Quality Assessment of Interlingual YouTube Auto-generated Closed Captions in Some Crime Narratives Applying the NTR Model

Rania Allam

Associate Professor,
Faculty of Languages,
October University
for Modern Sciences
and Arts (MSA),
Egypt.

Abstract

Audiovisual Translation (AVT) is a prolific milieu of cross-cultural communication. YouTube is one of the most prominent user-generated, social media streaming platforms. YouTube employs artificial intelligence (AI) devices namely, automatic speech recognition (ASR) and neural machine translation (NMT) to add auto-generated closed captions (CCs) as interlingual subtitles to its broadcasted videos. The present study attempts to assess the translation quality of these CCs to gauge their reliability as mediating tools enhancing culture and entertainment. Translated, auto-generated CCs on three YouTube videos on true crime channel entitled *Twisted Minds* are scrutinized in March 2023 applying Romero-Fresco and Pöchhacker's (2017) NTR model. Results show that the translated CCs are accurate with only **(95%)** approximately with a rate of less than the minimum starting point according to **(0/10)** scale suggested by the NTR mode. Errors of translation content and form as well as speech recognition errors are spotted, indicating a suboptimal translation quality. The auto-generated CCs display reasonable acceptability in what concerns AVT norms, yet with some deviations. Despite such acceptability and instances of positive effective editions of translational manipulation displayed in the CCs, the profuse errors mar the denotative and connotative meanings of the overall content of the crime narratives exhibiting semantic and pragmatic failures. A revisit analysis for the same data is conducted in December 2023, showing an accuracy rate of **(97.31%)** approximately with a rate of **(3+/10)** on the NTR model accuracy scale. Improvement is rather notable, yet the accuracy rate is still poor. This proves that seamless ongoing human intervention on the linguistic, semiotic and technical levels in the performance of the YouTube AI devices is much needed to achieve notable advancements in the quality of its auto-generated translated CCs, to be considered a reliable tool that can help demolish communication barriers.

Keywords: Audiovisual translation, YouTube closed captions, Artificial Intelligence, Automatic speech recognition, neural machine translation, NTR model

Quality Assessment of Interlingual YouTube Auto-generated Closed Captions in Some Crime Narratives Applying the NTR Model

Rania Allam

1. Introduction

Audiovisual content is a prolific semiotic milieu of cross-cultural communication. It involves entertainment, education and even self-expression. This entails the necessity of having translated, interlingual subtitles that boost comprehension and enhance mediation. However, in audiovisual translation (AVT) "the linguistic challenges of translation are supplemented by semiotic and technical issues that audiovisual translators have to deal with" (Malaczkov, 2020, p. 6).

Since the advent of the 21st century, the AVT scene has extended to include more forms other than the traditional, professional channels. "C(c)loud-based platforms have been progressively introduced in the translation workflow of audiovisual media conglomerates and localisation companies with an aim to reduce costs, increase productivity and optimise networked environments" (García-Escribano et al., 2021, p. 2).

Such technical multifaceted developments need to be explored. As "Web 2.0-mediated translation" has lately become a prolific research area in Translation Studies (Kraeva & Krasnopeyeva, 2020, p. 777). For instance, media broadcasting cloud platforms incorporate the option of adding automatically generated closed captions (CCs) to the videos by making use of artificial intelligence (AI). CCs are switched on and off by the video user and this service is provided as early as 2008 and 2009, by Google for YouTube user-generated videos. "L(l)ive closed captioning is a currently active area of study" (Ruiz-Arroyo et al., 2022, p. 1). Acoustic, auto-translated textual representation of the content presented on some YouTube videos can display a lot in such domain. Therefore, the prime focal point may be the quality of the translated output.

2. Literature Review

2.1 YouTube Auto-generated Captions

YouTube is one of the most popular channels of cloud communication. It is "the largest free video sharing site in the world" (Lee & Cha, 2020, p. 144). Being an open social platform that permits an unlimited number of video-sharing uploads, it provides myriad functions with accessible links and HTML codes. Such "accessibility touches upon various ways of communicating between people and, ultimately, all areas of life, including work, education, citizenship and societal participation, and culture" (Hirvonen & Kinnunen, 2021, p. 474).

Furthermore, YouTube is a user-generated pipeline that permits uploading of various genres of videos that exhibit cultural manifestation and artistic creation. It has become a "dominant part of the multi-platform digital media industry with revenues reportedly in the billions of US dollars" by 2017 (Burgess & Green, 2018, p. 40). The platform revenues are multiplied by including both auto-generated CCs (for hearing impaired persons) or auto-translated subtitles (for multilingual recipients) to the user / amateur-produced media content.

Hence, a means of achieving "linguistic accessibility in media could be to use AI technology like automatic speech recognition (ASR) and neural machine translation (MT) to provide automatic interlingual subtitles for audiovisual content" (Tuominen et al. 2023, p. 77). The inclusion of CCs as interlingual subtitles

into YouTube videos is actualized by merging two AI digital giant tools namely, **Google Voice Recognition** and **Google Translate**. Therefore, the process of fully automatic subtitling is actualized by combining a neural machine translation (NMT) device together with auto-spotting and automatic segmentation devices to achieve accurate linguistic synchronization with the acoustic media content (Karakanta, 2022, p. 89).

Such epochal technological leap manifested in speech-to-text (STT) translation "infiltrates not only the social life of the individual but also the way in which the external environment is being moulded" (García-Escribano et al., 2021, p. 6). This is because AI translated, auto-generated CCs spare the screenplays creators the effort of resorting to expensive, time-consuming traditional subtitling services. Since the percentage of views mounts by adding such intra and inter lingual service, its omnipresence entails particular concern about its quality. However, the assessment of the translation quality of such automatic subtitles "remains under-researched and calls for comprehensive context-dependent studies" (Kraeva & Krasnopeyeva, 2020, p. 777).

2.2 Automatic Speech Recognition

ASR is a type of STT "process in which a computer interprets a person's speech and then converts the contents into text" (Song et al. 2019, p. 1). It comprises an acoustic model of monolingual speech data transcribed as statistical sound representations of waveforms, a lexicon with a repository of words with their varieties of pronunciation and a language model with an n-gram of adjacent words; a data file of predefined set of words and another grammar file for correct combinations of such words (Carrier 2017; Ciobanu and Secară 2020).

ASR models meant for Voice-Based NMT are statistically converted into language models using certain algorithms like the Hidden Markov Model (HMM) Toolkit, for instance. However, such conversion is not free from errors which have an impact on the quality of the subtitling captions in the translation pipeline. Originally, ASR and NMT architectures are actualized by cascade systems in which the NMT receives a raw textual verbatim transcription from the ASR, that may exhibit some errors. Lack of identifying pauses/periods positions and specious sentence boundary detection may result in rendering multiple sentences to be identified as one, thus vitiating the performance of the MT system. "What's even worse is that YouTube manages subtitles by time slots, not by utterances" (Song et al. 2019, p. 2).

Attempts to improve the performance of the ASR are quite auspicious. For instance, Karakanta et al. **(**2020) contrive an ASR direct speech end-to-end ST systems rather than cascade pipelines of transcribing, segmenting and predicting timestamps by using speech signals and prosody to be later translated by NMT. Their proposed system works on an input with the temporal dimension of the speech, taking into consideration frequency, pitch and other prosodic elements**.** Therefore, their system is likely to produce subtitles that meet the subtitling industry guidelines. Nonetheless, despite such attempts to improve the verbatim outcome of the ASR setup with variant inputs of pronunciation, dialect varieties, syntactic and semantic language model variations, some errors still occur.

2.3 Neural Machine Translation

One of the elements of auto-generating CCs is NMT. It is one of "the leading workhorses" in the advancement transition that has taken place in AVT industry (Karakanta 2022, p. 90). Since YouTube has made use of Google Translate NMT device to convert the identified text from the ASR stream into subtitled captions, the binge of multilingual media content has blasted, yet, with a quality that is not above

questioning. The distinctive fabric of written-to-be-spoken language used for AVT scripts, the broad range of genres, the visual and physical contexts create tangible challenges to NMT. Besides, MT "output quality is affected by the grammatical irregularity in audiovisual texts and the language pair" (Xie 2022, p. 4). Therefore, the semiotic features of AVT can sometimes be dimmed or even marred by NMT.

Lately, there have been attempts to contrive NMT systems designed for subtitling, like using segmenter modules mimicking human segmentation decisions, for instance. Furthermore, speech translation (ST), image-guided translation (IGT), video-guided translation (VGT) have undergone notable hikes with the advancement of incorporating both visual and acoustic multimodal elements to the process of the NMT translation system. Such novel integration comprises "auto-spotting, shot change detection and audio-informed segmentation" (Karakanta, 2022, p. 2). However, the auto-generated translated captions still exhibit noticeable errors.

2.4 ASR and NMT

The usability of NMT subtitle productivity is explored more than once in projects like the *SUbtitling for MAchine Translation* (*SUMAT*), Translation for Massive Open Online Courses (TraMOOC) and later the European Multiple MOOC Aggregator project (EMMA) (Díaz Cintas and Massidda 2020, p. 262). Such projects evaluate the validity of combining NMT and ASR in the subtitling process and thus achieve notable strides in the AVT industry. With incessant developments in both techs, consistency, productivity and validity of the automated throughput have increased. Nevertheless, attempts to scrutinize the quality of the user-generated translation produced in socio-technical, multimodal translation orifices are still rather erratic.

3. Aim of Study

Research in social, multimodal media outlets is still in its infancy, as "audio captioning is a novel field of multi-modal translation" (Lipping et al., 2019, p. 1). Accordingly, to fulfill constant upgrading of the field of translation studies in general and AVT studies in particular ample, context-dependent studies are genuinely needed. Therefore, the present study aims to investigate the quality of the auto-generated CCs presented on some YouTube videos in an endeavor to add an extra contribution to such domain.

4. Research Questions

The quality assessment conducted in the present study seeks to answer the following research questions:

1) What is the quality of the translated auto-generated CCs presented in the videos under study in what concerns the performance of the NMT and ASR software devices according to the accuracy rates suggested by the NTR model?
2) What are the changes undergone to the Target texts (TTs) that can be considered more as strategic manipulations of AI AVT than deviations from the source texts (ST) narrated in the videos?
3) To what extent does the applied model of analysis prove to be a valid model for evaluating interlingual NMT subtitling quality?

5. Data of Study

True crime stories are a genre that "is arguably moving centre stage aligned to our recent obsession with the real life figure of the serial killer" (Peters, 2020, p. 23). The data selected for analysis in the present study are the translated, auto- generated CCs on three YouTube videos of the user-generated channel *Twisted Minds.* The total number of words understudy in the translated versions is more than 6000 words

approximately. Such number may be helpful in giving a wider view on the quality of the translated output.

The selected three episodes are true crime narratives about vicious murderers and persecuted victims. This genre is selected in particular as "d(D)igital platforms have heralded a renaissance in true crime production and engagement, providing previously marginalised voices with opportunities to challenge longstanding – and often problematic – genre conventions" (Hobbs & Hoffman, 2022, p. 26).

As a prime spot of audience attracting media genre, the popularity of crime genre entails a necessity of making it more accessible by adding auto-translated captions on the videos. Besides, the type of discourse employed in these crime narratives has "numerous traits of specific police discourse style such as specialized vocabulary, repetitive use of certain specific phrases and cohesion devices" (Filipović & Gascón, 2018, p. 70). Such pre-prepared scripts comprise rather well formulated, standardized linguistic content, nearly free from malformed utterances.

Furthermore, the opted for episodes under study are broadcast with studio-quality sound, appositely vocalized audio track and aptly narrated by a single speaker. The videos are almost free from overlapping, incomplete, spontaneous spoken discourse "such as disfluencies, repetitions, grammatically incorrect and/or incomplete sentences etc." (Stüker et al. 2007, p. 2). Except for some music tracks separating the narrative pauses, there are no inaudible, background noises or clamors. From the ASR angle, such content supposedly yields less faulty captions and thus, depict a clearer image about the true quality of the YouTube NMT.

The language pair in the present study is English-Arabic. Compared to other translation directions in the digital milieu, "translating from English into Arabic in particular is a rare translation direction and often yields significantly lower results" (Al-Obaidli et al., 2018, p. 4). Accordingly, the present study attempts to answer some research queries concerning this translational direction.

6. Methodology

To answer some of such queries, the auto-generated CCs on the video narratives needs to be probed using both a quantitative and qualitative approach, to assess the AI performance in what concerns auto translation quality and screen presentation. This entails a quality assessment model that is supposed to be meticulous and detailed rather than traditional and holistic model, merely inspecting translation equivalence. Nonetheless, AVT, since its moment of creation, has put the notion of mere equivalence against the ropes (Chaume, 2018, p. 84). Therefore, the assessment is to entail facets other than translational equivalence, despite being a major assessment aspect.

The model applied in the present study to assess the quality of interlingual YouTube CCs is Romero-Fresco and Pöchhacker (2017) **NTR model**. It is based on Romero-Fresco and Martínez's (2015) NER model, in which, N stands for Number of words, E for Editing errors and R for Recognition errors. NER gauges the accuracy rate in automated, intralingual live subtitles. It analyzes "the extent to which errors affect the coherence of the subtitled text or modify its content" (Romero-Fresco and Martínez, 2015, p. 1).

In Romero-Fresco and Pöchhacker's (2017) NTR model, (N stands for: Number of words in the subtitles, T stands for: Translation errors in content (omissions, additions or substitution), form (grammar, terminology) or style (naturalness, register) and R errors stand for: Recognition errors (the usage of software)). Accordingly, in the present study, a T error lies mainly in false NMT rendering of the ST linguistic content, while an R error lies in erroneous

ASR software recognition of the ST acoustic content.

The scale of both T and R errors severity is **minor**, **major** or **critical**. Minor errors are insignificant deviances that have trivial effect on the content or form. Major errors are detectable, yet the ST meaning is still perceptible and can be mentally rectifiable. Critical errors are conspicuous, upsetting the AVT outcome changing the ST meaning substantially. The whole model can be presented as follows:

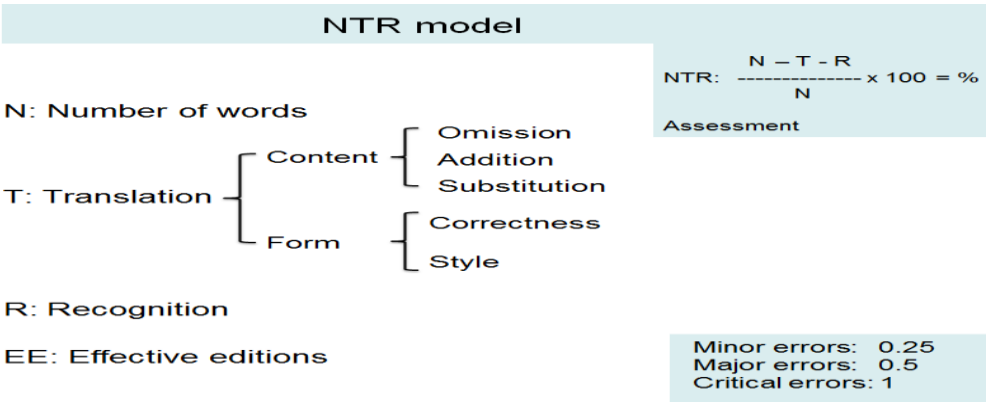

Figure (1): NTR model (Romero-Fresco and Pöchhacker, 2017, p. 159)

According to the NER model, the allowed percentage of errors in intralingual subtitling is **98%**-word accuracy rating. Similarly, in the NTR model, **98%** is the minimum accuracy rate for interlingual live. However, since this high percentage fits more intralingual live subtitling than the interlingual, the NTR model gauges the accuracy rate according to a more standard 10-point scale, presented as follows:

Accuracy (%)	10-point scale
< 96	0/10
96.40	1/10
96.80	2/10
97.20	3/10
97.60	4/10
98.00	5/10
98.40	6/10
98.80	7/10
99.20	8/10
99.60	9/10
100.00	10/10

Table (1): 10-point scale accuracy rate (Romero-Fresco and Pöchhacker 2017, pp. 159-160).

NTR model, however, does not only entail error counting accuracy rate, but also assessing positive facets. The model accounts for positive TT deviations from the ST termed as "effective editions" (EE). Such nonconformities do not involve a loss of information or marring of communicative effectiveness. On the contrary, they may be considered more as tactical strategies rendering ample output than translational deviances. Furthermore, the model has accessible parameters to gauge the different dimensions of AVT like; fluency, adequacy and AVT screen format **(**Karakanta, et al., 2020, p.67). However, being a human evaluation model, NTR may have "the benefit of rationale judgement but also the disadvantages derived from subjectivity the extent to which errors affect the coherence of the subtitled text or modify its content" (Romero-Fresco, 2020, p. 741). However, the rather accuracy rate formula of the NTR model renders it more objective than other assessment models. The model also affords a conclusion that may account for inconsistency between the translation quality and any deviation of the accurate

AVT norms and practices either in form or content.

To scrutinize the quality of the auto-translated CCs that appear on the videos under study, the original English ST transcripts are automatically downloaded from the YouTube auto-generated transcript option box for each video. Then, full Arabic scripts are manually transcribed for comparable evaluation. Detection of translation and recognition errors is perused, employing NTR parameters to reach an eventual number of errors. Penalty scale is calculated for each error respectively. Final calculations of accuracy scores together with overall comments of positive EE instances and AVT facets are to be concluded from the analysis. Effective editions are underlined and errors appear in italics, yet in the present study for further clarification errors are in bold also. Correct suggestions for erroneous CC sample segments are clarified (between inverted commas). These are the key abbreviations for the model analysis:

Translation = T	Content = Cont	Omission = omiss
Addition = add	Substitution = sub	Correction = corr
Recognition = R	Critical = crit	Major = Maj
Minor = Min	Errors = Es	Effective Editions = EE

7. Data Analysis

Employing the methodological NTR parameters suggested by Romero-Fresco and Pöchhacker (2017) a sample analysis of the quality of the auto-generated CCs of the three crime stories goes as follows:

7.1 Error Types

7.1.1 Translation Content Errors

ST	TT
The destruction of the house where the murders occurred ***only adds insult to injury it's as if*** ***the city has wiped the slate clean*** erasing any trace of the heinous act	فإن تدمير المنزل الذى وقعت في جرائم القتل ***يضيف فقط إهانة للإصابة كما لو*** ***أن مسحت المدينة تنظيف الأردواز ومحو*** أى أثر للفعل الشنيع
Main Error Type: ***Substitution*** **Errors:** [(MajT) (Cont-sub)] [(MajT) (Form-corr)] [(MajT) (Cont-sub)] [(MinT) (Form-corr)]	

The substitution of the ST lexemes into literal TT equivalents exhibits both semantic and pragmatic failures. The AI NMT fails to convey the idiomatic meaning of the ST phrases ***adds insult to injury*** and ***has wiped the slate clean***, which mean (to exacerbate an
existing disappointing situation) and (to forget previous conflicts and problems to start freshly) respectively (Collins COBUILD, n.d.). The faulty rendition of the ST idiomatic expression ***adds insult to injury*** as ***يضيف فقط إهانة للإصابة*** fails to express the connotative significance of the ST idiom, that can be translated into " زاد الطين بلة" indicating exacerbations of awfulness and pain. Similarly, the literal rendering of the ST idiom ***the city has wiped the slate clean*** as ***مسحت المدينة تنظيف الأردواز*** does not only fail to denote the preceding atrocious crime of a mother taking the lives of her four innocent sons, but also displays a flawed form structure by substituting the lexeme لوح with the TT

noun ***تنظيف*** and preceding the subject ***المدينة*** by the verbal phrase ***كما لو مسحت***. This aggravates the error, producing an even more bizarre TT equivalent, whose pragmatic, connotative meaning can be aptly conveyed as " محت المدينة آثار الماضى " المؤلم for instance, insinuating the mother's excruciating act.

ST	TT
Some killers dip their ***toes in the murky waters of taking lives*** but when it came *to* ***Metheny*** ***he took a run up and belly flopped right*** in there and what ***a splash*** the 450-pound man made	قام بعض القتلة بغمس ***أصابعهم*** فى ***المياة المظلمة لتوديع الأرواح، ولكنه*** عندما ***يتعلق بغاز الميثان***، فقد ركض وأخذت *بطنه تتخبط* هناك ***وما الذى*** صنعه الرجل الذى يبلغ وزنه 450 رطلا.
Main Error Type: ***Substitution*** **Errors:** [(MajT) (Cont-sub)] [(MajT) (Cont-sub)] [(MajT) (Cont-sub)][(MinT) (Form-corr)] [(MinT) (Cont-omiss)] [(MajT) (Cont-sub)][(MajT) (Cont-sub)] [(MajT) (Cont-omiss)] [(MinT) (Form-corr)]	

The substitution errors in this ST metaphorical image are quite profuse to even entail other translation (omission) and form ones (structure), drastically distorting its valid pragmatic, connotative implications delineated in the ST. First, the lexeme ***toes*** is imperfectly substituted with: ***أصابعهم*** instead of "أصابع أقدامهم". Translating the gerund ***taking lives*** as ***توديع الأرواح*** instead of "أخذ الأرواح" implies a paradoxical meaning of saying goodbye to the dying dear ones not ripping innocent souls. Moreover, the erroneous rendering of the phrase ***ولكنه عندما يتعلق*** instead of " ولكن عندما يتعلق الأمر" and the repeated misrecognition of the proper name Metheny as ***Methane*** (the poisonous gas) lead to a flawed TT phrase that displays both specious denotative and connotative failures. Finally, what completely dims the ST conceptual metaphorical image of diving into thick waters of vice and murder is the faulty grammatical structure and the lexical omission in translating the ST phrase what ***a splash*** the 450-pound man made as ***وما الذى*** صنعه الرجل الذى يبلغ وزنه 450 ويالها من دفقة صنعها الرجل الذى " instead of رطلا يبلغ وزنه 450 رطلا" for instance.

ST	TT
While being kept in a ***segregation unit*** in Renault Legere established a good relationship with the staff ***Warden*** Don Wheaton later commented that Legere had a Jekyll and Hyde personality	أثناء احتجازه فى وحدة ***الفصل العنصرى*** في رينو. أقام ليجير علاقة جيدة مع الموظفين وعلق ***واردن*** دون ويتون لاحقا بأن ليجير كان له شخصية جيكل وهايد
Main Error Type: ***Substitution*** **Errors:** [(CritT) (Cont-sub)] [(MajT) (Cont-sub)]	

The ST noun ***segregation*** is imprecisely substituted with the TT modified equivalent ***الفصل العنصرى*** . In this ST forensic register, the modified nominal group "الحبس الإنفرادى" would be a more pertinent equivalent, as the crime narrative has nothing to do with racial segregation. The full stop placed at the end of this

independent phrase produces inaccurate line segmentation rendering the readability of the whole caption flawed. Moreover, the faulty transliteration of the ST lexical item ***Warden*** (a job title) as a proper noun ***واردن*** instead of "حارس" hinders the overall contextual perception of the preceding and subsequent phrases.

ST	TT
was attempting to retrieve some clothes and tools from their ***residents he*** ran to a nearby gas station	وكان يحاول استعادة بعض الملابس والأدوات من منزلهم. ركض ***السكان*** إلى محطة وقود قريبة
Main Error Type: ***Addition*** **Errors:** [(CritT) (Cont-add)]	

Additional lexical items can also cause semantic and pragmatic failures. The ST noun ***السكان*** is imprecisely added after the ST noun resident which is accurately translated into منزلهم. The error is categorized as critical as the added nominal group ***السكان*** fits in as the subject of the phrase instead of the pronoun he (which refers to Dawson the boyfriend), rendering a wrong narrative plot that may cause subsequent confusion.

ST	TT
after ***the first failure he lured*** another woman ***which*** ended exactly the same	بعد الفشل الأول ***الذى أغوى أخرى المرأة التى*** انتهت بنفس الطريقة تماما
Main Error Type: ***Addition*** **Errors:** [(MajT) (Cont-add)] [(MajT) (Form-corr)] [(MinT) (Form-corr)] [(MinT) (Cont-add)]	

The additional relative pronoun ***الذى*** which precedes the verbal phrase ***أغوى أخرى المرأة***(which in turn is structurally erroneous, supposedly "أغوى أمرأة أخرى") causes semantic and pragmatic confusions. As instead of marking a start of a new clause relating the events of Metheny alluring another woman, the additional relative pronoun implies that his failure to find his eloped wife allures another woman. This consequently leads to the addition of another relative pronoun ***التى*** rendering the subsequent phrase even more confusing and specious.

ST	TT
The whole Miramichi Community was terrified ***kids were kept Indoors***	شعر مجتمع ميراميتشى بالرعب ***من أن*** الأطفال تم إبقاؤهم في منازلهم
Main Error Type: ***Addition*** **Errors:** [(CriT) (Cont-add)]	

The prepositional phrase ***من أن***is inadequately added preceding the nominal clause الأطفال تم إبقاؤهم في منازلهم. This mars the crime narrative which recounts the dread experienced by the residents of Miramichi community to let their children outdoors lest be murdered or kidnapped by the bolted serial killer. The prepositional phrase inaccurately denotes that the people are afraid from keeping their kids indoors; a meaning which is not only imprecise but totally contradictory to the one intended by the ST CC.

ST	TT
in a fit of Fury Susan had taken her 38 caliber revolver and fired it at the heads of the boys ***even reloading at one point***	فى نوبة غضب ***التقطها*** سوزان. مسدسها من عيار 38 ***وأطلقتها*** على رؤوس الأولاد
Main Error Type: ***Omission*** **Errors:** [(MinT) (Form-corr)] [(MinT) (Form-corr)] [(CritT) (Cont-omiss)]	

In omission errors, the missing lexical items render the TT CC vague and faulty. The drastic omission of the participial phrase ***even reloading at one point***, which can be translated into " حتى أنها أعادت تحميله", dims a vital forensic detail in the crime narrative. The fact of the mother reloading the crime weapon at a point is later employed by the prosecutors to prove that she deliberately murdered four sons. The crime is quite intended not a result of a block out or any kind of mental disturbance, thus she is sentenced to death.

ST	TT
The prostitute wasn't giving any answers ***frustrated he SA'd*** *her* before strangling her to death	ولكن العاهرة لم تقدم أى اجابات ***محبطة*** قبل أن *يخنقها* حتى الموت
Main Error Type: ***Omission*** **Errors:** [(CritT) (Cont-add)] [(MajT) (Cont-omiss)] [(MajT) (Cont-omiss)]	

Two major omission errors occur rendering the TT CC totally inaccurate. The ST abbreviated verbal group ***SA'd*** which means "sexually assault" is omitted dimming the meaning of the TT clause, as Metheny 's sexual assault against prostitutes constitutes the crime main forensic plot. Moreover, omitting the pronoun ***he*** which refers to Metheny does not imply that he is the perpetrator. This leads to another critical error of addition, in which the ST participle modifier ***frustrated*** referring to the male subject pronoun ***he*** (Metheny) is erroneously translated into ***محبطة***. The addition of the feminine inflection تاء مربوطة to the noun implies that the prostitute herself is the frustrated one. This dims the fact of Metheny's strangling the women to death and sexually assaulting her, thus implying a defective meaning marring the forensic facts.

ST	TT
Luckily for the kidnap pair Legere ***took his eye off the ball*** and the two managed to escape the car	لحسن الحظ، ***حيث قام الزوج المختطف بإخراج عينه، عن الكرة*** وتمكن الاثنان من الهروب من السيارة
Main Error Type: ***Omission*** **Errors:** [(CriT) (Cont-omiss)] [(MinT) (Form-corr)] [(MajT) (Cont-sub)]	

The drastic omission of the subject Legere (the culprit) falsely implies that the doer of the action is the kidnapped pair ***الزوج المختطف***; a meaning which is not only altered from the original, but also utterly incongruous. The mal structure is

heightened by the faulty misplacement of the lexical items. In addition, the idiomatic meaning in the ST idiom ***took his eye off the ball***, which means not to pay attention for a short while, resulting in negative consequences (Collins COBUILD, n.d.) is totally marred by the AI NMT. It is literally and invalidly translated into ***بإخراج عينه، عن الكرة*** instead of "وتحول انتباه ليجير لبرهة", for instance. What aggravates the distortion of the ST idiomatic phrase is the misplacement of the comma which precedes the prepositional phrase ***عن الكرة*** implying a fallacious, incomprehensible sense. Therefore, the ST meaning could be conveyed using the TT clause "لحسن الحظ الزوج المختطف حيث تمكن تحول انتباه ليجير لبرهة الاثنان من الهروب من السيارة".

7.1.2 Translation Form Errors

ST	TT
Dodson was living with ***Susan and her four boys they had taken*** some Valium and ***seemed to be having a*** good time	*كان دودسون يعيش مع سوزان* ***معها. أربعة أولاد أخذوا*** بعض الفاليوم وبدا ***أنهم يقضون*** وقتا ممتعا
Main Error Type: ***Correctness*** **Errors:** [(MajT) (Cont-sub)] [(MinT) (Form-corr)] [(MinT) (Form-corr)] [(MajT) (Form-corr)] [(MajT) (Form-corr)] [(MajT) (Form-corr)]	

The grammatical structure of the TT is not equivalent to the ST, despite the fact that, the sentence structure is not complicated. Nonetheless, the narrator's pauses may cause some confusion to the ASR segmenter, leading to substituting the linking word "و" with the adverbial phrase ***معها*** even putting a full stop after it. The faulty translation of the modified nominal group ***her four boys*** into ***معها. أربعة أولاد*** instead of"وأولادها الأربعة" conveys an imprecise meaning. In addition, the flawed usage of the masculine plural inflectional suffix واو الجماعة in ***أخذوا*** and ***يقضون*** instead of the masculine dual inflectional suffix "الف الأثنين" as in "أخذا يقضيان", and the usage of the masculine plural pronominal suffix هم in ***أنهم*** instead of the masculine dual pronominal suffix "أنهما" lead to an invalid, bizarre interpretation for the sentence; as if the mother, her boyfriend together with the four kids had some Valium, instead of "كان دودسون يعيش مع سوزان و أولادها الأربعة أخذا بعض الفاليوم وبدا أنهما يستمتعان بوقتهما".

ST	TT
Metheny was given the ***nickname Tiny*** which was a joke	وأطلق على ***الميثان اسما صغيرا*** كان مزحة
Main Error Type: ***Correctness*** **Errors:** [(MinT) (Form-corr)] [(MajT) (Form-corr)] [(CritT) (Cont-sub)]	

Changing the structure of the ST noun ***Tiny*** to be an adjective ***صغيراً*** to be modifying the nickname not Metheny himself is considered a major distortion of form, dimming the meaning of the ironical fact of Metheny's huge frame. The ST phrase can simply be translated into "وأطلق على *ميثيني* اسم الصغير وكان مزحة".

ST	TT
When a Killer is finally captured the communities ***in which they were lurking***	عندما يتم القبض على قاتل أخيرًا ، فإن المجتمعات التي ***كانوا يتربصون فيها*** تتنفس الصعداء

give a sigh of relief	
Main Error Type: ***Correctness*** **Errors:** [(MinT) (Form-corr)] [(MinT) (Form-corr)] [(MinT) (Form-corr)]	

The absence of the agreement between the TT subject***قاتل*** and the inflection on its verbal group ***كانوا يتربصون*** affects the syntactic correctness of the sentence. Besides, the erroneous usage of the TT preposition ***فيها*** instead of "بها" may render the structural fabric of the TT phrase flawed, though still conceivable.

ST	TT
The boyfriend told Eric that Susan ***had gone off the rails*** she had ***spiraled into madness***	أخبر صديقها إريك أن ***سوزان قد خرجت عن القضبان، فقد تصاعدت إلى الجنون***
Main Error Type: ***Style*** **Errors:** [(MajT) (Con-sub)] [(MajT) (Form-style)]	

The NMT software fails to recognize the ST idiom ***gone off the rails*** as an idiom, which means acting in a sociably unacceptable, deplorable way (Collins COBUILD, n.d.). Thus, literally translating the idiom as***قد خرجت عن القضبان*** instead of " " خرجت عن السيطرة exhibits both semantic and pragmatic failures. These failures extend to stylistic inconsistency in translating the ST verbal group ***had spiraled into*** as ***فقد تصاعدت إلى*** instead of "اقتربت بسرعة نحو "; a translation that may seem semantically comprehendible, yet stylistically awkward and unnatural in the TT.

ST	TT
after being freed from prison Methane was free to act upon his fantasies ***cooked up*** during his time behind bars	وكان الميثان حراً فى التصرف بناءً على تخيلاته التى ***تم طهيها*** خلال فترة وجوده خلف القبضان
Main Error Type: ***Style*** **Errors:** [(Form-style) (MajT)]	

The translation of the ST phrasal verb ***cooked up*** as ***تم طهيها*** is not semantically erroneous, yet stylistically rather influent. As the TT noun تخيلات is not naturally collocated with the verb تطهى. Besides, it could be more stylistically acceptable to retain the metaphorical shades of meaning in the TT by translating it as " " تم تصورها\نسجها so as to be more expediently appropriate.

ST	TT
However Michelle managed to contact the authorities police quickly ***poured into*** the area	لكن ميشيل تمكنت من الاتصال بالسلطات ***التى تدفقوا*** بسرعة على المنطقة
Main Error Type: ***Style*** **Errors:** [(Form-style) (MinT)] [(Form-corr) (MinT)]	

The translation of the ST phrasal verb ***poured into*** as ***تدفقوا***is a rather unnatural collocation with the subject noun سلطات. Police forces, in this criminal narrative context, can be better stylistically collocated with the verb "هرعت " for instance.

7.1.3 Speech Recognition Errors

ST	TT
Eric arrived at the house they left the house together	***وصلت أيريكا*** إلى المنزل وغادرا المنزل معا
Main Error Type: ***Speech Recognition*** **Errors:** (MajR)	

The proper noun Eric is identified by the ASR as ***Erica***, as it is followed by a verb with an initial vowel sound "**/ə**ˈ/". This led to the MT putting a feminine verb inflection to the verb ***وصل***. A major speech recognition error, yet not a critical one as there is no character in the crime narrative with such a name, and such an error can be detected by the viewer.

ST	TT
he ***SA'd*** them strangled and cut them up	حيث ***طلب منهم*** خنقهم وتقطيعهم
Main Error Type: ***Speech Recognition*** **Errors:** (CritR)	

The ASR software fails to recognize the abbreviation ***SA'd*** of the verbal group sexually assault, dimming a significant detail in the crime narrative plot, and instead recognizes it as the verbal group ***asked*** and thus mistakenly translating it as ***طلب منهم***. This misrecognition renders the meaning of the TT clause not only obscure but also illogic, as the perpetrator can never ask the victims to strangle them and chop their bodies.

ST	TT
That would be become surprising when looking at the ***MO*** of this monster	سيصبح مفاجئًا عند النظر ***إلى ذخيرة*** هذا الوحش
Main Error Type: ***Speech Recognition*** **Errors:** (CritR)	

The ASR software fails to recognize the abbreviations of the Latin words "modus operandi" which means a way of practice or method of doing something (Traub, n.d.). It recognizes it as the closest phonetic parallel "*ammo*" [ˈæməʊ] (which is an abbreviated form of ammunition) ***ذخيرة***. This can be considered a critical error as the murderous register of the narrative can permit such word thus resulting in a variant faulty meaning.

Furthermore, instances in which the ASR software displays mal function are sporadic, but truly operative in the NMT performance. For instance, in what concerns proper nouns the ASR exhibits notable inconsistency. Proper nouns like ***Eubanks***, ***Metheny*** and ***Legere*** are translated using different typographical

forms like ***يوبانك, يوبانكز*** and ***يوبانكس, ميثنى***, ***المیثان, ميثينى*** and***غاز الميثان*** and ***ليجير***, ***ليجيرى***and***ليغرى*** respectively. In other segments, the proper nouns are transferred into Arabic in Roman characters. Since these different forms refer to the same persons, they do contribute to a rather unsmooth viewing experience.

Besides, the ASR critical unrecognition of proper nouns causes both syntactic and pragmatic confusion, hindering the comprehension of the narrative plot, as the main actor/doer of the action is omitted. For instance, the omission of the proper noun *Legere* in unarmed guards escorted ***Legere to the Dr. George L Demond Hospital*** قام حراس غير مسلحين بمرافقة***الطبيب إلى مستشفى جورج إلى ديموند*** causes a critical error of confusing events of the crime narrative. It seems as if the guards escort a physician to a hospital called George L Demond, conveying an illogic denotative meaning, opposite to the whole forensic context.

Inconsistency of performance of both the ASR and the NMT devices can be spotted in falsely rendering proper nouns as ST lexical items to be translated into TT lexemes; like in translating ***Magaziner مجلة***, ***Olive الزيتون*** and ***Flame اللهب***. Congruently, transliteration errors can be detected like in treating the ST verb ***rob*** as a proper noun like in Todd Maskip and Scott Curtis ***rob*** the convenience store تود مسكيب وسكوت كيرتى المتجر ***روب***. Such misrecognition conveys a completely specious meaning causing semantic and pragmatic failures within the fabric of the narrative. Moreover, the standard cliché introduction of the crime story channel Welcome or welcome back to *Twisted Minds* is not consistently translated. In *Metheny*'s story it is translated into مرحبا أو مرحبا بك مرة أخرى إلى العقول الملتوية. In *Legere's* story it is rendered as ***Twisted Minds*** مرحبا أو مرحباً بكم مرة أخرى فى. However, the rather flawed transference is in *Eubanks*'s story: ***Twisted Minds ترحب أوترحب بالعودة إلى حلقة أخرى من***, maybe as a result of the preceding feminine lexeme ***والدتهم*** leading to inflecting the verb in conformity with feminine inflection. Inconsistency can be also distinguished in translating a single lexical item in various ways within the same video. For instance, in *Metheny*'s story the word ***pallet*** is translated into***البيليت*** , ***المنصات***and***الحبيبات*** ; a rendition that may cause confusion in comprehending the plot narrative.

In addition, YouTube restriction policy on using lexemes of sexual obscenity or physical abuse sometimes leads to the deliberate muting of such lexemes by the video creators. This results in changeable renderings in the TT. For instance, in *Metheny*'s story the word "prostitute" is sometimes muted and substituted by bracketed dots. In other instances, it is bluntly translated into العاهرة. In *Legere*'s story the word "raping" is muted in the ST video and thus untranslated in the TT CCs.

7.2 Accuracy Rates

The previous quantitative qualitative analysis for the YouTube auto-generated CCs in the three crime narrative videos yields some results that may be helpful in deciding their translation quality. The eventual calculations of accuracy rates resulted by applying the NTR model can be exhibited as follows according to the (0-10) scale:

Content Errors

[(MinT) (Cont-sub) = 68] + [(MinT) (Cont-add) = 47] + [(MinT) (Cont-omiss) = 40]

+

[(MajT) (Cont-sub) = 67] + [(MajT) (Cont-add) = 6] + [(MajT) (Cont-omiss) = 200]

+

[(CritT) (Cont-sub) = 13] + [(CritT) (Cont-add) = 5] + [(CritT) (Cont-omiss) = 2]

= [448]

Form Errors

[(MinT) (Form-corr) = 172] + [(MinT) (Form-style) = 11]

+

[(MajT) (Form-corr) = 102] + [(MajT) (Form-style) = 3]

= [288]

[Content Errors = 448] + [Form Errors = 288] = [Translation errors = 736]

Speech recognition errors

[MinR = 2] + [MajR = 2] + [CritR = 7]

= [11]

[736 Translation errors] + [11 Speech recognition errors] = [747 Number of errors]

NTR accuracy rate: $\frac{6093-293.5-8.5}{6093} \times 100\% = 95\%$ **approximately (-0/10)**

Errors Percentages

Translation/Content	
Error Type & Frequency	**Percentage%**
(MinT) (Cont-sub) 68	9.23%
(MinT) (Cont-add) 47	6.38%
(MinT) (Cont-omiss) 40	5.43%
(MajT) (Cont-sub) 67	9.10%
(MajT) (Cont-add) 6	0.81%
(MajT) (Cont-omiss) 200	27.17%
(CritT) (Cont-sub) 13	1.76%
(CritT) (Cont-add) 5	0.67%
(CritT) (Cont-omiss) 2	0.27%

Table (2)

Translation/Form	
Error Type & Frequency	**Percentage%**
(MinT) (Form-corr) 172	23.36%
(MinT) (Form-style) 11	1.49%
(MajT) (Form-corr) 102	13.85%
(MajT) (Form-style) 3	0.40%

Table (3)

Speech Recognition	
Error Type & Frequency	**Percentage%**
(MinR) 2	18.18%
(MajR) 2	18.18%
(CriT) 7	63.63%

Table (4)

Table (2): Content errors, Table (3): Form errors, Table (4): Speech recognition errors

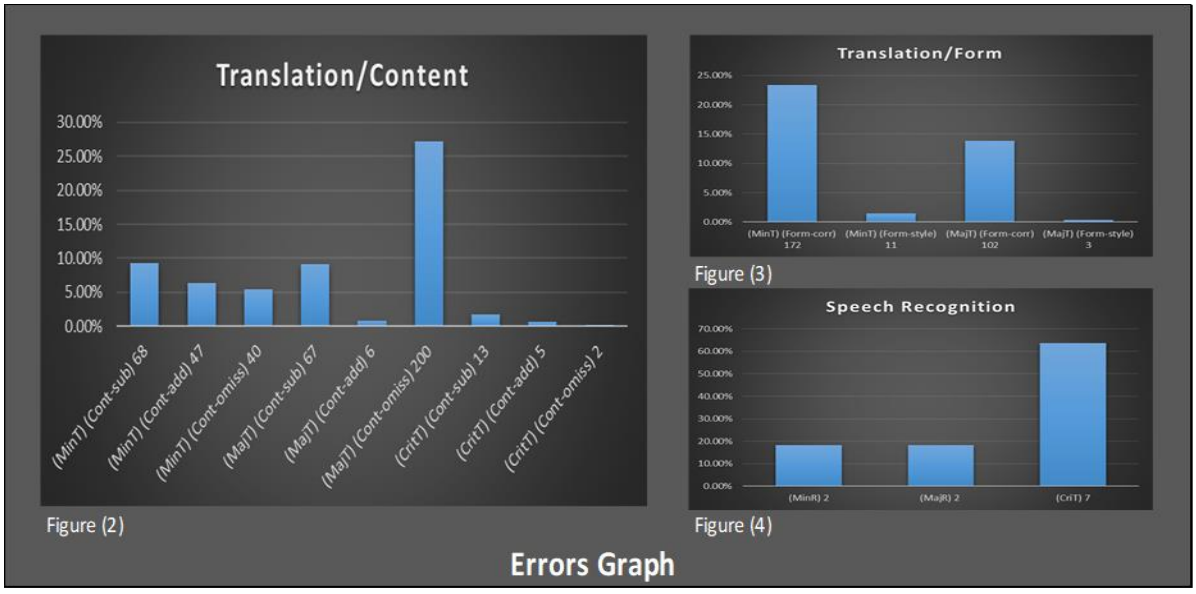

Figure (2): Content errors, Figure (3): Form errors, Figure (4): Speech recognition errors

Accordingly, it can be inferred that, the error accuracy rates of the translated AI CCs presented in this study are of a very poor quality according to the NTR accuracy rates. According to Romero-Fresco and Pöchhacker (2017) 10-point suggested scale, **95%** approximately is below the starting accuracy rate which is **<96%**. Accordingly, the auto-generated CCs translational lexical bulk is suboptimal, not even averagely accurate, and both the ASR and NMT devices need radical upgrades.

Semantic and pragmatic errors of content namely substitution, addition and omission constitute approximately (**60.86%)** of translation errors, thus the YouTube AI NMT device experience notable deficiency on the lexical level. The highest percentage of the content translation errors -**(54%)** approximately- is categorized as omission errors. Such errors cannot be categorized as speech recognition ones as their occurrence cannot be decisively attributed to a mal function in the ASR in particular. However, omission errors can anyway be categorized as translation content errors, signaling notable deficiency of performance of the NMT device.

The second highest percentage of the content translation errors is identified as substitution ones with (**33%)** approximately. In these instances, the NMT device is unable to opt for accurate equivalent TT lexemes, resulting in specious alteration of meaning and faulty creation of context. This is manifested by the fact that, the least percentage of content translation errors-(**13%)** approximately- is categorized as addition errors, most of them minor ones.

From the structural point of view, form correction errors constitute (**37.22%)** approximately of the translation errors. They include mainly incorrect word order, faulty structural fabric and mal inflection of verbal groups. Although the higher percentage of these form correction errors are minor ones; (**62.77%**) approximately, they negatively affect the quality of the produced CCs.

The other type of form errors, namely style errors, is not that profuse in the data under study. They constitute only **(4.86%)** approximately of the form errors and (**1.9%)** approximately only of the translation errors. This percentage does not have a radical deleterious effect on the CCs translation quality, yet it affects the fluency, smoothness and naturalness of the produced TTs and the AVT viewing experience as a whole.

Concerning, the ASR software device acoustic errors in the CCs, they are intermittent constituting a fringe percentage of (**1.47%)** approximately of the total errors. It is worth noting that, the ASR misrecognition errors are mainly critical ones constituting (**63.63%)** approximately of the ASR errors. This directly mars the content and the logic of the TT CCs.

7.3 Revisited Data Translation Accuracy Rates

Since translation errors comprise (**98.52%)** approximately of the total number of errors detected in the auto-generated CCs under study, this indicates that the YouTube AI software is in dire need for notable upgrades. This is rather manifested by the fact that, the quantitative and qualitative results of the previously analyzed crime narratives have been retrieved in March 2023 from the YouTube with a very poor translation accuracy rate of (**95%)** approximately.

However, when the same YouTube data is revisited by the researcher in December 2023, a remarkable variance of accuracy rates is spotted. The revisited December data estimated by the same NTR model scale increases to (**97.31%)** approximately with a scale of **(3+/10)**. Though the quality of the same auto-generated CCs is still rather poor, it has at least achieved a notable increase rate.

Some remarks can be distinguished from such revisited data. The huge percentage of omitted content data is conspicuously reduced from **(54%)** to **(25.71%)** approximately of content translation errors. This shows remarkable upgrading of the detection accuracy in the YouTube AI software. Translation errors in general are reduced from (**736)** error instances in March to (**402)** in December. Content translation errors in March constitute (**60.86%)** approximately of the translation errors**,** while content translation errors in December constitute (**43.53%)** approximately of the translation errors**.** Some of March data semantic and pragmatic failures are pertinently transferred into indicative TT equivalents in December data, exhibiting AI software prominent progression on the lexical and idiomatic levels, detected in some examples as follows:

ST
she's ***a little whacked***
when it comes to the ***appearance*** of Joe Metheny it would be easy to assume he is pure evil
After ***cutting the power*** the thieves broke into the store

March TT	**December TT**
لقد تعرضت ***للضرب قليلا***	لقد كانت محطمة قليلا
عندما يتعلق الأمر ***بظهور*** جو ماثينى، سيكون من السهل افتراض أنه شرخالص	عندما يتعلق الأمر بمظهر جوماثيني سيكون من السهل افتراض أنه شريرتماما
بعد ***قطع القوة التي اقتحمها*** اللصوص المتجر	بعد قطع التيار الكهربائى، اقتحم اللصوص المتجر

In March data translation form errors constitute (**39.13%)** approximately of the translation errors, while in December data they constitute **(56.46%)** approximately. This shows that, the NMT device does not exhibit notable advancement on the structural and syntactic levels as the one detected on the lexical one. On the contrary, it has deteriorated and thus needs much upgrading. Similarly, the ASR device performance has not exhibited any progression, even retrieving from a percentage of **(1.47%)** approximately of the total errors in March 2030 to **(2.89%)** approximately in December 2030.

7.4 Effective Renditions

However, in spite of the copious translation errors detected in March 2023 CCs data under study, there are examples of translational mastery, namely effective editions. For instance, some examples of forensic expressions are pertinently rendered with the exact terminology; like in translating She remains on death row في خطة premeditated plan, انتظار تنفيذ حكم الإعدام مع سبق الإصرار and a super maximum security penitentiary سجن شديد الحراسة. Instances of sematic and terminological accuracy can be detected in translating health department وزارة الصحة. As the lexical term department means ministry in the USA particularly. Similarly, the NMT device accurately identifies some terms as the acronym RCMP شرطة الخيالة الملكية الكندية which stand for (Royal Canadian Mounted Police) (Britannica, n.d.).

Besides, some instances of metaphorical ST verbal phrases are pragmatically translated into apt idiomatic TT equivalents like spread like wildfire كالنار في الهشيم, DNA testing was still in its infancy and didn't come up with any leads كان اختبار الحمض النووي لا يزال في مهده ولم يأت بأي خيوط. Such TT captions aptly convey the denotative and connotative meanings of the ST idioms into accurate and even metaphorical, TT equivalents.

7.5 AVT Norms

Concerning the semiotic readability aspects of the CCs in the three videos under study, they are considered moderately legible. When rather long captions appear in a certain screen frame, they are retained in the subsequent one to give viewers better opportunity of following the CCs sequence. In most cases, when a caption exceeds the limit of **37** characters with spaces, it is divided into two lines. The synchronization between the appearance of the translated CCs and the spoken narrated line is quite acceptable. Nevertheless, there are some deviations from the AVT norms, when some CCs appear ahead of the video sound. As captions are mainly recognized first by the ASR device then translated by the NMT software. The timing difference is a few seconds (not less than 0.25 of a second) compensated by the time needed for reading the caption.

The duration of the captions on the screen is linked with the utterance of the narrated lines, so some last on the screen for less than the AVT 6-second rule. For instance, the caption لو كان العالم قد انتهى بالنسبة لهؤلاء الأولاد ، فقد انتهى appears for only 4 seconds on the screen despite being **58** characters with spaces. The maximum characters per line especially for Semitic languages, as Arabic is up to **42** characters (Díaz Cintas and Remael, 2021, p. 98). Nonetheless, in some instances, the quality of produced translated linguistic content of the captions seem not to be that related to this AVT rule. For example, in Eubanks's video, **34** lines out of **465 (7.31%)** approximately appear in more than **42** characters, with one of the captions reaching **63** characters-length with spaces, rendering its readability rather difficult. In Metheny's video, **36** lines out of **332 (10.8%)** approximately are long lines with one of the lines comprising **65** characters with spaces. In Legere's video **29** lines out of **344 (8.4%)** approximately exceed the character maximum limit, with one of them

66 characters, constituting a difficult viewing experience.

Auto-spotting and line segmentation of the translated CCs of the three crime stories are generally in conformity with line utterances. However, some mal segmentation errors are quite critical to the extent of marring the denotative and the connotative meanings of the captions, like: The two began fighting ***well no big deal couples fight*** ***بشكل*** بدأ القتال ***جيد لا يقاتل الأزواج صفقة كبيرة***. The absence of the segmentation pause between the ST clause the two began fighting بدأ القتال and the adverb ***well*** leads to the NMT software rendering them as one clause. The critically incorrect translation of the adverb ***well*** ***بشكل جيد*** instead of "حسنا" rather distorts the TT content. Moreover, the erroneous placement of the negation article ***no*** which is a part of the idiomatic expression "no big deal" before the verbal phrase ***يقاتل الأزواج*** displays flawed semantic and syntactic manifestations. The specious meaning is aggravated by the literal translation of the ST idiom big deal ***صفقة كبيرة*** instead of " مشكلة كبيرة". This conveys a muddled rendition of a sentence that would be simply translated into " بدأ الشجار، حسناً ليست بالمشكلة الكبيرة فالأزواج عادة ما يتعاركون". The placement of a full stop separating the two nouns ***الهيروين. الادمان*** , in the CC while his mother was rarely around due to ***her heroin addiction*** بينما كانت والدته نادراً ما كانت موجودة بسبب ***الهيروين. الادمان*** and the addition of the identifiers ***ال*** to the noun ***إدمان*** , preceding the noun ***الهيروين*** produce a distorted sentence structure displaying denotative and connotative failures. Similarly, the misplacement of the full stop in the middle of the TT CC those living in the nearby community were lulled into a false sense of security. أولئك الذين يعيشون في المجتمع القريب ***هدأوا*** إلى شعور زائف الأمن displays a blunt segmentation error, that leads to difficult readability of the CC.

The three videos display overall coherence between the original image/sound and the CCs. However, there are some deviations from such unity like:

Figure 5: Image/sound coherence errors example 1

the ST CC How did a day that started with ***cheers for the Chargers*** end up with the ruthless murder of ***four innocent Souls foreign*** translated into كيف انتهى اليوم الذى بدأ ***بهتافات الشاحن*** بمقتل أربعة ***من الأبرياء من الأجانب***. The NMT software fails to achieve consistency between the original image and the CC. There is a failure on the visual level, as the video exhibits two glasses of drink being coerced to make a toast as a gesture of celebration not rallying. This is affirmed by the multimodal frame in which the scene participants are actually clinking their beer glasses. This is indicated by the ST lexical item ***cheers***, which is faultily translated into ***هتافات*** instead of "قرع الكؤوس". Besides, the ST lexical item ***Chargers***, which alludes to the Los Angeles American professional football team playing in the National Football League (Augustyn, n.d.), is literally translated into ***الشاحن*** instead of "فريق تشارجرز لكرة القدم الأمريكية".

Accordingly, the whole ST phrase denotation and connotation meanings are erroneously translated conveying misleading information, totally inconsistent with the semiotic level on the screen. On the audio level, there is a misrecognition of an extra wavelength by the ASR resulting in the appearance of the modifying adjective ***foreign***, which is not present in the ST. Therefore, it erroneously substitutes the noun ***أرواح*** which is already omitted and translated as a part of the previous sentence preceded by a preposition ***من الأجانب***. This rendition conveys deceitful TT contextual meaning, falsely implying that the four boys are foreigners.

Figure 6: Image/sound coherence errors example 2

In this example, the TT phrase not exactly ***the Brady bunch*** is translated into ليس بالضبط ***مجموعة البطلين***. The NMT fails to recognize the ST expression ***the Brady bunch*** as an allusion of a 1969 family sit.com, showing a group of people, enjoying happy family life (*The Brady Bunch*, n.d.). It is thus untenably translated into ***مجموعة البطلين***, exhibiting stark inconsistency with the visual semiotic level of a happy family riding in a car, which can be translated into "نموذج العائلة السعيدة" for instance.

Figure 7: Image/sound coherence errors example 3

Inconsistency between lexemes and images can be detected in the CC example Legere entered the church and crept around in the dark looking for anything valuable to steal discovering ***a safe*** *ودخل إلى الكنيسة وتسلل في الظلام بحثاً عن أي شيء ثمين لسرقته ليكتشف* ***مكاناً آمناً***. In this criminal context, the erroneous translation of the ST noun ***a safe***, which means "خزنة", into the modified noun ***مكانا آمناً*** starkly contradicts the visual frame exhibiting an image of an actual safe case.

8. Conclusion

The previous analysis for the YouTube closed captions within the fabric of the three crime narratives under study proves that there is notable insufficiency and inconsistency in the performance of the YouTube AI software device. Its linguistic content needs consistent feeding of sufficient software algorithms comprising vast parallel corpora of STs and their reasonably accurate equivalent TTs. Unfailing human intervention of constant editing and post editing is also rather imperative.

More specifically, context-oriented data can be installed in the translation memory of the YouTube AI configuration. This register-based terminological data can improve the quality of the lexical choices in certain genres of audiovisual material. For instance, medical, political, geographical, scientific, theological or forensic terminological lexemes can aid the AI device in opting for more accurate genre-oriented lexemes.

The very poor translation content accuracy rates of March 2023 data exhibited in the present study prove that most of the lexical changes undergone to the ST CCs can be considered more deviations rather than strategic manipulations on the part of the AI NMT device. Furthermore, with major form correction errors constituting approximately **(37.23%)** of the translation form errors, it can be deduced that the YouTube AI NMT device suffers a serious syntactic and structural problems. The distorted functional structure adversely hinders the operational flow of the semantic denotations and pragmatic connotations of any translated data. The rather improved translation accuracy rates of December 2023 revisited data prove the deficiency of the linguistic translational performance of the YouTube NMT device, which still needs prodigious steps of upgrading. Similarly, the ASR device still exhibits mal recognition of single phonemes and misrecognition of whole utterances and thus needs prominent enhancements.

Generally speaking, the YouTube auto-generated translated CCs under study meet the minimum technical requirement of AVT norms. As they are almost synchronized with the acoustic narrative content and segmented in a rather acceptable manner, except for some deviations. However, the previously shown technical errors may render the readability experience of such CCs rather difficult in some parts, proving that the YouTube segmenter device still needs further advancement.

As for the validity of Romero-Fresco and Pöchhacker (2017)'s NTR model as a model assessing the translation quality of the auto-generated YouTube CCs, it proves to be more functionally valid on the micro lexical individual level than the macro holistic clausal one. In spite of the fact that, the approximately **(95%)** accuracy rate calculated according to the suggested NTR equation may seem at first glance not a very low percentage, yet the notable pragmatic and semantic failures exhibited in the auto-generated YouTube CCs dim the fluency of a smooth, acceptable AVT viewing experience.

In what concerns fluency, naturalness and acceptability, the model needs extra detailed parameters for such aspects. The fringe error rate allocated to such important facets in the (form style) criteria is not indicative nor sufficient. The errors affecting aspects of pragmatic appropriateness and cultural nuances can have higher grade and more prominent representation in the suggested NTR formula. Some instances may not be penalized on the lexical level according to the proposed NTR error scale equation, yet they may have a grave impact on the connotative relevance of the overall acceptability of the produced translation quality.

Additionally, the model needs more specification of other AVT norms that directly connect the linguistic content with the audiovisual semiotic one. Subtitling screen norms of speed, synchronization, overall flow and coherence between the original image/sound and the subtitles are rather holistically presented in the model as additional comments with no accurate specification in the suggested equation.

Eventually, it can be inferred from the analysis conducted on the YouTube auto-generated CCs under study in the present paper that, although there are some instances of AI effective editions of positive translational renditions, such device still needs seamless measures of updating procedures to upgrade the performance of its NMT and ASR operative systems. Incessant human intervention exemplified in relentless feeding of parallel corpora for the AI translation memories and imperious post editing for the produced data is rather imperative. This is evident from the improvement of the assessed quality of the same interlingual data detected over time, signifying human intervention on the linguistic, semiotic and technical levels. Without such needed human interventions, the somehow suboptimal current performance of the AI YouTube NMT and ASR devices cannot be considered fully reliable. As improving the quality of the auto-generated interlingual YouTube CCs may help promote cross-cultural mediation and demolish communication barriers through this social media platform.

References

Al-Obaidli, F., Cox, S., Nakov, P. (2018). Bi-text Alignment of Movie Subtitles for Spoken English-Arabic Statistical Machine Translation. In: Gelbukh, A. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2016. Lecture Notes in Computer Science, 9624. Springer, Cham. https://doi.org/10.1007/978-3-319-75487-1_11

Augustyn, A. (n.d.). Los Angeles Chargers American football team. In *Britannica*. Www.britannica.com. https://www.britannica.com/topic/Los-Angeles-Chargers

Britannica. (n.d.). Royal Canadian Mounted Police, In Britannica. Retrieved October 18, 2023, from https://www.britannica.com/topic/Royal-Canadian-Mounted-Police

Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture*, (2nd ed.). Polity Press.

Carrier, M. (2017). Automated Speech Recognition in language learning: Potential models, benefits and impact. *Training Language and Culture*, *1*(1), 46–61. https://doi.org/10.29366/2017tlc.1.1.3

Chaume, F. (2018). Is audiovisual translation putting the concept of translation up against the ropes? *The Journal of Specialised Translation*, 8430: 84-104. https://www.jostrans.org/issue30/art_chaume.pdf

Ciobanu, D., & Secară A. (2020). Speech recognition and synthesis technologies in the translation workflow. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology*, 91–106. Routledge.

Collins COBUILD. (n.d.). Add insult to injury. In Collins COBUILD. Retrieved October 18, 2023, from https://www.collinsdictionary.com/dictionary/english/add-insult-to-injury

Collins COBUILD. (n.d.). Go off the rails. In Collins COBUILD. Retrieved October 18, 2023, from https://www.collinsdictionary.com/dictionary/english/go-off-the rails_1#:~:text=If%20someone%20goes%20off%20the,drugs%20or%20breaking%20the%20law.

Collins COBUILD. (n.d.). Take your eye off the ball. In Collins COBUILD. Retrieved October 18, 2023, from *https://www.collinsdictionary.com/dictionary/english/take-your-eye-off-the-ball*

Collins COBUILD. (n.d.). Wipe the slate clean. In Collins COBUILD. Retrieved October 18, 2023, from https://www.collinsdictionary.com/dictionary/english/wipe-the-slate-clean

Díaz Cintas, J., & Massidda, S. (2020). Technological advances in audiovisual translation. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology*, 255-270. Routledge.

Díaz Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and Practices*. Routledge.

Filipović, L., & Gascón, A. H. (2018). Interpreting meaning in police interviews: Applied Language Typology in a Forensic Linguistics context. *Vigo International Journal of Applied Linguistics*, *15*, 67–104. https://doi.org/10.35869/vial.v0i15.87

Hirvonen, M., & Kinnunen, T. (2021). Accessibility and linguistic rights. In K. Koskinen & N. K. Pokorn (Eds.), *The Routledge Handbook of Translation and Ethics*. 470–483. Routledge.

Hobbs, S., & Hoffman, M. (2022). "True Crime and…": The Hybridisation of True Crime Narratives on YouTube. *Crime Fiction Studies*, *3*(1), 26–41. https://doi.org/10.3366/cfs.2022.0058

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

Karakanta, A., Negri, M., & Turchi, M. (2020). Is 42 the Answer to Everything in Subtitling-oriented Speech Translation? *Proceedings of the 17th International Conference on Spoken Language Translation (IWSLT):* 209–219. DOI: 10.18653/v1/2020.iwslt-1.26

Kraeva, S., & Krasnopeyeva, E. (2020). Judging Translation On Social Media: A Pragmatic Look at Youtube Comment Section. *The European Proceedings of Social and Behavioural Sciences EpSBS*, 777–785. https://doi.org/10.15405/epsbs.2020.08.91

Lee, J.-H., & Cha, K.-W. (2020). An Analysis of the Errors in the Auto-Generated Captions of University Commencement Speeches on YouTube. *The Journal of Asia TEFL*, *17*(1), 143–159. https://doi.org/10.18823/asiatefl.2020.17.1.9.143

Lipping, S., Drossos, K., & Tuomas Virtanen. (2019). Crowdsourcing a Dataset of Audio Captions. *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019),* 139–143. https://doi.org/10.33682/sezz-vd31

Malaczkov, S. (2020). Subtitle revision in translator training: a case study of revisional modifications in TED translation crowdsourcing. *Bridge: Trends and Traditions in Translation and Interpreting Studies*, *1*(1), 3–22. https://www.bridge.ff.ukf.sk/index.php/bridge/issue/view/1

Peters, F. (2020). True Crime Narratives. *Crime Fiction Studies, 1*(1), 23–40. https://doi.org/10.3366/cfs.2020.0005

Romero-Fresco, P. (2020). Negotiating quality assessment in media accessibility: the case of live subtitling. *Universal Access in the Information Society*. 220:741–751. https://doi.org/10.1007/s10209-020-00735-6

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*, 28–50. Palgrave Macmillan.

Romero-Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, *16*. 149–167. https://doi.org/10.52034/lanstts.v16i0.438

Ruiz-Arroyo, A., Garcia-Crespo, A., Fuenmayor-Gonzalez, F., & Rodriguez-Goncalves, R. (2022). Comparative analysis between a respeaking captioning system and a captioning system without human intervention. *Universal Access in the Information Society*. 1-12. https://doi.org/10.1007/s10209-022-00926-3

Song, H.-J., Kim, H.-K., Kim, J.-D., Park, C.-Y., & Kim, Y.-S. (2019). Inter-Sentence Segmentation of YouTube Subtitles Using Long-Short Term Memory (LSTM). *Applied Sciences*, 9(7), 1504. https://doi.org/10.3390/app9071504

Stüker, S., Paulik, M., Muntsin Kolss, Fügen, C., & Waibel, A. (2007). *Speech Translation Enhanced ASR for European Parliament Speeches - On the Influence of ASR*

Performance on Speech Translation. IEEE International Conference 4, 1-4. https://doi.org/10.1109/icassp.2007.367314

TheFreeDictionary.com. (n.d.). The Brady Bunch. In TheFreeDictionary.com. Retrieved October 18, 2023, from https://encyclopedia.thefreedictionary.com/the+Brady+bunch

Traub, J. (n.d.). Modus operandi. In *Merriam-Webster*. https://www.merriam-webster.com/dictionary/modus%20operandi

Tuominen, T., Koponen, M., Vitikainen, K., & Tiedemann, J. (2023). Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution. *The Journal of Specialised Translation*, *39*, 77–98. https://jostrans.org/issue39/art_tuominen.pdf

Twisted Minds. (n.d.). *Escaped Prisoner Turned Psychotic Serial Killer.* [Video]. YouTube. https://www.youtube.com/watch?v=BKpNK7lK4BA&t=1s

Twisted Minds. (n.d.). *He Sold Human Burgers... | The Case of Joe Metheny*. [Video]. YouTube. https://www.youtube.com/watch?v=kwHcQ5WMFVA&t=36s

Twisted Minds. (n.d.). *The Mother Sentenced To DEATH.* [Video]. YouTube. https://www.youtube.com/watch?v=BKpNK7lK4BA&t=1s

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