

Key challenges in using automatic dubbing to translate educational YouTube videos

Rocío Baños

University College London

r.banos@ucl.ac.uk

<https://orcid.org/0000-0002-3026-6737>

Abstract

In an attempt to cope with the higher demand for dubbing experienced during the past few years, media companies have invested in Artificial Intelligence, with significant developments having been spearheaded by Google (Kottahachchi & Abeysinghe, 2022), Amazon (Federico et al., 2020; Lakew et al., 2021) and AppTek (Di Gangi et al., 2022). These organizations have seen the potential in automatic dubbing, understood as the combination of automatic speech recognition, machine translation and text-to-speech technologies to automatically replace the audio track of an original audiovisual text with synthetic speech in a different language, taking into consideration relevant synchronies. Given the latest developments in this field, this article sets out to provide an overview of automatic dubbing in the current mediascape and to identify some of the main challenges in its implementation, especially regarding the integration of MT and speech synthesis in the dubbing workflow. This is achieved by analysing the educational videos posted in the Spanish version of the YouTube Channel *Amoeba Sisters*¹ (Amoeba Sisters, n.d.), titled *Amoeba Sisters en Español* (n.d.), which have been dubbed using the tool Aloud². The evaluation focuses on the aspects users highlighted in their comments on this YouTube channel, which include a lack of naturalness and accuracy. Special attention is paid to the use of synthetic voices, which are heavily criticized by users. However, in line with the original intention of the Aloud tool developers of increasing accessibility, users also highlight the usefulness of automatic dubbing to students who are not fluent in English but who are interested in biology-related topics.

Keywords: automatic dubbing, machine translation, speech synthesis, dubbing, educational videos

¹ See youtube.com/@AmoebaSisters and youtube.com/@AmoebaSistersEspanol

² See aloud.area120.google.com

1. Introduction

Towards the end of the 2010s, dubbing seems to have entered its golden age, with a renewed interest in this audiovisual translation (AVT) mode both in academia and the media localization industry. In academia, scholars have recently investigated under-explored topics such as the role of dubbing in videogame localization (Mejías-Climent, 2021) and the dubbing of entertainment content into English (Hayes, 2021; Spiteri Miggiani, 2021). In addition, more recent research has increased our understanding of dubbing standards and workflows (Spiteri Miggiani, 2022) and of the way dubbed content is received and perceived by audiences (Ameri & Khoshsaligheh, 2018; Di Giovanni & Romero-Fresco, 2019; Pavesi & Zamora, 2021), among many other topics. The industry seems to be particularly concerned with how to cope with the higher demand for dubbing experienced during the past few years. To this end, media companies have invested effort both in Artificial Intelligence (AI) and in collaborations with academia to offer training coursesⁱ that help to solve the alleged talent crunch witnessed in the dubbing industry. Regarding the former, one of the technologies that has experienced significant development is automatic dubbing, which is the focus of this article.

The automation of revoicingⁱⁱ modes has received less attention in academia compared to other AVT modes such as subtitling, with only limited research and projects investigating the potential of technologies such as speech recognition, speech synthesis and machine translation (MT) (BBC Media Centre, 2015; Matamala, 2015). These projects have nevertheless paved the way for the latest developments we are witnessing today in automatic dubbing, spearheaded by companies such as Google (Kottahachchi & Abeysinghe, 2022), Amazon (Federico et al., 2020; Lakew et al., 2021) and AppTek (Di Gangi et al., 2022), who have seen great potential in this field. Against this backdrop, this article sets out (1) to provide an overview of automatic dubbing in the current mediascape and to contextualize this practice accordingly; and (2) to identify some of the key challenges in its implementation, especially as regards the integration of MT and speech synthesis in the dubbing workflow. To this end, in the study reported on in this article, the performance of the tool Aloud³ has been evaluated by analysing some of the videos currently available in the Spanish version of the YouTube Channel *Amoeba Sisters*⁴ (Amoeba Sisters, n.d.), titled *Amoeba Sisters en Español* (n.d.), which have been dubbed using this tool. The analysis is purely qualitative and exploratory, engaging as it does with the aspects highlighted by users in their comments on this YouTube channel. The aim is therefore not to provide an in-depth analysis or quality evaluation of the MT or speech-synthesis output, but rather to initiate a discussion about the key challenges and potential of these first implementations of automatic dubbing from the perspective of studies in AVT.

³ See aloud.area120.google.com

⁴ See youtube.com/@AmoebaSisters and youtube.com/@AmoebaSistersEspanol

2. Technology and dubbing

Technology has profoundly influenced not only the way in which audiovisual content is localized and distributed, but also the way audiences interact with such content. Whereas the role of technology in subtitling has been undeniable, it has evolved differently in the case of dubbing. Or, rather, it has had dissimilar effects on the human agents involved in the dubbing process. When comparing the use of technology in the subtitling and the dubbing industry, Bywood (2020) states that “the dubbing context demonstrates a substantially delayed adoption of technology for the translator” (p. 511). Many translators and dialogue writers working in the field of dubbing are still not required to use specialized software (unlike subtitlers); instead, they are asked to work mainly with text editors to produce a document which will then be managed accordingly by the dubbing studio. It is also surprising that popular tools used extensively in the translation industry (i.e., translation memory systems) are still not widely integrated into the AVT workflow, not only in revoicing but also in other AVT modes (Baños, 2018). The reluctance typical of the dubbing industry to embrace new technologies that was documented by Chaume (2007) more than 15 years ago seems still to be somewhat pertinent, with dubbing practices remaining conservative and resistant to homogenization. However, this scholar has recently referred to new developments that, while not being widespread yet, are progressively resulting in remarkable transformations:

Cloud dubbing is starting to find a place in the market, and is opening up possibilities for agents to work in a collective environment where the project manager can keep track of all stages of the dubbing process, glossaries are constantly updated and available to everybody, episodes can easily be split between several translators, who can work online and consult each other, voice talents can dub online under the supervision of the dubbing director and so forth. (Chaume, 2020, p. 104)

Regarding the impact of technology in other stages of the dubbing process, Chaume (2020, p. 108) contends that digital technology has transformed the way voices are recorded and synchronized in the dubbing studio. In similar vein, Bywood (2020, p. 511) argues that the main focus of technological development in the dubbing industry has been the process of synchronization. These developments have included the dedicated software used in the French and Québécois dubbing industry to insert the rhythm band or *bande rythmo*, a track added to an audiovisual programme for dubbing purposes only. Specialized software such as VoiceQ Writer⁵, and also more recent cloud dubbing platforms such as ZOODubs⁶, allows the rhythm band to be added to audiovisual texts to ensure that the translated dialogues are synchronized with the original before being revoiced by dubbing actors.

Technology has even claimed to be able to solve the complex challenges of synchrony and voice acting in dubbing. In the 2007 article cited above, Chaume (2007, pp. 213–214) highlighted the potential application of software such as Video Rewrite, which was capable of reusing existing footage to create a new clip featuring a person mouthing words not spoken in the original footage, or Reel Voice, a tool that relied on voice-conversion

⁵ See voiceq.com

⁶ See zoodigital.com/technology/zoodubs

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

technology to make a specific voice resemble another and therefore to avoid discrepancies between the voices of original and dubbing actors.

While these technologies did not succeed in the dubbing industry as its developers envisaged at the time, they were the precursors of current developments in this field, such as the implementation of AI to manipulate video images for synchronization purposes in dubbing or of voice-cloning technologies for speech synthesis. The former, often referred to in the media as ‘deepfake’, has been labelled by Chaume (2020) as a type of technology that favours “the adaptation of visuals to (translated) words” (p. 121), where the mouth movements of characters on screen are manipulated to match the text provided by the translator. As Chaume (2020) argues, these technological developments would render the dialogue writers’ job obsolete, given that image manipulation will ensure that any translation matches the articulatory movements of on-screen characters. This technology is already available commercially in the form of tools such as TrueSynch⁷, which its developers have defined as “the world’s first system that uses Artificial Intelligence to create perfectly lip-synced visualizations in multiple languages” (Flawless, 2022, para. 1). While this type of technology has been used for parodic purposes in some contexts (i.e., in the Spanish news satire television programme *El Intermedio*, as explained in laSexta (2022)), given the ethical issues surrounding the use of deep-learning technology to manipulate images and the consideration of some audiovisual content (especially films and TV series) as artistic work, the future of these tools in the dubbing sphere is rather uncertain.

As these applications show, AI has recently been integrated into dubbing workflows in different ways in an attempt to automate specific tasks. More recent developments in AI have focused on the full or nearly full automation of the dubbing process, through automatic dubbing solutions, which is discussed below.

2.1. Automatic dubbing: an overview

Automatic dubbing has been defined in recent research in AI and computational linguistics as “the task of automatically replacing the speech in a video document with [synthetic] speech in a different language, while preserving as much as possible the user experience of the original video” (Lakew et al., 2021, p. 7538). In the AI industry, automatic dubbing is often seen as an extension of speech-to-speech translation (S2ST) (Federico et al., 2020), with systems being a pipeline or combination of the following technologies: automatic speech recognition (ASR), used to transcribe speech to text in the source language (SL); MT, to translate the transcribed text in the SL automatically into a target language (TL); and text-to-speech (TTS) technology, to convert the MT output into speech. However, automatic dubbing differs from S2ST in that the main goal of S2ST is to produce an output that reflects the content of the original speech, whereas automatic dubbing should not only sound and look as natural as the original (Federico et al., 2020), but also “reproduce the original emotions, coherent with the body language, and be lip synchronized” (Di Gangi et al., 2022, p. 351).

⁷ See flawlessai.com/product

In order to tackle these challenges, the research on this technology has focused on the enhancement of traditional S2ST pipelines in different ways. For instance, Di Gangi et al. (2022) explain how speaker diarization is used in their ASR system in order to be more accurate when detecting who is speaking, “so that consecutive segments from the same speaker can be translated as coherent units, and each speaker is assigned a unique voice” (p. 351). In similar vein, Saboo and Baumann (2019) illustrate how to integrate into the MT paradigm the synchrony constraints dubbing is subject to in order to achieve “more ‘dubbable’ translations” (p. 94). The authors argue that training neural MT (NMT) engines on dubbed audiovisual corpora is not sufficient to yield translations that observe dubbing constraints. For instance, in order to maintain isochrony in automatic dubbing, speeding up the delivery of the MT output during the TTS phase does not suffice; it is also necessary to control the length of the MT output, making sure that the translation has a similar number of characters or syllables as the original (Lakew et al., 2021; Saboo & Baumann, 2019). However, the output should also respect the pauses in original utterances, which can be accomplished through prosodic alignment (Öktem et al., 2019). In order to achieve this, the transcript of the original clip should include time stamps, which can be provided by the ASR component or taken from a previously uploaded subtitle file. Prosodic alignment can have a negative impact on the fluency and prosody of the generated speech in the TL (Federico et al., 2020, p. 264), revealing the challenging nature of the task at hand and the need for very careful fine-tuning between the different technological components.

Achieving lip synchrony through automatic dubbing, especially during close-ups, is much more challenging and at the time of writing studies have suggested only potential solutions. As future developments, Saboo and Baumann (2019, p. 99) refer to the implementation of more powerful synchrony metrics that could take into consideration the syllables that are stressed in the original dialogue in addition to the presence of specific shots requiring lip synchrony (i.e., close-ups and extreme close-ups). The matter of lip synchrony has also been taken into consideration in recent studies which have highlighted the importance of distinguishing between on-screen and off-screen dubbing to devise viable solutions for “dubbing-customised NMT” (Karakanta et al., 2020, p. 4328).

Another important hurdle in automatic dubbing is the naturalness of the synthetic voices used, which are part of the TTS component. Di Gangi et al. (2022) explain that AppTek Dubbing resorts to speaker-adaptive TTS “to reproduce the voice features of the original actor” (p. 351) and acknowledge that the main limitation of their current system is “the synthesized voice speaking with a ‘flat’ tone, which does not match the emotions expressed in the original video” (p. 352). In order to overcome this limitation, they are planning to introduce improvements based on the annotation of emotions in the original audio and to support these through “emotion-aware” MT and TTS (Di Gangi et al., 2022, p. 352).

Some of the systems commercially available and those described in the studies cited above allow the user to skip a specific stage by, for example, inputting a subtitle file instead of resorting to ASR or inputting a translation undertaken or reviewed by human beings, therefore skipping the MT stage. Although most of the research cited above envisages scenarios where the dubbing process can be fully automated, there is always room for human intervention and quality control (QC). This is essential, given that these systems “can

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

make errors in multiple points of [their] pipeline, and the earlier the errors occur, the more harmful they can be for the final result” (Di Gangi et al., 2022, p. 352).

Even though the research undertaken in computational linguistics has to some extent taken into consideration the perception and disposition of audiences towards automatic dubbing, there does not seem to be much collaboration with dubbing professionals or with AVT scholars. In Translation Studies, with only a few exceptions, the research on automation and revoicing has been very limited. The ALST projectⁱⁱⁱ, targeted at investigating the use of MT, speech-recognition and speech-synthesis technologies for voiceover and audio description (AD) (Matamala, 2015), was pioneering in this field. This project failed to indicate the cost-effectiveness of ASR in the transcription of non-fictional content in comparison to other transcription methods. However, the results regarding the use of speech synthesis in AD and of statistical MT (SMT) in wildlife documentaries translated from English into Spanish were more encouraging, suggesting the potential application of technology in a wide range of AVT contexts and scenarios and highlighting the need to carry out further large-scale research in this field.

Martín-Mor and Sánchez-Gijón (2016) have also investigated the challenges of using MT and post-editing (PE) in translating different types of text from English into Spanish in a documentary film. In addition to concluding that MT (and SMT in particular) can be used successfully to translate audiovisual products containing formal and well-recorded speech (e.g., narration in documentaries), which also indicated the potential to translate training and instructional videos, the authors highlight some key challenges:

an important obstacle with audiovisual products is obtaining a written source text that can be fed into the MT system. Either if the audio is transcribed manually or by using voice-recognition technology, texts need heavy editing before they are ready for MT, especially where the soundtrack contains ambient noise. Hesitations and other features of oral discourse pose additional problems, since they are difficult to transcribe into text that can be processed by a MT system and are difficult for such systems to translate. (Martín-Mor & Sánchez-Gijón, 2016, p. 183)

These authors also expressed the need to compare the time required, the cognitive effort and the cost of human translation of a formal documentary versus MT followed by PE. Regarding the first parameter, a recent experiment undertaken by De los Reyes Lozano (2022) with undergraduate trainee translators working from English into Spanish and Catalan has suggested that human translation and the adaptation of fictional content for dubbing tend to take at least twice as long as the use of MT and full PE. The author emphasizes that, in the case of dubbing, MT should be seen as a technology not aimed at replacing translators but rather at optimizing and streamlining the translation and dialogue writing process with human intervention. The important role played by translators or language experts in automatic dubbing workflows is also acknowledged by some of the developers of the current solutions available in the market, as is discussed in the next subsection.

2.1.1. Overview of current automatic dubbing tools in the language industry and main content targeted

In addition to the organizations mentioned in the previous discussion (i.e., Google, Amazon and AppTek), many more companies working on AI and/or on media localization have

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

jumped on the automatic dubbing bandwagon, with many tools currently being available in the market. Table 1, compiled by undertaking independent research online and reviewing relevant publications and reports from market research companies such as Nimdzi (Akhulkova et al., 2022) and Slator (2021), provides an overview of some of the automatic dubbing solutions available at the time of writing.

Table 1

Overview of some of the most popular automatic dubbing solutions available in the market

Tool	Company	Main content/ market targeted
Aloud aloud.area120.google.com	Google's Area 120	Content creators posting on video sharing platforms such as YouTube Educational videos
AppTek Dubbing apptek.com/solutions/apptek-automatic-dubbing	AppTek	News and media User-generated content Education and E-learning
Deepdub deepdub.ai	Deepdub	Entertainment industry (i.e., Topic.com)
Dubdub appengine.ai/company/dubdub-ai	Dubdub AI	
Dubverse dubverse.ai	Dubverse	Educational and training videos Indie creators YouTubers
Klling klleon.io/product	Klleon	
Maestra Video Dubber maestrasuite.com/video-dubber	Maestra	Educational and training videos News and media
Mediacat xl8.ai	XL8	Media LSPs that need high-quality content Live events Markets that do not need post-editing
MW Vox mediawen.com/en/products/mwvox	Mediawen	Sports Corporate videos E-learning
Papercup papercup.com	Papercup	Sports and news Educational and training videos Entertainment
SmartDub straive.com/solutions/smartdub	Straive	Educational and training videos Documentaries Gaming and entertainment
Vidby vidby.com	Vidby	Start-ups and SMEs Large organizations NGOs Influencers and bloggers
VideoDubber videodubber.com	VideoDubber	News Educational and training videos Documentaries Talk shows and lifestyle shows

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

This list includes only those commercially available tools that integrate the three technologies mentioned above (ASR, MT and TTS) into automatic dubbing, even if these can be used separately by clients. In addition to these, there are other tools that focus on TTS solutions only (see, for example, matedub.com), as well as new developments that resort to AI to create videos from scratch (see, for example, synthesia.io), which could also be used in the dubbing industry.

Far from being exhaustive, this list is included to illustrate the fast pace at which these new developments are entering the market and to show the type of content being targeted. Although some developers claim that their automatic dubbing technologies are suitable for entertainment content such as films and TV series and provide relevant examples on their websites (notably DeepDub), the majority seem to suggest that these tools work better for non-fictional content. In particular, there seems to be an emphasis on educational and training materials, whether these are provided by educational institutions, corporations or independent creators (e.g., YouTubers). Independent creators are often identified as ideal users of some automatic dubbing tools. This is particularly the case of tools that are marketed as being free or relatively cheap, easy to use and quick, and that allow creators to skip the revision or QC stage if this is what they wish or can afford. While the FAST and AVOD markets, constituted by free ad-supported video content platforms such as YouTube, tend to do without QC processes or implement them only superficially or minimally, larger media companies that have bigger budgets might consider using automatic dubbing solutions that integrate more sophisticated and frequent QC processes. This is what Sky News did when it used Papercup to dub its content from English into Spanish.

As argued by Akhulkova et al. (2022), it is remarkable that these companies often refer to their technology as offering a ‘new way to dub’. Such references not only foreground the innovative nature of this technology but also its immaturity. Given the novelty of automatic dubbing tools, the challenges they pose to content creators, the language industry at large and viewers have not yet been explored. This article attempts to fill this gap, with a focus on the integration of MT and speech synthesis into the automatic dubbing workflow.

3. Investigating key challenges of using automatic dubbing to translate educational YouTube videos

In order to understand the key challenges that the technologies described above pose, this section evaluates the performance of Aloud, an automatic dubbing solution from Area 120, Google’s in-house incubator for experimental projects. This product was chosen because of its ethos and potential and because the audiovisual content dubbed with this tool is readily available on YouTube. With the aim of opening up “previously inaccessible information to the world” (Kottahachchi & Abeysinghe, 2022, section 5) and with a focus on informational content, Aloud developers have provided early access to their tool to audiovisual content creators who “previously considered dubbing too difficult or too costly” (Kottahachchi & Abeysinghe, 2022, section 1).

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

In particular, *Kings and Generals*⁸ (n.d.), a YouTube channel of animated historical documentaries with 2.8 million subscribers, and *Amoeba Sisters*⁹, a YouTube channel of animated videos on scientific topics with 1.48 million subscribers, have used Aloud to dub their content automatically. *Kings and Generals* seems to have used Aloud to dub only one video on Indo-European languages¹⁰ into Hindi, Indonesian, Portuguese and Spanish (*Kings and Generals*, 2022). This was done using a recently introduced YouTube feature that is currently being tested and which allows some content creators to enable multiple audio tracks for a single video. *Amoeba Sisters*, however, has created two separate YouTube channels where they post their videos dubbed into Spanish (*Amoeba Sisters en Español*¹¹ (n.d.), with 4,270 subscribers at the time of writing), and into Portuguese (*Amoeba Sisters em Português*¹² (n.d.)), with only 525 subscribers at the time of writing).

Considering the languages involved, the availability of the material, and the potential impact of the content dubbed, the analysis focused on the videos posted on *Amoeba Sisters en Español*. Aimed at “demystify[ing] science with humour and relevance” (*Amoeba Sisters*, n.d., info section), the videos from this channel are particularly suited to discussing key challenges such as the automatic dubbing of humour or specialized terminology and to investigating the way relevant synchronies are maintained in the final version. Targeted at high school students, the videos published by the two sisters behind this channel are usually under 10 minutes in length and focus on biology topics.

Given the experimental nature of the services provided by Aloud and the emphasis placed on viewers on their website (which features quotes from satisfied users), the analysis started by documenting the issues highlighted by YouTube users in their comments on the channel *Amoeba Sisters en Español*. To this end, all the comments included under the 29 videos posted on this channel were reviewed with the purpose of summarizing the main problems and positive comments highlighted by viewers. Each comment was entered in a spreadsheet and, in addition to indicating its authorship, the topics discussed were identified and recorded. If a comment discussed more than one issue, the issues were recorded separately by adding a new row (see Figure 1). As a result, the number of issues or topics identified exceeded the number of comments analysed.

⁸ See youtube.com/c/KingsandGenerals/videos

⁹ See youtube.com/c/amoebasisters

¹⁰ See youtube.com/watch?v=VpXgMdvLUXw

¹¹ See youtube.com/channel/UC1Njo3LBy53cOPngz6ArV8Q

¹² See youtube.com/channel/UCYTQPX2X_mXe0ZMPi0fXxbg

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

Figure 1

Spreadsheet illustrating the categorization of comments throughout the analysis

	A	B	C	D	E
1	Video	Comment	Author	Type of comment	Specific comments on MT/Voices
108	Introducción a las células	Some part of the translation is quite funny. They say "orgánulos unidos a la membrana" instead of "orgánulos envueltos por una membrana" it changes the meaning a lot! Anyway, I live in Barcelona and sometimes I use those accurate, cute and fantastic videos with my students(in english) They will be delighted to know that they can listen to them in spanish, And I could ask them to help you dubbing some videos into Catalan.	Users	Translation	MT - Inaccurate translations
109	Introducción a las células	Some part of the translation is quite funny. They say "orgánulos unidos a la membrana" instead of "orgánulos envueltos por una membrana" it changes the meaning a lot! Anyway, I live in Barcelona and sometimes I use those accurate, cute and fantastic videos with my students(in english) They will be delighted to know that they can listen to them in spanish, And I could ask them to help you dubbing some videos into Catalan.	Users	Translation	MT - Awkward sounding translations
110	Introducción a las células	Listo para reprobare el examen de mañanaaaaa It's just depressing. That monotonous, robotic voice completely kills these amazing cartoons	Users	Not related/relevant/clear Voice	
111	Introducción a las células	I offer to translate this better for you.	Users	Translation	Artificial/Robotic/Not natural
112	Introducción a las células	¿Introducción a las CELDAS? ¿En serio? In Spanish, biological cells are translated as "células". It's Basic biology. If you want to make science divulgation in Spanish, get and pay a good translator or DON'T DO IT.	Users	Translation	MT - Bad translation
113	Introducción a las células	>Thank you so much for this feedback; we apologize for the mistake. We think this was a mistake isolated with the title only and not in the video or description but we'll be double checking to confirm.	Users	Translation	MT - Inaccurate translations
114	Introducción a las células		Creators	Creators - reply	

The overview of viewer's perceptions was supported by data from a more detailed analysis of one of the videos whenever it was considered that this could help to illustrate and understand the challenges identified by viewers from a translation and dubbing perspective. The video chosen was the most popular one, entitled "Biomoléculas" (Amoeba Sisters en Español, 2021)¹³, which was published on 21/10/2021 and had achieved 30,942 views at the time of writing. As the material provided shows that *Amoeba Sisters* has uploaded an (almost verbatim) English subtitle file to the system instead of relying on Aloud's ASR component, the focus of the analysis was on the implementation of NMT and TTS. Considering the limitations of this exploratory study, the aim is not to provide an in-depth analysis or quality evaluation, but rather to bring to the fore the key challenges and potential of these first implementations of automatic dubbing.

3.1. Exploratory analysis of the *Amoeba Sisters* videos dubbed automatically into Spanish

Data analysis was undertaken from 15 September to 31 October 2022. At the time, 29 videos were available on the channel *Amoeba Sisters en Español* accompanied by 174 comments, that is, 36 creators' comments and 138 users' comments. As regards the former, the following comment, originally in Spanish, was pinned to the majority of the videos analysed (with only one exception): "This video has been dubbed into Spanish using an artificial voice with <https://aloud.area120.google.com> in order to increase accessibility" (my translation). With this comment, *Amoeba Sisters* meets Aloud's requirement that videos dubbed on their platform should state clearly that they are synthetic dubs (Kottahachchi & Abeyasinghe, 2022). In addition, the creators replied to specific comments from users (two replies) and posted the following comment (in both English and Spanish) on six videos in order to gather additional feedback, mainly on the use of synthetic voices:

¹³ Biomolecules, see [youtube.com/watch?v=gk9WmmRWgxA](https://www.youtube.com/watch?v=gk9WmmRWgxA)

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

This video has been dubbed into Spanish using an artificial voice. We'd love to hear your feedback: what is it like to watch this video? Would you like to see more videos dubbed into Spanish with this voice, or would you prefer the original video with Spanish closed captions? Why? (AmoebaSistersEspañol, 2021)

Table 2 includes the data gathered from the analysis of the 174 comments, distinguishing between those posted by the creators and by users. As indicated in the previous section, the total number of references (194) exceeds the number of comments because some users commented on more than one aspect in a single posting.

Table 2

Data analysis of the comments posted on the Amoeba Sisters videos dubbed into Spanish using Aloud

Comments from creators		Comments from users	
Comments with information about synthetic voices	28	Positive comments	72
Comments requesting specific feedback	6	Critical comments on synthetic voices	41
Replies to other comments	2	Not related/relevant comments	33
		Critical comments on the translation	12
Total	36	Total	158

As shown in Table 2, nearly half of the references identified in the comments made by users were positive (72), with many expressing their gratitude for the effort made by the content creators to dub their material into Spanish. However, some of these comments also indicate that the voices sound unnatural. The negative feedback from users tends to concentrate on the synthetic voices (41 references) and not so much on the actual translation (12 references). Regarding the former, as indicated in Table 3, most of the comments (28) refer to the synthetic voice being too artificial, robotic or unnatural, with some users specifically arguing that this renders the content not as appealing, engaging or easy to follow as the original videos (10 instances). It therefore seems that synthetic voices can impinge on comprehension both because they feel ‘boring’ or monotonous and because they are too fast to follow (two references). As discussed in section 2.1, this last criticism highlights the challenges of meeting isochrony when automatic dubbing is used, since, in this case, the generated audio track needs to be played at a higher speed to have a similar duration to the original. While the analysis undertaken has not identified any problems regarding action synchrony (Franco et al., 2013), that is, the synchronization of the generated audio track with the images shown on screen, this is also an area that needs attention in automatic dubbing.

One specific comment refers to the synthetic voice not pronouncing the initialism ARN (RNA in English) appropriately, a simple but serious mistake, especially if we consider the importance of accuracy in educational content and the fact that some of these videos have a strong focus on this molecule (i.e., the videos “DNA vs RNA” and “Biomolecules”). Unlike acronyms, initialisms are pronounced as separate letters, but TTS systems struggle to differentiate between them. Given that both acronyms and initialisms are extremely common in scientific communication, this problem is likely to recur in these videos. Indeed, we can find another example in the video “Biomoléculas”, where the initialisms CHO and CHONP are pronounced as an acronym and an initialism respectively.

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

Table 3

Critical comments on synthetic voices and translation-related aspects from YouTube users on the Amoeba Sisters videos dubbed into Spanish with Aloud

Critical comments	Specific references	Total
Critical comments on synthetic voices (41)	Voice being artificial/robotic/not natural	28
	Voice makes content not as appealing/engaging	10
	Voice speaking too fast	2
	Incorrect pronunciation (initialisms)	1
Critical comments on translation-related aspects (12)	MT – Awkward sounding translations	2
	MT – Bad translation	4
	MT – Inaccurate translations	3
	MT – Use of register	1
	Onscreen text left untranslated	2

While many comments (especially those that are only positive) are posted by users who consult these videos to support their learning, others come from tutors who consider using them in their classes. However, as the examples^{iv} below show, the artificial voice deters some of them from doing so:

- I always use your videos an images for my class (i always have to translate the images) i was very excited for this spanish videos but... this robotic voice is terrible, i love your Chanel because you love biology and explain it with so much love, i dont feel the same with this spanish videos... I belive you need help of a native spanish speaker that can do this for you insted of the robot (montsezaragoza9132, 2021).
- This is wonderful, but the voice sounds very robotic. I don't think that my students would want to listen to this voice. I would love to hear a real person in the translation. Thank you very much for doing this. If I were in your area I would do it for you. 😊 (lucyhernandez7672, 2021)

While other users explain that they would prefer it if a human voice were used in the dubs, they acknowledge that this is a very useful educational resource for those studying Biology. In addition, some non-native Spanish speakers indicate in the comments that these videos are also useful for language learning, highlighting the potential of YouTube videos to help individuals to learn foreign languages informally (Terantino, 2011) and to acquire vocabulary in specialized contexts (Macrea et al., 2023). Acknowledging the positive impact of these videos, many users show their willingness to help by lending their voices and their expertise. This is reflected in the following comment, which provides detailed feedback to help the wider community and, in doing so, encapsulates the key challenges automatic dubbing poses to translating these videos:

- Hey! Native Spanish speaker here, I'd love to give some feedback. Firstly, I'm a Mexican high school grad, and a big fan of your work! I'm thrilled that you guys dropped a Spanish channel! [...] Here are some points on how you could make it more fun for Spanish speakers to watch:
 - Something that makes your videos so enjoyable is Pinky's charismatic voice, and we can find artificial voice a bit less engaging. So, if you could put a natural voiceover in Spanish, it'll totally make the video easier to digest. [...]

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

- On to the illustrations: translating them to Spanish would make such a massive difference! Part of the fun of your videos are all those animated puns and jokes, it makes them so much fun and it's easier to recall the information! Also, all the text in English might confuse the viewers while trying to understand the concepts that are being visually represented. Also, if not possible, CC are okay.
- Regarding the translation, I only find one thing that makes it sound unnatural which is the use of the pronoun “usted” instead of “you” when addressing the viewer, as I find the latter one more appropriate to project friendliness and it's more common in most YT videos.
- I really appreciate you guys making this channel as it is extremely hard to find videos like this in Spanish. They are the reason I'm a great performing student now [...] And I'd love other students from Latin America to be able to get this kind of educational content [...] (abigarcia9588, 2022)

The reliance of educational videos on text on screen is a serious hurdle for automatic dubbing to overcome – and one that seems to be frequently disregarded by tool developers. The *Amoeba Sisters* videos illustrate the nature and relevance of this challenge because the text on screen is used both to reinforce and to complement the content communicated through the audio track, sometimes with a comic purpose or, more importantly, as a mnemonic device to help viewers remember scientific facts. Translating and replacing all instances of text on screen might be cumbersome and time-consuming, almost resulting in the creation of new content. Yet such an undertaking will arguably make for a more positive, complete, and engaging learning experience. Although many YouTubers who decide to dub their content seem to neglect the importance of the text on screen, leaving it untranslated (see, for instance, the content from the popular channel MrBeast¹⁴, dubbed into Spanish), some educational videos provided by projects supported by volunteers such as Khan Academy do translate it (see, for instance, the video “Introduction to the cell”¹⁵).

The use of an inappropriate register highlighted in the previous example is common in NMT and has been pointed out in similar studies (De Los Reyes Lozano, 2022). This issue is aggravated by the fact that NMT does not deal with register in a consistent manner. This is shown in the translation of the following sentence, from the clip “Biomoléculas”, where the first part addresses the viewer formally (*quiero que piense*) and the second part uses the informal *tu* (instead of the formal *su*), resulting in an inconsistent rendering:

Example 1	
Video	“Biomoléculas” (youtube.com/watch?v=gk9WmmRWgxA)
Timecode	00:09–00:12
Original version	I want you to think for a moment about your very favourite food.
Dubbed version & back translation	Quiero que piense por un momento en tu comida favorita. [I would like you to think for a moment about your favourite food.]

¹⁴ See youtube.com/user/MrBeast6000

¹⁵ See es.khanacademy.org/science/high-school-biology/hs-cells/hs-basic-cell-structures/v/introduction-to-the-cell?modal=1

Apart from this comment, no other users refer to this issue specifically, with translation-related aspects attracting less attention in comparison to synthetic voices. Some users simply refer to the translation being bad or suggest resorting to human beings to do it better (four references, including “just invest in a person who translates it properly ...”(angelguardiaardiaca7364, 2022)), while others are more specific and indicate that the issue lies in its accuracy (MT – Inaccurate translations) or fluency (MT – Awkward-sounding translations):

- First of all, congratulations on your channel. The dubbed translation in Spanish is quite bad,... I mean, we can understand the overall meaning, but some sentences sound weird, forced, and have incorrect scientific terms. I really wouldn't recommend the use of them to my students. It would be great if some Spanish speakers could revise the translation before you upload the videos (carneboix, 2021).

These matters of accuracy and fluency seem to go unnoticed by users who visit this YouTube channel to support their learning; moreover, these comments are posted by users with some kind of expertise in the topic (e.g., mostly teachers). The inaccuracies detected by users are concerned with the translation of key terminology (i.e., “cell” translated as *celda* instead of *célula*) and could be avoided if the NMT component could be further customized to consider the domain or topic discussed in the videos, which does not seem to be possible with Aloud. Other examples of the incorrect translation of polysemic terms identified in the analysis of the video “Biomoléculas” (e.g., “seal” is translated as *sello*, in the sense of stamp, instead of referring to the animal *foca*, at time code 02:36) could be resolved if the system could gather information from the visuals (since a white seal is shown on screen when this term is referred to in the original audio). These are also very obvious mistakes that could easily be detected by a post-editor, as indicated in the previous comment. The lack of post-editing in these scenarios is problematic, even if creators attempt to remedy them through crowdsourcing, as *Amoeba Sisters* has done through their website, asking followers to contribute to the translation of subtitle files or to the revision of the MT output produced by Aloud before it is revoiced by the system (Amoeba Sisters, 2022).

The analysis of the video “Biomoléculas” reveals that these problems are not isolated and, more importantly, that they have an impact on the features that make these videos unique, or at least appealing, to viewers. Users thank *Amoeba Sisters* for explaining complex concepts in a simple way and for making these fun and easy to remember. To achieve this, the original audio often resorts to parenthetical structures and explanations and also to mnemonic devices and wordplay. However, these devices are not always rendered by Aloud as originally intended, as illustrated in the following example:

Example 2	
Video	“Biomoléculas” (youtube.com/watch?v=gk9WmmRWgxA)
Timecode	05:01–05:30
Original version	Now, when we start talking about genes – the DNA genes, not the jeans you wear – the DNA codes for proteins that are very important for structure and function in the body. The last big biomolecule is known as a nucleic acid. [...] They have a monomer called a nucleotide. That's going to be an easy one for you to remember because nucleotide sounds a lot like nucleic acid.
Dubbed	Ahora, cuando empezamos a hablar de genes, los genes del ADN, no los <i>jeans</i> ,

<p>version & back translation</p>	<p>usamos los códigos de ADN para las proteínas que son muy importantes para la estructura y función del cuerpo. La última gran biomolécula se conoce como ácido nucleico. [...] Tienen un monómero llamado nucleótido. Será fácil de recordar porque los nucleótidos se parecen mucho a los ácidos nucleicos.</p> <p>[Now, when we start talking about genes, the DNA genes, not the jeans, we use the DNA codes for the proteins that are very important for the structure and function of the body. The last big biomolecule is known as a nucleic acid. [...] They have a monomer called nucleotide. It will be easy to remember because nucleotides look a lot like nucleic acids.]</p>
---------------------------------------	--

The rendering of both the wordplay (genes/jeans) and the parenthetical structure (the DNA genes, not the jeans you wear) does not make sense in the target version. This is not only because wordplay is challenging to translate (not only by NMT but by human beings too, especially if we consider that the audio is accompanied by an image referring to a gene within a DNA molecule and also a character wearing a pair of jeans), but also because the original is not grammatically correct. While understanding the original sentence is not an issue for English-speaking viewers and its translation would not be challenging for a human being, it is problematic for the NMT component. In this example, this results in the addition of the verb *usamos* in the dubbed version and therefore in an inaccurate translation and explanation. In addition, the explanation “the DNA genes not the jeans you wear” (Amoeba Sisters, 2016, 05:04) is not fluently revoiced as a parenthetical structure in the dubbed version, which leads to an even more confusing rendering.

In the second part of the example, Pinky encourages viewers to remember the word “nucleotide” by drawing attention to how similar it sounds to “nucleic acid”, with the letters “nucle” being highlighted on screen in a different colour when the relevant sentence is uttered. However, by referring to the physical resemblance in the dubbed version (*se parecen mucho* > they *look* alike), the mnemonic device is not only lost, but is also inaccurately rendered. Another aspect worth mentioning is the translation of the discourse marker “now”, which is translated as *ahora* in Spanish, where it usually functions as an adverb of time and not as an orality marker (Baños, 2014). This reveals another pain point for automatic dubbing, since many features of spontaneous discourse included in the original videos are sometimes translated into unnatural expressions by the NMT engine or revoiced in an unusual way (without appropriate pauses) by the TTS component. Again, this could probably be solved with human intervention (i.e., by changing punctuation), but this kind of tweaking is not possible without the appropriate linguistic knowledge.

4. Final remarks

This article has investigated automatic dubbing, understood as the combination of ASR, MT and TTS technologies to automatically replace the audio track of an original audiovisual text with synthetic speech in a different language, taking into consideration the relevant synchronies. In doing so, it has portrayed dubbing as an AVT mode inextricably linked to technological developments and reflected on the increasing demand for automatic dubbing services in the current mediascape. It has also provided an overview of the existing tools and current research on this topic. Although automatic dubbing tools have been tested and implemented in the language industry on a wide range of material and AVT modes (i.e.,

voiceover, off-screen dubbing and lip synch dubbing), educational training content to be translated through off-screen dubbing or voiceover has been recognized as being particularly suitable for the use of automatic dubbing tools. In line with this, the object of study in this article has been educational YouTube videos posted by *Amoeba Sisters* on their Spanish channel, which have been dubbed automatically using Aloud, apparently with very little human intervention.

The analysis undertaken has highlighted key challenges in the automatic dubbing of these videos, especially regarding the implementation of NMT and speech synthesis. Some of the findings highlight areas for improvement already identified in existing research; others suggest that more attention needs to be paid to accuracy and to the specificities of audiovisual texts, which have been only partially considered in previous studies. The analysis has shown that there is still work to be done regarding the naturalness of synthetic voices, and also the pace at which and the prosody with which the generated audio track is delivered. Impinging on fluency, these aspects also have an impact on comprehension and engagement, which is undesirable in the translation of educational and training content. Naturalness and fluency attract the attention not only of researchers, but also of viewers and users; yet the importance of accuracy in the educational material analysed seems to have been disregarded (mainly because the majority of users are learners and do not have the appropriate knowledge to identify these issues). In addition to inaccuracies, it is argued here that more attention should be paid to other components of the audiovisual text (not only to the dialogue), especially to on-screen images. Aspects that also need to be considered in the automatic dubbing workflow include the role of text on screen (which is left untranslated) and the way in which compliance with isochrony (sometimes achieved by speeding up the delivery of the generated audio track) might have an impact on action synchrony.

Developments in AI have undoubtedly facilitated the revoicing of audiovisual material – in our particular case by enabling Spanish-speaking YouTube users to enjoy educational content dubbed in their language, thus enhancing audiovisual accessibility and social inclusion. Valued by viewers and meeting specific demands in a burgeoning industry, automatic dubbing has great potential not only in commercial and educational environments, but also in other contexts such as crisis settings, where it could contribute to strengthening social resilience against adversity (Federici et al., 2023). In these scenarios, Federici et al. (2023) issue a warning about volunteerism and technology lulling users into a false sense of security regarding the (translation) quality achieved through automated processes. This cautious approach is equally applicable in the context discussed here.

The matters identified in the analysis have a negative impact on the features that make the *Amoeba Sisters* videos appealing to viewers. Content creators should therefore be aware of the limitations of automated dubbing tools and the need for QC to ensure that a more engaging and fulfilling experience is achieved. Disentangling the complexities of audiovisual texts and the interaction of the different codes at play in the creation of meaning (and even humour) is a challenging task, one in which some form of human intervention is vitally needed.

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

References

- abigarcia9588. (2022). Re: *Fotosíntesis*. [Video]. YouTube. <https://www.youtube.com/watch?v=2-lg8yOsZgA>
- Akhulkova, Y., Hickey, S., & Hynes, R. (2022). *The Nimdzi language technology atlas*. <https://www.nimdzi.com/language-technology-atlas>
- Ameri, S., & Khoshsaligheh, M. (2018). Exploring the attitudes and expectations of Iranian audiences in terms of professional dubbing into Persian. *HERMES – Journal of Language and Communication in Business*, 57, 175–193. <https://doi.org/10.7146/hjlc.v0i57.106206>
- Amoeba Sisters (2016) *Biomolecules (Older Video 2016)* [Video]. YouTube. <https://www.youtube.com/watch?v=YO244P1e9QM>
- Amoeba Sisters. (2022). *Pinky's ed tech tips: Translations*. <https://www.amoebasisters.com/pinkys-ed-tech-favorites/community-contributed-subtitles>
- Amoeba Sisters (n.d.) *Home* [Video]. YouTube. <https://www.youtube.com/@AmoebaSisters>
- Amoeba Sisters en Español (2021, October 21) *Biomoléculas* [Video]. YouTube. <https://www.youtube.com/watch?v=gk9WmmRWgxA>
- Amoeba Sisters en Español (n.d.) *Home* [Video]. YouTube. <https://youtube.com/@AmoebaSistersEspanol>
- Amoeba Sisters em Português (n.d.) *Home* [Video]. YouTube. <https://youtube.com/@AmoebaSistersPortuguese>
- AmoebaSistersEspanol. (2021). Re: *Replicación de ADN*. [Video]. YouTube. <https://www.youtube.com/watch?v=ddnul2kYJ6c&>
- angelguardiaardiaca7364. (2022). Re: *Respiración celular*. [Video]. YouTube. <https://www.youtube.com/watch?v=XZ246DX3zAU>
- Baños, R. (2014). Orality markers in Spanish native and dubbed sitcoms: Pretended spontaneity and prefabricated orality. *Meta*, 59(2), 406–435. <https://doi.org/10.7202/1027482ar>
- Baños, R. (2018). Technology and audiovisual translation. In C. Sin-wai (Ed.), *An Encyclopedia of practical translation and interpreting* (pp. 3–30). Chinese University Press.
- BBC Media Centre. (2015). *BBC introduces new automatic virtual voiceover translations: Media centre*. <https://www.bbc.co.uk/mediacentre/latestnews/2015/bbc-virtual-voiceover-translations>
- Bywood, L. (2020). Technology and audiovisual translation. In Ł. Bogucki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility* (pp. 503–517). Palgrave. https://doi.org/10.1007/978-3-030-42105-2_25
- carneboix. (2021) Re: *Comparación de mitosis y meiosis*. [Video]. YouTube. <https://www.youtube.com/watch?v=az1S54aMskI>
- Chaume, F. (2007). Dubbing practices in Europe: Localisation beats globalisation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 6, 203–217. <https://doi.org/10.52034/lanstts.v6i.188>
- Chaume, F. (2020). Dubbing. In Ł. Bogucki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility* (pp. 103–132). Palgrave. https://doi.org/10.1007/978-3-030-42105-2_6
- De Los Reyes Lozano, J. (2022). Traducción automática et doublage: Impressions d'une expérience d'enseignement. *Journal of Data Mining and Digital Humanities*. eISSN 2416-5999. <https://doi.org/10.46298/jdmdh.9119>
- Di Gangi, M., Rossenbach, N., Pérez, A., Bahar, P., Beck, E., Wilken, P., & Matusov, E. (2022). Automatic video dubbing at AppTek. *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, 351–352. <https://aclanthology.org/2022.eamt-1.65>

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

- Di Giovanni, E., & Romero-Fresco, P. (2019). Are we all together across languages? An eye tracking study of original and dubbed films. In I. Ranzato & S. Zanotti (Eds.), *Reassessing dubbing: Historical approaches and current trends* (pp. 125–144). John Benjamins. <https://doi.org/10.1017/btl.148.06di>
- Federici, F., Declercq, C., Díaz Cintas, J., & Baños Piñero, R. (2023). Ethics, automated processes, machine translation, and crises. In H. Moniz & C. Parra Escartín (Eds.), *Towards responsible machine translation: Ethical and legal considerations in machine translation* (pp. 135–156). Springer. https://doi.org/10.1007/978-3-031-14689-3_8
- Federico, M., Enyedi, R., Barra-Chicote, R., Giri, R., Isik, U., Krishnaswamy, A., & Sawaf, H. (2020). From speech-to-speech translation to automatic dubbing. *Proceedings of the 17th International Conference on Spoken Language Translation*, 257–264. <https://doi.org/10.18653/v1/2020.iwslt-1.31>
- Flawless (2022). *TrueSync*. <https://flawlessai.com/product>.
- Franco, E., Matamala, A., & Orero, P. (2013). *Voice-over translation: An overview*. Peter Lang. <https://doi.org/10.3726/978-3-0351-0630-5>
- Hayes, L. (2021). Netflix disrupting dubbing: English dubs and British accents. *Journal of Audiovisual Translation*, 4(1), 1–26. <https://doi.org/10.47476/JAT.V4I1.2021.148>
- Karakanta, A., Bhattacharya, S., Nayak, S., Baumann, T., Negri, M., & Turchi, M. (2020). The two shades of dubbing in neural machine translation. *Proceedings of the 28th International Conference on Computational Linguistics*, 4327–4333. <https://doi.org/10.18653/v1/2020.COLING-MAIN.382>
- Kings and Generals. (2022) *Evolution of the Indo-European Languages - Ancient Civilizations DOCUMENTARY* [Video]. YouTube. <https://www.youtube.com/watch?v=VpXgMdvLUXw>
- Kings and Generals. (n.d.) *Home* [Video]. YouTube. <https://www.youtube.com/c/KingsandGenerals/videos>
- Kottahachchi, B. & Abeyasinghe, S. (2022). *Overcoming the language barrier in videos with Aloud*. <https://blog.google/technology/area-120/aloud/>
- Lakew, S. M., Federico, M., Wang, Y., Hoang, C., Virkar, Y., Barra-Chicote, R., & Enyedi, R. (2021). Machine translation verbosity control for automatic dubbing. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings, 2021-June*, 7538–7542. <https://doi.org/10.48550/arxiv.2110.03847>
- laSexta (2022, July 2). *La 'magia' del deepfake: Un experto en efectos especiales desvela cómo convierten a Dani Mateo y Wyoming en 'Aznarito y Felipón'*. https://www.lasexta.com/programas/el-intermedio/experto-efectos-especiales-desvela-como-convierten-dani-mateo-wyoming-aznarito-felipon_2022070262be041203d1fb0001ef38b7.html
- lucyhernandez7672. (2021). *Re: Síntesis de proteínas*. [Video]. YouTube. <https://www.youtube.com/watch?v=6VlKfSvB02w>
- Macrea, C-I., Arias-Badia, B., & Torner Castells, S. (2023). El inglés biosanitario en YouTube: Recepción del uso de series y subtulado en vídeos de aprendizaje. *Panace@*, 23(56), 69–83.
- Martín-Mor, A., & Sánchez-Gijón, P. (2016). Machine translation and audiovisual products: A case study. *The Journal of Specialised Translation*, 26, 172–186. https://www.jostrans.org/issue26/art_martin.pdf
- Matamala, A. (2015). The ALST project: Technologies for audiovisual translation. *Proceedings of the 37th Conference Translation and the Computer-ASLING 2015*, 79–89. <https://aclanthology.org/2015.tc-1.11.pdf>
- Mejías-Climent, L. (2021). *Enhancing video game localization through dubbing*. Springer International. <https://doi.org/10.1007/978-3-030-88292-1>
- montsezaragoza9132. (2021). *Re: Síntesis de proteínas*. [Video]. YouTube. <https://www.youtube.com/watch?v=6VlKfSvB02w>

Baños, Rocío (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 61–79.

Öktem, A., Farrús, M., & Bonafonte, A. (2019). Prosodic phrase alignment for machine dubbing. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2019*, 4215–4219. <https://doi.org/10.21437/INTERSPEECH.2019-1621>

Pavesi, M., & Zamora, P. (2021). The reception of swearing in film dubbing: A cross-cultural case study. *Perspectives*, 30(3), 382–398. <https://doi.org/10.1080/0907676X.2021.1913199>

Saboo, A., & Baumann, T. (2019). Integration of dubbing constraints into machine translation. *Proceedings of the 4th Conference on Machine Translation*, 1, 94–101. <https://doi.org/10.18653/v1/W19-5210>

Slator. (2021). *Slator 2021 video localization report*. <https://slator.com/slator-2021-video-localization-report/>

Spiteri Miggiani, G. (2021). English-language dubbing: Challenges and quality standards of an emerging localisation trend. *The Journal of Specialised Translation*, 36a, 2–25. https://www.jos-trans.org/issue36/art_spiteri.pdf

Spiteri Miggiani, G. (2022). Measuring quality in translation for dubbing: A quality assessment model proposal for trainers and stakeholders. *XLinguae*, 15(2), 85–102. <https://doi.org/10.18355/XL.2022.15.02.07>

Terantino, J. (2011). Emerging technologies. YouTube for foreign languages: You have to see this video. *Language Learning & Technology*, 15(1), 10–16. <http://dx.doi.org/10125/44231>

ⁱ The Centre for Translation Studies at University College London offers the online intensive course “Translation and Adaptation for Dubbing”, sponsored by Netflix; and ZOO Digital has also partnered with the University of Sheffield to launch the short course “Adapting Dialogue for Dubbing”.

ⁱⁱ Dubbing and voiceover are the two most widely used interlingual revoicing modes. In dubbing, the original dialogue is completely replaced with a recording of the translated dialogue, whereas in voiceover the original and the translated tracks of dialogue are presented simultaneously, with the volume of the former lowered to avoid confusion.

ⁱⁱⁱ In Spanish, ALST stands for *Accesibilidad Lingüística y Sensorial: Tecnologías para la audiodescripción y las voces superpuestas* (Linguistic and Sensorial Accessibility: Technologies for audio description and voiceover).

^{iv} The examples have been reproduced as they appear on the YouTube channel, without any corrections. Being written by non-English native speakers, they contain grammatical mistakes; however, these do not impede understanding and have been maintained to preserve their authenticity.