

## Subtitles in the 2020s: The Influence of Machine Translation

 **Hanna Hagström** 

Stockholm University

 **Jan Pedersen** 

Stockholm University

---

### Abstract

Machine translation is now making serious inroads into the field of interlingual subtitling. This has been made possible by the use of template files and higher reading speeds. As we move into this new phase in the development of the subtitling process, the phase of machine-translated and postedited subtitles, it is highly pertinent to look at marks that this new process leaves on the subtitled product, i.e., the subtitles themselves. We conducted a diachronic study of subtitles before and after machine translation was part of the process. We did this by comparing a corpus of Swedish subtitles of Anglophone TV programmes produced after machine translation was introduced to a corpus of subtitles from before that period. We also took data from studies of earlier processes into account. When assessed using existing guidelines and the FAR model, the post-edited subtitles produced in the 2020s were found to be faster, more oral, less cohesive, less complete and with less meticulous punctuation and line-breaks than those produced in the 2010s. They were also of significantly lower quality in all areas investigated. Based on these results, we suggest that more research and development is needed to raise quality levels, and to make professional subtitlers augmented translators.

**Key words:** subtitling, machine translation, post-editing, subtitling processes, translation quality, quality assessment, FAR model, augmented translator.

Citation: Hagström, H., & Pedersen, J. (2022). Subtitles in the 2020s: The Influence of Machine Translation. *Journal of Audiovisual Translation*, 5(1), 207–225.

<https://doi.org/10.47476/jat.v5i1.2022.195>

**Editor(s):** J.L. Kruger

**Received:** July 10, 2021


**Accepted:** July 6, 2022

**Published:** December 14, 2022

**Copyright:** ©2022 Hagström & Pedersen. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/). This allows for unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

---

 [hanna.hagstrom@su.se](mailto:hanna.hagstrom@su.se), <https://orcid.org/0000-0002-5963-2446>

 [jan.pedersen@su.se](mailto:jan.pedersen@su.se), <https://orcid.org/0000-0002-9189-0330>

## 1. Introduction

Subtitling has developed in leaps and bounds ever since its introduction in the late 1920s (Ivarsson, 2002, p. 7). Many of these leaps and bounds are caused by changes in technology, as the subtitling process is technological in its very nature. Some of the technological innovations that can be seen as milestones are the word processor in the 1970s (Ivarsson, 2002, p. 10), the use of the electronic time code in the 1980s (Pedersen, 2011, p. 68), and the use of master template files in the 1990s (Pedersen, 2007, p. 41). Each development of the subtitling process has resulted in changes to the subtitled product, i.e., changes in the subtitles themselves and, in consequence, in how we are used to seeing them, which means changes in norms (cf. Pedersen, 2020).

The latest technological innovation for subtitling is the involvement of machine translation (MT) technology. MT has been used for some considerable time in other forms of translation, e.g., non-fiction translation and localization (Bywood et al., 2017), but less so in translating more creative texts, such as novels (but cf. e.g., Kenny & Winters, 2020; or Gueberhof-Arenas & Toral, 2022). Subtitling is not only creative in a way similar to literary translation, but the special conditions it entails, such as the switch from spoken to written language and the need for condensation, have previously made MT less conducive for subtitling. The key development that paved the way for MT in subtitling was the use of template files, which turned the spoken dialogue into written text that can be more easily processed by machines. Once subtitlers started making new translations based on previous translations or, more commonly, transcriptions of the original dialogue in subtitle format (henceforth called by their industry abbreviation EMT files – English master template files), the main practical obstacle for MT in subtitling was removed (Flanagan, 2009). In this way, the semiotic switch from spoken to written language was taken care of. After that, it was mainly a question of MT technology reaching satisfactory levels for the industry to start using it. Early attempts at MT in subtitling in the 1990s failed to reach these levels, but development continued (cf. e.g., Volk & Harder, 2007), and we have now reached a point where some of the leading companies in the business are using postedited MT for parts of their subtitling output (see below). The problem of condensation necessary for converting fast speech into comfortable reading speeds was dealt with in the same manner: condensation takes place when EMT is produced – or there might be no condensation at all, as expected reading speeds have increased dramatically over the last few years, thus making condensation less of an issue (Pedersen, 2018).

The careful reader will have noticed the word “postedited” in the penultimate sentence of the previous paragraph. Despite huge advances in MT technology and quality (as witnessed yearly in the World Conferences on Machine Translation, e.g., WMT21 [www.statmt.org/wmt21](http://www.statmt.org/wmt21)), humans are still needed for post-editing (PE) MT output before the subtitles are safe to air, just as humans are necessary for felicitous time-coding and postediting EMTs produced by automatic speech recognition. We are thus not suggesting that MT will replace humans in subtitle production; we are more interested in what happens when MT becomes part of the subtitling workflow.

When one of the world's leading subtitling companies (which will remain unnamed for reasons of anonymity) announced on their official website that they would be starting to integrate MT and PE into their subtitling process in Sweden in 2020, we were interested in seeing what the results of this most recent development would be. In order to do that, we assembled a corpus of TV programmes subtitled from English into Swedish after the process had been implemented and compared that to a similar corpus of programmes that had been subtitled in the 2010s, and also to some older data in the same language combination, to look at the development over time. We used these corpora to investigate the nature and quality of interlingual subtitles produced using a process that involved EMT files, MT and PE, compared to subtitles produced only using EMT files. The present study is thus a mixed-method descriptive empirical comparative product study of subtitles produced before and after a technical shift took place.

## 2. Subtitling Processes

In the present study, we are mainly interested in investigating the subtitle product, i.e., subtitles. However, the product is affected by the various processes involved, so in this section, we will give an overview of the main processes, actors, and roles involved in producing subtitles.

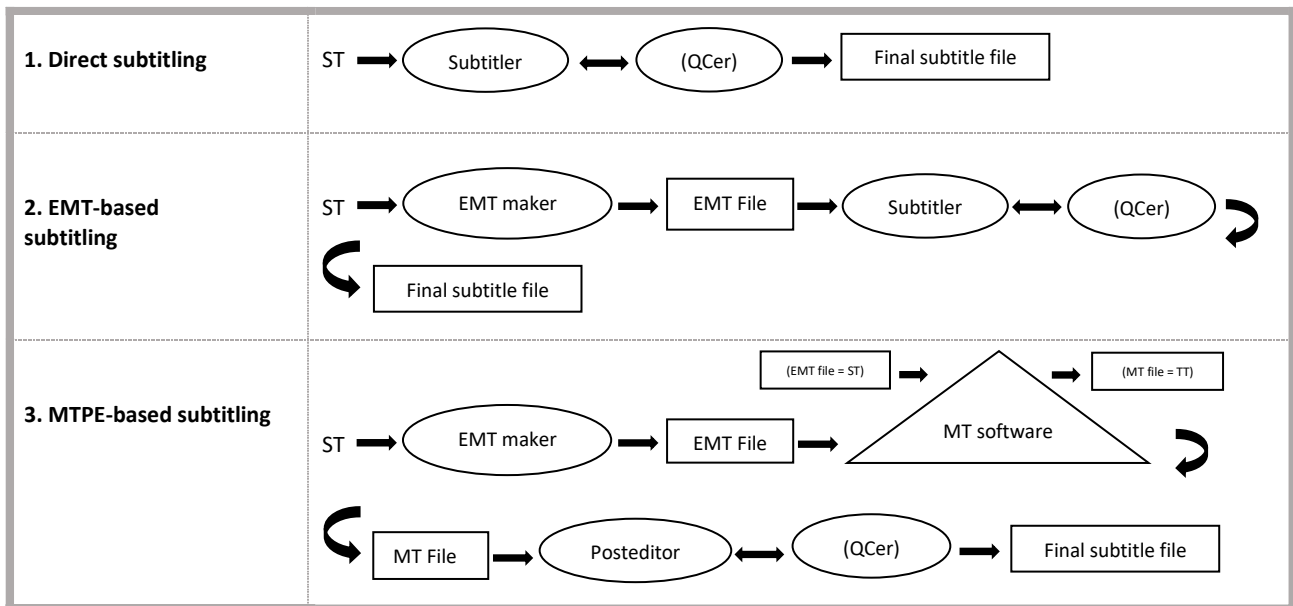
Toury (1995) distinguishes between the translation event and the translation act. The translation event is everything that happens from when a translation is commissioned to the final product. The translation act is the process of translating as such and is thus part of the translation event. Building on Nordman (2009), as outlined by Chesterman (2013), the translation event can be seen as a chain where various actors are involved in different stages of the translation process. In Nordman's (2009) study, the actors performed their tasks in sequences. In subtitling, interlingual translation can be carried out in different ways, and the translation act can be made up of various processes. The subtitling process is understood here to be the process of going from the audiovisual source text to the final subtitling file, which then becomes part of the target text, i.e., the subtitled audiovisual content. In this study, three interlingual subtitling processes are relevant. They are presented in a flow chart in Figure 1 below as chains in fixed sequential orders. In this figure, the roles performed by human actors are represented by oblongs, whereas the role performed by the machine is depicted as a triangle. The products (i.e., files) are represented by rectangles, and the arrows denote the typical order of the workflow. "QCer" stands for quality controller, an important role that ideally is played by a different person than the subtitler or posteditor.

In the first process, which we call direct subtitling, the subtitler is responsible for both spotting and translation, and this process is still used in subtitling workflows. In the second process, here called EMT-based subtitling, an EMT maker produces an EMT file, which is then translated by a subtitler. The third process is Machine Translated and Post-Edited, or MTPE-based, subtitling. PE can be explained as follows: "[T]he task of the posteditor is to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)" (Allen, 2003, p. 297). According to that definition, PE is not translation *per se* (though it is part of the

translation act), and Bywood et al. (2017) argue that the implementation of MTPE in subtitling introduces a new profession, the subtitler posteditor, henceforth posteditor. MTPE also means that one role (namely, translation) is performed by software, rather than a human agent. In the MTPE process, an EMT file is translated by an MT system, which produces a machine translation file (MT file). The MT file is then postedited by a human posteditor.

Figure 1.

*Interlingual Subtitling Processes in the Translation Act*



Note. Authors' own flowchart.

The roles involved in the three processes are subtitler, posteditor, quality controller, and EMT maker. It should be pointed out that one actor can have more than one role in these processes. In process 1, one or two roles are involved, and only one subtitle file is produced. In process 2, two or three roles are involved, and two subtitle files are produced, and in process 3, two or three roles, plus the MT software, are involved, and three subtitle files are produced.

In interlingual subtitling, EMT files should not be seen as source texts, as the source text in subtitling is the audiovisual material. Admittedly, in pivot translation, when subtitlers translate through an intermediate language (usually English) with no comprehension of the original dialogue (Díaz-Cintas & Remael, 2021, p. 43), the source text position of the EMT file becomes stronger; however, generally, the function of the EMT file is considered to be more of a guide for translation rather than an actual source text. Pedersen (2011) points out that subtitlers are paid less to translate via EMT files, and to save time, some will more likely translate from EMT files rather than the original dialogue. One might thus claim that these subtitlers treat EMT files as a kind of source text. Also, in the MTPE process, the EMT file is in fact a source text, as the MT software translates the EMT file and produces an MT file, which is the target text of the source EMT file. Nevertheless, in all three processes

depicted in Figure 1, it is the audiovisual material that is supposed to be the source text for subtitler, quality controller, and posteditor alike, and it is thus essential that all these actors have full access to the audiovisual content.

In all three processes, quality control is ideally involved, which means that from the process perspective there should be no major difference in the quality of the output. Also, the second and third processes have the same number of roles and the same number of eyes that can spot errors, at least when these roles are performed by different actors. Whether this affects quality is one of the main aspects investigated in the present study.

### 3. Quality in Subtitles

Quality in translation is a very complex issue, and in subtitling perhaps even more so. Subtitling quality is most often measured against in-house guidelines (Pedersen, 2017), which means that comparing the quality of subtitles produced by different companies, and/or at different times can be problematic. This is particularly true when in-house guidelines are not available to the person carrying out the quality assessment. For this purpose, the FAR (Functional equivalence – Acceptability – Readability) model was created (Pedersen, 2017). It is a generic model for assessing the quality of pre-prepared interlingual subtitles, and it has been used to measure quality in fansubs (Pedersen, 2019), as well as in professional subtitles, and is now an integrated part of the quality application of the Trados subtitling unit.

As we have not had access to the relevant in-house guidelines for the subtitles investigated in the present study, the FAR model has been used. The model is loosely based on the NER (Number of words – Edition – Recognition) model (Romero-Fresco & Martínez, 2015). Both models are based on error analysis, and have three penalty levels (minor, standard, and serious) depending on how serious the errors could be perceived by viewers. Each error can then be penalised by the FAR model with 0.25, 0.5, 1.0 – or 0.5, 1.0, 2.0 for semantic errors, as viewers are assumed to have lower tolerance for semantic errors in interlingual subtitles. Each episode or movie gets a score and quality can thus be compared between different translations.

The FAR model takes the relationship between interlingual subtitles and end consumer (viewer) into consideration. In Pedersen (2007), a tacit contract between the subtitler and the viewers is postulated, where the viewers agree to pretend that the subtitles are the actual dialogue. This contract, a contract of illusion, requires a “good deal of willing suspension of disbelief” (2007, p. 46) from the viewers, who have internalised it after having grown accustomed to subtitling. The basic unit of assessment in the FAR model is the subtitle itself. The subtitle is used as the basic unit of assessment over, for instance, the word, sentence, or minute of airtime, since subtitling involves verbal condensing, and dialogue intensity can vary greatly between different programs, and

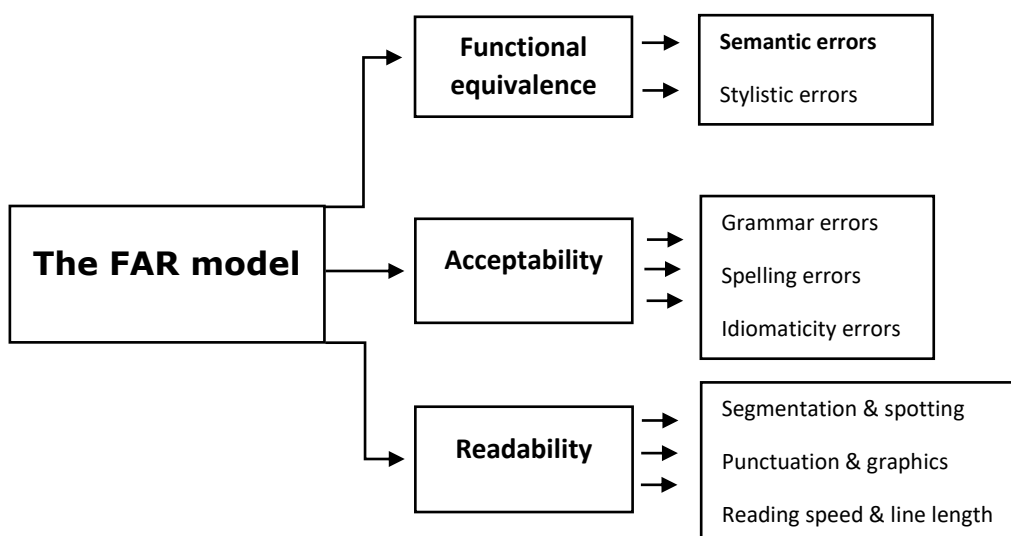
[...] for interlingual prepared subtitles, the subtitle as unit of assessment is not only intuitive, but also has other advantages. Firstly, it is a clearly and easily defined unit, which is also ideally

semantically and syntactically self-contained [...] Secondly, an error in a subtitle breaks the contract of illusion and makes the viewer aware that they are reading subtitles and that may affect not only a local word or phrase, but the processing of information in the whole subtitle. (Pedersen, 2017, p. 216)

The FAR model analyses three areas of quality from the viewers' perspective: functional equivalence, acceptability, and readability. Functional equivalence focuses on the message or meaning in the source text and how adequately it has been rendered into the target language. Acceptability assesses whether the norms of the target language have been adhered to, and readability determines the viewers' ability to read the subtitle and understand the message it conveys (Pedersen, 2017, pp. 217–223). This tripartite model also divides each area into sub-areas, which are illustrated in Figure 2 below.

Figure 2.

*The FAR Model*



*Note.* Authors' own flowchart, based on Pedersen, 2017.

As stated above, functional equivalence focuses on the message or meaning in the source text and on how adequately it has been rendered into the target language. Equivalence here is to be understood as pragmatic equivalence as subtitling entails condensing, and the intended meaning of the source text message matters more than the actual words used in the utterance. This area is divided into two sub-areas: semantic errors and stylistic errors. The second area is acceptability, which assesses whether subtitles adhere to the norms of the target language. Acceptability is divided into three sub-areas: grammar errors, spelling errors, and idiomacity errors. Grammar errors have to do with the grammar of the target language. However, subtitles “are seen as a hybrid form, containing some oral features in the written form” (Pedersen, 2017, p. 220), which means that, for instance, subject deletion and incomplete sentences would not automatically be analysed as errors,

as these features occur frequently in subtitling. Readability is the third and last area included in the model and it concerns the presentation of the subtitles and their technical aspects; it focuses on the viewers' assumed ability to read a subtitle while it is displayed on screen. Readability is divided into three sub-areas: segmentation and spotting, punctuation and graphics, and reading speed and line length.

The model gives a quantitative output in the form of three key figures for all three areas. The first figure represents the total number of errors. The second gives the error score, which is the number of errors multiplied by the gravity of each individual error. Finally, there is the approval rate, which is the error score divided by the number of subtitles. The approval rate is the most important key figure, as it tells you how good the subtitles really are, regardless of the length and the verbosity of the film or TV programme. In section 5, we present how the FAR model has been adapted to meet the conditions of the present study.

#### 4. Material

For the purposes of this study, we assembled two highly comparable corpora of TV series. The subtitles were commissioned by the same company before and after the time when MTPE was introduced into the subtitling process. We named the corpora 2010s subtitles and 2020s subtitles, respectively, and they each consist of the Swedish subtitles of 13 episodes of Anglophone TV series, and are presented below in Table 1.

Table 1.

##### *Material*

2010 episodes	Duration	Subtitles	2020 episodes	Duration	Subtitles
Chasing Classic Cars S8Ep1	43 min	483	Chasing Classic Cars S11Ep4	43 min	548
Chasing Classic Cars S8Ep2	43 min	523	Chasing Classic Cars S11Ep10	43 min	629
How the Universe Works S1Ep1	44 min	376	How the Universe Works S8Ep3	44 min	621
West Indochina S1Ep1	43 min	290	The Real Black Panther	45 min	341
Deadliest Catch S8Ep4	48 min	482	All on the Line S1Ep3	45 min	568
Deadliest Catch S8Ep4	48 min	425	All on the Line S1Ep4	45 min	578
Chopped S2Ep1	42 min	528	Bizarre Foods S4Ep11	42 min	606
Diners S17Ep8	21 min	252	The Zimmerman List S1Ep18	21 min	263
My 600lb Life S2Ep1	45 min	503	Dr Pimple Popper S2Ep1	45 min	587
Mountain Monsters S1Ep8	45 min	544	Mountain Monsters S2Ep8	44 min	492
Extreme Couponing S1Ep1	22 min	277	Flip or Flop S7Ep10	21 min	306
Hawaii Life S9Ep8	21 min	237	Hawaii Life S13Ep01	21 min	313
Grand Designs S17Ep1	43 min	506	100 Day Dream Home S1Ep6	43 min	629
Total	508 min	5 426		502 min	6 481
Total number Of subtitles					11 907

*Note.* Authors' own compilation.

The episodes were closely matched, in that they were either from the same series or comparable episodes of similar series. The 2010s episodes had been first broadcast in Sweden from 2012 to 2018,

i.e., before the MTPE process was implemented at this particular company. The 2020s episodes were all broadcast for the first time in Sweden between July and September 2020, after the company had announced that the process was in place. There are two genres represented in each corpus: reality series and documentaries. The reason for choosing these genres, which are part of what used to be referred to as daytime TV, is that they belong to the kind of programming on which processes such as EMT and MTPE would be used on. Had we instead chosen primetime TV with high-profile films, for instance, it would have been much harder to establish whether MTPE had been used.

As can be seen from Table 1, the two corpora are highly comparable, even though the 2020s episodes tend to have a higher number of subtitles, which we will discuss in section 6.1.1. All subtitles in both corpora in this study were produced by the same company, for similar clients, so it can be assumed that the same guidelines were followed. We used streaming services and TV to access and analyse the subtitled episodes. This means that we do not have electronic access to the subtitle files, which in turn makes it difficult to accurately pinpoint certain technical data, such as reading speeds. This point will be further developed in section 5, below.

## 5. Method

In this study, we applied two methods in parallel. We conducted a close reading, comparing the source texts (the whole polysemiotic texts, including non-verbal sound and vision) and the target texts, i.e., the source text plus the subtitles. This was done in order to find out what 2020s subtitles were like, when compared to subtitles produced using the old processes. At the same time, we applied the FAR model with the limitations described below. This was done to see if the 2020s subtitles differed in quality from subtitles produced during the previous period. To ensure the robustness of the investigation, we carried out these processes twice, and we alternated between investigating 2010s and 2020s subtitles, to avoid the accommodation factor that often appears with continuous reading of similar material. The evaluation was carried out by one of the authors of the present paper, who is not only a long-time professional subtitler and quality controller, but is also trained in using the FAR model, while the second author, who created the model, made a random sample evaluation to ensure reliability. This gave us an inter-rater reliability of more than 90 per cent.

As stated above, the FAR model was used for the quality assessment of the two corpora. The FAR model is fed local norms in order to carry out a fair assessment. These norms ideally come in the shape of in-house guidelines, but we did not have access to these. Instead, we used the national Swedish subtitling guidelines, which are available at [www.medietextarna.se](http://www.medietextarna.se). This seemed like a fair substitute as the company which produced the subtitles in our material co-created the guidelines, and could thus be assumed to stand by them.

As we did not have access to digital subtitle files, but instead accessed the published subtitles via streaming and TV, as any viewer would, there are some limitations to the quality assessment. Even though it is feasible to count all individual subtitles (which we did) it is much less feasible to manually



count line lengths and exposure times, so those are left out of the assessment, unless they were obvious enough to be detectable to the viewer/analyst. This is, however, true for both corpora, and will thus not affect the comparison. One thing that does affect the comparison, however, is a consequence of the FAR model using the subtitle as its basic unit. This means that if the subtitles are segmented so that there are many short subtitles rather than fewer and longer subtitles, the approval rate becomes higher (see section 3). This is probably unavoidable, as all other basic units (such as a word, phrase, sentence, utterance, minute), would be even more problematic, and should be borne in mind when interpreting the results.

## 6. Results and Analysis

In this section, we present the results and analyses. In the first subsection (6.1), we give the results of the investigation into the nature of subtitles produced in the 2020s. In section 6.2, we present the results of the quality comparison of 2010s and 2020s subtitles.

### 6.1. The Nature of Subtitles from the Early 2020s

Our close reading of the 2020s corpus, when compared to 2010s subtitles, reveals six traits in which the former differ considerably from the latter: higher subtitle density and more one-liners, increased orality in subtitles, omissions of plot-relevant information, unconventional line breaks, punctuation errors, and lower cohesion.

#### 6.1.1. Subtitle Density and Percentage of One-Liners

Subtitle density is defined as

the number of subtitles the TT translation is divided into, as measured by the number of subtitles per minute. This measurement is based on subtitle quantity (i.e. the number of subtitles per TT), divided by the length of the TT in minutes. (Pedersen, 2011, p. 130)

By counting one-, and two-liners in this study, we were able to calculate the percentage of one-liners and subtitle density in both corpora. When it comes to this measurement, we were fortunate to have access to the data of an earlier study on similar material from the noughties (Pedersen, 2007). By using the findings of Pedersen (2007), and our two corpora, we can see developments over time, and check whether these figures are affected by the three processes presented in section 2. The results of the analyses are displayed in Table 2 below.

Table 2.

*Subtitle Density and Percentage of One-Liners*

	2000s data	2010 data	2020s data
% one-liners	27.3	29.2	32.0
Subtitle density, subtitle/minute	9.2	10.6	12.9

*Note.* Authors' own compilation.

The average episode in Pedersen's 2000s study, which were mainly produced using direct subtitling, had 27.3% one-liners, compared to the average 2010s corpus, where 29.2% of the subtitles were one-liners. Already here, we have an increase of 7.0 %. The average 2020s episode had 32% one-liners, an increase of 9.6% compared to the 2010s data. In other words, the percentage of one-liners increases with every new subtitling process, and between 2000 and 2020, the increase is 17.2%.

Table 2 also shows an increase in subtitle density over the twenty years. In Pedersen's study, the subtitle density was 9.2 subtitles per minute in an average episode. The subtitle density was a bit higher in the 2010s, with 10.6 subtitles per minute, and the 2020s corpus showed a subtitle density of 12.9 subtitles per minute. This means an average of 2.3 subtitles more per minute, or an increased subtitle density of 21.7% in the 2020s data when compared to the 2010s data. This, and the increased percentage of one-liners, explains why the 2020s corpus have 1,055 more subtitles than the 2010s corpus. Like the percentage of one-liners, subtitle density also increases with every new subtitling process, and the increase from 2000 to 2020 is 3.7 subtitles per minute, or 40.2%. It is thus very clear that viewers now are exposed to more and shorter subtitles than in the mid-noughties when Pedersen's data was assembled.

### 6.1.2. Increased Orality

By increased orality, we mean that more spoken language features are rendered in the subtitles. The Swedish national guidelines urge subtitlers to "condense the content rather than reflect the speech verbatim. [...] repetitions and filler words can easily be left out" (Medietextarna, 2020, p. 8). The Code of Good Subtitling Practice also recommends this (Ivarsson & Carroll, 1988, pp. 85–87). Mattsson found that discourse particles (in that case "well", "you know", "I mean", and "like") were mainly not translated (2009, pp. 273–276). Nonetheless, in the 2020s corpus there is a strong tendency to render filler words or backchannels such as "yes", "well", "I know", and even an occasional "aha". Redundant oral features have, in other words, not been omitted. Example 1 illustrates a translation that resulted in two subtitles, where one would probably have sufficed for rendering the source text dialogue, if condensed. Two men are talking about a vintage car and comparing it to another make and model. One of them says that he favours the convertible in front of them, and the other man agrees.

(1) *Chasing Classic Cars* S11Ep4

English original version:

-Me too. They're more drivable. (Speaker 1)

-Slabside... (Speaker 2)

-Yeah. (Speaker 1)

-You know, taller wheels. (Speaker 2)

Subtitle:

-Jag med. De är mer körbara. (Speaker 1)

-Slabside... (Speaker. 2)

-Ja. (Speaker 1)

-...de har större hjul. (Speaker 2)

English back translation

-Me too. They are more drivable. (Speaker1)

-Slabside... (Speaker 2)

-Yeah. (Speaker1)

-...they have taller wheels. (Speaker 2)

Example 1 shows an oral feature that is unusual and perhaps even hard to follow in writing. Subtitles traditionally tend to be condensed and reformulated for the written format, as in our alternative version of example 1a below.

(1a)

Subtitle:

-Jag med. De är mer körbara.

-Slabside har större hjul.

English back translation

-Me too. They are more drivable.

-Slabside have taller wheels.

Interestingly, this tendency is only evident in the 2020s corpus. This feature may be caused by the transcription in the EMT file rendered in Swedish by the MT software and being left unedited.

### 6.1.3. Omissions of Plot-Relevant Information

Information interpreted as plot pertinent has been more frequently omitted in the 2020s subtitles when compared to the 2010s ones. Condensation and reduction are common features in subtitling due to time-and-space constraints, and De Linde claims that reductions are systematic and not random (1995, p. 19), but in the 2020s corpus, several examples to the contrary can be seen.

Example 2 is a typical example of these partially rendered source text messages, where the missing information gave the subtitle a very different meaning from the original. A married couple is building a new home. An architect is sharing his vision of a new outdoor living area. His plan entails a large deck with multiple seating and lounge areas.

(2) *100 Day Dream Home S1Ep6*

English original version:

This thing is going to be massive. Maybe some outdoor dining here, like a cool sitting area here...

Subtitle:

Den blir enorm.

English back translation

It is going to be massive.

The viewer does not know what “den” (it) refers to, because it has not been mentioned previously. Something in the yard will be massive, but neither the subtitles nor the visuals explain what. In addition, since no work has been carried out in the yard, viewers would have a hard time trying to figure out what the man is actually talking about.

One explanation for this tendency may be found in Koponen et al.’s (2020) experimental subtitling study, where the content in the MT files was so condensed that the subtitlers added both, textual information missing from the MT file and new subtitles. On the other hand, Bywood et al.’s SUMAT study found that source content words were missing from the MT files due to the statistical MT system used in their study (2017, p. 501), and neural MT systems are known to “sacrifice adequacy for the sake of fluency” (Koehn & Knowles, 2017, p. 28). Consequently, omissions in this study might be a result of MT errors which the posteditors have not noticed. An explanation might be that information has not been transcribed in the MT input, i.e., EMT file. This needs more research to be verified, however.

#### 6.1.4. Unconventional Line Breaks

Line breaks are an important aspect of subtitling, at least from the points of view of subtitling agencies, practitioners, and subtitling organizations. How to break lines is explained in several national subtitling guidelines<sup>1</sup>, the Swedish ones included (Medietextarna, 2020, pp. 3, 5); Ivarsson and Carroll devote two pages to line breaks (1988, pp. 90–91). The general rule is to break the line at the highest possible syntactic node (cf. Karamitrouglou, 1998). While the impact of this on viewers is contested (cf. e.g., Szarkowska et al., 2018), it is still part of the guidelines that both corpora in the present study were supposed to adhere to.

In the present study, both corpora make use of the pre-existing spotting and segmentation in EMT files. For some reason, however, unconventional line breaks, while still rare, were three times as common in the 2020s subtitles than in the earlier subtitles (129 vs. 43). One possible explanation might be that posteditors are affected by the MT solutions to the extent that they let unconventional line breaks pass without reflecting on it. Another explanation might be that in postediting, the text is already there, as opposed to being keyed in by a subtitler. We would like to suggest that for subtitlers, line-breaking is an internalised action to some extent. That is, when subtitlers write the texts, they line-break in a semi-automated fashion, as they construct the text. Arguably, this, does not come naturally when revising and editing MT files.

#### 6.1.5. Punctuation Errors

As we will show in section 6.2, the 2010s material contained 2 punctuation errors, while 72 errors were found in the 2020s subtitles. The majority of the minor errors were made up of the use of dashes instead of ellipsis, and most standard errors consisted of missing or unconventional use of speaker dashes. There were examples of two people speaking, but the absence of speaker dashes signalled only one speaker. There were also instances where one person was speaking, but two dialogue dashes were used, making it seem like two people were in fact speaking.

#### 6.1.6. Low Cohesion

The last trait concerns translation flow. Several examples of renderings broken where there were no pauses in the source text dialogue were found in the 2020s subtitles, which gives these subtitles a somewhat telegraphic style. In many of these examples, full stops are used between two main clauses instead of conjunctions, creating shorter sentences. There were also a substantial number of main clauses separated from subclauses by a full stop so that one subtitle contained only a subclause, as in example 3 below. A man has a big round lump growing on his head. He consults a doctor to get

---

<sup>1</sup> E.g., Croatia, Denmark, Finland, France, Netherlands, and Norway.

it removed and expresses his excitement over being “normal” again. His utterance contains no pauses.

(3) *Dr Pimple Popper* S2Ep1

English original version:

I’m kind of excited but a little nervous too, but I’m gonna be excited cause they’re gonna take this ball away from me.

Subtitle:

Jag är glad, men lite nervös.

Men jag är glad!

För de ska ta bort bollen.

English back translation

I’m happy but a little nervous.

But I’m happy!

Because they’re going to remove the ball.

From the above tendencies we can conclude that subtitles from the early 2020s are more fast-paced, more oral, less cohesive, less complete and with less meticulous punctuation and line-breaks than subtitles from several years before.

## 6.2. The Quality of 2020s Subtitles

As explained in section 5, the FAR model was applied to assess the quality of the two corpora. The model had to be adjusted due to the lack of digital subtitle files, and thus errors in line lengths (in characters per line; CPL) and expected reading pace (in characters per second; CPS) were noted only when they were obvious enough to be detectable to the viewer/analyst. The results of the quality assessment are presented in Table 3 below.

Table 3 clearly shows that the quality is lower for the 2020s subtitles, when compared to the 2010s subtitles. There are 6.5 times more errors in the 2020s subtitles. The error score is 8.5 times higher for the 2020s subtitles, meaning that the errors are not only more numerous, but also more serious. The 2020s subtitles ( $M = 91.75$ ,  $SD = 93.17$ ,  $n = 8$ ) have more errors in all areas and sub-areas than the 2010s subtitles ( $M = 14$ ,  $SD = 21.65$ ,  $n = 8$ ). A two-sample t-test on the number of errors in each area and sub-area showed that the differences are significant,  $t(14) = 1.76$ ,  $p = 0.2$  (1 tail). There are significantly more semantic errors, grammar errors, spelling errors, idiomaticity errors, spotting errors, segmentation errors, and punctuation errors. Only stylistic errors show a minor difference between the two corpora. The greatest difference in the number of errors was found when assessing readability, there were 69 errors in the 2010s episodes compared to 369 errors in the 2020s episodes, where spotting and inappropriate line breaks dominated.

Table 3.

*Number of Errors, Error Rates, and Approval Rates Comparison of Both Corpora*

2010s corpus Area/sub-area	Number of errors	Error score	2020s corpus Area/sub-area	Number of errors	Error score
Functional equivalence			Functional equivalence		
Semantic subtotal	11	5.5	Semantic subtotal	135	102.5
- Minor	11	5.5	- Minor	99	49.5
- Standard	-	-	- Standard	19	19
- Serious	-	-	- Serious	17	34
Stylistic subtotal	2	0.5	Stylistic subtotal	13	3.75
- Minor	2	0.5	- Minor	11	2.75
- Standard	-	-	- Standard	2	1
FE Subtotal	13	6.00	FE Subtotal	148	106.25
Approval rate: 99.88%			Approval rate: 98.23%		
Acceptability	30	7.5	Acceptability	217	56.75
Grammar subtotal	16	4.00	Grammar subtotal	85	21.25
- Minor	16	4.00	- Minor	85	21.25
Spelling subtotal	9	2.25	Spelling subtotal	38	10.00
- Minor	9	2.25	- Minor	36	9
- Standard	-	-	- Standard	2	1
Idiomacity subtotal	5	1.25	Idiomacity subtotal	94	25.5
- Minor	5	1.25	- Minor	86	21.5
- Standard	-	-	- Standard	8	4.00
Acceptability subtotal	30	7.5	Acceptability subtotal	217	56.75
Approval rate: 99.84%			Approval rate: 99.06%		
Readability			Readability		
Segment/spotting subtotal	66	16.5	Segment/spotting subtotal	295	75.5
- Minor	66	16.5	- Minor	292	73
- Standard	-	-	- Standard	1	0.5
- Serious	-	-	- Serious	2	2
Punctuation & graphics	2	0.75	Punctuation & graphics	72	25.25
- Minor	1	0.25	- Minor	43	10.75
- Standard	1	0.5	- Standard	29	14.5
CPL & CPS subtotal	1	0.25	CPL & CPS subtotal	2	0.5
- Minor	1	0.25	- Minor	2	0.5
Readability subtotal	69	17.5	Readability subtotal	369	101.25
Approval rate: 99.65%			Approval rate: 98.38%		
Total	112	31.00	Total	734	264.25
Total approval rate: 99.38%			Total approval rate: 95.71%		

*Note.* Authors' own study.

As can be seen in Table 3, the approval rate of the 2010s subtitles is 3.65 percentage points higher than for the 2020s subtitles. This may not sound like much, but it means that the 2010s subtitles contained 0.62% disapproved subtitles, while the 2020s subtitles had 4.29% disapproved subtitles,

i.e., almost seven times as many. As the FAR model is a generic tool for assessing any subtitles, it does not include a pre-set lowest acceptable level of approval rate; it may be worth mentioning, however, that the NER model, which works in a similar way, has a cut-off point at 98% (Romero-Fresco & Martínez, 2015, p. 4), so if we were to apply that standard here, the 2020s subtitles would not be deemed acceptable. We would also like to reiterate that using the subtitle as the basic unit, which the FAR model does, means that the approval rate is higher for texts with more subtitles. Thus, if the 2020s corpus had not contained 1,055 subtitles more, its approval score would have been even lower.

Perhaps the most striking difference when it comes to error score involves semantic errors, most of which resulted from missing pertinent information, where the error score was 17.7 times higher for the 2020s subtitles. A likely explanation for this is that the MT software comes up with infelicitous or erroneous translations that the posteditors fail to spot. Example 4 illustrates the most common type, a minor semantic error, in which a couple is viewing an apartment and they discuss the best place in which to put the washer and dryer. The husband is not too keen on having them in the bathroom.

(4) *Hawaii Life* S13Ep1

English original version:

It's new equipment, but I really don't like it in the bathroom like this.

Subtitle:

De är nya men jag gillar inte  
att ha dem i toaletten så här.

English back translation

It's new but I don't like it  
in the toilet like this.

By translating verbatim here and keeping the source text preposition (in), the meaning changes from *inside the bathroom* to *in the toilet*. Most viewers are probably aware of the fact that a washer and a dryer would not fit into a normal sized toilet (bowl), and since all the appliances are visible, this error is only analysed as minor.

When it comes to acceptability, the most common errors are minor grammar and idiomaticity errors. It can thus be said that the 2020s subtitles are not as fluent as the 2010s ones, and show a higher degree of what Gellerstam (1986) calls “translationese”.

As for readability, the analysis is somewhat stunted, as we could not investigate this parameter thoroughly, as explained above. Of the errors we found, the most common ones in the 2020s subtitles were micro-segmentation errors, i.e., inappropriate line breaks, as discussed in section 6.1.4. After that, the most common error was another of the special features of 2020s subtitling, namely the missing speaker dashes and erroneous use of ellipses and dashes (cf. section 6.1.5).



It is thus abundantly clear from the above that the quality of the 2020s subtitles compares poorly to their predecessors. The 2020s subtitles in our material contain a significantly higher number of errors than the 2010s subtitles do in every area investigated.

## 7. Discussion and Conclusion

To sum up, we found that the 2020s subtitles in our material were more fast-paced, more oral, less cohesive, less complete, and constructed with less meticulous punctuation and line-breaks than their predecessors. They were also of significantly lower quality in all areas investigated. We want to stress here that despite the names we have given our two corpora, we do not claim that these differences apply to all subtitles produced in the 2010s and the 2020s, far from it, but they apply specifically to those of our investigation, which are subtitles of daytime TV produced in the mid to late teens and 2020, respectively. It is perhaps striking that the differences were so great, both in quality and in other features. After all, the material in the two corpora was produced with no considerable distance in time, in some cases with a time difference of only two years (cf. section 3). Most other variables were kept constant: the language combination, the producing company, (most likely) the guidelines, the genres, and often the individual TV series were all the same. The reason for the differences is thus most likely the change in process that the company announced in 2020. There can be many explanations for the differences that were caused by the process change. The MT systems are perhaps not as good or as efficient as they (hopefully) will be in the future. It is also possible that the subtitlers that carried out the postediting had not yet been properly trained or had not obtained the necessary skill levels as posteditors, which is a different task from subtitling.

Many of the traits we found to be typical for the 2020s subtitles are also linked to quality, particularly the punctuation, omissions, line break, orality, and cohesion issues. If these remain traits characterising MTPE subtitles, it seems likely that subtitling norms will evolve to incorporate them, meaning that standard subtitles will be different from what Swedish viewers are used to today. This will probably also be the case for the increased number of one-liners and subtitle density. As with the previous technical developments mentioned in the introduction, like the PC and electronic time-coding, there are bound to be teething problems in the processes that involve them. Our investigation, which we realise is too detailed to be carried out every day in a professional environment, has not looked at the quality of machine translation per se, but rather at the output of a process that involves MT, which means that there are many stages in which errors can occur. Machine translation is perhaps best seen as a tool like many others, including translation memories or automatic speech recognition. As such, it requires training to be used skilfully, and should be applied wisely, to help professional subtitlers become augmented translators. We leave that for others to investigate, and for our own future research, we aim to investigate how the traits displayed by the 2020s subtitles affect viewers. We think that this is particularly important if what we have found in the present study in fact signals a shift in norms.

## References

- Allen, J. (2003). Post-editing. In Somers, H. L. (Ed.) *Computers and translation: a Translator's guide*. John Benjamins.
- Bywood, L., Georgakopoulou, P., & Etchegoyhen, T. (2017). Embracing the threat: Machine translation as a solution for subtitling. *Perspectives: Studies in Translatology*, 3, 492–508.
- Chesterman, A. (2013). Models of what processes? *Translation & Interpreting Studies: The Journal of the American Translation & Interpreting Studies Association*, 8(2), 155–168.
- De Linde, Z. (1995). Read my lips: Subtitling principles, practices and problems. *Perspectives: Studies in Translatology*, 1, 9–20.
- Díaz-Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices*. Routledge.
- Flanagan, M. (2009). *Recycling texts: Human evaluation of example-based machine translation subtitles for DVD* [Doctoral dissertation, Dublin City University].  
<http://doras.dcu.ie/14842/1/MarianFlanaganPhD.pdf>
- Gellerstam, M. (1986). Translationese in Swedish Novels translated from English. In L. Wollin, & H. Lindquist (Eds). *Translation Studies in Scandinavia. Proceedings From the Scandinavian Symposium on Translation Theory (SSOTT) II. Lund 14–15 June, 1985* (pp. 88–95). (Lund Studies in English 75). Liber/Gleerup.
- Gueberhof-Arenas, A., & Toral, A. (2022). Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces*. <https://doi.org/10.1075/ts.21025.gue>
- Ivarsson, J., & Carroll, M. (1998). *Subtitling*. TransEdit.
- Ivarsson, J. (2002). Subtitling through the ages: A technical history of subtitles in Europe. *Language International*, 14(2), 6–10.
- Karamitroglou, F. (1998). A proposed set of subtitling standards for Europe. *The Translation Journal*, 2(2). <http://accurapid.com/journal/04stndrd.htm>
- Kenny, D., & Winters, M. (2020). Machine translation, ethics and the literary translator's voice. *Translation Spaces*, 9(1), 123–149.
- Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation* (pp.28–39). Association for Computational Linguistics. <https://www.aclweb.org/anthology/W17-3204.pdf>
- Koponen, M., Sulubacak, U., Vitikainen, K., & Tiedemann, J. (2020). MT for subtitling: User evaluation of post-editing productivity. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 115–124). European Association for Machine Translation. <https://aclanthology.org/2020.eamt-1.13.pdf>
- Mattsson, J. (2009). *The subtitling of discourse particles: A corpus-based study of well, you know, I mean, and like, and their Swedish translations in ten American films* [Doctoral dissertation, University of Gothenburg]. <https://gupea.ub.gu.se/handle/2077/21007>
- Medietextarna. (2020). *Guidelines for subtitling in Sweden*. Journalistförbundet. <http://www.est2019.com/wp-content/uploads/2018/10/Research-ethics-in-translation-studies.pdf>
- Nordman, L. (2009). *Lagöversättning som process och produkt* [PLEASE ADD ENGLISH TRANSLATION HERE] [Doctoral dissertation, Helsingfors universitet]. <https://helda.helsinki.fi/handle/10138/19639>
- Pedersen, J. (2007). *Scandinavian subtitles: A comparative study of subtitling norms in Sweden and Denmark with a focus on extralinguistic cultural references*. Stockholm University.
- Pedersen, J. (2011). *Subtitling norms for television an exploration focussing on extralinguistic cultural references*. John Benjamins.

- Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *The Journal of Specialised Translation*, 28, 210–229.
- Pedersen, J. (2018). From old tricks to Netflix: How local are interlingual subtitling norms for streamed television? *Journal of Audiovisual Translation*, 1(1), 81–100.  
<https://doi.org/10.47476/jat.v1i1.46>
- Pedersen, J. (2019). Fansubbing in subtitling land: An investigation into the nature of fansubs in Sweden, *Target*, 31(1), 50–76. <https://doi.org/10.1075/target.18017.ped>
- Pedersen, J. (2020). Audiovisual translation norms and guidelines. In Ł. Bogucki & M. Deckert, (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility* (pp. 417–436). Palgrave Macmillan.
- Romero-Fresco, P., & Martínez, J. (2015). Accuracy rate in live subtitling: The NER model. In J. Díaz-Cintas & R. Baños (Eds.), *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28–50). Palgrave Macmillan.
- Szarkowska, A., Gerber-Morón, O., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research*, 11(4).  
<https://dx.doi.org/10.16910%2Fjemr.11.4.2>
- Toury, G. (1995). *Descriptive translation studies and beyond*. John Benjamins.
- Volk, M., & Harder, S. (2007). Evaluating MT with translations or translators. What is the difference?. *Proceedings of MT Summit*. Copenhagen.  
[http://ling16.ling.su.se:8080/new\\_PubDB/doc\\_repository/239\\_volk\\_harder\\_subtitle\\_translation.pdf](http://ling16.ling.su.se:8080/new_PubDB/doc_repository/239_volk_harder_subtitle_translation.pdf)
- WMT21. (2021). *Conference on Machine Translation EMNLP 2021*. <https://www.statmt.org/wmt21>