

Online Automatic Subtitling Platforms and Machine Translation. An Analysis of Quality in AVT

Cristina VARGA*

Abstract: The current paper aims to analyse the main features and the limitations of the online automatic subtitling platforms. Based on different applications such as: video-to-text transcription programs, machine translation programs, and text segmenters, automatic subtitling involves a complex workflow and is meant to enhance the productivity of the professional subtitler. There are very few studies about the online subtitling platforms, therefore, the analysis we carried on will provide comprehensive empirical data and will contribute to a better knowledge of these innovative systems.

Keywords: automatic transcription, machine translation, automatic subtitling, video-to-text conversion, subtitling online platforms

1. Introduction

Machine translation (MT) systems are used on a daily base by worldwide users as they provide rapid linguistic solutions and allow the users to avoid linguistic barriers. MT systems are also used by professionals and companies in the field of specialised translation in order to improve their efficiency. The current paper investigates whether such a successful and popular system like machine translation can be used satisfactorily in automatic translation of movie subtitles.

During the last decade, subtitling changed drastically from the traditional workflow to a different framework involving CAT tools. Considered as a further step from CAT tools (Diaz Cintas&Remael, 2021; Volk, 2008, 2010), machine translation

* Senior Lecturer, Department of Applied Modern Languages, Faculty of Letters, Romania. Babeș-Bolyai University, Cluj-Napoca, Senior Lecturer, Department of Professional and Specialized Translation, Catholic Faculty of Humanities, University of the West, Angers, France. E-mail: cristina.varga@gmail.com

was implemented recently in subtitling aiming to enhance productivity. The first studies on this topic Lavecchia et al. (2007), Volk&Harder (2007), Tiedemann (2007) focus the use of parallel corpora in order to training and calibrating the MT for subtitling. Different European projects such as *MUSA (2002-2004)*, *eTITLE (2004-2006)*, *SUMAT (2011-2014)*, Georgakopoulou & Bywood, (2014), *TransLectures (2011-2014)* were also meant to explore the development automatic subtitling. The outcome of these highly ambitious European programs consisted of different automatic subtitling applications using rule-base approaches or statistical translation methods. Nowadays most of their webpages are inactive and none of them are considered functional interfaces.

From the part of the industry, automatic subtitles were implemented by *Google* in *YouTube*, in 2006 and 2009, as announced on Google official blog (Harrenstien, 2006; 2009). This system was enhanced in 2013 with a voice recognition system based on Deep Neural Network (Hank Liao et al., 2013).

Nowadays, with the cloud technology and the virtualisation of specific applications there is a great variety of online platforms for *automatic subtitling* accessible for the Internet users. They all offer the same service namely *automatic subtitling* and their functioning and their results were not evaluated.

The aim of the current paper is to analyse and observe the output of the online automatic subtitling platforms, to point out their advantages and their limitations. This paper is intended to provide an overview of the state of the art of online automatic subtitle platforms considered as a new development stage in the field.

2. Work Methodology

The work methodology of the current research focused on the basic structure of the automatic subtitle systems which are basically formed by three applications namely: *a video-to-text transcriber* based on a speech recognition system; a machine translation software, and a *subtitle generator*, basically a text segmenter application which converts the transcribed text into subtitles and synchronises them with the movie. Each component of the online subtitling platforms should be analysed in order to observe the accuracy of each task and the correctness of the output.

A list of potential online subtitling platforms has been identified. We focused on the following application: *Temi*, *TransLectures*, *360converter*, *Veed.io*, *HappyScribe*, *Amberscript*, *Sonix.ai*, *Trint*, *Descript*. The list is not exhaustive.

The *machine translation* was tested for a less known and spoken language (Romanian), in order to observe the real performances of the MT systems. Since the online platforms use also the public MT systems, a list of the most popular online MT applications has been created and used in automatic subtitling translation.

Last but not least, the *segmenter* of each subtitling platform mentioned above was tested in order to assess the quality of the subtitles generated by the different platforms. The outcome was contrasted with the standard requirements of the professional standards in subtitling.

All these online platforms were tested with the same video file and their outcome was analysed from a quantitative and qualitative point of view in order to highlight the prevalent features of each platform. In order to fully assess their capabilities, the videoclip we selected was the opening scene of Quentin Tarantino's movie *Reservoir Dogs*. It is a complex fragment with a duration of 9.13 minutes, consisting of seven characters speaking, in a noisy restaurant setup, using slang, interrupting each other, and referring in their dialogues to culture-specific elements.

And it is intended to measure the capabilities of the automatic subtitling platforms. It is also important to clarify that this is not a comparative study, therefore, the results of our analysis will not point out to an individual online platform. We will intend to draw a general panorama of some of the most important features and limitations of the milieu.

3. Video-to-Text. Automatic Transcription

The 1st phase of the automatic subtitling consists of a video-to-text transcription. The text produced by the online platform represents the base of the entire automatic subtitling process and the final product.

From a quantitative point of view, the results of the analysis are the following. In terms of *functionality* the online platforms responded in very different manners to the user's requirements. Since all the platforms were fully-functional it was surprising to observe that their responsiveness was different in terms of *user-software interaction*. One of the platforms, turn out to be completely irresponsive. The platform uploaded the video file without providing any automatic transcription. Other platforms were moderately responsive. Approximately 30% of the platforms didn't work at the first attempt. In these cases, the video files had to be uploaded repeatedly and the automatic transcript had to be requested recurrently. Only 60% proved to be fully responsive and performed without error messages, technical problems nor delays in carrying out the task requested.

In terms of *completeness* of the task, two platforms worked incompletely reporting they performed the entire transcription. Actually, only 10% and respectively 61% of the text was transcribed as observed during the analysis of transcriptions. Nevertheless, the majority of the platforms (80%) completed properly the transcription of the video.

It could be also important to observe that even if the audio-visual text was the same for all the platforms, they needed different lengths of time for achieving the same transcription. The majority of the platforms (70%) achieved in few minutes the automatic transcription of the video material consisting of 1312 words (approx. 2,8 pages). The rest of 30% of the platforms needed more time for the same task. The maximum length of time required to produce a complete transcription was 30 minutes.

Concerning the *accuracy* of the transcription, the length of text produced by the platforms varies considerably. The shortest transcription consists of 836 words

and it represents only a part of the audio-visual text. The other platforms produced transcription consisting of different text length comprised in a range between 1242 and 1382 words. These differences can be explained by different behaviour and capabilities of their inner voice recognition systems. Thus, their performance is uneven and the resulting text is far from being exact and coherent. It is also important to mention that none of the online transcription tools analysed during our research is error-free.

In this context, analysing the *accuracy* in automatic video-to-text transcription is basically a matter of determining the amount of errors made by the online transcription tools while performing a requested task. The empirical data provided by our research show that the degree of *accuracy* determined in automatic transcription is comprised in a range between 4% and 42%, which means that, according to these statistic data, some of the automatic transcription tool provide decent results.

Data provided by the quantitative analysis illustrate partially the effectiveness of the automatic transcription tools. Qualitative analysis provides complementary data allowing us to enhance the analysis based on quantitative data and have a broader view of the situation. In order to provide comprehensive information about the quality of automatic transcription of audiovisual texts, we focussed our research on following categories of transcription errors: *text coherence*, *missing text*, *text layout*, *speaker recognition*, *spelling errors*, and *punctuation errors*. All mentioned categories will be illustrated with examples observed during the phase of testing and provided by the online tools quoted above.

The *text coherence* is one of the most important aspect of *movie-to-text transcription* since an accurate transcript guarantees a good quality of the translation and of the subtitling. The online transcription tools we analysed produced a coherent and understandable text with two exceptions where the produced text had no relation whatsoever with the source text and the resulted transcript was unusable for translation purposes.

In what follows we will comment different errors and limitations of the automatic transcription systems. *Incoherent transcript* is a serious error as it can be observed in the following example. It occurs when the film is too noisy and many characters are talking simultaneously:

Speaker 100:01:20 schools, ect. ECT ect ect. ECT ect. Ce dialogue discret jusqu'en searching yo yo. Benjo, toi toi obéit aux besoins Obi-Wan Kenobi Voir tomber Tchang dans la baie de Ha! Ha!

A different degree of incoherence in transcription can be observed in the following example. The text can be read, the phrases present structure, punctuation marks, and the produced text is similar but not identical with the source text. In this case, because of the lack of accuracy it can't be used for translation:

To a lack of urgency that it's all about a girl who takes a guy with a big person in the middle of a big night, [...] Whoa, whoa, whoa, whoa, whoa. Time out. [...] could like a bridge is not about some sensitive girl who needs a nice fella. That's what you're good about. No, no argument about that because True-Blue blue or true blue, I guess even farther up to the corruption.

The majority of the transcription tools we analysed, produced a text is generally coherent with inconsistent fragments, such as the following examples: *It's cool things to me. Give me my dollar back. Hey, [00:06:28] If the dollar is there, [...]*

Apparently, the text has a consistent structure and a clear meaning. In fact, a part of the source text is different: *"He's convinced me. Give me my dollar back. Hey. Leave the dollars there"*.

Sometimes the transcription is very similar to the source text, but contains different small errors altering the meaning:

All right, everybody cough up some *Greenford* the lady. Come on, throw in a *book*. *I don't you know*, I don't believe in it. *I believe* in tipping. You know, these chicks make *them make shit*.

These brief inconsistencies such as *Greenford*, correctly *"green for [the little lady]"*, *book* instead of *buck*, *"I don't you know"* for *"I don't tip. / You don't tip?"* can be corrected by a professional proofreader.

Sometimes brief fragments of transcription are also very difficult to understand especially dialogues involving more than one speaker as in the following example:

Put in throw in Mr. Pink with the pink eye not you don't tip you don't really

The source text of this transcription consists of 8 dialogue lines but the absence of correct segmentation, the lack of speaker identification, and the absence of punctuation marks make the text unintelligible.

- Who didn't throw in? / - Mr. Pink. / - Mr. Pink? / - Why not? / - He don't tip. / - He don't tip? / - You don't tip? / - He don't believe in it.

Another category of errors that can be observed in automatic transcription online is related to the *missing text*. Online transcription tools eventually mark in diverse manners the fact that different fragments of text are missing or are inaccurate. This function is not available on all online transcription tools, only a part of them use this marking system. This function is important for professional translators and proofreaders allowing them to identify and correct the gaps in an automatically transcribed text.

Different marks are used by the transcription tools in order to indicate the missing/unintelligible audiovisual text. Some of them add *subtitle time codes* mentioning the number of seconds of non-transcribed text:

Oh, Tommy's that little Chinese [00:01:00] girl. What's her last name? What's that? I [00:05:00] mean, when I order a coffee, I want it filled six times, six times.

Others add the mention *[inaudible]* directly in the text: This one I'm talking about *[inaudible]* during that part, I'll take care of the check.

And in some cases, online transcription tools change the colour of a missing or inaccurately transcribed text. These marks allow the proofreader to focus on the most vulnerable parts of the transcribed text.

The *layout* of the transcribed text is another important issue since the script is a textual genre that has a particular layout. Most of online transcription tools organise the text in paragraphs only and some of them transcribe everything as a block of text without any structure making the text more difficult to comprehend and translate.

The *identification of the speakers* is another important issue. Nowadays, this aspect seems to be a very difficult task and the technology has to be more developed in order to obtain better results. Even if some of the online transcription tools intend to identify automatically the speakers, the results are inconsistent. Since the source text we used in our research represents a movie scene involving seven different characters, most of them taking part actively in the conversation, the identification of each speaker is essential for the correct comprehension and further translation of the dialogues.

As an example, the following automatically transcribed text represents the dialogue between two interlocutors, the first part of the text being uttered by one character, and the second part of the text, in italics, representing the reply of the second speaker. Without a clear *identification of the speakers* in the transcription, it is impossible to observe that “now for the past 15 minutes” is uttered for another speaker:

What do you mean? We leave? Give me it back *now for the past 15 minutes*. Now, *you've been droning on about name's, Toby, Toby*, [...]

The most recurrent errors in automatic transcription are *spelling errors*. In general, they are caused by the *homophony*. Likewise, different disturbing elements in a movie such as *background noise*, *accents*, *cluttering*, and so on can be at the origin of these *spelling errors*. In the following examples, “*Wait for singers*” and “*Well, Christina*” are two different ways to transcribe the word *Waitressing* pronounced unclearly by the movie character. In the 3rd example, “*Good day*” is actually a misspelled transcription of “*I'd go*” also, a better transcription of “*12 percent*” would be *12%*.

Wait for singers. [...] The number one occupation for female non-college graduates in this country.

Well, Christina is the number one occupation for female non-college graduates in this country.

Good day over *12 percent* for that.

Another possible source of *spelling errors* are the *proper names*. In the source text are mentioned singers, actors, songs, and the title of a movie, which, depending on the online transcription tools, were correctly/partially identified. In the

following example we can see the name *Obiwan*, the name of a character from another movie, instead of the name “*Toby Wong*” mentioned in the source text.

[...] names. Obiwan. / Obiwan, Obiwan. / Toby Chung [...]

Proper names of singers and actors mentioned in the source text are, in general, correctly transcribed. It is the case of the names *Madonna* and *Charles Bronson*, recognised by most of the online transcription tools. The situation is different in the case of the title of the song “*True Blue*” or the title of the movie “*The Great Escape*”, both mentioned in the source text but inconsistently transcribed most of the time.

...*Charles Bronson* in the grand scheme.
...like *Charles Bronson*. *immigrant escape*.
...is like, *Charles. Bronson* in the *Great Escape*...
...like *Charles Bonds* and *integrated scape*.

Punctuation in automatic transcribed texts is mostly inconsistent and can generate errors in translation. The most frequent errors we observed during the current analysis are: *missing punctuation marks* and *mistaken punctuation marks*:

[...] this is the world’s smallest violin playing Just For The Waitresses [*missing period mark*] you don’t have any idea [*missing period mark*] what you’re talking about [*question mark*] These people bust their ass.
But no society says, [*missing quotation mark*] don’t tip these guys over here [*erroneous period mark*] But tip these guys over here [*missing quotation mark*] [*missing period mark*] that’s bullshit.
I don’t tip [*missing period mark*] you don’t tip [*question mark*]

The analysis of the online transcription tools allows us to state that nowadays the online platforms produce inconsistent automatic video-to-text transcription for complex videoclips. For a suitable transcription human professional assistance is highly recommended.

4. Machine Translation for subtitling purposes

The second stage of the automatic subtitling process is *machine translation*. Depending on the platform, *machine translation* functionality can be available as a paid service. Nonetheless, there are platforms that do not offer this service, in which case, it is possible to exploit one of the online machine translation applications. Actually, it is quite common for CAT tools and online translation platforms to connect to machine translation servers such as *Google Translate* or *Bing Microsoft Translator* but, not only. In our analysis of machine translation for subtitling we tested the following applications: *Google Translate*, *Systran*, *Bing Microsoft Translator*, *Yandex Translate*, *Deepl*, *Kantan*, *Wordlingo*, and *MyMemory*.

In order to have a complete collection of empirical data on the accuracy of machine translation tools that can be used in the process of automatic subtitling, we

automatically translated the script of the same videoclip we used during the previous stage. The script was elaborated by the author of the current research for this occasion from the same video file hosted on *Youtube* we used for transcription.

The translations produced by the MT online tools were analysed and the following inconsistencies were reported: *literal translation, word order, language register, noun-adjective agreement, punctuation* and *mistranslation*. Each category will be illustrated and commented in what follows.

One of the most common mistakes that machine translation does regularly is *literal translation*. All idioms and collocations in the text were literally translated such as: “*Voi puteți lua bacșișul.*” instead of “*Voi vă ocupați de bacșiș*”. Also the translation “*În regulă, toată lumea să tușească / niște verde pentru micuța doamnă.*” is inconsistent in Romanian. A correct translation would be: “*În regulă, toată lumea să scoată / niște verzișori pentru domnișoara.*”. Even if the phrase is correct apparently “*Așteptați un minut. / Cine nu a aruncat?*” doesn’t communicate the same message as the source text. In this case the correct translation is: “*Ia stați! Cine nu a pus?*”

Word order is also important, any error can affect the communication of meaning. MT in this example is particularly poor: “*Lasă-mă să-ți spun ce / „Like A Virgin” este despre*”.

Language register is an expression of the level of formality in a specific situation of communication. The inappropriate use of language register, the alternance of registers in the same communication is perceived as a lack of quality in translation. In the chosen examples, within the same phrase can be observed marks of *formal* and *familiar register*: “*Scuzați-mă, domnule Pink, / dar ultimul lucru de care ai nevoie este...*” or “*Nu le dați bacșiș la tipii ăștia de aici, / dar dă-le bacșiș la ăștia.*”

Since Romanian is a Latin language, the *noun-adjective agreement* is a complex problem, very difficult to manage for a machine or software. Therefore, very many errors in translation are caused by the impossibility to determine precisely the correct grammatical form of the *noun-adjective agreement* such as: “*„Like A Virgin” nu este despre unele / fată sensibilă care întâlnește un tip drăguț*”. *Noun-adjective agreement* is also quite difficult to be properly used in machine translation if the adjective refers to a visual element on the screen. For example, the reply: “*Nu am fost destul de norocoasă să am o slujbă*” is apparently correct excepting the fact that the adjective has a feminine form and in the movie the character speaking about himself is a man.

Punctuation is a matter of correctness; therefore, it is important to adapt the quotation marks to the Romanian instead of writing them as in the source text: “*Like A Virgin*” *nu este despre un / o fată sensibilă care întâlnește un tip drăguț*.

Mistranslation is the most frequent error of MT tools. It causes negative impact on the reception of the translation and of the movie. There are different degrees of mistranslation. A light error could be, for example the use of an incorrect preposition: “*A fost un mare succes pentru Madonna.*” Here, in the context of a music hit is suited a different translation “*A fost un mare succes al Madonnei*”.

MT tools usually translate the names. It is a very common error largely commented in the literature. In the text, the name *Chew* is translated by the MT in the following example: “*Mestecă? Toby Chew?*”.

The use of loan translations (calques) in MT is another cause of translation errors. In this example, *argument* is a word with a different meaning in Romanian and therefore the translation is inconsistent: “*Nici un argument despre asta.*”.

The different types of errors illustrated above and their rate of recurrence observed during the current research allow us to state that MT is a technology that needs further development. We think that MT system may be better trained and calibrated for subtitling purposes in order to assist the subtitler not to replace it. Human involvement is essential in a good quality AVT.

5. Automatic Subtitling

As mentioned above, our research is focused on the analysis of 9 online subtitling platforms, of which only 5 offer *free automatic subtitles* of different quality. The analysis of empirical data gathered during the carried out research, allow us to assess some components of quality of automatic subtitling produced by the online platforms.

From a quantitative point of view, it is important to mention that different *automatic subtitles tools* produced different automatic subtitles. In the first place, the number of subtitles produced for the same videoclip is different, they are comprised in a range between 124 and 249 subtitles. Correspondingly, the amount of errors observed in automatic subtitles is different, starting from 24% and up to 61% of the generated subtitles.

Other quantitative empirical data gathered during the research concern *the reading speed, the maximum and minimum duration of a subtitle on the screen, and the number of characters per line*. The number of characters per line is excellently done and no error was observed in the automatic subtitles. The situation is different in the case of the *reading speed* where the amount of errors is significant. They start from 42 errors in a subtitle file and go up to 122 errors, with a rate raging between 16 CPS and 27 CPS. The management of *minimum and maximum duration of subtitles* is uneven on different online platforms. Some of them manage perfectly *the minimum* or *the maximum duration of subtitles* generating 0 errors. Unfortunately, no platform manages to realise correctly both *minimum and maximum durations*. Other platforms instead produce errors on both durations, starting from 1 and up to 65 erroneous subtitles. The *minimum duration* of an automatic subtitle is 0,15 seconds and the *maximum duration* of a subtitle in the same material is 9 seconds, both are values being considered as out of range.

From a qualitative point of view, several different issues can be identified in automatic subtitling. The main cause of these issues such as *missing text* or *erroneous subtitling* is the poor quality of automatic transcription. Other issues concerning the integrity and correctness of automatic subtitles are the following: *punctuation*,

segmentation of a subtitle, line length, spoken utterances and subtitled sentences, taboo words, and culture-specific elements.

Reduction as a strategy of AVT is used frequently in subtitling. If a *taboo word* or another word is used recurrently in the dialogues of a movie, in the subtitling the word is repeated, but for reasons of text economy its apparition on the screen will be reduced. An automatic subtitling tool, cannot evaluate correctly the linguistic context and the *interjections* or *the taboo words* in the movie are repeated without reduction in subtitling as in the following example: “*Whoa, whoa, whoa, whoa, whoa. Time out.*”

Other issues of automatic subtitling concern the *segmentation of a subtitle*. The subtitler creates a two-lines subtitle when the uttered text is long and a one-line subtitle is not enough to render all the spoken text. An automatic subtitling tool cannot weigh correctly the linguistic context; therefore, sometimes a *regular one-line subtitle* uttered by the same movie character is divided with no reason in two different subtitles. In our example, each subtitle has a duration of less than one second: “*It's my book. / Now*”. In this case we can highlight the presence of multiples errors in the same subtitle concerning *punctuation, segmentation, and duration*.

Another automatic subtitling problem that can be highlighted because is recurrent is the *line length*. In subtitling, the literature of the field states that in a two-line subtitle the lines should be equal or proportional in length. This rule is generally followed by the automatic subtitling tool but not always. Sometimes the film spectator can observe a full subtitle on the 1st line and only one very short word on the second line, such in the following examples:

That's [...] to the tourists. / Hold.
I asked is, how's it going in the world? / But

A different segmentation error can be observed in the example: *I mean, when I order a coffee, I / want it filled six times, six times*. Here we have the dialogue line of one character which is incorrectly segmented at the end of the 1st line. As it can be observed, the segmentation splits the subject and the verb on two different lines. Another segmentation problem can be observed in the case of the text “*six times*”. This is not a mere repetition but the beginning of the dialogue line of another character. This kind of segmentation error can make the spectator to misunderstand the text on the screen.

Culture-specific elements can also be a problem in automatic subtitling. When dealing with culture-specific element, the subtitler can use many different strategies for the linguistic transfer such as: *cultural transfer, transposition, explanation, neutralisation, and omission*. In automatic subtitling *transposition* is the only strategy applied in translation and sometimes the cultural element is mistaken because of the homophony such in the following example: *being the world's biggest McDonald's / personally I can do without, or I*

In the subtitle above is mentioned the brand name *McDonald's*. In this case, *transposition* is the correct strategy to use since it is a worldwide famous brand name.

The error consists of the fact that the *culture-specific element* is mistaken. In the movie the characters around the table are talking about *Madonna* and her songs and *McDonald's* is only a misinterpretation of the online platform.

As it can be observed, the automatic subtitling is an innovative technology involving complex processes and a sequential workflow. Nevertheless, the quality of the results is far from being satisfactory from a professional point of view.

6. Conclusions

The automatic subtitling represents an innovative and technological translation activity that shapes new ways of working for the professionals in the field. It presents a complex workflow involving many different online tools designed to perform specific tasks such as: *automatic transcription*, *machine translation*, and *automatic subtitling*. Testing their performance is a method to measure the efficiency of those systems and to assess their limitations.

After testing the online automatic transcription tools, we can understand better how these applications function and what are their specific weaknesses. The different categories of errors highlighted by the empiric data such as: *text coherence*, *missing text*, *text layout*, *speaker recognition*; *spelling errors*, and *punctuation errors*, provide us with comprehensive information about the state of the art of automatic transcription technology. We think that our analysis demonstrated that these platforms are not autonomous and that they depend on a professional in order to generate satisfactory quality of the transcription.

Concerning the use of *machine translation* in subtitling, the conclusion of our research is that the MT online applications are not sufficiently trained nor properly calibrated. Postediting is highly recommended in order to obtain a good quality of the translation, especially in languages that are less known or used.

Concerning *automatic subtitling*, as it results the text segmenters are still struggling with spatial and temporal limitations specific to the subtitling. The many errors observed on the screen, namely the reading speed, the maximum and the minimum duration of a subtitle, the segmentation, and the line length to name just a few, allow us to conclude that this technology needs to be further developed and enhanced.

We hope that within our research and analysis we have demonstrated that the results obtained in audiovisual translation are not very satisfactory from a professional point of view. Maybe these new automatization technologies of AVT should be presented as an opportunity for the professionals and their clients to enhance quality and efficiency in the field and not as an economic alternative to the professional subtitler.

References

1. Burry, M., How to decipher a job advertisement, January 2020, retrieved from <https://www.thebalancecareers.com/how-to-decode-a-job-advertisement-2061002>
2. Georgakopoulou, Panayota, Bywood, Lindsay, 'MT in subtitling and the rising profile of the post-editor' in *Multilingual*, 2014, 25 (1), pp. 24-28.
3. H. Liao, E. McDermott and A. Senior, "Large scale deep neural network acoustic modeling with semi-supervised training data for YouTube video transcription," in *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2013, pp. 368-373.
4. Jorge Díaz Cintas, Aline Remael, *Subtitling: Concepts and Practices*, London: Routledge, 2021, p. 273.
5. Lavecchia, Caroline, Kamel Smali, David Langlois, 'Machine translation of movie subtitles', in *Proceedings of Translating and the Computer 29*, London. Aslib. 2007.
6. Tiedemann, Jörg, 'Improved sentence alignment for movie subtitles', in *Proceedings of RANLP*, Borovets: Bulgaria, 2007, pp. 576-582. Available at: <https://lml.bas.bg/ranlp2007/>.
7. Volk, Martin, 'The Automatic Translation of Film Subtitles. A machine Translation Success Story?' in Nivre., Joakim, Dahllöfand, Mats, Megyesi, Beáta (eds) *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein*, Uppsala: University of Uppsala, 2008, p. 202–214. URL: uu.diva-portal.org/smash/record.jsf?pid=diva2:131983.
8. Volk, Martin, Sennrich, Rico, Hardmeier, Christian and Tidström, Frida, 'Machine Translation of TV Subtitles for Large-Scale Production', in *Proceedings of the Second Joint EM+/CNGL Workshop on Bringing MT to the User: Research on Integrating MT in the Translation Industry*, Denver, 2010, 53–62. Available at: web.me.com/emcnglworkshop/JEC2010_Proceedings.pdf. <https://otter.ai/login>.
9. Volk, Martin.; Harder, Søren, 'Evaluating MT with translations or translators. What is the difference?', in *Machine Translation Summit XI Proceedings*, Copenhagen: European Association of Machine Translation, 2007, pp. 499-506.
10. Youtube, Reservoir Dogs, by Quentin Tarantino (1992) - Opening scene <https://www.youtube.com/watch?v=hvFZUOn1TCs&t=4s> (last accessed on 17.05.20121).
11. Harrenstien, Ken, Finally, Caption Playback, *Official Google Video Blog*, 19 September 2006. <http://googlevideo.blogspot.com/2006/09/finally-caption-playback.html>.
12. Harrenstien, Ken (2009) Automatic Captions in YouTube, *Official Google Video Blog*, 19 November 2009. <http://googleblog.blogspot.com/2009/11/automatic-captions-in-youtube.html>.

Online automatic subtitling platforms

1. Amberscript Audio To Text - Fast and Automatic, <https://app.amberscript.com/register?lang=fr> (last accessed on 17.05.20121).
2. Descript, <https://www.descript.com/transcription> (last accessed on 17.05.20121).
3. FoxA Video 360 Converter, <https://www.360converter.com/conversion/> (last accessed on 17.05.20121).

4. Sonix, <https://sonix.ai/accounts/> (last accessed on 17.05.20121).
5. Temi, <https://www.temi.com/> (last accessed on 17.05.20121).
6. Transcription et Sous-titres - Happy Scribe, <https://www.happyscribe.com/> (last accessed on 17.05.20121).
7. *TransLectures*, <https://www.mllp.upv.es/transcription-and-translation-platform/> (last accessed on 17.05.20121).
8. Trint: Audio Transcription Software, <https://app.trint.com/> (last accessed on 17.05.20121).
9. VEED - Online Video Editor - Video Editing Made Simple, <https://www.veed.io/> (last accessed on 17.05.20121).