the three genres used in this thesis.

Translating musicals, or expressive texts in general, is a considerably demanding task. While all genres of translation come with their own limitations and difficulties, expressive texts are mainly difficult because of their dependence on style just as much as content to convey their message and achieve their function to entertain. According to Haque, literary translation is often considered "one of the highest forms of rendition because it is more than simply the translation of text. A literary translator must also be skilled enough to translate feelings, cultural nuances, humour and other delicate elements of a piece of work" (2012, p. 97). To make matters more difficult, poetry translators must, according to Boase-Beier, adhere to the style of the source poem, because the style shows the writer's attitude towards the content of the text (2004). While this is true for all texts to a certain extent, in poetry translation, a translator has to take into account a possible rhyme scheme and meter. Holmes (1988) proposes three approaches to this problem:

- A mimetic approach: in which the original form is replicated in the target language. While this remains the closest to the original style, it runs the risk that the poem is not pleasurable or understandable for its target audience.
- An analogical approach: in which the form is adapted to one that is found more often in the target culture. This approach makes sure that the target audience will experience the poem as best as they can, but removes the original form entirely.
- An organic approach: in which a form is chosen that the translator finds best suited for the source text. While this looks specifically at each individual poem and finds a solution for the problem of a specific form not present in the target culture, it is also subjective to the personal preferences of the translator (1988, p. 25).

Song translation poses additional limitations to that of poetry translation. According to Hieble, this is mainly, but not limited to:

> special problems of metre, rhyme, singability, matching musical notes with corresponding syllables or changing the one or the other, onomatopoetic considerations, or finding a whole series of similar words, where two, three and four persons sing together. (1958, p. 235)

While these specific problems can create interesting fields of study, there has been little research on the topic of song translation until quite recently. Many have described the relationship between text and music, which together work to produce the impact of the song, which therefore makes the translation of song lyrics considerably difficult (Tråvén, 2005; Apter, 1985; Golomb, 2005).

Franzon (2008) considers five different choices that a translator can make when starting to translate a song:

> 1. Leaving the song untranslated;
> 2. Translating the lyrics but not taking the music into account;
> 3. Writing new lyrics to the original music with no overt relation to the original lyrics;
> 4. Translating the lyrics and adapting the music accordingly – sometimes to the extent that a brand new composition is deemed necessary;
> 5. Adapting the translation to the original music. (Franzon 2008, p. 376)

These choices are in part linked to the function or *skopos* of the translation. For example, option two is not favourable when translating for a singable translation, but is a useful option when translating for an accompanying leaflet for a theatre production, or for a lyrics booklet found in CDs.

Low (2005) developed an approach to assess existing song translations, as well to test one's own translations during the translation process. His Pentathlon Approach focuses on the following five criteria which a translator must pay attention to, in order to achieve a functional translation:

- Singability: The translation must be able to fit on the existing music and can be sung as though it was the original work.
- Sense: The original sense, or meaning, of the text is preserved in the translation.
- Naturalness: The translation is in natural speech and follows correct lexis and syntax.
- Rhythm: The translation follows the rhythm of the original.
- Rhyme: The translation follows the rhyme scheme of the original.

While Low's approach was developed specifically for song translation, it could arguably be used for poetry translation as well. However, with the exception of his criteria rhyme there are no specifics given into the scoring method of a translated text, apart from stating these five categories and what they are supposed to test. Therefore, the pentathlon may not be as easy to use as an assessment method when comparing different song translations.

1.2 Machine Translation

1.2.1 A Brief History of Machine Translation

From the 1930's onwards, researchers have envisioned the possibilities of using automatic machines in the aid of translation (Henisz-Dostert, 1979). In this beginning fase, researchers believed that machines could eventually be taught to translate without the need for human interference and solve the problem of communication between people who did not speak the same language in a world that was striving towards global unity. Warren Weaver, one of the first researchers to be fully dedicated to machine translation, said on this topic:

> There is no need to do more than mention the obvious fact that a multiplicity of languages impedes cultural interchange between the peoples of the earth, and is a serious deterrent to international understanding. The present memorandum, assuming the validity and importance of this fact, contains some comments and suggestions bearing on the possibility of contributing at least something to the solution of the world-wide translation problem through the use of electronic computers of great capacity, flexibility, and speed. (1949, p. 15)

Although no ‘perfect’ machine has been made, as early researchers had strived to achieve, current machine translation has been working towards being functional in a professional setting, albeit with the aid of a human post-editor.

The history of machine translation can be divided into four main parts:

- 1950’s - 1980’s: Rule-Based Machine Translation (RBMT). Researchers believed that a machine could function in the same way as a human brain. The systems they made consisted of two components: the ‘rules’ in which the syntactic knowledge of the MT was stored, and a lexicon, which dealt with morphological, syntactic and semantic information.
- 1980’s - 1990’s: Example-Based Machine Translation (EBMT). This type of MT makes use of a bilingual corpus with parallel texts of a certain language pair, from which the MT makes possible matches to translate a new text. This type was more successful than RBMT, especially for language pairs that have vastly different grammatical rules, as the MT can now make use of previously translated phrases and sentences, instead of translating words that would then have to adhere to the new rules. However, EBMT needs a very extensive database in order to function reliably,

as each utterance needs an example. For language pairs with little reference material, this is not ideal.

- 1990's - 2015: Statistical Machine Translation (SMT). Research shifted its focus from a machine doing what a human brain can do, towards what it is actually good at: statistics. The MT would now be given a database of options and would calculate the probability of a word or phrase being translated in a certain manner, and then pick the option with the highest probability. SMT has gone through three phases: word-based, in which the MT looked at each individual word in its calculation; syntax-based, in which the computer looked at syntax-phrases; and finally, phrase-based, in which the MT first calculated the possibility of certain words being a phrase, and then calculating the probability of a translation. While compared to rule-based systems, SMT is significantly better in semantics (Costa-Jussà et al., 2012), but has certain limitations (e.g word order, subject-verb agreement, and tense) which do not allow it to work to its full potential (Vanmassenhove et al., 2016).
- 2015 - onwards: Neural Machine Translation (NMT). The current stage of machine translation, Neural Machine Translation makes use of an artificial neural network that predicts the likelihood of certain translated sequences throughout a document, instead of translating on a phrase or sentence level. Furthermore, NMT allows the computer to learn from previous translations and adapt without human interference.

NMT has been greatly influencing machine translation research and shows much potential. Compared to its predecessors, NMT translations contain fewer errors and result in a more fluent end product, as well as translate far less literal than the previous stages (Castilho et al., 2017a). For example, research by Constantine regarding the use of Google Translate for literary texts, specifically the translation of Voltaire into English, shows that contrary to the common conception that Machine Translation is only functional with short, non-creative

text, it is still competent in producing an adequate literary translation (2019). Furthermore, in their postscript, they iterate that a few weeks after completing their research, they re-entered the text used into Google Translate, which resulted in adapted and improved translations. This shows the most important aspect of NMT, it being able to learn and to adapt in a short timeframe.

However, it is not yet without flaws: while the errors made by NMT are significantly less severe, it does make them more difficult to identify (Castilho et al., 2017b). This might cause errors to go unidentified, which would result in a product that might contain small errors that a human translator might not have made. Another research by Abdulfattah and Yasser (2020), which compared machine translations performed on Google Translate and QTranslate to human made translations for J. K. Rowling's novel *Harry Potter and the Philosopher's Stone* (1997). and Edgar Allan Poe's short story *The Black Cat* (1843). From their analysis, which used a macro textual method concerning specifically the function, purpose, and effect of the machine translations. With this method, they found that roughly 44% of all errors made by both Google Translate and Qtranslate were lexical errors, roughly 30% in structural errors and 20% in semantic and pragmatic errors. This mostly affects the cohesion and comprehension of both translations and, Abdulfattah and Yasser argue, is due to the linguistic particularities of literary texts, which is in line with previous beliefs held by scholars. They nevertheless continue to make an argument that while the quality of their findings suggests many errors, the translated texts provide a basic, if not high, understanding of the text, which can be guided to a functioning translation with the help of human intervention. As even human translators often struggle with the complexities of literary translation, "it makes it highly understandable, in the aim of accuracy and completeness, that the two methodologies can be used hand in hand for the ultimate perfection" (Abdulfattah Yasser, 2020, p. 233).

Machine translation (MT) has seen a considerable amount of development in the last few decades and has greatly improved from its starting position. Although the eventual ideal of machine translation is to create a flawless translation without human interference, the reality sees that MT might only be usable as an aid of a human post-editor in order to create a functional translation (Hutchins, 1994). Although yet imperfect, researchers argue that if translators would be more willing to use machine translation in their practice and strive towards bettering these programs, machine translation could function as an aid in the translation process on a much larger scale (Läubli, S & Orrego-Carmona, D, 2017). However, some practitioners refrain from using machine translation, because it could "diminish translator's professional autonomy," as well as diminishing translation quality and, in the long run, might endanger translation as a profession (Taivalkoski-Shilov, 2018).

1.2.2 The Reception of Machine Translation as a Productivity Tool

Although research has shown the possibilities of machine translation as an aid for translators, and MT's are used widely to increase a human translator's productivity and consistency (Moorkens et al., 2018), the reception of MT use within a professional setting has not been completely positive. On the one hand, many translators make considerable use of technological tools like online dictionaries, and some find that the use of translation memory software helps them be more effective in their work. CAT-tools can, according to Taivalkoski-Shilov, "decrease the load on translators' working and long-term memory and release their cognitive resources for complex tasks by relieving them of repetitive and boring tasks" (2019, p. 691). On the other hand, those against the use of machine translation find that translation technology can "diminish translator's professional autonomy", and "might negatively influence the quality of translated texts" (Taivalkoski-Shilov, 2019, p. 691).

Furthermore, many professional translators still fear that MT might become too advanced, and human translators will be replaced completely by machines. However, as the International Federation of Translators states in one of their papers on MT:

> MT is unlikely to completely replace human translators in the foreseeable future. Leaving aside the area where MT is a feasible option, there will continue to be plenty of work for them. Professional translators, who have the appropriate skills and qualifications, will still be needed for demanding, high-quality products. (Heard, 2017, p. 131).

While this fear of being replaced by machine translation is understandable, the current quality of MT is not good enough to replace all translation efforts, and it is not expected to work on the same level as a human translator. However, MT could function as an aid for translators, and in some areas of translation could be used with the aid of a human post-editor.

1.3 The Assessment of Literary and Musical Machine Translation

1.3.1 Research on Machine Translation and Assessment Methods in General

Research within the field of machine translation is greatly diverse. First, there has been some research on the differences between SMT and NMT, for example the research done by Daems and Macken, in which they compared the translation quality of SMT and NMT output as well as their "perceived usability of SMT and NMT for post-editing according to professional translators" (2019, p. 118).

Another area in which research is being performed is that of quality assessment of several MT output. Daems et al. (2017) reported on the post-editing effort of MT output, both for beginning student translators and professionals. In their research, they looked at keystrokes and eye tracking for both groups, in order to evaluate how the quality of the

output influences the post-editing effort, as well as which type of errors were made by the MT. They concluded that while the quality of the machine translation output impacts both product and process post-editing effort, excluding average fixation duration and the pause ratio, "a more fine-grained MT quality analysis is needed to correctly estimate actual post-editing effort" (Daems et al., 2017, p. 1). Others, such as Snyman and Naudé (2003), focus specifically on the accuracy of the output. In their research using newspaper articles as well as minutes of meetings and reports, they mainly focussed both on whether the meaning of the text was preserved, and whether the output was grammatically correct, which they tested using a 10-point value scale which was then calculated into percentages. They concluded that, as the preservation of meaning was achieved in 60% of the cases but the text was grammatically correct only 40% of the text, machine translation still has a long way to go before it can be accurate enough to function without assistance.

Next to quality assessment, research within machine translation also focuses on how MT could help in a professional environment. Läubli et al. (2019) had four professional translators translate four German texts from Migros Bank, either using their own workbench, or post-editing MT-output. Their translated texts were then evaluated by two experts. They concluded in terms of speed that "while the minimum speed hardly differed between TM-ONLY and POST-EDIT, the latter allowed for higher average and maximum speed" (Läubli et al., 2019, p. 4). In terms of quality, they concluded that NMT post-editing can lead to texts that are of equal or slightly better quality when compared to human-translated texts. As mentioned before, most of these types of assessment are done by human evaluators. There are, however, some researchers, such as Ueffing & Ney (2007), who consider the possibility of creating an automatic assessment for translation output.

Most existing methods for the assessment of MT output are not necessarily aimed towards the translation of expressive texts. For my purposes, these methods described would

not be as useful, as they focus more on the literal adequacy, and not on style, which is an important part of literary translation. However, there is some research that does focus on the use of MT for literary texts, which will be described in the following section.

1.3.2 Previous Research on Machine Translation within Literature

There is little research on how machine translation could function as an aid within literary translation specifically. This is mostly due to the general understanding that MT is most successful when used on short sentences that can be translated in a literal manner, and not on expressive texts that use more complex and perhaps ambiguous language (Moorkens et al., 2018). Research by Toral and Way (2018), which evaluated NMT and PBSMT on literary translation and allowed consumers to assess NMT translation output compared to human translations, not only showed a significant improvement in NMT compared to PBSMT, but also showed that in one third of the cases, human evaluators assessed the NMT to be of equal quality of texts done by human translators.

Similarly, Tezcan, Hoste and Macken created the SCATE (Smart Computer-Aided Translation Environment) taxonomy, in which errors in MT output are classified into either fluency errors, which can be detected by only looking at the output, and accuracy errors, which can be detected by comparing the output with the original text (2017). These two classes are then further divided into several categories to better show in which area of fluency or accuracy the output is lacking. In a later article by Fonteyne, Tezcan and Macken, they described this taxonomy as “a hierarchical, fine-grained error taxonomy based on the well-known distinction between fluency and accuracy errors” (2020, p. 3783). In this same research, they improved on their SCATE error taxonomy by adding more specific categories in terms of style & register, as well as coherence. In this paper, they specifically translated and post-edited Agatha Christie’s novel *The Mysterious Affair at Styles* (1920) using Google

Translate. First, they used the improved SCATE error taxonomy in order to show the accuracy and fluency errors on the three levels specified. Two professional annotators worked on this project. They concluded that in this novel, 44% of all sentences did not contain any errors in its output, and most errors were made in the subcategory 'mistranslation' in terms of accuracy, followed by the fluency subcategories 'coherence' and 'style & register' (2020).

For other expressive texts, such as poetry or song lyrics, it is to be expected that, given the restrictive style of these types of texts as well as the singable quality of songs, MT would be even less suited for these translations. However, Lee (2011) has shown that for avant-garde poetry, machine translation can create unexpected new meanings and can aid in properly conveying the meaning of a poem. Furthermore, research by Genzel, Uszkoreit & Och (2010) focuses on making a system to be able to use an MT for poetic translation, in which the MT is able to create a text following the given stress pattern and rhyme scheme. In their research, they trained a French-English MT to translate several forms of poetry using the correct rhythm and stress patterns. Although the MT was technically able to conform to these constraints some of the time, they concluded that, considering both the stress-patterns and the system's ability to match a given constraint, their system was too slow and not impactful enough to work in the present iteration.

According to Toral & Way (2018), two developments in the past years have given a new perspective on the possibility of MT as an aid for literary translation specifically. First, the introduction of Neural Machine Translation. This new development in MT not only has great potential for translation in general, but also shows a better understanding of cultural references and does not depend on literal translations, which especially for expressive texts will prove more useful than previous statistical MT's. Secondly, the increased presence of electronic books on the Internet, which may not seem that impactful for translation practice,

ensures digital access to many electronic novels as well as their translations. As, according to Toral & Way, “the main resource required to train statistical MT systems is bilingual parallel text” (2018, 2), electronic books can provide a more complete and faster to create database consisting of novels, which would make it much more feasible to make MT systems specific to novels.

1.3.3 Moorkens’ Assessment of MT Output

While not many researchers consider literary machine translation, there are some theories on the assessment of machine translation output, as well as certain methods on how to achieve this assessment. Moorkens’ method (2018) consists of three steps of assessment: adequacy, error typology, and post-editing effort. His method offers an easy way to assess MT output, not only by scoring the raw output but also by timing how long it takes to post-edit the output. His method was tested initially by his own students, who were asked to translate and assess a Wikipedia article of choice. While his steps are easy to follow, they do not consider the style of a text but focus only on to what degree the MT maintains the meaning of the text, which is scored under Adequacy.

Adequacy, according to Moorkens, is used in this research as “a functional measure of equivalence between source and target text” (2018, p. 9). This part of the research is therefore mainly meant to assess how much the MT output corresponds in meaning with the source text. While the preservation of meaning is important, and one of the primary efforts of machine translation should be to correctly convey the meaning of the text, it is still desirable to maintain the style as well for expressive texts.

Furthermore, Moorkens’ categories for Error Typology are not ideal. For example, Moorkens has one category which contains all mistranslations of the output, including literal translations, translations that are simply wrong, and translations that are technically correct but put in the wrong gender, number, or case. Although this category shows that a text has

many mistranslations, it does not properly convey what kinds of mistranslations occur. A system in which there is a distinction between different types of mistranslations could clarify whether an MT produces too literal translations, incorrect grammar, or semantically incorrect segments. Farrús et al. (2012), while only focussing on the errors found in raw output, and not considering the more practical post-editing effort, offer a more complete method in terms of error typology. In their research, they considered the errors found in the Spanish-to-Catalan output of 711 sentences, which were taken from two newspapers. They make use of five main error categories: orthographic, morphological, lexical, semantic and syntactic. This method considers mistranslations at different levels of language, which makes it easier to determine which aspects of language are difficult for the MT in different types of expressive texts.

Moorkens himself uses his method primarily to assess the output of literary MT, but this method could be used for all types of MT output. However, in the case of expressive texts where form can be restrictive, such as certain types of poetry or song lyrics, it should be taken into account whether we can expect MT to replicate this into the target language. Nevertheless, while the quality of MT output has increased considerably, it is still imperfect and needs a human post-editor to improve the output.

To sum up, this chapter has given an overview of the history of machine translation and its various phases, focussing mostly on the current state of research within NMT. Many agree that although NMT is not yet perfected and is unable to function as a method of translation without human interference at some point in the process, NMT has made remarkable progressions in the last few decades. Going against the common belief that NMT is only of use in texts that do not require creative skill or adherence to style to be effective, there has been research into NMT for the field of Literature that shows promises. However, there has been little research done on the specific uses of NMT within the translation for

musical theatre. In the following chapter, I will discuss the method used for this thesis, based on the method previously described by Moorkens.

2. Method

2.1 Materials

T*he Strange Case of Dr. Jekyll and Mr. Hyde* was written by Robert Louis Stevenson and first published in 1886. The novel is predominantly told through the viewpoint of legal practitioner Gabriel John Utterson, a good friend of the scientist Henry Jekyll. Dr. Jekyll researches the duality of the human spirit and attempts to separate a human soul in two parts: good and evil. When his request for human testing is denied, he decides to test his theory on himself. By drinking a self-made serum, he becomes the evil Mr. Hyde for a limited time and does all kinds of wicked things his mind normally suppresses. The more he takes of this serum, however, the more difficult it becomes to turn back to Dr. Jekyll. Mr. Hyde becomes more prominent in his life and takes over Dr. Jekyll's body more and more, using it to kill innocent people. Eventually, Dr. Jekyll confesses of his crimes in a letter to his friend Gabriel, ending this letter and the novel with the line "[h]ere then, as I lay down the pen and proceed to seal up my confession, I bring the life of that unhappy Henry Jekyll to an end" (Stevenson, 1886, pp. 66), likely committing suicide soon after. Although early reviews of the book were uncertain how to treat the work, as it was unlike anything Stevenson had written before, it soon became a commercial success (Luckhurst, 2006).

The Musical *Jekyll & Hyde* was first produced in 1990 by Frank Wildhorn and Steve Cudon, and is loosely based on the novel by Stevenson. Although the original production had over 1500 performances on Broadway, it did not obtain the same commercial success as the novel, instead losing $1.5 million at closing on its first release (Idbd, 2020). Robert Cuccioli won an Outer Critics Circle Award as Best Leading Actor for his Dr. Jekyll, and Christiane

Noll won a Friends of New York Theatre Award for her Emma Carew. In 2013, the musical was revived for the Marquis theatre, and later for an American tour, in which many of the songs and scenes were altered, which was received much more positively by audiences and critics (Stewart, 2013). This new version will be used for this thesis. The original musical was translated into 14 languages, including a Flemish version in 1997. This version was not adopted by any of the large Dutch producers, and the musical was only performed in the Netherlands by a small entertainment company called Storytellers which produced the revived musical from December 20, 2019 till January 2, 2020 (Storytellers, 2020).

For this thesis, three excerpts from the novel were selected. Each of these excerpts are roughly 400 words and are all beginning paragraphs for important chapters. They were taken from the 2006 reprint by Oxford World's classics. The selected excerpts from the novel are as follows:

- *Chapter 1: Search for Mr. Hyde*, in which Mr. Utterson meets Mr. Hyde for the first time (p. 11).
- *Chapter 4: The Carew Murder Case*, in which Sir Danvers Carew was murdered by the hands of Mr. Hyde (pp. 20-21).
- *Chapter 10: Henry Jekyll's full statement of the case,* in which Dr. Jekyll explains in a letter to Mr. Utterson how his experiment with the duality of the human spirit has gone wrong (pp. 52-53).

The novel, as mentioned previously, was first published in 1886. As such, the style of the novel is slightly archaic. Most sentences are long, using several embedded clauses. Moreover, there is a marked to the general sentence structure. For example, in chapter one, line 2 states: "It was his custom of a Sunday, when this meal was over, to sit close by the fire, a volume of some dry divinity on his reading-desk, until the clock of the neighbouring church rang out the hour of twelve, when he would go soberly and gratefully to bed." Not only does this line have

several embedded clauses, there is little cohesion between the clauses, making it difficult for the MT to translate the structure correctly in correct Dutch

On the other hand, there are few metaphors that are unable to translate to Dutch, and little culture-specific or time-specific words or phrases. This would mean that the lexical style and figures of speech would be relatively easy to translate.

From the musical *Jekyll & Hyde*, three songs and three excerpts from the dialogue were selected based on their relevance within the musical. The following two factors were taken into consideration when selecting the songs: the number of present singers, and their melody. The three songs all contain run-on sentences, and while the grammar is easy to understand, the references and metaphors might prove difficult to be translated correctly by the MT. The dialogue excerpts were selected mainly based on how many different characters are talking in it to portray as many different classes and accents within the excerpts as possible. In general, the excerpts from the musical dialogue consist of smaller, mostly easy to follow sentences. This is mainly because the text is spoken out loud, and therefore needs to be understandable at the first time hearing it and should be easy to pronounce and remember for the actors. The dialogue excerpts were taken from a live recording, featuring Constantine Maroulis as Dr. Jekyll and Mr. Hyde. The dialogue from this performance (Walking The Way To Dawn # SoRiku, 2016) was transcribed specifically for this thesis.

The songs selected for the analysis are the following:

- *Facade*: In this song, the inhabitants of London describe how every person alive pretends to be something that they are not. This song is not strictly necessary to advance the plot but conveys the underlying message of Dr. Jekyll's work and general theme of the musical: every person has a version of themselves that they show to the world, and a version that they would rather hide. The song is repeated several times throughout the musical, often functioning as a bridge between different scenes. The

song is sung largely by the ensemble, while the four members of the board of Governors of the St. Jude's hospital have some solo lines. It contains both faster and slow parts and does not have any particular rhyme scheme. In terms of grammar, most stanzas consist of a singular, grammatically correct sentence, often with several embedded clauses. The lyrics often refer to the general theme of the musical: the duality of mankind. As such, there are many references to 'another self' and conflicting desires.

- *Someone like you*: This song is performed by Lucy, a prostitute, who falls in love with Dr. Jekyll because of the kindness he has shown her. It is a ballad, in which she laments her life, and starts to dream of a life in which Dr. Jekyll 'finds' her and falls in love with her as well. The melody is quite slow, and most vowels are elongated. In terms of stylistics, a large part of the text refers to flying or taking wing, as Lucy sees Dr. Jekyll as some way to be freed of her current life. Many lines are repeated or are only minorly adapted.
- *Your work and nothing more*: This song is made up of two duet-couples, Dr. Jekyll and John Utterson, and Emma Carew and Sir Danvers Carew. The melody of this song is fragmented, and each character has its own melody. The text mostly consists of run-on sentence and does not have a specific rhyme-scheme, instead relying on repeating vowels to create assonance.

The following excerpts were selected from the musical dialogue:

- *Act one, scene one*: Dr. Jekyll's proposal for his research is put before the board of Governors of St. Jude's hospital, and subsequently rejected for ethical reasons. The first half of the excerpt uses slightly official language, while the rest of the excerpt, when the governors talk among themselves, the characters often make quips and wordplays while making fun of Dr. Jekyll and the other people around them.

- *Act one, scene three*: Dr. Jekyll meets Lucy Harris in the brothel "The Red Rat". They have a short conversation, after which Dr. Jekyll goes home, leaving Lucy a business card in case she would ever 'need a friend'. In this excerpt, Lucy and the other woman, as well as the brother owner Spider, speak in a rough accent indicating they were not well-brought up, while Dr. Jekyll continues to speak politely.
- *Act two, scene one*: The board of Governors is murdered one by one by Mr. Hyde. This scene is split up in smaller parts in between the song "Murder, murder". This excerpt consists of several smaller parts, each providing different aspects that need to be considered. The parts with the Board of Governors mostly consist of quips and sarcastic comments, while the part in which Dr. Jekyll asks for certain chemicals is much more serious. Additionally, Mr. Hyde mostly delivers brooding and sarcastic witticisms.

Based on the materials, we can hypothesize that:

- The quality of the MT output of the prose excerpts might be compromised by the marked language of the text.
- The simplicity of the dialogue might cause these excerpts to be translated, though literal, very correctly and completely, which would indicate very little post-editing effort.
- The song lyrics might be easy to be translated by the MT, as the text itself is relatively easy. On the other hand, as the text depends on metaphors and short, incomplete utterances, the text might not be grammatically correct or hard to understand.

2.2. Method

All of the selected excerpts were translated using Google Translate on the 21st of May, 2020. Given the speed at which NMT's such as Google Translate improve, it is likely that any

replication of this analysis will result in different output. Song lyrics lines were occasionally grouped together in one sentence. This was done in order to ensure that the MT recognised run-on sentences as one complete sentence, instead of translating each excerpt individually.

Google Translate was used for various reasons. First, I did not have the means to create a specialised MT for the genres used, nor access to any existing MT's. Second, beginning translators will likely choose this MT as well, for the same reasons, as they are unlikely to receive a specially trained MT by their employer. While the built-in MT plugins in CAT-tools like MemoQ or SDL Trados could also be used, these are less suitable for the current project considering that they have limited data available due to their protection of client data. Google Translate will have much more reference data at its disposal and will therefore likely perform better than these CAT-tool plugins.

Before assessing the MT output, there were some preliminary efforts needing to be performed before. First, as there is no official playbook that can be accessed online, I transcribed the excerpts for translation from a recording available on YouTube (Walking The Way To Dawn # SoRiku, 2016). Secondly, I established the rhyme scheme and metre of the lyrics, while also paying attention to the melody of the song in terms of speed, to properly address these matters while translating. This was a necessary step and, considering a human translator would need to perform these efforts themself as well or would be provided with a scansion overview before translating, is not included within the assignment itself. As mentioned previously, I will use an adapted version of Moorkens' (2018) method for evaluating MT output for this thesis. Moorkens' method consists of three steps:

1. Adequacy: The MT output is assessed and graded on each sentence or line how well it conveys the original content of the text.

2. Error Typology: Translation errors in the MT output are categorised and counted by category or error.
3. Post-editing effort: The MT output is edited by a human editor and it is timed how long each excerpt takes to post-edit.

In his own research, Moorkens set out to compare NMT output with SMT output, using his own students to evaluate and post-edit both outputs in a two-hour timeframe. This exercise accounted for 20% of their final grade within the course. In total, 46 students in their second year of their bachelor in Translation at Dublin University were given the assignment to evaluate and post-edit 20 segments from a wikipedia page of their choice, using Microsoft Translator MT. Of these 20 segments, 10 were translated using NMT, while the other 10 were translated using SMT. This task was later repeated by 9 PhD-candidates.

It should be mentioned that, in his own research, Moorkens started by giving students the MT output and asked them to perform the third step, post-editing effort, first and then moving on to adequacy and error typology. Furthermore, his method is a general method for the MT translation of literature and would not reflect the effort that is specific for the translation of song lyrics, nor the specific translation problems that an MT cannot provide, such as rhythm or rhyme. Therefore, I will also test adequacy in terms of style, instead of only in terms of content.

2.1.1. Adequacy

First, the excerpts are given an average grade out of four in content, as well as four categories of stylistics. Content is graded on how much of the original content of the text was correctly translated by the MT. Moorkens (2018) used only content as an assessment in terms of adequacy. However, as the style of expressive texts is of equal importance. In the current research, this was altered by adding a new category to the model, in which the output will

also be scored on its preservation of style. Leech and Short's (2007) checklist for stylistic features can be used to analyse style in terms of four categories of language: lexical categories, grammatical categories, figures of speech, and cohesion and context. While these categories are not specific for machine translation, they can be used to systematically test in which way the style of the MT output is lacking.

Style depends on several categories of language, and it is expected that the different genres tested in this thesis will have different categories in which it falters. Therefore, style is graded on a scale of one to four on each of the following categories by Leech and Short (2007):

- Lexical categories: Considers the vocabulary and register of the translation.
- Grammatical categories: Considers the syntax and grammar of the translation.
- Figures of speech: Considers how the MT has translated specific figures of speech.
- Cohesion and context: Considers the general cohesion between phrases, sentences and in how much the translation fits into the context of the source text.

The grading mostly depends on the errors made in that category, which lowers the grade. Occasionally, the grade is also lowered not because of an error, but because there are other alternatives that would have fit the text better. After grading each category, an average grade for style is calculated.

2.2.2. Error Typology

Moorkens (2018) distinguishes between four different types of errors. However, for the use in this thesis, this might not be specific enough to differentiate between the text genres. Therefore, the error typology by Farrús et al. (2018) will be used instead, as it makes a distinction between which error category is found in the translation. The categories are as follows:

- Orthographic errors: Errors made in terms of spelling or punctuation.

- Morphological errors: Grammatical errors made on word level, in which the MT does translate a word with the correct translation, but the grammatical structure is wrong in terms of gender, case or number.
- Syntactic errors: Grammatical errors made on sentence level, such as word order or the grammatical structure between clauses
- Lexical errors: Errors made on word-level, in which the MT mistranslates or does not translate certain words.
- Semantic errors: Errors made in semantics, in which the MT chooses the wrong translation when a certain word has two different meanings in the target text and vice versa.
- Syntactic errors: Errors made in grammar on sentence level, such as word order or the grammatical structure between clauses.

In the table below all categories are set out once again, showing all subtypes of errors that are included under them:

Orthographic errors	**Morphological errors**	**Lexical errors**	**Semantic errors**	**Syntactic errors**
Punctuation marks	lack of gender concordance	Incorrect words	polysemy	prepositions
accents	lack of number concordance	unknown words	homonymy	verbal periphrases
capital and lowercase letters	verbal morphology	missing target words	-	clitics
joined words	lexical morphology	extra target words	-	reordering
extra spaces	-	-	-	-
apostrophes	-	-	-	-

Table 1. Error Typology

Each excerpt was assessed on these error categories, and the errors were counted and then averaged per genre (novel, song, dialogue).

2.2.3. Post-editing effort

In Moorkens' research (2018), students had to time how long they spent on editing the MT output. The aim of the post-editing assignment is to deliver a publishable end-product. The original assignment by Moorkens did not include the post-edited product from his students, and he did not give any specifics how well his students performed this assignment other than the time it took to perform, which I will not do either. In this part of the assignment, only the effort it takes is important to assess the quality of the MT, which does not depend on the quality of the human post-editors. The time of each excerpt, as well as an average time per genre, will be included in the results. I will further discuss the difficulties and considerations of post-editing for all three genres.

2.3 Expectations and Limitations

Based on the three research questions discussed before, as well as previous research within Machine Translation, I expect the following to occur:

RQ1: Is the quality of MT output of song lyrics and dialogue comparable to that of written prose and what are the main difficulties encountered by machine translation in these three areas?

I expect the MT output of both song texts and dialogue to be quite literal. For dialogue, this does not directly mean the translation will be unsuccessful and will in many areas read as (almost) normal speech. The main difficulties will be altering cultural references or metaphors that are not present in the target culture. However, as the songs are fitted to music, and need to follow melody, rhythm and, for these songs, have some dependency on rhyme, it will not function as song lyrics. While there still will be metaphors and other

references that need changing as well, the main difficulties for song translation will be the song aspect.

RQ2: Is the post-editing effort of MT output on song texts and dialogue comparable to written prose, and is the post-editing effort of MT output less demanding than translating from scratch?

I expect the post-editing effort of the dialogue to be the lowest of the three genres, as the relatively easy spoken text will likely be unproblematic for the MT to translate correctly, and the text mainly needs to be fitted to the target culture, as well as altering some word choices. The editing effort of song lyrics will be much higher, as a literal translation simply will not be functional in that state. However, it is expected that the editing effort will be lower than a hypothetical translation made from scratch, as the MT output functions as a preliminary literal translation that the translator does not need to do themselves anymore.

RQ3: Can machine translation function as an aid for (the beginning) translator in musical translation?

Especially for less experienced translators, MT output could function as a starting point for translation. The MT provides a translator with a literal translation, which especially for song translation is the first step in the translation progress. A translator then does not need to consider the literal meaning as much anymore, and instead focus on fitting the text to the melody and the rhyme scheme. This could possibly cut down the time needed to translate songs, as the translator no longer needs to create a literal base-line translation first.

It should be mentioned that this thesis will be focussing on a single musical, and only specific excerpts and songs from this musical. Therefore, though this research will be able to give an idea if Machine Translation can be used as an aid in song translation, more research will be required to come to a general conclusion. Furthermore, as I am the only assessor in

this thesis, it is considerably subjective, and any conclusions drawn should ideally be confirmed by more extensive replication studies using more data and more assessors.

3. Results

In this master thesis, I will assess the MT output of Google Translate for 3 excerpts from the novel *The Strange Case of dr. Jekyll and mr. Hyde* (Stevenson, 1886), as well as three song lyrics and three dialogue scenes from the musical *Jekyll & Hyde* (2013). This will be achieved using Moorkens' methodology for Machine Translation assessment (2018), adapted using Leech and Short's stylistic categories (2007) for adequacy, and the error typology by Farrús et al. (2018). In this chapter I will discuss the results of my assessment, which will be the foundation to answer the three research questions:

> RQ1: Is the quality of MT output of song lyrics and dialogue comparable to that of written prose and what are the main difficulties encountered by machine translation in these three areas?
>
> RQ2: Is the post-editing effort of MT output for song lyrics and dialogue comparable to written prose, and is the post-editing effort of MT output less demanding translating from scratch?
>
> RQ3: In what ways can machine translation function as an aid for (the beginning) translator in musical translation?

I will first discuss the results found in the adequacy section of the assessment, moving towards error typology, and finally discuss the post-editing effort.

3.1 Adequacy

For adequacy, I have adapted Moorkens' methodology to reflect not only on how adequately the MT output portrays the content of the source, but also to reflect on the style of the source. To do this, I have used Leech and Short's stylistic categories (2007), which consists of four categories: Lexical, Grammatical, Figures of Speech and Cohesion and Context.

Each sentence or line was given a separate score from one to four, after which an average was calculated for each category. The grading for this was as follows:

- 1: Content or stylistic was severely compromised in MT output; output would need to be fully edited to be functional.
- 2: Content or stylistic was majorly compromised in MT output; more than half of output would need to be edited to be functional.
- 3: Content or stylistic was almost completely intact in MT output; some edits needed.
- 4: Content or stylistic was completely intact in MT output; little to no edits needed.

For example, if a specific excerpt scores a 2.5 in the category Grammatical style, it can be assumed that the quality in most sentences for grammar is adequate, but still contains some errors and would need significant editing time in order to be functional.

Overall, all three genres were given considerably high marks. In the table below the scores for each excerpt are included, as well as an average for each genre for an easier comparison:

	Content	**Lexical style**	**Grammatical style**	**Figures of speech**	**Cohesion and context**
Prose 1	2.8	2.4	2.08	3.25	2.67
Prose 2	2.81	2.56	3.19	2.94	3.06
Prose 3	3.18	3.09	3.45	3.73	3.64
Prose average	**2.93**	**2.68**	**2.91**	**3.31**	**3.12**
Dialogue 1	3.26	2.95	3.36	3.18	3.13
Dialogue 2	2.97	2.67	3.24	2.97	2.88
Dialogue 3	3.47	3.41	3.56	3.47	3.44
Dialogue average	**3.23**	**3.01**	**3.39**	**3.21**	**3.15**
Song 1	2.99	2.92	3.09	2.95	2.96
Song 2	2.92	3	2.89	3.06	2.94
Song 3	3.19	3.29	3.17	3.45	3.1
Song lyrics average	**3.03**	**3.07**	**3.05**	**3.15**	**3**

Table 1: Score adequacy

3.1.1 Prose

Based on these results it could be tempting to conclude that in terms of content, as well as lexical style and grammatical style, prose MT translation seems to be less adequate than the MT output for the selected dialogue and song lyrics. However, it should be mentioned that the age difference between when the prose was written (1886) and when the musical was written (1990) may very well be one of the main reasons why the prose output seems to be less adequate. The MT has little reference to data from prose written and translated 19th century, and as such bases its decision on modern language data. It is possible to train an MT using data from this time period specifically, but for the scope of this thesis, this was not

possible. Furthermore, the prose excerpts chosen were grammatically quite complex, using many embedded clauses and complex sentence structures. These clauses will likely not be translated correctly, an example of this occurs in the second excerpt, line 14:

The old gentleman took a step back, with the air of one very much surprised and a trifle hurt; and at that Mr. Hyde broke out of all bounds and clubbed him to the earth.	De oude heer deed een stap achteruit, met de uitstraling van een zeer verbaasde en een klein beetje gekwetst; en toen brak meneer Hyde alle grenzen uit en knuppelde hem op de aarde.

The clause ‘with the air of one’ is the most prominent instigator of this mistranslation. Where a human translator would have translated this in the spirit of ‘als iemand’, or else translate possibly the following clause as ‘een zeer verbaasde en enigszins gekwetste man’, the MT omitted this which resulted in the rest of this clause to be mistranslated. While this last proposal would not be feasible to expect from an MT, the first translation is still literal, and could have been possible for the MT to perform. This error is counted into two categories, mainly in grammatical style, which was given a 2 out of 4. One point was deducted from lexical style as well, considering that the mistranslation of the word ‘een’ caused the rest of the sentence to be grammatically incorrect as well.

Comparing the three genres, prose performed the worst in lexical style. This is mainly caused by the age of the text. Furthermore, the scores given for grammatical style and overall content were comparatively low as well. This is due to the fact that the prose excerpts were considerably more complex than the other two genres, and often contained several embedded clauses in a single sentence. Having established this, the MT performed surprisingly well in certain longer sentences, while making several large errors in relatively small sentences, suggesting that MT output quality depends largely on sentence complexity, rather than length. For example, the first line of the first excerpt is relatively short, but due to its relative complexity contains several mistranslations:

THAT evening Mr. Utterson came home to his bachelor house in sombre spirits and sat down to dinner without relish.	Die avond kwam meneer Utterson thuis bij zijn vrijgezel huis in sombere geesten en ging zitten zonder diner.

Here, the clauses ‘bachelor house’, ‘sombre spirits’ and ‘without relish’ were all mistranslated, meaning roughly half of this excerpt is mistranslated.

On the other hand, consider the following example from the same excerpt. This example is all one line, in which I showed the errors made by the MT in bold due to the length of the sentence and the relatively small and inconspicuous errors:

The will was holograph, for Mr. Utterson, though he took charge of it now that it was made, had refused to lend the least assistance in the making of it; it provided not only that, in case of the decease of Henry Jekyll, M.D., D.C.L., LL.D., F.R.S., etc., all his possessions were to pass into the hands of his ‘friend and benefactor Edward Hyde,’ but that in case of Dr. Jekyll’s ‘disappearance or unexplained absence for any period exceeding three calendar months,’ the said Edward Hyde should step into the said Henry Jekyll’s shoes without further delay and free from any burthen or obligation, beyond the payment of a few small sums to the members of the doctor’s household.	Het testament was **holografisch**, **voor** Mr. Utterson, hoewel hij er de leiding over nam nu het gemaakt was, had geweigerd de minste hulp te verlenen bij het maken ervan; het voorzag niet alleen (**dat**) in het geval van het overlijden van Henry Jekyll, M.D., D.C.L., LL.D., F.R.S., etc., al zijn bezittingen zouden overgaan in de handen van zijn 'vriend en weldoener Edward Hyde, 'maar dat in het geval van Dr. Jekyll's 'verdwijning of onverklaarbare afwezigheid voor elke periode langer dan drie kalendermaanden ', **zei** Edward Hyde **moet** in de schoenen van de genoemde Henry Jekyll stappen zonder verdere vertraging en vrij van enige last of verplichting, dan de betaling van een paar kleine bedragen aan de leden van het huis van de dokter.

While this is a very long sentence, there are just little inconsistencies or errors within the structure, and the few errors found are on a lexical level. First, the word ‘holograph’ was translated using the wrong sense, namely holographic. The other mistakes were relatively small, such as the omitting of the word ‘dat’ to make the underlying grammatical structure correct or translating the clause ‘said Edward Hyde’ in a too literal manner.

3.1.2 Dialogue

When assessing these excerpts, it should be taken into account that these are not fully representative of real speech dialogue, which is often filled with hesitations, sentence structure changes and filler words. This would make automated translating quite difficult to

achieve, as the output would be equally riddled with such spoken errors. However, fictional dialogue such as discussed in this master thesis often does not contain such interjections and may therefore be easier to translate. This is also true for two of the chosen excerpts, which are more stylised than regular spoken speech would be. There are very little interjections or structural changes, which would be necessary for the actors in order to memorize and perform the lines. Each line is complete and grammatically correct, while still being relatively simpler than for example the prose excerpts.

Most importantly, there is a difference in register and style between two groups of characters. One the one hand, there is the significantly larger group of Victorian English upper-class characters, which includes Dr. Jekyll, mostly represented in excerpt one and three. On the other hand, there is a group of lower-class characters, mostly represented by Lucy, and some minor characters found in excerpt two, detailing the first meeting of Dr. Jekyll and Lucy. In order to portray the socio-economic differences between Lucy and the upper class which Dr. Jekyll is usually involved with, she and the inhabitants of The Red Rat speak with an accent, but also use a much looser grammatical structure compared to the other characters. For example, when Dr. Jekyll asks Lucy how she ended up in the brothel, she responds:

> “One fine night the prince of Wales wasn’t going to make me queen of England, I wanted to look for something more romantic. Which is why I chose this little be rathole for all my dreams to come true. Know what I mean?”

The language presented by the lower class could be considered closer to spoken language as compared to the upper class. As in this example, the grammar is imperfect and tends to be using little linking words to glue unrelated sentences together.

As discussed in the Method chapter, I expected dialogue to result in the best MT output of the three genres, which it did indeed for most categories. Most of the errors made were lexical and continuity errors, which consequently also affected cohesion. For example,

all three excerpts result in problems with the difference between formal and informal ‘you’ in Dutch. In English, there is no difference between the formal and informal use of ‘you’, while in Dutch, its translation can either be the formal ‘*u*’ or informal, ‘*jij*’/‘*je*’. On many occasions, the translation contained both the formal and informal option of the translation for ‘you’, even when they were in the same paragraph. This resulted in the following translation:

I’ll have to put up a statue then, for your general memory. You miserable old bugger. You’re next, Savage.	Ik zal dan een standbeeld moeten neerzetten, voor uw algemene herinnering. Jij ellendige oude klootzak. Jij bent de volgende, Savage.

The inconsistency in the translation of the word ‘you’ takes away from the cohesion of the text and does need to be edited in the later stages. In terms of adequacy, this results in lower scores for ‘context & cohesion’.

Having said that, most of the grammar and sentence structure were very well executed by the MT for this genre. This was mainly because, as the lines need to be spoken out loud and be understood in one sitting by an audience, the dialogue mainly consists of short and simple sentences. For example, at the start of excerpt two, three characters have the following exchange:

NELLIE: Hey, watch out. Spider’s in an evil mood.	NELLIE: Hé, kijk uit. Spider is in een slechte bui.
LUCY: And why should tonight be different?	LUCY: En waarom zou vanavond anders zijn?
SPIDER: A gentleman customer, let’s see. Business, before pleasure.	SPIDER: Laten we eens kijken. Zaken gaan voor plezier.
LUCY: Don’t worry Spider, pleasure is my business.	LUCY: Maak je geen zorgen Spider, plezier is mijn zaak.
SPIDER: Quite right, my dear. And business is my pleasure. Be good to him, Luce, some men know quality when they see it.	SPIDER: Precies, mijn liefste. En zaken doen is mij een genoegen. Wees lief voor hem, Luce, sommige mannen kennen kwaliteit als ze het zien.

There are some errors in lexical style, as well as in cohesion and context, most notably Spider’s first line in which roughly a third of the sentence was omitted. However, the grammatical structure of the source was largely maintained in the output. This demonstrates

that, while still in need of post-editing due to lexical stylistics and occasional awkward literal translations, the overall structure lends itself to being translated by an MT.

3.1.3 Song lyrics

In general, the three song lyrics were relatively simple in form. This is due mainly because there is a necessity for this to be singable and, similarly to the dialogue excerpts, they should also be easy to remember. Therefore, the songs mostly have short sentences with very little clauses, as well as many repetitions both on a syntactic and a lexical level. For example, in the song *Your work and nothing more*, each of the characters has one phrase that they repeat in various ways throughout the songs.

It should be mentioned that I did not take the rhyme and rhythm of the source text into account when assessing the lyrics output because of two reasons. Firstly, as Google Translate cannot be altered or limited in its output to adhere to a rhyme scheme or a certain rhythm, it could not be reasonably expected of the MT to do this by itself. Secondly, as the other two genres do not have to adhere to such a restrictive form, it would in all cases be graded down just for this reason, while the rest of the output might be of the same quality as the other two genres.

Next to this, I have combined some of the lines into one sentence before putting the lyrics through the MT. This was done because, separately, the MT output was unusable. Without combining the lines into complete sentences, verbs would often be translated in an infinitive form and cohesion was non-existent. For example:

This increasing Isolation only adds To your frustration And it could Endanger Your Career...	Dit neemt toe Isolatie voegt alleen toe Tot je frustratie En dat zou kunnen Breng uw carrière in gevaar...

While this is expected with MT output, as there often is little coherence between separated lines, it would create an unfair assessment as this would render most of the excerpts incorrect or of low quality, while the other two genres do not have this problem. When these sentences were combined, the MT performed more efficiently:

This increasing Isolation only adds to your frustration And it could endanger your career...	Deze toenemende isolatie voegt alleen maar toe Tot je frustratie En het kan je carrière in gevaar brengen

This is an interesting phenomenon, as this suggests that not only the language itself is important for the quality of the MT output, but also the way in which it is inserted into the MT. Especially for song lyrics, but also for other types of text, it might prove useful to allow time for a human pre-editor. They can eliminate problematic structures and clauses before the MT output is created, which would greatly influence the quality and, afterwards, decrease post-editing effort.

After pre-editing, the MT translation of the song lyrics was of a surprisingly high quality. This can be explained by the text being, similar to the dialogue, grammatically simple. This means that, in most cases, the grammatical structure will be translated correctly, and only small errors occur in other areas. Of the three songs, *Facade* was given the lowest score. This was mainly due to the MT mistranslating the word ‘facade’ as ‘gevel’ nine out of the ten occurrences (the tenth time it did translate correctly to ‘façade’). The translation also did not convey the correct use of accents. Many lines conveyed the singing accent of the characters with the replacement of ‘g’ with an accent. The MT likely does not have any reference to such occurrences, which resulted in the translation not reflecting this and keeping the source in the translated text. For example, consider the following lines:

That's how our little -	Dat is hoe onze kleine -
Game is played,	Game wordt gespeeld,
Livin' like a masquerade	**Livin** 'als een maskerade
Actin' a bizarre charade -	**Actin** 'a bizarre charade -
While playing the saint!	Tijdens het spelen van de heilige

This excerpt shows twice that when the apostrophe is used as an indication for accent, the MT does not translate the words that go in front of it. This can occur more regularly in fiction dialogues as well, but is also used in songs where two words are sung together to fit the melody. In such cases, it might prove useful to pre-edit such occurrences to greatly improve the output.

The third song, *Someone Like You*, resulted in the best output. This was the least complex song of the three, with many repetitions and straightforward clauses. Especially in terms of content and cohesion, this song scored higher than average. While it was not effective to be sung immediately, it would not need a lot of post-editing to make the lyrics fit properly on the melody. However, there were still many mistakes in the translation, specifically in the translation of verbs. This was likely due to the subject of the sentence being in one line, while the verb only occurred in the line after. For example, the second verse illustrates this as follows:

The past is holding me	Het verleden houdt me vast,
Keeping life at bay	Het leven op afstand houden,
I wander lost in yesterday	Ik dwaal verloren gisteren,
Wanting to fly	Willen vliegen -
But scared to try	Maar bang om te proberen.

In this line, ‘the past’ functions as a subject for the both the first and the second line. Implied in the second line is ‘the past is keeping life at bay’. With this implied meaning, the

translation should be along the lines of ‘houdt het leven op afstand’. The output does not reflect this implied meaning, and instead considers the two lines as separate units.

To sum up the first part of this three-part study, the adequacy of the MT output was quite high for all three genres, with the highest adequacy being in the dialogue excerpts and song lyrics. This is mainly due to the relative simplicity of these texts in terms of grammatical structure. It proved useful to expand the method to look not only at overall content, but also to style, as this provides a larger understanding on which part of the segment the MT output was insufficient. Overall, it can be said that mainly lexical style proved to be less sufficient in the output compared to figures of speech and grammatical style, especially within the dialogue excerpts.

One caveat to mention is that the content category in this study often acted as an average score for the other four categories. To expand, if a segment were to be lacking in lexical style or grammatical style and being given a 1 or 2, but still having higher points in cohesion or figures of speech, the content score would often end as the average between those categories.

3.2 Error Typology

In the second part of this study, error typology, I adjusted Moorkens’ method using the categories from Fargus et al. Moorkens’ method used four error types: word order errors, mistranslations, omissions, and additions. This was adapted into five categories that more specifically consider the type of mistranslation. The three genres were assessed, and errors were picked out, giving the total results in the table below.

Overall, it should be mentioned that in all excerpts (excluding the third song *Someone like you*) the most common error was on a lexical level. This error type consists of mistranslations, untranslated target words, as well as extra or missing target words. Arguably,

this category might be further divided into actual lexical mistranslations and untranslated words, and extra or missing target words. Regardless, these two categories will still be the two most seen errors, so it would only be for clarity's sake.

Included on the next page are the total and average number of errors per genre. Appendix A, B and C show a complete overview of all errors made for each individual genre and excerpt.

	Orthographic errors	**Morphological errors**	**Lexical errors**	**Semantic errors**	**Syntactic errors**	**Total errors**
Prose 1	5	4	13	10	10	**42**
Prose 2	2	3	13	3	7	**28**
Prose 3	2	2	9	5	0	**17**
Prose average	**3**	**3**	**11.7**	**6**	**5.7**	**29**
Dialogue 1	1	5	15	11	2	**32**
Dialogue 2	0	0	12	4	2	**17**
Dialogue 3	0	6	10	4	4	**24**
Dialogue average	**0.33**	**3.66**	**12.33**	**6.33**	**2.66**	**24.33**
Song 1	10	3	34	3	5	**55**
Song 2	1	9	21	7	5	**45**
Song 3	0	5	2	2	3	**11**
Song lyrics average	**3.67**	**5.66**	**19**	**4**	**4.33**	**37**

Table 2: Error typology total scores of each category and genre

3.2.1 Prose

As mentioned previously, most errors were caused by the relatively older and complex English used. This genre contained the highest total of syntactic errors. This was caused mostly because of the generally complex sentences compared to the other two genres which primarily consist of short or simple clauses, but also because the original English follows less common syntax, which the MT did not alter to correct Dutch syntax. For example, consider the eighth sentence of the second excerpt:

And as she so sat she became aware of an aged and beautiful gentleman with white hair, drawing near along the lane; and advancing to meet him, another and very small gentleman, to whom at first she paid less attention.	En terwijl ze daar zat, werd ze zich bewust van een oude en mooie heer met wit haar, die langs de laan naderde; en ging op weg om hem te ontmoeten, een andere en heel kleine heer, aan wie ze eerst besteedde minder aandacht.

The final clause of this segment is grammatically correct in English, although most people would consider rewriting this line as: “advancing to meet him was another and very small gentleman, to whom she paid less attention at first.” The underlying syntax was not picked up by the MT and translated the same way as the original was phrased. However, in Dutch this structure is incorrect and should look somewhat like the following: “en op weg om hem te ontmoeten, kwam een andere en heel kleine heer, aan wie ze eerst minder aandacht besteedde.”

Although semantic errors occurred in roughly similar amounts in all three genres, one of the more interesting cases was in the title of the second excerpt, which translated “the Carew murder case” as “de Carew moordkoffer”. The word ‘case’ can have several meanings, importantly for this error either ‘suitcase’ or ‘murder case’. These two meanings are respectively translated into Dutch as ‘koffer’ and ‘moordzaak’. The MT made use of the wrong translation, which is interesting as the word murder should have acted as an indicator which sense of the word to use. When putting only the world murder case through Google Translate, it does translate it correctly to ‘moordzaak’.

Lastly, there was a distinct difference in output quality between the first two excerpts and the third. In total, the first excerpt has 43 errors, the second 28 and the third 19. The third excerpt did not contain any syntax errors. This might be explained because the final excerpt was told by dr. Henry Jekyll in the form of a letter, instead of following Mr. Utterson. Henry Jekyll used slightly shorter lines, with fewer embedded clauses. Even when embedded clauses are present, they are less complex than found earlier in the text. For example:

In this case, I was driven to reflect deeply and inveterately on that hard law of life, which lies at the root of religion and is one of the most plentiful springs of distress.	In dit geval was ik gedreven om diep en onverbiddelijk na te denken over die harde levenswet, die aan de wortel van religie ligt en een van de meest overvloedige bronnen van nood is.

While there are still some errors, such as the use of ‘diep en onverbiddelijk’ and stylistic choices that a post-editor would make to create a better flowing product, this is a quite strong and simple translation with little errors on a syntactic or lexical level. This resulted in an easier text, which the MT was more likely to follow. Overall, the average of each error category was lowered by this excerpt.

3.2.2 Dialogue

Dialogue contained the fewest total errors, and the errors that were made were mostly small. For example, several lexical errors was attributed to the MT not being consistent with whether or not it translated words such as Sir and Mr. While this is understandable, as the instances are separated by different character lines and might cause the MT to translate them separate them, it does lower the continuity of the entire text. As such, I decided to count each time the MT maintained the English words.

Otherwise, the dialogue excerpts mostly other small errors where one word was mistranslated or added articles where they should not be added. There were some larger errors, such as the following line by Lucy in the second excerpt:

Hello handsome. I'm Lucy. Got a name, brown eyes?	Hallo knap. Ik ben Lucy. Ik heb een naam, bruine ogen.

There are several errors in this example. First, there is the semantic error in which the English text means 'handsome' as a noun (meaning 'a handsome person') but the MT has translated it as the adjective 'handsome'. Secondly, the MT did not understand that the last line of the example was a question, and instead translated it as though Lucy was stating she had a name, instead of asking Henry Jekyll for his name. This counts as both a lexical and semantic error.

As a result of the shortness of sentences, previously discussed under adequacy, some lines were too short to correctly portray the intended meaning, causing severe lexical errors. For example, when introducing the scene in excerpt one of the dialogue, Simon Stride utters:

Let's get on with it. Proposition 929, presented by doctor Henry Jekyll. Order! The Chairman will address the board.	Laten we doorgaan. Stelling 929, ingediend door dokter Henry Jekyll. **Bestellen**! De voorzitter spreekt het bestuur toe.

In this excerpt, the segment 'order!' is too short to determine whether this should be translated as 'orde' or 'bestellen'. A human translator, seeing the entire play and the setting of this excerpt, knows the correct translation is 'orde', but the MT does not have this ability.

Similarly, in the third excerpt, Dr. Jekyll waits for a chemist called Bisit, to receive what he needs for the potion:

Bisit. Do you have the chemicals?	**Bezoek**. Heeft u de chemicaliën?

In this excerpt, Bisit is a name, and should remain as in the source text. However, as it is not an existing name, it was not picked up as such by the MT. Bisit was likely picked up as 'visit' as the closest reference within the MT, which resulted in the translation as 'bezoek'.

The output for the dialogue segments is quite good, with some minor awkward phrasing a human editor would need to amend, and there are little errors in syntax or

morphology. Although ideally none of these mistakes would be made, such errors are easier to spot than slight grammatical errors, which makes post-editing easier and less straining.

3.2.3 Song lyrics

Song lyrics contained the highest number of errors, which was mainly caused by the lexical errors. The first song, *Façade*, contained 38 lexical errors alone, which was mainly because the MT continued to mistranslate the same words. This was more prevalent here compared to the other two genres, which can be explained by the higher level of repetition found in song lyrics. While this may look like song lyrics are generally more prone to errors, this would not require so much post-editing effort, as the correct translation needs to be found once and then quickly checked whether it fits in every context, compared to many different errors that require a new translation every time.

The song lyric excerpts also contained a comparatively high number of orthographic errors. This was mainly because the song is sung with an accent, which was written out as an eye dialect for easy reference in reproductions. This meant that the output looked like this:

Nearly everyone you see -	Bijna iedereen die je ziet -
Like him an' her,	Zoals hij een 'haar,
An' you, an' me	Een 'jij, een' ik -

The MT translated the ‘an’ not as ‘and’, which was implied, but as an indefinite article. Every occurrence of this resulted in an orthographic error. This also caused other, mainly lexical, errors to occur:

May 'ave saintly looks -	May 'ave heilige looks -
But they're sinners an' crooks!	Maar ze zijn zondaars en 'boeven!

Here, the MT did not understand that ‘ave’ was meant to mean ‘have’ and, not finding any appropriate translation, did not translate it. Interestingly, it neither translated the word before it nor the last word of the line.

To sum up, most errors in all three genres were on a lexical level, with semantic errors the second highest. In prose, there were also quite a few syntactic errors, which was mostly due to the complexity of the text. Orthographic errors occurred occasionally when the use of accents was represented in text. This is in line with other research on MT output, such as that of Abdulfattah and Yasser (2020).

Even though the adequacy scores for prose were lower, there were less errors in this genre. This was mostly because whenever a small error occurred in song lyrics, this error was then repeated several times. While the errors made in prose were often more severe or complex, they often only resulted in one large error which objectively meant less errors overall.

The use of Farrus et al. instead of the original error typology suggested by Moorkens was, in my opinion, a useful decision. Had the original typology been used, there would have been many errors in the mistranslations category, with quite a few in omissions and word order errors, and little to no additions. While this would have shown a distinction, it would not be as extensive as the current assessment, showing no distinction between what kind or mistranslation was found.

3.3 Post-editing Effort

Post-editing time	**excerpt 1**	**excerpt 2**	**excerpt 3**	**Average**
Prose	19:50	15:30	9:00	14:42
Dialogue	22:40	12:00	15:30	16:43
Song lyrics	2:35:00	2:10:00	59:00	1:54:40

Table 3: Post-editing time

3.3.1 Prose

The prose excerpts were overall easy to post-edit, but for several parts the quality of the output was so low that it was necessary to translate from scratch. For example:

On this night, however, as soon as the cloth was taken away, he took up a candle and went into his business-room.	Op deze avond echter zodra de stof werd weggenomen, hij pakte een kaars op en ging naar binnen zijn zakenkamer.

Not only were certain words translated incorrectly, such as ‘cloth’ and ‘business-room’, but there were also some errors in the underlying structure, which can be seen in incorrect verb-placement of ‘hij pakte’ [he took]. Again, these parts are also likely to have been worsened by the fact that this is an older text, and sometimes uses sentence structures that are not often used anymore. If a newer text was used, these problems might be lessened, or, in the case of older texts that need to be translated, a specialized MT in such language might be able to help as well. Having said that, other areas contained fewer and less severe errors, which only needed some small alterations. For example, the closing lines of the third excerpt:

With every day, and from both sides of my intelligence, the moral and the intellectual, I thus drew steadily nearer to that truth, by whose partial discovery I have been doomed to such a dreadful shipwreck: that man is not truly one, but truly two. I say two, because the state of my own knowledge does not pass beyond that point.	Met elke dag, en van beide kanten van mijn intelligentie, de morele en de intellectuele, kwam ik dus steeds dichter bij die waarheid, door wiens gedeeltelijke ontdekking ik gedoemd ben tot zo'n vreselijke schipbreuk: die man is niet echt één, maar echt twee. Ik zeg er twee, omdat de staat van mijn eigen kennis niet verder gaat dan dat punt.

While this is a slightly longer line, the MT translated it with few mistakes. There are still some stylistic irregularities that a post-editor would still need to closely assess such output on small mistakes, which does take up some time, but it is certainly less time than it would take to translate this line from scratch.

3.3.2 Dialogue

Dialogue was the easiest to post-edit because the high quality of most lines resulted in most cases to only having a change in the sentence structure or a different, more fitting synonym. However, the dialogue did contain certain phrases and abbreviations that the MT did not translate, and do not have a Dutch counterpart, meaning a Dutch audience will not be able to understand them in the same way as an English audience. For example, when Simon Stride introduces the board of examiners:

The right, honourable sir Archibald Proops, QC. And sir Danvers Carew, KBE, Chairman.	De juiste, geachte heer Archibald Proops, QC. En meneer Danvers Carew, KBE, voorzitter.

Sir Archibald Proops is a QC (Queen's Counsel), which means he is a lawyer appointed by the country's monarch. As there is no equivalent in Dutch, and this abbreviation has no meaning in Dutch, it took a while to come up with an appropriate solution. However, since the information is not important to the plot, it was eventually omitted in the post-editing. Similarly, Sir Danvers Carew is described with the abbreviation KBE. This most likely stands for Knight Commander of the Order of the British Empire (K.n, 2020). This order was established in 1917, so it is unlikely that he had this title in the book as well, but since his role in the story has greatly been expanded compared to the book, this was likely added in the musical to give his character more background. A human translator or editor could remove this again, as this knowledge does not add much to a Dutch audience, or decide to leave it in to create more prestige around this character. It would then be advisable that it is written and spoken in full.

3.3.3 Song Lyrics

Unsurprisingly, the post-editing of the song lyrics output took the longest of the three genres. While the post-edited output might not be completely of singable quality, for the sake of this thesis and in light of my inexperience with translating this genre, it is good enough for the

purpose. The long post-editing time was not due to the quality of the output, but rather having to conform to the existing melody and rhyme scheme.

Overall, the post-editing of the song lyrics went easier than I had expected. Especially compared to having to start from scratch, I believe that having a literal translation of the original lyrics can greatly help with starting to edit. This is mainly because some of the lines are in part already on the melody, and only need certain words to be reorganized, altered or omitted. For example, the chorus of *Someone like you*:

But if someone like you	Maar als iemand zoals jij
Found someone like me,	Iemand zoals ik gevonden,
Then suddenly	Dan plotseling
Nothing would ever be the same!	Niets zou ooit hetzelfde zijn!

To correctly fit this on the melody, only very few changes need to be made.This chorus would likely take longer when having to translate from scratch, and in many cases a beginning translator would benefit from having the MT output as a first draft to build on.

In short, in all three stages of this thesis, the dialogue excerpts were of a higher quality than the other two. In all excerpts, lexical errors occurred the most frequently, shortly followed by semantic errors. Furthermore, while the post-editing of the song lyrics took significantly longer than that of prose and dialogue, it is still expected to take less time than translating from scratch, especially for a beginning translator.

4. Discussion

To sum up the first part of this three-part study, the adequacy of the MT output was relatively high for all three genres, with the highest adequacy being in the dialogue excerpts and song lyrics. This is mainly due to the relative simplicity of these genres in terms of grammatical structure. The use of Leech and Short's stylistic categories (2007) proved to be, in my opinion, quite useful, as it was more clear in which area of stylistics the translation was

lacking. In the case of these three genres, this was mainly in lexical style, with the prose excerpts receiving an especially low score in grammatical style as well. With only an overall score for content, this would not have shown in the results. Having said this, the use of the content category in addition to style was now less useful, as this category often acted as an average score of the other four categories.

Additionally, the use of the error typology Farrus et al. (2020), proved to be a good addition to Moorkens' methodology as well, as it provided a larger understanding on the type of mistranslations present in the MT output. Most errors in all three genres were on a lexical level, with semantic errors the second highest. In prose, there were also quite a few syntactic errors, which was mostly due to the complexity of the text. Orthographic errors occurred occasionally when the use of accents was represented in text. This is in line with other research on MT output, such as that of Abdulfattah and Yasser (2020), which showed roughly 44% of all errors as lexical and roughly 30% in syntax as well.

Even though the adequacy scores for prose were lower, there were less errors in this genre, mostly because whenever a small error occurred in song lyrics, this error was then repeated several times. While the errors made in prose were often more severe or complex, they often only resulted in one large error which objectively meant less errors overall. Castilho et al. (2017) on the quality differences between NMT and SMT already made the case that while the errors made by NMT were significantly less severe, they were more difficult to identify, which could result in small errors still being present in the final product. Similarly, I believe that this is also true for these genres. One error being repeated many times throughout a text will very likely stand out to an editor, which will be picked up and amended. However, in a long text with a lot of different errors, it is likely small grammatical errors will be missed.

Finally, while the post-editing effort of the song lyrics was significantly higher than that of the prose and dialogue, it was still expected it would take less effort for a beginning translator than to translate from scratch.

In the introduction to this thesis, I set out three research questions to answer within this thesis. Based on the results found and considering the literature discussed, these can be answered thus:

RQ1: Is the quality of MT output of song texts and dialogue comparable to that of written prose and what are the main difficulties of machine translation in these three areas?

In this thesis, both the song lyrics and dialogue outperformed the prose excerpts. In part this can be attributed to the age of the novel, but even then, it is not expected that a newer version or a newer book would greatly outperform the other two genres. This was an expected outcome, since prose text tends to use longer, more complex sentences than dialogue or song lyrics would. These last genres both contain relatively easy syntax because of their spoken and sung quality, meaning that an audience can only hear it once over, and not re-read the same line as it would in a novel.

Furthermore, the lexis was relatively easier as well for these two genres, which makes less mistranslations or incorrect literal translations. Having said this, the output of these two genres did contain much more orthographic errors, and errors in which words were not translated. This was mainly due to the use of eye dialect in these genres to portray the way the text is supposed to be said or sung.

The dialogue excerpts were very well translated by the MT. The errors made were mostly small, such as cultural references like abbreviations that were untranslated, or certain words it switched around causing wrong syntax. This often happened when the characters were speaking with an accent or making jokes. Furthermore, because each spoken line was separated from the rest of the text, it often switched between translating words such as Sir to

Meneer, and leaving the Sir untranslated in the text, causing a lower average regarding context and cohesion. The song lyrics, while not functional as a singable translation, can be very useful as a working sheet for a translator. While a translator could also translate from scratch, this does make it easier to focus on only the music, instead of having to look up suitable translations for the entire text itself.

Hutchins (1994) surmised already that it would be impossible to reach an ideal Machine Translation that would not be needing any human interference. This is here the case as well. While it was unusable as a piece of creative writing or as a piece of performing arts as raw output, the MT output could potentially serve as an aid for a human post-editor to create a functionable translation in less time and effort than translating without any machine aid.

RQ2: Is the editing effort of MT output on song texts and dialogue comparable to written prose, and is the editing effort of MT output lower than translating from scratch?

The editing effort of MT output of the dialogue excerpts used in this thesis are comparable to that of the prose excerpts. Taking in mind that both genres have different areas in which they are harder to edit (the relative older age of the prose resulting in difficult to understand sentences, and the dialect in which the dialogue was often written resulting in the MT misunderstanding what is said), these difficulties cancel each other out, resulting in almost identical editing times.

Song lyrics, on the other hand, took almost six times longer to edit than the other two genres. It should be mentioned that this is in part since I am most inexperienced in this genre, and that the many different aspects that song translation needs to adhere to compared to the other two genres, mostly caused the delay and not the MT output itself. Because of this, I expect the editing effort of the song lyrics to be lower than translating from scratch. While I was not able to properly test this in this thesis, I assume this due to the head start that the MT

output provides. While a translator does not need to account for fixing errors in the way that a post-editor does, they would need to look up far more words and think about far more synonyms than a post-editor would, as they do not have a rough literal translation to work with. Especially a beginning translator would likely first translate literally, before fitting the text to music, meaning that the time spent on this first draft would add to roughly the same editing time as I have experienced with the MT translation. Having said that, a more experienced translator might find the MT output to be limiting and would prefer to write from scratch, being able to diverge from the source text more easily. Further research may be able to test this in a larger research using both post-editors and translators working on the same piece.

For businesses, there will be another advantage for the use of Machine Translation in this way. One of the students in Moorkens' original research stated that the German-English NMT output was of a quality sufficient to make monolingual post-editing feasible "by someone with no knowledge of German at all, removing the need for a German-speaking post-editor" (2018, p. 16). This could potentially mean that, in order to save time and resources, a company might be able to use not a translator specialised in the language pair, but rather a general linguist for the target language only. For rare language pairs, this will remove the need for translation processes in which the text is first translated into a larger, more generally used language, before being translated into the actual target language, and greatly open the market for target language linguists. Within my research, I do believe this might be feasible, providing the text will be pre-edited by a source language linguist who is specified in Machine Translation, and able to discern most problem areas.

RQ3: Can machine translation function as an aid for (the beginning) translator in musical translation?

Based on my own experiences within this thesis, I do expect that machine translation is able to function as an aid for beginning translators within musical translations. Especially when translating dialogue, I believe that the MT output is of such quality that it would greatly reduce the time spent on translation, since many sentences need little to no editing. While a beginning translator would need to be aware of small errors that might be overlooked, it would be able to function as a first draft of which only certain elements would need to be altered.

Using machine translation for song lyrics would, in my opinion, depend on the translator. While I found it helpful to be able to work with a literal translation I only had to fit to music with different words, structure or sometimes a different idea, others might find this method too limiting for their creative process, such as established by Taivalkoski-Shilov (2018). A beginning translator might find this aid useful, while more experienced song translators would instead find it limiting and would prefer to translate from scratch as they are used to. Nevertheless, machine translation could provide an option for all translators when they find certain sentences or clauses that they are unsure of how to translate, providing a fast literal translation on which they can build further.

While this thesis might function as an example how the translation of musicals might benefit from the use of machine translation, it is not enough to completely prove this viewpoint. First, the sample size is relatively small, only analysing one book and musical, of which three excerpts were taken of the three genres. Secondly, as I am the only evaluator, the results are subjective, and a more reliable result can be achieved with a larger evaluator group. Further research is needed in which this topic is further explored and tested with both a larger testing area as well as a larger evaluator group.

5. Conclusion

In this master thesis, I aimed to provide new insights into how NMT can be used for the translation of creative texts, more specifically the translation of musical song lyrics and dialogue. To do this, I first assessed the MT output of Google Translate from three excerpts of the book *The Strange Case of dr. Jekyll and mr. Hyde* (Stevenson, 1886), as well as three song texts and three dialogue scenes from the Musical *Jekyll & Hyde* (2013). The MT output was assessed using a methodology largely based on Moorkens (2018) method for assessing MT output, which was adapted for a more specific assessment on style as well as content by using the four categories of language (Leech & Short, 2007), and the error typology by Farrús et al. (2012).

The findings suggest that the quality of MT output depends on complexity, rather than length, of the input, in order to portray correct grammar in the target language. Next to this, it is mostly important to establish and, when possible, eliminate problem areas before the texts are put through an NMT by pre-editing. Examples of this found in this thesis are line breaks within song lyrics, as well as abbreviations and marked sentence structures in dialogue and prose texts, or any words with two senses in the target language.

Of the three genres used in this thesis, the dialogue excerpts are especially fitted to be translated using NMT technology, as long as a human pre- and posteditor is used. The other two genres contain more errors and would need more post-editing effort for them to be effective but will still be functionable as a form of drafting. In line with other research, there is significant proof that NMT software can be suitable for creative texts, certainly no any less suitable than other types of text. While the need of a post-editor will always be needed here, the output is suitable enough to create an understandable, relatively clean first copy, which would greatly reduce translation effort compared to translating from scratch, especially for a beginning translator.

The limitations in terms of size and objectivity mean that this thesis should not be considered as conclusive evidence that MT can be a useful aid. In the future, more research is needed using a larger database, as well as a larger testing group. This will not only create a better understanding of how useful MT is in the aid of translating creative texts, but also create a further understanding of which areas of language an MT would still need to improve upon before being able to be used on a large scale within the field of translation. Nevertheless, some guidelines can already be formulated.

First, pre-editing can be especially helpful, mostly for song lyrics in which the lines can be merged to create a more complete grammatical line to avoid incorrect grammar in the output. This would equally be useful for prose, in which complex lines with a lot of embedded clauses can cause problems. Grammatical errors in which the word order was incorrect might be resolved in certain language pairs. In the case of English to Dutch it will be better suited to remove any instances in which the English makes use of a grammatical structure that is not commonly used.

Furthermore, instances of eye dialect are best removed in a pre-editing stage as well. Abbreviations or other words that are known to have no counterpart in the target language should either be written out, omitted, or flagged for post-editing. Consistency errors will likely not be able to be avoided but should be paid special attention to when post-editing. The same can be said for common lexical errors, as well as any words that have two meanings in your target language.

References

Abdulfattah, O., Yasser G. (2020). The Machine Translation of Literature: Implications for Translation Pedagogy. *International Journal of Emerging Technologies in Learning, 15*(11), 228-235.

Boase-Beier (2004). 'Knowing and not knowing: style, intention and the translation of a Holocaust poem', *Language and Literature* 13.1: 25–35.

Brouwer, R. (1983). *De musical – Haar oorsprong, haar definitie en haar ontwikkeling in Nederland.* Instituut voor theaterwetenschap. [Doctoraalscriptie] UvA.

Buhler, F. (2017). The Pitfalls of Musical Translation. *TranscUlturAl, Vol.9*(1), 7-26.

Busscher. (2018). *Een vertaalgeschiedenis van de musical in Nederlands*. [Scriptie]. Universiteit Utrecht.

Castilho, S., Moorkens J., Gaspari F., Calixto I., Tinsley J., and Way A.. (2017a). Is Neural Machine Translation the New State-of-the-Art? *The Prague Bulletin of Mathematical Linguistics 108*: 109-120. Doi: 10.1515/pralin-2017-0013

Castilho, S., Moorkens J., Gaspari F., Sennrich R., Sosoni V., Georgakopoulou P., Lohar P., Way A., Valerio Miceli Barone A., and Gialama M.. (2017b). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In: *Proceedings of MT Summit 2017*.

Constantine, P. (2019). Google Translate Gets Voltaire: Literary Translation and the Age of Artificial Intelligence. *Contemporary French and Francophone Studies*, 23(4), 471-479.

Costa-Jussà, M., Farrús, M., Mariño, J., Fonollosa, J. (2012). Study and comparison of rule-based and statistical Catalan-Spanish machine translation systems. *Comput Inform 31*(2):245–270.

Croonen, E. (2011) *Een Nederlandse musicaltraditie: van Mjoezikul tot Musical* [Scriptie]. Universiteit Utrecht.

Daems, J., & Macken L. (2019). Interactive adaptive SMT versus interactive adaptive NMT: A user experience evaluation. *Machine Translation, 33*(1-2), 117-134.

Daems J., Vandepitte S., Hartsuiker R., and Macken L. (2017). Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort. *Front. Psychol. 8*:1282. doi: 10.3389/fpsyg.2017.01282

Farrús, M., Costa-jussà, M.R., Popovic, M. (2012). Study and Correlation Analysis of Linguistic, Perceptual, and Automatic Machine Translation Evaluations. *Journal of the American Society for Information Science and Technology, 63* (1), 174-184.

Fonteyne, M., Tezcan, A., Macken, L. (2020). Literary machine translation under the magnifying glass : assessing the quality of an NMT-translated detective novel on document level. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, … S. Piperidis (Eds.), *12th International Conference on Language Resources and Evaluation Conference (LREC 2020), Proceedings* (pp. 3783–3791). Paris, France: European Language Resources Association (ELRA). http://hdl.handle.net/1854/LU-8662553

Franzon, J. (2008). Choices in Song Translation: Singability in Print, Subtitles and Sung Performance. *The Translator: Translation and Music*, Vol.14(2), pp.373-399

Genzel, D., Uszkoreit J., Och F. (2010). Poetic" Statistical Machine Translation: Rhyme and Meter. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP '10.* 158–166.

Haque, M. (2012). Translating literary prose: Problems and solutions. *International Journal of English Linguistics*, *2*(6), 97.

Heard, R. (2017). FIT position paper on machine translation. *Babel 63*(1):130–135.

Henisz-Dostert, B., Macdonald, R., & Zarechnak, M. (1979). *Machine translation*. Trends in linguistics. Studies and monographs; 11. The Hague ; New York: Mouton.

Hieble, J. (1958). Should operas, lyric songs, and plays be presented in a foreign language? *Modern Language Journal 42*: 235–7.

Holmes, J. (1988). *Translated! Papers on Literary Translation and Translation Studies.* Amsterdam: Rodopi.

Holmström, J. (1951). *Report on the interlingual scientific and technical dictionaries*. Paris, Unesco.

Idbd, 2020. *Jekyll & Hyde*. Retrieved from: https://www.ibdb.com/broadway-production/jekyll-hyde-4750#Statistics.

Jones, R., Irvine, A. (2013). The (Un)faithful Machine Translator. *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sci-ences, and Humanities*. pp. 96–101.

Hutchins, J. (1994). Machine translation: History and general principles. *The encyclopedia of languages and linguistics, 5,* 2322-2332.

"K, n.". (June, 2020) OED Online. Oxford University Press. Web. 27 August 2020.

Läubli, S. and Orrego-Carmona, D. (2017). When Google Translate is better than Some Human Colleagues, those People are no longer Colleagues. *Proceedings of the 39th Conference Translating and the Computer*, pages 59–69. London, UK. November 16-17, 2017.

Läubli, S., Amrhein, C., Düggelin, P., Gonzalez, B., Zwahlen, A., & Volk, M. (2019). *Post-editing Productivity with Neural Machine Translation: An Empirical Assessment of Speed and Quality in the Banking and Finance Domain.* Retrieved from: https://arxiv.org/abs/1906.01685

Lee, T. (2011). The death of the translator in machine translation: A bilingual poetry project. *Target. International Journal of Translation Studies*, *23*(1), 92-112.

Leech, G., & Short, M. (2007). *Style in fiction: A linguistic introduction to English fictional prose* (No. 13). Pearson Education.

Linders-Nouwens, J. (1999). *Doe Nooit Wat Je Moeder Zegt: Annie M.G. Schmidt, de geschiedenis van haar schrijverschap.* Amsterdam: Querido.

Low, P. (2005). The Pentathlon Principle to Translating Songs. In Gorlée D. L. (Ed.) *Song and Significance : Virtues and Vices of Vocal Translation* (235-262). Amsterdam: Rodopi. Retrieved from EBSCOhost, login.ezproxy.leidenuniv.nl:2443/login?URL=http://search.ebscohost.com.ezproxy.leidenuniv.nl:2048/login.aspx?direct=true&db=nlebk&AN=160120&site=ehost-live.

Low, P. (2008). Translating Songs that Rhyme. *Perspectives: Studies in Translatology, 16,* 1-2, 1-20, DOI: 10.1080/13670050802364437

Luckhurst, R. (2006). *Introduction*. in: Strange case of dr Jekyll and Mr Hyde and Other Tales. Oxford: Oxford University Press.

Moorkens, J. (2018). What to Expect from Neural Machine Translation: A Practical In- class Translation Evaluation Exercise. *The Interpreter and Translator Trainer, 12*(4), 375-387.

Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' Perceptions of Literary Post-editing using Statistical and Neural Machine Translation. Manuscript submitted for publication in: *Translation Spaces, 7*(2), 240-262.

Reiss, K. (1981). Type, Kind and Individuality of Text: Decision Making in Translation. *Poetics Today, 2*(4), 121-131. doi:10.2307/1772491

Russell, Z. (2018). *Translating Broadway: the translation of musical theatre into Spanish.* http://hdl.handle.net/10230/36040

Snyman, F., & Naudé, J. (2003). The assessment of translation accuracy of the Lexica Machine Translation System. *Southern African Linguistics and Applied Language Studies, 21*(4), 295-306, DOI: 10.2989/16073610309486350

Smola, K. (2011). Translator's freedom and the most suitable method in music-linked translation - On the basis of A. Marianowicz's fiddler on the roof Translation. *Anglica Wratislaviensia 49*, 107-114.

StageEntertainment. (2020). *Archief*. Retrieved from: https://www2.stage-entertainment.nl/nl/archief.

Stevenson, R. L. (1886). *The Strange Case of dr Jekyll and Mr Hyde.* In: Strange case of dr Jekyll and Mr Hyde and Other Tales, ed. Luckhurst, R (2006). Oxford: Oxford University Press.

Stewart, Z. (2013, April 18). *Jekyll & Hyde*. Retrieved from: https://www.theatermania.com

Storytellers. (2020). *Jekyll & Hyde.* Retrieved from: http://storytellers.nl/producties-bezoeken/jekyll-hyde/

Taivalkoski-Shilov, K. (2018). Ethical Issues Regarding Machine(-assisted) Translation of Literary Texts. Manuscript submitted for publication in: *Perspectives*, 1-15.

Tezcan, A., Hoste, V., & Macken, L. (2017). "Chapter 10: scate Taxonomy and Corpus of Machine Translation Errors". In *Trends in E-Tools and Resources for Translators and Interpreters*. Leiden, The Netherlands: Brill | Rodopi. doi: https://doi.org/10.1163/9789004351790_012

Toral, A., Way, A. (2018). What Level of Quality can Neural Machine Translation Attain on Literary Text? In: Moorkens, J., Castilho, S., Gaspari, F. & Doherty, S. (Eds.), *Translation Quality Assessment* (1-26). New York, NY: Springer Publishing.

Ruth Torres Domınguez. 2012. *The 2012 use of translation technologies survey.* Retrieved from: http://mozgorilla.com/download/19/.

Ueffing, Nicola, & Ney, Hermann. (2007). Word-Level Confidence Estimation for Machine Translation. *Computational Linguistics, 33*(1), 9-40.

Vanmassenhove E., Du J., Way A. (2016). Improving subject-verb agreement in SMT. In: *Proceedings of the fifth workshop on hybrid approaches to translation*. Riga, Latvia.

Walking The Way To Dawn #SoRiku (2016, May 10). Jekyll and Hyde the Musical Act 1 ((Constantine Maroulis)) [video file]. Retrieved from: https://www.youtube.com/watch?v=nOccf22dUi0&t=2161s

Walking The Way To Dawn #SoRiku (2016, May 10). Jekyll and Hyde the Musical Act 2 ((Constantine Maroulis)) [video file]. Retrieved from: https://www.youtube.com/watch?v=J0bBvXGErIk&list=PL7O-TEZeYJ04zRyoVdA-edEktbP946OOk&index=2

Weaver, W. (1955). Translation. *Machine translation of languages*, *14.*

Appendix A: Prose

Excerpt 1	English	MT output
Semantic Errors (10)	spirits (2) rang out (3) soberly (3) cloth (4) a clouded brow (5) took charge (6) the said (6) beyond (6) his knowledge (9) obnoxious (12)	Geesten belde nuchter stof vertroebeld voorhoofd leiding zei dan zijn kennis onaangenaam
Lexical Errors (14)	SEARCH (1) bachelor house (2) dinner without relish (2) his custom of a Sunday (3) the hour of twelve (3) business-room (4) - (5) envelope (5) Will (5) holograph (6) of which he could learn no more (10) to be clothed upon (11) - (12)	ZOEK vrijgezel huis zonder diner gewoonte van een zondag het uur van twaalf zakenkamer is envelop Will holografisch van die hij niet meer kon leren bekleed (terugplaatste)
Syntactic Errors (10)	- (1) on his (3) would go (3) he took up (4) went into his (4) should (6) the (8) the immodest (8) it began to be (11) it is (12)	(De) de zijne zou gaan hij pakte ging naar binnen zijn moet (het) de onfatsoenlijke zijn is
Orthographic Errors (5)	Mr. Utterson (2) private part (5) Dr. (5) Mr. (6) Mr. (9)	meneer Utterson privé-gedeelte Dr Mr meneer
Morphological Errors (3)	the (5) had been (7)	de het had

	sides (8)	kanten

Excerpt 2	English	Output
Semantic Errors (3)	MURDER CASE (1) men (7) speech (9)	MOORDKOFFER mannen spraak
Lexical Errors (13)	given (6) a dream of musing (6) tears (7) - (7) - (7) felt (7) pretty manner (9) indeed (10) conceived (11) trifling (12) ill-contained (12) one (14) out of all bounds (14)	gegeven een droom van mijmeren tranen over nooit gevoeld mooie inderdaad bedacht spotten ongeduldig man alle grenzen uit
Syntactic Errors (7)	October, 18 — (2) rendered (2) she used to say (7) another and very small gentleman (8) to whom at first she paid less attention. (8) to (14) under (15)	18 oktober werd het ze zei een andere en heel kleine heer ze eerst besteedde minder aandacht op te
Orthographic Errors (2)	Mr. (11) Mr. (14)	meneer meneer
Morphological Errors (3)	pointing (10) other (11) trifle hurt (14)	aanwijzing andere gekwetst

Excerpt 3	English	Output
Semantic Errors (5)	parts (2) the wise (2) good (2) trench (6) the mystic (9)	delen wijzen goeden greppel de mysticus

Lexical Errors (9)	besides (2) - (2) such as I (3) than commonly grave (3) as I was guilty of (5) dealer (8) in the eye of day (8) among (9) members (9)	en de net meer dan gewoonlijk graf als ik schuldig was dealer het oog van de dag onder leden
Syntactic errors (0)		
Orthographic (2)	and, (6) I (8)	en, Ik
Morphological (2)	found (3) the (6)	vond het

Appendix B: Song Lyrics

Excerpt 1	English	Output
Semantic Errors (3)	find (76) mean (79) mean (80)	vinden bedoelen bedoelen
Lexical Errors (34)	facade (6) facade (12) Game (18) Livin’ (19) Actin’ a (20) While (21) found (29) - (30) is (37) an’ (43) an’ (44) an’ (44) pillar (46) An’ (49) Notoriety (51) an’ (52) may ‘ave (54)	gevel gevel Game Livin Actin 'a Tijdens het gevonden het is een Een een pijler Een Berucht een May 'ave

	looks (54) facade (63) fine (66) facade (69) Might look kinda well-to-do (71) Right down to their boots (73) mankind (75) they’re all in cahoots (78) clean (82) facade (84) facade (85) fine (88) cross (89) discard (91) facade (93) facade (94) facade (95)	looks gevel fijne gevel Ziet er misschien goed uit om te doen - Recht naar beneden zijn ze laarzen! mensheid ze zitten allemaal in de maling schoon gevel gevel fijne oversteken weggooien gevel gevel gevel
Syntactic errors (5)	lurking beneath (31) That (67) Is it’s (82) Is it’s (83) all (89)	ligt Dat Is het Is het allemaal
Orthographic (10)	like (19) ‘Cause (35) she (43) you (44) me (44) an’ piety (49) ‘gents (52) an’ crooks (55) ‘cause (63) an’ he (88)	'als 'Want 'haar 'jij, ik - 'vroomheid 'heren’ 'boeven 'Omdat 'hij
Morphological (3)	add (15) be (16) break through (69)	toevoegen wees breken

Excerpt 2	English	Output
Semantic Errors (7)	known (14) you (38) your (38) you (44) your (44)	geweten u uw u uw

	may (78) Dear (104)	misschien Beste
Lexical Errors (21)	been (6) before - (7) on (9) - (13) - (14) overall (27) don't (31) May (41) may (44) way (60) - (60) - (60) to be (64) to be (72) the (75) all (75) haven’t (82) in over his head (83) - (98) - (98) May (106)	geweest eerder (op) de het algemeen doet het May Misschien manier waarop het (om te) (om te) (Het) (enkel) niets boven zijn hoofd zit (die) het May
Syntactic errors (5)	- (9) find (41) find (44) you know (58) I’ve (82)	hebt vind vindt weet je Ik heb
Orthographic (0)		
Morphological (9)	has (14) you (15) to live for (29) find (41) forgiven (64) forgiven (72) may find (78) pursuing (89) find (106)	hebben jij moet vind vergeven vergeven vindt nastreven Vind

Excerpt 3	English	Output
Semantic Errors (2)	someone there (21) set (25)	daar iemand laten
Lexical Errors (2)	in (7)	(in)

	- (19)	ernaar
Syntactic Errors (3)	found me (17) have (23) found me (34)	vond me heb vond me
Orthographic Errors (0)		
Morphological Errors (5)	keeping (6) found (11) have needed (20) found (28) found (36)	houden gevonden nodig heb gevonden gevonden

Appendix C: Dialogue

Excerpt 1	English	Output
Semantic Errors (11)	just (6) Order (11) amount (14) happily (15) those (19) those (20) your (25) take to (26) mad (31) powerful (35) You (36)	juiste bestellen bedrag gelukkig die die je meenemen gek krachtige U
Lexical Errors (15)	- (1) Grace (2) Lord (3) honourable (6) Sir (6) GC (6) KBE (7) order of business (8) accord (15) accord (16) Aye (19) Sir (22) Abstained (23) to none (24) abstention (24)	de Grace Lord geachte heer GC KBE agenda stemmen stem Aye Sir Onthield voor onthouding
Syntactic Errors (2)	with (24) babbling (27)	bij babbelen

Orthographic Errors (1)	saint jude's hospital (13)	Sint-Judeziekenhuis
Morphological Errors (5)	His (2) address (12) - (23) - (26) have had (27)	Zijne spreekt zich de had

Excerpt 2	English	Output
Semantic Errors (4)	handsome (6) nice (8) alright (21) soliciting (26)	knap aardige oké vragen
Lexical Errors (12)	be different (2) let's see (3) my dear (5) Got a name? (6) wasn't (11) rathole (13) Don't I know it (17) street (25) I'll just work the doctor all the day (25) nobbing (29) They'll cost you (30) Lucky me (33)	anders zijn Laten we eens kijken liefste Ik heb een naam geen beestje Ik weet het niet straat ik werk de hele dag bij de dokter nobbing ze kosten je Lucky me
Syntactic Errors (2)	you (16) I'm (32)	u ik heb
Orthographic Errors (0)		
Morphological Errors (0)		

Excerpt 3	English	Output
Semantic Errors (4)	remember (1) principle (1) agree (19) actually (28)	gedenken principe akkoord eigenlijk
Lexical Errors (10)	straight (3)	recht

	good (4) damn- damn- (6) Bisit (9) fail (12) they (12) what there (13) - (17) be off with you (22) incolence (22)	goede damn- damn Bezoek falen die daar je ga met je mee incolentie
Syntactic Errors (4)	they’re (13) go (17) hope (20) probably (26)	zijn gaan geloof waarschijnlijk
Orthographic Errors (0)		
Morphological Errors (6)	you (9) go (17) is (20) should wonder (22) you (23) your (23)	U gaan is afvragen je jouw