MASTERARBEIT | MASTER'S THESIS

Titel | Title

The Quality and Reception of Machine Translated Subtitles

verfasst von | submitted by

Sophie Kalny BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 070 331 342
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Translation Deutsch Englisch
Betreut von | Supervisor:	Univ.-Prof. Dragos Ioan Ciobanu PhD

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Univ.-Prof. Dragos Ioan Ciobanu, PhD, for his invaluable guidance and insightful feedback throughout the process of writing this master's thesis.

I am deeply thankful to Raluca-Maria Chereji, BA MSc for her assistance and expertise during the execution of my eye-tracking experiment. I am also grateful to all the participants who generously dedicated their time and efforts to take part in the experiment. Their participation was essential in gathering the data necessary for this study.

Lastly, I would like to mention my family and friends who supported me morally throughout the writing process and provided helpful feedback.

Abstract (English)

This paper examined the quality and user experience of machine translated subtitles of a TED Talk in the domain of environmental protection in the language combination English-German. Given that machine translation has recently gained importance in this field and the users' reception is still rarely taken account of in similar studies, this thesis aimed at contributing to bridging this gap.

The MQM framework was used to analyse the quality of the machine translated subtitles. To study the user experience, I measured the cognitive load with eye-tracking parameters while the participants were watching the subtitled videos. Additionally, their attitude towards the subtitles was measured with a post-task questionnaire. To see how well the subtitles were understood by the participants, comprehension questions were used as well. I expected the eye-tracking parameters to show a higher cognitive load when the quality was worse. In this case, I also expected fewer correct answers in the comprehension questionnaire. A worse rating in the attitude questions was expected when the cognitive load was higher as the reading process must have been more tedious.

In this study, most of these hypotheses were confirmed. However, some were only partially confirmed or the differences between the different kinds of subtitles were not very large. The results lead to the conclusion that viewers could deal with bad quality subtitles as comprehension was not impaired to a great extent. Although they might not catch every detail, the participants were still able to get the gist. However, there was a great difference in viewing experience between human translated and machine translated subtitles. It was more exhausting to watch machine translated subtitles, which were of worse quality. Therefore, I concluded that it depends on the situation whether general-purpose machine translation models are an appropriate choice for the translation of subtitles. Conducting this research with fine-tuned MT models may lead to different results.

Abstract (German)

In dieser Masterarbeit wurden die Qualität und die Nutzererfahrung von maschinell übersetzten Untertiteln eines TED-Talks zum Thema Umweltschutz in der Sprachkombination Englisch-Deutsch untersucht. Da maschinelles Übersetzen im Bereich der Untertitelung in den letzten Jahren an Bedeutung gewonnen hat und die Erfahrungen der Nutzer*innen in ähnlichen Studien noch selten berücksichtigt werden, hatte diese Arbeit zum Ziel, diese Forschungslücke zu beleuchten.

Zur Analyse der Qualität der maschinell übersetzten Untertitel wurde die Multidimensionale Qualitätsmetrik (MQM) verwendet. Um die Nutzererfahrung zu analysieren, wurde die kognitive Belastung mittels Eye-Tracking untersucht, während die Teilnehmenden die untertitelten Videos anschauten. Mittels Einstellungsfragen wurde auch die Meinung der Teilnehmenden zu den Untertiteln ermittelt. Zusätzlich wurden Verständnisfragen eingesetzt, um zu prüfen, wie gut die Teilnehmenden den Inhalt verstanden haben. Im Vorfeld wurde davon ausgegangen, dass die Eye-Tracking-Parameter auf eine höhere kognitive Belastung schließen lassen, wenn die Qualität schlechter ist. In diesem Fall wurden auch weniger richtige Antworten auf die Verständnisfragen erwartet, was ebenfalls ein Indikator für eine höhere kognitive Belastung ist. Des Weiteren wurde bei höherer kognitiver Belastung angenommen, dass die Teilnehmenden die Untertitel bei den Einstellungsfragen schlechter bewerten würden, da der Leseprozess mühsamer gewesen sein muss.

In dieser Arbeit wurden die meisten dieser Hypothesen bestätigt. Einige wurden jedoch nur teilweise bestätigt oder die Unterschiede zwischen den verschiedenen Arten von Untertiteln waren nicht besonders groß. Die Ergebnisse lassen darauf schließen, dass die Zuschauer*innen mit Untertiteln von schlechterer Qualität umgehen können, da das Verständnis nicht in hohem Maße beeinträchtigt wird. Auch wenn nicht jedes Detail verstanden wurde, konnten die Teilnehmenden dennoch das Wesentliche erfassen. Allerdings gab es einen deutlichen Unterschied in der Nutzererfahrung zwischen dem Lesen von menschlich und maschinell übersetzten Untertiteln. Es war anstrengender, maschinell übersetzte Untertitel zu lesen, die von schlechterer Qualität waren. Daher wurde das Fazit gezogen, dass es von der jeweiligen Situation abhängt, ob maschinelle Übersetzungssysteme für die Übersetzung von Untertiteln geeignet sind. Die Durchführung dieser Untersuchung mit feingetunten maschinellen Übersetzungsmodellen könnte jedoch zu anderen Ergebnissen führen.

Contents

List of Figures

List of Tables

Introduction

Over the past years, the volume of audiovisual content has surged, especially online due to the increased popularity of video and streaming platforms as well as the use of videos on social media platforms. Given that this type of content can be accessed from nearly all over the world, the need to make it available to people speaking different languages has become more and more prominent. However, it would be very expensive and time-consuming to professionally translate such large amounts into so many different languages, which is why machine translation seems to be an ideal solution to this issue as it is a very fast and cheap form of translation. Nevertheless, other aspects of machine translation, such as quality issues, might impair its usability.

The quality of machine translation output is oftentimes not as good as a translation made by a human professional. Nevertheless, it should be acknowledged that there have been studies where MT was on par with professional human translations. One example would be a study by Fischer and Läubli (2020) where a domain specific MT system performed better than human translation in one out of three language combinations.

Moreover, the translation of audiovisual content comes with its own challenges that might be even more challenging for machine translation systems than the translation of standard text. Therefore, the aim of this thesis is to find out more about how suitable general-purpose machine translation systems are for this type of translation; more precisely, how well machine translation systems can translate subtitles from English into German and how well those subtitles are processed by the viewers.

The quality of machine translation can be assessed by automatic or human evaluation methods. Both evaluation methods have their own advantages and disadvantages, which means that the right choice of method highly depends on the purpose of the evaluation (see chapter 3). In the experiment of the present paper, the goal was to obtain a detailed evaluation of the quality, which can be achieved by using human evaluation methods. In this case, the MQM framework was chosen for quality analysis because it is a well-established error typology that has been widely used and refined over time. It also allows customisation, making it suitable for the requirements of different kinds of projects. For example, it is possible to choose how fine grained the analysis should be or to add customised error types.

While there is a great deal of literature on the quality of machine translated subtitles, the users' opinion on and their cognitive processing of this type of subtitles has only been studied by a small number of researchers. Due to this lack of research on this aspect of quality, I deemed

it important to incorporate it into the present paper. Two methods were chosen to acquire information on the user experience of machine translated subtitles. On the one hand, the user experience was examined by using eye-tracking because this technology allows to indirectly gain insights into the cognitive load. On the other hand, a questionnaire was used to gain insights into the consciously perceived attitude of the viewers towards the machine translated subtitles. The questionnaire also contained comprehension questions to compare the results against the results of the eye-tracker as good/bad comprehension can also be a result of low/high cognitive load.

The combination of these methods was used to answer the following research questions:

1. How good is the quality of the MT systems Google Translate and DeepL in the translation of the subtitles of a TED Talk in the domain of environmental protection from English into German and what kinds of mistakes do these systems make?
2. What effect does the overall quality of the machine translated subtitles have on the cognitive load and the attitude towards the subtitles?

The first part of this thesis consists of a literature review summarising the necessary background information and current state of research. This theoretical part is divided into four chapters.

Chapter 1 gives an overview of what exactly subtitles are, their most important characteristics related to space and timing, and some translation strategies that can be used to respect their characteristics.

Chapter 2 revolves around machine translation. The focus lies on machine translation in the context of subtitling, which is why the chapter starts with the challenges for machine translation in subtitling. In order to take into consideration the current state of research, the second section of chapter 2 deals with recent technological developments in this field.

Chapter 3 deals with the quality evaluation of machine translation output. The first section discusses the reasons for the necessity of evaluating the quality of machine translation output. The following sections describe different methods as well as when and why they are used.

The last chapter of the theoretical part deals with cognitive load and eye-tracking. The first two sections revolve around the theoretical construct of cognitive load and how it can be measured. The last section deals with eye-tracking, which is one way to measure cognitive load. It shows what exactly eye-tracking is and how it has been implemented in other similar studies.

The practical part of this thesis consists of 4 chapters, starting with chapter 5, the Methodology. This chapter, will delve into the methods employed, outline the hypotheses,

describe the data collection process and elaborate on the data analysis. In chapter 6, the results will be described in detail, with each section treating the results of one part of the experiment. Chapter 7 is the discussion and will interpret the results. It will also try to find possible explanations for unexpected results and establish relations to previous papers. Finally, it will shortly state the limitations of the present paper and make some recommendations for future work before the conclusion will sum up the most important findings of this master's thesis.

1 Subtitling

Before dealing with MT in the field of subtitling, this chapter goes into subtitling as such and gives an overview of the most important aspects of subtitling. The first part defines the term subtitling and summarises the different types of subtitling. The second part revolves around subtitling guidelines and temporal and spatial characteristics inherent to subtitles. The final section will present some translation strategies that can be used to respect subtitles' characteristics.

1.1 Definitions

Subtitling is a type of audiovisual translation (AVT) alongside many others such as dubbing or voice-over. Although AVT has existed since the invention of the cinema around 1900, it has not received considerable attention in translation studies until the mid-1990s. The increased scholarly attention to AVT coincides with the beginning of digitisation and the increasing production of audiovisual material (Díaz-Cintas & Remael, 2021, p. 1). According to Díaz-Cintas and Remael (2021, p. 3), the lack of academic interest in AVT might have resulted from the opinion of many academics that AVT was not a form of translation proper due to the restrictions that are involved in this form of translation. Audiovisual material consists of two semiotic layers, namely image and sound, that make up its meaning. Therefore, the interplay of these two communication channels must be taken into consideration when translating, which means that the translator must deal with spatial and temporal constraints resulting from the multimodal nature of audiovisual material (Diaz-Cintas, 2020, pp. 149–150).

The following two definitions include the main features of subtitling as a concept and practice. Diaz-Cintas (2020, p. 150) defines subtitling as

> a translation practice that consists of rendering in writing, usually at the bottom of the screen, the translation into a target language of the original dialogue exchanges uttered by different speakers, as well as all other verbal information that appears written on-screen (letters, banners, inserts) or is transmitted aurally in the soundtrack (song lyrics, voices off).

This means that an additional element, the subtitles, is added to the spoken language and the image.

The second definition by Chaume contains another important aspect of subtitling that is not mentioned in the first one, namely synchrony with the actors' utterances. Chaume (2013, p. 112) defines subtitling as "incorporating a written text (subtitle) in the target language on the screen where an original version film is shown, such that the subtitles coincide approximately with the screen actors' dialogues".

Subtitles can be categorised in many different ways. Díaz-Cintas and Remael (2021, p. 11) note that it is difficult to create a fixed classification of subtitles due to the constantly changing mediascape and technologies involved. However, they still describe some of the existing categories as they are useful to get an overview of subtitles. I will shortly describe the categorisation according to linguistic parameters to further define the term *subtitle* like it will be used in this paper. Firstly, subtitles can be divided into intralingual and interlingual subtitles. In intralingual subtitles, also called *captions*, the spoken dialogue is in the same language as the subtitles, whereas interlingual subtitling means that the subtitles are translated into another language. Interlingual subtitles can be further divided into monolingual subtitles, which consist of one target language, and bilingual subtitles, which means that two different languages are displayed at the same time (Díaz-Cintas & Remael, 2021, pp. 11–19). This paper revolves around *monolingual interlingual subtitles*.

1.2 Characteristics

As already briefly outlined in the previous section, audiovisual translations have certain characteristics that are related to their multimodal nature. This section will give a short overview of subtitling guidelines and describe the most important characteristics of subtitles and their translation, as well as ways in which translators can deal with them.

1.2.1 Guidelines

In general, there is a lack of standardisation when it comes to how subtitles are presented on screen and a variety of practices regarding parameters like line length and duration, display rate and line breaks, have developed over time. This circumstance is often regarded as a lack of quality, which is why there have been several attempts to establish general subtitling guidelines. However, up to now, there is no universally valid set of rules. Nevertheless, there is a certain framework within which most guidelines are situated (Díaz-Cintas & Remael, 2021, pp. 91–92).

Many companies that offer subtitling services have their own style guides for the creation of subtitles. Traditionally, they have only been available to the companies' employees.

Netflix, however, was one of the first companies that made its style guide available to the general public by publishing it on its website (Díaz-Cintas & Remael, 2021, p. 47). In order to illustrate the different guidelines of different companies and organisations, a few points of the guidelines published by Netflix (2022) and TED (n. d., Subtitling Tips) will be compared briefly. The first one is reading speed. With a maximum reading speed of 21 characters per second (cps), TED's guidelines allow for a higher reading speed than Netflix, whose guidelines set a limit of 17 cps for adult programmes. A possible explanation for the higher reading speed limit on TED might be the fact that it is a platform for the dissemination of innovative ideas, which means that the focus is on the content and not on the enjoyment. Netflix, however, is a streaming platform that people usually use for entertainment purposes, which might be the reason why the focus is rather on enabling the viewers to read the subtitles more comfortably and give them enough time to also follow the visual component more easily. Moreover, Netflix offers a great variety of film and show genres and therefore addresses a very heterogenous target group, while TED's educational content rather targets people with a higher level of literacy, who are more likely to be able to cope with the higher reading speed. As far as line breaks are concerned, both accept a maximum of two lines per subtitle. However, their guidelines differ in the expectations regarding the shape of the subtitle blocks. According to Netflix, the translator should aim for a shorter upper line and a longer lower line. TED's guidelines instruct the translator to make sure that one line is not more than 50% shorter than the other line, but it does not matter which one is the longer one. One point that both guidelines have in common is that they allow a maximum of 42 characters per line.

This comparison shows that there is not just one single way of how subtitles are presented perfectly. Instead, there is a certain scope, within which most companies and organisations choose rules that fit their purpose best. As will be discussed in section 2.2, some of these guidelines are easier to implement in MT systems than others. The next two sections will deal with characteristics of subtitles that are related to space and time in more depth and review recent literature on this topic.

1.2.2 Characteristics Related to Space

Díaz-Cintas and Remael (2021) describe the most important aspects of subtitling that are related to the limited space. As already mentioned in the previous section, they acknowledge that there are no fixed rules but a "variation within generally accepted practice" (Díaz-Cintas & Remael, 2021, p. 93). In general, subtitles should not attract unnecessary attention, which is why their configuration is rather simple and unobtrusive. In terms of layout, this means that subtitles are

usually located at the bottom of the screen and centre justified as this part of the screen is less important for the action in most cases. The font is typically a neutral one without serifs (Díaz-Cintas & Remael, 2021, pp. 93–96).

Line length and line number seem to be a lively research topic at the moment. Certain standards that have developed over the years are usually applied in the industry, but for the past few years they have been increasingly questioned by researchers (e.g. (Szarkowska & Gerber-Morón, 2019; Zahedi & Khoshsaligheh, 2021).

Traditionally, a maximum of 37 characters per line (cpl) was allowed as monospaced fonts were used, i.e. fonts where every character takes up the same amount of space. Nowadays, with the use of proportionally spaced fonts, a limit on the cpl is not so relevant anymore as the individual letters take up different amounts of space, meaning that lines of the same length can accommodate different numbers of characters. Nevertheless, many companies still set limits on the maximum number of cpl (Díaz-Cintas & Remael, 2021, pp. 97–98). This can be seen, for example, in the guidelines of Netflix and TED (see section 1.2.1) where both set the limit to 42 cpl.

The number of lines is generally limited to a maximum of two lines per subtitle; however, one-liners are often preferred. In the industry, there is a variety of practices regarding line length and how lines should be broken up. In certain settings, it is even common to use more than two lines, such as on the Internet, in bilingual subtitles or in intralingual subtitles for the deaf and hard of hearing (SDH) (Díaz-Cintas & Remael, 2021, pp. 99–100).

Szarkowska and Gerber-Morón (2019) observed a frequent use of three-liners in British SDH as well, leading them to the question whether three lines could be used in interlingual subtitling too. They conducted an experiment on the cognitive load on viewers when watching subtitles made up of two and three lines respectively, as well as on their preferences regarding the number of lines. They asked their participants to watch two short videos, one with subtitles made up of two lines and one with subtitles made up of three lines. The subtitles contained exactly the same words; they were just split up differently as can be seen in Figure 1. They recorded the participants' eye movements in an eye-tracking lab, conducted interviews about the viewing experience and used a questionnaire for a self-assessment of the cognitive load and enjoyment, as well as for a comprehension test. The results of this study suggest that the cognitive load is higher in the case of three-liners. The difference is not very big, but the videos used in the study were very short, and the authors note that the difference might be bigger in longer videos. Comprehension is not negatively affected by the higher number of lines; however, the participants clearly prefer two-liners.

No.	Two-line subtitles	Three-line subtitles
1	I've invited my doctor of ten years here	I've invited my doctor of ten years here and instructed him to bring all of my medical records.
2	and instructed him to bring all of my medical records.	Hopefully, by exposing my very own medical history, I will inspire our future leaders to do the same.
3	Hopefully, by exposing my very own medical history,	It's the old: "I'll show you mine if you show me yours."
4	I will inspire our future leaders to do the same.	
5	It's the old "I'll show you mine if you show me yours."	

Figure 1: How subtitles were split into 2 and 3 lines respectively (Szarkowska & Gerber-Morón 2019, p. 148)

In my opinion, this study supports current practices in the interlingual subtitling industry as three lines reduce the enjoyment and increase the effort the viewer must put into the programme. However, I agree with the authors (Szarkowska & Gerber-Morón, 2019, p. 145) that it might make sense in information-packed programmes where the enjoyment is less important than the information, such as the newscast.

Whenever a subtitle consists of two lines, the translator must decide where to break it. According to Díaz-Cintas and Remael (2021, p. 100), both the shape and the syntax are usually considered. However, they say that it is more important to respect the syntax than the shape. This is in line with findings by Gerber-Morón and Szarkowska (2018) who conducted an eye-tracking experiment testing viewers' preferences on segmentation and the readability of subtitles. They showed the participants 30 screenshots of subtitle pairs where each pair contained the same subtitle twice, once syntactically segmented and once non-syntactically segmented. In syntactically segmented subtitles, linguistic units, like article and noun or auxiliary and lexical verb, are not separated. The participants had to select the version that they preferred and answer questions about the factors leading them to their decision; in other words, whether the shape or the syntax was more important in their decision-making process. The gathered data showed that the participants preferred the syntactically segmented subtitles although the percentage varied depending on the linguistic units, mother tongue and hearing status. Moreover, the syntax mattered more than the shape for the majority of the participants when making a decision on the preferred subtitle. This preference might have also had an effect

on the results of the above experiment where the participants preferred the two liners as they were segmented syntactically as opposed to the three liners.

In a similar experiment with different groups of participants based on their hearing status or mother tongue, Gerber-Morón et al. (2018) studied how syntactically segmented subtitles influenced the viewers' cognitive load and comprehension. Their results suggest that syntactically segmented subtitles indeed reduce the cognitive load. They found no evidence that non-syntactically segmented subtitles had a negative influence on comprehension. They also realised that Spanish participants, who are more used to dubbing, had the highest cognitive load, the lowest comprehension scores and spent more time reading the subtitles than participants of the other groups. Based on their results, they insist on sticking to existing segmentation guidelines in order to provide users with easily readable subtitles (Gerber-Morón et al., 2018, pp. 9–14).

I agree with the authors that in light of their results, it would be best to stick to existing guidelines that allocate more relevance to syntactic line breaks than to the shape of subtitles. However, the results also highlight how important it is to take into account the needs of the target groups and how much they have been exposed to the different types of AVT when stipulating guidelines.

1.2.3 Characteristics Related to Timing

Temporal synchronisation is an important characteristic of subtitles and means that the content of a subtitle should match what is being said in the audio. It is very important that subtitles are accurately timed in order to enhance semiotic cohesion, i.e. the relation between the different semiotic channels, in the end product. Moreover, viewers can understand more easily who is saying what if subtitles appear and disappear at the right time. The process of determining the exact in and out times of subtitles is called spotting. Sometimes perfect synchronisation is not possible, especially in fast-paced dialogues, which is why a certain degree of flexibility is allowed (Díaz-Cintas & Remael, 2021, pp. 101–102).

There are also conventions about how long a subtitle can stay on the screen. Usually, a subtitle that consists of two full lines is allowed to stay on the screen for a maximum of six seconds. If a subtitle is visible for a longer period, viewers might read it twice, which should be avoided. There is also a minimum exposure time of one second for subtitles to make sure that viewers have enough time to notice the new subtitle and to read it. The subtitle speed can be measured in characters per second (cps) or words per minute (wpm), however, cps is more

accurate as the length of words can vary considerably (Díaz-Cintas & Remael, 2021, pp. 105–108).

Díaz-Cintas and Remael (2021, pp. 109–110) remark that there is a constant evolution in subtitling standards and conventions. With new media service providers like Netflix, the maximum number of cps has increased, and nowadays 17 cps or even more are frequently used as opposed to the traditional 12 cps in TV or 15 cps on DVDs. The fact that the maximum number of cps has increased, might be beneficial for MT as it sometimes has difficulties to adhere to very low subtitle speeds. Subtitle speed seems to be a lively research topic at the moment, which is why I will present two studies with interesting results on the effects of higher subtitle speed in the following.

The first study was conducted by Szarkowska and Gerber-Morón (2018) who wanted to find out whether viewers could deal with higher subtitle speed. In their experiment, they asked different groups of participants to watch three videos where each one had a different subtitle speed (12, 16, and 20 cps). They used eye-tracking, questionnaires and interviews to assess the effect of subtitle speed on different factors, such as comprehension, enjoyment, and scene and subtitle recognition.

The results of the comprehension test, scene recognition and self-reports on the reading experience showed that the participants had enough time to read and understand the subtitles of all three speeds. The eye-tracking results showed a higher mean fixation duration for the slowest subtitles, which indicates a higher cognitive load. Possible reasons might be the higher degree of text reduction in the slowest subtitles, rendering them less internally cohesive, as well as the fact that there were more discrepancies between the spoken dialogues and the subtitle text. Another finding was that participants revisited the subtitle area more often in the video with slow subtitle speed. When the number of revisits was higher, the self-reported enjoyment was lower and the difficulty higher. The authors suggest that the reading process might be interrupted if the subtitle is on the screen for too long because viewers look at the subtitle area expecting a new subtitle but only find the one that they have already read. Moreover, the eye-tracking data showed that the participants did not spend more time reading the slower subtitles although they were on the screen for longer. The authors interpret this finding as an indication that faster subtitles are read more efficiently (Szarkowska & Gerber-Morón, 2018, pp. 24–25).

Kruger et al. (2022, p. 215) criticise the authors for their interpretation of these eye-tracking data. They think that proportional reading time is not a useful measure in the context of subtitle speed. In their opinion the participants spent proportionally more time reading the

faster subtitles as more text had to be read in less time, which means that there was proportionally less time for each character.

Kruger et al. (2022, p. 221) conducted an experiment on the effects of higher subtitle speed on comprehension too, but they used a different set of eye-tracking measures to see differences on the word level, as well as a comprehension questionnaire. The participants were asked to watch three videos with different subtitle speeds (12, 20 and 28 cps). For each subtitle speed, the authors determined to what extent words were skipped and reread, both of which are frequently used measurements in studies on static reading.

The results show a significant difference in comprehension accuracy between 12 cps and 28 cps, which indicates that a higher number of characters per second hampers comprehension. A higher subtitle speed also leads to more words being skipped, especially towards the end of a subtitle. This means that fewer words are processed, and fewer subtitles can be processed completely when the subtitle speed is higher, leading to potentially important information being missed. As far as revisits are concerned, the results show that the higher the subtitle speed is, the fewer revisits occur. This was observed in both horizonal eye movements, i.e. going to back to another word in the subtitle, and vertical eye movement, i.e. looking at the image. According to the authors, this means that the participants are not able to properly process the fastest subtitles as revisits are important to catch up on missed information (Kruger et al., 2022, pp. 227–228).

Subtitle speed is a very interesting research topic where much more research is still needed as can be seen from the very different results of the two studies, one of which suggesting that higher subtitle speed is reasonable while the other one is rather critical of that. However, the research design of these two studies varies a great deal, for example regarding the maximum number of cps used, which is why the results are difficult to compare. In my opinion, more scientific findings on how fast subtitles should be displayed at most would also be very interesting in the light of machine translated subtitles as machine translation (MT) cannot always condense the text by itself if it exceeds the recommended limit of cps (see section 2.1).

1.3 Translation Strategies

This section deals with possible translation strategies that are available to translators in order to deal with the typical constraints occurring in audiovisual texts. Díaz-Cintas and Remael (2021, p. 4) note that one challenge is that translators can usually only work with the dialogue although audiovisual material is composed of other codes as well.

In order to shorten subtitles when they do not conform to the guidelines, translators can render them more concise by condensing them or, when necessary, they can even omit certain parts. Both condensation and omission can be done at word and sentence level (Díaz-Cintas & Remael, 2021, p. 147).

There are various possibilities of how translators can condense the text in a subtitle. One example for a condensation at word level is translating the expression "mother and father" in the subtitle "You lied to us, son. Your own mother and father" as "parents" in French, resulting in the translation saying "Tu nous a menti, à nous, tes parents." Another example would be to use shorter near-synonyms or equivalent expressions, such as rendering "He's got lots of money" into "Il est riche" in French. Translators might as well change the word class, for instance by changing a verb into a noun like it was done in the following example. "Je me suis mis à travailler", which means "I have started working" was translated as "I found a job". By changing the verb for a noun, the subtitle only has 13 characters instead of 22 (Díaz-Cintas & Remael, 2021, pp. 152–159).

There are also numerous ways of how to condense subtitles at a sentence level, one of which is turning negative sentences into affirmative ones. This strategy has been used in the following sentence where the speaker talks about the place he used to live in. "OK, on n'habitait pas dans un palais…", which literally translates into English as "OK, we did not live in a palace…", was condensed into "OK, the place was small…". Another strategy to make subtitles shorter is to turn direct speech into indirect speech as shown in Table 1. By using this strategy, the number of characters in this subtitle was reduced from 73 to 52 (Díaz-Cintas & Remael, 2021, pp. 152–159).

Table 1: Turning direct speech into indirect speech, example taken from Díaz-Cintas & Remael (2021, p. 156)

Source Text (French)	**Literal Translation (English)**	**Condensed Translation (English)**
Souvent, je me dis : « Tant mieux qu'elle soit partie, comme ça, on est soulagés ».	I often tell myself: "Good thing she went, we're more at ease like this."	Sometimes I'm glad she went. It makes things easier.

When deciding to omit certain parts of a subtitle, the translator must make sure that this information is not necessary for understanding the content. The deleted information is often repeated in another subtitle or can be seen in the image. Omission at word level can consist of

leaving out adjectives or adverbs. Also, phatic expressions are commonly omitted in subtitles as can be seen in Table 2. The expressions *mais enfin* et *comme ça* emphasise the emphatic nature of this statement (Díaz-Cintas & Remael, 2021, pp. 162–163). By leaving these expressions out, the content is not changed, and the emphatic nature will probably also be transferred by the actors' gestures and the way they talk.

Table 2: Omission of phatic phrases, example taken from Díaz-Cintas & Remael (2021, p. 163)

Source Text (French)	Literal Translation (English)	Condensed Translation (English)
Mais enfin, Norah, on n'abandonne pas un bébé, comme ça, pendant des heures !	In heaven's name, Norah, one does not abandon a baby, just like that, for hours!	You don't abandon a baby for hours!

Omission at sentence level should only be done when it is unavoidable. For example, if a crowded scene with a lot of speakers is supposed to create a specific atmosphere, not everything has to be subtitled (Díaz-Cintas & Remael, 2021, pp. 164–165).

This section showed a selection of options a translator has if they need to shorten a subtitle due to spatial and/or temporal restrictions. These decisions require creativity, finesse, and a profound understanding of what is happening in the scene. As these are qualities that only human translators have, condensation and omission cannot be expected from a MT system to the same extent. However, there have been attempts to train MT systems specifically for translating subtitles, which will be discussed in the next chapter.

2 Machine Translation in Subtitling

Due to the special characteristics of subtitling, MT has long not been applied in this field as much as it has been in the translation of other types of text. However, in the last couple of years, there has been much more research on machine translated subtitles, and MT has been used increasingly in the subtitling industry. Many authors (Brendel & Vela, 2022; Burchard et al., 2018; Karakanta et al., 2020b) justify the need for more research on MT in subtitling by the increasing amount of audiovisual content that is produced and consumed. As a result, there is an increasing demand for audiovisual translation, and MT is used more and more often in order to be able to cope with this amount. Karakanta et al. (2020b, p. 63) note that audiovisual content is omnipresent in today's world. It is not only displayed on classic TV screens for entertainment, but people can access it from wherever they are on their smartphones and tablets. Also, many new forms of audiovisual content have developed over time, such as home-made videos, online courses or virtual conferences.

To give an example of the increasing production and consumption of audiovisual content over the last few years, I will present some numbers of the video streaming platforms YouTube and Netflix. In February 2020, around 500 hours of video content was uploaded to YouTube every minute, which is a 40% increase compared to 2014. Also in 2020, YouTube was estimated to have more than 2.1 billion users, who watched about one billion hours of videos per day (Statista, 2022). The growth in demand for audiovisual content can also be seen from the growth in the number of Netflix subscribers. While Netflix had around 50.65 million subscribers in the third quarter of 2014, they had over 223 million in the third quarter of 2022.

Although machine translation has been used increasingly in subtitling and there have been a lot of interesting research findings, machine translation does still struggle with certain characteristics inherent to subtitling. The next section will address some of these challenges.

2.1 Challenges for MT in Subtitling

In this section, I will briefly describe challenges that MT faces in the field of subtitling that are related to training data and characteristics of audiovisual translation.

For MT in general, large amounts of high-quality training data that are suitable for the intended purpose are crucial (Karakanta et al., 2020a, p. 3727). Bywood et al. (2017, p. 496) note that building a good corpus for subtitling is not easy due to copyright issues of the audiovisual material. Many subtitling companies have large high-quality corpora, but they are not always accessible to researchers as they are the property of the companies and their clients.

A challenge regarding corpora that has emerged more recently, is the need for corpora that help develop and enhance neural machine translation (NMT) systems that can fully or partially automate the task of subtitling. This means that not only the text itself is translated, but also other necessary steps such as spotting are done automatically. To train such systems, parallel corpora do not only have to contain a huge amount of high-quality data, but they must also contain sound and line break information. In many corpora, the latter information is lost as subtitles are merged into complete sentences before they are aligned. (Karakanta et al., 2020a, pp. 3727–3729). In their work, Karakanta et al. (2020a, pp. 3729–3733) present a corpus that they created for this specific purpose. It is called MuST-Cinema and is based on the MuST-C corpus, which is a multilingual corpus containing not only translations but also audios and transcriptions. This means that it has the potential to be used to train end-to-end subtitling systems. However, one drawback is that the subtitles were merged into full sentences before the alignment. Therefore, the authors annotated the transcriptions and translations in this corpus with symbols to add the missing information on line breaks and block breaks. They used the symbol <*eol*> to indicate a line break and <*eob*> to a indicate a block break, i.e. the end of a subtitle. Additionally, they developed a model that is able to insert these symbols into existing corpora. This model can considerably increase the number of lines that conform to the limit of 42 cpl. The success varied between the seven different languages, but the best results were achieved for the Italian corpus where after two rounds of training 92% of lines did not exceed 42 cpl. Before the training, it was around 30%. I think this is very useful as already existing parallel corpora can be adapted to the specific needs of new automatic machine translation systems, which enables more effective training.

Some other challenges for MT in subtitling result from some characteristic features of subtitling. Of course, challenges that MT faces in general also apply to the translation of subtitles. Gupta et al. (2019) list and briefly explain 27 mistakes that they encountered frequently in their research on machine translated subtitles. They divided the mistakes into three categories and conducted an experiment to find out which errors occurred most in which language combination. In the following, I will describe a few of these mistakes for each category as describing all of them would go beyond the scope of this section. Moreover, I will explore the types of errors that were most commonly encountered by the authors in the English-German language combination during their experiment.

The first two error types are in the category called *problems related to subtitle creation guidelines*. First of all, the machine cannot condense the text of subtitles, which is a frequently used technique to be able to stick to the spatial and temporal constraints (see section 1.3). For

instance, Gupta et al. (2019, p. 2) notice that machine translation engines often fail to avoid unnecessary repetitions. One example is the subtitle “Get over there! Hurry up! Get over there!”, which was translated into “Geh da rüber! Beeil dich! Geh da rüber!” into German by the MT engine, while the human translator left out the repetition and translated it into “Geht da rüber! Beeilt euch!”. It does also not change the number of subtitle blocks. Human translators might decide to merge or split subtitle blocks, for example, to conform to the allowed subtitle speed or to the limit of allowed cpl (Gupta et al., 2019, p. 3).

The second category that the authors identified is called *problems related to textual translation*. The inability to produce paraphrased translations in order to stick to the maximum number of cpl is another type of mistake that shows that MT is not capable of condensing text. For example, the subtitle saying “Come by and drive it whenever you want.” was translated into “Kommen Sie vorbei und fahren Sie es, wann immer Sie wollen.” by a machine. Although the translation is technically correct, it is way longer than the source text, which can be a problem if there is not enough time to display a translation that is longer than the original. It consists of 59 characters whereas the original subtitle consists of 39 characters. The human translator paraphrased the subtitle and translated it as “Komm jederzeit zum Fahren vorbei.”, which consists of 33 characters (Gupta et al., 2019, p. 3).

Another frequent issue that the authors observed was the fact that the source text can have different meanings and the MT engine often picks the wrong one as it cannot understand contextual information. In the case of subtitling, this is even more of an issue than in normal texts as the relevant contextual information might also be in the video (see section 1.1). Furthermore, the authors encountered word order errors in machine translated subtitles (Gupta et al., 2019, pp. 4–5). Table 3 shows one of their examples for this issue. If this phrase was a whole sentence, the machine translation would be grammatically correct. However, it is just a part of a sentence because the English source text does not start with a capitalized letter. The whole sentence might start with something like “Over the last x years…”. In English, this would not change the word order of the rest of this sentence. In German, however, the conjugated verb has to be at the second position of the sentence in main clauses, which is “hat” in this case. In the machine translation “die CO2-Menge in der Atmosphäre” is at position one and “hat” is at position two. However, if the previous subtitle already contains position one of the sentence, this subtitle has to start with the conjugated verb. The authors do not elaborate on the reasons why MT is not able to translate this correctly. I assume that each subtitle block is treated individually by the MT system, and it does not recognise sentences that stretch across several subtitle blocks.

Table 3: Word order errors in MT, example taken from Gupta et al. (2019, p. 5)

Source Text (English)	Machine Translation (German)	Human Translation (German)
the amount of CO2 in the atmosphere has increased nearly 40%,	die CO2-Menge in der Atmosphäre hat fast 40% zugenommen,	hat die CO2-Menge in der Atmosphäre fast 40% zugenommen

The third category, which the authors call *machine translation adaptability problems*, revolves around background knowledge that would be needed for a correct translation. Among others, this category encompasses errors related to cultural nuances, such as not differentiating between British and American English (Gupta et al., 2019, p. 6). However, I think that this strongly depends on the MT system used and whether it has been trained with exclusively British or American English or both. Wrong lexical translations and word structure errors are another two examples of errors in this category. Wrong lexical translations are defined as errors that do not conform with certain standards, such as glossaries, standard language, or industry usage. Word structure errors are morphological errors in words that are otherwise correct in terms of grammar and technical specifications. One example would be the translation of "deathly silent", which was translated as "Todesstille" by MT although it should actually be called "Totenstille" (Gupta et al., 2019, pp. 6–7).

The authors conducted an experiment to find out which error types occurred most in which language combination. They had a corpus of nearly 18,000 English subtitle blocks and machine translated them into six other languages, one of which was German. They asked professional subtitlers to classify the mistakes made by the MT system and to correct them. As the language combination of the present study will be English-German as well, only the results of this language combination will be reported here. A lack of paraphrasing was the biggest issue. It seems very logical to me that this is one of the biggest issues for MT because there is no pattern that you can always follow when a subtitle is too long. Word structure errors and word order errors predominantly occurred in the German target text. Lexical errors were also found frequently in this language combination.

2.2 Recent Technological Advances

This section will deal with promising recent developments in MT for subtitling. MT in the field of subtitling has traditionally been approached as text translation, meaning that other necessary tasks in the subtitling process such as spotting usually still have to be done by humans. This

considerable human effort that is still required to edit the MT output according to subtitling guidelines, as well as the fact that this approach does not consider the multimodal nature of subtitling, limit the efficiency of the use of MT in this field. Recently, there has been a great deal of research on how to further automate the subtitling process (Karakanta, 2022, p. 90). In the following, I will present some of these new developments.

One step of further automating interlingual subtitling is to generate the transcript, i.e. the source text, via Speech to Text (STT) technology. This technology has already been implemented on YouTube years ago. On YouTube, intralingual subtitle files are created automatically via voice data and are then translated by MT technology. However, the STT technology cannot recognize individual sentences, which has a negative impact on the quality of the MT output on YouTube (Song et al., 2019, pp. 1–2). In order to tackle this problem, Song et al. (2019) developed a sentence segmentation approach that predicts sentences and inserts period marks based on a neural network. They conducted an experiment to test their system and reported that it could accurately predict 70.84% of period marks. In their future work, the authors would like to further improve the accuracy by using other neural net models and more training data. Moreover, they are planning to build a system that is able to create subtitles for YouTube videos and translate them into Korean.

Matusov et al. (2019, pp. 82–85) built an NMT system in the language combination English and Spanish that is specialized in translating subtitles in the field of entertainment as the existing off-the-shelf NMT systems are not able to fully consider the specific features of subtitling. Therefore, they trained a baseline NMT model and adapted it to subtitling content and style. For example, they customised their system to shorter sentences and brief utterances, which is a typical feature of subtitles, by training it with several parallel corpora consisting of subtitles. The typically short sentences often provide little context for certain words such as pronouns, meaning that it is not clear from a sentence itself what is meant, but this information might be contained in the preceding sentences. To mitigate this issue, the authors built a training corpus where they spliced two or more successive sentences and their translations by using a special separator symbol. The NMT system can learn how to insert this symbol and use the provided context information.

In addition to a few other adaptations, Matusov et al. (2019, p. 85) introduced a new segmentation algorithm to insert line breaks in appropriate positions. When inserting line breaks, a few aspects must be considered (see section 1.2.2). The maximum number of characters and lines, as well as the subtitle speed can be computed as hard rules. However, it is not so easy to implement such rules for syntactically and semantically appropriate line breaks

that do not disturb the viewers' reading flow. Therefore, the authors developed a neural model that is able to predict segmentation positions.

The authors conducted an experiment to test the quality of the subtitles translated with their customised NMT system. They used a documentary and a sitcom as source material and divided each one into three parts. They asked two experienced translators to translate one part from English into Spanish and post-edit the other two parts, one of which had been translated by the baseline NMT system and the other one by the customised NMT system (Matusov et al., 2019, p. 86).

In a first step, the authors used the automatic quality metrics BLEU, TER and characTER on both machine translated parts of the documentary and the sitcom. In all cases, the quality of the subtitles translated by the adapted NMT system was better. However, the difference in quality between the two MT systems is higher in the sitcom; probably because the style differs more from the baseline NMT system's training data (Matusov et al., 2019, p. 87).

In a second step, they measured the post-editing (PE) effort by using automatic metrics and measuring the work time. Additionally, they also conducted a survey with the two post-editors. The results show that the number of corrections was significantly lower in the output of the adapted MT system. Compared to the documentary, the number of corrections was generally higher in the sitcom, but the reduction in PE effort was still very high. Post-editor 1, for example, reached a Human Translation Edit Rate (HTER) score of 36.7% when post-editing the baseline MT output of the documentary and 27.8% when translating the adapted MT output. In the case of the sitcom, the HTER score of post-editor 1 was 73.8% for the baseline MT output and 44% for the adapted MT output. Moreover, both translators needed less time for PE than for translating from scratch. Both translators needed less time for post-editing the adapted MT output. However, the authors stress that this was a very small experiment with only two participants, which means that no general conclusions can be made from these findings (Matusov et al., 2019, pp. 88–90).

In my opinion this sounds like a very promising approach that could increase translators' efficiency to a large extent. However, the study did not analyse whether the quality between the post-edited target texts and the translation from scratch is equally good. As PE and translating are two different kinds of task, this aspect should be considered as well when comparing them. Moreover, at the time the paper was published, this adapted system existed for only one language pair, which means that a great deal of work is still needed to make this system available in more languages.

Papi et al. (2022) developed an NMT system for translating subtitles too, which is trained for more language combinations. The authors describe their system as the first tool that is able to perform all the steps that are necessary in subtitling, including the generation of timestamps. They conducted an experiment to evaluate the quality of their system's translations of subtitles, timing and segmentation in comparison to five cascaded systems that split the process of subtitling into different phases. The before-mentioned NMT system that was customised to subtitling by Matusov et al. (2019) was one of the five tested systems. However, all systems were anonymised in this study, which means that it is not clear how it performed in this experiment. Several metrics, namely BLEU, Sigma and SubER, were used to evaluate the translation and segmentation quality of all systems for different languages in different settings. Additionally, they calculated how many subtitles did not exceed the maximum number of cpl and cps. The results show that their system performed better than the cascaded systems when translating TED talks. In the other scenarios other systems achieved better results, but usually the difference was not very high. In the cps category, especially, the other systems performed better.

These studies show that quality evaluation of machine translation systems is very important as it shows what can be expected from the respective systems and what aspects need to be improved. Therefore, the next chapter will revolve around different ways of evaluating the quality of machine translation systems, especially in the field of subtitling.

3 Quality Evaluation of Machine Translation Output

This chapter deals with the evaluation of machine translation quality, which is a very important task and a vital field of research (Popović, 2018, pp. 130–131). The first section will briefly discuss the reasons why MT evaluation is a very important task. The following sections will deal with the different types of MT evaluation in more detail.

3.1 Reasons for Evaluating the Quality of Machine Translation Output

MT evaluation is often done by researchers or MT system developers, who are building or improving their system. There are several reasons why MT systems are evaluated, and the best evaluation method depends on the purpose of the evaluation. This section will give an overview of the possible reasons for the evaluation of MT.

According to Chauhan and Daniel (2022, pp. 4–5), there are three main tasks of MT evaluation. It can be used, for example, to analyse errors. In this case, a type of error in the MT model is identified. After finding possible reasons for this problem and developing a potential solution, MT developers can implement it in their model and train it. Then they evaluate the model again to determine whether the problem has been fixed. Another task of MT evaluation is the comparison of the performance of MT models by comparing their outputs. The third task is optimization of the internal parameters of the MT model to improve the quality of the output.

Popović (2018, pp. 130–131) distinguishes between two types of quality evaluation that are performed for different reasons. The first type are overall scores and rankings of different translations. They provide valuable information about the overall quality and help researchers and MT developers to improve MT systems. However, they sometimes need more detailed information, which can be collected via the second type, namely error classification and analysis. As the name already suggests, the actual errors in the MT output are identified and classified. Then this information is used to create error profiles, to see how errors are distributed over specific error classes or to compare the output of different MT systems. This more detailed kind of evaluation can give information on, for example, strengths and weaknesses of a particular MT system or its biggest problems. It can also answer more in-depth questions, such as whether a system that has been ranked worse performs better in one specific aspect.

This section outlined that machine translation evaluation is very important for the improvement of MT systems and can analyse many different aspects. The following sections will discuss a selection of human and automatic evaluation methods, including some very recent ones that were specifically created for the evaluation of machine translated subtitles. In the end,

the approach of taking the viewers into consideration when evaluating the quality of subtitles will be discussed as well.

3.2 Human Evaluation

There are several ways how the quality of MT can be evaluated by humans. For example, humans can evaluate the quality of an MT output by rating it on a scale or by conducting a more detailed quality evaluation in the form of error annotation.

When rating the output, evaluators can rate different aspects of quality. For example, they can rate how correct the meaning has been transferred, i.e. the adequacy of the translation. They can also rate how understandable or how fluent the target text is (Sanders et al., 2011, p. 807). Besides rating translation outputs, there are two main approaches when humans evaluate the quality of machine translations in more detail. One approach is to identify mistakes in the output and categorise them. During this task, the annotator should have access to the source text, a reference translation or both. With the increasing importance of PE, another approach, namely the analysis of the PE activity, has become more common. In this approach, each PE operation is assigned an error category. This is useful, for example, to study the relation between certain types of mistakes and PE effort (Popović, 2018, pp. 131–132).

The human evaluation of MT output, in particular error annotation, is a very time-consuming and therefore expensive task because a person must analyse a text in great detail (Popović, 2018, p. 142). Moreover, this process is unreproducible and there is a certain degree of inconsistency involved as different annotators will interpret mistakes differently. One possibility to overcome this issue is to ask more than one person to annotate errors in order to have more opinions and even out differences. Whether or not the annotator is familiar with the subject matter or the type of language used in the text, will influence the annotation decisions as well (Sanders et al., 2011, p. 808). Moreover, the definition of an error typology that is suitable for the intended use is already a difficult task itself. For example, the number of error categories might have an impact on the outcome of the annotation as a higher number of error categories can lead to a lower consistency in the annotation results of different annotators (Popović, 2018, p. 154).

Despite these drawbacks, human evaluation of MT output is still reasonable in some settings as it provides a much more detailed analysis than automatic evaluation (Popović, 2018, p. 154). Moreover, only humans have the ability to really understand the texts. Therefore, human judgment alone can tell how severe a mistake really is because the severity strongly depends on the context and type of text. It can also be argued that translations are produced for

humans, which makes them the ideal measure for the quality (Sanders et al., 2011, pp. 807–808).

The Multidimensional Quality Metrics (MQM) is an error typology that has been used very frequently in studies on the evaluation of MT or subtitles (Matusov et al., 2019; Xie, 2022). It was developed between 2012 and 2014 in the context of an EU-funded project and has been harmonized with the DQF error typology (Lommel, 2018, p. 110). One important characteristic of MQM is that it aims at standardisation while still considering the individual purpose of a translation by providing a framework out of which an evaluator can choose the error types that are important to their specific needs, instead of suggesting one list of error types that applies to all kinds of translations. Moreover, MQM is based on a hierarchical approach, and for each project, it can be decided which level of detail is appropriate. For example, in some cases, it might be enough to categorise an error as grammatical error, and in other cases, it might be important to know which kind of grammatical error it is (Lommel, 2018, pp. 113–114). In general, metrics generated via the MQM framework "should be as course-grained as possible while still serving their purpose" (Lommel, 2018, p. 116) because it is more difficult for annotators to agree on the error type if they are more fine-grained, as already mentioned above when discussing the drawbacks of human evaluation. On the top level, the MQM framework has seven so-called dimensions (terminology, accuracy, linguistic conventions, style, locale conventions, audience appropriateness & design and markup), which each contain a certain number of error types (MQM, n. d.).

Other important characteristics of the MQM are severities and weights. There are four severity levels that refer to the errors themselves and how severe their effect on the usability is. Weights can be assigned to a whole dimension. If, for example, the correct use of terminology is very important, the weight of this dimension can be adjusted and every error in this dimension counts more than errors in the other dimensions (Lommel, 2018, pp. 120–121). These two characteristics further add to the customization possibilities of the MQM framework. After categorizing the errors, a score can be calculated if desired. To get an overall score, the penalty points have to be calculated in a first step by multiplying each error by its severity value and its weight. Then the penalty points must be divided by the word count and subtracted from the number one (Lommel, 2018, pp. 121–122).

The MQM seems to be a very useful tool for the evaluation of quality in translations as it can be adjusted to different needs by choosing the relevant dimensions. The hierarchical structure, the severity levels and the weights allow for further customization. In my research, I found a great number of articles where the MQM was used for the evaluation of both MT and

human translations in a variety of settings and for different text types including subtitles. Hence, this framework appears to be a very commonly used method of human evaluation. Nevertheless, I found attempts of establishing a new manual quality metric by Brendel and Vela (2022) that was specifically created for the evaluation of machine translated subtitles.

Brendel and Vela (2022) developed a new method for assessing the quality of machine translated subtitles to improve MT systems and the quality assessment process in subtitling. Their new approach consists of three steps. The first one is manual error annotation. The authors use an annotation scheme that was developed by Martínez and Vela (2016, p. 2248) and consists of the four dimensions content, language, format and semiotics. The first two dimensions could be used for any kind of translation, while the last two dimensions designate subtitling issues such as maximum cpl and cps or inter-semiotic coherence. After the annotation process, the acceptability of the found errors is evaluated by using the following four categories: perfect (no modification needed), acceptable (only small mistakes that do not require modification), revisable (errors that require PE) and unacceptable (retranslation required). To be more efficient in finding the strengths and weaknesses of an MT system and to better predict the quality and PE effort in the output, Brendel and Vela (2022, p. 55) added some error types that occur frequently in machine translated subtitles, according to the findings of Gupta et al. (2019).

The second step is to measure the inter-annotator agreement (IAA), which describes how well the annotations of different annotators match. The third step, determining the PE effort, depends on the IAA. The authors developed a new method where they use the interquartile range and the median to determine the IAA per line for each subtitle. In their model, the median ranges from 1 to 4 where 1 is perfect quality, 2 is acceptable quality, 3 is revisable quality and 4 is unacceptable quality. Figure 2 shows the decision tree for the determination of the IAA and the PE effort. If the interquartile range, which is 50 % of all data, is < 1 the annotators' opinions lie close together, which means that their opinion is reliable. This means, for example, that if the IQR is < 1 and the median is 1, there is a substantial agreement that the quality of the respective line is perfect, which means that it does not require any PE. If the IQR is < 1 and the median is 2, low PE effort is required. If the median is 3, medium PE effort is required and if the median is 4, the subtitle line needs to be translated from scratch. If the IQR is between 1 and 2, the subtitle line requires PE. In this case, it must be checked whether the 75% quartile is 2, 3 or 4 to determine the PE effort (low, medium or retranslation). Perfect quality is not possible anymore as the opinions of the annotators lie too far apart to be of sufficient reliability. If the IQR > 2 the subtitle line needs to be retranslated (Brendel & Vela, 2022, pp. 55–57).

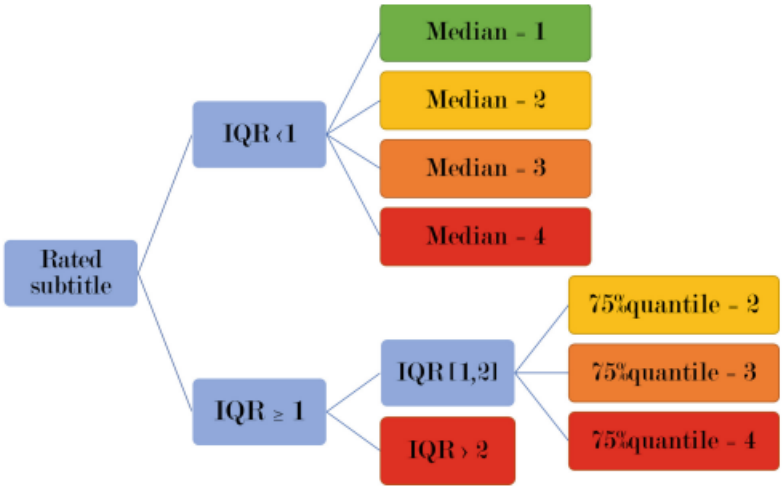

Figure 2: How to determine the IAA and PE Effort (Brendel & Vela, 2022, p. 56)

When comparing this method to the MQM, the advantage of this approach is that subtitles which contain mistakes that are difficult to categorise correctly or mistakes that are easily overlooked and therefore result in a low IAA, can be found efficiently. However, it involves many resources as there must be several annotators in order to be able to calculate the IAA. When using the MQM, it is possible to only have one annotator or more if necessary, which makes it suitable for a wider range of settings. However, there was no information on the exact number of annotators involved in this study. Moreover, the MQM framework is more suited to create a fine-grained analysis of the mistakes made due to the higher number of dimensions and error types and its customisability.

3.3 Automatic Evaluation

The quality of MT output can also be evaluated automatically. There has been a lot of research in automatic quality metrics, which is why a great number of metrics have developed over time. Automatic quality metrics usually calculate a score by comparing the MT output to a reference translation created by a human translator. In general, the score is either based on the percentage of matched n-grams, which is a sequence of a specific number of words, or the edit distance, which represents how many changes have to be made to match the reference translation (Popović, 2018, p. 130).

As opposed to human evaluation methods, the advantages of automatic quality metrics are that they are very fast and cheap. Moreover, there is no subjective judgment involved, which

enables more consistent results. However, they cannot provide the same level of detail as human error annotation (Popović, 2018, p. 130). Consequently, the choice of evaluation method depends on the circumstances and the goal of the evaluation.

In the following, I will describe two automatic quality metrics that are used very frequently and that I encountered in some studies on the quality of machine translated subtitles (Karakanta et al., 2020a; Matusov et al., 2019; Papi et al., 2022). Then I will describe a novel neural metric and one that has been developed specifically to be used on subtitles.

The first one is called BLEU (Bilingual Evaluation Understudy). It is based on the comparison of n-grams and deemed very reliable (Chauhan & Daniel, 2022). It was developed by Papineni et al. in 2002 and is based on the thought that "The closer a machine translation is to a professional human translation, the better it is" (Papineni et al., 2002, p. 311). Accordingly, BLEU is a numerical metric that measures translation closeness to a corpus of high-quality reference translations created by humans. It works on the basis of modified precision of n-grams. To calculate precision, the number of words of a machine translated sentence that also appear in a reference sentence must be divided by the total number of words of the machine translated sentence. However, this is not a suitable measure as it can result in high-precision scores for bad-quality output. For example, if the reference sentence was "The cat is on the mat." and the MT output was "the the the the the the the ", the precision would be 1 (7 divided by 7) because each word of the machine translated sentence is also in the reference sentence. As 1 is the best possible score, it is clear that this is not an adequate score for this example. Therefore, the BLEU metric uses modified precision, in which the credit a machine translated sentence can receive is limited by the maximum number of times a word appears in either of the reference sentences. In this example, the word "the" appears twice in the reference which means that the modified precision of the machine translated sentence is 2/7. In more realistic examples, where the machine translation output consists of several words, the same is done for each word of the MT output and then the results for each word are added up. The same can be done for larger n-grams such as 2-grams (Papineni et al., 2002, p. 312).

The BLEU metric also contains a brevity penalty because it is easier for short machine translated sentences to reach a higher score as there are fewer words that might not be in the reference translation. The brevity penalty is applied to the whole corpus and not to individual sentences (Papineni et al., 2002, p. 315).

The second metric is called TER (Translation Edit Rate) and measures how much PE a human would have to do in order to make the MT output look exactly like the reference translation. All edit operations, namely insertion, deletion, substitution of single words and

shifts of individual words or several words in a row, weigh equally. The calculation of the number of edits is done in two phases. In a first step, a greedy search algorithm is used to find the number of shifts by selecting all shifts that lead to smaller number of insertions, deletions and substitutions. In a second step, dynamic programming is applied to calculate the number of insertions, deletions and substitutions. If there are several reference translations, this is done for every reference and the best score is used. As the greedy search does not work so well in longer sentences, beam search is used as well (Snover et al., 2006, pp. 223–226).

Although still used frequently, Rei et al. (2020, n. p.) state that with the advent of NMT, traditional metrics such as BLEU are not up to date anymore as they can only work on the lexical level. Therefore, they developed COMET (Cross-Lingual Optimized Metric for Evaluation of Translation), a neural framework for training MT evaluation models that can be used as a metric for MT evaluation. COMET is based on cross-lingual language modelling, which has seen considerable breakthroughs shortly before the metric has been developed, and predicts human judgments. It can be based on different human judgment metrics and frameworks such as the MQM framework. In contrast to traditional automatic quality evaluation metrics, COMET evaluates the quality of MT output by using not only a reference translation but also the source text and achieves a higher correlation with human judgment.

According to Wilken et al. (2022), metrics that were drawn up for normal texts are not really suitable for the quality evaluation of automatically created subtitles as they only evaluate plain text and do not consider subtitle-specific characteristics such as timing information or line breaks. For this reason, they developed SubER (subtitle edit rate), an automatic quality metric that is segmentation- and timing-aware. It can be applied to subtitles generated by automatic speech recognition and to machine translated subtitles. SubER is based on TER due to its interpretability and because segmentation and timing information can be integrated easily. It is calculated by adding up the number of word edits, the number of break edits and the number of shifts and then dividing it by the sum of the number of reference words and the number of reference breaks. There are three possible word edits: insertions, deletions and substitutions. Break edits can also consist of insertions, deletions and substitutions. Shifts are defined as the movement of a sequence of tokens to a matching phrase in the reference. A shift may consist of any number of words or breaks. Wilken et al. (2022) conducted an experiment to evaluate their new metric where they measured how well the automatic evaluation correlates with human judgment. In their experiment, it correlated well with human judgment, and it showed a higher correlation than metrics that only consider plain text or integrate timing and segmentation information in different ways. However, this metric was published not very long ago, and more

studies are needed to confirm these results. The authors encourage the usage of and research on their new metric by publicly releasing its source code.

3.4 User-centred Evaluation

Research on the evaluation of the quality of machine translated subtitles usually focuses on the language itself, i.e. accuracy and fluency, and on formal parameters like segmentation and timing. There are only few studies in this field that focus on the viewers' reception of machine translated subtitles. In this section, I will present two of those studies.

The first one was conducted by Hu et al. (2020) who compared Chinese viewers' reception of machine translated subtitles, post-edited subtitles and human translated subtitles of Massive Open Online Courses (MOOC). They based their choice of methodology on Gambier's model of reception. According to Gambier (2009, p. 53), reception consists of three parts. The first one is response which is the perceptual decoding. The second one is reaction which refers to the processing effort, and the third one is repercussion and refers to the attitude and the sociocultural background. In order to measure the response, Hu et al. (2020, p. 523) used the eye-tracking parameters glance count, glance duration, and fixation duration because they show where the viewers direct their attention. To receive data on the reaction, they measured the average fixation duration. In addition, they used comprehension questions. For data on the repercussion, the authors asked attitude questions. There were 66 participants who were divided into three groups, and each group watched an MOOC with one of the three different kinds of subtitles.

The results of the study show that most statistically significant differences in the above-mentioned eye-tracking parameters appear in the comparison of the raw machine translated subtitles and the human translation. In terms of comprehension questions, the post-edited subtitles scored the highest. However, the authors only found a statistically significant difference between the group that watched the post-edited subtitles and the one that watched the raw machine translated output. The results of the attitude questions show that the participants generally have a positive attitude towards the kind of subtitle they were watching. The post-edited subtitles again showed the best results. To the surprise of the authors, the subtitles translated by a human translator received the worst rating (Hu et al., 2020, pp. 527–532). One of the study's limitations, the fact that the human translation was not created by a professional translator, might be the reason for the unexpected results (Hu et al., 2020, p. 534).

The second study, which was conducted by Xie (2022), is a pilot study that compares the quality and users' reception of machine translated subtitles on YouTube and Bilibili, which

is a Chinese video platform that has recently introduced machine translation. To evaluate the quality, the author used the MQM framework. He used the dimensions accuracy (addition, mistranslation, omission), design (segment length), fluency (grammar, spelling, typography), style (register, collocations) and timing (cps). To evaluate the viewers' reception, he used the same kind of questionnaire as Hu et al. (2020), i.e. comprehension and attitude questions. However, unlike Hu et al., he did not use eye-tracking. 44 participants were divided into two groups where one group watched a video with subtitles generated on YouTube and the other one watched a video with subtitles generated on Bilibili (Xie, 2022, p. 5).

The results of the quality evaluation in this study show that the subtitles generated by YouTube have fewer mistakes in the dimensions of accuracy and fluency (13 mistakes as opposed to 18 in the subtitles by Bilibili). In the dimensions of design, style and timing, there are no statistically significant differences. However, the author does not state any specific numbers for these two dimensions. Also, the results of the comprehension questionnaire in the two groups are very similar, which indicates that accuracy and fluency might not have an impact on comprehension. However, the video shown in this study was only two minutes long and only two comprehension questions were asked, which limits the generalisability of the results. The results might be different for longer videos. In both groups, the majority of the participants has a positive attitude towards the subtitles of the video although the percentage is higher for the group that watched the video with subtitles generated by YouTube (Xie, 2022, pp. 6–11).

The tentative results of these two studies show that machine translated subtitles in the language combination English – Simplified Chinese are comprehensible for humans. In addition, the viewers had a predominantly positive attitude towards them. One study also showed that the difference in reception between post-edited subtitles and subtitles created by humans or machines is not significant for many parameters measured in the study. All in all, machine translated subtitles performed quite well in these two studies, however, it must be taken into account that both studies have important limitations and to my knowledge, there are no other studies that confirm or rebut these results. The next chapter will go into more detail on the concept of cognitive load and how it has been studied in the context of subtitling.

4 Cognitive Load and Eye-tracking

This chapter revolves around the theoretical construct of cognitive load and how it can be measured. As eye-tracking is one of the most common methods for the assessment of cognitive load and has also been used more frequently in the field of subtitling, it will be dealt with in this chapter as well.

4.1 Cognitive Load – a Theoretical Construct

Cognitive load is an active field of research and has evolved a great deal over the past decades. I will start this section by describing one theory of cognitive load of the 1990s because I have come across elements of this theory in many eye-tracking studies on subtitling. Then I will proceed with a more recent theory.

According to Paas and Merriënboer (1994, p. 353), cognitive load is "a multidimensional construct that represents the load that performing a particular task imposes on the cognitive system of a learner". Figure 3 shows a representation of this construct, which consists of causal factors and assessment factors. Causal factors affect the cognitive load and can either stem from the task, the subject that is performing the task, or interactions between the task and the subject. Examples of causal factors relating to the task are time pressure or task novelty. This means that the need to perform a specific task under time pressure or doing something for the first time will likely increase the cognitive load. Causal factors relating to the subject, such as cognitive capabilities and previous knowledge, are usually stable and not likely to change fast. In the case of subject-task interactions, the cognitive load can for example be affected by the interaction between the type of task and the subject's preferences (Paas, Fred G. W. C. & van Merriënboer, 1994, pp. 353–354).

There are three assessment factors that can be used to determine the cognitive load. As can be seen in Figure 3, the assessment factors comprise mental load, mental effort, and performance. Mental load is predetermined by the task itself or the environment in which it is performed. If a task is more complex, it will usually induce a higher mental load no matter what the subject's characteristics are. Mental effort refers to the number of cognitive resources that a person actually invests to perform a task. The performance on the task can also be used as an indicator of cognitive load (Paas, Fred G. W. C. & van Merriënboer, 1994, pp. 354–355).

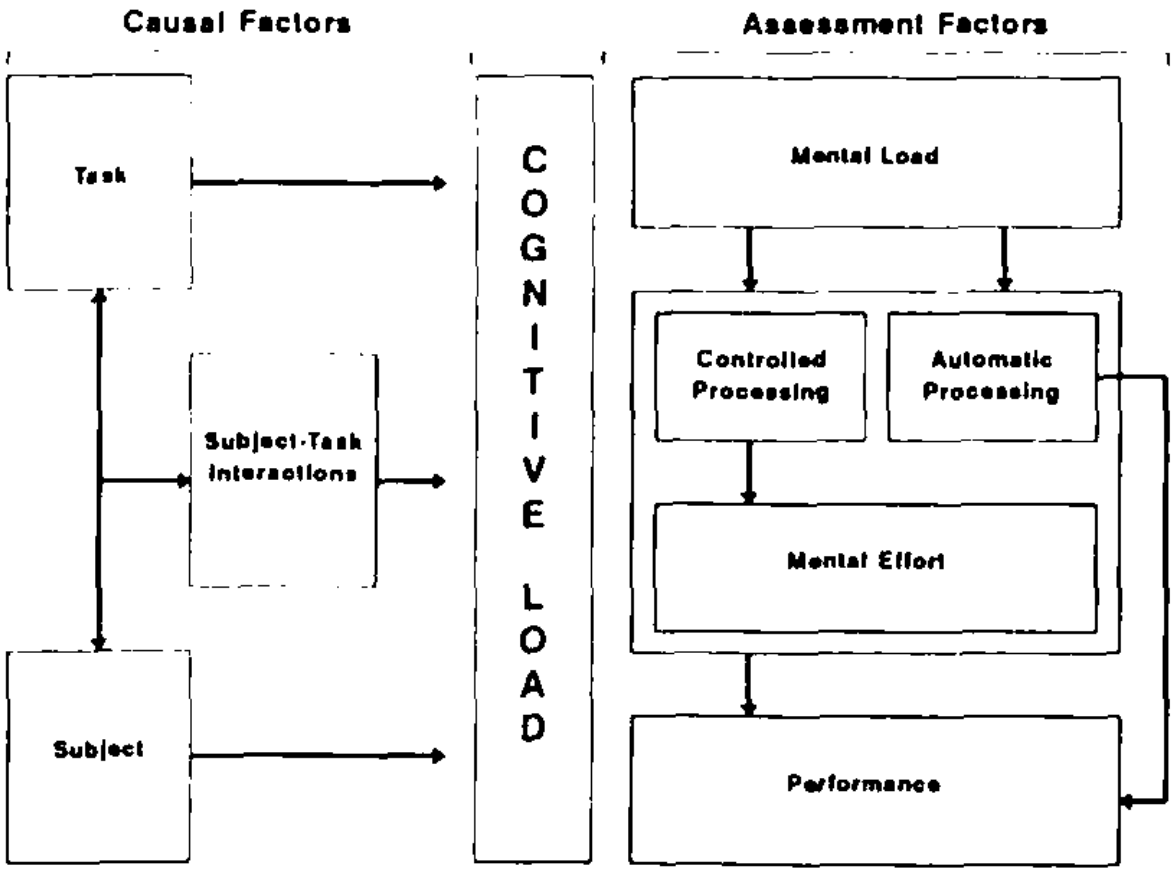

Figure 3: Multidimensional construct of cognitive load (Paas, Fred G. W. C. & van Merriënboer, 1994, p. 353)

According to Paas and van Merriënboer (1993, p. 742), to get the most accurate results of the cognitive load during a certain task, one has to measure a combination of mental effort and performance rather than using only one of them. There are several methods of measuring mental effort, including self-reports, EEG or eye-tracking. Performance is usually measured with a test score where a lower score is expected if the task is more difficult and more cognitive resources are needed to solve the task (Szarkowska & Gerber-Morón, 2019, p. 147).

Sweller et al. (2019) discuss the last 20 years of developments in cognitive load theory, which they first developed in 1998. They define the aim of cognitive load theory as "explain[ing] how the information processing load induced by learning tasks can affect students' ability to process new information and to construct knowledge in long-term memory" (Sweller et al., 2019, pp. 161–162). As opposed to the theory described above, the terms long-term memory and working memory come into play. This theory relies on the assumption that the working memory limits cognitive processing as it can only process a small number of information elements at once. This means that if unnecessary elements are presented to the

learner, their cognitive load will increase. If it becomes too high, there will be negative effects on the learning process. The long-term memory, on the contrary, for which there is no known limit, contains huge amounts of already cognitively processed and organised information. This means that if information ought to be stored in the long-term memory, it must be presented in a way that the working memory can process it effectively (Sweller et al., 2019, pp. 262–263).

Originally, there were three categories of cognitive load, namely intrinsic, extraneous, and germane cognitive load. Intrinsic cognitive load is the level of complexity of the processed information. The complexity cannot be determined by simply evaluating the characteristics of this information. The knowledge of the person that processes the information must also be taken into account as knowledge stored in the long-term memory has an influence on how easily information can be processed by the working memory. This means that intrinsic cognitive load can be changed either by changing the learning matter or by changing the knowledge background of the learner (Sweller et al., 2019, p. 264).

Extraneous cognitive load refers to how the information that has to be learned is presented and how the learner is instructed. It can be changed easily by changing the instructional procedure (Sweller et al., 2019, p. 264).

In the original theory, germane cognitive load referred to the working memory resources that are invested in dealing with the intrinsic cognitive load. If more resources needed to be devoted to the extraneous cognitive load, fewer resources could be invested in the intrinsic cognitive load, which resulted in worse learning outcomes. In this definition of germane cognitive load, it was assumed that the germane cognitive load substitutes for the extraneous cognitive load. Over the years, the authors adapted this part of the theory in that germane cognitive load does not contribute to the total cognitive load but rather redistributes working memory resources from extraneous cognitive load to intrinsic cognitive load (Sweller et al., 2019, p. 264).

4.2 Measuring Cognitive Load

There are many ways of how cognitive load can be measured. Brünken et al. (2010) categorise these methods of determining cognitive load in subjective, objective, and combined methods. In this section, I will describe a few methods of these three categories.

Subjective methods typically involve asking the participants either for their perceived cognitive load or for task difficulty. Rating the perceived cognitive load is usually done via a 7- or 9-point Likert scale. The main benefit of this method is that it is very simple and easily feasible. The drawbacks of using this method include that it usually evaluates the task as a

whole and not specific elements of the task and that it is not always clear which category of cognitive load it measures. Rating self-perceived cognitive load is often combined with a rating of task difficulty. It can be used to evaluate the extraneous cognitive load if it is rated differently when presented in different forms, i.e. when the extraneous cognitive load is changed. However, it can also be used as an indicator for intrinsic cognitive load, e.g. when the participants' level of expertise varies (Brünken et al., 2010, pp. 182–184).

Objective methods can be divided into measures that relate to the learning process in outcome variables, input variables, and process-related behavioural variables. Outcome variables refer to learning outcomes. These measures should only be used in combination with other methods to assess cognitive load as it cannot be ruled out that the differences in learning outcomes are caused by, for example, differences in motivation and not in differences in cognitive load. Input variables refer to task complexity or task difficulty. According to cognitive load theory, the intrinsic cognitive load must be higher in more complex tasks. Finally, behavioural variables refer to the participants' behaviour during the learning process. There are several behavioural parameters, including functional magnetic resonance imaging (fMRI), heart rate, and eye-tracking (Brünken et al., 2010, pp. 184–187).

Combined methods calculate the relation of mental effort and performance. These methods enable researchers to optimise the instructional design by identifying efficient and inefficient aspects of the instructional design (Brünken et al., 2010, pp. 191–192).

According to the authors, there is no best or worst measure for cognitive load. Each of them has advantages and disadvantages, and choosing the best method strongly depends on the aim and research question of a study (Brünken et al., 2010, p. 194). The following section will discuss eye-tracking in more detail as it is used to assess cognitive load in the present study.

4.3 Eye-Tracking

By tracking eye movements, researchers can draw conclusions on how participants process certain stimuli. This research method has been used increasingly in subtitling studies, such as in Hu et al. (see section 3.4) whose goal was to learn about the perception and cognitive processing of machine translated, post-edited and human translated subtitles. However, eye-tracking has also been used much by other researchers studying other subtitling aspects. Before describing some of these studies in detail, I will briefly explain what exactly eye-tracking is and how it can be used to measure cognitive load in translation studies.

4.3.1 Eye-Tracking – Definition and Parameters

Eye-tracking is a technology that records the movements of the eyes while they are reacting to a certain stimulus (Conklin et al., 2018). These data on eye movements show what a person is paying attention to and how much cognitive effort is needed to process the input material (Rayner, 2009, p. 1487). When reading, the eyes do not move forward continuously, but there are different types of eye movements called saccades, fixations, and regressions. Saccades occur when the eyes are moving forward and last on average between 30 and 40 milliseconds. In between the saccades, the eyes stop for 200-300 milliseconds, which is called a fixation. In general, readers only obtain new information during a fixation. Regressions are backward movements of the eyes and occur about 10 to 15 percent of the time when reading (Rayner, 2009, pp. 1458–1460). As these eye movements cannot be controlled and occur automatically, eye-tracking provides information on unconscious behaviour (Conklin et al., 2018, p. 2).

There are several eye-tracking measures that can be used in subtitling studies. Doherty and Kruger (2018, p. 48) categorise them in primary and secondary measures. Primary measures can be derived directly from the raw data, while secondary measures can be derived by combining at least two primary measures. According to the authors, fixation count, i.e. the number of fixations that occur within a specific Area of Interest (AOI), and fixation duration, i.e. the length of a fixation, are two commonly used primary measures in subtitling studies. It is believed that the higher the number of fixations and the longer the fixations are the more cognitive effort is needed to process a stimulus. They are often used in combination with the secondary measure gaze time, which is the sum of all fixation and saccade durations in a specific AOI. It is also called dwell time or visit duration. Another common secondary measure in subtitling eye-tracking studies is skipped subtitles, which is the number of subtitles that were fixated not even once (Doherty & Kruger, 2018, pp. 49–51).

Some less frequently used measures are saccadic measures, such as saccade count and saccade length as well as pupil dilation. Pupil dilation is in fact a widely used measure for cognitive load in general; however, it is not commonly used in subtitling studies because the dynamic nature of media can cause changes in pupil diameter that are not related to cognitive load but are rather a physiological response. Other eye-tracking measures that are not very common in subtitling studies include, for example, the time to the first fixation in an AOI, glance duration, and revisit duration (Doherty & Kruger, 2018, pp. 49–51).

Conklin et al. (2018, p. 6) list four advantages of using eye-tracking to study cognitive processing. Firstly, eye-tracking allows to measure cognitive processing while a specific task

is done, rather than evaluating the results of some kind of task that is done afterwards. Secondly, eye movements that are recorded with eye-tracking cannot be controlled by the participants as opposed to conscious judgments. Thirdly, eye-tracking provides temporal precision. Finally viewers can engage with the visual stimulus, e.g. a video or a website, as they would in a real situation. However, this also depends on the research design.

4.3.2 Examples of Measuring Cognitive Load with Eye-tracking

This section discusses how eye-tracking can be applied in subtitling studies to indicate the cognitive load of viewers by describing some examples. The first study was conducted by Szarkowska and Gerber-Morón (2019) and examines how two and three lines respectively are cognitively processed by the viewers and what their preferences are (see chapter 1.2.2). They conducted the same experiment twice, once with three groups of participants that each had a different mother tongue and once with participants with different hearing abilities.

The authors used a mixed-methods approach with eye-tracking, a comprehension test, and a questionnaire on self-reported cognitive load, enjoyment, and preferences. If three-liners cause a higher cognitive load than two-liners, the authors expect lower comprehension scores and lower enjoyment. Based on previous research, they expect longer mean fixation durations and more time spent on reading the subtitles if the cognitive load is higher. If three-liners induce a higher cognitive load, the authors also expect a proportionally higher reading time and more revisits, which indicate that the subtitles might be read less fluently (Szarkowska & Gerber-Morón, 2019, pp. 147–148).

The participants were shown two videos, one with subtitles consisting of two lines and one with three-line subtitles. Right after watching each part, the participants answered the comprehension questions and self-reported their cognitive load and enjoyment. After they had watched both videos, the participants stated which type of subtitles they liked better. The authors also conducted a short semi-structured interview with them to talk about their choices and views in more depth (Szarkowska & Gerber-Morón, 2019, pp. 150–151).

The results of the comprehension test showed that the number of lines did not make a significant difference. There were only slight differences in comprehension between the different groups of participants in each of the two experiments. In terms of self-reported cognitive load, however, significant differences could be found. In experiment 1, where each group of participants had a different mother tongue, all three indicators of cognitive load, namely difficulty, effort, and frustration, were significantly higher in three-line subtitles. In experiment two, the three indicators were higher in three-liners for hearing, deaf, and hard of

hearing participants as well, except for difficulty in the deaf participants. The self-reported enjoyment was higher in the case of two-liners in experiment 1, while no significant difference could be observed in experiment 2. In both experiments, most participants preferred two-line subtitles over three-liners. The eye-tracking results in experiment 1 showed that the absolute reading time and the proportional reading time was higher in three-line subtitles. In contrast, the number of revisits was similar in both conditions. In terms of mean fixation duration, there was only a significant difference in the group of English-speaking participants. There was no significant difference in participants speaking Spanish or Polish. In experiment 2, some eye-tracking results differed from the results in experiment 1. The absolute and proportional reading time are also higher in three-liners. However, the number of lines had no effect on the mean fixation duration in any of the three groups. The number of revisits was higher in three-line subtitles, especially among deaf participants (Szarkowska & Gerber-Morón, 2019, pp. 151–157). Although not all of the authors' hypotheses could be confirmed in the study, the results still show that the cognitive processing of three-line subtitles is more demanding.

Chan et al. (2022) also used eye-tracking in combination with other research methods to measure the impact of subtitles when learning something in one's second language. They compared the cognitive processing of subtitles in the participants' first language (L1), their second language (L2) and no subtitles at all in a video lecture held in their second language.

70 Chinese participants with a high proficiency in English participated in this study. They were divided into three groups and were asked to watch five 7-minute parts of an English video lecture. One group watched the videos without subtitles, one group watched them with English subtitles and the last group watched them with subtitles in Simplified Chinese. After each of the five videos, the participants had to answer five multiple-choice questions on the content and rate their self-perceived cognitive load. The experiment took place in an eye-tracking lab and the authors measured the mean fixation duration, the fixation count and the number of skipped subtitles (Chan et al., 2022, pp. 160–164).

For some parameters, namely comprehension score, mean fixation duration, and fixation count, the results showed statistically significant differences. The group that read subtitles in their L1 scored significantly better in the comprehension test, which suggests that readers benefit from L1 subtitles. L2 subtitles induced longer mean fixation duration and a higher fixation count, which means that L2 subtitles were harder to process than L1 subtitles. However, there were no significant differences in the self-perceived cognitive load and in the number of skipped subtitles (Chan et al., 2022, pp. 166–170).

The next study was conducted by Kasperavičienė et al. (2020) and does not investigate subtitles but the Lithuanian MT output of a news type article published by a well-known organization in English. The authors wanted to study the MT output of a small language and learn which errors in MT output caused comprehension problems while reading. Therefore, they studied whether errors in machine translated texts required more cognitive effort. They also wanted to know which error types increased the eye-tracking parameters gaze time and fixation count as well as how acceptable the machine translated text was to the readers. They used the eye-tracking parameters gaze time and fixation count as well as a post-task survey asking comprehension and acceptability questions to answer these research questions (Kasperavičienė et al., 2020, pp. 272–273).

In a first step, the authors assessed the quality of the MT output drawing on the works of different researchers (Petkevičiūtė and Tamulynas, 2011; Temnikova, 2010, 2016; Vilar et al., 2006) who established evaluation criteria for MT output. They divided the text into 58 segments and identified errors in 12 segments. These segments were the AOI in their experiment (Kasperavičienė et al., 2020, pp. 276–277).

They asked 14 participants to read the text translated by Google Translate. Afterwards, they were asked to answer three open comprehension questions and six five-point Likert scale statements on the acceptability of the text. While reading, the participants' eye movements were recorded. The authors analysed the parameters gaze time and number of fixations. They found that the gaze time and the fixation count in segments with errors was significantly higher than in segments without errors. Additionally, segments containing lexical errors (e.g. untranslated phrases, literal translations) induced the longest mean gaze time and the highest mean fixation count. Linguistic morphological errors (e.g. errors in case, number, person, gender) caused medium mean gaze time and mean fixation count, while systemic errors (e.g. missing words) induced the lowest mean gaze time and mean fixation count. The authors could not find a correlation between overall acceptability and the eye-tracking parameters (Kasperavičienė et al., 2020, pp. 277–281).

Overall, the authors concluded that when processing machine translated texts, the cognitive effort is higher in segments containing errors than in segments without errors. However, according to the authors, bigger studies are needed to confirm these results and further research on the acceptability of machine translated texts would be interesting as well (Kasperavičienė et al., 2020, p. 281).

All three studies discussed in this section used eye-tracking in combination with comprehension tests and some sort of self-assessment to gain insights on cognitive processes

while certain stimuli were processed. As cognitive load is a theoretical construct and hard to measure objectively, I think it is helpful to try to measure it from different perspectives. In these studies, the authors tried to measure it physiologically by tracking eye movements, while also taking into account the participants' views and their knowledge of the content after watching the videos. While the methodologies of the studies seem similar at first, the details are quite different. Although all studies examined the cognitive load during reading, they used different sets of eye-tracking parameters and self-assessment questions. As far as the comprehension questionnaires are concerned, two of the studies use multiple-choice questions; however, they might still be quite different from each other, for example, in the level of difficulty.

5 Research Design

This chapter gives a detailed overview of the research design employed in this thesis to answer the following research questions:

1. How good is the quality of the MT systems Google Translate and DeepL in the translation of the subtitles of a TED Talk in the domain of environmental protection from English into German and what kinds of mistakes do these systems make?
2. What effect does the overall quality of the machine translated subtitles have on the cognitive load and the attitude towards the subtitles?

In the first section, I will describe which methods I used to answer these research questions and why I used them. In the second section, I will go into the data collection process by describing the materials used, the participants, the eye-tracking apparatus and the procedure. Finally, I will explain how I analysed the data in the third section.

5.1 Methods

The research design is primarily based on Hu et al. (2020) but also draws on elements of Szarkowska and Gerber-Morón (2019). This experiment consists of two main parts, namely the quality analysis of the machine translated subtitles and the analysis of the cognitive load induced by these subtitles as opposed to human translated subtitles.

To analyse the quality of the machine translated subtitles, the MQM framework was employed in this experiment. I opted for a human evaluation method because it allows for a detailed in-depth analysis. The MQM framework not only allows the calculation of an overall score, but also requires the annotation and classification of errors, which results in a fine-grained analysis of the MT systems' performances. Moreover, the MQM framework is very versatile and can be customised to any project's requirements (see chapter 3.2). This feature is very useful for this project because issues that are specific to subtitling can be included in this metric. As can be seen in Figure 4, the entire set of dimensions was used as the goal was to get a comprehensive picture of the overall quality and the mistakes made. Generally, two hierarchy levels were used in order to get the desired level of detail. A third customised hierarchy level was only added to the error type Truncation/Text expansion in the dimension of Design and Markup, to include the error types *too many cpl* and *too many cps*. These two error types are very important when assessing the quality of subtitles because too many cpl or cps can hamper the reading process and cause unnecessary cognitive load in viewers. I decided to use TED's

subtitling guidelines, which stipulate a maximum of 42 cpl and 21 cps (see section 1.2.1), because the source material is a TED Talk, which is why the official translation has also been created according to these rules. Although MT systems cannot be expected to perform text condensation to the same extent as human translators can, it is still interesting how many subtitles adhere to these subtitling conventions because it has an effect on the viewers' cognitive load and general viewing experience.

In the MQM framework, the pass/fail threshold must also be customised according to the conditions of a specific project. The MQM website does not give any recommendations for thresholds for different kinds of projects. Therefore, I chose a threshold of 80% for acceptable quality. This threshold allows for quite a few minor errors, which are to be expected in a machine translation. 80% is also typically the lower end of the second-best school grade in Austria and is a value that would be acceptable in some situations but not good enough for a professional translation.

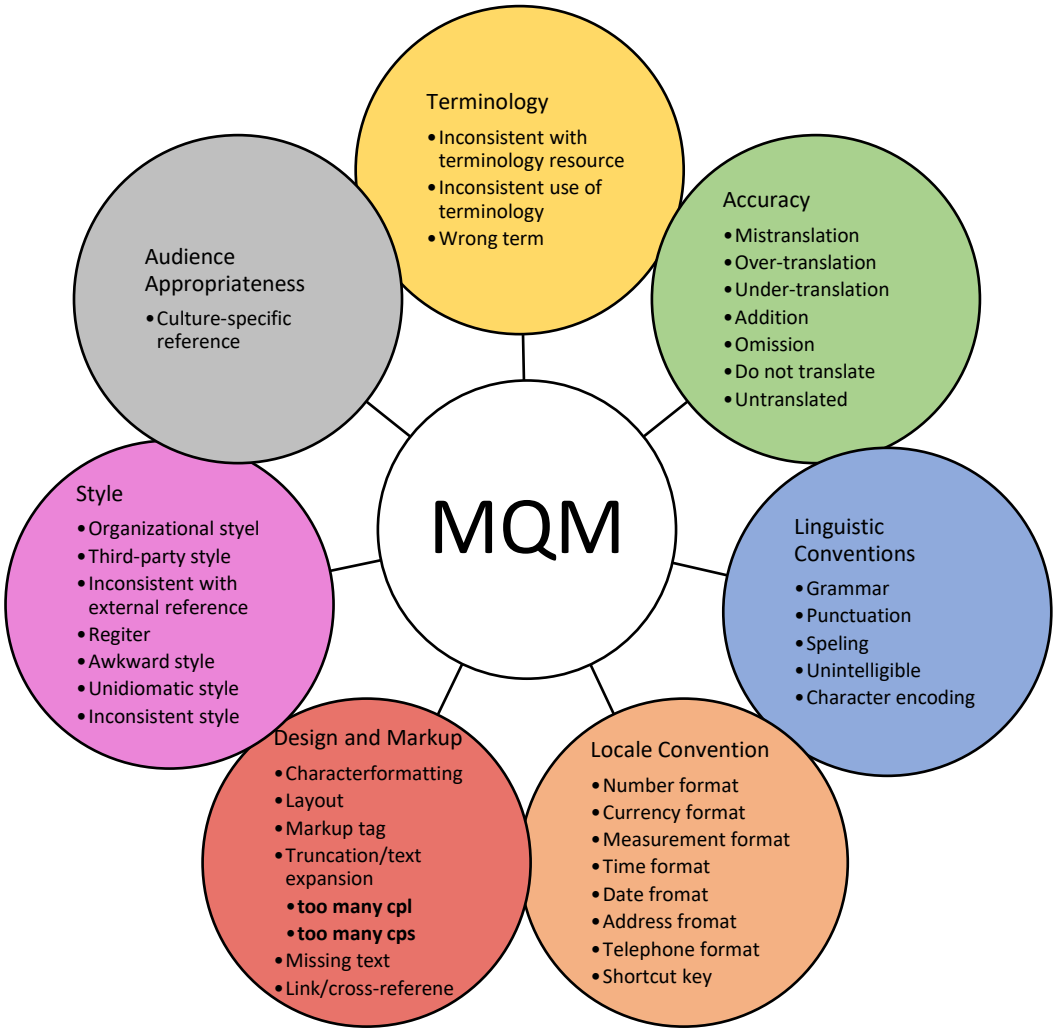

Figure 4: Experiment error typology based on the MQM typology

To assess the cognitive load, this study uses a quantitative approach, namely eye-tracking A mixed methods approach is used to determine the participants' attitude towards the subtitles, namely via a questionnaire consisting of a Likert scale and an open question. These methods were selected because the combination of them is frequently used in other subtitling experiments where differences in the appearance or production of subtitles are studied (e.g. (Chan et al., 2022; Hu et al., 2020; Szarkowska & Gerber-Morón, 2019). Eye-tracking allows to record eye movements, which give information about the cognitive load during the reading process. The choice of eye-tracking metrics, namely fixation count, mean fixation duration, and dwell time, is based on Doherty & Kruger (2018), who classified these parameters as commonly used in subtitling studies in their overview work.

Comprehension questions allow to obtain additional insights on how well the stimulus was processed by the participants, i.e. if the quality is good enough to allow an adequate level understanding.

Attitude questions or other forms of self-reported information, such as self-reported cognitive load, provide an interesting insight what the participants think of the subtitles. The comprehension and attitude questions are based on Hu et al. (2020). The comprehension questions are multiple-choice questions with four answer options and one correct answer. The attitude towards the subtitles is evaluated by five-point Likert scale statements. The statements were taken from Hu et al. (2020); however, not all 14 statements were taken because some of them did not seem appropriate in this study as the research design is different. Statement 1 asked the participants for their self-perceived comprehension of the content. The next three statements asked for how the participants perceived the reading process, and the last statement asked for how the participants would evaluate the quality of the subtitles. Moreover, one open question was included at the end of the questionnaire so that the participants had the opportunity to make comments or share their thoughts if they wished to do so.

5.2 Hypotheses

In this section, I will present the hypothesis for the experiment of the present paper. Based on the information collected in the literature review, the following hypotheses were developed.

In chapter 2, I described some new developments in machine translation in the field of subtitling, which showed that a great deal of work has been done recently to improve machine translated subtitles. As machine translation in general has experienced great progress over the last years, I hypothesise that the quality of the machine translated subtitles will be acceptable. However, I also assume that there will be clear differences in quality when comparing it to a

human translation because subtitling requires certain translation strategies in order to adhere to industry standards, which MT struggles with.

> Hypothesis 1a: Overall, the quality will be acceptable; however it will be clearly worse than the human translation.

As I am not using MT systems that have been trained specifically for the translation of subtitles, I assume that they will struggle with certain subtitling characteristics. Google Translate and DeepL are free standard MT systems that will likely not adhere to all subtitling industry standards related to time and space.

> Hypothesis 1b: Both MT systems will struggle to adhere to the subtitle characteristics related to time and space.

I expect the two MT systems to have different strengths and weaknesses because they must have been trained differently. Probably, the number of mistakes will be distributed differently among the dimensions and error categories between Google Translate and DeepL.

> Hypothesis 1c) DeepL and Google Translate will have different strengths and weaknesses, meaning that the errors they make will be distributed differently among the dimensions and error types.

As far as cognitive load is concerned, I predict that the eye-tracking parameters used will indicate a higher cognitive load in the participants if the quality of the subtitles is worse. I assume that comprehension will not be affected as much by the quality as eye-tracking parameters because it was not affected strongly in other similar studies either (e.g. Szarkowska & Gerber-Morón, 2019).

> Hypothesis 1d: The quality of the machine translated subtitles will have an effect on comprehension scores, but only a very slim one.

> Hypothesis 2a: If the quality of the subtitles is lower, the eye-tracking parameters fixation count, mean fixation duration, and dwell time will indicate a higher cognitive load.

Regarding the participants' attitude towards the machine translated subtitles, I assume that it will be worse if the quality of the subtitles is lower because the standards in subtitling have developed over time and have been adapted to the viewers' needs. There has been much research recently, which challenges current subtitling standards. However, these studies usually come to the conclusion that the participants prefer the subtitles that adhere to current standards. Therefore, I think that errors of different kinds, including errors relating to space and time restrictions, will have a negative impact on the participants' attitude.

> Hypothesis 2b: If the quality of the subtitles is lower, the attitude towards the subtitles will be worse.

5.3 Data Collection Process

This section provides a description of the data collection process. Firstly, it will explain which materials were used and why they were used. Secondly, it will briefly describe the participants and the eye-tracking apparatus. Finally, it will go into the procedure that was applied to gather the needed data.

5.3.1 Materials

A TED Talk called *How is your city tackling the climate crisis?* by Marvin Rees was used as a stimulus in this experiment because the subtitles published on the TED platform are generally of high quality and could therefore be used as a gold standard. Besides the requirement that the source language of the video was English and that German subtitles were available on the TED platform, there were three other criteria for choosing this video. Firstly, it was important that the video was as recent as possible to minimise the risk that MT engines had already been trained with the video. Secondly, the speaker should have English as their native language to minimise the risk of mistakes in the source text. And thirdly, the video should have a length of 10-15 minutes so that it is long enough to be split into three parts. There were only very few videos that adhered to all these criteria. I chose this video because it was quite dense in content, which would allow me to come up with a sufficient number of questions for the comprehension questionnaire.

After choosing the video, I divided the video into three parts that were about the same length each and prepared the questionnaire. I watched the English video, wrote down the key points and came up with the questions and four answer options for each question. One of them was correct and three were false. I created the questions before engaging too much with the

machine translations in order to avoid knowing the mistakes and being biased. There are no questions on what is going on in the image, as it is a presentation where not much is happening in the image. For each of the three parts of the video, I created an online questionnaire with LamaPoll, containing five multiple-choice questions, five five-point Likert scale attitude statements and one open question on the participants' opinion on the subtitles.

Google Translate and DeepL were used to translate the SRT-file. The whole files were used for the quality analysis. For the eye-tracking experiment, the video was divided into three parts of about the same length, and each part had a different kind of subtitles. The order was chosen randomly and eventually, part 1 had subtitles by DeepL, part 2 had the human translation, and part 3 had the subtitles by Google Translate. The software VideoPad was used to insert the subtitles and cut the video.

5.3.2 Participants and Apparatus

In total, 5 participants took part in this experiment. They are all aged between 23 and 33. Their first language is German, and they are currently enrolled in a master's programme or have graduated from university recently. They were roughly informed about the purpose of this thesis prior to the experiment. Before the experiment had started, they were asked explicitly for their consent.

The eye-tracking experiment was conducted with the EyeLink Portable Duo eye-tracker. It was used in a remote mode, which means that the participants' heads were not stabilised. Instead, a target sticker was used to show the eye-tracker the position of the eyes.

5.3.3 Procedure

The eye-tracking experiment was conducted on two days and took approximately one hour per participant. It took place at the eye-tracking lab at the Zentrum für Translationswissenschaft in Vienna. The participants were tested individually. At first, the participants were asked to sign a consent form, allowing me to use their data within the scope of this thesis and assuring them that their data is treated confidentially. Afterwards the participants were given the instructions. Before the experiment started, the eye-tracker had to be adjusted and positioned appropriately. Before every video, the eye-tracker was recalibrated to ensure good quality data.

The participants were asked to watch the three parts of the video, and their eye movements were recorded while they were watching it. The sound was turned off because the source language was English, a language which all the participants understood very well. Right after watching each part, they filled out an online questionnaire with five multiple choice

questions on the content and some attitude questions. They were offered a short break before the second and third video.

For the quality analysis of the two sets of subtitles, I familiarised myself with the dimensions and error types of the MQM framework. Then I copied the subtitles of DeepL and Google Translate into two separate Excel files and annotated the errors in the files. I used the official human translation as a gold standard to avoid bias because I was the only annotator. Prior to the annotation process, I checked the human translation for mistakes to make sure that it did not contain any serious mistakes that would render it unsuitable as a gold standard. It contained a very small number of minor issues, such as typing errors, that I corrected (see Annex 1). However, there were no serious mistakes. Errors were annotated by comparing the machine translated versions to the gold standard clause by clause. If there were deviations from the gold standard that I deemed equally good alternatives, I assigned the severity level neutral, which resulted in no penalty points. In some cases, the two MT systems made the same errors or very similar ones, which added an additional form of control and consistency to the annotation of errors.

5.4 Data Analysis

The eye-tracking data was analysed with the tool DataViewer, which is a software that allows users to work with eye-tracking data collected with EyeLink eye-trackers. The data collected from all five participants was imported one after the other in individual projects. In a first step, the image was divided into two AOIs, namely the video area and the subtitle area, which covered a little bit more than the subtitles displayed in order to compensate for minor inaccuracies of the eye-tracker. Subsequently, the exact start and end of the videos was annotated in the recording. These annotations were used to define Interest Periods (IP), which are the parts of the recording that contain relevant eye movements. Afterwards, the videos and reports containing relevant eye movement data were exported. The two reports, namely a Fixation report and an Interest Area report, were both exported for the left and the right eye, resulting in four reports in total for each participant. By checking for which eye the number of recorded fixations was higher, I determined which of the two eyes had better tracking data for each of the participants. Then the data in the reports were used to determine the fixation count, mean fixation duration, and dwell time in the subtitle area and the video area.

For the quality analysis, the errors of each error type and severity level were counted and filled into the MQM scorecard, which calculated the overall quality score for each of the

two machine translations. Additionally, tables showing the number of mistakes and their severity level were created for each dimension.

Turning now to the questionnaire, the comprehension questions were analysed by comparing the total number of correctly answered questions across all participants per part. I also looked at some statistical measures like Hu et al. (2020) did in their study, namely the minimum and the maximum of correct answers per question for each part as well as the mode, mean and standard deviation of correctly answered questions.

The five-point Likert scale statements were analysed by compiling tables containing the absolute frequencies of each value for each statement and the median for each value. In the literature, there is an ongoing controversy on whether Likert scales can only be analysed by the median or whether the mean can also be used (e.g. Carifio & Perla, 2008; Jamieson, 2004). I opted for the median because, in my opinion, the most conclusive argument is that it cannot be assumed that the distances between the values of my Likert scale (stimme überhaupt nicht zu, stimme nicht zu, neutral, stimme zu, stimme voll und ganz zu) are exactly the same, which means that it would not make a lot of sense to calculate the mean. Owing to the fact that the open question was only answered five times (2 answers in the questionnaire for the DeepL version and 3 answers in the one for the Google Translate version), I will discuss each answer individually in the results chapter.

6 Results

This chapter outlines the results of this study. It is divided into three sections, which summarise the main results of the quality analysis, the eye-tracking experiment, and the questionnaire.

6.1 Quality Analysis Results

The quality analysis of the machine translated subtitles revealed that the subtitles by DeepL scored better than the ones by Google Translate, as can be seen in Table 4. The Absolute Penalty Total, which is the sum of all penalty points, is 449 for the translation by Google Translate and 334 for the translation by DeepL, which results in a difference of 115 penalty points. The Per-Word Penalty Total is the Absolute Penalty Total divided by the Evaluation Word Count, which is usually the source text word count (MQM, n. d., Values and Scores). This means that on average, every word is accountable for 0.18% of all errors in the translation by DeepL and 0.24% in the translation by Google Translate. In other words, there is one mistake after every 5.7 words in the DeepL version and one mistake after every 4.24 words in the Google Translate version. The overall quality score is 82.45% for the DeepL version and 76.41% for the Google Translate version. It is calculated by multiplying the Per-Word Penalty Score by 100 and subtracting it from 1. If the target text word count was taken for calculating the Per-Word Penalty Total, the difference would be even bigger as the DeepL translation contains 112 words more than the Google Translate version.

This result partly confirms hypothesis 1a saying that overall, the quality will be acceptable; however, it will be clearly worse than the human translation. Only one MT system, namely DeepL, produced a translation that had an Overall Quality Score of more than 80%, which was determined as threshold for acceptable quality. The Google Translate version went below the 80%-threshold by less than four percent. The second part of this hypothesis can be confirmed as both Overall Quality Scores are clearly lower than 99,58%, which is the Overall Quality Score of the human reference translation.

Table 4: Scores and Totals for the Translations by DeepL and Google Translate

Scores and Totals	DeepL	Google Translate
Total number of words in source text	1,903	1,903
Total number of words in target text	1,941	1,829
Total number of errors	459	498
Absolute Penalty Total	334.00	449.00
Per-Word Penalty Total	0.18%	0.24%
Overall Quality Score	82.45%	76.41%

Errors were found only in four out of the seven dimensions, namely the dimensions Accuracy, Linguistic Conventions, Style, and Design and Markup. As shown in Tables 5 and 6, there were 60 errors in total in the dimension of Accuracy in the DeepL translation and 62 errors in the Google Translate version. The total number of errors in this dimension is nearly the same. The most striking difference, however, lies in the number of major mistakes. The Google Translate version contains 19 major errors, 18 of which are mistranslations, while the DeepL version only contains 4 major errors. This results in a very high difference in penalty points as major errors each account for five penalty points, while minor mistakes only account for one penalty point each. One example of a major mistranslation in the Google Translate version is subtitle 15 where "The buck stops with me" is translated as "Bei mir hört der Bock auf". This is a literal translation of an English saying that does not exist in German. Therefore, viewers will not understand at all what the speaker is saying, which is why this mistake was categorised as major mistake.

The translation by DeepL has more addition and omission errors, namely 20 addition errors and 12 omission errors. Three out of the 12 omission errors are major ones. The Google Translate version only contains three addition errors and two omission errors. Most addition errors consist of words at the end of the first line that are repeated at the beginning of the second line. One example is subtitle 124 of the DeepL version where the word *Städte* is repeated:

123 „Wir haben einen fantastischen Ruf
124 als eine der grünsten *Städte*
Städte in Europa."

Omission errors mostly consist of missing verbs. Subtitle 104 in the DeepL version illustrates this error type clearly:

103 „Sie beabsichtigen, zu 100 Prozent
mit erneuerbaren Energien
104 oder rückgewonnener Wärme bis 2030."

In this sentence, the translation of the verb "powered" is missing. In the remaining error types of the dimension Accuracy (overtranslation, undertranslation, and do not translate), the number of errors is very low and quite similar in both versions.

Table 5: Errors in the Dimension Accuracy made by DeepL

DeepL				
Error Type	Neutral	Minor	Major	Penalty Points
Accuracy	**6**	**50**	**4**	**70**
Mistranslation	3	19	1	24
Overtranslation	1	1	-	1
Undertranslation	1	1	-	1
Addition	-	20	-	20
Omission	1	8	3	23
Do not translate	-	1	-	1

Table 6: Errors in the Dimension Accuracy made by Google Translate

Google Translate				
Error Type	Neutral	Minor	Major	Penalty Points
Accuracy	**13**	**30**	**19**	**125**
Mistranslation	8	26	18	116
Overtranslation	1	-	-	0
Undertranslation	3	-	-	0
Addition	-	3	-	3
Omission	1	1	-	1
Do not translate	-	-	1	5

It can be seen from the data in Table 7 and 8 that there are 130 errors in total in the dimension of Linguistic Conventions in the DeepL version and 199 errors in the Google Translate version. Contrary to the previous dimension, the total number of errors differs greatly. Also the total number of penalty points differs greatly, with Google Translate receiving 68 penalty points

more than DeepL. In terms of the number of errors, the biggest difference lies in the minor severity level with a deviation of 68 errors. What stands out in this dimension is that Google Translate made a very large number of grammar mistakes. A large portion of these 128 grammar errors occurred due to the wrong positioning of the verb. This is, for example, the case in subtitle 17 where the verb *habe* should be at the very end of the sentence because, in general, the conjugated verb must be at the end of the clause in subordinate clauses:

16 Und das ist manchmal sogar für Probleme
17 über die ich sehr wenig *habe*
echte Kontrolle oder Macht.

This issue also happened in the DeepL version but not nearly as often as in the Google Translate version. Other types of grammar errors included the use of wrong prepositions, errors in capitalisation, and agreement of adjectives or verbs.

In terms of punctuation errors, DeepL made slightly more punctuation errors than Google Translate. Punctuation errors usually occurred before conjunctions, such as *because*, or before relative clauses where a comma is always needed in German, while no comma is used in English. This can be seen in the previous example where *über die* introduces a relative clause and should be preceded by a comma.

Table 7: Errors in the Dimension Linguistic Conventions made by DeepL

DeepL				
Error Type	Neutral	Minor	Major	Penalty Points
Linguistic Conventions	**21**	**108**	**1**	**113**
Grammar	13	51	1	56
Punctuation	8	57	-	57

Table 8: Errors in the Dimension Linguistic Conventions made by Google Translate

Google Translate				
Error Type	Neutral	Minor	Major	Penalty Points
Linguistic Conventions	**22**	**176**	**1**	**181**
Grammar	13	128	1	133
Punctuation	9	48	-	48

Turning now to the dimension of style, it can be seen in Tables 9 and 10 that DeepL made more errors in this dimension, even though the difference is not big. However, DeepL made more neutral errors, which do not result in penalty points. Therefore, the Google Translate version receives more penalty points although there are fewer mistakes in total. Both translations contain so many neutral mistakes because clauses that differed from the human translation but were still grammatically correct and accurately translated were categorised as either awkward or unidiomatic style. As long as the variations were deemed equally good as the human translation, the errors were assigned the neutral severity level.

Table 9: Errors in the Dimension Style made by DeepL

DeepL				
Error Type	Neutral	Minor	Major	Penalty Points
Style	**117**	**64**	**-**	**64**
Register	-	2	-	2
Awkward Style	63	46	-	46
Unidiomatic Style	54	16	-	16

Table 10: Errors in the Dimension Style made by Google Translate

Google Translate				
Error Type	Neutral	Minor	Major	Penalty Points
Style	**97**	**67**	**1**	**72**
Register	1	1	-	1
Awkward Style	51	45	1	50
Unidiomatic Style	45	21	-	21

Comparing the results of DeepL and Google Translate in the dimension of Design and Markup, it can be seen that DeepL made more mistakes than Google Translate. DeepL also scored lower in this dimension with a difference of 16 penalty points. As shown in Tables 11 and 12, both have only one error in Character Encoding, but in the DeepL version there are more lines that contain too many cpl. Interestingly, the Google Translate version does not contain a single subtitle with a too high subtitle speed. The maximum subtitle speed in the Google Translate version is 21 cps, while it is up to 25 cps in the DeepL version. What stands out is that this is the only dimension without any major mistakes.

Table 11: Errors in the Dimension Design and Markup made by DeepL

DeepL				
Error Type	Neutral	Minor	Major	Penalty Points
Design and Markup	**1**	**87**	**-**	**87**
Character Encoding	1	-	-	0
Too many cpl	-	82	-	82
Too many cps	-	5	-	5

Table 12: Errors in the Dimension Design and Markup made by Google Translate

Google Translate				
Error Type	Neutral	Minor	Major	Penalty Points
Design and Markup	**1**	**71**	**-**	**71**
Character Encoding	1	-	-	0
Too many cpl	-	71	-	71
Too many cps	-	-	-	0

Hypotheses 1b and 1c, which refer to the second part of Research Question 1 can be confirmed by the results of the quality analysis. As far as hypothesis 1b, which says that both MT systems will struggle to adhere to the subtitle characteristics related to time and space, is concerned, the results show that both systems did struggle and received a high number of penalty points for this kind of mistake. However, DeepL produced more subtitles that contained too many cpl according to the TED subtitling rules. Moreover, too many cps were found only in the DeepL version. This means that time and space restrictions were a bigger problem for DeepL than they were for Google Translate. The quality analysis of both MT systems can be found in Annex 2 and 3; the MQM scorecards for all three translations can be found in Annex 4, 5, and 6.

Hypothesis 1c states that DeepL and Google Translate will have different strengths and weaknesses and perform differently in the different dimensions and error types. Overall, DeepL outperformed Google Translate by far in the dimensions of Accuracy and Linguistic Conventions both in terms of number of errors and penalty points. However, there were still some error types in which Google Translate performed better, including addition, omission, and punctuation. Google Translate outperformed DeepL in the dimension of Design and Markup as already outlined in the previous paragraph. However, the difference in penalty points is not as high as in the dimensions of Accuracy and Linguistic Conventions. In the dimension of Style, Google Translate made fewer mistakes than DeepL; however, DeepL made more neutral mistakes and fewer major mistakes, resulting in a lower number of penalty points for DeepL.

6.2 Questionnaire Results

This section reports the results of the questionnaires. The three questionnaires can be found in Annex 7. In each of the questionnaires, the first five questions aimed at testing the participants'

comprehension of the three videos. Tables 13, 14, and 15 show the results of these questions. Each questionnaire contained five multiple choice questions on the content with one correct answer per question. As there were five participants that each answered all five questions, the maximum of correctly answered questions possible per part is 25. Tables 13 and 14 show that the total number of correctly answered questions in the human translation and the DeepL version is exactly the same. On the contrary, the participants chose fewer correct answers after watching the video with the Google Translate subtitles as shown in Table 15. In other words, the participants got 88% of questions right in Part 1 and Part 2, while they got 72% of questions right in Part 3, resulting in a difference of 16%.

Table 13: Number of Correct Answers to Questions 1-5 in Questionnaire 1

Questions in Part 1 (DeepL)	**Correct answers (highest possible per question: 5)**
Question 1	4
Question 2	5
Question 3	5
Question 4	3
Question 5	5
Total	22

Table 14: Number of Correct Answers to Questions 1-5 in Questionnaire 2

Questions in Part 2 (human)	**Correct answers (highest possible per question: 5)**
Question 1	5
Question 2	5
Question 3	3
Question 4	5
Question 5	4
Total	22

Table 15: Number of Correct Answers to Questions 1-5 in Questionnaire 3

Questions in Part 3 (Google Translate)	**Correct answers (highest possible per question: 5)**
Question 1	5
Question 2	3
Question 3	3
Question 4	2
Question 5	5
Total	18

The statistical data in Table 16 also shows that the participants scored worse in Part 3, which had the subtitles by Google Translate. While the maximum number of correct answers per question is 5 in all three parts, the minimum number of correct answers per question differs. It is 3 in Part 1 and 2 but only comes to 2 in Part 3. The mode in Part 3 is 3 and 5 as opposed to only 5 in Parts 1 and 2, also indicating that participants scored lower in Part 3. With a value of 3.6, the mean is also lower in Part 3 than in Parts 1 and 2 where it is 4.4. The standard deviation, which indicates how far each value lies from the mean on average, is higher in Part 3. This means that the number of correct answers per question is on average farer away from the mean in Part 3 than in Parts 1 and 2.

Table 16: Statistical Data on Answers to Questions 1-5

	Min	Max	Mode	Mean	SD
Part 1 (DeepL)	3	5	5	4,4	0,89442719
Part 2 (Human)	3	5	5	4,4	0,89442719
Part 3 (Google Translate)	2	5	3 and 5	3,6	1,34164079

Hypothesis 1d, which states that the quality of the machine translated subtitles will only have a very slim effect on comprehension scores, can be confirmed partly. In the case of DeepL, the difference in quality compared to the human translation did not result in any difference in comprehension scores. In the case of Google Translate, the participants scored 16% worse which can be considered a high difference, especially since the results were equal for the DeepL translation and the human translation.

The next section of the survey dealt with the participants' attitude towards the subtitles of each part. As shown in Tables 17, 18, and 19, the results of the Likert scale show clear differences in attitude. The human translation was ranked the highest and was clearly perceived as a high-quality translation by everyone. All participants chose either *stimme zu* (agree) or *stimme voll und ganz zu* (fully agree) for all five statements, resulting in *stimme zu* being selected 11 times and *stimme voll und ganz zu* being selected 14 times. The other three options were not chosen at all. Therefore, the median is high with a value of 5.

The subtitles translated by DeepL scored second-best. Every option was selected at least once. However, the options in the middle of the spectrum (2-4) were selected the most, while the extremes were only chosen very few times. This resulted in a median of 3, which means that the translation must have been perceived as moderate. What stands out is that statement 1, which asked the participants how well they think they had understood the content of the video, has the highest median. All other statements that were more about the effort involved in reading and the participants' contentment scored worse. All in all, the participants seem to think that they understood the content well, which is true as can be seen in the results of the comprehension questions, but they still did not enjoy reading them.

The Google Translate version received the worst score. Interestingly, no one chose *stimme voll und ganz zu*. The majority of participants chose *stimme überhaupt nicht zu* (do not agree at all) and *stimme nicht zu* (do not agree), showing that the attitude towards these subtitles is clearly worse than towards the other two translations. With a value of 2, the median is lower than the one for the human and the DeepL translation. For three statements, the median is even as low as 1, which further highlights the negative attitude towards this set of subtitles.

Table 17: Results of the Likert Scale for Questionnaire 1 (DeepL)

Statements	**1**	**2**	**3**	**4**	**5**	**Median**
(1) Ich habe das Gefühl, den Inhalt des Videos im Großen und Ganzen verstanden zu haben.	0	0	2	3	0	4
(2) Die Untertitel sind leicht zu verstehen.	0	2	1	1	1	3
(3) Es erfordert keine große Anstrengung die Untertitel zu lesen.	0	2	2	1	0	3
(4) Die Untertitel waren angenehm zu lesen.	0	2	2	0	1	3
(5) Ich bin mit der Qualität der Untertitel zufrieden.	1	2	0	1	1	2
Total	1	8	7	6	3	**3**
1 = stimme überhaupt nicht zu, 2 = stimme nicht zu, 3 = neutral, 4 = stimme zu, 5 = stimme voll und ganz zu						

Table 18: Results of the Likert Scale for Questionnaire 2 (human Translation)

Statements	**1**	**2**	**3**	**4**	**5**	**Median**
(1) Ich habe das Gefühl, den Inhalt des Videos im Großen und Ganzen verstanden zu haben.	0	0	0	4	1	4
(2) Die Untertitel sind leicht zu verstehen.	0	0	0	2	3	5
(3) Es erfordert keine große Anstrengung die Untertitel zu lesen.	0	0	0	2	3	5
(4) Die Untertitel waren angenehm zu lesen.	0	0	0	2	3	5
(5) Ich bin mit der Qualität der Untertitel zufrieden.	0	0	0	1	4	5
Total	0	0	0	11	14	**5**
1 = stimme überhaupt nicht zu, 2 = stimme nicht zu, 3 = neutral, 4 = stimme zu, 5 = stimme voll und ganz zu						

Table 19: Results of the Likert Scale for Questionnaire 3 (Google Translate)

Statements	**1**	**2**	**3**	**4**	**5**	**Median**
(1) Ich habe das Gefühl, den Inhalt des Videos im Großen und Ganzen verstanden zu haben.	0	4	0	1	0	2
(2) Die Untertitel sind leicht zu verstehen.	3	1	0	1	0	1
(3) Es erfordert keine große Anstrengung die Untertitel zu lesen.	2	1	1	1	0	2
(4) Die Untertitel waren angenehm zu lesen.	4	0	1	0	0	1
(5) Ich bin mit der Qualität der Untertitel zufrieden.	3	1	0	1	0	1
Total	12	7	2	4	0	**2**
1 = stimme überhaupt nicht zu, 2 = stimme nicht zu, 3 = neutral, 4 = stimme zu, 5 = stimme voll und ganz zu						

In the final part of the questionnaires, the respondents were asked if there was anything left that they would like to comment on regarding their attitude on each set of subtitles. This question was answered by two participants in the first questionnaire and by three participants in the third questionnaire. It was not answered by any participant in the second questionnaire.

The answers received in the first questionnaire are listed in Table 20. One participant noted that they could understand the content on the whole, but they did not catch all the details because they did not understand some sentences. They also missed some information when it took them too long to decipher certain mistakes. This answer is in line with results of the Likert scale where the participants stated that they understood the content quite well, but they still did not enjoy reading the subtitles a great deal. Another participant commented that they noticed many grammar mistakes and incomplete sentences. This perception can be confirmed by the results of the quality analysis where many grammar mistakes and some omissions were found.

Table 20: Answers to the Open Question in Questionnaire 1

	Answers Questionnaire 1
P01	Im Großen und Ganzen kann man dem Inhalt zwar folgen, aber manche Teile bzw. Details bekommt man nicht so gut mit, wenn man an Fehlern in den Untertiteln hängen bleibt oder einzelne Sätze teilweise schwer verständlich sind.
P03	Sehr viele Grammatikfehler, unvollständige Sätze

The comments on the Google Translate version are clearly more negative, as can be seen in Table 21. One participant said that it took them a great deal of attention and energy to understand the subtitles. As a result, they had the feeling that their understanding of the content was very superficial. Even if the results of the comprehension questionnaire are the worst for Part 3, they are still not that bad. It seems that this participant thought that they would score worse than they actually did. Another participant commented that many subtitles did not make sense. Again, these statements are in line with the results of the Likert scale where most participants expressed a rather negative attitude towards this set of subtitles. The last comment in questionnaire three focused more on the language than on the reading experience itself. It says that the sentence structure is absurd, some translations are too literal, there are repetitions, and there are many grammar mistakes.

Table 21: Answers to the Open Question in Questionnaire 3

	Answers Questionnaire 3
P01	Dadurch, dass die Sätze teilweise so verdreht waren, hat man viel Aufmerksamkeit und Energie gebraucht, um sie richtig zu verstehen. Ich habe das Gefühl, nur recht Grob verstanden zu haben, um was es geht.
P03	Völlig absurde Satzstruktur, wahrscheinlich zu wörtliche Übersetzung von Eigennamen, Wiederholungen, Grammatikfehler,...
P04	Haben häufig keinen Sinn ergeben

All in all, the comments in questionnaire three showed more negative feelings towards the subtitles than the comments in questionnaire one. Moreover, more participants felt the need to comment something in the third questionnaire.

The results of the second part of the questionnaire fully confirm hypothesis 2b, which says that the lower the quality of the subtitles the lower the participants' attitude towards them. The results of both the Likert scale and the open question rank the three translations in the same order as the quality analysis, with the human translation being ranked the highest and Google Translate the lowest.

6.3 Eye-Tracking Results

The eye-tracking experiment aimed at collecting data on the cognitive load of the participants while they were watching the three videos. Three parameters, namely fixation count, fixation duration, and time spent in the subtitle AOI, were collected.

Table 22 shows the fixation count in the subtitle AOI, i.e. the total number of fixations made in the subtitle AOI. It was adapted to a reference time of five minutes to make the values comparable as the lengths of the three videos vary slightly. The colours represent the order of the fixation count for each participant, with green representing the lowest fixation count, yellow the medium fixation count, and red the highest fixation count. This table shows that the majority of participants (4/5) had the lowest number of fixations when watching the video with the human translation. Only one participant had the lowest fixation count in the DeepL version, and no one had their lowest fixation count in the Google Translate version.

The red cells in this table show that three out of five participants had the highest number of fixations in Part 3. Two participants had their highest number of fixations when watching Part 1, leaving Part 2 with no red cell at all. What stands out here, is that the highest fixation counts per participant can only be found in the machine translated versions. This indicates that the machine translated subtitles induced a higher cognitive load than the human translation. The subtitles by Google Translate seem to have caused a higher cognitive load than the DeepL version. These results are in line with the results of the Likert scale where the participants stated that they enjoyed the subtitles by Google Translate the least and the human translation the most. However, it should also be acknowledged that the differences between the fixation counts are not very big. The maximum differences between green and yellow cells as well as yellow and red cells are around 10%.

Table 22: Fixation Count in the Subtitle Area of Interest (per 5 minutes)

	P01	P02	P03	P04	P05
DeepL	886	678	693	730	780
Human	785	646	589	659	809
Google Translate	847	734	604	800	880

I created the same kind of table to illustrate the mean fixation duration in the subtitle AOI, i.e. the sum of the durations of all fixations in the subtitle AOI divided by the total number of fixations in the subtitle AOI. Contrary to the fixation count, the results of the mean fixation duration shown in Table 23 do not indicate the lowest cognitive load in Part 2. In four out of the five participants, the highest mean fixation duration occurs in Part 2, i.e. the human translation, indicating that the majority of participants experienced the highest cognitive load in Part 2. This means that Table 22 and 23 show conflicting results.

Another striking result is that four out of five participants show exactly the same pattern of mean fixation duration distribution. For each of them, the lowest mean fixation duration occurred when watching the subtitles translated by DeepL, and the highest one when watching the human translation. Surprisingly, only P04 shows the same pattern as in the previous table.

Table 23: Mean Fixation Duration in the Subtitle Area of Interest (in Milliseconds)

	P01	P02	P03	P04	P05
DeepL	275,43	208,12	307,43	269,96	219,53
Human	298,56	230,96	349,49	261,39	279,8
Google Translate	298,08	228,37	339,53	273,14	250,33

The next two tables show the dwell time in percent in the subtitle AOI and in the video AOI, i.e. how much time the participants spent in the subtitle AOI and in the video AOI. Table 24 shows that three participants spent the shortest amount of time reading the subtitles in the human translation and also three participants spent the most time reading the ones translated by Google Translate. This result is in line with the result of the fixation count as a shorter dwell time indicates less cognitive load. What stands out as well is that none of the participants spent the shortest amount of time reading the subtitles by Google Translate; however, one participant's highest dwell time is in Part 2.

Another interesting result is how far apart the dwell times for some of the participants are. P01 and P02 have very high and very low dwell times compared to the other three participants. P01 spent over 90% of the time reading the subtitles translated by DeepL and Google Translate and around 87% to read the human translation, while P02 spent less than 60%

in the subtitle AOI of Part 1 and 2, and around 64% in Part 3. This means that the differences in dwell time for P01 and P02 are always around 30%.

When looking at Table 25, it can be seen that four participants spent the least amount of time in the video AOI in Part 3. Moreover, four participants spent the most time watching the image of Part 2, indicating like Table 22 that the cognitive load was the highest in Part 3 and the lowest in Part 2.

Table 24: Dwell Time in Subtitle Area of Interest (in Percent)

	P01	P02	P03	P04	P05
DeepL	92,23%	56,29%	79,48%	77,85%	68,02%
Human	87,28%	57,01%	75,34%	67,09%	87,82%
Google Translate	93,99%	64,31%	75,87%	86,37%	86,8%

Table 25: Dwell Time in Video Area of Interest (in Percent)

	P01	P02	P03	P04	P05
DeepL	7,72%	35,74%	20,18%	19,19%	30,92%
Human	12,54%	42,39%	24,64%	32,84%	7,68%
Google Translate	5,46%	32,21%	22,89%	12,30%	6,31%

Hypothesis 2a, stating that if the quality of the subtitles is lower, the eye-tracking parameters fixation count, mean fixation duration, and dwell time will indicate a higher cognitive load, can be confirmed partly. Two out of three eye-tracking parameters (fixation count and dwell time) confirm this hypothesis and suggest the lowest cognitive load for reading the human translation and the highest cognitive load for reading the Google Translate version.

Finally, this section will provide a short analysis of some subtitle parameters to see whether they influenced the eye-tracking results as well. As can be seen in Table 26, the DeepL version contains the highest number of words. Although it is also the longest video, the number of words per second is still the highest, meaning that the number of words is proportionally higher compared to the other two videos. The version containing the human translation contains the least number of words in absolute numbers and in proportion to the length of the video. The DeepL version, however, consists of a higher percentage of one liners than the other two versions which might compensate for the higher density of words. In the Google Translate version and the human translation, the percentage of one and two liners is nearly the same. As these parameters of the human translation and the translation by Google Translate are very similar, I don’t think that they have had an impact on the results of the experiment. The higher proportion of one liners in the DeepL version might have helped the viewers to read them more

comfortably and, therefore, might be an explanation for the good performance of the DeepL version in some eye-tracking parameters and participants.

Table 26: Subtitle Analysis

	DeepL	Human	Google Translate
Total number of words	663	505	620
Length of the video	4 min, 47 s	4 min, 33 s	4 min, 43 s
Words per second	2.31	1.85	2.19
Total number of subtitles	92	79	78
Total number of one liners (percentage of one liners)	52 (57%)	35 (44%)	33 (42%)
Total number of two liners (percentage of two liners)	40 (43%)	44 (56%)	45 (58%)

7 Discussion

This chapter will delve further into the meaning of the results described in the previous chapter. The first section provides a short summary of the results. Subsequently, the results are dealt with in more detail by discussing their implications and relating them to each other. Unexpected results will also be examined more closely in a separate section. Another section is dedicated to discussing relations of the present results and methods to the ones of previous studies. Finally, the last section will briefly describe the limitations of this study and give recommendations for future related work.

7.1 Summary of the Results

This section summarises the above-mentioned results. Overall, the experiment shows that the human translation scored best in most of the three parts of this experiment. It contained only a few minor mistakes and served as a gold standard for the quality analysis. In terms of overall quality, the DeepL version achieved 82.45% and the Google Translate version received a score of 76.41%, meaning that DeepL scored better with a difference of 6.04 percentage points. Although DeepL made more mistakes in half of the dimensions used (Style and Design and Markup), Google Translate received more penalty points in three dimensions because it made much more severe errors, and DeepL made more neutral errors.

The comprehension part of the questionnaire showed that considering all participants, there was no difference in comprehension between the human translation and the translation by DeepL. In both versions, all participants taken together answered 22 out of 25 questions correctly, while they chose only 18 correct answers after reading the subtitles by Google Translate. Even though there was no difference in comprehension between the subtitles translated by a human and by DeepL, the second part of the questionnaire revealed a clear difference in attitudes towards the different kinds of subtitles. The median of all Likert scale statements is two for the subtitles by Google Translate, three for the ones by DeepL and five for the human translation. The open question at the end of the questionnaire was answered by two participants for the DeepL version and by three participants for the Google Translate version. All five answers expressed negative feelings towards the subtitles, which means that Google Translate received the most negative rating again, while the human version was ranked best and the DeepL version second-best.

The eye-tracking parameters suggest that the cognitive load is the lowest for most participants when watching the video with the human translated subtitles and the highest when watching the Google Translate version.

In total, these results provide interesting insights into the quality and reception of machine translated subtitles, which will be interpreted in more detail in the next section.

7.2 Interpretation of the Results

In this section, the results of this experiment will be interpreted by providing possible explanations for the results and relating them to each other. This is a very small study, which is why the results can only provide a small insight into the quality and reception of machine translated subtitles in the language combination English-German. However, more and bigger studies similar to this one could show the state of the art and help evaluate which tools are best used for which purpose.

The results show that DeepL performed better than Google Translate. The quality of the DeepL version is only a little bit better; however, it seems to be enough to enable a better understanding of the content and to induce a better attitude towards the subtitles. Although Google Translate scored slightly better in the error types that were explicitly related to subtitling characteristics, namely *too many cpl* and *too many cps*, which would mean that it might be able to handle subtitles better than DeepL, Google Translate received a great number more penalty points in most of the other dimensions. For this reason, DeepL produced a translation of a higher overall quality despite struggling more with the maximum number of cpl and cps.

A possible explanation for the difference in comprehension scores might be the higher number of major errors in the dimension of *Accuracy* in the Google Translate version. Google Translate made 19 major mistranslation errors, while DeepL only made four major mistranslation errors. However, it should be noted that not all 19 errors were made in the part that was shown to the participants. As mistranslations contain wrong information, it might have contributed to comprehension difficulties.

Another striking difference between the two MT systems is the number of grammar errors they made. DeepL made 51 minor grammar errors, and Google Translate made 128, more than double the amount that DeepL made. This immense difference may have influenced comprehension as well because the additional errors in the Google Translate version might have impaired the participants' reading process. At least, some participants made such comments in the attitude questionnaire. It should be noted, however, that comments like this were made for both machine translations, but the comments for Google Translate were more negative and

higher in number. One participant noticed grammar mistakes and incomplete sentences, which probably refers to omission errors, in the DeepL version. In the Google Translate version, the same participant noticed even more types of mistakes, including sentence structure, literal translations, repetitions, and grammar mistakes. This indicates that the participants were aware of the many mistakes in both translations, and they also noticed that the quality of the Google Translate version was worse than that of the DeepL version.

One of the five Likert scale statements asked the participants about their opinion on the quality of the subtitles (5) and one for their self-perceived comprehension (1) of the subtitles. Interestingly, the results of these two Likert scale statements are in line with the quality analysis and the comprehension questions. The median for Likert scale statement 1 is 4 for the DeepL version and for the human translation. The median for both versions is quite high and the same, just as the comprehension score of 22/25. However, the distribution of the individual values is different, with the values for the DeepL version (3,3,4,4,4) being lower than the ones for the human translation (4,4,4,4,5). This shows that more participants must have been less confident about how much they understood when watching the video with subtitles by DeepL, as also shown in the answers to the open questions discussed before.

Statement 5 reflects the results of the quality analysis. The median is 2 for the DeepL version, 5 for the human translation, and 1 for the Google Translate version. Similarly to the quality analysis, the human translation received the best rating by far. Also, the difference in perceived quality between the translations by DeepL and Google Translate is not very big, just as it was not so big in the quality analysis.

When looking at the other three Likert scale statements that were asking the participants for their attitude towards the reading process, the same ranking as in the quality analysis can be observed. All three statements for the human translation have a median of 5, while the DeepL version has a median of 3 and the Google Translate version has a median of 1 for two statements and a median of 2 for one statement. This means that the attitude towards the subtitles is in line with the quality of the subtitles, indicating that viewers notice quality differences and are more content with good-quality subtitles.

The eye-tracking parameters generally confirm the findings of the Likert scale. The parameters fixation count and dwell time imply that the cognitive load must have been the lowest for most participants when watching Part 2 with the subtitles translated by a human, the second-lowest for Part 1 containing the DeepL version, and the highest for Part 3 containing the subtitles by Google Translate.

The results suggest that the machine translated subtitles contain many issues that should be corrected when the subtitles ought to be of good quality and easy to read. However, despite all the mistakes and the higher cognitive load induced in comparison to the human translation, the machine translated subtitles still enabled the participants to understand the gist of the videos. Overall, DeepL performed better than Google Translate in this experiment. However, the results of such a small experiment cannot provide enough evidence to generally recommend DeepL rather than Google Translate for the translation of subtitles. In future studies, it would be interesting to find out whether DeepL generally delivers higher quality translations of subtitles in this language combination or if results might be different for other domains.

7.3 Unexpected Results

In this section, I will discuss some unexpected results. Generally, all hypotheses were confirmed at least partly. Accordingly, most results were expected although there were some surprising elements. For example, according to hypothesis 1a, both MT systems should have produced acceptable translations which were defined as translations with an overall quality score above 80%. (see section 5.1). However, the Google Translate version had a score of around 76%, which is lower than 80%; but it did not fall below dramatically.

Another unexpected finding was the extent to which the dwell time in the subtitle AOI varied among the participants. P01 had by far the highest dwell times with values between 87.28% and 93.99%, while P02 had the lowest dwell times with values between 56.29% and 64.31%. I did not expect to see such big differences. Therefore, it would have been a good idea to test the individual participants' reading speeds before the experiment to see if they vary. Participants might use subtitles to varying extents in their daily lives. Therefore, these high differences in dwell time might be caused by differences in the consumption of subtitles. P02 might read subtitles more often than P01 and have thus gained more practice, which could be an explanation for the faster reading. Another reason for the high difference might be that P01 put more effort into understanding and remembering as much as possible of the subtitles because they wanted to score better in the questionnaire.

7.4 Relations to Previous Work

This section compares the results of this study to the ones of similar previously conducted studies. It will primarily focus on the study by Hu et al. as it revolved around a very similar research question. However, the employed research methods will also be discussed with regard

to other subtitling studies that focused on different aspects of subtitles but used a similar research design.

7.4.1 Quality Analysis

This part focuses on the quality analysis and whether the results confirm or rebut the results of previous studies. Similarly to Xie (2022), this study used the MQM framework to assess the quality of the subtitles. However, Xie used a different set of dimensions and error types as well as a different language combination (English-Simplified Chinese), which makes it difficult to compare the results. Moreover, the author unfortunately does not discuss the quality analysis results in very much detail. However, it seems that the subtitles on Youtube, which are generated by Google Translate, performed better in the dimensions of *Accuracy* and *Fluency* (called *Linguistic Conventions* in this study) than the ones created by Bilibili. In this current study, Google Translate does not do that well compared to DeepL, especially in those two dimensions. There are a few possible explanations for these differences in performance. Firstly, Google Translate might perform better in the language combination English-Simplified Chinese than in the language combination English-German. Secondly, Bilibili might perform worse than Google Translate because it has only been released quite recently and therefore has not yet been trained as much as Google Translate or DeepL. Moreover, the domain of the videos might also have had an impact on Google Translate's performances. In terms of errors in the dimension of *Style*, no comparison can be made between the two studies because Xie does not provide any details of errors made in this dimension. In the present study, the dimension of *Style* accounts for 72 out of 449 penalty points (16%) in the Google Translate version and 64 out of 334 penalty points (19%) in the DeepL version, which constitutes a large part of penalty points. Therefore, it would have been interesting to also see a comparison between Youtube and Bilibili in this regard.

7.4.2 Eye-Tracking

Eye-tracking studies in the field of subtitling seem to have become more and more popular over the last couple of years. In many of these studies, researchers change one subtitling parameter, such as segment length or number of lines, and try to find out whether these changes have effects on certain eye movements, which would allow them to draw conclusions on the cognitive load induced by the change in subtitle presentation. Interestingly, I could not find any previous work that dealt with the quality of machine translated subtitles and the cognitive load

induced by them in the language combination English-German. However, there is a great number of subtitling studies that use very similar methods.

When reviewing the existing literature on cognitive load research in subtitling, it was astonishing to see how differently eye-tracking has been implemented in different studies. For example, some studies used parameters that referred to each individual subtitle, some studies used parameters that referred to the video as a whole, and some studies used both. Szarkowska and Gerber-Morón (2019, p. 151), are researchers who used both, for example, the parameter proportional reading time, which they defined as "The percentage of time that a participant spent in the AOI as a function of subtitle display time, where 100% is the total subtitle display time" (Szarkowska & Gerber-Morón, 2019, p. 151). In other words, they calculated the proportional reading time by dividing the actual time a participant spent in a subtitle by the total display time of the subtitle, meaning that they looked at each subtitle individually. Hu et al. (2020, p. 524), in turn, used parameters that only require to divide the video into an image AOI and a subtitle AOI as well as to define one Interest Period. For example, they used the parameter glance count, which they defined as the number of times that a participant moves their eyes from the image AOI into the subtitle AOI. In this case, it does not matter how often the viewers revisit each individual subtitle; it only matters how often they revisit the subtitle AOI. In addition, researchers use very diverse sets of eye-tracking parameters. Usually, three or four parameters are analysed and interpreted to gain insights on cognitive processes. In this research area, there seems to be no consent yet on which eye-tracking parameters or combination of parameters are most suitable for assessing the cognitive load in viewers while they are watching a subtitled video. This can make it difficult to compare the results of studies. For example, Hu et al. (2020), who compared machine translated subtitles to post-edited and human translated subtitles, used the parameters glance count and fixation count, whereas the present study used fixation count, fixation duration, and dwell time. Apparently, more research on how eye-tracking parameters are best combined is still needed. Doherty and Kruger (2018), whose research I used a basis for deciding which parameters to use, made a step in this direction by reviewing which parameters have been used frequently and which ones have not been commonly used in subtitling studies.

7.4.3 Comprehension Questions

Comprehension questions are frequently used alongside eye-tracking to assess the cognitive load from different perspectives or alongside a quality analysis. This section discusses how it

was used in the present experiment and how it has been used in previous work as well as how researchers have interpreted it.

In this experiment and many related studies (Hu et al., 2020; Xie, 2022) that I have read, comprehension was tested by using multiple choice questions or similar forms of questions, such as true/false questions (Szarkowska & Gerber-Morón, 2019). The usage of these types of questions is in line with research on reading comprehension testing. According to Day and Park (2005, pp. 65–66), it is recommended to use multiple choice questions or true/false questions to test reading comprehension in situations where the readers do not engage further with the source material. Therefore, the answers should be directly and explicitly in the text. In the case of post-task comprehension testing, the situation is of course a little bit different because the readers have to answer the questions from what they remember after reading the text once because they cannot go back to the text. Therefore, I think that these types of questions are appropriate for this specific situation because it might be easier for participants to recall what they have read if they see it again in one of the answer options. However, I also think that it depends on how the individual questions are worded and how much level of detail they are asking for.

During my research and while analysing the results of my own experiment, it became apparent that comprehension is a robust indicator that is not influenced easily in most cases. One example where comprehension did not show a significant difference in different subtitling conditions is the study conducted by Szarkowska and Gerber-Morón (2019) where they wanted to find out whether two or three lines per subtitle are easier to process. In terms of the susceptibility of comprehension to changes in the conditions of subtitles, the two authors made the same observation as I did during my research. They hypothesise that the reason might be that viewers have been previously exposed to both subtitles of high quality and subtitles of low quality, which allowed them to learn how to process different kinds of subtitles efficiently (Szarkowska & Gerber-Morón, 2019, pp. 158–159). In my opinion this assumption is plausible because people are exposed increasingly to subtitles and MT, especially in the context of social media and video platforms. It seems like the different subtitling conditions that are compared, e.g. number of lines, MT system, or characters per line, are too similar to reduce the level of comprehension although at the same time, they might increase the cognitive load and are perceived as being of worse quality.

One study that saw a difference in comprehension is the one conducted by Chan et al. (2022). The reason for this difference might be that their aim was to find out what effects subtitles in different languages as well as the absence of subtitles had on the learning outcome

of the participants. In contrast to the other studies that I dealt with in the literature review, in this study, the difference between the different kinds of subtitles is not a formal parameter, the number of mistakes, or stylistic issues but the language itself, which constitutes a completely different baseline.

7.4.4 Attitude Questions

It seems to be very common in cognitive load studies to ask the participants explicitly for their self-perceived evaluation of either their cognitive load or their opinion towards the subtitles. This is often done by using a Likert scale. However, it varies how exactly researchers employ them in their studies. This section will discuss how the results of the present study relate to the results of the study by Hu et al. (2020).

This study drew on the attitude questions in the form of a Likert scale used by Hu et al. (2020). In their study, in which they compared human translated subtitles, raw machine translated subtitles, and post-edited subtitles, all three kinds of subtitles received a fairly high attitude rating by the participants as can be seen in Figure 5. More participants even chose "agree" or "strongly agree" for the post-edited subtitles than for the human translation.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Group PE (24)	33.04% (114)	49.28% (170)	13.04% (45)	4.35% (15)	0.29% (1)
Group RAW (22)	17.21% (53)	57.79% (178)	18.51% (57)	6.49% (20)	0
Group HT (15)	27.05% (56)	39.61% (82)	24.16% (50)	9.18% (19)	0

Note: One participant in Group HT failed to answer Statements 25, 26, 27, and 28. Hence, the number of respondents for the four questions in this group is 14.

Figure 5: Results of Attitude Questions in Hu et al. (2020, p. 528)

In the present study, however, the participants' attitude towards machine translated subtitles strongly differed from those results. To make the results easier to compare, I calculated the percentages for each Likert scale value for each video like Hu et al. did and compiled them in Table 27. It is surprising how differently viewers evaluated the machine translated and human translated subtitles in the two studies, especially when looking at the Google Translate version. This version received "agree" or "strongly agree" in 16% of the votes in this experiment, whereas the raw machine translation in the study by Hu et al. received a total of 75% in those two values. The rating of the human translation is also very different in these two studies. While the human translated subtitles were rated positively (strongly agree or agree) by all participants, only 66.66% chose the first two options in Hu et al.'s study. To summarise, the three different kinds of subtitles in Hu et al. are rated much more similarly than in the present study.

Table 27: Results of Attitude Questions in Percent

	Stimme voll und ganz zu (strongly agree)	Stimme zu (agree)	Neutral (neutral)	Stimme nicht zu (disagree)	Stimme überhaupt nicht zu (strongly disagree)
DeepL	12%	24%	28%	32%	4%
Human	56%	44%	0%	0%	0%
Google Translate	0%	16%	8%	28%	48%

Unfortunately, the authors did not do a quality analysis in their study, which could have shown whether the machine translated subtitles and the human translated subtitles in their study were of similar quality as in the present paper. The reason for the strong deviation in attitude ratings might be differences in quality, especially since the human translation in Hu et al. was not done by a professional translator. Therefore, it is difficult to know whether it adheres to professional standards. Even though the official TED subtitles are not necessarily done by professionals either, there are still steps that are taken to ensure a certain quality standard. Other possible reasons for the different attitude ratings could be cultural differences of the viewers or different levels of previous exposure to subtitles.

Attitude questions or a combination of self-reported cognitive load and some sort of attitude questions in the form of Likert scales have been used by all studies that I have discussed in chapter 4.3.2. The exact design differed slightly, but it seems to be considered important to not only measure the performance of MT objectively but also incorporate the viewers' opinions and impressions. I also think that the users' perspectives are very useful in this context because machine translations are eventually made for humans. And therefore, it is important to not only evaluate a translation by a number without taking into consideration which types of people will use the end product in which kinds of situations and under which circumstances. For example, the satisfaction with and acceptability for machine translated subtitles might be influenced by the age group of the viewers or by the country they live in and how much they have been exposed to subtitles. To conclude, an MT with a Total Quality Score of 80 might be accompanied by a better attitude and therefore higher acceptance in one language than in another. In the present study, the attitude questions delivered valuable additional information, which helped to support the findings collected via other methods.

7.5 Limitations and Recommendations for Future Work

As with the majority of studies, the design of the current study is subject to limitations, which will be discussed briefly in this section. It will also go into possible directions for future related work.

The first limitation is the size of the experiment. With only five participants that took part in this study, the results can only give some first hints, but they cannot be generalised. Therefore, it would be interesting to see if similar studies with a bigger sample size would have the same results. It would also be interesting to compare the performances of more MT systems or other Artificial Intelligences in general, such as ChatGPT. Another possibility for future research in this field would be to compare the performance of MT systems for videos of different domains as well as videos which are less scripted or not scripted at all, such as vlogs. I think it would be interesting to see how well MT can cope with spontaneous speech that includes incomplete sentences, more colloquial language and oftentimes faster speakers.

The next limitation concerns the absence of sound during the experiment. I had to choose the language combination English-German because those two are my strongest languages. Although I am also fluent in French, I am more capable to do a good quality analysis in English, which in turn increases the quality of the results. As it is difficult to find participants that have to solely rely on the subtitles to understand the content if the source language is English, I turned off the sound to circumvent this issue.

The third limitation concerns the annotation process. I was the only annotator; however, it would have been ideal to have at least two annotators to mitigate the risk of subjectivity when annotating errors. Instead, I used the official TED translation as a gold standard against which I compared the two MT versions. As already briefly mentioned in section 5.3.3, I think that it also helped me to make more objective choices that I annotated two translations of the same text because there were many instances where Google Translate and DeepL made either the same or very similar mistakes. I made sure to categorise them uniformly.

Another limitation is that the participants' reading speed was not tested prior to the experiment. Knowing their individual reading speeds could have helped drawing more conclusions from their eye-tracking results.

While analysing the data and interpreting the results of the present study, I encountered several topics that could be addressed in future work. One interesting aspect was the hypothesis about why people might understand subtitles of different quality levels equally well made by Szarkowska and Gerber-Morón (2019) that I briefly discussed in section 7.4.3. They think that

the reason might be that people have been exposed to subtitles of good quality as well as subtitles of bad quality, which might have helped them at learning how to process bad-quality subtitles. It would be interesting to further investigate those ideas and see how comprehension levels and cognitive load change over time. If people become more exposed to machine translated subtitles over the next years, the increased exposure might change the cognitive load induced and comprehension scores. As discussed in section 1.2.2, it has already happened that subtitling guidelines have been changed due to changes in viewing behaviour. Nowadays, the maximum number of cpl and cps is higher than it used to be.

8 Conclusion

This study set out to analyse the quality and reception of machine translated subtitles in the language combination English-German. It compared the performances of two very frequently used MT systems, Google Translate and DeepL, in translating the subtitles of a TED Talk in the domain of environmental protection by using the MQM framework, a questionnaire and eye-tracking. Research question 1 aimed at finding out how good the quality of the translations by Google Translate and DeepL was and what kinds of mistakes the two MT systems made. Research question 2 targeted the analysis and comparison of the reception of machine translated and human translated subtitles. It aimed at assessing the cognitive load induced by them as well as the participants' attitude towards the different kinds of subtitles. Eye-tracking was used to assess the cognitive load, and a Likert scale in combination with one open question was used to measure the attitude of the participants towards human and machine translated subtitles.

On the whole, the results of the study were expectable although there were some surprising aspects. The quality analysis has shown that there is a clear difference in quality between MT and human translation. DeepL performed slightly better than Google Translate, both on the whole and in most dimensions of the MQM. However, Google Translate outperformed DeepL in terms of maximum number of cpl and cps, which are two error types that refer directly to the peculiarities of subtitle presentation on screen. Despite the relatively high number of penalty points, which exposes the fact the subtitles were translated by MT systems, the two MTs were still of use to the participants as shown by the results of the comprehension questionnaire. Although some participants expressed doubts about how much they understood in the attitude questions, the comprehension questionnaire revealed that on the whole, the participants understood the gist of all three parts. However, it is important to note that most participants received a lower number of points when answering content questions for the translation by Google Translate, which indicates they must have struggled more to understand the video when reading those subtitles.

The results of the attitude questions show that the participants rated the human translation the highest. They felt like they understood the content the best, and they enjoyed reading those subtitles the most. They also rated the quality of the human translation the best. The Google Translate version received the worst rating in all categories. These results show that the viewers were able to tell the differences in quality between the three translations although the two MTs were not that far apart in the quality analysis. Also, the quality has effects on how the translation is perceived by the viewers in terms of reading experience.

The eye-tracking experiment provides interesting insights into the cognitive load of the participants while they were watching the three different translations. Although not all eye-tracking parameters indicated that the human translation induced the lowest cognitive load, and the differences between the three types of subtitles were often not very large, they still showed a trend that is in line with the results of the other parts of the experiment, namely that the human translation, which is of the highest quality, is also the one that is most easy to read.

All in all, the answer to research question 1 is that DeepL reached a total quality score of 82.45%, while Google Translate reached a score of 76.41%, which means that overall, DeepL performed better. Both MT systems made mistakes in the same four dimensions of the MQM framework, namely *Accuracy*, *Linguistic Conventions*, *Style*, and *Design and Markup*. In the dimension *Accuracy*, DeepL struggled most with *mistranslations*, *additions* and *omissions*. For Google Translate, *mistranslations* were a big issue, especially since it made many major errors in this error type. *Grammar* and *punctuation* in the dimension *Linguistic Conventions* were big sources of errors for both MT systems; however, Google Translate struggled more with *grammar* than DeepL did. In the dimension *Style*, both MT systems received most penalty points for *unidiomatic style* and *awkward style*. In the dimension *Design and Markup*, most errors were made in the error type *too many cpl*. However, DeepL had more difficulties adhering to the maximum number of cpl. In addition, DeepL was the only MT system that made errors in the error type *too many cps*. The results of the comprehension questionnaire also show that the quality of the Google Translate subtitles must have been worse as the participants seem to have had more difficulties understanding them.

The results of the experiment showed that, at least in this case, the overall quality of the subtitles had an effect on the cognitive load of the participants and the attitude towards the subtitles. Therefore, the answer to research question 2 is that the higher the quality of the subtitles, the lower the cognitive load. This was shown in part by the eye-tracking parameters. In addition, the effect on the attitude was even higher in this study. Participants enjoyed the subtitles by Google Translate less than the ones by DeepL. The gap between the human translation and machine translation was much clearer in the attitude questions than in the cognitive load experiment. This leads me to the conclusion that viewers can deal with subtitles of bad quality quite well (to a certain extend). However, they notice the difference in quality and clearly prefer better-quality subtitles.

In total, these results suggest that the machine translated subtitles were of use to the participants to a certain extent. However, I think that machine translated subtitles of the same quality level as in the present study would only be of use in certain real-life scenarios. For

example, machine translated subtitles make larger volumes of online audio-visual content accessible for more people as it would not be possible for human professional translators to translate such large amounts. Therefore, MT can be very helpful at rendering audiovisual content of social media platforms or video platforms accessible to more people around the world. I believe that in this case, perfect quality is not always more important than quantity to the users because such large volumes of content are consumed every day. In TV shows and films, however, I think quality is very important because people often watch them to relax or enjoy art. Therefore, bad-quality subtitles could ruin the experience. However, it should be investigated further in future studies which quality expectations there are for which scenarios.

In my opinion, this study gave an interesting insight not only into the ability of two very commonly used MT systems to translate subtitles but also into the user experience induced by those subtitles. This should not be neglected because subtitles are made for certain users, and if the quality does not meet the expectations and the capabilities of the intended users, the content that the subtitles are trying to make accessible might not be consumed or understood properly. Therefore, experiments on the user experience of MT or translation in general seems essential to me if the goal is to produce translations that are useful to the intended audience.

The methodology used in the present study was able to include both objective criteria for quality and different perspectives of user experience. It dealt with user experience on a conscious level by using a questionnaire where the participants could express their opinions and on an unconscious level by using eye-tracking, resulting in a comprehensive representation of the user experience. The results of both were also compared to the results of the quality analysis and comprehension questions, which enabled me to examine the effects of quality on user experience.

All in all, this experiment shows a snapshot of the ability of DeepL and Google Translate to translate subtitles of a scripted video on the topic of environmental protection from English into German. I think it is useful to know which aspects they struggled with and how well users could deal with those subtitles. However, MT is a very lively research topic and a field that is constantly evolving. Therefore, research on this topic can be outdated fast, and it is important to keep up with new developments. As already mentioned in the previous chapter, there are numerous topics for related research, and I am excited to read what future studies will discover in the field of MT for subtitling.

9 References

Brendel, J., & Vela, M. (2022). Quality assessment of subtitles: Challenges and strategies. *Text, Speech, and Dialogue (25th International Conference, TSD 2022)*, 52–63. https://link-springer-com.uaccess.univie.ac.at/book/10.1007/978-3-031-16270-1

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring Cognitive Load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 181–202). Cambridge University Press.

Burchard, A., Lommel, A., Bywood, L., Harris, K., & Popović, M. (2018). Machine translation quality in an audiovisual context. In Y. Gambier & S. R. Pinto (Eds.), *Audiovisual Translation: Theoretical and Methodological Challenges* (pp. 23–38). John Benjamins Publishing Company. https://ebookcentral-proquest-com.uaccess.univie.ac.at/lib/univie/reader.action?docID=5398427&ppg=6

Bywood, L., Georgakopoulou, P., & Etchegoyhen, T. (2017). Embracing the threat: Machine translation as a solution for subtitling. *Perspectives*, *25*(3), 492–508.

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, *42*(12), 1150–1152.

Chan, W. S., Kruger, J.-L., & Doherty, S. (2022). An investigation of subtitles as learning support in university education. *The Journal of Specialised Translation*, *38*, 155–179. https://jostrans.org/issue38/art_chan.pdf

Chauhan, S., & Daniel, P. (2022). A comprehensive survey on various fully automatic machine translation evaluation metrics. *Neural Processing Letters*. https://link-springer-com.uaccess.univie.ac.at/article/10.1007/s11063-022-10835-4#citeas

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press. https://www-cambridge-org.uaccess.univie.ac.at/core/books/eyetracking/8E99E0D09594B0098D81E7DC56F0D604

Day, R. R., & Park, J.-s. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, *17*(1), 60–73.

Diaz-Cintas, J. (2020). The name and nature of subtitling. In L. Bogucki & M. Deckert (Eds.), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (pp. 149–171). Palgrave MacMillan.

Díaz-Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices* (1st ed.). Routledge. https://www-taylorfrancis-

com.uaccess.univie.ac.at/books/mono/10.4324/9781315674278/subtitling-jorge-d%C3%ADaz-cintas-aline-remael

Doherty, S., & Kruger, J.-L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In T. Dwyer, C. Perkins, S. Redmond, & J. Sita (Eds.), *Seeing into screens: Eye tracking and the moving image* (pp. 46–64). Bloomsbury Academic.

Fischer, L., & Läubli, S. (2020). What's the Difference Between Professional Human and Machine Translation? A Blind Multi-language Study on Domain-specific MT. https://doi.org/10.48550/arXiv.2006.04781

Gambier, Y. (2009). Perception and reception of audiovisual translation: Implications and challenges. In H. C. Omar, H. Haaroon, & A. A. Ghani (Eds.), *The 12th international conference on translation: The sustainablility of the translation field* (pp. 40–57). Malaysian Translators Association.

Gerber-Morón, O., & Szarkowska, A. (2018). Line breaks in subtitling: An eye tracking study on viewer preferences. *Journal of Eye Movement Research, 11*(3), 1–22.

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research, 11*(4), 1–18.

Gupta, P., Sharma, M., Pitale, K., & Kumar, K. (2019). *Problems with automating translation of movie/TV show subtitles*. arXiv:1909.05362

Hu, K., O'Brien, S., & Kenny, D. (2020). A reception study of machine translated subtitles for MOOCs. *Perspectives*, *28*(4), 521–538. https://doi.org/10.1080/0907676X.2019.1595069

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical Education*, *38*(12), 1217–1218.

Karakanta, A. (2022). Experimental research in automatic subtitling: At the crossroads between machine translation and audiovisual translation. *Translation Spaces, 11*(1), 89–112. https://doi.org/10.1075/ts.21021.kar

Karakanta, A., Negri, M., & Turchi, M. (2020a). MuST-Cinema: a speech-to-subtitles corpus. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 3727–3734.

Karakanta, A., Negri, M., & Turchi, M. (2020b). Towards automatic subtitling: Assessing the quality of old and new resources. *Italian Journal of Computational Linguistics*, *6*(1), 63–76.

Kasperavičienė, R., Motiejūnienė, J., & Patašienė, I. (2020). Quality assessment of machine translation output: Cognitive evaluation approach in an eye tracking experiment. *Belo Horizonte*, *13*(2), 271–285.

Kruger, J.-L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: Evidence from word skipping and rereading. *Applied Psycholinguistics*(43), 211–236.

Lommel, A. (2018). Metrics for translation quality assessment: A case for standardising error typologies. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 109–128). Springer.

Martínez, J. M., & Vela, M. (2016). SubCo: A learner translation corpus of human and machine subtitles. *Proceedings of the 10th International Conference on Language Resources and Evaluation*, 2246–2254.

Matusov, E., Wilken, P., & Georgakopoulou, P. (2019). Customizing nerual machine translation for subtitling. *Proceedings of the Fourth Conference on Machine Translation*, *1*, 82–93. https://aclanthology.org/W19-5209.pdf

MQM. (n. d.–a). *The MQM Typology*. https://themqm.org/the-mqm-typology/

MQM. (n. d.–b). *Values and Scores*. https://themqm.org/error-types-2/1_scorecards/values-and-scores/

Netflix. (2022). *Timed text style guide: Subtitle templates*. https://partnerhelp.netflixstudios.com/hc/en-us/articles/219375728-Timed-Text-Style-Guide-Subtitle-Templates

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737–743. https://www.researchgate.net/publication/258138558_The_Efficiency_of_Instructional_Conditions_An_Approach_to_Combine_Mental_Effort_and_Performance_Measures

Paas, Fred G. W. C., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371. https://www-jstor-org.uaccess.univie.ac.at/stable/23359294

Papi, S., Gaido, M., Karakanta, A., Cettolo, M., Negri, M., & Turchi, M. (2022). *Direct speech translation for automatic subtitling*. arXiv:2209.13192

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. *: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 311–318.

Popović, M. (2018). Error classification and analysis for machine translation quality. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 129–158). Springer.

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search, *62*(8), 1457-1406. https://journals-sagepub-com.uaccess.univie.ac.at/doi/epdf/10.1080/17470210902816461

Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for MT evaluation, n. p. https://arxiv.org/abs/2009.09025

Sanders, G., Przybocki, M., Madnani, N., & Snover, M. (2011). Human subjective judgments. In Joseph Olive, Caitline Christianson, & John McCary (Eds.), *Handbook of natural language processing and machine translation: DARPA Global Autonomous Language Exploitation* (1st ed., pp. 801–894). Springer New York.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, 223–231. https://aclanthology.org/2006.amta-papers.25/

Song, H.-J., Kim, H.-K., Kim, J.-D., Park, C.-Y., & Kim, Y.-S. (2019). Inter-sentence segmentation of YouTube subtitles using long-short Term Memory (LSTM). *Applied Sciences*, *9*(7), 1–10.

Statista. (2022). *Hours of video uploaded to YouTube every minute 2007-2020.* https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292.

Szarkowska, A., & Gerber-Morón, O. (2018). Viewers can keep up with fast subtitles: Evidence from eye movements. *PLoS ONE*, *13*(6), 1–30.

Szarkowska, A., & Gerber-Morón, O. (2019). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives*, *27*(1), 144–164.

TED. (n. d.). *Subtitling tips*. Retrieved December 17, 2022, from https://www.ted.com/participate/translate/subtitling-tips

Wilken, P., Georgakopoulu, P., & Natusov, E. (2022). *SubER: A metric for automatic evaluation of subtitle quality*. arXiv:2205.05805

Xie, B. (2022). *A comparative study of machine translated subtitles based on the user-centered approach: A case study between Bilibili and YouTube*. 10.21203/rs.3.rs-2179598/v1 https://doi.org/10.21203/rs.3.rs-2179598/v1

Zahedi, S., & Khoshsaligheh, M. (2021). Eyetracking the impact of subtitle length and line number on viewers allocation of visual attention. *Translation, Cognition & Behavior*, *4*(2), 331–352.

Annex

Annex 1: Quality Analysis Human Translation

Annex 2: Quality Analysis DeepL

Annex 3: Quality Analysis Google Translate

Annex 4: MQM Scorecard Human Translation

Annex 5: MQM Scorecard DeepL

Annex 6: MQM Scorecard Google Translate

Annex 7: Questionnaires

Annex 1

Quality Analysis Human Translation (TED)			
Subtitle	**Error Type**	**Dimension**	**Severity Level**
1			
00:00:00.0 --> 00:00:07.000			
Übersetzung: Birgit Ackermann			
Lektorat: Andrea Hielscher			
2			
00:00:04.334 --> 00:00:07.879			
Mir wurde eine der größten Ehren zuteil,			
3			
00:00:07.879 --> 00:00:10.966			
die einem Politiker wohl widerfahren kann:			
4			
00:00:10.966 --> 00:00:13.510			
Ich wurde zum Bürgermeister			
der Stadt gewählt,			
5			
00:00:13.51 --> 00:00:15.720			
in dem ich geboren wurde und aufwuchs.	grammar	linguistic conventions	minor
correct: der			
6			
00:00:15.72 --> 00:00:18.056			
Das ist Bristol in Großbritannien.			
7			
00:00:18.056 --> 00:00:19.099			
Danke schön!	spelling	linguistic conventions	minor
correct: Dankeschön!			
8			
00:00:19.099 --> 00:00:21.393			
(Beifall)			
9			
00:00:21.56 --> 00:00:23.144			
Ds ist eine große Aufgabe.	spelling	linguistic conventions	minor
correct: Das			
10			
00:00:23.144 --> 00:00:25.313			
Von Bildung über Wohnen,			

11
00:00:25.772 --> 00:00:27.857
Budgets und Müllabfuhr

12
00:00:28.233 --> 00:00:31.194
bis hin zu Protesten
und Gegendemonstrationen

13
00:00:31.486 --> 00:00:33.780
sind Städte komplizierte Organismen.

14
00:00:34.155 --> 00:00:37.909
Sie können turbulent sein
und voller Widersprüche.

15
00:00:38.743 --> 00:00:42.455
Als Bürgermeister bin ich
die verantwortliche Person in Bristol.

16
00:00:42.455 --> 00:00:44.124
Ich bin Rechenschaft schuldig,

17
00:00:44.124 --> 00:00:46.209
manchmal sogar bei Angelegenheiten,

18
00:00:46.209 --> 00:00:50.630
über die ich sehr wenig
wirkliche Kontrolle oder Macht habe.

19
00:00:51.089 --> 00:00:52.299
Das ist in Ordnung.

20
00:00:52.299 --> 00:00:53.466
So ist das Leben.

21
00:00:53.675 --> 00:00:56.720
Aber ich schulde

einem weiteren Wahlkreis Rechenschaft,

22
00:00:56.72 --> 00:00:59.014
der sich zwar in meiner Stadt befindet,

23
00:00:59.514 --> 00:01:02.225
aber über die Stadtgrenzen hinausreicht.

24
00:01:02.309 --> 00:01:03.852
Das sind unser Planet

25
00:01:03.852 --> 00:01:08.440
und die 7,9 Milliarden Mitmenschen,
deren Überleben von ihm abhängt.

26
00:01:09.357 --> 00:01:13.445
Wir haben uns in eine Situation gebracht,
in der es 1,7 Erden bräuchte,

27
00:01:13.778 --> 00:01:16.740
um die derzeitige Lebensweise
nachhaltig zu gestalten.

28
00:01:16.823 --> 00:01:18.950
Es muss also etwas passieren.

29
00:01:19.284 --> 00:01:22.954
Denn wir alle wissen, dass wir
keine weitere Erde bekommen, oder?

30
00:01:22.954 --> 00:01:26.625
Also steht es außer Frage,
dass wir uns ändern müssen.

31
00:01:26.958 --> 00:01:31.212
Während wir die internationalen
Verhandlungen verfolgen,

32				
00:01:31.671 --> 00:01:35.050				
während wir allzu oft				
nationale Untätigkeit beobachten,				
33				
00:01:35.05 --> 00:01:38.553				
fragen sich wohl viele,				
wie wir diese Herausforderung meistern				
34				
00:01:38.803 --> 00:01:41.765				
und sogar, ob wir erfolgreich sein werden.				
35				
00:01:41.806 --> 00:01:43.016				
Ich verstehe,				
36				
00:01:43.391 --> 00:01:46.394				
dass so viele Menschen weltweit				
die Hoffnung verlieren.				
37				
00:01:46.519 --> 00:01:48.521				
Aber meine heutige Botschaft lautet:				
38				
00:01:48.521 --> 00:01:51.066				
Es gibt Hoffnung				
und sie liegt im Verborgenen.	undertranslation	accuracy	neutral	
correct (e.g.): ist nicht auf dem ersten Blick erkennbar / ist nicht sofort sichtbar				
39				
00:01:51.316 --> 00:01:54.569				
Ich glaube, unsere Städte				
verfügen über riesiges Potenzial.				
40				
00:01:54.861 --> 00:01:56.946				
Schauen Sie sich diese vier Zahlen an.				
41				
00:01:57.322 --> 00:01:58.406				
Drei,				
42				
00:01:58.657 --> 00:01:59.824				

55

43
00:02:00.075 --> 00:02:01.242
75

44
00:02:01.618 --> 00:02:02.702
80.

45
00:02:02.994 --> 00:02:07.499
Städte nehmen weniger als drei Prozent
der Landfläche der Erde ein.

46
00:02:07.499 --> 00:02:10.502
Wir haben also einen
kleinen geografischen Fußabdruck.

47
00:02:10.919 --> 00:02:12.712
Nähme man alle Städte der Welt,

48
00:02:12.712 --> 00:02:15.173
bedeckten sie gerade die Fläche Indiens.

49
00:02:15.757 --> 00:02:18.968
Und doch lebt in Städten über die Hälfte,

50
00:02:19.344 --> 00:02:22.138
nämlich 55 Prozent, der Weltbevölkerung.

51
00:02:22.514 --> 00:02:24.391
Voraussichtlich steigt diese Zahl

52
00:02:24.391 --> 00:02:27.394
bis Mitte des Jahrhunderts
auf zwei Drittel.

53
00:02:27.852 --> 00:02:31.981
Städte sind verantwortlich

für rund 75 Prozent der CO_2-Emissionen.

54
00:02:32.44 --> 00:02:37.320
Außerdem stoßen sie in erheblichem Maße
Stickstoffdioxid und Methan aus.

55
00:02:37.779 --> 00:02:41.700
Städte verbrauchen weltweit
80 Prozent der Energie.

56
00:02:42.075 --> 00:02:43.743
Aber bedenken Sie Folgendes.

57
00:02:44.244 --> 00:02:46.788
Gerade die Merkmale von Städten --

58
00:02:46.788 --> 00:02:50.083
Einflussbereich, Größe, Dichte,

59
00:02:50.583 --> 00:02:54.087
unmittelbare Nähe der Führungsriege
zu den Menschen,

60
00:02:54.713 --> 00:02:59.050
Anpassungsfähigkeit
und ihr Vermögen, sich neu zu erfinden --

61
00:02:59.467 --> 00:03:03.263
all das bedeutet, dass diese Zahlen
tatsächlich beherrschbar sind.

62
00:03:03.471 --> 00:03:06.725
Das heißt, dank unserer Städte
können wir planen,

63
00:03:06.725 --> 00:03:10.437
für mehr Menschen mehr zu leisten,
aber mit weniger Ressourcen.

64
00:03:10.77 --> 00:03:14.190
Deshalb sind Städte für mich
eins der effektivsten Werkzeuge,

65
00:03:14.19 --> 00:03:16.025
die uns zur Verfügung stehen,

66
00:03:16.443 --> 00:03:19.779
um unsere Beziehungen
zu Grund und Boden, Energie und Müll

67
00:03:20.28 --> 00:03:22.907
effizienter zu gestalten.

68
00:03:23.742 --> 00:03:25.243
In unseren Städten

69
00:03:25.243 --> 00:03:29.122
können wir die Effizienz
von mehr Menschenleben schneller steigern

70
00:03:29.122 --> 00:03:33.084
als durch jede andere Form
menschlicher Organisation.

71
00:03:33.96 --> 00:03:35.670
So können wir zum Beispiel

72
00:03:35.67 --> 00:03:38.882
mehr Menschen auf weniger Land
Wohnraum und Arbeit bieten

73
00:03:39.34 --> 00:03:44.596
und so der Zersiedelung entgegenwirken,
die in Land und Natur eingreift,

74

00:03:44.596 --> 00:03:46.723				
dabei auch die Entfernungen reduzieren,				
75				
00:03:46.723 --> 00:03:50.226				
die man zur Befriedigung				
der Grundbedürfnisse zurücklegen muss.				
76				
00:03:50.351 --> 00:03:54.230				
In Städten können Menschen				
durch gemeinsame Nutzung von Gebäuden				
77				
00:03:54.23 --> 00:03:57.901				
und durch intelligente Neuerungen				
wie Fernwärme Energie teilen.				
78				
00:03:58.401 --> 00:04:01.529				
Die Dichte unserer Städte				
macht den öffentlichen Verkehr				
79				
00:04:01.78 --> 00:04:04.783				
zugänglicher und kostengünstiger.				
80				
00:04:05.7 --> 00:04:09.913				
Durch unsere Städte können wir				
unser Verhältnis zur Energie ändern.				
81				
00:04:09.913 --> 00:04:12.665				
Gerade jetzt	mistransatlation	accuracy	minor	
brauchen wir Energiesicherheit.				
correct: Wir brauchen sofort Energiesicherheit				
82				
00:04:12.665 --> 00:04:15.418				
Aber Städte bieten Märkte				
von solcher Dimension,				
83				
00:04:15.418 --> 00:04:19.130				
dass Investitionen in erneuerbare Energien				
attraktiver werden.				

84
00:04:19.547 --> 00:04:22.133
Bedenken Sie die Möglichkeiten
beim Thema Müll.

85
00:04:22.133 --> 00:04:23.551
Wir können die Effizienz

86
00:04:23.551 --> 00:04:26.429
beim Sammeln und der Verarbeitung
von Müll steigern

87
00:04:26.513 --> 00:04:31.267
und gleichzeitig Kreislaufwirtschaft
in großem Maßstab einführen,

88
00:04:31.267 --> 00:04:33.853
um Ressourcen zu recyceln,

89
00:04:33.853 --> 00:04:35.480
Güter wiederzuverwenden

90
00:04:35.48 --> 00:04:38.983
und unvermeidbare Abfälle
in die Energieverwendung zu stecken,

91
00:04:38.983 --> 00:04:41.903
etwa Lebensmittelabfälle als Düngemittel.

92
00:04:42.612 --> 00:04:47.659
Stellen Sie sich das globale Potenzial
eines weltweiten Städtenetzwerks vor,

93
00:04:47.951 --> 00:04:50.703
das diese Art von Effizienzsteigerungen

94
00:04:50.703 --> 00:04:56.167
für über die Hälfte und bald zwei Drittel

der Weltbevölkerung erreicht.

95
00:04:57.085 --> 00:05:00.296
Hier ist die Hoffnung,
die ich eingangs erwähnte.

96
00:05:01.047 --> 00:05:03.216
Sie müssen sich das nicht nur vorstellen.

97
00:05:03.341 --> 00:05:05.301
Von Freetown bis Los Angeles,

98
00:05:05.635 --> 00:05:07.428
von Kampala bis London,

99
00:05:07.762 --> 00:05:10.181
und in vielen, vielen Städten dazwischen

100
00:05:10.598 --> 00:05:14.686
ergreifen Bürgermeister
und Stadtoberhäupter Maßnahmen,

101
00:05:14.686 --> 00:05:17.522
um sich aktuellen Herausforderungen
zu stellen.

102
00:05:18.064 --> 00:05:22.235
Nehmen wir Malmö,
eine Stadt mit knapp 350.000 Einwohnern.

103
00:05:23.069 --> 00:05:25.613
Dort wurde ein Fernwärmesystem entwickelt,

104
00:05:25.613 --> 00:05:29.117
das mit Wärme aus verarbeiteten
Abfällen gespeist wird.

105

00:05:29.909 --> 00:05:32.829
Malmö plant bis 2030
die 100-prozentige Versorgung

106
00:05:33.037 --> 00:05:36.749
mit erneuerbarer oder recycelter Wärme.

107
00:05:37.834 --> 00:05:42.422
Die Stadt Oslo subventioniert
Elektrofahrzeuge und Ladestationen.

108
00:05:43.298 --> 00:05:46.426
Es gibt dort
ein Kreislaufwirtschaftssystem für Müll.

109
00:05:46.759 --> 00:05:48.678
Man hat eine Biogasanlage gekauft

110
00:05:48.678 --> 00:05:52.682
und fast 50 Prozent
aller Lebensmittelabfälle werden recycelt.

111
00:05:53.391 --> 00:06:00.106
(Beifall)

112
00:06:00.607 --> 00:06:04.235
Singapur ist eine der am dichtesten
besiedelten Städte der Welt,

113
00:06:04.402 --> 00:06:06.988
aber ein Vorbild für Grünflächenplanung.

114
00:06:07.071 --> 00:06:10.783
In den letzten Jahren wurden
riesige Süßwasserreserven

115
00:06:10.783 --> 00:06:14.329
und städtische Gärten angelegt,

die die Lunge der Stadt bilden.

116
00:06:14.621 --> 00:06:16.956
Ich habe auch
große Bewunderung für Bogota,

117
00:06:16.956 --> 00:06:20.126
eine der am dichtesten besiedelten
Städte Lateinamerikas.

118
00:06:20.335 --> 00:06:22.962
Dort wurde
ein Schnellbussystem eingeführt.

119
00:06:22.962 --> 00:06:25.423
Zu-Fuß-Gehen und Radfahren
werden gefördert

120
00:06:25.423 --> 00:06:27.216
und die Stadt verfügt heute

121
00:06:27.216 --> 00:06:30.887
über eine der größten
Elektrobusflotten Lateinamerikas.

122
00:06:32.055 --> 00:06:33.598
Solange ich das Wort habe,

123
00:06:33.598 --> 00:06:37.185
lassen Sie mich ein wenig
meine eigene Stadt loben, Bristol,

124
00:06:37.477 --> 00:06:41.481
Heimat für 465.901 Menschen,

125
00:06:41.481 --> 00:06:43.399
von denen einer heute hier ist.

126
00:06:43.733 --> 00:06:47.278
Wir haben einen fantastischen Ruf

127
00:06:47.487 --> 00:06:49.656
als eine der grünsten Städte Europas.

128
00:06:50.239 --> 00:06:52.617
In Bristol gibt es eine Wohnungskrise.

129
00:06:53.159 --> 00:06:54.953
Wir müssen Wohnungen bauen.

130
00:06:55.578 --> 00:06:57.956
Aber wir wissen ganz genau:

131
00:06:57.956 --> 00:07:01.000
Die Art Häuser,
die wir bauen und wo wir sie bauen,

132
00:07:01.292 --> 00:07:04.295
wird einer der größten Faktoren
für den Preis sein,

133
00:07:04.295 --> 00:07:06.547
den der Planet für unser Wachstum zahlt.

134
00:07:06.756 --> 00:07:11.552
Unser Ziel ist also
Netto-Null-Wohnraum in größerer Dichte

135
00:07:11.552 --> 00:07:14.430
auf altem Industriegelände
mitten in der Stadt.

136
00:07:15.098 --> 00:07:18.768
Das verringert den Druck
durch Zersiedelung.

137
00:07:18.768 --> 00:07:21.479
So können wir aktive Mobilität fördern

138
00:07:21.771 --> 00:07:24.357
und die Abhängigkeit vom Auto verringern.

139
00:07:25.191 --> 00:07:26.985
Wir ergreifen sogar Maßnahmen

140
00:07:26.985 --> 00:07:29.696
gegen die Klimafolgen
von banalen Toiletten.

141
00:07:30.405 --> 00:07:32.699
In unserem öffentlichen Wohnungsbestand

142
00:07:32.699 --> 00:07:34.117
erneuern wir die Bäder

143
00:07:34.117 --> 00:07:36.619
und ersetzen dabei die Armaturen

144
00:07:36.619 --> 00:07:38.788
durch wassersparendere Alternativen,

145
00:07:39.038 --> 00:07:42.291
wassersparendere Duschen,
Waschbecken und Wasserhähne

146
00:07:42.625 --> 00:07:45.003
sowie wassersparendere Toiletten.

147
00:07:45.378 --> 00:07:48.131
Nicht alle Klimamaßnahmen
sind spektakulär, oder?

148
00:07:49.465 --> 00:07:52.260
Aber wir Bürgermeister
konzentrieren uns nicht nur

149
00:07:52.26 --> 00:07:54.929
auf das Geschehen
innerhalb der Stadtgrenzen.

150
00:07:55.388 --> 00:08:00.393
Bürgermeister auf der ganzen Welt
sind über ihre Befugnisse hinaus tätig.

151
00:08:00.393 --> 00:08:03.146
Sie bilden internationale Netzwerke,

152
00:08:03.146 --> 00:08:05.815
setzen sich feste Dekarbonisierungsziele

153
00:08:05.815 --> 00:08:09.152
und überwachen gegenseitig
deren Einhaltung.

154
00:08:09.152 --> 00:08:10.653
Weltweit gibt es derzeit

155
00:08:10.653 --> 00:08:13.990
Hunderte dieser städtischen
Solidaritätsnetzwerke.

156
00:08:14.782 --> 00:08:18.286
Der Globale Konvent
der Bürgermeister für Klima und Energie

157
00:08:18.286 --> 00:08:20.997
ist ein Netzwerk von rund 12.000 Städten.

158
00:08:21.372 --> 00:08:24.500

Sie haben sich gemeinsam				
zu Maßnahmen verpflichtet,				
159				
00:08:24.5 --> 00:08:29.589				
um ihre Emissionen bis 2030	awkward style	style	minor	
um fast zwei Gigatonnen zu senken,				
160				
00:08:29.589 --> 00:08:33.217				
was nicht der Fall wäre,				
wenn alles beim Alten bliebe.				
correct: die sicherstellen, dass ihre Emissionen bis 2030 um zwei Gigatonnen niederiger sind als sie es wären,				
161				
00:08:33.718 --> 00:08:34.761				
Als Bürgermeister				
162				
00:08:34.761 --> 00:08:38.598				
bemühen wir uns auch um Einfluss				
auf internationale Organisationen				
163				
00:08:38.598 --> 00:08:42.643				
und um Unterstützung der globalen Politik				
zur Umsetzung dieser Maßnahmen.				
164				
00:08:43.394 --> 00:08:46.773				
Die C40 ist ein Netzwerk				
von fast 100 Bürgermeistern,				
165				
00:08:46.898 --> 00:08:50.860				
die die größten Städte der Welt vertreten.				
166				
00:08:51.569 --> 00:08:54.405				
Im Fokus ihrer Arbeit				
steht City-Diplomatie.				
167				
00:08:54.781 --> 00:08:58.326				
Sie nehmen an nationalen				
und internationalen Verhandlungen teil,				

168
00:08:58.326 --> 00:09:02.330
um Entscheidungsfindung und
globale Verpflichtungen zu beeinflussen.

169
00:09:02.914 --> 00:09:04.248
Ich versichere Ihnen:

170
00:09:04.248 --> 00:09:08.419
Bürgermeister engagieren sich dafür,

171
00:09:08.836 --> 00:09:13.674
diese globalen Verpflichtungen
von Worten in Taten umzusetzen.

172
00:09:14.3 --> 00:09:19.639
Die Nähe zu unseren Einwohnern bedeutet:
Wir sind direkt verantwortlich

173
00:09:19.931 --> 00:09:24.477
für die Umsetzung von Veränderungen,
die die Menschen sehen und erleben können.

174
00:09:26.02 --> 00:09:28.689
Hier stoßen wir
auf eine neue Herausforderung.

175
00:09:29.565 --> 00:09:33.986
Dieses notwendige weltweite Netzwerk
effizienter Städte

176
00:09:33.986 --> 00:09:36.989
bekommen wir nur
mit größeren Investitionen.

177
00:09:37.907 --> 00:09:41.994
Wir bekommen kein weltweites Netzwerk
dekarbonisierter Städte,

178
00:09:42.245 --> 00:09:43.704
nur weil wir es brauchen,

179
00:09:43.996 --> 00:09:45.164
weil wir es wollen

180
00:09:45.248 --> 00:09:48.167
oder weil wir darüber
blumige Erklärungen abgeben.

181
00:09:49.001 --> 00:09:52.964
Wir bekommen es nur,
wenn wir es planen und dafür bezahlen.

182
00:09:54.34 --> 00:09:56.342
Letztlich erhalten Stadtoberhäupter

183
00:09:56.384 --> 00:09:59.011
weltweit nur schwer Zugang
zu Finanzmitteln,

184
00:09:59.011 --> 00:10:01.848
um das volle Potenzial ihrer Stadt
zu erschließen.

185
00:10:02.557 --> 00:10:04.684
Selbst hier sind wir nicht untätig.

186
00:10:04.767 --> 00:10:07.228
Im Vereinigten Königreich bin ich Mitglied

187
00:10:07.228 --> 00:10:09.480
der UK Cities
Climate Investment Commission.

188
00:10:09.48 --> 00:10:12.233
Wir wollen

den größten Städten Großbritanniens			
189			
00:10:12.233 --> 00:10:17.155			
zu den nötigen Finanzmitteln verhelfen,			
um dieses Potenzial freizusetzen.			
190			
00:10:17.655 --> 00:10:21.659			
Wir haben zur Dekarbonisierung			
in Großbritannien			
191			
00:10:21.659 --> 00:10:24.745			
Möglichkeiten im Wert von			
206 Milliarden Pfund ermittelt,			
192			
00:10:24.745 --> 00:10:28.541			
Nachrüstung, erneuerbare Energien,			
Umstellung auf Elektro-Fuhrparks.			
193			
00:10:29.167 --> 00:10:30.376			
Und wir informieren			
194			
00:10:30.376 --> 00:10:34.380			
öffentliche und private Investoren			
in ganz Großbritannien	mistranslation	accuracy	neutral
correct: im ganzen Vereinigten Königreich			
195			
00:10:34.589 --> 00:10:36.132			
über diese Möglichkeiten.			
196			
00:10:37.049 --> 00:10:38.593			
Die Sache ist also die.			
197			
00:10:39.51 --> 00:10:41.679			
Bürgermeister, Stadtoberhäupter,			
198			
00:10:41.679 --> 00:10:46.976			
es ist keine Zeit für abstrakte Debatten			
oder blumige Erklärungen.			

199
00:10:47.435 --> 00:10:50.605
Unsere Bevölkerungen
fordern Veränderungen heute.

200
00:10:50.98 --> 00:10:52.940
Am liebsten schon gestern.

201
00:10:53.441 --> 00:10:56.652
Die derzeitige Klimakrise
erfordert Führung.

202
00:10:57.153 --> 00:11:02.116
Bürgermeister überall auf der Welt
stellen sich dieser Situation

203
00:11:02.325 --> 00:11:03.492
und gehen sie an.

204
00:11:04.076 --> 00:11:05.620
Wir wollen und brauchen

205
00:11:05.828 --> 00:11:09.040
nationale Regierungen
und internationale Organisationen,

206
00:11:09.04 --> 00:11:12.126
die mit uns arbeiten,
uns bestärken, uns unterstützen.

207
00:11:12.251 --> 00:11:14.253
Aber wir können nicht auf sie warten.

208
00:11:14.503 --> 00:11:16.672
Laut der weltweit besten Wissenschaftler

209
00:11:16.672 --> 00:11:19.467
haben wir zehn Jahre,

um das Blatt zu wenden.

210
00:11:20.259 --> 00:11:24.889
Aufgrund der bisherigen Untätigkeit,
mangelnden Leistungsbereitschaft

211
00:11:24.889 --> 00:11:27.016
und schleppenden Entscheidungsfindung

212
00:11:27.016 --> 00:11:30.436
müssen wir gerade jetzt hoch pokern

213
00:11:31.187 --> 00:11:34.190
und mit diesem Einsatz
Maßnahmen ergreifen,

214
00:11:34.357 --> 00:11:38.152
die schnelle und umfangreiche
Veränderungen bewirken.

215
00:11:38.819 --> 00:11:41.572
Ich denke, unsere Städte
bieten gute Gewinnchancen.

216
00:11:42.281 --> 00:11:44.075
Ich rufe also zum Handeln auf.

217
00:11:45.993 --> 00:11:51.874
Wir müssen mit Bürgermeistern weltweit
einen globalen Plan für Städte entwickeln.

218
00:11:53.0 --> 00:11:55.878
Der Plan muss
den Ausstoß von Kohlendioxid reduzieren,

219
00:11:56.087 --> 00:11:58.965
die Effizienz bestehender Städte steigern,

220			
00:11:59.382 --> 00:12:01.008			
dabei aber auch garantieren,			
221			
00:12:01.008 --> 00:12:06.013			
dass zukünftige Urbanisierungsprozesse			
die Effizienz von Städten maximieren.			
222			
00:12:06.806 --> 00:12:08.432			
Und wenn ich “global” sage,			
223			
00:12:08.641 --> 00:12:10.309			
dann meine ich auch “global”.			
224			
00:12:10.518 --> 00:12:13.938			
Dieser Plan muss			
über nationale Grenzen hinausgehen.			
225			
00:12:14.355 --> 00:12:16.190			
Er muss sidh im Globalen Norden	spelling	linguistic conventions	minor
correct: sich			
226			
00:12:16.357 --> 00:12:18.609			
und auch im Globalen Süden realisieren,			
227			
00:12:18.609 --> 00:12:23.072			
wo sich 90 Prozent der künftigen			
Urbanisierung abspielen werden.			
228			
00:12:24.031 --> 00:12:25.241			
Es ist ein Plan,			
229			
00:12:25.241 --> 00:12:31.038			
bei dem wir über unser enges nationales			
Eigeninteresse hinausgehen müssen.			
230			
00:12:31.08 --> 00:12:35.418			
Wir müssen lernen, die Städte der Welt			
als internationales Gut zu betrachten			

231
00:12:35.543 --> 00:12:37.545
statt als nationale Besitztümer.

232
00:12:38.212 --> 00:12:39.797
Wir müssen erkennen:

233
00:12:39.797 --> 00:12:43.759
Investitionen in die Effizienzsteigerung
der Städte weltweit

234
00:12:44.135 --> 00:12:47.096
sind der Schlüssel zu unserer Zukunft.

235
00:12:47.972 --> 00:12:51.267
Der Schlüssel,
um unser Potenzial freizusetzen,

236
00:12:51.35 --> 00:12:54.478
und eine Investition
in unser globales Gemeinwohl.

237
00:12:56.314 --> 00:12:59.442
Als ich 2016 zum Bürgermeister
von Bristol gewählt wurde,

238
00:12:59.775 --> 00:13:06.115
hatte ich nur eine begrenzte Vorstellung
von der globalen Rolle der Stadt.

239
00:13:06.991 --> 00:13:08.534
Daher hatte ich auch

240
00:13:08.534 --> 00:13:12.330
nur eine begrenzte Vorstellung
des Maßes an Verantwortung,

241

00:13:12.747 --> 00:13:16.917
die ich als neu gewählter Bürgermeister
einer britischen Großstadt

242
00:13:17.168 --> 00:13:19.045
würde tragen müssen.

243
00:13:20.254 --> 00:13:22.798
In den Folgejahren habe ich begriffen,

244
00:13:22.798 --> 00:13:23.841
dass Städte,

245
00:13:24.175 --> 00:13:27.094
die Art ihrer Planung,
ihres Funktionierens,

246
00:13:27.553 --> 00:13:30.348
ihres Wachstums und ihrer Erneuerung,

247
00:13:31.182 --> 00:13:35.311
der Schlüssel unseres Erfolgs
oder Misserfolgs

248
00:13:35.311 --> 00:13:37.271
bei der Aufgabe sind,

249
00:13:37.646 --> 00:13:40.691
die Flut des globalen Klimawandels
einzudämmen.

250
00:13:41.65 --> 00:13:46.322
Gelingt es uns, das volle Potenzial
unserer Städte zu erschließen,

251
00:13:46.322 --> 00:13:48.324
können wir den Preis minimieren,

252
00:13:48.324 --> 00:13:51.911
den der Planet
für die wachsende Bevölkerung zahlt.

253
00:13:52.828 --> 00:13:58.417
Effiziente Städte könnten hier eines
unserer wirksamsten Instrumente sein.

254
00:13:58.667 --> 00:14:01.712
Ich bitte Sie also,
sie gemeinsam mit uns zu gestalten.

255
00:14:01.921 --> 00:14:03.089
Ich danke Ihnen.

256
00:14:03.089 --> 00:14:08.928
(Beifall)

Annex 2

Quality Analysis DeepL			
Subitle	**Error Type**	**Dimension**	**Severity Level**
1			
00:00:04.334 --> 00:00:08.630			
Ich habe eine der größten Ehrungen	unidiomatic style punctuation	style linguistic conventions	minor minor
2			
00:00:08.63 --> 00:00:11.049			
die ich glaube, dass jeder Politiker haben könnte,	awkward style unidiomatic style cpl	style style design and markup	minor minor minor
3			
00:00:11.091 --> 00:00:14.803			
dass ich zum Bürgermeister gewählt wurde	awkward style grammar	style linguistic conventions	minor minor
des Ortes, in dem ich geboren wurde	unidiomatic style	style	neutral
4			
00:00:14.844 --> 00:00:16.012			
und aufgewachsen bin.			
5			
00:00:16.012 --> 00:00:18.056			
Das ist also Bristol in Großbritannien.	unidiomatic style	style	neutral
6			
00:00:18.056 --> 00:00:19.224			
Nun, danke schön.	unidiomatic style	style	neutral
7			
00:00:19.266 --> 00:00:21.518			
(Beifall)			
8			
00:00:21.56 --> 00:00:23.144			
Und es ist eine große Aufgabe.	awkward style	style	neutral
9			
00:00:23.186 --> 00:00:25.772			
Von Bildung bis Wohnen,	awkward style	style	minor
10			

00:00:25.814 --> 00:00:28.233			
bis hin zu den Haushalten und der Müllabfuhr	unidiomatic style cpl	style design and markup	neutral minor
11			
00:00:28.275 --> 00:00:31.486			
zu Protesten und Gegenprotesten,			
12			
00:00:31.528 --> 00:00:34.155			
Städte sind komplizierte Organismen.	grammar	linguistic conventions	minor
13			
00:00:34.197 --> 00:00:37.909			
Sie können stürmisch sein,	mistranslation punctuation	accuracy linguistic conventions	minor minor
und sie können voll von Widersprüchen sein.	awkward style cpl	style design and markup	neutral minor
14			
00:00:38.743 --> 00:00:42.455			
Und als Bürgermeisterin bin ich die verantwortliche	mistranslation awkward style cpl	accuracy style design and markup	minor neutral mior
Person in Bristol.			
15			
00:00:42.455 --> 00:00:44.207			
Die Verantwortung liegt bei mir.	unidiomatic style	style	neutral
16			
00:00:44.207 --> 00:00:46.167			
Und das ist manchmal sogar für Probleme	awkward style punctuation	style linguistic conventions	minor minor
17			
00:00:46.209 --> 00:00:51.089			
über die ich sehr wenig			
wirkliche Kontrolle oder Macht habe.			
18			
00:00:51.131 --> 00:00:52.299			
Und das ist gut so.	mistranslation	accuracy	minor
19			
00:00:52.34 --> 00:00:53.466			
So ist das Leben.			

20			
00:00:53.842 --> 00:00:56.803			
Aber ich habe eine andere Wählerschaft	awkward style	style	minor
dem ich rechenschaftspflichtig bin,	grammar unidiomatic style	linguistic conventions style	minor minor
21			
00:00:56.845 --> 00:00:59.472			
und das ist einer, der in meiner Stadt liegt,	grammar awkward style undertranslation cpl	linguistic conventions style accuracy design and markup	minor neutral neutral minor
22			
00:00:59.514 --> 00:01:02.267			
aber über meine Stadtgrenzen hinausreicht.			
23			
00:01:02.309 --> 00:01:03.852			
Und das ist der Planet	grammar awkward style	linguistic conventions style	neutral neutral
24			
00:01:03.893 --> 00:01:08.607			
und die 7,9 Milliarden anderen Menschen	unidiomatic style punctuation	style linguistic conventions	neutral minor
die für ihr Überleben auf ihn angewiesen sind.	awkward style cpl	style design and markup	neutral minor
25			
00:01:09.482 --> 00:01:13.778			
Wir haben uns in eine Situation gebracht	punctuation	linguistic conventions	minor
in der wir 1,7 Erden brauchen würden	awkward style punctuation	style linguistic conventions	neutral minor
26			
00:01:13.778 --> 00:01:16.823			
damit unsere derzeitige Lebensweise	awkward style	style	neutral
nachhaltig wäre.			
27			
00:01:16.865 --> 00:01:18.783			
Also muss etwas nachgeben.	unidiomatic style	style	minor
28			
00:01:19.367 --> 00:01:22.954			
Und ich denke, wir alle wissen	awkward style punctuation	style linguistic conventions	neutral minor

dass wir keine weiteren Erden mehr bekommen, richtig?	grammar unidiomatic style cpl	linguistic conventions style design and markup	neutral neutral minor
29			
00:01:22.954 --> 00:01:26.625			
Also sind wir es zwangsläufig, die sich ändern müssen.	grammar awkward style cpl	linguistic conventions style design and markup	minor neutral minor
30			
00:01:26.958 --> 00:01:31.671			
Und wenn wir uns umsehen und zuhören	awkward style grammar	style linguistic conventions	minor minor
auf die internationalen Verhandlungen,	grammar	linguistic conventions	minor
31			
00:01:31.713 --> 00:01:35.050			
wenn wir die nationale Untätigkeit sehen, die zu oft vorkommt,	awkward style cpl	style design and markup	minor minor
32			
00:01:35.091 --> 00:01:38.762			
wird es viele Menschen geben, die sich fragen	awkward style punctuation cpl	style linguistic conventions design and markup	neutral minor minor
wie wir diese Herausforderung angehen wollen.	unidiomatic style punctuation cpl	style linguistic conventions design and markup	neutral minor minor
33			
00:01:38.762 --> 00:01:41.765			
Und sogar, ob wir erfolgreich sein werden.	grammar	linguistic conventions	minor
34			
00:01:41.806 --> 00:01:43.391			
Und ich verstehe es,	awkward style	style	neutral
35			
00:01:43.391 --> 00:01:46.186			
so viele Menschen auf der ganzen Welt	grammar awkward style	linguistic conventions style	minor neutral
die Hoffnung verlieren werden.	grammar	linguistic conventions	neutral
36			
00:01:46.936 --> 00:01:51.274			

Aber meine Botschaft heute ist: Es gibt Hoffnung,	unidiomatic style punctuation cpl	style linguistic conventions design and markup	neutral minor minor
und sie versteckt sich im Verborgenen.	awkward style	style	minor
37			
00:01:51.316 --> 00:01:54.277			
Und ich glaube, es gibt große			
Hoffnung in unseren Städten.	unidiomatic style	style	neutral
38			
00:01:54.861 --> 00:01:56.946			
Betrachten Sie also diese vier Zahlen.	unidiomatic style	style	neutral
39			
00:01:57.322 --> 00:01:58.615			
Drei,			
40			
00:01:58.657 --> 00:02:00.033			
55	punctuation	linguistic conventions	neutral
41			
00:02:00.075 --> 00:02:01.576			
75	punctuation	linguistic conventions	neutral
42			
00:02:01.618 --> 00:02:02.994			
80.			
43			
00:02:02.994 --> 00:02:07.499			
Städte bedecken weniger als drei Prozent	mistranslation	accuracy	minor
der Landfläche der Erde ein.			
44			
00:02:07.499 --> 00:02:10.502			
Wir haben also einen kleinen geografischen Fußabdruck.	cpl	design and markup	minor
45			
00:02:10.96 --> 00:02:13.672			
In der Tat, wenn man alle Städte	awkward style	style	minor
der Welt zusammen,	grammar	linguistic conventions	minor
46			
00:02:13.672 --> 00:02:15.715			
könnte man sie in Indien unterbringen.	unidiomatic style	style	neutral

47			
00:02:15.757 --> 00:02:19.344			
Und doch leben mehr als die Hälfte in Städten,	grammar grammar cpl	linguistic conventions linguistic conventions design and markup	minor minor minor
48			
00:02:19.344 --> 00:02:22.305			
55 Prozent der Weltbevölkerung.	punctuation unidiomatic style	linguistic conventions style	minor neutral
49			
00:02:22.639 --> 00:02:25.350			
Und wir erwarten, dass	awkward style	style	neutral
dass das auf zwei Drittel anwachsen wird	grammar grammar awkward style	linguistic conventions linguistic conventions style	minor minor neutral
50			
00:02:25.392 --> 00:02:27.394			
bis zur Mitte dieses Jahrhunderts.			
51			
00:02:27.852 --> 00:02:31.981			
Städte sind verantwortlich für etwa			
75 Prozent der CO2-Emissionen.	character formatting	design and markup	neutral
52			
00:02:32.44 --> 00:02:37.320			
Und wir sind auch große Emittenten von	unidiomatic style	style	minor
von Stickstoffdioxid und Methan.			
53			
00:02:37.779 --> 00:02:41.700			
Und Städte verbrauchen 80 Prozent			
der Energie der Welt.	unidiomatic style	style	neutral
54			
00:02:42.075 --> 00:02:43.743			
Aber denk mal darüber nach.	register unidiomatic style	style style	minor neutral
55			
00:02:44.244 --> 00:02:46.996			
Dass es die Eigenschaften von Städten sind.	punctuation grammar cpl	linguistic conventions linguistic conventions design and markup	minor minor minor

56			
00:02:47.038 --> 00:02:50.583			
ihre Reichweite, ihre Größe, ihre Dichte,	mistranslation awkward style	accuracy style	neutral neutral
57			
00:02:50.625 --> 00:02:56.256			
die große Nähe der Führung	unidiomatic style	style	minor
zu den Menschen, ihre Anpassungsfähigkeit			
58			
00:02:56.256 --> 00:02:59.467			
und ihre Fähigkeit zur Neuerfindung --	unidiomatic style	style	minor
59			
00:02:59.467 --> 00:03:03.179			
was bedeutet, dass wir tatsächlich planen können	grammar addition cpl	linguistic conventions accuracy design and markup	minor minor minor
planen können, diese Zahlen zu verwalten.	awkward style	style	minor
60			
00:03:04.013 --> 00:03:05.473			
Das heißt, durch unsere Städte,	unidiomatic style punctuation	style linguistic conventions	neutral minor
61			
00:03:05.515 --> 00:03:10.311			
können wir tatsächlich planen, mehr zu tun,	cpl	design and markup	minor
für mehr Menschen, mit weniger.	awkward style	style	minor
62			
00:03:10.979 --> 00:03:14.149			
Und deshalb sage ich, dass Städte	awkward style	style	neutral
eines der effektivsten Werkzeuge sind	punctuation	linguistic conventions	minor
63			
00:03:14.149 --> 00:03:18.486			
die wir zur Verfügung haben	punctuation awkward style	linguistic conventions style	minor neutral
um die Effizienz zu steigern	grammar awkward style	linguistic conventions style	minor minor
64			
00:03:18.486 --> 00:03:21.239			
in unseren Beziehungen mit dem Land,	mistranslation	accuracy	minor
65			

00:03:21.281 --> 00:03:23.199			
Energie und Abfall.			
66			
00:03:23.742 --> 00:03:25.326			
Durch unsere Städte,	punctuation grammar	linguistic conventions linguistic conventions	minor neutral
67			
00:03:25.368 --> 00:03:30.415			
können wir die Effizienz steigern	addition	accuracy	minor
von mehr Menschenleben schneller steigern	cpl	design and markup	minor
68			
00:03:30.457 --> 00:03:33.418			
als durch jede andere Form			
der menschlichen Organisation.	grammar	linguistic conventions	neutral
69			
00:03:33.96 --> 00:03:35.670			
Wir können also zum Beispiel,	punctuation unidiomatic style	linguistic conventions style	minor neutral
70			
00:03:35.712 --> 00:03:39.340			
mehr Menschen auf weniger Land unterbringen und beschäftigen,	unidiomatic style cpl	style design and markup	neutral minor
71			
00:03:39.382 --> 00:03:45.138			
Minimierung des Drucks auf die Zersiedelung der Landschaft	grammar punctuation cpl	linguistic conventions linguistic conventions design and markup	major minor minor
der dann um Land und Natur konkurriert,	grammar	linguistic conventions	minor
72			
00:03:45.138 --> 00:03:47.974			
während die Entfernungen minimiert werden	awkward style punctuation	style linguistic conventions	minor minor
die Menschen zurücklegen müssen	punctuation awkward style cps	linguistic conventions style design and markup	minor neutral minor
73			
00:03:47.974 --> 00:03:49.726			
um ihre Grundbedürfnisse zu befriedigen.			
74			

00:03:50.435 --> 00:03:55.315			
Durch Städte können wir Menschen dazu bringen	grammar awkward style cpl	linguistic conventions style design and markup	minor neutral minor
Energie gemeinsam nutzen, indem sie Gebäude gemeinsam nutzen	cpl	design and markup	minor
75			
00:03:55.356 --> 00:03:57.901			
und durch intelligente Innovationen	awkward style	style	minor
wie Wärmenetze.			
76			
00:03:58.401 --> 00:04:03.239			
Die Dichte in unseren Städten	unidiomatic style	style	neutral
macht öffentliche Verkehrsmittel besser zugänglich	unidiomatic style cpl	style design and markup	neutral minor
77			
00:04:03.281 --> 00:04:05.033			
und kostengünstiger.			
78			
00:04:05.7 --> 00:04:09.913			
Und durch unsere Städte können wir	awkward style	style	neutral
unsere Beziehung zur Energie verändern.	unidiomatic style	style	neutral
79			
00:04:09.913 --> 00:04:12.665			
Wir brauchen jetzt Energiesicherheit.	unidiomatic style	style	neutral
80			
00:04:12.665 --> 00:04:15.794			
Aber Städte bieten Märkte von solcher Größe	punctuation unidiomatic style cpl	linguistic conventions style design and markup	minor neutral minor
81			
00:04:15.835 --> 00:04:19.380			
dass sie die Investition in erneuerbare Energien	awkward style cpl	style design and markup	neutral minor
finanziell attraktiver machen.			
82			
00:04:19.798 --> 00:04:22.133			
Und denken Sie über	grammar awkward style	linguistic conventions style	minor neutral
die Möglichkeiten mit Abfall.			

83			
00:04:22.175 --> 00:04:24.803			
Wir können die Effizienz			
in der Sammlung	unidiomatic style	style	neutral
84			
00:04:24.844 --> 00:04:26.513			
und Verarbeitung von Abfällen	omission unidiomatic style	accuracy style	major neutral
85			
00:04:26.513 --> 00:04:31.267			
bei gleichzeitiger Einführung der Grundsätze	unidiomatic style cpl	style design and markup	neutral minor
der Kreislaufwirtschaft in großem Maßstab	punctuation	linguistic conventions	minor
86			
00:04:31.309 --> 00:04:33.853			
damit die Ressourcen wiederverwertet werden,	awkward style cpl	style design and markup	neutral minor
87			
00:04:33.895 --> 00:04:35.480			
Güter wiederverwendet werden			
88			
00:04:35.522 --> 00:04:38.983			
und nicht vermeidbare Abfälle	awkward style	style	minor
wird zur Energiegewinnung verarbeitet,	grammar	linguistic conventions	minor
89			
00:04:39.025 --> 00:04:41.903			
zum Beispiel Lebensmittelabfälle als Dünger.	awkward style cpl	style design and markup	neutral minor
90			
00:04:42.612 --> 00:04:45.490			
Man denke nur an das globale Potenzial	unidiomatic style	style	neutral
91			
00:04:45.532 --> 00:04:50.745			
eines weltweiten Netzwerks von Städten	punctuation	linguistic conventions	minor
das diese Art von Effizienz steigert	unidiomatic style grammar	style linguistic conventions	neutral minor
92			
00:04:50.787 --> 00:04:56.376			
für mehr als die Hälfte und bald zwei Drittel	cpl	design and markup	minor

der Weltbevölkerung.			
93			
00:04:57.085 --> 00:05:00.588			
Und hier ist die Hoffnung	awkward style punctuation	style linguistic convention	neutral minor
die ich eingangs erwähnte.			
94			
00:05:01.047 --> 00:05:02.841			
Man muss sich das nicht nur vorstellen.	unidiomatic style	style	neutral
95			
00:05:03.341 --> 00:05:05.635			
Von Freetown nach Los Angeles,	mistranslation	accuracy	minor
96			
00:05:05.635 --> 00:05:07.846			
von Kampala nach London,	mistranslation	accuracy	minor
97			
00:05:07.887 --> 00:05:10.598			
und in vielen, vielen Städten dazwischen,	punctuation	linguistic conventions	minor
98			
00:05:10.598 --> 00:05:14.936			
Bürgermeister, Stadtoberhäupter	grammar awkward style	linguistic conventions style	minor minor
werden aktiv und ergreifen Maßnahmen	punctuation	linguistic conventions	minor
99			
00:05:14.978 --> 00:05:17.397			
um sich den aktuellen Herausforderungen zu stellen.	cpl	design and markup	minor
100			
00:05:17.981 --> 00:05:22.402			
Nehmen wir also Malmö, eine Stadt	awkward style	style	neutral
mit knapp 350.000 Einwohnern.			
101			
00:05:23.069 --> 00:05:26.948			
Sie haben ein Wärmenetz entwickelt	unidiomatic style punctuation	style linguistic conventions	neutral minor
das durch Wärme gespeist wird	grammar	linguistic conventions	minor
102			
00:05:26.99 --> 00:05:29.325			

die durch verarbeitete Abfälle erzeugt wird.	awkward style cpl	style design and markup	minor minor
103			
00:05:29.909 --> 00:05:34.205			
Sie beabsichtigen, zu 100 Prozent	unidiomatic style punctuation	style linguistic conventions	neutral minor
mit erneuerbaren Energien	overtranslation	accuracy	minor
104			
00:05:34.247 --> 00:05:37.041			
oder rückgewonnener Wärme bis 2030.	omission	accuracy	major
105			
00:05:37.834 --> 00:05:42.714			
Oslo ist eine Stadt, die Subventionen für	grammar awkward style	linguistic conventions style	minor neutral
Elektrofahrzeuge und Ladestationen.			
106			
00:05:43.298 --> 00:05:46.426			
Sie haben ein Kreislaufsystem	mistranslation unidiomatic style	accuracy style	minor neutral
Abfallmanagementsystem eingeführt.			
107			
00:05:46.759 --> 00:05:48.678			
Sie haben eine Biogasanlage gekauft,	punctuation unidiomatic style	linguistic conventions style	minor neutral
108			
00:05:48.72 --> 00:05:52.891			
und fast 50 Prozent			
aller Lebensmittelabfälle werden recycelt.			
109			
00:05:53.433 --> 00:06:00.106			
(Beifall)			
110			
00:06:00.69 --> 00:06:04.402			
Singapur ist eine der dichtesten	unidiomatic style	style	neutral
Städte der Welt,			
111			
00:06:04.444 --> 00:06:06.696			

aber sie ist ein Musterbeispiel für grüne Planung.	unidiomatic style mistranslation cpl	style accuracy design and markup	neutral minor minor
112			
00:06:07.113 --> 00:06:11.034			
In den letzten Jahren haben sie	unidiomatic style	style	minor
riesige Süßwasserreserven			
113			
00:06:11.075 --> 00:06:14.329			
und städtische Gärten	omission punctuation	accuracy linguistic conventions	minor minor
die wie die Lungen der Stadt wirken.	unidiomatic style	style	neutral
114			
00:06:14.704 --> 00:06:17.415			
Und ich habe eine große Bewunderung			
Bewunderung für Bogota,	addition	accuracy	minor
115			
00:06:17.457 --> 00:06:19.959			
eine der dichtesten Städte	unidiomatic style	style	neutral
in Lateinamerika.			
116			
00:06:20.335 --> 00:06:22.962			
Sie haben ein	unidiomatic style	style	neutral
das Busschnellbahnsystem eingeführt.	mistranslation	accuracy	minor
117			
00:06:23.004 --> 00:06:25.423			
Sie machen das Gehen	unidiomatic style	style	neutral
und Radfahren zugänglicher,	punctuation	linguistic conventions	minor
118			
00:06:25.423 --> 00:06:31.137			
und haben heute eine der größten Flotten	awkward style	style	minor
von Elektrobussen in Lateinamerika.			
119			
00:06:32.055 --> 00:06:33.598			
Und während ich die Bühne habe,	unidiomatic style	style	minor
120			
00:06:33.64 --> 00:06:37.477			

lasst mich ein bisschen mit meiner Stadt angeben	register unidiomatic style cpl	style style design and markup	minor minor minor
über meine eigene Stadt, Bristol,	addition	accuracy	minor
121			
00:06:37.518 --> 00:06:41.898			
Heimat von 465.901 Menschen,	grammar	linguistic conventions	neutral
122			
00:06:41.94 --> 00:06:43.399			
einer von ihnen ist heute hier.	grammar	linguistic conventions	minor
123			
00:06:43.733 --> 00:06:47.487			
Wir haben einen fantastischen Ruf			
124			
00:06:47.487 --> 00:06:49.822			
als eine der grünsten Städte			
Städte in Europa.	unidiomatic style addition	style accuracy	neutral minor
125			
00:06:50.239 --> 00:06:52.784			
In Bristol haben wir eine Wohnungskrise.	unidiomatic style	style	neutral
126			
00:06:53.159 --> 00:06:55.328			
Wir müssen Wohnungen bauen.			
127			
00:06:55.745 --> 00:06:59.958			
Aber wir sind uns der Tatsache sehr bewusst	punctuation awkward style cpl	linguistic conventions style design and markup	minor neutral minor
dass die Art von Häusern, die wir bauen			
128			
00:06:59.958 --> 00:07:01.334			
und wo wir sie bauen	punctuation	linguistic conventions	minor
129			
00:07:01.376 --> 00:07:03.294			
wird eine der größten Determinanten sein	unidiomatic style grammar	style linguistic conventions	minor minor
130			

00:07:03.336 --> 00:07:06.547			
des Preises, den der Planet für unser			
für unser Wachstum.	addition omission	accuracy accuracy	minor minor
131			
00:07:07.09 --> 00:07:11.844			
Wir konzentrieren uns also auf die Bereitstellung von	awkward style cpl	style design and markup	minor minor
Netto-Null-Wohnungen mit höherer Dichte	grammar	linguistic conventions	minor
132			
00:07:11.886 --> 00:07:14.555			
auf altem Industriegelände			
in der Mitte der Stadt.	awkward style	style	neutral
133			
00:07:15.098 --> 00:07:18.768			
So können wir den Druck	awkward style	style	minor
den Druck der Zersiedelung zu verringern.	addition	accuracy	minor
134			
00:07:18.81 --> 00:07:21.771			
Es erlaubt uns, aktiven Verkehr einzubauen	mistranslation awkward style	accuracy style	major minor
135			
00:07:21.771 --> 00:07:24.607			
und die Abhängigkeit vom Auto zu verringern.	cpl	design and markup	minor
136			
00:07:25.191 --> 00:07:28.403			
Wir ergreifen sogar Maßnahmen			
gegen die Klimafolgen			
137			
00:07:28.403 --> 00:07:30.029			
der bescheidenen Toilette.	unidiomatic style	style	neutral
138			
00:07:30.53 --> 00:07:32.615			
In unserem gesamten öffentlichen Wohnungsbestand,	punctuation awkward style cpl	linguistic conventions style design and markup	minor neutral minor
139			
00:07:32.657 --> 00:07:34.075			
ersetzen wir die Badezimmer	mistranslation	accuracy	neutral

140			
00:07:34.117 --> 00:07:36.828			
und wir nutzen die Gelegenheit	awkward styke	style	neutral
um die Armaturen zu ersetzen	addition	accuracy	minor
141			
00:07:36.869 --> 00:07:39.038			
durch wassersparende Alternativen zu ersetzen,	grammar cpl	linguistic conventions design and markup	neutral minor
142			
00:07:39.08 --> 00:07:42.625			
wassersparende Duschen,	grammar	linguistic conventions	minor
Waschbecken und Wasserhähne,	punctuation	linguistic conventions	minor
143			
00:07:42.667 --> 00:07:45.003			
und mehr wassersparende Toiletten.	unidiomatic style	style	minor
144			
00:07:45.336 --> 00:07:48.339			
Also sind nicht alle Maßnahmen zum Klimawandel	unidiomatic style mistranslation cpl	style accuracy design and markup	neutral minor minor
ist voller Glamour, oder?	grammar	linguistic conventions	minor
145			
00:07:49.841 --> 00:07:52.260			
Aber wir Bürgermeister sind nicht nur auf	omission addition	accuracy accuracy	minor minor
146			
00:07:52.301 --> 00:07:54.929			
auf das, was innerhalb	awkward style	style	neutral
unserer Stadtgrenzen passiert.			
147			
00:07:55.388 --> 00:08:00.560			
Überall auf der Welt gibt es Bürgermeister, die	awkward style cpl	style design and markup	neutral minor
jenseits ihrer Befugnisse führen.	unidiomatic style	style	minor
148			
00:08:00.56 --> 00:08:03.146			
Sie kommen zusammen	unidiomatic style grammar	style linguistic conventions	neutral minor
in internationalen Netzwerken	punctuation	linguistic conventions	minor

149			
00:08:03.146 --> 00:08:05.857			
um harte Ziele für die Dekarbonisierung zu setzen	unidiomatic style grammar cpl	style linguistic conventions design and markup	neutral neutral minor
150			
00:08:05.898 --> 00:08:09.152			
für das sie sich gegenseitig	awkward style	style	minor
gegenseitig rechenschaftspflichtig sind.	addition	accuracy	minor
151			
00:08:09.152 --> 00:08:12.697			
Sie werden Hunderte	awkward style	style	minor
dieser städtischen Solidaritätsnetzwerke	omission	accuracy	minor
152			
00:08:12.697 --> 00:08:14.365			
auf der ganzen Welt.			
153			
00:08:14.824 --> 00:08:17.577			
Der Globale Konvent der Bürgermeister			
für Klima und Energie			
154			
00:08:17.577 --> 00:08:20.997			
ist ein Netzwerk von rund 12.000 Städten.			
155			
00:08:21.664 --> 00:08:24.500			
Sie haben sich gemeinsam verpflichtet			
verpflichtet, Maßnahmen zu ergreifen	addition awkward style punctuation cps	accuracy style linguistic convention design and markup	minor neutral minor minor
156			
00:08:24.542 --> 00:08:29.589			
um sicherzustellen, dass die Emissionen bis 2030	awkward style cpl	style design and markup	minor minor
fast zwei Gigatonnen niedriger sind			
157			
00:08:29.63 --> 00:08:33.217			
als sie es sonst wären	punctuation	linguistic conventions	minor
wenn wir so weitermachen wie bisher.	awkward style	style	neutral

158			
00:08:33.718 --> 00:08:35.136			
Und als Bürgermeister,	punctuation awkward style	linguistic conventions style	minor neutral
159			
00:08:35.178 --> 00:08:38.723			
nehmen wir auch verstärkt Einfluss auf	awkward style	style	neutral
internationalen Organisationen	grammar	linguistic conventions	minor
160			
00:08:38.723 --> 00:08:42.810			
und die globale Politik	punctuation mistranslation	linguistic conventions accuracy	minor neutral
die uns dabei unterstützen kann, Maßnahmen zu ergreifen.	awkward style cpl	style design and markup	neutral minor
161			
00:08:43.394 --> 00:08:46.898			
Die C40 ist ein Netzwerk von fast 100 Bürgermeistern	punctuation cpl	linguistic conventions design and markup	minor minor
162			
00:08:46.939 --> 00:08:51.152			
die die größten Städte der Welt			
größten globalen Städte.	omission addition	accuracy accuracy	major minor
163			
00:08:51.569 --> 00:08:54.614			
Städtediplomatie ist ein zentraler Bestandteil ihrer Arbeit.	awkward style cpl	style design and markup	neutral minor
164			
00:08:54.947 --> 00:08:58.409			
Mitglieder nehmen an nationalen	awkward style	style	neutral
und internationalen Verhandlungen	omission	accuracy	minor
165			
00:08:58.451 --> 00:09:02.497			
mit dem Ziel, Einfluss auf die Entscheidungsfindung	mistranslation awkward style cpl	accuracy style design and markup	minor neutral minor
Entscheidungsfindung und globale Verpflichtungen zu beeinflussen.	addition cpl cps	accuracy design and markup design and markup	minor minor minor

166			
00:09:02.955 --> 00:09:04.207			
Und lassen Sie mich nur sagen,	unidiomatic style punctuation	style linguistic conventions	neutral neutral
167			
00:09:04.207 --> 00:09:07.376			
was man mit Bürgermeistern bekommt, ist eine Verpflichtung	awkward style cpl	style design and markup	minor minor
168			
00:09:07.418 --> 00:09:13.925			
um sicherzustellen, dass diese globalen Verpflichtungen	cpl	design and markup	minor
von Worten in Taten umgesetzt werden.			
169			
00:09:14.383 --> 00:09:20.014			
Unsere Nähe zu unseren Bewohnern	unidiomatic style	style	neutral
bedeutet, dass wir unmittelbar rechenschaftspflichtig sind	awkward style punctuation cpl	style linguistic conventions design and markup	minor neutral minor
170			
00:09:20.056 --> 00:09:24.602			
um Veränderungen zu bewirken	punctuation	linguistic conventions	minor
die die Menschen sehen und erleben können.			
171			
00:09:26.104 --> 00:09:28.523			
Aber hier stoßen wir auf eine weitere Herausforderung.	awkward style cpl	style design and markup	neutral minor
172			
00:09:29.649 --> 00:09:35.363			
Wir werden nicht das weltweite Netz	awkward style grammar grammar	style linguistic conventions linguistic conventions	minor minor neutral
von effizienten Städten bekommen, das wir brauchen	unidiomatic style cpl	style design and markup	neutral minor
173			
00:09:35.404 --> 00:09:37.490			
ohne große Investitionen.	grammar	linguistic conventions	neutral
174			
00:09:37.907 --> 00:09:42.245			
Wir werden kein weltweites Netzwerk	omission	accuracy	minor

von dekarbonisierten Städten	grammar punctuation grammar	linguistic conventions linguistic conventions linguistic conventions	neutral minor neutral
175			
00:09:42.286 --> 00:09:43.955			
nur weil wir es brauchen,			
176			
00:09:43.996 --> 00:09:45.248			
wir es wollen	omission	accuracy	neutral
177			
00:09:45.248 --> 00:09:48.334			
oder weil wir blumige			
blumige Erklärungen darüber abgeben.			
178			
00:09:49.001 --> 00:09:53.172			
Wir werden sie nur bekommen	grammar punctuation grammar	linguistic conventions linguistic conventions linguistic conventions	minor minor neutral
wenn wir es planen und dann dafür bezahlen.	cpl	design and markup	minor
179			
00:09:54.298 --> 00:09:57.301			
Letztendlich kämpfen die Stadtverwaltungen in der ganzen Welt	mistranslation awkward style grammar cpl	accuracy style linguistic conventions design and markup	minor minor minor minor
um den Zugang zu erhalten	cps	design and markup	minor
180			
00:09:57.343 --> 00:10:02.014			
zu den Finanzmitteln, die sie brauchen	punctuation awkward style	linguistic conventions style	minor minor
um das volle Potenzial ihrer Stadt zu erschließen.	cpl	design and markup	minor
181			
00:10:02.64 --> 00:10:04.767			
Und dennoch, auch hier,	awkward style grammar punctuation	style linguistic conventions linguistic conventions	minor minor minor
warten wir nicht herum.			
182			
00:10:05.268 --> 00:10:07.145			

In Großbritannien bin ich Teil von etwas, das sich	undertranslation awkward style cpl	accuracy style design and markup	minor minor minor
183			
00:10:07.186 --> 00:10:09.480			
der UK Cities Climate			
Investitionskommission.	do not translate omission	accuracy accuracy	minor minor
184			
00:10:09.522 --> 00:10:13.192			
Es ist unser Ziel, die	awkward style	style	neutral
die größten Städte Großbritanniens in Kontakt zu bringen	addition grammar cpl	accuracy linguistic conventions design and markup	minor minor minor
185			
00:10:13.234 --> 00:10:17.280			
mit der Finanzierung, die sie brauchen	awkward style	style	neutral
damit dieses Potenzial freigesetzt werden kann.	awkward style cpl	style design and markup	neutral minor
186			
00:10:17.655 --> 00:10:23.452			
Wir haben Möglichkeiten zur Dekarbonisierung im Wert von 206 Milliarden Pfund identifiziert	awkward style grammar cpl	style linguistic conventions design and markup	neutral minor minor
an Dekarbonisierungsmöglichkeiten	addition	accuracy	minor
187			
00:10:23.452 --> 00:10:24.745			
im gesamten Vereinigten Königreich,	overtranslation cps	accuracy design and markup	neutral minor
188			
00:10:24.745 --> 00:10:28.666			
Nachrüstung, erneuerbare Energien,			
Umstellung der Flotte auf Elektroantrieb.	mistranslation	accuracy	minor
189			
00:10:29.292 --> 00:10:30.501			
Und wir stellen sicher	awkward style	style	neutral
190			
00:10:30.543 --> 00:10:34.589			
dass öffentliche und private Investoren			
sich dieser Möglichkeiten bewusst sind	grammar	linguistic conventions	minor

191			
00:10:34.589 --> 00:10:36.090			
im gesamten Vereinigten Königreich.			
192			
00:10:37.049 --> 00:10:38.593			
Also, hier ist die Sache.	unidiomatic style	style	minor
193			
00:10:39.51 --> 00:10:41.679			
Bürgermeister, Stadtoberhäupter,			
194			
00:10:41.721 --> 00:10:47.435			
wir haben keine Zeit für abstrakte Debatten	unidiomatic style cpl	style design and markup	neutral minor
oder blumige Erklärungen.			
195			
00:10:47.435 --> 00:10:50.980			
Unsere Bevölkerungen wollen heute Veränderungen.	unidiomatic style cpl	style design and markup	neutral minor
196			
00:10:51.022 --> 00:10:53.441			
Sie wollen den Wandel gestern.	unidiomatic style	style	minor
197			
00:10:53.441 --> 00:10:57.361			
Die Klimakrise, in der wir stecken	awkward style	style	minor
erfordert Führung.			
198			
00:10:57.403 --> 00:11:02.408			
Und die Bürgermeister, die ich überall auf der Welt treffe	awkward style cpl	style design and markup	minor minor
treten in diesen Bereich vor	mistranslation	accuracy	minor
199			
00:11:02.45 --> 00:11:03.701			
um den Moment zu treffen.			
200			
00:11:04.076 --> 00:11:09.040			
Wir wollen und brauchen nationale Regierungen	cpl	design and markup	minor
und internationale Organisationen	punctuation	linguistic conventions	minor

201			
00:11:09.081 --> 00:11:12.376			
die mit uns zusammenarbeiten, uns unterstützen,	mistranslation cpl	accuracy design and markup	minor minor
uns zu unterstützen.	grammar	linguistic conventions	minor
202			
00:11:12.376 --> 00:11:14.253			
Aber wir können nicht auf sie warten.			
203			
00:11:14.587 --> 00:11:16.589			
Die besten Wissenschaftler der Welt sagen uns	awkward style cpl	style design and markup	neutral minor
204			
00:11:16.631 --> 00:11:19.675			
wir haben zehn Jahre Zeit	punctuation grammar	linguistic conventions linguistic conventions	minor minor
um die Sache umzukehren.	unidiomatic style	style	minor
205			
00:11:20.259 --> 00:11:24.889			
Und wegen einer Geschichte der Untätigkeit,	awkward style cpl	style design and markup	minor minor
unzureichender Leistung,			
206			
00:11:24.931 --> 00:11:27.016			
Schwülstigkeit bei der Entscheidungsfindung,	punctuation cpl	linguistic conventions design and markup	minor minor
207			
00:11:27.016 --> 00:11:31.187			
Das bedeutet, dass wir gerade jetzt	grammar	linguistic conventions	minor
wir einige große Wetten eingehen müssen	awkward style	style	minor
208			
00:11:31.229 --> 00:11:34.357			
und wir müssen			
große Wetten auf Interventionen abschließen	punctuation cpl	linguistic conventions design and markup	minor minor
209			
00:11:34.398 --> 00:11:38.361			
die einen Wandel bewirken werden	grammar awkward style	linguistic conventions style	minor minor
in großem Umfang und Tempo.			

210			
00:11:38.819 --> 00:11:41.572			
Und ich denke, unsere Städte geben uns gute Chancen.	awkward style cpl	style design and markup	neutral minor
211			
00:11:42.281 --> 00:11:44.158			
Hier ist also mein Aufruf zum Handeln.	awkward style	style	neutral
212			
00:11:45.91 --> 00:11:52.041			
Wir müssen mit den Bürgermeistern der Welt zusammenarbeiten	awkward style cpl	style design and markup	neutral minor
um einen globalen Plan für Städte zu entwickeln.	cpl	design and markup	minor
213			
00:11:53.0 --> 00:11:59.382			
Es ist ein Plan, der die Dekarbonisierung	awkward style	style	minor
und Effizienz in bestehende Städte einbauen,	omission cpl	accuracy design and markup	minor minor
214			
00:11:59.423 --> 00:12:03.719			
sondern um sicherzustellen, dass die zukünftigen	mistranslation cpl	accuracy design and markup	minor minor
Prozesse der Urbanisierung	awkward style	style	minor
215			
00:12:03.719 --> 00:12:06.389			
die Effizienz der Stadt maximieren.	grammar	linguistic conventions	minor
216			
00:12:06.806 --> 00:12:08.641			
Und wenn ich global sage,	punctuation	linguistic conventions	neutral
217			
00:12:08.683 --> 00:12:10.142			
meine ich wirklich global.	punctuation unidiomatic style	linguistic conventions style	neutral neutral
218			
00:12:10.518 --> 00:12:13.938			
Es ist ein Plan, der über die	awkward style	style	minor
nationale Grenzen überschreiten muss.	addition	accuracy	minor
219			
00:12:14.355 --> 00:12:16.357			

Er muss im Globalen Norden stattfinden,	awkward style punctuation	style linguistic conventions	neutral minor
220			
00:12:16.357 --> 00:12:18.567			
und es muss im Globalen Süden sein,			
221			
00:12:18.609 --> 00:12:23.322			
wo 90 Prozent der zukünftigen Urbanisierung	unidiomatic style cpl	style design and markup	neutral minor
stattfinden wird.			
222			
00:12:24.031 --> 00:12:25.241			
Und es ist ein Plan	punctuation awkward style	linguistic conventions style	minor neutral
223			
00:12:25.241 --> 00:12:31.038			
der uns dazu bringen muss, über	awkward style addition	style accuracy	minor minor
unser enges nationales Eigeninteresse überschreiten.	grammar cpl	linguistic conventions design and markup	minor minor
224			
00:12:31.08 --> 00:12:35.543			
Wir müssen lernen, die Städte der Welt	unidiomatic style	style	neutral
als internationales Kapital sehen	grammar	linguistic conventions	minor
225			
00:12:35.543 --> 00:12:37.837			
und nicht als nationale Besitztümer.			
226			
00:12:38.212 --> 00:12:39.797			
Wir müssen zu der Einsicht kommen	awkward style punctuation	style linguistic conventions	neutral neutral
227			
00:12:39.797 --> 00:12:44.135			
dass Investitionen in die Steigerung der Effizienz	awkward style cpl	style design and markup	neutral minor
der Städte auf der Welt			
228			
00:12:44.135 --> 00:12:47.305			
der Schlüssel zu unserer Zukunft ist.	grammar	linguistic conventions	minor

229			
00:12:47.972 --> 00:12:51.309			
Es wird der Schlüssel sein	awkward style punctuation	style linguistic conventions	neutral minor
um unser Potenzial freizusetzen.	punctuation	linguistic conventions	neutral
230			
00:12:51.35 --> 00:12:54.770			
Und es ist eine Investition	awkward style	style	neutral
in unser globales Gemeinwohl.			
231			
00:12:56.314 --> 00:12:59.775			
Als ich gewählt wurde	grammar awkward style	linguistic conventions style	minor neutral
Bürgermeister von Bristol im Jahr 2016,			
232			
00:12:59.817 --> 00:13:06.115			
hatte ich eine begrenzte Vorstellung	unidiomatic style	style	neutral
über die globale Rolle der Stadt.			
233			
00:13:06.991 --> 00:13:10.453			
Und im weiteren Sinne,	awkward style punctuation	style linguistic conventions	minor minor
hatte ich eine begrenzte Wertschätzung	mistranslation	accuracy	minor
234			
00:13:10.453 --> 00:13:15.374			
für das Ausmaß der Verantwortung	punctuation	linguistic conventions	minor
die damit auf meinen Schultern lastete	grammar awkward style	linguistic conventions style	minor neutral
235			
00:13:15.374 --> 00:13:19.337			
als neu gewählter Bürgermeister			
einer britischen Großstadt.			
236			
00:13:20.338 --> 00:13:21.505			
In den Jahren seither,	awkward style punctuation	style linguistic conventions	minor minor
237			
00:13:21.505 --> 00:13:24.175			
habe ich gelernt, dass Städte,			

238			
00:13:24.216 --> 00:13:27.553			
die Art, wie sie geplant sind,	awkward style	style	minor
die Art und Weise, wie sie funktionieren,			
239			
00:13:27.595 --> 00:13:31.182			
die Art und Weise, wie sie wachsen			
und die Art und Weise, wie sie innovativ sind	cpl	design and markup	minor
240			
00:13:31.182 --> 00:13:35.311			
wird der Schlüssel dazu sein, ob wir	awkward style	style	minor
oder nicht erfolgreich sind	grammar mistranslation	linguistic conventions accuracy	minor minor
241			
00:13:35.311 --> 00:13:37.646			
diese Herausforderung anzunehmen	punctuation	linguistic conventions	minor
242			
00:13:37.688 --> 00:13:40.691			
um den globalen Klimawandel aufzuhalten.	awkward style	style	minor
243			
00:13:41.65 --> 00:13:46.322			
Wenn wir das volle Potenzial unserer Städte freisetzen	addition cpl	accuracy design and markup	minor minor
Potenzial unserer Städte erschließen,			
244			
00:13:46.322 --> 00:13:52.078			
können wir den Preis minimieren, den der Planet dafür zahlt	grammar addition cpl	linguistic conventions accuracy design and markup	minor minor minor
für die wachsende Zahl von Menschen.	awkward style	style	minor
245			
00:13:52.828 --> 00:13:58.959			
Ich denke, effiziente Städte könnten eines der	awkward style cpl	style design and markup	neutral minor
eines der effektivsten Werkzeuge sein, die wir haben.	addition awkward style cpl	accuracy style design and markup	minor neutral minor
246			
00:13:59.001 --> 00:14:01.837			

Ich bitte Sie also, mit uns zusammenzuarbeiten	awkward style cpl	style design and markup	neutral minor
um sie aufzubauen.			
247			
00:14:01.879 --> 00:14:03.089			
Ich danke Ihnen.			
248			
00:14:03.089 --> 00:14:08.928			
(Beifall)			

Annex 3

Quality Analysis Google			
Subtitle	**Error Type**	**Dimension**	**Severity Level**
1			
00:00:04.334 --> 00:00:08.630			
Ich habe eine der größten Ehrungen	unidiomatic style punctuation	style linguistic conventions	minor minor
2			
00:00:08.63 --> 00:00:11.049			
das glaube ich jeder Politiker haben könnte,	grammar unidiomatic style cpl	linguistic conventions style design and markup	minor minor minor
3			
00:00:11.091 --> 00:00:14.803			
indem ich zum Bürgermeister gewählt wurde	grammar grammar awkward style	linguistic conventions linguistic conventions style	minor minor neutral
des Ortes, an dem ich geboren wurde			
4			
00:00:14.844 --> 00:00:16.012			
und erzogen.	mistranslation	accuracy	neutral
5			
00:00:16.012 --> 00:00:18.056			
Das ist also Bristol in Großbritannien.	unidiomatic style	style	neutral
6			
00:00:18.056 --> 00:00:19.224			
Danke.			
7			
00:00:19.266 --> 00:00:21.518			
(Beifall)			
8			
00:00:21.56 --> 00:00:23.144			
Und es ist eine große Aufgabe.	awkward style	style	neutral
9			
00:00:23.186 --> 00:00:25.772			
Von Bildung bis Wohnen,	awkward style	style	minor

10			
00:00:25.814 --> 00:00:28.233			
von Budgets bis zur Müllabfuhr			
11			
00:00:28.275 --> 00:00:31.486			
zu Protesten zu Gegenprotesten,			
12			
00:00:31.528 --> 00:00:34.155			
Städte sind komplizierte Organismen.	grammar	linguistic conventions	minor
13			
00:00:34.197 --> 00:00:37.909			
Sie können turbulent sein,	punctuation	linguistic conventions	minor
und sie können voller Widersprüche sein.	awkward style	style	neutral
14			
00:00:38.743 --> 00:00:42.455			
Und als Bürgermeister bin ich der Verantwortliche	grammar awkward style cpl	linguistic conventions style design and markup	minor neutral minor
Person in Bristol.			
15			
00:00:42.455 --> 00:00:44.207			
Bei mir hört der Bock auf.	mistranslation	accuracy	major
16			
00:00:44.207 --> 00:00:46.167			
Und das ist manchmal sogar für Probleme	awkward style punctuation	style linguistic conventions	minor minor
17			
00:00:46.209 --> 00:00:51.089			
über die ich sehr wenig habe	grammar	linguistic conventions	minor
echte Kontrolle oder Macht.	mistranslation	accuracy	neutral
18			
00:00:51.131 --> 00:00:52.299			
Und das ist in Ordnung.	awkward style	style	neutral
19			
00:00:52.34 --> 00:00:53.466			
So ist das Leben.			
20			
00:00:53.842 --> 00:00:56.803			

Aber ich habe einen anderen Wahlkreis	awkward style	style	minor
wem ich Rechenschaft schulde,	grammar	linguistic conventions	minor
21			
00:00:56.845 --> 00:00:59.472			
und das ist einer, der in meiner Stadt ist,	undertranslation awkward style cpl	accuracy style design and markup	neutral neutral minor
22			
00:00:59.514 --> 00:01:02.267			
sondern reicht über meine Stadtgrenzen hinaus.	mistranslation grammar cpl	accuracy linguistic conventions design and markup	minor minor minor
23			
00:01:02.309 --> 00:01:03.852			
Und das ist der Planet	grammar awkward style	linguistic conventions style	neutral neutral
24			
00:01:03.893 --> 00:01:08.607			
und die 7,9 Milliarden anderen Menschen	unidiomatic style punctuation	style linguistic conventions	neutral minor
die für ihr Überleben darauf angewiesen sind.	awkward style cpl	style design and markup	neutral minor
25			
00:01:09.482 --> 00:01:13.778			
Wir haben uns in eine Situation gebracht	punctuation	linguistic conventions	minor
wo es 1,7 Erden dauern würde	mistranslation punctuation	accuracy linguistic conventions	major minor
26			
00:01:13.778 --> 00:01:16.823			
für unsere heutige Lebensweise	awkward style	style	major
nachhaltig sein.	grammar	linguistic conventions	minor
27			
00:01:16.865 --> 00:01:18.783			
Es muss also etwas gegeben werden.	mistranslation	accuracy	major
28			
00:01:19.367 --> 00:01:22.954			
Und ich denke, wir alle wissen es	awkward style punctuation	style linguistic conventions	neutral minor

Wir bekommen keine Erden mehr, richtig?	grammar grammar unidiomatic style	linguistic conventions linguistic conventions style	minor neutral neutral
29			
00:01:22.954 --> 00:01:26.625			
Also sind es zwangsläufig wir, die sich ändern müssen.	awkward style cpl	style design and markup	neutral minor
30			
00:01:26.958 --> 00:01:31.671			
Und während wir uns umschauen und zuhören	awkward style grammar	style linguistic conventions	minor minor
zu den internationalen Verhandlungen,	grammar	linguistic conventions	minor
31			
00:01:31.713 --> 00:01:35.050			
da wir zu oft auf nationale Untätigkeit blicken,	mistranslation cpl	accuracy design and markup	major minor
32			
00:01:35.091 --> 00:01:38.762			
da werden sich viele fragen	punctuation awkward style	linguistic conventions style	minor neutral
wie wir diese Herausforderung annehmen werden.	unidiomatic style punctuation cpl	style linguistic conventions design and markup	neutral minor minor
33			
00:01:38.762 --> 00:01:41.765			
Und selbst wenn wir erfolgreich sein werden.	mistranslation grammar cpl	accuracy linguistic conventions design and markup	major minor minor
34			
00:01:41.806 --> 00:01:43.391			
Und ich verstehe es,	awkward style	style	neutral
35			
00:01:43.391 --> 00:01:46.186			
so viele Menschen auf der ganzen Welt	grammar awkward style	linguistic conventions style	minor neutral
wird die Hoffnung verlieren.	grammar	linguistic conventions	minor
36			
00:01:46.936 --> 00:01:51.274			

Aber meine Botschaft heute ist, es gibt Hoffnung,	unidiomatic style punctuation punctuation cpl	style linguistic conventions linguistic conventions design and markup	neutral minor neutral minor
und es versteckt sich vor aller Augen.	grammar awkward style	linguistic conventions style	minor neutral
37			
00:01:51.316 --> 00:01:54.277			
Und ich glaube, es ist riesig	unidiomatic style grammar	style linguistic conventions	neutral minor
Hoffnung in unseren Städten.			
38			
00:01:54.861 --> 00:01:56.946			
Betrachten Sie also diese vier Zahlen.	unidiomatic style	style	neutral
39			
00:01:57.322 --> 00:01:58.615			
Drei,			
40			
00:01:58.657 --> 00:02:00.033			
55	punctuation	linguistic conventions	neutral
41			
00:02:00.075 --> 00:02:01.576			
75	punctuation	linguistic conventions	neutral
42			
00:02:01.618 --> 00:02:02.994			
80.			
43			
00:02:02.994 --> 00:02:07.499			
Städte nehmen weniger als drei Prozent ein	grammar	linguistic conventions	minor
der Landoberfläche der Erde.			
44			
00:02:07.499 --> 00:02:10.502			
Wir haben also einen kleinen geografischen Fußabdruck.	cpl	design and markup	minor
45			
00:02:10.96 --> 00:02:13.672			
In der Tat, wenn Sie alle Städte setzen	awkward style grammar	style linguistic conventions	minor minor
der Welt zusammen,			

46			
00:02:13.672 --> 00:02:15.715			
Sie könnten sie in Indien unterbringen.	grammar unidiomatic style	linguistic conventions style	minor neutral
47			
00:02:15.757 --> 00:02:19.344			
Und doch leben in Städten mehr als die Hälfte,	grammar cpl	linguistic conventions design and markup	minor minor
48			
00:02:19.344 --> 00:02:22.305			
55 Prozent der Weltbevölkerung.	punctuation unidiomatic style	linguistic conventions style	minor neutral
49			
00:02:22.639 --> 00:02:25.350			
Und wir antizipieren	awkward style punctuation	style linguistic conventions	neutral minor
das wird auf zwei Drittel anwachsen	awkward style grammar	style linguistic conventions	neutral minor
50			
00:02:25.392 --> 00:02:27.394			
bis Mitte dieses Jahrhunderts.			
51			
00:02:27.852 --> 00:02:31.981			
Die Städte sind für rund verantwortlich	grammar	linguistic conventions	minor
75 Prozent der CO2-Emissionen.	character formatting	design and markup	neutral
52			
00:02:32.44 --> 00:02:37.320			
Und wir sind auch erstaunliche Emittenten	unidiomatic style	style	minor
von Stickstoffdioxid und Methan.			
53			
00:02:37.779 --> 00:02:41.700			
Und Städte verbrauchen 80 Prozent			
der Weltenergie.	unidiomatic style	style	minor
54			
00:02:42.075 --> 00:02:43.743			
Aber denken Sie darüber nach.	unidiomatic style	style	neutral
55			
00:02:44.244 --> 00:02:46.996			

Dass es die Eigenschaften von Städten sind --	grammar cpl	linguistic conventions design and markup	minor minor
56			
00:02:47.038 --> 00:02:50.583			
ihre Reichweite, ihre Größe, ihre Dichte,	mistranslation awkward style	accuracy style	neutral neutral
57			
00:02:50.625 --> 00:02:56.256			
Nähe der Führung	awkward style	style	minor
an die Menschen, ihre Anpassungsfähigkeit			
58			
00:02:56.256 --> 00:02:59.467			
und ihre Fähigkeit zur Neuerfindung --	unidiomatic style	style	minor
59			
00:02:59.467 --> 00:03:03.179			
das heißt, wir können es tatsächlich	awkward style	style	minor
planen, diese Zahlen zu verwalten.			
60			
00:03:04.013 --> 00:03:05.473			
Das heißt durch unsere Städte,	unidiomatic style punctuation	style linguistic conventions	neutral minor
61			
00:03:05.515 --> 00:03:10.311			
wir können tatsächlich planen, mehr zu tun,	grammar cpl	linguistic conventions design and markup	minor minor
für mehr Menschen mit weniger.	awkward style	style	minor
62			
00:03:10.979 --> 00:03:14.149			
Und deshalb sage ich Städte	punctuation awkward style	linguistic conventions style	minor neutral
sind eines der effektivsten Werkzeuge	punctuation	linguistic conventions	minor
63			
00:03:14.149 --> 00:03:18.486			
wir zur Verfügung haben	grammar grammar awkward style	linguistic conventions linguistic conventions style	minor minor neutral
zur Effizienzsteigerung			
64			
00:03:18.486 --> 00:03:21.239			

in unsere Beziehungen zu Land,	mistranslation	accuracy	minor
65			
00:03:21.281 --> 00:03:23.199			
Energie und Abfall.			
66			
00:03:23.742 --> 00:03:25.326			
Durch unsere Städte,	punctuation grammar	linguistic conventions linguistic conventions	minor neutral
67			
00:03:25.368 --> 00:03:30.415			
Wir können die Effizienz steigern	grammar grammar	linguistic conventions linguistic conventions	minor minor
von mehr Menschenleben schneller			
68			
00:03:30.457 --> 00:03:33.418			
als durch jede andere Form			
der menschlichen Organisation.	grammar	linguistic conventions	neutral
69			
00:03:33.96 --> 00:03:35.670			
So können wir bspw.	register	style	neutral
70			
00:03:35.712 --> 00:03:39.340			
mehr Menschen auf weniger Land beherbergen und beschäftigen,	unidiomatic style cpl	style design and markup	neutral minor
71			
00:03:39.382 --> 00:03:45.138			
Minimierung des Drucks auf die Zersiedelung der Städte	grammar punctuation cpl	linguistic conventions linguistic conventions design and markup	major minor minor
die dann um Land, um Natur konkurriert,			
72			
00:03:45.138 --> 00:03:47.974			
bei gleichzeitiger Minimierung der Abstände	awkward style punctuation cpl	style linguistic conventions design and markup	minor minor minor
Menschen müssen reisen	grammar punctuation grammar awkward style	linguistic conventions linguistic conventions linguistic conventions style	minor minor minor minor

73			
00:03:47.974 --> 00:03:49.726			
um ihre Grundbedürfnisse zu befriedigen.			
74			
00:03:50.435 --> 00:03:55.315			
Durch Städte können wir Menschen haben	grammar mistranslation	linguistic conventions accuracy	minor major
Teilen von Energie durch Teilen von Gebäuden	grammar unidiomatic style cpl	linguistic conventions style design and markup	minor minor minor
75			
00:03:55.356 --> 00:03:57.901			
und durch smarte Innovationen			
wie Wärmenetze.			
76			
00:03:58.401 --> 00:04:03.239			
Die Dichte in unseren Städten	unidiomatic style	style	neutral
macht den öffentlichen Nahverkehr zugänglicher	overtranslation cpl	accuracy design and markup	neutral minor
77			
00:04:03.281 --> 00:04:05.033			
und kostengünstiger.			
78			
00:04:05.7 --> 00:04:09.913			
Und durch unsere Städte können wir uns verändern	grammar addition awkward style cpl	linguistic conventions accuracy style design and markup	minor minor neutral minor
unsere Beziehung zu Energie.			
79			
00:04:09.913 --> 00:04:12.665			
Wir brauchen jetzt Energiesicherheit.	unidiomatic style	style	neutral
80			
00:04:12.665 --> 00:04:15.794			
Aber Städte bieten Märkte dieser Größenordnung	mistranslation punctuation cpl	accuracy linguistic conventions design and markup	major minor minor
81			
00:04:15.835 --> 00:04:19.380			

dass sie in erneuerbare Energien investieren	awkward style grammar cpl	style linguistic conventions design and markup	minor minor minor
finanziell attraktiver.	grammar	linguistic conventions	minor
82			
00:04:19.798 --> 00:04:22.133			
Und denken Sie darüber nach	awkward style grammar	style linguistic conventions	minor minor
die Möglichkeiten mit Abfall.			
83			
00:04:22.175 --> 00:04:24.803			
Wir können Effizienz hebeln	mistranslation grammar	accuracy linguistic conventions	major minor
in die Sammlung			
84			
00:04:24.844 --> 00:04:26.513			
und Verarbeitung von Abfällen	unidiomatic style	style	neutral
85			
00:04:26.513 --> 00:04:31.267			
bei der Einführung der Prinzipien	mistranslation	accuracy	minor
der Kreislaufwirtschaft in großem Maßstab	punctuation	linguistic conventions	minor
86			
00:04:31.309 --> 00:04:33.853			
damit Ressourcen recycelt werden,	unidiomatic style	style	neutral
87			
00:04:33.895 --> 00:04:35.480			
Waren werden wiederverwendet	grammar unidiomatic style	linguistic conventions style	minor neutral
88			
00:04:35.522 --> 00:04:38.983			
und unvermeidbare Verschwendung	mistranslation	accuracy	major
zu Energie verarbeitet wird,	awkward style	style	neutral
89			
00:04:39.025 --> 00:04:41.903			
zum Beispiel Lebensmittelabfälle für Dünger.	awkward style cpl	style design and markup	neutral minor
90			
00:04:42.612 --> 00:04:45.490			

Denken Sie jetzt nur an das globale Potenzial	unidiomatic style cpl	style design and markup	neutral minor
91			
00:04:45.532 --> 00:04:50.745			
eines weltweiten Städtenetzes			
Skalierung dieser Arten von Effizienzen	mistranslation	accuracy	major
92			
00:04:50.787 --> 00:04:56.376			
für über die Hälfte und kommen auf zwei Drittel	mistranslation cpl	accuracy design and markup	major minor
der Weltbevölkerung.			
93			
00:04:57.085 --> 00:05:00.588			
Und hier ist die Hoffnung	punctuation awkward style	linguistic conventions style	minor neutral
Ich erwähnte am Anfang.	grammar grammar	linguistic conventions linguistic conventions	minor minor
94			
00:05:01.047 --> 00:05:02.841			
Das muss man sich nicht nur einbilden.	mistranslation unidiomatic style	accuracy style	minor neutral
95			
00:05:03.341 --> 00:05:05.635			
Von Freetown bis Los Angeles,			
96			
00:05:05.635 --> 00:05:07.846			
von Kampala nach London,	mistranslation	accuracy	minor
97			
00:05:07.887 --> 00:05:10.598			
und in vielen, vielen Städten dazwischen,	punctuation	linguistic conventions	minor
98			
00:05:10.598 --> 00:05:14.936			
Bürgermeister, Stadtoberhäupter	awkward style grammar	style linguistic conventions	minor minior
schreiten voran und handeln	punctuation	linguistic conventions	minior
99			
00:05:14.978 --> 00:05:17.397			
um die Herausforderung des Augenblicks zu meistern.	awkward style cpl	style design and markup	minor minor

100			
00:05:17.981 --> 00:05:22.402			
Nehmen Sie also Malmö, eine Stadt	unidiomatic style	style	minor
von knapp 350.000 Menschen.	grammar	linguistic conventions	neutral
101			
00:05:23.069 --> 00:05:26.948			
Sie haben ein Wärmenetz aufgebaut	unidiomatic style punctuation	style linguistic conventions	neutral minor
die durch Wärme gespeist wird	grammar grammar	linguistic conventions linguistic conventions	minor minor
102			
00:05:26.99 --> 00:05:29.325			
durch verarbeiteten Abfall erzeugt.	awkward style	style	minor
103			
00:05:29.909 --> 00:05:34.205			
Sie wollen 100 Prozent sein	grammar grammar	linguistic conventions linguistic conventions	minor minor
angetrieben durch Erneuerbare	unidiomatic style grammar	style linguistic conventions	neutral minor
104			
00:05:34.247 --> 00:05:37.041			
oder recycelte Wärme bis 2030.			
105			
00:05:37.834 --> 00:05:42.714			
Oslo ist eine subventionierende Stadt	mistranslation	accuracy	minor
Elektrofahrzeuge und Ladestationen.			
106			
00:05:43.298 --> 00:05:46.426			
Sie haben ein Rundschreiben eingeführt	unidiomatic style mistranslation grammar	style accuracy linguistic conventions	neutral major minor
Abfallwirtschaftssystem.			
107			
00:05:46.759 --> 00:05:48.678			
Sie haben eine Biogasanlage gekauft,	unidiomatic style punctuation	style linguistic conventions	neutral minor
108			
00:05:48.72 --> 00:05:52.891			
und fast 50 Prozent			

aller Lebensmittelabfälle werden recycelt.			
109			
00:05:53.433 --> 00:06:00.106			
(Beifall)			
110			
00:06:00.69 --> 00:06:04.402			
Singapur ist eines der dichtesten	grammar unidiomatic style	linguistic conventions style	minor neutral
Städte der Welt,			
111			
00:06:04.444 --> 00:06:06.696			
aber sie sind ein Modell der grünen Planung.	mistranslation unidiomatic style cpl	accuracy style design and markup	minor minor minor
112			
00:06:07.113 --> 00:06:11.034			
In den letzten Jahren haben sie eingeführt	unidiomatic style mistranslation grammar	style accuracy linguistic conventions	minor minor minor
riesige Süßwasserreserven			
113			
00:06:11.075 --> 00:06:14.329			
und Stadtgärten	punctuation	linguistic conventions	minor
die als Lungen der Stadt fungieren.	unidiomatic style	style	neutral
114			
00:06:14.704 --> 00:06:17.415	awkward style	style	minor
Und ich habe eine riesige Menge			
der Bewunderung für Bogota,			
115			
00:06:17.457 --> 00:06:19.959			
eine der dichtesten Städte	unidiomatic style	style	neutral
in Lateinamerika.			
116			
00:06:20.335 --> 00:06:22.962			
Sie haben eingeführt	unidiomatic style grammar	style linguistic conventions	neutral minor
das Bus-Rapid-Transit-System.	mistranslation	accuracy	minor
117			
00:06:23.004 --> 00:06:25.423			

Sie machen zu Fuß	mistranslation unidiomatic style	accuracy style	minor neutral
und Radfahren zugänglicher,			
118			
00:06:25.423 --> 00:06:31.137			
und haben heute eine der größten Flotten	awkward style	style	minor
von Elektrobussen in Lateinamerika.			
119			
00:06:32.055 --> 00:06:33.598			
Und während ich die Bühne habe,	unidiomatic style	style	minor
120			
00:06:33.64 --> 00:06:37.477			
lass mich ein bisschen angeben	register grammar unidiomatic style	style linguistic conventions style	minor minor minor
über meine eigene Stadt, Bristol,			
121			
00:06:37.518 --> 00:06:41.898			
Heimat für 465.901 Menschen,			
122			
00:06:41.94 --> 00:06:43.399			
einer von ihnen ist heute hier.	grammar	linguistic conventions	minor
123			
00:06:43.733 --> 00:06:47.487			
Wir haben einen fantastischen Ruf			
124			
00:06:47.487 --> 00:06:49.822			
dafür, einer der grünsten zu sein	grammar grammar mistranslation	linguistic conventions linguistic conventions accuracy	minor minor minor
Städte in Europa.			
125			
00:06:50.239 --> 00:06:52.784			
In Bristol haben wir eine Immobilienkrise.	unidiomatic style	style	neutral
126			
00:06:53.159 --> 00:06:55.328			
Wir müssen Häuser bauen.	mistranslation	accuracy	neutral
127			

00:06:55.745 --> 00:06:59.958			
Aber wir sind uns dessen sehr bewusst	punctuation awkward style	linguistic conventions style	minor minor
dass die Art von Häusern, die wir bauen			
128			
00:06:59.958 --> 00:07:01.334			
und wo wir sie bauen	punctuation	linguistic conventions	minor
129			
00:07:01.376 --> 00:07:03.294			
wird eine der größten Determinanten sein	grammar unidiomatic style	linguistic conventions style	minor minor
130			
00:07:03.336 --> 00:07:06.547			
des Preises, den der Planet zahlt	grammar	linguistic conventions	minor
für unser Wachstum.			
131			
00:07:07.09 --> 00:07:11.844			
Wir konzentrieren uns also auf die Lieferung	awkward style cpl	style design and markup	minor minor
Netto-Null-Häuser mit höherer Dichte	grammar	linguistic conventions	minor
132			
00:07:11.886 --> 00:07:14.555			
auf altem Industriegelände			
mitten in der Stadt.			
133			
00:07:15.098 --> 00:07:18.768			
Dadurch können wir entlasten	grammar awkward style	linguistic conventions style	minor minor
der Zersiedelungsdruck.	grammar	linguistic conventions	minor
134			
00:07:18.81 --> 00:07:21.771			
Es ermöglicht uns, aktives Reisen zu gestalten	awkward style mistranslation mistranslation cpl	style accuracy accuracy design and markup	neutral minor minor minor
135			
00:07:21.771 --> 00:07:24.607			
und Autoabhängigkeit ausgestalten.	mistranslation	accuracy	major
136			

00:07:25.191 --> 00:07:28.403			
Wir ergreifen sogar Maßnahmen			
zu den Klimafolgen	grammar	linguistic conventions	minor
137			
00:07:28.403 --> 00:07:30.029			
der bescheidenen Toilette.	unidiomatic style	style	neutral
138			
00:07:30.53 --> 00:07:32.615			
In unserem öffentlichen Wohnungsbestand			
139			
00:07:32.657 --> 00:07:34.075			
Wir ersetzen Badezimmer	grammar mistranslation	linguistic conventions accuracy	minor neutral
140			
00:07:34.117 --> 00:07:36.828			
und wir nutzen die Gelegenheit	awkward style	style	neutral
um die Beschläge auszutauschen	mistranslation grammar	accuracy linguistic conventions	major minor
141			
00:07:36.869 --> 00:07:39.038			
mit wassereffizienteren Alternativen,	unidiomatic style	style	minor
142			
00:07:39.08 --> 00:07:42.625			
wassersparendere Duschen,			
Waschbecken und Wasserhähne,	punctuation	linguistic conventions	minor
143			
00:07:42.667 --> 00:07:45.003			
und wassersparendere Toiletten.			
144			
00:07:45.336 --> 00:07:48.339			
Also nicht alle Klimaschutzmaßnahmen	unidiomatic style	style	minor
ist voller Glamour, oder?	grammar	linguistic conventions	minor
145			
00:07:49.841 --> 00:07:52.260			
Aber wir Bürgermeister sind nicht nur fokussiert	unidiomatic style grammar cpl	style linguistic conventions design and markup	neutral minor minor
146			

00:07:52.301 --> 00:07:54.929			
auf das, was drinnen passiert	grammar grammar	linguistic conventions linguistic conventions	minor minor
unsere Stadtgrenze.			
147			
00:07:55.388 --> 00:08:00.560			
Sie finden Bürgermeister auf der ganzen Welt	unidiomatic style cpl	style design and markup	minor minor
führen über ihre Autorität hinaus.	grammar	linguistic conventions	minor
148			
00:08:00.56 --> 00:08:03.146			
Sie kommen zusammen	unidiomatic style grammar	style linguistic conventions	neutral minor
in internationalen Netzwerken	punctuation	linguistic conventions	minor
149			
00:08:03.146 --> 00:08:05.857			
hartes Ziel für die Dekarbonisierung zu setzen	grammar unidiomatic style cpl	linguistic conventions style design and markup	minor neutral minor
150			
00:08:05.898 --> 00:08:09.152			
an denen sie sich festhalten	mistranslation	accuracy	minor
gegenseitig verantwortlich.			
151			
00:08:09.152 --> 00:08:12.697			
Sie werden Hunderte finden	awkward style grammar	style linguistic conventions	minor minor
dieser städtischen Solidaritätsnetzwerke			
152			
00:08:12.697 --> 00:08:14.365			
lebe gerade auf der Welt.	mistranslation	accuracy	major
153			
00:08:14.824 --> 00:08:17.577			
Der globale Konvent der Bürgermeister			
für Klima und Energie			
154			
00:08:17.577 --> 00:08:20.997			
ist ein Netzwerk aus rund 12.000 Städten.	grammar	linguistic conventions	neutral
155			

00:08:21.664 --> 00:08:24.500			
Sie haben ein Kollektiv gebildet	awkward style grammar	style linguistic conventions	minor minor
Verpflichtung, Maßnahmen zu ergreifen	punctuation	linguistic conventions	minor
156			
00:08:24.542 --> 00:08:29.589			
um sicherzustellen, dass 2030 Emissionen	awkward style grammar	style linguistic conventions	minor minor
sind fast zwei Gigatonnen niedriger			
157			
00:08:29.63 --> 00:08:33.217			
als sie es sonst wären	awkward style punctuation	style linguistic conventions	neutral minor
wenn wir so weitermachen wie bisher.			
158			
00:08:33.718 --> 00:08:35.136			
Und als Bürgermeister	awkward style	style	neutral
159			
00:08:35.178 --> 00:08:38.723			
Wir verstärken auch unseren Einfluss	grammar grammar	linguistic conventions linguistic conventions	minor minor
Internationale Organisationen			
160			
00:08:38.723 --> 00:08:42.810			
und die Weltpolitik	unidiomatic style punctuation mistranslation	style linguistic conventions accuracy	neutral minor neutral
die uns beim Handeln unterstützen können.	awkward style	style	minor
161			
00:08:43.394 --> 00:08:46.898			
Die C40 ist ein Netzwerk von fast 100 Bürgermeistern	punctuation cpl	linguistic conventions design and markup	minor minor
162			
00:08:46.939 --> 00:08:51.152			
repräsentiert die Welt	addition grammar grammar	accuracy linguistic conventions linguistic conventions	minor minor minor
größten Weltstädte.	unidiomatic style	style	minor
163			
00:08:51.569 --> 00:08:54.614			

Die Stadtdiplomatie steht im Mittelpunkt ihrer Arbeit.	awkward style cpl	style design and markup	neutral minor
164			
00:08:54.947 --> 00:08:58.409			
Mitglieder besuchen national	awkward style grammar	style linguistic conventions	neutral minor
und internationalen Verhandlungen	omission	accuracy	minor
165			
00:08:58.451 --> 00:09:02.497			
mit dem Ziel, Entscheidungen zu beeinflussen	awkward style cpl	style design and markup	neutral minor
Eingehen und globale Verpflichtungen.	mistranslation	accuracy	major
166			
00:09:02.955 --> 00:09:04.207			
Und lassen Sie mich nur sagen,	unidiomatic style punctuation	style linguistic conventions	neutral neutral
167			
00:09:04.207 --> 00:09:07.376			
Was Sie von Bürgermeistern bekommen, ist eine Verpflichtung	grammar awkward style cpl	linguistic conventions style design and markup	minor minor minor
168			
00:09:07.418 --> 00:09:13.925			
um sicherzustellen, dass diese globalen Verpflichtungen	grammar grammar cpl	linguistic conventions linguistic conventions design and markup	minor minor minor
werden aus Worten Taten.	grammar	linguistic conventions	minor
169			
00:09:14.383 --> 00:09:20.014			
Unsere Nähe zu unseren Bewohnern	unidiomatic style	style	neutral
bedeutet, dass wir sofort verantwortlich sind	punctuation unidiomatic style cpl	linguistic conventions style design and markup	neutral minor minor
170			
00:09:20.056 --> 00:09:24.602			
für die Bereitstellung von Änderungen	mistranslation punctuation	accuracy grammar	minor minor
die Menschen sehen und erleben können.			
171			
00:09:26.104 --> 00:09:28.523			

Aber hier stoßen wir auf eine andere Herausforderung.	awkward style cpl	style design and markup	neutral minor
172			
00:09:29.649 --> 00:09:35.363			
Wir werden das weltweite Netzwerk nicht bekommen	unidiomatic style grammar grammar cpl	style linguistic conventions linguistic conventions design and markup	neutral minor neutral minor
von effizienten Städten, die wir brauchen	awkward style	style	minor
173			
00:09:35.404 --> 00:09:37.490			
ohne große Investitionen.	grammar	linguistic conventions	neutral
174			
00:09:37.907 --> 00:09:42.245			
Wir werden kein weltweites Netzwerk bekommen	grammar grammar punctuation cpl	linguistic conventions linguistic conventions linguistic conventions design and markup	neutral minor minor minor
von dekarbonisierten Städten	grammar	linguistic conventions	neutral
175			
00:09:42.286 --> 00:09:43.955			
nur weil wir es brauchen,			
176			
00:09:43.996 --> 00:09:45.248			
wir wollen es	grammar omission	linguistic conventions accuracy	minor neutral
177			
00:09:45.248 --> 00:09:48.334			
oder weil wir machen	grammar unidiomatic style	linguistic conventions style	minor neutral
blumige Erklärungen darüber.			
178			
00:09:49.001 --> 00:09:53.172			
Wir werden sie nur bekommen	grammar punctuation grammar	linguistic conventions linguistic conventions linguistic conventions	minor minor neutral
wenn wir es planen und dann dafür bezahlen.	cpl	design and markup	minor
179			
00:09:54.298 --> 00:09:57.301			

Am Ende, Stadtführer auf der ganzen Welt	awkward style grammar punctuation	style linguistic conventions linguistic conventions	minor minor minor
kämpfen, um Zugang zu bekommen	awkward style	style	minor
180			
00:09:57.343 --> 00:10:02.014			
auf die Art der Finanzierung, die sie benötigen	cpl	design and markup	minor
um das volle Potenzial ihrer Stadt auszuschöpfen.	awkward style cpl	style design and markup	minor minor
181			
00:10:02.64 --> 00:10:04.767			
Und doch auch hier	awkward style	style	minor
wir warten nicht herum.	grammar	linguistic conventions	minor
182			
00:10:05.268 --> 00:10:07.145			
In Großbritannien bin ich Teil von etwas namens	undertranslation awkward style cpl	accuracy style design and markup	neutral minor minor
183			
00:10:07.186 --> 00:10:09.480			
das britische Städteklima	do not translate	accuracy	major
Anlagekommission.			
184			
00:10:09.522 --> 00:10:13.192			
Es ist unser Ziel zu setzen	awkward style grammar	style linguistic conventions	neutral minor
die größten Städte Großbritanniens in Kontakt	cpl	design and markup	minor
185			
00:10:13.234 --> 00:10:17.280			
mit der Finanzierung, die sie brauchen	awkward style	style	neutral
damit dieses Potenzial freigesetzt wird.	awkward style	style	neutral
186			
00:10:17.655 --> 00:10:23.452			
Wir haben einen Wert von 206 Milliarden Pfund identifiziert	awkward style grammar cpl	style linguistic conventions design and markup	neutral minor minor
von Dekarbonisierungsmöglichkeiten			
187			
00:10:23.452 --> 00:10:24.745			
in ganz Großbritannien,			

188			
00:10:24.745 --> 00:10:28.666			
Nachrüstung, Erneuerbare,	unidiomatic style	style	neutral
Umstellung des Fuhrparks auf Elektro.			
189			
00:10:29.292 --> 00:10:30.501			
Und wir sorgen dafür	awkward style	style	neutral
190			
00:10:30.543 --> 00:10:34.589			
dass öffentliche und private Investoren			
sind sich dieser Möglichkeiten bewusst	grammar	linguistic conventions	minor
191			
00:10:34.589 --> 00:10:36.090			
in ganz Großbritannien.	grammar undertranslation	linguistic convventions accuracy	minor neutral
192			
00:10:37.049 --> 00:10:38.593			
Also hier ist das Ding.	unidiomatic style	style	minor
193			
00:10:39.51 --> 00:10:41.679			
Bürgermeister, Stadtvorsteher,	mistranslation	accuracy	minor
194			
00:10:41.721 --> 00:10:47.435			
wir haben keine Zeit für abstrakte Debatten	unidiomatic style cpl	style design and markup	neutral minor
oder nur blumige Erklärungen.			
195			
00:10:47.435 --> 00:10:50.980			
Unsere Bevölkerung will heute Veränderungen.	unidiomatic style cpl	style design and markup	neutral minor
196			
00:10:51.022 --> 00:10:53.441			
Sie wollen Veränderung gestern.	unidiomatic style	style	minor
197			
00:10:53.441 --> 00:10:57.361			
Die Klimakrise, in der wir uns befinden	awkward style	style	minor
fordert Führung.			

198			
00:10:57.403 --> 00:11:02.408			
Und Bürgermeister, die ich auf der ganzen Welt treffe	awkward style cpl	style design and markup	minor minor
treten in diesen Raum ein	mistranslation	accuracy	major
199			
00:11:02.45 --> 00:11:03.701			
dem Moment begegnen.			
200			
00:11:04.076 --> 00:11:09.040			
Wir wollen und brauchen nationale Regierungen	cpl	design and markup	minor
und internationalen Organisationen	punctuation grammar	linguistic conventions linguistic conventions	minor minor
201			
00:11:09.081 --> 00:11:12.376			
mit uns zu arbeiten, uns zu unterstützen,	grammar mistranslation	linguistic conventions accuracy	minor minor
um uns zu unterstützen.	grammar	linguistic conventions	minor
202			
00:11:12.376 --> 00:11:14.253			
Aber wir können nicht auf sie warten.			
203			
00:11:14.587 --> 00:11:16.589			
Die besten Wissenschaftler der Welt sagen es uns	awkward style cpl	style design and markup	neutral minor
204			
00:11:16.631 --> 00:11:19.675			
Wir haben zehn Jahre	grammar punctuation	linguistic conventions linguistic conventions	minor minor
um das Ding umzudrehen.	unidiomatic style	style	minor
205			
00:11:20.259 --> 00:11:24.889			
Und aufgrund einer Geschichte der Untätigkeit,	awkward style cpl	style design and markup	minor minor
Unterleistung,	mistranslation	accuracy	minor
206			
00:11:24.931 --> 00:11:27.016			
Schwerfälligkeit in der Entscheidungsfindung,	punctuation cpl	linguistic conventions design and markup	minor minor

207			
00:11:27.016 --> 00:11:31.187			
das heißt jetzt gleich	grammar	linguistic conventions	minor
wir müssen einige große Wetten machen	awkward style	style	minor
208			
00:11:31.229 --> 00:11:34.357			
und wir müssen machen	grammar	linguistic conventions	minor
einige große Wetten auf Interventionen	punctuation	linguistic conventions	minor
209			
00:11:34.398 --> 00:11:38.361			
das bringt Veränderung	grammar grammar awkward style	linguistic conventions linguistic conventions style	minor minor minor
im Maßstab und Tempo.	mistranslation	accuracy	minor
210			
00:11:38.819 --> 00:11:41.572			
Und ich denke, unsere Städte bieten uns gute Chancen.	awkward style cpl	style design and markup	neutral minor
211			
00:11:42.281 --> 00:11:44.158			
Hier ist also mein Aufruf zum Handeln.	awkward style	style	neutral
212			
00:11:45.91 --> 00:11:52.041			
Wir müssen mit den Bürgermeistern der Welt zusammenarbeiten	awkward style cpl	style design and markup	neutral minor
einen globalen Plan für Städte zu entwickeln.	grammar cpl	linguistic conventions design and markup	minor minor
213			
00:11:53.0 --> 00:11:59.382			
Es ist ein Plan, der dekarbonisieren muss	awkward style grammar	style linguistic conventions	minor minor
und Effizienz in bestehende Städte einbauen,	mistranslation cpl	accuracy design and markup	minor minor
214			
00:11:59.423 --> 00:12:03.719			
aber um sicherzustellen, dass die Zukunft	mistranslation mistranstlation	accuracy accuracy	major minor
Prozesse der Urbanisierung	awkward style	style	minor
215			

00:12:03.719 --> 00:12:06.389			
Maximieren Sie die Effizienz der Stadt.	grammar grammar grammar	linguistic conventions linguistic conventions linguistic conventions	minor minor minor
216			
00:12:06.806 --> 00:12:08.641			
Und wenn ich global sage,	punctuation	linguistic conventions	neutral
217			
00:12:08.683 --> 00:12:10.142			
Ich meine global.	grammar punctuation unidiomatic style	linguistic conventions linguistic conventions style	minor neutral neutral
218			
00:12:10.518 --> 00:12:13.938			
Es ist ein Plan, der transzendieren muss	awkward style grammar	style linguistic conventions	minor minor
nationale Grenzen.			
219			
00:12:14.355 --> 00:12:16.357			
Es muss im globalen Norden sein,	awkward style punctuation	style linguistic conventions	neutral minor
220			
00:12:16.357 --> 00:12:18.567			
und es muss im globalen Süden sein,			
221			
00:12:18.609 --> 00:12:23.322			
wo 90 Prozent der zukünftigen Urbanisierung	unidiomatic style cpl	style design and markup	neutral minor
wird stattfinden.	grammar	linguistic conventions	minor
222			
00:12:24.031 --> 00:12:25.241			
Und es ist ein Plan	punctuation awkward style	linguistic conventions style	minor neutral
223			
00:12:25.241 --> 00:12:31.038			
das muss uns dazu bringen, darüber hinauszugehen	grammar grammar awkward style cpl	linguistic conventions linguistic conventions style design and markup	minor minor minor minor

unser engstirniges eigenstaatliches Interesse.	mistranslation cpl	accuracy design and markup	neutral minor
224			
00:12:31.08 --> 00:12:35.543			
Wir müssen kommen, um die Städte der Welt zu sehen	mistranslation grammar unidiomatic style cpl	accuracy linguistic conventions style design and markup	minor minor minor minor
als internationale Vermögenswerte			
225			
00:12:35.543 --> 00:12:37.837			
statt nationalem Besitz.	grammar	linguistic conventions	neutral
226			
00:12:38.212 --> 00:12:39.797			
Wir müssen kommen, um zu sehen	mistranslation puncuation	accuracy linguistic conventions	minor neutral
227			
00:12:39.797 --> 00:12:44.135			
dass in die gesteigerte Effizienz investiert wird	grammar grammar cpl	linguistic conventions linguistic conventions design and markup	minor minor minor
der Städte der Welt			
228			
00:12:44.135 --> 00:12:47.305			
ist der Schlüssel zu unserer Zukunft.	grammar	linguistic conventions	minor
229			
00:12:47.972 --> 00:12:51.309			
Es wird der Schlüssel sein	awkward style punctuation	style linguistic conventions	neutral minor
um unser Potenzial freizusetzen.	punctuation	linguistic conventions	neutral
230			
00:12:51.35 --> 00:12:54.770			
Und es ist eine Investition	awkward style	style	neutral
in unserem globalen Gemeinwohl.	grammar	linguistic conventions	minor
231			
00:12:56.314 --> 00:12:59.775			
Als ich gewählt wurde	grammar awkward style	linguistic conventions style	minor neutral
Bürgermeister von Bristol im Jahr 2016,			

232			
00:12:59.817 --> 00:13:06.115			
Ich hatte eine begrenzte Wertschätzung	grammar mistranslation	linguistic conventions accuracy	minor minor
der globalen Rolle der Stadt.	grammar	linguistic conventions	neutral
233			
00:13:06.991 --> 00:13:10.453			
Und als Erweiterung,	awkward style punctuation	style linguistic convention	minor minor
Ich hatte eine begrenzte Wertschätzung	grammar mistranslation	linguistic conventions accuracy	minor neutral
234			
00:13:10.453 --> 00:13:15.374			
der Verantwortungsebene	mistranslation punctuation	accuracy linguistic conventions	minor minor
das fiel also auf meine Schultern	grammar grammar awkward style	linguistic conventions linguistic conventions style	minor minor neutral
235			
00:13:15.374 --> 00:13:19.337			
als neu gewählter Bürgermeister			
einer britischen Großstadt.			
236			
00:13:20.338 --> 00:13:21.505			
In den Jahren seitdem	awkward style	style	minor
237			
00:13:21.505 --> 00:13:24.175			
Ich habe verstanden, dass Städte,	grammar	linguistic conventions	minor
238			
00:13:24.216 --> 00:13:27.553			
wie sie geplant sind,	awkward style	style	neutral
wie sie funktionieren,			
239			
00:13:27.595 --> 00:13:31.182			
wie sie wachsen			
und die Art und Weise, wie sie innovativ sind	cpl	design and markup	minor
240			
00:13:31.182 --> 00:13:35.311			

wird der Schlüssel dazu sein, ob wir es sind	grammar awkward style cpl	linguistic conventions style design and markup	minor minor minor
oder nicht erfolgreich sind	mistranslation	accuracy	minor
241			
00:13:35.311 --> 00:13:37.646			
diese Herausforderung anzunehmen	punctuation	linguistic conventions	minor
242			
00:13:37.688 --> 00:13:40.691			
um die Flut des globalen Klimawandels einzudämmen.	awkward style cpl	style design and markup	minor minor
243			
00:13:41.65 --> 00:13:46.322			
Wenn wir das voll freischalten können	awkward style grammar grammar	style linguistic conventions linguistic conventions	neutral minor minor
Potenzial unserer Städte,			
244			
00:13:46.322 --> 00:13:52.078			
Wir können den Preis minimieren, den der Planet zahlt	grammar grammar awkward style cpl	linguistic conventions linguistic conventions style design and markup	minor minor minor minor
dafür, dass Sie uns in unserer wachsenden Zahl aufnehmen.	mistranslation cpl	accuracy design and markup	minor minor
245			
00:13:52.828 --> 00:13:58.959			
Ich denke, effiziente Städte könnten eine sein	awkward style addition grammar cpl	style accuracy linguistic conventions design and markup	neutral minor minor minor
eines der effektivsten Tools, die wir haben.	unidiomatic style cpl	style design and markup	neutral minor
246			
00:13:59.001 --> 00:14:01.837			
Deshalb bitte ich Sie, mit uns zusammenzuarbeiten	awkward style cpl	style design and markup	neutral minor
sie zu bauen.	awkward style	style	neutral
247			
00:14:01.879 --> 00:14:03.089			
Danke schön.	unidiomatic style	style	neutral

248
00:14:03.089 --> 00:14:08.928
(Beifall)

Annex 4

MQM Scorecard: Full MQM-Core Error Typology with 4 Severity Levels (TED)							
	Error Severity Levels:	**Neutral**	**Minor**	**Major**	**Critical**	**Error Type Penalty Total**	
	Severity Multipliers:	0	1	5	25		
ET Nos	**Error Types**	**Error Counts**				**ET Weights**	**ETPTs**
1	**Terminology**	0	0	0	0	1,0	0,0
1.1	Inconsistent with terminology resource	0	0	0	0	1,0	0,0
1.2	Inconsistent use of terminology	0	0	0	0	1,0	0,0
1.3	Wrong term	0	0	0	0	1,0	0,0
2	**Accuracy**	0	0	0	0	1,0	0,0
2.1	Mistranslation	0	1	0	0	1,0	1,0
2.1.?	MT Hallucination	0	0	0	0	1,0	0,0
2.2	Overtranslation	0	0	0	0	1,0	0,0
2.3	Undertranslation	2	0	0	0	1,0	0,0
2.4	Addition	0	0	0	0	1,0	0,0
2.5	Omission	0	0	0	0	1,0	0,0
2.6	Do not translate (DNT)	0	0	0	0	1,0	0,0
2.7	Untranslated	0	0	0	0	1,0	0,0
3	**Linguistic conventions (was Fluency)**	0	0	0	0	1,0	0,0
3.1	Grammar	0	1	0	0	1,0	1,0
3.2	Punctuation	0	0	0	0	1,0	0,0
3.3	Spelling	0	3	0	0	1,0	3,0
3.4	Unintelligible	0	0	0	0	1,0	0,0
3.5	Character encoding	0	0	0	0	1,0	0,0
4	**Style**	0	0	0	0	1,0	0,0
4.1	Organizational style	0	0	0	0	1,0	0,0
4.2	Third-party style	0	0	0	0	1,0	0,0
4.3	Inconsistent with external reference	0	0	0	0	1,0	0,0
4.4	Register	0	2	0	0	1,0	2,0
4.5	Awkward style	0	1	0	0	1,0	1,0
4.6	Unidiomatic style	0	0	0	0	1,0	0,0
4.7	Inconsistent style	0	0	0	0	1,0	0,0
5	**Locale convention**	0	0	0	0	1,0	0,0
5.1	Number format	0	0	0	0	1,0	0,0
5.2	Currency format	0	0	0	0	1,0	0,0
5.3	Measurement format	0	0	0	0	1,0	0,0
5.4	Time format	0	0	0	0	1,0	0,0
5.5	Date format	0	0	0	0	1,0	0,0
5.6	Address format	0	0	0	0	1,0	0,0
5.7	Telephone format	0	0	0	0	1,0	0,0
5.8	Shortcut key	0	0	0	0	1,0	0,0
6	**Audience appropriateness**	0	0	0	0	1,0	0,0
6.1	Culture-specific reference	0	0	0	0	1,0	0,0
7	**Design and markup**	0	0	0	0	1,0	0,0
7.1	Character formatting	1	0	0	0	1,0	0,0
7.2	Layout	0	0	0	0	1,0	0,0
7.3	Markup tag	0	0	0	0	1,0	0,0
7.4	Truncation/text expansion	0	0	0	0	1,0	0,0
7.4.1	Too many characters per line	0	0	0	0	1,0	0,0
7.4.2	Too many characters per second	0	0	0	0	1,0	0,0
7.5	Missing text	0	0	0	0	1,0	0,0
7.6	Link/cross-reference	0	0	0	0	1,0	0,0
8	**Custom**	0	0	0	0	1,0	0,0
					Absolute Penalty Total (APT):		8,00
	Evaluation Word Count (EWC):	1903			**Per-Word Penalty Total (PWPT):**		0,0042
	Reference Word Count (RWC):	1000			**Overall Normed Penalty Total (ONPT):**		4,20
	Penalty Scaler (PS):	1,00			**Overall Quality Score (OQS):**		99,58
	Max. Score Value (MSV):	100,00					
					Overall Quality Fraction (OQF):		1,00

Annex 5

MQM Scorecard: Full MQM-Core Error Typology with 4 Severity Levels (DeepL)							
	Error Severity Levels:	Neutral	Minor	Major	Critical	Error Type Penalty Total	
	Severity Multipliers:	0	1	5	25		
ET Nos	Error Types	Error Counts				ET Weights	ETPTs
1	**Terminology**	0	0	0	0	1,0	0,0
1.1	Inconsistent with terminology resource	0	0	0	0	1,0	0,0
1.2	Inconsistent use of terminology	0	0	0	0	1,0	0,0
1.3	Wrong term	0	0	0	0	1,0	0,0
2	**Accuracy**	0	0	0	0	1,0	0,0
2.1	Mistranslation	3	19	1	0	1,0	24,0
2.1.?	MT Hallucination	0	0	0	0	1,0	0,0
2.2	Overtranslation	1	1	0	0	1,0	1,0
2.3	Undertranslation	1	1	0	0	1,0	1,0
2.4	Addition	0	20	0	0	1,0	20,0
2.5	Omission	1	8	3	0	1,0	23,0
2.6	Do not translate (DNT)	0	1	0	0	1,0	1,0
2.7	Untranslated	0	0	0	0	1,0	0,0
3	**Linguistic conventions (was Fluency)**	0	0	0	0	1,0	0,0
3.1	Grammar	13	51	1	0	1,0	56,0
3.2	Punctuation	8	57	0	0	1,0	57,0
3.3	Spelling	0	0	0	0	1,0	0,0
3.4	Unintelligible	0	0	0	0	1,0	0,0
3.5	Character encoding	0	0	0	0	1,0	0,0
4	**Style**	0	0	0	0	1,0	0,0
4.1	Organizational style	0	0	0	0	1,0	0,0
4.2	Third-party style	0	0	0	0	1,0	0,0
4.3	Inconsistent with external reference	0	0	0	0	1,0	0,0
4.4	Register	0	2	0	0	1,0	2,0
4.5	Awkward style	63	46	0	0	1,0	46,0
4.6	Unidiomatic style	54	16	0	0	1,0	16,0
4.7	Inconsistent style	0	0	0	0	1,0	0,0
5	**Locale convention**	0	0	0	0	1,0	0,0
5.1	Number format	0	0	0	0	1,0	0,0
5.2	Currency format	0	0	0	0	1,0	0,0
5.3	Measurement format	0	0	0	0	1,0	0,0
5.4	Time format	0	0	0	0	1,0	0,0
5.5	Date format	0	0	0	0	1,0	0,0
5.6	Address format	0	0	0	0	1,0	0,0
5.7	Telephone format	0	0	0	0	1,0	0,0
5.8	Shortcut key	0	0	0	0	1,0	0,0
6	**Audience appropriateness**	0	0	0	0	1,0	0,0
6.1	Culture-specific reference	0	0	0	0	1,0	0,0
7	**Design and markup**	0	0	0	0	1,0	0,0
7.1	Character formatting	1	0	0	0	1,0	0,0
7.2	Layout	0	0	0	0	1,0	0,0
7.3	Markup tag	0	0	0	0	1,0	0,0
7.4	Truncation/text expansion	0	0	0	0	1,0	0,0
7.4.1	Too many characters per line	0	82	0	0	1,0	82,0
7.4.2	Too many characters per second	0	5	0	0	1,0	5,0
7.5	Missing text	0	0	0	0	1,0	0,0
7.6	Link/cross-reference	0	0	0	0	1,0	0,0
8	**Custom**	0	0	0	0	1,0	0,0
					Absolute Penalty Total (APT):		334,00
	Evaluation Word Count (EWC):	1903			**Per-Word Penalty Total (PWPT):**		0,1755
	Reference Word Count (RWC):	1000			**Overall Normed Penalty Total (ONPT):**		175,51
	Penalty Scaler (PS):	1,00			**Overall Quality Score (OQS):**		82,45
	Max. Score Value (MSV):	100,00					
					Overall Quality Fraction (OQF):		0,82

Annex 6

MQM Scorecard: Full MQM-Core Error Typology with 4 Severity Levels (Google Translate)							
	Error Severity Levels:	Neutral	Minor	Major	Critical	Error Type Penalty Total	
	Severity Multipliers:	0	1	5	25		
ET Nos	Error Types	Error Counts				ET Weights	ETPTs
1	**Terminology**	0	0	0	0	1,0	0,0
1.1	Inconsistent with terminology resource	0	0	0	0	1,0	0,0
1.2	Inconsistent use of terminology	0	0	0	0	1,0	0,0
1.3	Wrong term	0	0	0	0	1,0	0,0
2	**Accuracy**	0	0	0	0	1,0	0,0
2.1	Mistranslation	8	26	18	0	1,0	116,0
2.1.?	MT Hallucination	0	0	0	0		
2.2	Overtranslation	1	0	0	0	1,0	0,0
2.3	Undertranslation	3	0	0	0	1,0	0,0
2.4	Addition	0	3	0	0	1,0	3,0
2.5	Omission	1	1	0	0	1,0	1,0
2.6	Do not translate (DNT)	0	0	1	0	1,0	5,0
2.7	Untranslated	0	0	0	0	1,0	0,0
3	**Linguistic conventions (was Fluency)**	0	0	0	0	1,0	0,0
3.1	Grammar	13	128	1	0	1,0	133,0
3.2	Punctuation	9	48	0	0	1,0	48,0
3.3	Spelling	0	0	0	0	1,0	0,0
3.4	Unintelligible	0	0	0	0	1,0	0,0
3.5	Character encoding	0	0	0	0	1,0	0,0
4	**Style**	0	0	0	0	1,0	0,0
4.1	Organizational style	0	0	0	0	1,0	0,0
4.2	Third-party style	0	0	0	0	1,0	0,0
4.3	Inconsistent with external reference	0	0	0	0	1,0	0,0
4.4	Register	1	1	0	0	1,0	1,0
4.5	Awkward style	51	45	1	0	1,0	50,0
4.6	Unidiomatic style	45	21	0	0	1,0	21,0
4.7	Inconsistent style	0	0	0	0	1,0	0,0
5	**Locale convention**	0	0	0	0	1,0	0,0
5.1	Number format	0	0	0	0	1,0	0,0
5.2	Currency format	0	0	0	0	1,0	0,0
5.3	Measurement format	0	0	0	0	1,0	0,0
5.4	Time format	0	0	0	0	1,0	0,0
5.5	Date format	0	0	0	0	1,0	0,0
5.6	Address format	0	0	0	0	1,0	0,0
5.7	Telephone format	0	0	0	0	1,0	0,0
5.8	Shortcut key	0	0	0	0	1,0	0,0
6	**Audience appropriateness**	0	0	0	0	1,0	0,0
6.1	Culture-specific reference	0	0	0	0	1,0	0,0
7	**Design and markup**	0	0	0	0	1,0	0,0
7.1	Character formatting	1	0	0	0	1,0	0,0
7.2	Layout	0	0	0	0	1,0	0,0
7.3	Markup tag	0	0	0	0	1,0	0,0
7.4	Truncation/text expansion	0	0	0	0	1,0	0,0
7.4.1	Too many characters per line	0	71	0	0	1,0	71,0
7.4.2	Too many characters per second	0	0	0	0	1,0	0,0
7.5	Missing text	0	0	0	0	1,0	0,0
7.6	Link/cross-reference	0	0	0	0	1,0	0,0
8	**Custom**	0	0	0	0	1,0	0,0
						Absolute Penalty Total (APT):	449,00
	Evaluation Word Count (EWC):	1903				**Per-Word Penalty Total (PWPT):**	0,2359
	Reference Word Count (RWC):	1000				**Overall Normed Penalty Total (ONPT):**	235,94
	Penalty Scaler (PS):	1,00				**Overall Quality Score (OQS):**	76,41
	Max. Score Value (MSV):	100,00					
						Overall Quality Fraction (OQF):	0,76

Annex 7

Umfrage zu Part 1

Beantworten Sie bitte die folgenden Fragen zum Inhalt des Videos sowie zu Ihrer Einstellung zu den Untertiteln des Videos.

Es folgen fünf Multiple-Choice-Fragen zum Inhalt des Videos. Bei jeder Frage ist nur eine der vier Antwortmöglichkeiten richtig.

1.

Welche Aussage über den Sprecher ist richtig?

- (•) Er ist in Bristol aufgewachsen und zurzeit Bürgermeister dieser Stadt.
- ( ) Er ist Politiker in Großbritannien und hat vor kurzem eine Reise nach Bristol unternommen.
- ( ) Er lebt erst seit einigen Jahren in Bristol und wurde vor kurzem Bürgermeister.
- ( ) Er kandidiert für das Amt des Bürgermeisters von Bristol.

2.

Wie beschreibt der Sprecher seinen Job?

- ( ) Es stresst ihn, dass er für so viele verschiedene Bereiche verantwortlich ist.
- (•) Er ist für viele verschiedene Bereiche verantwortlich, auch wenn er nicht immer Kontrolle über sie hat.
- ( ) Er kritisiert die große Verantwortung, die er in seinem Beruf tragen muss.
- ( ) Er hat über fast keine Bereiche, für die er verantwortlich ist, Kontrolle und möchte seinen Verantwortungsbereich daher genauer definieren.

3.

Welche Botschaft versucht der Sprecher in diesem Video zu vermitteln?

- ( ) Der Klimawandel kann in den Städten kaum noch gestoppt werden.
- (•) Man soll nicht die Hoffnung verlieren, dass eine Kehrtwende in Sachen Nachhaltigkeit noch möglich ist.
- ( ) So wie viele Menschen verliert der Sprecher die Hoffnung, dass der Klimawandel gestoppt werden kann.
- ( ) Er fühlt sich machtlos, weil auf nationaler Ebene so wenig gegen den Klimawandel unternommen wird.

4.

Was beträgt laut dem Sprecher 75 %?

- Auf 75 % der weltweiten Landfläche sind Großstädte zu finden.
- Städte sind für rund 75 % der CO2-Emissionen verantwortlich.
- In Städten wird rund 75 % der weltweit verbrauchten Energie konsumiert.
- 75 % der Weltbevölkerung lebt in Städten.

5.

Warum sind Städte laut dem Sprecher eines der effektivsten Werkzeuge für mehr Effizienz hinsichtlich Boden, Energie und Müll?

- In Städten gibt es mehr Infrastruktur, die modernisiert werden kann.
- Aufgrund der Merkmale von Städten kann mit weniger Ressourcen für eine größere Anzahl an Menschen mehr geleistet werden.
- Mehr Menschen sind willig ihren Alltag zu ändern.
- Die öffentlichen Verkehrsmittel bieten großes Einsparungspotenzial.

6.

Geben Sie bitte an, wie sehr Sie den folgenden Aussagen auf einer Skala von 1 bis 5 zustimmen.

1 = stimme überhaupt nicht zu
2 = stimme nicht zu
3 = neutral
4 = stimme zu
5 = stimme voll und ganz zu

Ich habe das Gefühl, den Inhalt des Videos im Großen und Ganzen verstanden zu haben.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Die Untertitel sind leicht zu verstehen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Es erfordert keine große Anstrengung die Untertitel zu lesen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Die Untertitel waren angenehm zu lesen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Ich bin mit der Qualität der Untertitel zufrieden.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

7.

Gibt es etwas, das Sie zu den Untertiteln noch sagen möchten?

Umfrage zu Part 2

Beantworten Sie bitte die folgenden Fragen zum Inhalt des Videos sowie zu Ihrer Einstellung zu den Untertiteln des Videos.

Es folgen fünf Multiple-Choice-Fragen zum Inhalt des Videos. Bei jeder Frage ist nur eine der vier Antwortmöglichkeiten richtig.

1.

Als Beispiel für welche Aussage nennt der Sprecher die Städte Malmö, Oslo, Singapur und Bogota?

- ○ Der Sprecher möchte Projekte, die in diesen Städten umgesetzt wurden, auch in Bristol umsetzen.
- ○ Die Bürgermeister dieser Städte haben eine Organisation für Umweltschutz gegründet.
- ○ Die Bürgermeister vieler Städte sehen es nicht in ihrer Verantwortung, Umweltschutzprojekte umzusetzen.
- ◉ Bürgermeister in vielen Städten konnten schon wichtige Schritte in Richtung Nachhaltigkeit machen.

2.

Welchen Ruf hat Bristol laut dem Sprecher?

- ○ Bristol hat einen überdurchschnittlich hohen CO2-Ausstoß.
- ◉ Bristol ist eine der grünsten Städte Europas.
- ○ Bristol wird seinem guten Ruf nicht gerecht.
- ○ Bristol ist eine sehr innovative Stadt.

3.

Wie möchte Bristol der Immobilienkrise entgegenwirken und dabei auf Nachhaltigkeit achten?

- ○ Das Bauen von Netto-Null-Häusern wird mit Förderungen unterstützt.
- ◉ Auf altem Industriegelände sollen Netto-Null-Häuser gebaut werden.
- ○ Bereits existierende Wohnhäuser werden umgebaut.
- ○ Alte Fabriken sollen in Wohnhäuser umgebaut werden.

4.

Wie werden existierende städtische Wohnhäuser in Bristol modernisiert?

- ○ Wasser wird durch erneuerbare Energie gewärmt.
- ○ Der Wasserverbrauch in Bädern wird kontrolliert.

- [x] In Bädern werden wassersparende Alternativen eingesetzt.
- [ ] Alle Toiletten, die älter als 20 Jahre sind, müssen durch wassersparende Modelle ersetzt werden.

5.

Was ist das Ziel des Globalen Konvents der Bürgermeister für Klima und Energie?

- [ ] Die teilnehmenden Bürgermeister tauschen sich mit Klimaaktivisten aus.
- [ ] Die teilnehmenden Bürgermeister haben sich zum Ziel gesetzt, ihre Städte bis zum Jahr 2030 klimaneutral zu machen.
- [ ] Die teilnehmenden Bürgermeister reisen zu internationalen Verhandlungen, um Einfluss auf die dort getroffenen Entscheidungen zu nehmen.
- [x] Die teilnehmenden Bürgermeister sind die Verpflichtung eingegangen, die Treibhausgase in ihren Städten bis 2030 zu senken.

6.

Geben Sie bitte an, wie sehr Sie den folgenden Aussagen auf einer Skala von 1 bis 5 zustimmen.

1 = stimme überhaupt nicht zu
2 = stimme nicht zu
3 = neutral
4 = stimme zu
5 = stimme voll und ganz zu

Ich habe das Gefühl, den Inhalt des Videos im Großen und Ganzen verstanden zu haben.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Die Untertitel sind leicht zu verstehen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Es erfordert keine große Anstrengung die Untertitel zu lesen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Die Untertitel waren angenehm zu lesen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Ich bin mit der Qualität der Untertitel zufrieden.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

7.

Gibt es etwas, das Sie zu den Untertiteln noch sagen möchten?

Umfrage zu Part 3

Beantworten Sie bitte die folgenden Fragen zum Inhalt des Videos sowie zu Ihrer Einstellung zu den Untertiteln des Videos.

Es folgen fünf Multiple-Choice-Fragen zum Inhalt des Videos. Bei jeder Frage ist nur eine der vier Antwortmöglichkeiten richtig.

1.

Von welchen Herausforderungen für Bürgermeister erzählt der Sprecher zu Beginn des Abschnitts?

- ○ Es gibt nicht genug öffentliche Reden und Veranstaltungen, in denen auf die Wichtigkeit der Treibhausgasverringerung aufmerksam gemacht wird.
- ○ Aufgrund von nicht ausreichenden finanziellen Mitteln konnte er in Bristol noch keine Projekte für den Klimaschutz umsetzen.
- ○ Die neuen Technologien, welche die Verringerung der Treibhausgase ermöglichen, stecken noch in den Kinderschuhen.
- ◉ Es ist nicht leicht, an die finanziellen Mittel für Projekte zur Treibhausgasverringerung zu kommen.

2.

Welches Ziel hat die UK Cities Climate Investment Commission?

- ○ Sie unterstützt Forscher in Großbritannien bei der Entwicklung neuer Umwelttechnologien.
- ◉ Sie verhilft den größten Städten Großbritanniens zu den notwendigen finanziellen Mitteln, um die Dekarbonisierung voranzutreiben.
- ○ Sie organisiert Treffen zwischen den Bürgermeistern der größten Städte Großbritanniens, um sich gegenseitig zu vernetzen und bei der Umsetzung von Maßnahmen zur Dekarbonisierung zu unterstützen.
- ○ Sie unterstützt kleine Städte in Großbritannien bei Projekten zur Dekarbonisierung.

3.

Was sagt der Sprecher über den zeitlichen Horizont der Umkehr der Klimakrise?

- ○ Man muss auch außerhalb der Städte Veränderungen einleiten, um die Kehrtwende in den nächsten 10 Jahren zu schaffen.
- ○ Der Sprecher befürchtet, dass auch in den nächsten 10 Jahren Untätigkeit die Kehrtwende erschweren wird.
- ○ Laut internationalen Organisationen und Regierungen haben wir noch 10 Jahre Zeit, um die Klimakrise zu verhindern.
- ◉ Es muss sofort mit der Umsetzung umfangreicher Maßnahmen zur Bekämpfung des Klimawandels begonnen werden.

4.

Wie sollen Städte laut dem Sprecher gesehen werden?

- ○ Ort der Innovation
- ◉ internationales Gut
- ○ nationales Eigentum
- ○ Globaler Norden

5.

Was hat sich geändert seit der Sprecher im Jahr 2016 Bürgermeister wurde?

- ○ Die Art und Weise, wie Städte funktionieren
- ◉ Die Einstellung des Sprechers in Hinblick auf die globale Rolle von Städten
- ○ Die Rolle, die Städte in der Weltpolitik spielen
- ○ Die Aufgabenbereiche seiner Rolle als Bürgermeister

6.

Geben Sie bitte an, wie sehr Sie den folgenden Aussagen auf einer Skala von 1 bis 5 zustimmen.

1 = stimme überhaupt nicht zu
2 = stimme nicht zu
3 = neutral
4 = stimme zu
5 = stimme voll und ganz zu

Ich habe das Gefühl, den Inhalt des Videos im Großen und Ganzen verstanden zu haben.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Die Untertitel sind leicht zu verstehen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Es erfordert keine große Anstrengung die Untertitel zu lesen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Die Untertitel waren angenehm zu lesen.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Ich bin mit der Qualität der Untertitel zufrieden.
1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

7.

Gibt es etwas, das Sie zu den Untertiteln noch sagen möchten?