YouTube Videos Subtitling in the Age of AI

Zainab Fahmy Ahmed ⓘ

Department of Translation / College of Arts /
University of Mosul / Mosul -Iraq

Abdul Rahman Ahmed Abdul Rahman ⓘ

Department of Translation / College of Arts /
University of Mosul/ Mosul-Iraq

Article Information	Abstract

This study examines AI-generated subtitling, using YouTube's videos as case study. Recent advancements in technology have led to significant improvements in subtitling, making the content more accessible to global audience. YouTube's auto-generated captions combine Automatic Speech Recognition (ASR) and machine translation (MT). The study hypothesizes that these AI-generated subtitles frequently contain errors; therefore, human post editing, segmentation and synchronization remain essential to correct these errors and ensure overall quality. This study adopts the NTR model suggested by Romero-Fresco and Pöchhacker (2017), which categorizes errors as minor, major and critical, to assess the subtitles accuracy. Additionally, it incorporates the technical parameters suggested by Robert and Remael (2017) to provide a comprehensive assessment of subtitle quality. The study concludes that AI technology is beneficial for making subtitling faster and simpler; however, the essential role of human oversight in achieving true accuracy cannot be overlooked.

ترجمة المحتوى المرئي الفديوي لليوتيوب في عصر الذكاء الاصطناعي

زينب فهمي احمد*   عبد الرحمن احمد عبد الرحمن*

المستخلص:

تبحث هذه الدراسة الترجمة النصية للمحتوى المرئي الفديوي المولدة بالذكاء الاصطناعي باستخدام مقاطع يوتيوب كدراسة حالة. وقد أدت التطورات التكنولوجية الحديثة إلى تحسينات كبيرة في مجال ترجمة المحتوى المرئي، مما جعله أكثر اتاحة للمشاهدين من مختلف أنحاء العالم. تجمع الترجمة التلقائية في اليوتيوب بين تقنيات التعرف التلقائي على الكلام (ASR) والترجمة الآلية (MT). تفترض الدراسة ان الترجمات النصية المولدة تلقائيا تحتوي على عدة أخطاء لذا تبقى عملية المراجعة البشرية اللاحقة والتحسين والتقسيم والمزامنة ضرورية لتصحيح هذه الأخطاء وضمان جودة النصوص. تعتمد هذه الدراسة نموذج NTR الذي اقترحه روميرو فريسكو وبوخهاكر (2017) والذي يصنف الأخطاء الى بسيطة وجسيمة وفادحة وذلك لتقييم دقة تلك الترجمات. كما تتبنى الدراسة المعايير التقنية التي

* قسم الترجمة /كلية الاداب/ جامعة الموصل / الموصل – العراق.
* قسم الترجمة /كلية الاداب/ جامعة الموصل/ الموصل – العراق .

اقترحها روبرت ورامل (2017) لتقديم تقييم شامل لجودة الترجمات. تخلص الدراسة إلى نتيجة مفادها أن تقنية الذكاء الاصطناعي مفيدة بلا شك لجعل ترجمة المحتوى المرئي أسرع وأبسط، ومع ذلك، لا يمكن إغفال الدور الأساسي للتدخل البشري في تحقيق الدقة الكاملة.

الكلمات المفتاحية : الترجمة المصاحبة، الترجمة النصية، الذكاء الاصطناعي والترجمة

1.Introduction

Subtitling has progressed from a specialized type of human translation to a process heavily impacted by advances in technology, particularly Artificial Intelligence (AI) and Automatic Speech Recognition (ASR). This progress has revolutionized content accessibility, enabling a worldwide audience to interact with audiovisual media in multiple languages. However, relying on AI-powered solutions, such as YouTube's auto-captioning, creates new issues and implications. This research investigates the advantages and disadvantages of AI-generated subtitles, utilizing YouTube's automated captioning as a case study.

1.1Research Questions

The following questions are going to be addressed in this study:

1. Is there any disparity between the accuracy rate and the technical parameters in AI-generates subtitles?
2. Does the low effective editions rate mean that those subtitles are not reliable?
3. Do auto captions need human segmentation and synchronization to ensure their quality?
4. Does the low rate of effective editions mean that machine fails sometimes to solve critical issues?

1.Subtitling

Cintas and Remael (2014, p.8) define subtitling as the process in which the translation of the original dialogue, on-screen text and audio information are all shown, generally at the bottom of the screen. It is not only a unique kind of translation, but it also constitutes a distinct translational genre because it translates speech in written form, which helps define the shape subtitles eventually take. (Costan Davara, 2020, p.50). In subtitling, the TT does not claim the position of an original; however, subtitling doesn't only simplify the original foreign text for the local audience but also allows watchers to experience different cultures (Chiaro, 2009, p.160). It keeps the "foreignness" of the original text by hearing the real voices of the characters. Subtitlers help viewers understand the dialogue and story better, and also provide important hints about the characters' social status, class, and relationships (Mera, 1999, p.75).

The subtitling process is often handled by a single translator, who oversees all aspects of subtitle development. Pedersen's Subtitling Norms for Television (2011) outlines the subtitling process. The translation agency starts the process by assigning a project to a translator who receives a copy of the video and, sometimes, a script. If there's no script, the translator has to transcribe the original audio and check for correct names and cultural references. Then he translates the content and synchronize the subtitles with the speech, a process called spotting. The translator sets time codes for each subtitle to match the dialogue, using a digital clock provided by the agency. Translators might also use other online tools to manage the timing and character limits of the subtitles. After translating and creating the subtitles, the translator edits the text to make it concise. The finished subtitles are then sent back to the agency. Nowadays, Technological advancements and new software have streamlined this process.

1.1 Subtitling Strategies

A subtitler should focus on crucial information and source language features like vocabulary, register, and dialogue styles. Although subtitling is difficult due to space and time constraints, there are ways to modify target text while retaining meaning and content (Tosatto, 2019, p.32). Cintas and Remael (2014) suggest the following solution to focus on necessary information:

2.1.1. Text reduction

Text reduction is a process that involves reducing a text by removing unnecessary information or omitting lexical items. It can be partial or total, with partial reduction achieved through condensation and reformulation; whereas total reduction through deletion or omission. Subtitling involves combining these processes, resulting in rewriting, which is a common feature of translation. (Cintas and Remael, 2014, p.145).

2.   1. 1. 1. Condensation and Reformulation (partial)

Tosatto (2019, p.34) believes that these methodologies allow for word, phrase, and sentence summarization based on the linguistic distinctions between SL and the target TL. Thus, the translator must be fluent in both languages. The simplest and most typical method is simplifying verbal phrases when translating a huge periphrasis that takes up too much space for a subtitle. They are often simplified and replaced by shorter, equivalent phrases. The second method is the generalization, which is used when the dialogue includes a list of names or places that will not be in the subtitle because the names can be replaced with their pronouns.

2.   1. 1. 2. Omission or Deletion (total)

Tosatto (2019, p.38) mentions that subtitling eliminates features for spatial reasons and to reduce repetition between subtitles and pictures. Some words, phrases, or expressions can be removed since they exist in many sentences. Word- and sentence-level omission might include various conversation aspects. The eliminated items should not be crucial to the message. Words such as greetings, interjections, vocatives, politeness formulae, hesitations, colloquialisms, and tag inquiries can be removed if they do not affect the final message (Cintas and Remael, 2014, p.164).

3.   YouTube Videos

YouTube is a large, international platform that offers a place to share and access an enormous amount of information. With more than 500 hours of video being uploaded per minute creating an overwhelming amount of information that consumers may find difficult to go through (Singh et.al, 2023, p.817). According to Almurashi (2016, p.34), YouTube has become one of the most accessible and easy forms of social media because of Google's video-sharing platform, which allows users to upload, search, and share short videos. Gray (2017, p.12) indicates that technology has advanced to the extent where YouTube can generate subtitles for all videos in most languages. They are produced by machine-learning algorithms; hence their quality is not consistently better or equivalent to that of expert translations. Google and YouTube have teamed together to offer subtitles with automated translation or machine translation as part of YouTube closed captions (CC), making the material more accessible to users. (The YouTube Team, 2008). There are still certain restrictions with auto-captioning, even if it offers a lot of languages for the user to choose from. Subtitles are generated by the system algorithm used by the auto-captioning system.

This might lead to subtitles that aren't correct or do not fit the context. If the term isn't already in the system, the automatic captioning system might not be able to identify it. On the other hand, if the user loses a lot of words, the message won't go over (The YouTube Team, 2008). YouTube's automatic captioning exemplifies the application of Automatic Speech Recognition (ASR) in subtitling. Google's voice recognition technology automatically transcribes spoken language from a sound into text, which is then shown as subtitles or captions on screen, with timing also calculated automatically (Chavez, 2009).

4.   AI and Auto-Generated Subtitles

In recent years, the Auto captions technology has reached significant progress, thanks to Artificial Intelligence (AI) and it is now widely used as a starting point for professional translation services (Gregori, 2021, p.69). Automatic subtitling is commonly viewed as a combination of the following technologies: automatic speech recognition (ASR), which transcribes voice to text in the source language (SL); and machine translation (MT), which automatically translates the transcribed text in the SL into a target language (TL) (Lewis, 2015, p.59). Speech recognition problems appear to be the most frequent cause of translation errors in automatic translation, and the speaker's accent, audio quality, and speech rate all have a significant impact on the translation's quality (Karakanta, 2022, p.263). Xie (2023, p.54) argues that automatic subtitling requires an advanced workflow that includes auto-transcription, auto segmentation, auto-spotting, and machine translation. Although inconsistencies were found in literal translation, word order, linguistic register, punctuation, and mistranslation, most subtitlers have a good attitude toward automatic subtitling since it allows subtitlers to save time and effort.

Varga (2021, p.46) believes that there are numerous variables that influence subtitles, including the maximum and minimum duration of a subtitle on the screen, the number of characters per line, segmentation of a subtitle, line length, prohibited words, and culture-specific elements. Another problem that can be observed in automated transcription online is related to the absence of text. The fact that various fragments of text are either missing or inaccurate is ultimately indicated in a variety of ways by online transcription tools.

5.   Error Classification of Auto-Generated Subtitles

Romero-Fresco and Pérez (2015, p.30)**,** Romero-Fresco and Pöchhacker, (2017, p.152) classify auto-generated errors into three types:

1. Critical errors; these errors, which frequently result from replacements or misunderstandings, change the original meaning. They result in omissions or false information that seriously distorts the original message.

2. Major errors; normal errors cause information to be omitted, but they do not alter the main meaning (e.g., missing sentences regarding time or location). They break the flow, making rebuilding challenging and sometimes surprising the audience. Compared to serious errors, they are less disruptive but still apparent.

3. Minor errors; these are tiny mistakes that often do not interfere with comprehension or text reconstruction, including improper capitalization, omitted apostrophes, or minor insertions. They frequently rely on own speaking habits.

6.   Assessing Auto-Generated Subtitles

Romero-Fresco and Pöchhacker (2017, p.158) suggest the NTR model which examines source and target texts. It employs a formula to calculate accuracy rate and an error grading. Initially, this model identifies two categories of problems: translation errors and recognition errors. Translation errors are those that result from the translation process itself and can be related to either content or form. Misrecognitions that occur when the target text is dictated to the ASR program are known as recognition errors. To calculate the accuracy rate, Romero-Fresco and Pöchhacker (2017, p.159) presented the following formula

$$\underline{N - T - R} * 100 = \dots\%$$
$$N$$

According to this model, **N** represents the number of words in the subtitles, **T** is a translation error, **T** errors are classified into two types: content errors, which include omissions, additions, or substitutions (usually mistranslations), and form errors, which impact the accuracy of the subtitles like correctness (grammar, punctuation) or style (appropriateness, naturalness), **R** is a recognition by the speech-recognition software. The NTR model has three-level grading system and are used to classify both translation errors **(TEs)** and recognition errors **(REs)** by their severity (**critical given 1, major given 0.5 and minor given 0.25**). **EE**, Effective Editions, those are obviously strategic solutions on the part of the subtitler but are not considered into the formula for calculating the accuracy rate, however, they do contribute to the overall rating (Romero-Fresco and Pöchhacker, 2017, p.159).

The NTR model presented by Romero-Fresco and Pöchhacker (2017, p.159)
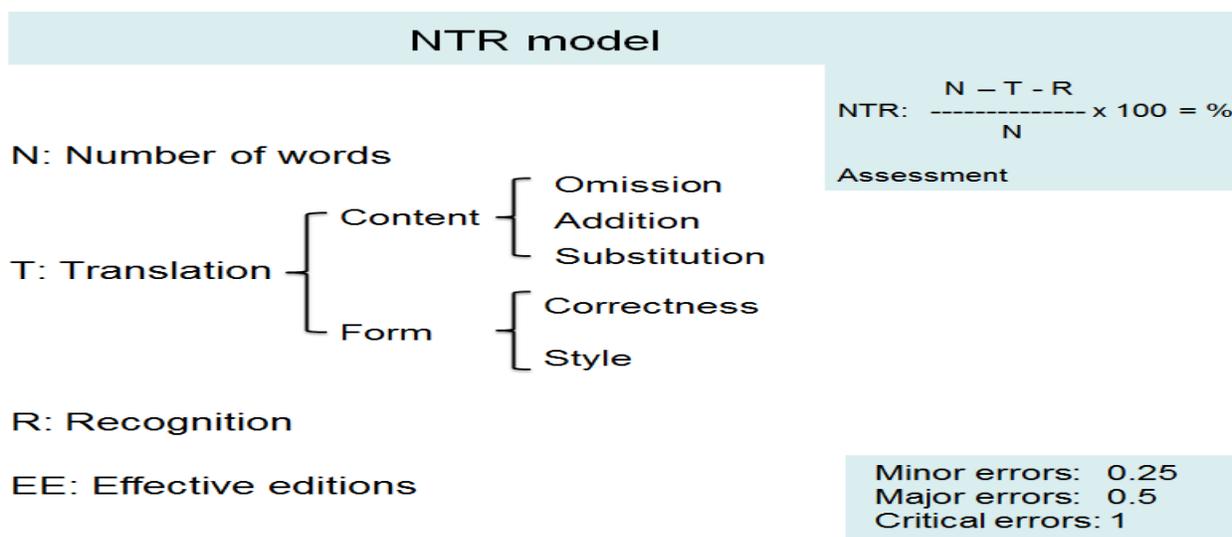
In cases where disparity arises between the accuracy rate and the conclusions derived form the broader assessment (for instance, when synchronicity or segmentation of highly accurate subtitles is deemed unacceptable), the final conclusion- rather than accuracy rate- serves as the representation of subtitle quality as assessed through the NTR model (Romero-Fresco and Pöchhacker, 2017, p.159).

To have a comprehensive assessment, Robert and Remael (2017, 170) suggest technical parameters for the final assessment, including: formatting (segmentation and layout); spotting (synchronicity with the spoken text and reading speed), and readability (in terms of font type and size).

7. Methodology and Data Analysis

This study employs a comprehensive, two-step approach to assess the quality of auto-generated subtitles. The primary phase involves applying the NTR, proposed by Romero-Fresco and Pöchhacker (2017), to evaluate the quantitative accuracy of the subtitles. This will be achieved by using their established formula to calculate an accuracy rate based on a detailed classification of errors including critical, major, and minor errors. Following the quantitative assessment, the second phase will broaden the scope of the analysis to include a qualitative evaluation of additional technical parameters. Drawing upon the insights of Robert and Remael (2017). This methodology seeks to evaluate the entire auto-subtitling process by combining the objective accuracy assessment of the NTR model with a qualitative review of these technical parameters. This approach goes beyond merely finding errors and provides a deep understanding of how AI-generated subtitles work in actual situations.

Using a case study methodology, the researcher qualitatively examined a YouTube video called (Return to Iraq: Life in the Kurdish Militia) https://www.youtube.com/watch?v=r-JzrtalwLw to recognize and classify particular issues with the automatically generated subtitles. This targeted analysis reveals more recurring issues than a broad evaluation of accuracy. This approach enables a thorough investigation of problems that might not be picked up by just quantitative measures.

Errors appearing in subtitles will be identified and categorized as either **recognition errors** (originating from Automatic Speech Recognition technology) or **translation errors** (related to content or form). Each detected error will be then classified by its severity: **minor, major, or critical**. This process serves as the foundational first step for applying the accuracy rate formula to assess auto-generated subtitles.

Errors Detecting Proces

ST	ST by (ASR) (auto-generated)	TT (auto-generated)	Comments

00:46 Salam, i'm in a town called al'amadiya, do you know where the mosul gate is?	so i'm in a town called ahmadiyya do you know where the mosul gate is so	لــذا فأنــا فــي مدينــة تسـمى الأحمديـة هـل تعرف أين تقع بوابة الموصل حتى	**RE…So is Salam in the ST. Minor (MinR)** **RE العمادية instead of الاحمدية.** **Critical (CritR)** **RE..So (the last word) is addition. Minor (MinR)** **TE of form…Punctuation** **There should be comma after العماديـة and a question mark at the end. Minor (MinT)**
00:51 it goes straight down, do you speak English?	it goes straight down do you speak English	تنـزل مباشـرة هـل تتحدث الإنجليزية	**TE of form (correctness) Punctuation, full stop or comma after مباشــرة and question mark after الانجليزيـة Minor (MinT)**
00:55 No, you do, where are you sending me? is that the	no you do where are you sending me is that the	لا انت تفعل إلـى أيـن ترسلني هل هذه بوابة	**TE of form. (correctness) Punctuation marks. We have two speakers** **no,** **you do, where are you sending me?** **Minor (MinT)**
00:58 Mosul gate okay cool yeah that's it Mosul gate this is	muzzle gate okay cool yeah that's it muzzle gate this is	الفوهة حسنا رائع نعم هذا هو بوابـة الفوهـة هذا هو	**RE الموصل instead of الفوهة …** **Critical (CritR)* 2** **TE of form (style)…okay, cool, yeah are not necessarily translated. Minor (MinT)**
01:08 this gate was built in the 14th century, the sun represents life-giving energy	this gate was built in the 14th century the sun represents life-giving energy	تم بناء هذه البوابة فـي القــرن الرابــع عشــر تمثـل الشــمس طاقـة الحياة	**TE of form (correctness) Punctuation, there should be comma after عشــر. Minor (MinT)**
01:14 and the snakes protection from	and the snakes protection from	وحمايـة الثعـابين مـن الخطر	

danger 01:20 over there is Iran, behind me is Turkey, just pushed over	danger over there is iran behind me is turkey just pushed over	هناك إيران خلفي هـي تركيا تم دفعها للتو	**TE of content (substitution)** وتمثل الثعابين الحماية من الخطر **Major (MajT)** **TE of form (correctness)** Punctuation. There should be comma after ايـران and تركيـا. **Minor (MinT)**
01:37 My name's Adnan Sawa in 2003 I served as a British soldier in the Iraq war	[Laughter] my name's adnan sawa in 2003 i served as a british soldier in the iraq war	[ضحك] اسمي عدنان خدمت كجندي سوا، بريطـاني فـي حـرب العراق عام 2003	**TE of content (substitution)** تم **. دفعها للتو Major (MajT)** سـاوا..**TE of form (correctness)** instead of سوا**. Minor (MinT)** **TE of form (style)…Laughter doesn't need to be translated. Minor (MinT)**
02:10 and its people, you welcome anytime buddy, like never before	and its people you woke up anytime buddy like never before	وشعبها، لقد استيقظت فـــي أي وقـــت يـــا صديقي، كما لم يحدث من قبل	**RE….welcome instead of woke up. Major (MajR)**
02:30 from the mountains in the north via the capital Baghdad	from the mountains in the north via the capital baghdad	من الجبال في الشمـال عبر العاصمة بغداد	**TE of content (mistranslation)** والى العاصمة**. Minor (MinT)**
02:50 I'll make friends, come face to face with old enemies	i'll make friends come face to face with old enemies	سـأجعل الأصـدقاء يواجهون أعداء قدامى وجها لوجه	**TE of content (substitution). Major (MajT)* 2** سأصنع صداقات I'll make friends, والتقي come face to face with… وجها لوجه بأعداء قدامى there is a silence between the two texts
02:55 I'm going to meet one of the people	i'm going to meet one of the people	سـأقابل أحـد الأشـخاص الـذين قصـفوني أنـا	**RE…ask… I will ask in the**

who bombed me and my friends, and ask if	who bombed me and my friends and asked if	وأصـدقائي وسـألوني عما إذا كان	**present tense which affects the accuracy of the TT..the meaning ,... اذا عمـا واسـأل.. Minor (MjnR)**
03:00 the country can escape its cycle of violence?	the country can escape its cycle of violence	بإمكان البلاد الهروب من دائرة العنف	**TE of form (correctness) punctuation. It needs a full stop or comma after اصدقائي.**
03:04 four men are going to die for their crime, the door to war has always been open in	four men are going to die for the door to wall has always been open in	أربعة رجال سيموتون بسـبب جـريمتهم كـان البــاب إلــى الجــدار مفتوحا دائما في	**TE of form (correctness) punctuation...this should be a question not a statement because he enquires the possibility. Minor (MinT)** **TE of form (correctness) punctuation..there should be comma or full stop after جريمتهم . Minor (MinT)** **RE...wall is war..Critical (CritR)**
03:20 Journey in the danger zone: Iraq	………… …………..	………… …………	
03:29 twice the size of the UK, Iraq is home to 37 million people	twice the size of the uk iraq is home to 37 million people	ضـعف حجـم المملكـة المتحـدة العـراق هـو مـوطن لـ 37 مليـون شخص	**RE on screen text not detected by ASR, Major (MajR)* 5 (missing words) منطقـة الـى رحلـة الخطر: العراق**
03:35 it's divided into 18 provinces, this is Iraqi Kurdistan	it's divided into 18 provinces this is iraqi kurdistan	مقسم إلى 18 محافظـة هـذه هــي كردسـتان العراق	**TE of form (correctness) punctuation it needs comma after المتحـدة المملكــة. Minor (MinT)**
03:41 split across Iraq Syria, Turkey and Iran, many Kurdish people long for an 03:47	split across iraq syria turkey and iran many kurdish people long for an	مقسـمة بـين العـراق وسـوريا وتركيــا وإيـران يتـوق العديـد من الأكراد إلى	**TE of form (correctness) punctuation, there is a silence between محافظة and هذه, there should be comma. Minor (MinT)**

independent Kurdistan, in Iraq they've suffered decades of	independent kurdistan in iraq they've suffered decades of	كردستان مستقلة في العراق لقد عانوا من عقود من	**TE of form (correctness) punctuation. There should be comma or full stop after ايران. Minor (MinT)**
04:00 in a very special convoy up to mount Gara	in a very special convoy up to mount gara	في قافلة خاصة جدا حتى جبل غارا	**TE of form (correctness), punctuation, there should be a comma after مستقلة, Minor (MinT)**
04:07 look at this look at the female Peshmerga flying the flag	look at this look at the female fish burger flying the flag	انظر إلى هذا انظر الى برجر السمك الانثوي الذي يرفع العلم	**TE of content (mistranslation) up to should be translated into اعلى. Minor (MinT)**
04:13 this is crazy this is crazy we've got the Peshmerga and we've got a bear	this is crazy this is crazy we've got the fresh murgo and we've got a bear	هذا جنون هذا جنون لقد حصلنا على مورغو الطازج ولدينا دب	**RE fish burger is misrecognized by ASR it is Peshmerga البيشمركة. Critical (CritR)**
04:17 Riding alongside a bear, look at that, do you think he's enjoying it? look it's	riding alongside a bear look at that do you think he's enjoying it look it's	يركب بجانب دب انظر الى ذلك هل تعتقد انه يستمتع به انظر انه	**RE fresh murgo is misrecognized by ASR it is Peshmerga. Critical (CritR)**
04:25 he's loving it I'm with animal activist Blend	he's loving it i'm with animal activist blend	انه يحبه أنا blend مع ناشط الحيوان	**TE of content (mistranslation) يركب is mistranslated, it should be نقود بجانب. Minor (MinT)** **TE of form (correctness), there should be a comma after دب and ذلك and a question mark after يستمتع به . Minor (MinT)**
04:34 bears that have			**TE of form (style)…Blend**

been held in captivity into the wild	bears that have been held in captivity into the wild	الدبــــة التـــي تـــم احتجازها في البرية	**Brifcany is the name of the activist.. Blend comes at the beginning of the text in the TT, confusing translation. Major (MajT)**
04:39 Blend with all the trouble that's going on in	blend with all the trouble that's going on in	مع كل المشـاكل التـي تحدث فيblend	**TE of content (mistranslation)** الدببة التي تم احتجازها لتطلق في البرية **Major (MajT)**
04:42 Iraq, why are you doing this? this must be down the list of priorities rescuing	iraq why are you doing this this must be down the list of priorities rescuing	العراق لماذا تفعل هذا يجب أن يكون هذا في أسـفل قائمـة أولويـات انقاذ	**TE of content (mistranslation) blend again couldn't be translated as a name, Minor (MinT)**
04:48 We mustn't stop We mustn't say the situation is bad, there is war so I'll stay at home, No, everyone must do what they can	……………… ……………… ……………….. …………… …………… …………	…………… …………… …………… …………… ……………… ……………… ……………..…… ……………	**TE of form (correctness), punctuation, there should be question mark after** لمــاذا تفعل هذا؟. **Minor (MinT)** **RE no screen texts not detected by ASR.. Minor (MinR)* 20 (missing words)**
04:59 Brown bears are native to these mountains but hunting and habitat loss	ground bears are native to these mountains but hunting and habitat loss	الدببــة الأرضـية هـي موطنها الأصـلي لهـذه الجبــال ولكـن الصـيد وفقدان الموائل	يجب الا نتوقف، يجب الا نقول ان الوضع سيء، هنـاك حرب فسـأبقى فـي البيـت، لا، على كل شخص ان يفعل ما بوسعه
05:03 means there's hardly any left and blend wants to change that	means there's hardly any left and blend wants	blendتغييــر ذلــك يعني أنه لم يتبـق منهـا	**RE of content (mistranslation) the translation is not correct** تعد الجبـال المـوطن الأصلي للدببـة البنيـة.

	to change that	سوى القليل ويريد	**Major (MajR)**
05:14 now it's a tourist spot with a twist, this is all a minefield	now it's a tourist spot with a twist this is all a minefield	إنهــا منطقــة الآن سياحية مـع لمسـة هذا كله حقل ألغام	**TE of form (style), Minor (MinT)** **TE of content (mistranslation) blend is a name, Minor (MinT)** **TE of form (style) the translation is confusing the reader, Major (MajT)**
05:23 Mines, Mines, they're in a minefield	mines mines they're in a minefield	الألغام الألغام انهم فـي حقل ألغام	**TE of form (correctness), punctuation, there should be comma or full stop after منطقة سياحية, Minor (MinT)** **TE of content (mistranslation) منطقـة سـياحية بطـابع متميـز, Minor (MinT)**
05:36 and cleared whole villages to make way for one of his palaces	and cleared whole villages to make way for one of his palaces	وطهـر قـرى بأكملهـا لإفسـاح المجـال لأحـد القصور	**TE of form (style) no need for repetition it could be translated into الغـام فـي كـل مكـان, Minor (MinT)**
05:42 so what what's that there? Saddam's house	so what what's that there saddam's house	اذن مـاذا هنـاك منـزل صدام	**TE of content (substitution) his palaces should be translated into احد قصوره, Minor (MinT)**
05:45 Saddam's house on top of that mountain yeah we're going to release them here	salem's house on top of that mountain yeah we're going to release them here	منـزل سـالم علـى قمـة هذا الجبل، نعم سنطلق سراحهم هنا	**TE of form (correctness), punctuation, this is a question and answer it should be like منزل صدام....ماذا هناك؟ Minor (MinT)**
04:49 yes, well let's go release some bears oh we're going to go	yes well let's go release some bears oh we're gonna go	نعم حسنا دعنـا نـذهب لإطـلاق سراح بعض الدببة أوه سنذهب	**RE salem is saddam, Critical (CritR)**

05:53 words got out about today's event and there's a scrum of locals and media	words got out about today's event and there's a scrum of locals and media	انتشرت كلمات حول حدث اليوم وهناك تدافع من السكان المحليين ووسائل الإعلام	**TE of form (style) no need for translation of oh, Minor (MinT)**
06:00 the bears have been kept as pets and Blend paid thousands of dollars for them	the bears have been kept as pets and blend paid thousands of dollars for them	تم الاحتفاظ بالدببة كحيوانات أليفة ودفع مزيج آلاف الدولارات مقابلها	**TE of content, (mistranslation), Minor (MinT),انتشرت الاخبار**
06:05 what is happening now Blend	what is happening now blend	ما يحدث الآن مزيج	**TE of content (mistranslation) blend is a name should not be translated, Major (MajT)**
06:07 We are all happy right now all the volunteers We'll set these animals free and make Kurdistan more beautiful	……………….. ……………….. ……………….. ……………… ……………… ……………….. ………………..	……………… ……………… ……………… ……………… ……………… ……………… ………………	**TE of content (mistranslation) blend is a name, Minor (MinT)** **TE of form (correctness), punctuation, there should be a question mark at the end, Minor (MinT)**
06:16 first of for release	………………..	………………	**RE no screen text not detected by ASR, Minor (MinR)* 13** **نحن جميعا سعداء, (missing words) كل المتطوعين, سوف نطلق سراح هذه الحيوانات ونجعل كردستان اجمل**
06:18 is the younger bear nearly there	is the younger bear nearly there	الدب الأصغر سنا تقريبا هناك نذهب	

there we go 06:38 that looks like she's never been out to the wild doesn't it 06:41 they're doing the big one, the locks are 06:44 yeah the door is on the bear is out 06:53 Screams and Noises 07:08 Screams and Noises 07:18 i mean he didn't he didn't actually attack that person, he just dived on top of them 07:25 Screams 07:35 it's become this spectacle you know and the bear is confused,	there we go that looks like she's never been out of the wild doesn't it they're doing the big one the locks are yeah the door is off the bear is out [Music] [Applause] i mean he didn't he didn't actually attack that person he just dived on top of them [Applause] it's become this spectacle you know and the bear is confused	 يبدو أنها لم تخرج مـن البرية أبدا أليس كذلك يفعلـون الشـيء الكبيـر الأقفال نعم الباب مغلق الـدب بالخارج [موسيقى] [تصفيق] أعني أنـه لـم يفعل لـم يهاجم هذا الشخص في الواقع لقد انقض عليهم فقط انهم [تصفيق] وأصـبح هـذا المشـهد انـت اعـرف والـدب مرتبك إنه يتجول فقط	**RE of content (omission), not detected by ASR, Minor (MinR)* 4 (missing word),** اول من سنطلق سراحه **TE of form (style)** الدب الأصغر سنا هيا ننطلق **Minor (MinT)** **RE out to the wild** لم تخرج الى البرية, **Major (MajR)** **TE of content (mistranslation)** انهـم يحـررون الـدب الكبيـر, **Major (MajT)** **RE the door is on, it should be translated as** البـاب مفتـوح , **Major (MajR)** **RE screams not music, Minor (MinR)** **RE screams not applause Minor (MinR)** **TE of form (style), no need for repetition of** لم يفعل لم يهاجم **It should be translated into only** لم يهاجم, **Minor (MinT)**

it's just going around	it's just going around		**RE screams not applause, Minor (MinR)**
07:41 he doesn't know what to do look it's sniffing that tripod	he doesn't know what to do look it's sniffing that tripod	ولا يعـرف مـاذا يفعـل انظر إنـه يشـم الحامـل ثلاثي القوائم	**TE of content (mistranslation) either it shouldn't be translated or it is translated into كمـا ترى, Minor (MinT)**
07:55 look it's walking on the road it's going to Saddam's palace	look it's walking on the road it's going to saddam's old hands	انظـر إنـه يسـير فـي الطريق إنه ذاهب إلـى أيدي صدام القديمة	**TE of form (correctness), punctuation, there should be comma after يفعل, Minor (MinT)**
08:01 it's like a Monty python sketch isn't it this is crazy	it's like a monty python sketch isn't it this is crazy	إنـه مثـل رسـم مـونتي بايثون أليس كذلك هـذا جنون	**RE, old hands is old palace, Major (MajR)**
08:16 undeterred by near disaster Blend gives me the honor of releasing a Chukar partridge	undeterred by near disaster blend gives me the honor of releasing a chukka partridge	شـرف اطـلاق سـراح غيـر عـابئ Blend بالكارثـــة الوشـــيكة يمنحني	**TE of content (mistranslation), It should be translated into انه يشبه مشهد كوميدي من مونتي بايثن مشهد كوميدي Or** **Major (MajT)**
08:20 the national bird of Iraq, you ready? yes ready	the national bird of iraq you ready yes ready	الطائر الوطني للعراق هـل انـت مسـتعد نعـم مستعد	**TE of form (style) confusing translation,** **غير عابئ بالكارثـة الوشيكة.. منحني بلند شرف اطلاق سراح طائرالحجل التشوكر, Major (MajT)** **RE the name of the bird is misrecognized, Minor (MinR)** **TE of content (mistranslation) the name Blend, Minor (MinT)**
08:33 it's not surprising	it's not surprising	وأصـــدقائه يقـــدرون	

Blend and his friends value freedom so much	blend and his friends value freedom so much	Blendالحرية كثيرا انه ليس مفاجئا ان	TE of form (correctness), punctuation,**هل انت مستعد؟** نعم**, Minor (MinT)**
09:14 ISIS also known as Daesh held Mosul for three brutal years	isis also known as daesh held mosul for three brutal years	داعش المعروفة أيضــا باسـم داعـش سـيطرت علـى الموصـل لمــدة ثلاث سنوات وحشية	**TE of form (style) confusing translation, Major (MajT)** **TE of content (mistranslation), the name blend, Minor (MinT)**
09:32 ISIS's last stand was in the west of the city across the river Tigris	isis's last stand was in the west of the city across the river tigris	كـــان آخـــر موقـــف لـــداعش فـــي غـــرب المدينة عبر نهر دجلة،	**TE of form (style), no need for this addition to the translation, isis is known as داعش, Minor (MinT)** **TE of content (mistranslation) Minor (MinT) كانت اخر مواجهة,**

Errors number and weight

Now after detecting different kinds of errors, the NTR formula is adopted to assess the accuracy of about 10 minutes of the video. Giving each error its weight ranging between (0,25- 0.5 to 1). **Total Number of words = 880**

NTR Accuracy $\underline{880-15.5-23.5} * 100 = 95.7\%$

$$880$$

The second stage of the assessment is for other parameters. The technical parameters will be evaluated based on subtitling norms to determine their acceptability or unacceptability. This assessment will identify whether the subtitles are of high quality or if they require improvements. The   final

Translation Errors (TEs)	Automatic Speech Recognition Errors (REs)
MinT 44* 0.25= 11	MinR 40* 0.25= 10
MajT 9* 0.5= 4.5	MajR 13* 0.5= 6.5
CritT 0	CritR 7*1= 7
Total TEs = 15.5	Total REs = 23.5

judgment will be a clear determination of their adherence to established standards.

Technical and Readability Parameters

Category	Parameters	STs	TTs	Grading	Comments
A Formatting	**Segmentation**	01:57 this is incredible	هـذا امـر لا يصدق، لقد	They are all illogical line breaks and poor segmented. Too short or too long. They need resegmenting in acceptable and logical way.	لقد belongs to the next segment
		02:55 and asked if	وسألوني عما إذا كان		عمـــا اذا كـــان belongs to the next segment
		03:09 iraq and forge a better future	العـــراق ويصــــنعون مستقبلا افضل		العـراق belongs the previous segment
		03:15 really wrong	خطأ كبيـرا	Unacceptable line breaks, they need improvements	خطـــا كبيـــرا belongs to the previous segment
		03:47 independent kurdistan in iraq	كردســـتان مســـتقلة فـــي العراق		كردستان مستقلة belongs to the previous segment
		03:52 repression under saddam hussein	القمـــع في عهد صدام		القمـــع فـــي... belongs to the previous segment
		04:17 he's enjoying it look it's	يسـتمتع بـه انظر انه		انـه belongs to the next segment
		04:30 briefcarney and	مــــن briefcarney		Briefcany belongs to the

		04:42 iraq why are you doing this	<u>العراق</u> لمـاذا تفعل هذا		previous segment العـراق belongs to the previous segment الحيوانـــــــــات belongs to the previous segment
		04:47 animals	<u>الحيوانات</u>		
		05:25 throwing snowballs at each other	كـرات الـثلج علـى بعضـهم البعض <u>زرع</u>		زرع belongs to the next segment
		06:41 they're doing the big one the locks are	يفعلون الشيء الكبير <u>الأقفال</u>		الاقفـال belongs to the next segment
		07:29 uncomfortable about this because	عدم الارتياح حيال هذا <u>لأنه</u>		لأنـه belongs to the next segment
	Layout	Acceptable, positioned consistently in the middle and down of the screen, clear display and easy to see against the background			
B Spotting	**Synchronicity**	08:20 the national bird of Iraq you ready yes ready	الطائر الـوطني للعراق هل انت مسـتعد نعـم مستعد	Unacceptable Bothering readers, it needs improvements	Stayed 5 sec after speaker stops talking
		08:42 their land from invaders and occupiers for generations	أرضـهم مـن الغــــزاة و المحتلــــين لأجيال	Unacceptable	Stayed 4 sec ……

		09:00 today I'm leaving Kurdistan and entering Mosul in Iraq's Nineveh province	اليــوم أغــادر كردســــــتان وأدخـــــــل الموصــل فــي محافظة نينـوى العراقية	Unacceptable	Stayed 5 sec ……
		09:56 this is this is the worst thing I've seen in my life	وهــذا أســوأ شيء رأيته فــي حياتي	Unacceptable	Stayed 5 sec ……
	Reading Speed	Acceptable, comfortable reading pace			
C Readability	Font type and size	Acceptable, Legible			

According to an examination of the auto-generated subtitles' technical specifications, segmentation and synchronicity were the main problems. The largest frequency of errors was found in these two areas, suggesting a major difficulty in the automated process of breaking up speech into coherent subtitle units and accurately aligning them with the audio. Other technical factors, on the other hand, such reading speed and font type and size, were generally deemed acceptable. The system seems to control these parameters within accepted standards, indicating that although the fundamental aspects of subtitle display are managed well, the intricate and challenging jobs of timing and segmentation continue to be the main areas where auto-generated subtitling has to be improved. The following table will show the metrics of unacceptable segmentation and synchronicity based on the number of subtitles (102).

Subtitle Segmentation and Synchronization Metrics

Metric	Count	Percentage
Total number of subtitles	102	100.0%
Segmentation, unacceptable	12	11.7%
Segmentation, acceptable	90	88.3%

Synchronicity, unacceptable	4	3.9%
Synchronicity, acceptable	98	96.1%

For accurate subtitle assessment, each of the five technical parameters will be assigned a weight of 20 points, totaling 100 points. This weighting system ensures that the technical aspects are balanced with the overall accuracy of the subtitles. This will facilitate a comprehensive evaluation of the final product.

Technical Parameters Metrics

Parameters	Out of 20
Segmentation	17.7
Layout	20
Synchronicity	19.3
Reading speed	20
Font type and size	20
Total rate	97%

The final stage of the evaluation will incorporate the effective editions made in generated subtitles. This involves calculating the number of changes that significantly improve the quality and value of the subtitles. The following table will clarify the areas of affective editions.

Effective Editions EEs

STs (ASR generated)	TTs	Comments on EEs
01:08 this gate was built in the 14th century the sun represents life-<u>giving</u> energy	تم بنـاء هـذه البوابـة في القرن الرابـع عشر تمثل الشمس <u>طاقة الحياة</u>	<u>giving</u> in the ST is omitted in the TT but the meaning is correct (effective omission)
01:48 now 15 years on <u>hello</u> hello i'm back come on	،والآن بعـد 15 عامـا، <u>مرحبـا</u> لقـد عدت، هيا	(effective omission) of one <u>hello</u>
02:05 you could not believe that this was the <u>right</u> to <u>get to</u> know this land wow	لا يمكنك تصديق أن هذا كان من <u>حقك</u> <u>التعرف</u> على هذه الأرض، يا إلهي	(effective من الحـق instead <u>من حقك</u> substitution) <u>to get to know</u> is translated as <u>التعرف</u> (effective omission)
02:19 with its borders <u>drawn by</u> the british empire after world war one	بحدودها التـي <u>رسـمتها</u> الإمبراطوريـة البريطانية بعد الحرب العالمية الأولى	converting passive into active in Arabic
02:42 i'll learn about life here hello <u>hey</u> what's up from the iraqis themselves	سـأتعلم عـن الحيـاة هنـا، مرحبـا، مـا الاخبار من العراقيين أنفسهم	(effective omission) of <u>hey</u> because it doesn't need to be translated

03:04 four men are going to die for the door to wall has always been open in	أربعة رجال سيموتون بسبب جريمتهم كان الباب إلى الجدار مفتوحا دائما في	for their crime although not detected by ASR but translated in the TT (effective addition)
03:52 repression under saddam hussein and more recently fought off isis	القمع في عهد صدام حسين وقاتلوا داعش مؤخرا	under is translated into في عهد (effective substitution)
09:56 this is this is the worst thing i've seen in my life	وهذا أسوأ شيء رأيته في حياتي	no repetition of this is in the TT, (effective omission)

Upon evaluation, only 9 effective editions were identified across a total of 102 subtitles, resulting in an effective editing rate of **8.8%.** This low percentage suggests that machine translation is often unable to resolve complex subtitling issues on its own, highlighting the continued need for human post-editing to enhance quality.

Balancing Table: Technical Parameters, Accuracy, and Effective Edits

Dimension	Metric / Score	Interpretation
Technical parameters	97%	High level of technical soundness in auto-generated captions; however, the need improvement in segmentation and synchronicity
Accuracy	95.7%	Good accuracy in transcription/recognition and content handling; however the need post edition with recognition and translation errors
Effective edits	8.8%	Very low rate of edits that meaningfully improve quality
Overall implication	Reliance on auto-generated captions without review is not advisable High technical/accuracy alone does not ensure quality	

8. **Findings**

The study found the following:

1. A discrepancy between the high scores for technical parameters (97%) and accuracy (95.7%) and the very low rate of effective editions (8.8%).

2. Auto-generated captions are reliable from a purely technical standpoint. They can accelerate the subtitling workflow but are not sufficient on their own. Auto segmenting and synchronizing still need human intervention.

3. Auto captions indicate generally good content fidelity but do not guarantee error-free outputs; they still require human verification.

4. Low rate of effective editions suggests that machine-produced captions frequently fail to correct critical issues; substantial human post-editing is necessary.

9. **Conclusion**

This study investigated AI-generated subtitling that integrates Automatic Speech Recognition (ASR) and machine translation (MT). The study's major hypotheses—that these AI-generated subtitles often contain various kinds of errors and that human post editing, segmentation, and synchronization are still necessary to fix these errors and guarantee overall quality—were supported by the study's findings. In order to evaluate the accuracy of the subtitles, this study used the NTR model, which classifies errors into three categories: minor, major, and critical. It also included technical parameters in order to offer a comprehensive assessment of the quality of the subtitles. The study further found a disparity between accuracy and technical parameters on the one side and automatic effective editions on translation on the other side which shows the inability of machine and AI systems to solve some critical errors in translation. The study came to the conclusion that while AI technology might speed up and simplify subtitling, human control is crucial to obtaining true accuracy.

References

1. Almurashi, W. (2016). The Effective Use Of Youtube Videos For Teaching English Language In Classrooms As Supplementary Material At Taibah University In Alula. International Journal of English Language and Linguistics Research, 4–4(3), 32–47.
2. Chiaro, D. (2009). Issues in audiovisual translation. In Munday, J., The Routledge companion to translation studies (pp. 155-179). Routledge.
3. Cintas, J. D., & Remael, A. (2014). Audiovisual translation: subtitling. Routledge.
4. Costan Davara, G. (2020). Audiovisual Translation: Subtitling Netflix documentary a Black Hole Apocalypse, Dissertation.
5. Gray, A. (2017). These are the world's most popular websites. In World Economic Forum (Vol. 10).
6. Gregori, A. (2021). Automatic Speech Recognition (ASR) and NMT for Interlingual and Intralingual Communication: Speech to Text Technology for Live Subtitling and Accessibility, University of Bologna.
7. Karakanta, A. (2022). Experimental research in automatic subtitling: At the crossroads between machine translation and audiovisual translation. Translation spaces, 11(1), 89-112.
8. Lewis, W. D. (2015). Skype translator: Breaking down language and hearing barriers. *Translating and the Computer (TC37)*, *10*, 125-149.
9. Mera, M. (1999). Read my lips: Re-evaluating subtitling and dubbing in Europe. Links & Letters, 73-85.
10. Pedersen, J. (2011). Subtitling norms for television: An exploration focusing on extralinguistic cultural references. Stockholm University.
11. Robert, I. S., & Remael, A. (2017). Assessing quality in live interlingual subtitling: A new challenge. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16*.

12. Romero-Fresco, P., & Pérez, J. M. (2015). Accuracy rate in live subtitling: The NER model. *Audiovisual translation in a global context: Mapping an ever-changing landscape*, 28-50.

13. Romero-Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16 pp.149–167*.

14. Singh, Y., Kumar, R., Kabdal, S., & Upadhyay, P. (2023). Youtube video summarizer using NLP: A review. *International Journal of Performability Engineering*, *19*(12), 817.

15. Tosatto, C. (2019). Audiovisual Translation: An Analysis of Subtitling in the TED Talks. In Università degli Studi di Padova & Dipartimento di Studi Linguistici e Letterari, Tesi Di Laurea [Thesis].

16. Varga, C. (2021). Online Automatic Subtitling Platforms and Machine Translation. Buletinul Stiintific al Universitatii Politehnica din Timisoara, Seria Limbi Moderne, (20), 37-49.

17. Xie, B. (2023, September). Machine Translation Implementation in Automatic Subtitling from a Subtitlers' Perspective. In Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track (pp. 54-64).

Online References

18. Chavez, P. (November 19, 2009), Automatic captions on YouTube, https://publicpolicy.googleblog.com/2009/11/automatic-captions-on-youtube.html

19. Team, T. Y. T. (2008). Auto translate now available for videos with captions. Retrieved from YouTube Official blog, from https://blog.youtube/news-and-events/auto-translate-now-available-for-videos/