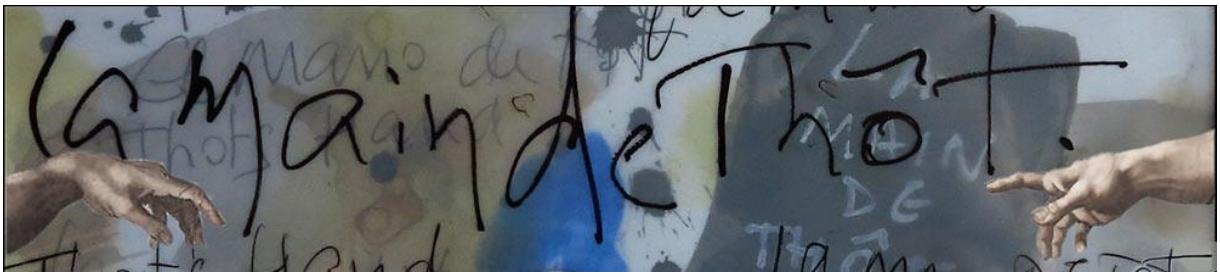


Is automatic subtitling a new technology that professional adapters can use?

Sabrina Baldo-de Brébisson
Université d'Évry Val d'Essonne
Maître de conférences
sabrina.baldo@univ-evry.fr

Many computer programs offer automatic subtitling services on the Internet. Such is the case of YouTube, which provides us with the opportunity to do automatic captioning for videos. The service can operate thanks to the coupling of two systems: Google Voice and Google Translate. So, the subtitles which are first generated by the technology of the vocal recognition system (Google Voice) are in a second phase automatically translated by the machine translation tool (Google Translate).

In our study, we will describe the use of such a service which, although it was created in 2009, is still not very well known to the general public or to professional adapters. We will seek to study the quality of vocal recognition and machine translation proposed by the two systems, thanks to a comparative analysis with the work of a professional adapter. Our purpose is to discover whether this technology is exploitable for a professional adapter or not.

Introduction

We decided to expand our research on Google's automatic subtitling system because it has the advantage of coupling two fields of application: speech recognition, through the Google Voice tool, and machine translations, through Google Translate.

First, we will present the automatic subtitling tool provided by Google on YouTube and its ease of use. Then, we will present the results of our speech recognition tests with Google Voice, and automatic subtitling tests with both Google Voice and Google Translate, and then with Google Translate alone. We will offer a brief comparative analysis between human translations done by a professional adapteri and the automatic translations provided by Google. The objective of this research is to establish whether Google's automatic translations of videos are a pure utopia or if they can be used by professional adapters.

1. The use of automatic subtitling on YouTube: a learning process that is stunningly simple?

Searches on the Internet have revealed the existence of many software applications that produce subtitles for videos, such as Jubler, Time Adjuster, Subtitle Workshop, Sublight, Aegisub, Kijio, Subtitle Translation Wizard, Handbrake and Any Video Converter. They provide many features such as the extraction of files from a video, the editing of these files, the viewing of a video's subtitles, the insertion of subtitles into a video, the synchronization of subtitles with images, the search for existing subtitle files on the internet and even machine translations.

Google had a great idea. They uploaded a tutorial (in English) to YouTube that describes the successive steps required to enable the automatic subtitling of a videoii. At the end of the two and a half minutes of explanation, the user is likely to be convinced of the extreme simplicity of the program: "All you need is a simple text transcript, no time-codes required, and Google will do the rest". The tutorial's host and specialist wants to be very reassuring: after a few commented clicks, he says, "your work is done!" with a rather attractive result, "Sometimes, the automatic captions are pretty good".

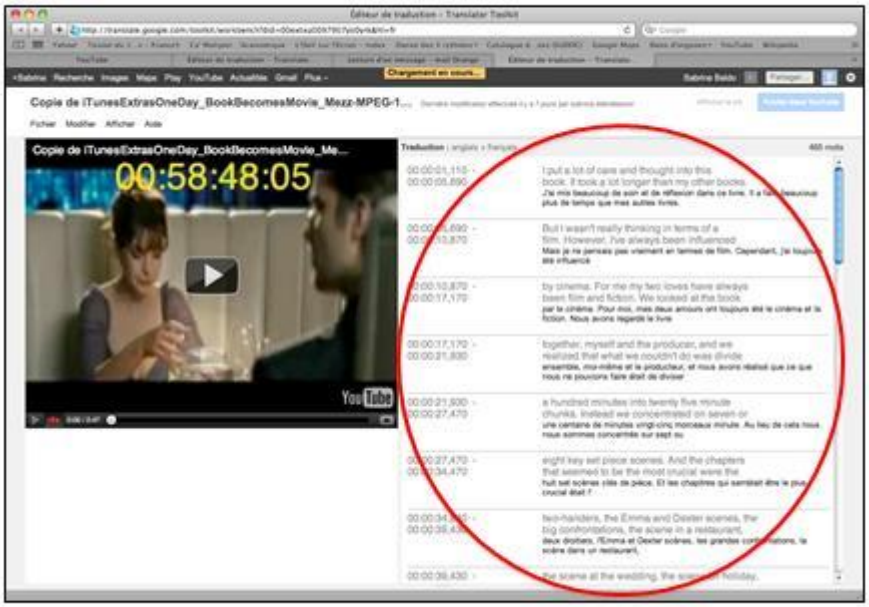
When we list the steps needed in order to obtain the automatic subtitling of an English video in French, only six steps – through a few clicks – are needed:

1. Click on the "Sign in" buttoniii on YouTube;
2. Click on "Add a video";

3. Go to the privacy Settings and click on "Public";
4. Go to Modify and click on "Subtitles";
5. Go to active Tracks and click on "Automatic English subtitles";
6. Go to Translate and click on "French".

The first five steps activate Google Voice and result in a "track", which refers to the text produced by the speech recognition in the same language as that of the video, in this case English. The sixth and last step is intended for Google Translate and aims to produce a second track: a machine translation in French of the first track. This last track features a time-coded version in two columns: on the left is the time-code and across from it, on the right, is the English text and its corresponding automatic translation, sentence by sentence. Here is a screenshot to illustrate this:

Fig.1



The tutorial describes only one automatic subtitling process even though our research has enabled us to discover other options. For example, users can intervene between the work of Google Voice and Google Translate in order to correct the first track generated by speech recognition, before it is submitted to machine translation. This capability is key in order to obtain even better results. However, the tutorial does not mention this.

Another option that we found within the system consists of downloading the file with the English written version (when the user has it) and submitting it to Google Translate directly. This alternative is a good one:

From a practical standpoint, as adapters who typically receive an English transcript of the video that they need to translate have no reason to use Google Voice;

From a qualitative standpoint, it is an option that shouldn't be overlooked as it can only improve the final quality, because – in this specific case – Google Translate works from an error-free track and not from an imperfect track produced by Google Voice.

Thus, the explanations outlined by the tutorial are far from complete because they only apply to a single type of situation, naturally the easiest and fastest one that can be explained: Google Voice submits a track that is directly translated by Google Translate. The tutorial apparently wants to be attractive and enticing to non-experienced users and, as such, it deliberately ignores the variants that the system provides and which are nevertheless key in terms of the final quality. Lastly, we can consider that the path outlined by the tutorial does more harm than good because it does not reflect reality. The explanations – which are simplistic and incomplete – do not enable the user to use the program the way that a more complete, more technical and more professional tutorial would. In fact, as we will explain, actual use of Google's guided tutorial proved to be quite a challenge.

2. Understanding the system or the path of a determined adapter

Before we got our initial results with Google Voice and Google Translate, we had to face a few unpredictable and complex obstacles. Here, in chronological order, are the details of the tricky situations that we experienced.

2.1 First obstacle: Google Voice freezing

We noticed that Google Voice was able to launch instantly with some videos, but with others it would freeze. After multiple attempts and much pondering, it occurred to us that all of the videoswhich caused the system to freeze, had something in common: they all began with a silent section lasting several seconds. In an empirical manner, we edited the problematic videos in order to reduce this initial silent section. When we resubmitted the problematic videos to Google Voice, it no longer froze up and it delivered a transcribed version of the videos that we were forced to shorten. The first difficulty had therefore been surmounted thanks to a comparative analysis of various media combined with the intervention of an IT specialist, whose idea it was to cut the video.

Even if a professional adapter is willing to devote a great deal of his/her time and energy, there is no guarantee that he/she would be able to overcome Google Voice's freezing problem. This first program malfunctionwhich we faced is very troublesome as it requires a certain level of computer skills (knowing how to edit a video) for the adapter to overcome it. However, if the adapter fails to un-freeze Google Voice, there is no way for him/her to proceed and he/she will be in a bind.

2.2. Second obstacle: Google Translate freezing

After successfully obtaining the first track produced by the speech recognition – with or without any required cutting – we proceeded to follow the video's instructions: "Go to Translate and click on 'French'". We then hit another brick wall: we were unable to activate the second track required to obtain an automatic translation. After trying everything, we finally stumbled upon a solution that would not be considered to be very professional: a modification of the first track through a random correction, such as the replacement of a small letter with a capital letter, or a punctuation mark with another one, or a word with a synonym. We concluded that any change, no matter how small, at least enabled us to generate a new voice recognition track, an essential condition for Google Translate to launch.

Once it is laid out, the solution we provide here may appear to be simple, at least in terms of its implementation, for a professional adapter without advanced IT skills. Nevertheless, we still had to try a multitude of more or less risky tricks before we were able to overcome this second computer bug. Surprisingly enough, the tutorial ignores this mandatory step: the creation of a new track from the first one in order for the second one to launch.

2.3 Third obstacle: Google Translate, the timeless one

After having successively overcome the first two bugs, we were finally confronted with a third difficulty that we were unable to overcome: Google Translate's time processing turned out to be extremely variable from one video to another. For example, sometimes the system submitted a track for certain videos in just a few seconds, which is a good performance. However, with other videos, we sometimes had to wait several hours, sometimes all night, before Google Translate decided to provide the automatically translated track.

Our tests with different videos did not allow us to elucidate this time variance problem, hindering us from fixing it. This last difficulty

remains a mystery and is above all a fatal flaw. One cannot imagine that a professional adapter would subject himself to Google Translate's highly random processing times, which range from a few seconds to ten hours, especially in light of the fact that adapters face short translation deadlines.

The use of both of these tools, Google Voice and Google Translate, provided by YouTube, therefore proved to be chaotic, unmanageable and therefore unusable by a professional adapter. As Google Translate's highly variable processing time is impossible to evaluate, we weren't able to carry out a comparative study of the quality-time ratio between:

The "all-human" factor: a professional adaptation based on a human transcription source followed by a target language simulationiv;

The "semi-automated" factor with Google Translate: the importing of the human transcription of the video in its source language, the synchronization and creation of automatic subtitles in the source language, followed by a human revision of the subtitles in the target language;

The "all-automated" factor with Google Voice and Google Translate: the automatic subtitling in the source language through voice recognition and the automatic translation of the subtitles, followed by a human revision of the subtitles in the target language.

We therefore had to limit our research to a qualitative study of speech recognition (automatic transcription), automatic translations using automatic transcriptions (voice recognition) and automatic translations using human transcriptions (source scripts). The media we used were provided by a professional adapterv: an English video file of "One day. A novel becomes a movie, by David Nicholls", a file with an English transcript of the video file and a file with the human subtitles simulated in French.

3. Automatic transcription quality (using speech recognition): an unusable track

The automatic transcription track obtained from the "One day. A novel becomes a movie" video file turned out to be a failure: only 227 words of 392 (57%) were identified and most of the transcribed passages were plagued by phonetic confusion that made them incomprehensible. For proof, just read the below wording and the transcript generated by Google Voice:

> We realized that what we couldn't do was divide a hundred minutes into twenty-five minute chunksvi.

We realize that will be couldn't it was defiant hundred minutes into twenty-five minute chocs.

Despite the evolution of speech recognition systems over the last thirty years and their rapid growth, much research needs to be done to improve their robustness. Speech recognition remains a complex multidisciplinary domain (involving cognitive science, neuroscience, computer science, mathematics, signage, phonetics and linguistics). The results can be highly variable from one video to another and they depend on many parameters, from the quality of speakers' enunciation to the sonic environment, not to mention the recording quality of the medium itself.

The video that we submitted to Google Voice does not appear to have any features that would prevent the system from proceeding with phonetic decoding:

-There is only one speaker, director David Nicholls;

-His delivery is clear;

-His speed of speech is average, sometimes slow;

-The lexicon used is simple and general;

-The sonic environment is correct (no noise is layered onto his words);

-The quality of the recording does not seem particularly bad.

However, the outcome of Google Voice's speech recognition is very disappointing and it is not operational, due to an error rate that represents 43% of the words.

4. The quality of a machine translation from an automatic transcription: pure madness!

Google Translate's error rate increases dramatically when non-corrected speech recognition is used. Obviously, this is because further difficulties are introduced. The end result was clear: we were unable to produce a single sensible sentence out of the 392 translated words...

Here is an example of the output of a track that was automatically translated without any human pre-intervention:

I put a lot of care and thought into this book. (Video)

Care and gordon's this book. (Google Voice)

Entretien et Gordon ce livre. (Google Translate)

It is absolutely essential that the first track generated by the voice recognition system be corrected in order to avoid the processing of an unusable track. Without a doubt, a machine translation system cannot

produce meaningful text from text which is senseless. As was mentioned earlier, speech recognition research faces many challenges and the same is true for machine translations. In the light of this, it would have made sense for YouTube's tutorial to mention that human correction of the first track is essential before it is submitted to the machine translation system.

5. The quality of a machine translation from a human transcript: sweet madness...

The processing time required by Google Translate for some videos sometimes proved to be disconcerting as the translations can be nearly instantaneous (a few seconds), which is not in itself a guarantee of quality. Here is one of the sections that was best translated by the system:

> I put a lot of care and thought into this book. It took a lot longer than my other books. (Video)
>
> J'ai mis beaucoup de soin et de réflexion dans ce livre. Il a fallu beaucoup plus de temps que mes autres livres. (Google Translate)
>
> J'ai consacré du temps à ce livre, plus qu'aux précédents. (Human adaptation)

We decided to bring up this section because it is highly significant. The proposed automatic translation is not fundamentally incorrect: although hardly idiomatic, it is free of grammar, vocabulary and spelling errors, false statements and – more importantly – nonsensical ones. Here, the machine translation can be considered to be relatively acceptable.

In absolute terms, the human translation J'ai consacré du temps à ce livre, plus qu'aux précédents could be deemed an unsatisfactory and below par translation, since the concepts of "care" and "thought" are literally omitted. We could even go as far as to say that the machine translation is better than the human translation here since it is more complete and more respectful of the original text. However, in the context of adaptation, the automatic translation would probably not be acceptable as a subtitle due to its length; it would be hard to read on the screen. The fundamental areas of divergence between translations and adaptations explain this sort of qualitative back-and-forth action, which depends on the context:

-In the context of written translations, machine translations seem to be more faithful to the original meaning and more acceptable than human translations;

-In the context of adaptations, machine translations would not be acceptable, unlike human translations.

Hence, although Google provides the automatic translation of subtitles, they fully ignored the fundamental differences that exist between a written translation that is not a subtitle and a subtitle translation. It is no coincidence that the term "adaptation" was chosen for this type of translation, which is subject to specific spatial and temporal constraints. A professional adapter therefore has to constantly stray away from the source message through specific techniques that enable him to produce subtitles that are stripped, general and sometimes disloyal to the original meaning (BALDO, 2009, 157-167). This explains why one of the main skills expected of an adapter is the ability to express an idea in a limited number of words – according to a maximum amount of time and space – to the detriment of some semantic information that is deliberately set aside. For this reason, the concepts of "care" and "thought" were deliberately excluded from the statement, at the risk of sacrificing meaning. This notion has been described by Henrik Gottlieb:

> "In subtitling, the speech act is always in focus, intentions and effects are more important than isolated lexical elements. This pragmatic dimension leaves the subtitler free to take certain linguistic liberties, bearing in mind that each subtitle must be phrased and cued as part of a larger polysemiotic whole aimed at unimpeded audience reception". (GOTTLIEB, 1998, 246).

An analysis of this statement highlights one flaw: by providing a service with a tool that was not designed for adaptation, Google is shooting itself in the foot. Assuming that Google Translate is able to provide an acceptable translation, it will not be admissible as a subtitle. Antonini places emphasis on the extremely condensed nature of subtitles: "the words contained in the original dialogues tend to be reduced by between 40 to 75 percent in order to give viewers the chance of reading the subtitles while watching the film at the same time" (ANTONINI, 2005, 213). Thus, even though the expected automatic translation result can be obtained, it is inappropriate for the application, which in this case is the adaptation.

Generally speaking, one must not forget that machine translations typically offer literal and linear translations. However, as we have explained, adapters take a different approach, as they play around with words and ideas. The semantic omissions that they make are not considered to be errors (although they would normally be considered the worst type of translation errors), but commendable choices. In short, we can say that there is something incongruous about wanting to use automatic translations in a field as specific as that of subtitling. This incongruity is probably due to a lack of knowledge regarding the

difference between translations and adaptations. As Tony Hartley highlights:

> "TMvii is not used in literary translation, nor is it common to incorporate it into the subtitling process, no doubt because of the relatively low incidence of repetitions within this genre and the context-bound nature of the equivalence between subtitles in different languages. So, dedicated subtitling tools provide no help for the core task of finding the right words". (HARTLEY, 2009, 120).

Conclusion

All this confirms that Google's automatic subtitling system is an unsuitable technology for professional adapters:

-Using it a genuine technological feat;

-The final quality is unusable (with or without Google Voice);

-Even when the machine translation is acceptable, it does not meet the sui generis requirements of adaptation.

This probably explains why Google is so unfamiliar and so rarely used by professional adapters: "Although prototypes exist for the automatic subtitling of films and video documentaries, to our knowledge no automatic subtitling system is widely used among industry professionals" (HATON, 2006, 293). So why does Google provide this overly ambitious system?

To understand all of this, let's go back to the beginning of YouTube: this online platform allows all of its visitors to store videos, and even create channels on it that are similar to TV stations, which are supplied with videos on a regular and monitored basis. As a result, hundreds of millions of videos are viewed on YouTube around the world for entertainment purposes (music videos, for example), and for information purposes, through video reports depicting the conflicts and wars in countries where the flow of information is restricted by the current regimes. Of course, YouTube is not entirely free of charge to its users, and the maintenance of this type of technology platform is extremely expensive. So, how does Google – which owns YouTube – make any money from this? Simply by selling advertising space at the beginning of its videos: accessing a video requires that you view a short fifteen-second ad first. These fifteen seconds that are sold to advertisers are quite expensive. In this context, the more accessible the videos are, the more likely they are to be viewed, and therefore the more the advertisements that are attached will be viewed. Each time a video is subtitled in a foreign language, it can be found in the results of a search made in this language, and advertising can be associated with it in this language...

The main goal of the automatic subtitling system that YouTube provides may be to increase the potential viewership of the original video as well as the accompanying commercials. Incidentally, as with all translations performed with Google, it can also enrich Google Translate's translation engine, which uses bilingual texts in order to improve its results. Thus, none of this is motivated by the desire to be useful to professional adapters who, consequently, are not really competing with this tool.