

YouTube's automatic subtitles in the ESL/ESP classroom

ILARIA PARINI

Università del Piemonte Orientale

Received 17 September 2024; accepted 21 February 2025

ABSTRACT

EN YouTube's technology allows users to automatically obtain intralingual subtitles for the videos that are uploaded on the platform. As YouTube is owned by Google, it uses the Google Automatic Speech Recognition system (Google Voice) to synchronize the subtitles with the speakers' utterances. In order to go one step further and produce automatic interlingual subtitles within a wide range of foreign languages, Google combines its ASR system with its Machine Translation system (Google Translate). However, despite ongoing improvements in technology, automatic captioning can fail to convey the message accurately. This can cause a series of cascading errors, leading the Machine Translation system to mistranslate or to hinder overall accuracy. This paper presents the results of a study conducted with a group of students at the University of Turin on the topic of marketing, to focus on the potential of using YouTube's automatic subtitles to enhance their listening skills and use of specialized vocabulary.

Key words: AUDIOVISUAL TRANSLATION, FOREIGN LANGUAGE EDUCATION, ESP, MARKETING, AUTOMATIC SUBTITLING

ES La tecnología de YouTube permite a los usuarios obtener automáticamente subtítulos intralingüísticos para los videos subidos a la plataforma. Como YouTube es propiedad de Google, utiliza el sistema de reconocimiento automático de voz de Google (Google Voice) para sincronizar los subtítulos con las intervenciones de los hablantes. Para ir un paso más allá y generar subtítulos automáticos interlingüísticos en una amplia gama de idiomas extranjeros, Google combina su sistema de reconocimiento de voz con su sistema de traducción automática (Google Translate). Sin embargo, a pesar de los avances tecnológicos, los subtítulos automáticos pueden no transmitir el mensaje con precisión. Esto puede provocar una serie de errores en cascada, lo que lleva al sistema de traducción automática a traducir incorrectamente o afectar la precisión general. Este trabajo presenta los resultados de un estudio realizado con un grupo de estudiantes de la Universidad de Turín sobre el tema del marketing, centrado en el potencial del uso de los subtítulos automáticos de YouTube para mejorar sus habilidades de escucha y el uso de vocabulario especializado.

Palabras clave: TRADUCCIÓN AUDIOVISUAL, ENSEÑANZA DE LENGUAS EXTRANJERAS, ESP, MARKETING, SUBTITULACIÓN AUTOMÁTICA

IT La tecnologia di YouTube consente agli utenti di ottenere automaticamente sottotitoli intralinguistici per i video caricati sulla piattaforma. Poiché YouTube è di proprietà di Google, utilizza il sistema di riconoscimento vocale automatico di Google (Google Voice) per sincronizzare i sottotitoli con gli interventi dei parlanti. Per fare un ulteriore passo avanti e generare sottotitoli automatici interlinguistici in un'ampia gamma di lingue straniere, Google combina il suo sistema di riconoscimento vocale con il sistema di traduzione automatica (Google Translate). Tuttavia, nonostante i progressi tecnologici, i sottotitoli automatici non sono sempre in grado di trasmettere il messaggio con precisione. Questo può causare una serie di errori a cascata, portando il sistema di traduzione automatica a tradurre in modo errato o a compromettere l'accuratezza complessiva. Questo saggio presenta i risultati di uno studio condotto con un gruppo di studenti dell'Università di Torino sul tema del marketing, e si focalizza sul potenziale dell'uso dei sottotitoli automatici di YouTube per migliorare le competenze di ascolto e l'uso del vocabolario specialistico.

Parole chiave: TRADUZIONE AUDIOVISIVA, INSEGNAMENTO DELLA LINGUA STRANIERA, ESP, MARKETING, SOTTOTITOLAZIONE AUTOMATICA

1. Introduction

This paper intends to contribute to the field of foreign language teaching and learning through the use of Audiovisual Translation. More specifically, it focuses on the use of automatic subtitles in the ESL/ESP classroom in order to enhance students' listening skills and acquisition of specialized terminology. Indeed, in the course of the last few decades, Audiovisual translation (AVT) has emerged as one of the most dynamic and rapidly expanding domains within the discipline of Translation Studies. This does not really come as a surprise if we consider how saturated our society is with moving images on the screen and how the latter influence our reception and consumption of information, working and social relationships, education, and entertainment. Due to the current proliferation of audiovisual texts and products, today's language learners are undoubtedly proficient with a wide range of digital media. This exposure, and the extremely important role that social media and digital interactions play in our lives, results in language learners' proficiency in the multimodal nature of audiovisual material, which employs various codes to construct meaning.

It has long been acknowledged that audiovisual works can have multiple educational benefits. As stressed by Díaz Cintas and Fernández Cruz (2008, 201–202), "Video permits students to see and appreciate how native speakers interact in everyday conversation, providing them with linguistic cues (regional accents, registers, and grammatical and syntactical structures), as well as paralinguistic cues (body language, gestures), allowing them to see language in use in a cultural context". Consequently, the integration of videos has become increasingly common in foreign language classrooms, especially with the advent of digitization and the development of technology.

The availability of audiovisual content has further increased due to the growth of online video-sharing and social media platforms, whether they are free, like YouTube or Vimeo, or on payment, like Prime, Disney+ or Netflix. These sites give users the option to watch their favourite films, TV shows, or documentaries in their original languages and in their translated versions, whether dubbed or subtitled. This offers both teachers and students the possibility to use this extensive collection of content for teaching and learning purposes (see Adams & Díaz Cintas, 2022, p. 10).

Despite its undeniable cross-cultural significance, the role of translation in foreign language classrooms has historically been ambiguous. This ambiguity can be largely attributed to the dominance of educational paradigms like the communicative approach, which emphasizes the exclusive use of the target language in the classroom. Consequently, translation was banned from many foreign language teaching and learning contexts for a long time (Adams & Díaz Cintas, 2022, p. 11). However, the value of translation in the foreign language curriculum has been increasingly recognized over the past few decades. Recent literature on language teaching indicates a resurgence in the use of translation activities in classrooms. This shift can be largely attributed to the implementation of the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001), which not only emphasizes the acquisition of the four foundational language skills—reading, writing, speaking, and listening—but also highlights the practical application of these skills in a multilingual world. The CEFR stresses the importance of learners' ability to mediate between speakers of different languages who cannot communicate directly, especially through activities of interpretation and translation.

This emphasis on mediation is further supported by a report from the European Commission (2013), which presupposes that if language learning enables learners to perform mediation activities, then explicit teaching of translation and interpretation could enhance these abilities alongside other areas of language use. The report suggests that translation activities can foster a deeper understanding of both source and target languages, as well as their cultural contexts (Adams & Díaz Cintas, 2022, p. 11).

The return to translation in language education has contributed to a wider recognition of its benefits. Undoubtedly, translation can foster a more comprehensive understanding of linguistic structures and cultural nuances, and consequently, improve learners' overall language competence (Cook, 2010). Moreover, it allows learners to draw connections between their native language and the foreign language, thus promoting bilingual and bicultural proficiency (Carreres, 2006). More recent studies also highlight the cognitive benefits of translation in language learning. In fact, it encourages analytical thinking and problem-solving, as learners must handle linguistic and cultural differences to convey meaning accurately (Laviosa, 2014). Furthermore, translation exercises can improve attention to detail and precision in language use as students engage with different aspects of grammar, vocabulary, and syntax.

Indeed, the evolving role of translation in foreign language education reflects a shift towards more integrated and practical approaches to language learning. By including translation in the curriculum, educators can provide learners with valuable skills that are increasingly relevant in today's interconnected world. This

approach not only enriches linguistic competence but also prepares learners to effectively communicate across cultural and linguistic boundaries.

Translation, including Audiovisual Translation (AVT), has become increasingly popular in second language acquisition since the turn of the century. The emergence of didactic AVT (DAT) as a distinct field, as noted by Talaván (2020), has demonstrated the benefits of subtitling and dubbing audiovisual texts in developing traditional language skills and promoting intercultural competence. These practices, in fact, conform to the updated versions of the CEFR (Council of Europe, 2018, 2020), which emphasize intercultural competency. As a consequence, the use of AVT in foreign language education has become a significant area of research, marked by experimental studies which explore the impact of various AVT practices like subtitling, dubbing, and audio description in language classrooms. Such activities not only motivate students but also enhance their linguistic and intercultural competences. Indeed, students' exposure to digital technologies and various media formats during AVT activities contributes to the development of transferable skills and reinforces the educational value of incorporating AVT practices in language learning environments (Adam & Díaz Cintas 2022).

One of the main reasons to introduce audiovisual texts into educational settings is to provide learners with examples of oral communication in realistic contexts. This was traditionally achieved through passive observation, in contrast, recent research and practice (Sokoli & Zabalbeascoa Terrán, 2019) have shifted towards engaging learners actively through action-oriented tasks. Educators and researchers who implement active audiovisual translation tasks have observed that this not only promotes receptive and productive language skills but also improves critical thinking, pragmatic and intercultural awareness in both the first (L1) and second languages (L2). Furthermore, such tasks develop learners' ability to extract and infer information from multisemiotic texts combining multiple modes of communication, such as visual and auditory elements (Bolaños-García-Escribano & Navarrete 2022, p. 104).

The adoption of AVT tasks that involve active engagement, such as dubbing and subtitling video clips, has become rather popular in recent years. This approach has become particularly influential in foreign language education (FLE) scholarship over the past few decades (Incalcaterra McLoughlin, Lertola, & Talaván, 2018, 2020), with projects such as SubLanLearn (2009–2012), LeVis (2006–2008), ClipFlair (funded by the European Commission from 2011 to 2014), PluriTAV (2016–2019), Babelium (2013–2015), and, finally, the TRADILEX project (2019 - ongoing). All these projects have significantly contributed to various lines of enquiry and experimental research within the field and represent a remarkable step forward in the use of audiovisual translation as a didactic tool in foreign language education. By promoting action-oriented tasks and emphasizing mediation skills, they aim to enhance language competence and prepare students for effective communication in a globalized world (Lertola, 2019; Navarrete & Bolaños García-Escribano 2022; Talaván, 2020).

2. Subtitling in the foreign language classroom

In recent decades, the use of subtitles as a tool to improve foreign language learning (FLL) has received considerable attention from researchers all over the world. Europe, indeed, has emerged as a prominent centre for investigating the efficacy of subtitles in foreign language education, mostly due to the proactive initiatives led by the European Commission aimed at promoting multilingualism (Gambier, 2008, p. 12). Early research conducted by scholars such as Vanderplank (1988, 1990), Danan (1992), and d'Ydewalle and Van de Poel (1999) provided evidence that subtitled videos can significantly help language learning. These studies demonstrated that the integration of auditory and visual elements can effectively "facilitate formal and incidental language learning" (Danan, 2015, p. 41). Vanderplank (1988) had already stressed the advantages of using subtitles by showing how they help learners associate spoken language with its written form, thus improving both listening and reading skills. His subsequent research (1990) further emphasized the practical applications of subtitled texts in reinforcing language learning. Moreover, Danan (1992) investigated how subtitles improve vocabulary acquisition by providing contextual clues that help in understanding and learning new words. Similarly, d'Ydewalle and Van de Poel (1999) highlighted the role that subtitles play in conveying cultural differences and social contexts, which are crucial for developing pragmatic and intercultural competence.

The European Commission's efforts to promote multilingualism have significantly contributed to the widespread adoption of subtitles in language education. By supporting projects that integrate subtitles into FLL, the Commission has established new best practices and fostered innovative approaches within this field (Council of Europe, 2001). Research, indeed, shows that subtitles can substantially improve language learning

by providing learners with visual and auditory stimuli that reinforce vocabulary acquisition, listening comprehension, and overall language proficiency (Vanderplank, 2016). Subtitles, in fact, offer contextualized language exposure, enabling learners to see and hear words as they are used in authentic settings, which is crucial for developing a deeper understanding of the target language.

Contemporary studies continue to explore and confirm the advantages of using subtitles in language learning. For example, Pujadas and Muñoz (2019) found that subtitled videos significantly improve listening comprehension and vocabulary acquisition among language learners. Moreover, subtitles also serve as an effective bridge between spoken and written language, helping learners to develop the skills which are necessary to decode and understand linguistic structures (Díaz Cintas & Remael, 2007).

The European Commission's initiatives have also highlighted the role of subtitles in promoting cultural awareness and intercultural competence. By exposing learners to diverse linguistic and cultural contexts, subtitles help them develop a distinct understanding of different cultures and communication styles (Kruger, Szarkowska, & Krejtz, 2015). This cultural dimension is particularly important in a globalized world, where effective communication across languages and cultures is increasingly essential. In addition, the integration of subtitles into digital learning platforms and online resources has made language education more accessible and adaptable to different learning needs (Kruger, Hefer, & Matthew, 2013).

In general, it is possible to claim that the role of subtitles in foreign language learning has evolved from a niche interest to a widely recognized and researched pedagogical tool. As research and practice continue to evolve, the role of subtitles is likely to expand, offering innovative solutions for language educators and learners.

3. Technological advancements in subtitling

Subtitling has become increasingly popular in recent years, both as a professional practice and as an activity undertaken by amateurs. This growth in importance can be attributed to the rapid advancement of technology, which has helped the development of automatic subtitling solutions. These innovations, which will be examined in detail in this paper, aim to provide fully automated subtitles that enable quick understanding and an overall comprehension of audiovisual products.

Audiovisual Machine Translation projects have evolved in distinct phases that correspond to advancements in Machine Translation technology: rule-based, statistical, and neural methodologies (Karakanta 2022: 92). Each phase marks significant progress in the pursuit of effective Audiovisual Machine Translation solutions. The initial phase of Audiovisual Machine Translation projects focused on Rule-Based Machine Translation (RBMT). These early systems lacked the flexibility and sophistication necessary to handle the complexities of audiovisual content, such as contextual subtleties and accurate timing alignment. The advent of Statistical Machine Translation marked the next significant phase. The SUMAT project (2011-2014) exemplifies this stage, where Statistical Machine Translation systems were trained using human-transcribed, compressed, and time-coded subtitles for eleven different language pairs. While Statistical Machine Translation technology actually brought some productivity gains, it still lacked the precision required for professional subtitling (Karakanta, 2022, p. 94). The main limitation of Statistical Machine Translation was its inability to consistently produce translations that were both contextually accurate and synchronized with audiovisual content. Subsequent projects, including EU-Bridge (2012-2015) and ALST (2013-2015), moved beyond pure translation to integrate Speech Recognition and Automatic Speech Recognition with Machine Translation. These initiatives enhanced the subtitling process by combining the advantages of Automatic Speech Recognition technologies with advanced Machine Translation systems. The goal was to automate the transcription of spoken content and then translate it, streamlining the entire workflow from speech to subtitled text. This integration represents a more holistic approach to Audiovisual Machine Translation, addressing both the transcription and translation phases within a unified system. The most recent phase involves the adoption of Neural Machine Translation, which offers significant improvements over previous Machine Translation technologies. In fact, Neural Machine Translation systems use deep learning techniques to produce more accurate and contextually appropriate translations.

Juang and Rabiner (2005, p. 2) assert that the ambition to automate Speech Recognition has been a technological goal for a long time. In fact, Speech Recognition technology started to be popular in the mid-2000s, with a peak in popularity around 2006. During this period, the respeaking technique emerged as a prevalent method and effectively replaced manual stenotyping. Respeaking involves a trained professional, known as a "respeaker," who listens to the source audio and clearly repeats it into a Speech Recognition engine that utilizes acoustic and language models, and subsequently transcribes the repeated speech into text. Since

the respeaking process depends on a human mediator to convert speech into intralingual subtitles, it is classified as a “speaker-dependent” Speech Recognition technology. The initial transcription produced through this method captures the core content of the spoken language but requires additional processing to refine the text. This refinement is essential to transform the raw transcription into precise and well-formed subtitles suitable for viewers. Therefore, the role of a post-editor is crucial in this phase. The post-editor undertakes several detailed tasks to ensure the subtitles are accurate and synchronized with the audio. This includes segmentation (where the text is divided into manageable units), alignment (ensuring the text matches the spoken dialogue), and timecoding (to synchronize the subtitles with the audio and visual elements on the screen). These perfectionist processes are necessary to meet the visual constraints of the medium and to provide a smooth viewing experience.

Unlike respeaking, Automatic Speech Recognition technologies operate without the need for a human professional. These technologies use an acoustic model, a lexicon model, and a language model, which collectively process and analyze the audio input received via a microphone. Because of this independence from a human mediator, Automatic Speech Recognition is categorized as a “speaker-independent speech recognition technology.” Automatic Speech Recognition technologies, described by Ciobanu and Secară (2019, p. 92) as “language-specific computer programs that convert spoken input into written text in the language of the original speech,” have significantly transformed our interaction with digital content. They have become integral to numerous applications, enhancing accessibility and user experience across various domains. In fact, they are now commonly embedded in everyday tools such as virtual assistants (e.g., Siri, Alexa), transcription services, and language learning applications (Parini forthcoming). The development of these technologies was mainly steered by the desire to minimize human involvement in the transcription process. By automating the conversion of speech to text, they have streamlined many processes, making them a crucial component in modern digital applications.

As far as translation is concerned, efforts to integrate Automatic Speech Recognition into interlingual translation processes actually started when this technology was in its early developmental stages (Aliprandi et al. 2014, p. 2) and the advancement of technology has had its positive outcomes. Indeed, the “technological turn” described by Jiménez-Crespo (2019, p. 26) has had a deep impact on the field of Translation Studies, with a particularly significant effect on the Audiovisual Translation sector. As global media consumption habits continue to evolve and the demand for localized content increases, translation tools have had to undergo constant upgrading to keep pace (Burchardt, 2016, p. 206). As a consequence of this dynamic environment, Machine Translation has become an integral component of the subtitling process. Traditional MT systems are being replaced by avant-garde technologies such as Neural Machine Translation and Speech Translation. These innovations aim to simplify and accelerate the complex, traditionally human-driven workflows involved in subtitling, thus promoting the development of fully automatic subtitling solutions. In fact, the integration of Neural Machine Translation and Speech Translation technologies into subtitling not only increases efficiency but also ensures that high volumes of audiovisual content can be localized quickly to meet the increasing demand.

However, despite these advancements, the role of human subtitlers remains indispensable. Automatic subtitling solutions, although efficient, do not always grasp the nuances and subtleties of language, cultural references, or humour. Indeed, professional subtitlers bring a level of expertise and sensitivity to these aspects, ensuring that the subtitles not only convey the literal meaning but also the intended tone and context, which is something that often cannot be achieved through automatic subtitles. In fact, even if fully automatic subtitling systems offer promising capabilities in terms of speed and efficiency, the role of human supervision is essential to maintain the quality and cultural relevance of translations. This is why hybrid models that combine Neural Machine Translation and post-editing by human translators are being increasingly adopted. These models take advantage of machines’ rapid processing capabilities while ensuring that the final output is polished and culturally appropriate.

In brief, the “technological turn” has significantly transformed Translation Studies and the AVT sector, which was driven by evolving media consumption patterns and the increasing demand for localized content. The integration of advanced Machine Translation technologies into subtitling workflows marks a significant step forward in the field. However, as these technologies continue to develop, further academic research is essential to fully understand their impact and optimize their implementation in the translation industry (Jiménez-Crespo 2019, p. 26; Burchardt 2016; Karakanta 2020; Ferrara 2023, p. 32, Parini forthcoming). Research and development therefore aim at improving the accuracy and context-awareness of Machine Translation systems for subtitling. Innovations such as Neural Machine Translation and Speech Translation are

at the forefront of these efforts, with the purpose of enhancing the quality of automated subtitles (Karakanta 2020). Nevertheless, even if Neural Machine Translation has greatly improved the quality of translations, it is necessary to keep in mind that subtitling is a written form of spoken language and functions within the dynamic and diverse realm of entertainment, which includes, among others, various lexicons from fields like news, education, lifestyle, sports, law, and teenage slang (Bywood et al. 2017, p. 495). These distinguishing characteristics are responsible for the fact that technology is still not fully capable of handling the specific demands of subtitling and has to rely on extensive human intervention. Indeed, as will be seen in the following sections of this paper, automatic intralingual and interlingual subtitles are still flawed to various extents, and human intervention in a post-editing phase is still necessary if we want to obtain an accurate text.

4. Applications of automatic speech recognition technologies and machine translation: the case of YouTube

One notable application of Automatic Speech Recognition technologies is YouTube's auto-captioning system, introduced in 2009 and continuously refined since then. The evolution of YouTube as an open video platform reflects broader shifts in the media landscape and changes in content distribution methods. This transformation has significantly challenged traditional mass media, which relied on a linear, top-down model of content dissemination (Lee 2021, p. 2; Ferrara 2023, p. 54; Parini forthcoming). Historically, the global media market was divided into territories based on cultural and linguistic differences, where each country was associated with specific languages and media content.

However, digital and technological developments have radically changed the way audiences engage with media. These innovations have led to the rise of various online platforms that enable the immediate distribution of media content (Lee 2021, p. 1). This shift has moved away from the traditional top-down approach, allowing users to engage with and distribute content in a more interactive manner, and redefining the media landscape as a result.

YouTube's development occurred within a participatory culture, which emphasizes collective effort and voluntary user contributions (O'Hagan 2009, p. 97). This culture signifies a shift where consumers become active contributors and producers of content (Burgess & Green 2018, p. 12). These users, known as "prosumers," blur the distinction between content creators and audiences by engaging in both the production and consumption of audiovisual content.

The launch of YouTube in May 2005 marked a significant moment in digital media evolution. Founded by PayPal employees Chad Hurley, Steve Chen, and Jawed Karim, YouTube aimed to democratize video sharing by making it accessible to non-expert users (Burgess & Green 2018, p. 2). In fact, its user-friendly interface allowed anyone to upload and watch videos. The success of YouTube was accelerated when a Nike advertisement became the first video to reach one million views, highlighting the platform's potential for making content viral and fostering its appeal to advertisers (McFadden, 2023).

YouTube also introduced interactive and social features that facilitated user engagement and community building. This means that users could subscribe to channels and share videos via embedded links, thus promoting a networked environment of content creators and viewers (Burgess & Green, 2018, p. 2). The platform's early success exemplifies how technological innovation can transform media landscapes by enabling individuals to actively participate in digital culture. It has played a pioneering role in video-sharing platforms and it continues to influence global media consumption patterns.

Moreover, YouTube offers features that enhance user interaction. In fact, users can curate playlists, rate videos with a 5-star system, and view their personal watch history (McFadden, 2023). The platform also allows for direct interaction between viewers and creators through comments, which have been the primary mode of communication and engagement since its inception (Lee 2021, p. 4). These features foster a sense of community, allowing users to not only consume content but also engage in discussions, share feedback, and contribute to the evolving content landscape. The ability to rate videos and create playlists personalizes the viewing experience, while commenting simplifies the dialogue between creators and audiences, enriching the platform's social dimensions (Ferrara 2023, p. 55).

In October 2006, Google purchased YouTube, thus boosting its popularity. Since 2007, YouTube has launched several initiatives to enhance accessibility and support content creators. A significant development was the introduction of a captioning method to make web-based video content accessible to deaf and hard-of-hearing users (Smith, Crocker & Allman 2017, p. 115). During this period, YouTube also launched its partner programme, allowing content creators to monetize their videos through semi-transparent banner

advertisements. Through its collaboration with Google, YouTube raised its video quality standards and integrated licensed music features. These enhancements not only broadened YouTube's capabilities but also reinforced its position as a transformative force in the digital landscape by prioritizing accessibility, monetization opportunities, and content diversity (Ferrara 2023, p. 56).

YouTube has long been associated with globalization and "neo-Babelianism," highlighting the importance of mutual understanding among people of different languages (Cronin 2003, p.59; Ferrara 2023, p. 56). The platform's main aim has always been to ensure its content is comprehensible and accessible worldwide, an objective which was helped by the integration of automated technologies. This is why YouTube integrated Automatic Speech Recognition between 2008 and 2009 to provide automatic intralingual subtitles and improve accessibility.

Auto-captioning was initially limited to English-language videos (Tokusei, 2010). Later, Machine Translation was implemented on the platform, with the purpose of providing interlingual subtitles in various languages. Therefore, Google integrated its Automatic Speech Recognition system, Google Voice with its Machine Translation system, Google Translate, which was a Statistical Machine Translation system at the time (Turovsky, 2016). This integration marked a significant step as it allowed users to obtain automatic intralingual and interlingual subtitles, supporting YouTube's mission to improve global accessibility and inclusivity through technological innovation by overcoming language barriers.

By 2011, YouTube had enabled access to intralingual auto-captions for videos, providing nearly 40 million videos with captions that were automatically generated from their audio content (Ellis 2011). After the integration of automatic translation, it became able to produce subtitles in over 50 languages, which created videos that were comprehensible to audiences from different linguistic backgrounds (Ferrara, 2023, p. 57; Parini, forthcoming).

However, in spite of these technological advancements, it is not possible to claim that automatic subtitles, whether intralingual or interlingual, manage to convey the message accurately. In fact, even though Harrenstien (2010) reported that speech recognition algorithms had achieved a 20% reduction in error rates by 2010, and Varga (2021, p. 38) highlighted remarkable improvements in 2013, when YouTube's automatic subtitling integrated a speech recognition system based on deep neural networks, the quality of subtitles produced by YouTube is still far from being high. This fact has led to widespread dissatisfaction among users and content creators who recommended manual captioning, believing it would produce more accurate and reliable subtitles compared to those generated automatically.

Researchers and companies, including Google, have recognized the limitations of automatic captioning (Smith, Crocker & Allman, 2017, p. 117). Automatic Speech Recognition errors during auto-captioning often result in inaccuracies or misspellings, thus affecting the quality of both the captions and translated subtitles. Google Translate, which currently supports 133 languages (Caswell, 2022), relies heavily on accurate initial captions in order to produce reliable translations. Consequently, flaws in intralingual caption quality have a direct impact on the accuracy of translated content, which in turn affects comprehension across different linguistic audiences. In short, even if fully automatic subtitling solutions improve the global accessibility and dissemination of multimedia content, it often fails to provide viewers with accurate understanding.

Effective subtitle translation necessitates both linguistic accuracy and synchronization with the audiovisual context, ensuring that subtitles appear and disappear in harmony with the spoken dialogue. The absence of accurate time-coding can result in subtitles that are out of sync with the audio, which affects viewer comprehension. In fact, Google Translate faces difficulties in processing segmented strings of completed sentences, which is essential for generating accurate interlingual subtitles. Moreover, inadequate punctuation in intralingual subtitles often results in translated sentences that are syntactically inaccurate, leading to errors and compromising clarity. In order to maintain meaning and readability, it is necessary to ensure that Google Translate handle text segments that are structured and punctuated according to subtitle formatting.

Furthermore, Automatic Speech Recognition technologies face significant linguistic challenges, including disambiguating homophones, recognizing named entities (people, institutions, brands), and identifying technical or specialized language (Bolaños García-Escribano and Díaz-Cintas, 2023; Ciobanu and Secară, 2019, p. 98; Freitas and Liu, 2017; Karakanta, Negri & Turki, 2020, p. 63; Karakanta, 2022; O'Brien, 2022; Volk, 2008), as well as dealing with variations in accents, pronunciations, and common speech patterns like pauses, hesitations and repetitions, which are typical of spoken language. This can cause a series of cascading errors, leading the Machine Translation system to mistranslate or to affect overall accuracy.

The final part of this paper will present results from a study conducted with a group of students at the University of Turin, Italy, to illustrate how these new technologies in the field of Audiovisual Translation can be a useful instrument in teaching and learning marketing discourse for foreign language learners.

5. A practice-based study

This last part of the paper will present the results of a practice-based study involving 60 Italian students enrolled in the 2nd year of the MA degree programme in Foreign Languages for International Communication of the University of Turin, Italy. As reported on the webpage of the degree programme: “the degree course aims to build professional figures who interact with companies and international organizations in intercultural situations: linguistic and cultural mediators (communication area) in Italian and foreign companies, in public administration, in local authorities; management, responsibility or consultancy functions in international organizations, and in international cooperation projects¹.” The topic of marketing discourse, therefore, is undoubtedly of deep interest for the students. The study took place within the context of the course in English Language and Linguistics during Academic Year 2023-2024. The course is composed of two modules, taught by other professors. The former focuses on Corpus Linguistics, whereas the latter deals with Machine Translation. The last three lessons (2 hours each) of the second module – taught by the author – were dedicated to audiovisual translation and specifically automatic subtitling. It is important to stress that the three lessons took place on three subsequent days during the same week, so the students did not have much time to assimilate and process the information required for the study. These three lessons were aimed at showing the students the differences between translating and post-editing a written text with the aid of Machine Translation (which was the focus of the second module, which lasted 24 hours altogether) and post-editing and translating the subtitles of a spoken text while improving their listening skills and their competence in specialized terminology.

In fact, according to Karakanta (2020), the application of Machine Translation in subtitling closely mirrors its use in conventional text translation. Machine Translation systems can process subtitle content either as individual subtitle lines or as consolidated blocks of text that encompass several subtitle lines. The distinction lies in how the content is prepared and presented to the Machine Translation system. When Machine Translation is used with a time-coded transcription, it benefits from the preparatory work of professionals who have segmented the text into appropriate units and provided precise timing cues. This pre-segmentation is crucial, as it ensures that the Machine Translation system can accurately align the translated text with the corresponding audiovisual content. The timing cues help in maintaining synchronization between the subtitles and the video, which is a fundamental requirement for effective subtitling.

However, in the absence of such preparatory work, as in the case of YouTube's intralingual subtitles, Machine Translation systems process the text similarly to any other translation task. This means that the system does not inherently account for the specific requirements of subtitling, such as the necessity to match the duration and pace of the spoken dialogue with the on-screen text. Consequently, without professional input to segment and time-code the text, the Machine Translation output often does not meet the distinct demands of subtitling, leading to issues in readability and synchronization. Moreover, as will be seen in further detail, the fact that the Machine Translation system operates starting from a text which was originally spoken, gives rise to further difficulties.

5.1. Methodology

The students had no training in Audiovisual Translation and professional subtitling, so the first lesson was dedicated to the introduction of the main features and the main constraints which should ideally be kept in consideration when producing professional subtitles, such as the length of each subtitle, its duration on the screen, and other rules related to line breaks (making reference to the Netflix guidelines, which can be retrieved online)². They were then introduced to the main features of both intralingual and interlingual subtitles in the specific case of YouTube. Finally, we watched a video, which is a TED talk entitled “12 secrets of marketing and you won't believe what happens next”, which lasts 13.02 minutes and can be retrieved at the following link: <https://www.youtube.com/watch?v=3gEOP8Th478>. TED talks are delivered live to an audience during conferences organized by the TED organization. A diverse group of speakers presents their ideas and

¹ https://www.lingue.unito.it/do/home.pl/View?doc=/i_corsi/profili_professionali.html&tab=6 (translated from Italian)

² <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215758617-Timed-Text-Style-Guide-General-Requirements> (accessed November 2024)

experiences on various topics to educate both the live audience and viewers at home. In fact, these talks are not confined to the event location, as they are also recorded and made available online. Specifically, they are published on the official website (www.TED.com) and platforms like Facebook and, most importantly for the study presented in this paper, YouTube. The Internet and the shift towards online videos have enabled widespread distribution, resulting in TED talks gaining increasing attention and interest from global viewers and listeners. Thus, TED fulfils its mission of sharing knowledge and connecting people through its mantra, namely, “ideas worth spreading.” The talk, released in 2015, is delivered by Naimul Huq, an alumnus of the Morehead-Cain Scholars Program at UNC, who is currently working as a Social Media Associate at WCG in Austin. In the video, he focuses on bettering the tools available to companies, specifically health-care companies, and making them more user-friendly in order to encourage customer interaction. The videoclip was divided into 4 segments (each one was about 3-minute long). The students were divided into groups of 4-5 people. Each group was instructed to work on a specific segment (more groups could work on the same segment).

During the second lesson, we first watched the video together in class with the intralingual subtitles activated (performed by Google Voice). Secondly, students were provided with a template on Google Drive with the transcription of the intralingual subtitles (which can be downloaded through various online websites). At this point they were supposed to 1) identify the various mistakes, 2) classify them, 3) edit and correct the subtitles and provide their version on the template. This phase of the study particularly aimed at honing the students’ listening skills, as quite a few of the mistakes were due to the Speech Recognition system’s misunderstanding of the spoken output of the TED presenter.

During the third lesson, the students were asked to activate the Italian interlingual subtitles, download them, and 1) identify the mistakes, 2) classify them, 3) edit and correct the subtitles and provide their version on the template on Google Drive. Finally, the students were tasked with examining the root causes of the errors found in the interlingual subtitles and discussing them in class. Their goal was to corroborate the hypothesis according to which these mistakes predominantly stemmed from word misrecognitions by the Automatic Speech Recognition system. These initial misrecognitions were believed to trigger a chain reaction of subsequent errors (the so-called “cascading errors”, Soffritti, 2018, p. 337).

Besides the more general purpose of identifying and classifying the inaccuracies of both interlingual and intralingual mistakes (which can be found in any video uploaded on YouTube), which presupposes a proficient level of the students’ mastery of the language, and in particular their listening skills, the study aimed at focusing on specific terminology related to marketing discourse. In fact, the goal was also to enhance students’ awareness in relation to the topics presented, and to improve their familiarity with them and the related lexis.

5.2. Analysis of intralingual subtitles

The students identified four main problems in the intralingual subtitles: lack of adherence to the norms related to the formal requirements of subtitles, misrecognition of common words, misrecognition of entity names, and misrecognition of specialized terminology.

- 1) Formal requirements (lack of punctuation; timing; line breaks)

The most immediate and evident problem was the absence of punctuation in the intralingual subtitles. This omission meant that the subtitles were not structured as complete sentences. Additionally, the subtitles failed to meet general standards regarding duration, which stipulate a minimum of 1 second and a maximum of 6-7 seconds per subtitle, as well as line treatment principles, which guide the proper way to break text into two lines³. Essentially, these intralingual subtitles were mere transcriptions of every spoken word by the speaker, which were roughly but often inaccurately synchronized with his speech. Moreover, the lack of punctuation not only disrupts sentence structure but also affects the readability and comprehension of the

³ For example, according to the guidelines provided by Netflix, the line should be broken after punctuation marks, before conjunctions, and before prepositions, whereas the line break should not separate a noun from an article, a noun from an adjective, a first name from a last name, a verb from a subject pronoun, a prepositional verb from its preposition, or a verb from an auxiliary, reflexive pronoun or negation.

subtitles. Without clear sentence demarcations, viewers struggle to follow the intended meaning, leading to confusion and a diminished viewing experience. In addition, failing to adhere to duration guidelines results in subtitles that are either too brief to be read comfortably or too lengthy to keep pace with the spoken content. This inconsistency in timing disrupts the natural flow of watching and understanding the video. The improper handling of line breaks can also lead to awkward splits in the text, making it harder for viewers to read and comprehend the subtitles smoothly (Parini forthcoming).

Therefore, when required to post-edit the intralingual subtitles, the first thing that the students had to do was to reconstruct the sentences by adding punctuation in order to make them meaningful. The table below shows an example from the very incipit of the talk:

Table 1.

Transcription of intralingual subtitles showing lack of formal requirements and post-edited version

Intralingual subtitles	Post-edited subtitles
I'm going to reveal to you the 12 steps to winning at marketing and what happens next is that well there are only four steps create share some kind of magical process and profit in order to succeed in marketing you have to create something representative of yourself and your audience and distribute it in the right way in reality there is some assembly required the one idea I want to express today is that marketing is complex because it strives to understand human beings and human beings are perplexing but you have the tools to understand them you just need to look for the right things and there are 12 steps we'll start with the Voice	I'm going to reveal to you the 12 steps to winning at marketing, and what happens next is that, well, there are only four steps: create, share, some kind of magical process, and profit. In order to succeed in marketing, you have to create something representative of yourself and your audience, and distribute it in the right way. In reality, there is some assembly required. The one idea I want to express today is that marketing is complex, because it strives to understand human beings, and human beings are perplexing. But you have the tools to understand them, you just need to look for the right things. And there are 12 steps. We'll start with the voice.

As can be observed, although the post-edited subtitles do not follow the guidelines that are usually required in professional subtitling, as far as time and space constraints are concerned (we should not forget that these students have no real training in subtitling, and that this is not the focus of the study), punctuation was added so the that sentences could provide a complete sense of what the presenter is actually saying, even in a written form.

2) Misrecognition of common words

The intralingual subtitles contain instances of misrecognition of words by Google Voice. Some examples are reported in Table 2 below:

Table 2

Transcription of intralingual subtitles containing examples of misrecognition of common words and post-edited version

Intralingual subtitles	Post-edited subtitles
Red Bull has a loud voice but it's not just loud it's one of the strongest brand voices on Earth not just among beverages or pickme UPS their promise is Wings an alternate reality where you can freef fall from the edge of space at the speed of sound or perfect that kickflip she actually drew it on a napkin in less	Red Bull has a loud voice, but it's not just loud, it's one of the strongest brand voices on Earth. Not just among beverages or pick-me-ups , their promise is "wings": an alternate reality where you can free fall from the edge of space at the speed of sound or perfect that kickflip. she actually drew it on a napkin in less than a minute

<p>than a minute and then she spent a whole year with Bor justifying her intuition they made a laptop for young people it's a tablet and a computer and one social conversation is volunteered on the other hand it's a public Declaration of our beliefs we check into our favorite haunts we spit vitrio about cable companies</p>	<p>and then she spent a whole year with boards justifying her intuition They made a laptop for young people: it's a tablet and a computer in one. Social conversation is volunteered. On the other hand, it's a public declaration of our beliefs. We check into our favorite haunts. We spit vitriol about cable companies.</p>
---	---

Overall, Google Voice rather accurately transcribed the words uttered by the presenter. No real cases of misrecognition of homophones were found, although some words were misunderstood and mistranscribed, as will also be seen in the following sections.

3) Misrecognition of entity names

As previously mentioned, proper names, institutions, and brands can be difficult for Automatic Speech Recognition technologies to recognize and accurately transcribe. In the videoclip that the students analysed in class, this is probably the category which presented the highest number of errors, which are reported in Table 3.

Table 3
Transcription of intralingual subtitles containing examples of misrecognition of entity names and post-edited version

Intralingual subtitles	Post-edited subtitles
<p>the field of visual identity was pioneered by Gordon lipincott liping cot introduced Evolution Paul aare is one of the most prolific and active designers in the world when polisher was tasked with designing the logo of City Bank she actually drew it on a napkin in less than a minute many of you will recognize this building the Morhead planetarium houses the offices of the oldest merit scholarship in the country the Morhead Kane's audience is adaptive and The Branding needed to be adaptable it needed to be built on the latest Technologies the website needed to be current responsive and flexible marketing used to be driven by focus groups but like the elementary particles in Heisenberg's UNC C principal experiments when we're observed we change our behaviors fer got rid of the Viva Viagra ads when they realized that men were searching on Google about questions about Ed that were completely different from what they were saying on forums</p>	<p>the field of visual identity was pioneered by Gordon Lippincott Lippincott introduced evolution Paula Share is one of the most prolific and active designers in the world When Paula Share was tasked with designing the logo of Citi Bank she actually drew it on a napkin in less than a minute Many of you will recognize this building. The Morehead planetarium houses the offices of the oldest merit scholarship in the country the Morehead Cain's audience is adaptive and the branding needed to be adaptable, it needed to be built on the latest technologies, the website needed to be current responsive and flexible. marketing used to be driven by focus groups but like the elementary particles in Heisenberg's Uncertainty principle experiments when we're observed we change our behaviors. Pfizer got rid of the Viva Viagra ads when they realized that men were searching on Google about questions about ED that were completely different from what they were saying on forums.</p>

<p>they wanted to prove how powerful it was so they recruited Clams Casino to produce and Vic Mensa to MC this was the genius of the marketers over there they called it us today and they said how do you figure out what you're going to include in your wrapup</p> <p>Proctor and Gamble was going to scrap the yellow spice brand and my friend at widen Kennedy knew that the audience which was made up mostly of women who make the consumer decisions in the household needed to be talked to so they came up with this campaign and it had uh kind of Keo ree sort of character Surfer Dude speaking directly to the audience language itself is evolving the language the English that we use today is different from the English of baywolf</p>	<p>They wanted to prove how powerful it was So they recruited Clams Casino to produce and Vic Mensa to MC</p> <p>This was the genius of the marketers over there. They called USA Today and they said: "How do you figure out what you're going to include in your wrap-up?"</p> <p>Procter and Gamble was going to scrap the Old Spice Brand and my friend at Wieden-Kennedy knew that the audience, which was made up mostly of women, who make the consumer decisions in the household, needed to be talked to so they came up with this campaign and it had uh kind of Keanu Reeves sort of character Surfer Dude speaking directly to the audience Language itself is evolving. The English language that we use today is different from the English of Beowulf.</p>
---	--

It is interesting to observe that the misrecognition of entity names includes proper names of people whose activities were/are connected to the marketing world (Gordon Lippicott, Paula Share), actors (Keanu Reeves), musicians/singers (Vic Mensa), brands (Procter and Gamble, Old Spice, Pfizer), institutions (Morehead Cain), advertising agencies (Wieden-Kennedy), newspapers (USA Today), theories in physics (Heisenberg's Uncertainty principle), as well as Old English epic poems (Beowulf).

It is also worth noting that in the video of the TED talk the camera focuses on the presenter most of the time, mainly with close-up shots, but from time to time it is possible to see that there is a big screen behind him which projects some slides related to what the presenter is talking about. This sometimes turned out to be helpful for the students, as they could see some of the names (specifically, Paula Share's and Morehead Cain's) on the screen, and were therefore facilitated in the task of post-editing.

4) Misrecognition of specialized terminology

As already pointed out, one of the main objectives of this study was to focus on the specialized terminology related to the field of marketing. In this case, the students managed to identify quite a few words or phrases which had not been accurately caught by Google Voice. In some cases, the students were not familiar with the terminology: therefore the proposed exercise turned out to be useful not only to enhance their listening skills but also to improve their lexical competence, as well as their knowledge of topics belonging to a specialized semantic field. The table below reports some examples:

Table 4.

Transcription of intralingual subtitles containing examples of misrecognition of specialized terminology and post-edited version

Intralingual subtitles	Post-edited subtitles
if we're in a focused group of moms we're unlikely to admit to sharing junk food with our kids	if we're in a focus group of moms we're unlikely to admit to sharing junk food with our kids
she actually drew it on a napkin in less than a minute and then she spent a whole year with Bor justifying her intuition declining sales and Sheriff voice meant that they were losing out if you haven't seen a peso model in your marketing classes yet you will it sensor paid earn shared and own and it stands for your advertising which is scalable your external endorsements which are credible your your shared media property like social which makes you authentic and your website which is where you define Who You Are	she actually drew it on a napkin in less than a minute and then she spent a whole year with boards justifying her intuition declining sales and share of voice meant that they were losing out if you haven't seen a PESO model in your marketing classes yet you will. It stands for PAID, EARN, SHARED and OWN , and it stands for your advertising, which is scalable, your external endorsements, which are credible, your shared media property, like social, which makes you authentic, and your website, which is where you define who you are.
we worked with 1 adla on a campaign called two days beat for HP	we worked with one ADLA on a campaign called #2DAYSBEAT for HP

A "focus group" is defined by the Online Cambridge Dictionary as "a group of people brought together to discuss what they think about a particular product, advertisement, or subject, as part of a company's market research"⁴ and it is labelled as belonging to the specialized terminology related to the marketing semantic field.

A "board" (a polysemic word) may refer to various referents. With reference to the semantic field of business, the Online Cambridge Dictionary defines it as "the group of people who are responsible for controlling and organizing a company or organization"⁵

The "share of voice" is defined by the Online Cambridge Dictionary as "the degree to which people see a company's advertisements, etc. in a particular market, compared to those of competitors"⁶. This phrase is also labelled by the dictionary as belonging to the specialized terminology related to the marketing semantic field. As far as the PESO model is concerned, Kernez (2022) claims:

The PESO model in marketing is important because it combines various marketing techniques and activities to create and build a strong presence. Marketing using the PESO model optimizes your content in every sphere, from sharing your content to engaging your customers. The PESO model consists of four arenas: paid media, earned media, shared media and owned media.

In this case, it is worth noting that the students did not manage to report the verbs that contribute to the creation of the acronym in their correct form. In fact, they put the verbs "earn" and "own" in their base or present form, whereas the verbs should all be in the past participle form (paid, earned, shared, owned). This is even more interesting if we consider the fact that in that specific moment of the video the camera focuses on the slide on the screen behind the presenter, where the following image is projected (and the four verbs are clearly visible):

⁴ <https://dictionary.cambridge.org/dictionary/english/focus-group>

⁵ <https://dictionary.cambridge.org/dictionary/english/board>

⁶ <https://dictionary.cambridge.org/dictionary/english/share-of-voice>

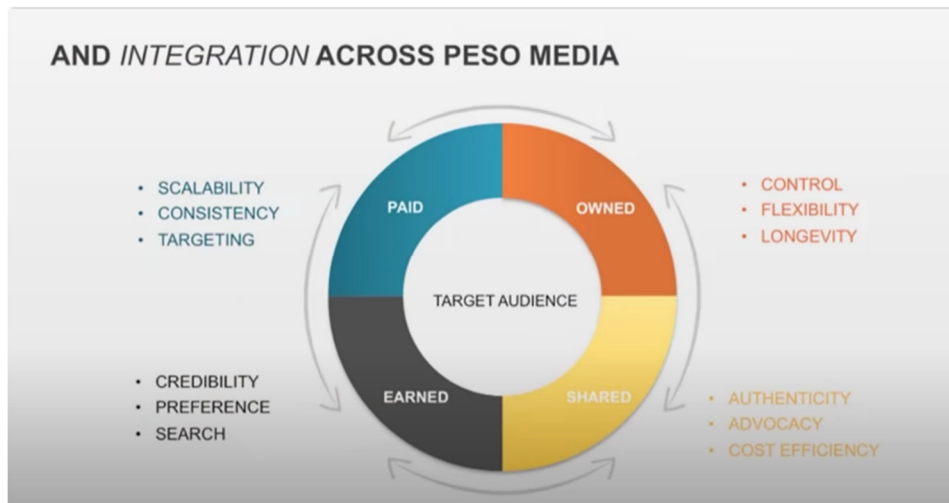


Figure 1. PESO model as represented on the screen behind the presenter.

Finally, the initialism ADLA (which pertains to the field of the specialized terminology used in marketing) stands for “Awareness, Demand, Lead, Acquisition.” This framework outlines the steps of a customer’s journey from first encountering a product or service to making a purchase. Below is an overview of each stage (Hussin & Hishan, 2022, p. 1538):

- 1) **Awareness:** The initial phase where potential customers become aware of the product or service. Marketing efforts focus on increasing visibility and recognition through advertising, social media, public relations, and content marketing.
- 2) **Demand:** Following awareness, the goal is to create interest and desire for the product or service. This involves educating potential customers about its benefits and value through targeted campaigns, informative content, and engaging promotions.
- 3) **Lead:** In this phase, interested potential customers (leads) are identified and captured, typically by collecting contact information via forms, sign-ups, or other lead generation tactics. The focus is on nurturing these leads to convert them into potential buyers.
- 4) **Acquisition:** The final stage is converting leads into paying customers. This involves closing the sale through effective sales strategies, personalized follow-ups, special offers, and ensuring a smooth purchasing process. The goal is to secure the customer’s commitment, leading to a completed transaction.

By using the ADLA framework, marketers can strategically guide potential customers through each stage of their journey, optimizing efforts to drive sales and business growth. Therefore, it is essential to stress how this exercise has managed to enhance the students’ knowledge of some topics related to their field of study.

5.3. Analysis of interlingual subtitles

As previously discussed, since interlingual subtitles are generated by Google Translate starting from the intralingual subtitles created by Google Voice, the inaccuracies noted earlier will inevitably affect the accuracy of the interlingual subtitles. This phenomenon is referred to as “cascading errors” (Soffritti, 2018, p. 337).

- 1) Formal requirements (lack of punctuation, timing, line breaks)

The intralingual subtitles of the video clip, as noted, lack punctuation, resulting in subtitles that are not formed as complete sentences. The intralingual subtitles are essentially just transcriptions of every word the speaker utters that are somewhat synchronized with his speech but not always accurate. This lack of structure in the intralingual subtitles impacts the interlingual subtitles significantly. The translation ends up being a

sequence of words rather than coherent sentences. Furthermore, the machine translation often struggles with grammatical and syntactic accuracy because it cannot determine the subject or object of the verbs. Consequently, the translated sentences frequently contain errors that render them incomprehensible. Here follows the automatic interlingual subtitles of the passage that was analysed (and post-edited) in the previous section:

Table 5.

Transcription of intralingual subtitles showing lack of formal requirements, the corresponding interlingual (Italian) version and their post-edited version

Intralingual subtitles	Interlingual subtitles	Post-edited subtitles
I'm going to reveal to you the 12 steps to winning at marketing and what happens next is that well there are only four steps create share some kind of magical process and profit in order to succeed in marketing you have to create something representative of yourself and your audience and distribute it in the right way in reality there is some assembly required the one idea I want to express today is that marketing is complex because it strives to understand human beings and human beings are perplexing but you have the tools to understand them you just need to look for the right things and there are 12 steps we'll start with the Voice	Ti svelerò i 12 passaggi per vincere nel marketing e quello che succede dopo è che beh, ci sono solo quattro passaggi per creare una sorta di processo magico e condividere un profitto. Per avere successo nel marketing devi creare qualcosa rappresentativo di te stesso e del tuo pubblico e distribuirlo nel modo giusto in realtà è necessario un po' di assemblaggio l'idea che voglio esprimere oggi è che il marketing è complesso perché si sforza di comprendere gli esseri umani e gli esseri umani lasciano perplessi ma tu hai gli strumenti per capirli devi solo cercare le cose giuste e ci sono 12 passaggi inizieremo con la Voce	Vi svelerò i 12 passaggi per avere successo nel marketing. E quello che succede dopo è che beh, ci sono solo quattro passaggi: creare, condividere, una sorta di passaggio magico e profitto. Per avere successo nel marketing, bisogna creare qualcosa che rappresenti te stesso e il tuo pubblico e distribuirlo nel modo migliore. In realtà, è richiesto una sorta di assemblaggio. L'idea che voglio esprimere oggi, è che il marketing è complesso perché si sforza di comprendere gli esseri umani, e gli esseri umani lasciano perplessi. Ma voi avete gli strumenti per comprenderli, dovete solo osservare le cose giuste. Ci sono 12 passaggi. Iniziamo con la voce.

Indeed, apart from the presence of a comma in line 3 and a full stop in line 5, punctuation is completely missing. The passage, reflecting the original source text, sounds rather like a sort of “stream of consciousness”. However, there are other oddities that can be noticed from a syntactic point of view. First of all, the speaker is talking directly to his entire audience (who are actually present in the room) using the second personal pronoun, which can be interpreted as a singular or as a plural pronoun in English. In the interlingual subtitles, Google Translate always rendered it with the Italian singular pronoun (“ti svelerò...”, “devi creare...”, “di te stesso e del tuo pubblico...”, “tu hai gli strumenti per capirli...”, “devi solo cercare le cose giuste...”, etc.), which, in this context, does not sound accurate, as he is speaking to a large number of people. Moreover, at times the lack of punctuation is a source of syntactic ambiguity. For example, in the sentence “Per avere successo nel marketing devi creare qualcosa rappresentativo di te stesso e del tuo pubblico e distribuirlo nel modo giusto”, the verb “distribuirlo” is syntactically incorrect within the context that is provided (the correct form would be “distribuirlo”, as it depends on the Italian modal verb “devi” and should consequently be in the infinitive form).

If we observe the post-edited version of the subtitles of the passage performed by one group of students, first of all we can observe that the form of the verbs in the second pronoun has often been rendered as a plural pronoun (“Vi svelerò i 12 passaggi...”, “voi avete gli strumenti...”, “dovete solo osservare...”). However, surprisingly, it was sometimes rendered with the singular form (“bisogna creare qualcosa che rappresenti te stesso e il tuo pubblico”), which turns out to be rather inconsistent.

2) Misrecognition of common words

The intralingual subtitles contain few instances of misrecognition of words by Google Voice. Some examples are reported in Table 6:

Table 6

Transcription of intralingual subtitles containing examples of misrecognition of common words, the corresponding interlingual (Italian) version and their post-edited version

Intralingual subtitles	Interlingual subtitles	Post-edited subtitles
Red Bull has a loud voice but it's not just loud it's one of the strongest brand voices on Earth not just among beverages or pickme UPS their promise is Wings an alternate reality where you can freef fall from the edge of space at the speed of sound or perfect that kickflip	Red Bull ha una voce forte ma non è solo forte è una delle voci di marchio più forti sulla Terra non solo tra le bevande o i pickme UPS la loro promessa è Wings , una realtà alternativa in cui puoi cadere in libertà dal limite dello spazio alla velocità del suono o con il perfetto kickflip .	Red Bull lascia un segno, ma non è tutto, è uno dei marchi pionieri in ambito globale, non solo tra le bevande o energizzanti . La loro promessa è un " paio di ali ", una realtà alternativa in cui puoi cadere liberamente dai confini dello spazio alla velocità del suono o eseguire perfettamente un trick .
they made a laptop for young people it's a tablet and a computer and one social conversation is volunteered on the other hand it's a public Declaration of our beliefs we check into our favorite haunts we spit vitrio about cable companies	per cui hanno realizzato un laptop i giovani sono un tablet, un computer e uno la conversazione sociale è volontaria d'altro canto è una dichiarazione pubblica delle nostre convinzioni controlliamo nei nostri luoghi preferiti sputiamo al vetriolo sulle società via cavo	È stato realizzato un computer portatile per i giovani che è un tablet e un computer e uno . la conversazione sociale è volontaria. D'altro canto è una dichiarazione pubblica delle nostre convinzioni. Ci ritroviamo nei nostri luoghi preferiti. Sputiamo vetriolo sulle società via cavo.

In the first excerpt, we can note that Google Translate left the phrase "pickme UPS" unaltered, as the Machine Translation probably associated UPS with the shipping company, which, combined with "pick me" probably seemed plausible to the software. However, the students who post-edited this chunk of speech correctly translated it into "energizzanti". A "pick-me-up", indeed, is a noun compound which is defined by the Online Collins Dictionary as "something, often food or drink, that gives you more energy or makes you feel better"⁷, and since the speaker is talking about Red Bull, the students could easily detect the mistake. In the same passage, Google Translate left the noun "Wings" untranslated. This noun obviously refers to Red Bull's famous claim "Red Bull gives you wings", which, however, in Italian sounds like "Ti mette le ali". Leaving the noun in English, therefore, does not achieve an effect of functional equivalence, as the Italian spectator is used to reading/hearing the word "ali" when talking about Red Bull, and not "wings". This is a clear example of a cascading error, where the misrecognition of the word by Google Voice has an impact on its translation performed by Google Translate. The Italian students who worked on this passage managed to grasp this nuance and translated the word as "un paio d'ali" (literally, "a pair of wings"). Interestingly, the misspelling of the word "free" as "freef" in the intralingual subtitles did not have any repercussion on the interlingual subtitles ("free" was translated as "cadere", which is the exact equivalent of "free"). However, "to free fall" is commonly translated as "lanciarsi in caduta libera" in Italian, whereas the interlingual subtitles rendered the phrase with the literal, but less fluent, phrase "cadere in libertà", and the Italian students post-edited the subtitles with the likewise uncommon phrase "cadere liberamente". Moreover, the students (mis)translated the word "kickflip" (which had been maintained in the interlingual subtitles, as it is an anglicism used in Italian in the field of skateboarding⁸) with the generic word "trick" (which is still an anglicism, but with a much more generic meaning).

⁷ <https://dictionary.cambridge.org/dictionary/english/pick-me-up>

⁸ <https://www.fillow.it/blog/2021/10/dizionario-dello-skateboard/>

In the second passage, the interlingual subtitles present more than one inaccuracy. First of all, the preposition “for”, which is part of the benefactive complement phrase “for young people”, was completely ignored, so that the phrase was translated into “un laptop i giovani” (instead of “per i giovani”) in the Italian interlingual subtitles. Secondly, the lack of punctuation, already observed in the previous sections, produces a completely meaningless sentence (the subject of the verb “be” is not the laptop about which the speaker is talking, but the young people for whom the tablet has been made): “i giovani **sono** un tablet, un computer e uno”). Finally, the misrecognition of the word “in” as “one”, which had already been pointed out in the previous section, caused a cascading error in the interlingual subtitles, as it was translated literally as “uno”. Once again, the final outcome is a meaningless sentence. It is worth noting that the Italian students who worked on the post-edited version of this passage did not identify this error and translated “and one” literally as “e uno”.

In the third passage, we can see that the word “vitrìol”, which was mistranscribed in the intralingual subtitles as “vitrio”, was correctly translated as “vetriolo” in the interlingual subtitles. However, Google Translate performed a syntactic mistake by adding the preposition “al” (“sputare al vetriolo”), whereas, from a syntactic point of view, the correct phrase in Italian would be “sputare vetriolo”, as “vetriolo” becomes the direct object of the verb “sputare”. This is exactly how the group of Italian students who worked on this passage translated the phrase. However, it ought to be noted that the most common collocation in Italian is “sputare veleno”, whereas “vetriolo” usually collocates with “al” (e.g. “commenti al vetriolo”), though not with the verb “sputare”. This is probably why Google Translate added the preposition “al”⁹.

3) Misrecognition of entity names

As previously mentioned, proper names, institutions, and brands can be difficult for Automatic Speech Recognition technologies to recognize and accurately transcribe. In the videoclip the students analysed in class, this is probably the category that presented the highest number of errors, which are reported in Table 7.

Table 7.

Transcription of intralingual subtitles containing examples of misrecognition of entity names, the corresponding interlingual (Italian) version and their post-edited version

Intralingual subtitles	Interlingual subtitles	Post-edited subtitles
the field of visual identity was pioneered by Gordon lipincott	il campo dell'identità visiva è stato aperto da Gordon lipincott	Il campo della visual identity è stato lanciato da Gordon Lippincott .
liping cot introduced Evolution	<i>Interlingual subtitles missing</i>	Lippincott ha portato ad un'evoluzione
Paul aare is one of the most prolific and active designers in the world	Paul Aare è uno dei designer più prolifici e attivi al mondo	Paula Share è una delle designer più prolifiche e attive al mondo
when polisher was tasked with designing the logo of City Bank she actually drew it on a napkin in less than a minute	quando la lucidatrice è stata incaricata di disegnare il logo della City Bank lo ha effettivamente disegnato su un tovagliolo in meno di un minuto	Quando Paula Share è stata incaricata di disegnare il logo della Citi Bank lo ha disegnato su un tovagliolo in meno di un minuto
many of you will recognize this building the Morhead planetarium houses the offices of the oldest merit scholarship in the country the Morhead Kane 's audience is adaptive and The Branding needed to be adaptable it needed to be built on the latest Technologies the website needed	molti di riconoscerai questo edificio, il planetario Morhead ospita gli uffici della più antica borsa di studio al merito del paese il pubblico di Morhead Kane è adattivo e il marchio doveva essere adattabile doveva essere costruito sulle tecnologie più recenti il sito web doveva essere reattivo e flessibile	Molti di voi riconosceranno Il Morhead , un planetario Che ospita gli uffici della più antica Borsa di studio per merito del paese. Il pubblico di Morehead Cane è flessibile e il marchio doveva essere versatile. Doveva essere costruito sulle tecnologie più recenti.

⁹ The Corpus Italian Web 2020 on Sketch Engine reports only 5 instances of concordances of the verb “sputare” followed by the noun “vetriolo”, whereas it reports 1,641 instances of concordances of the verb “sputare” followed by the noun “veleno”.

<p>to be current responsive and flexible</p>	<p>il marketing era guidato da focus group ma, come le particelle elementari negli esperimenti principali UNC C di Heisenberg, quando siamo osservati, cambiamo i nostri comportamenti di cui sbarazzarsi gli annunci di Viva Viagra quando si sono resi conto che gli uomini cercavano su Google domande su Ed che erano completamente diverse da ciò che dicevano sui forum volevano dimostrare quanto fosse potente, quindi hanno reclutato Clams Casino come produttore e Vic Mena come MC questo è stato il genio degli esperti di marketing laggiù, ci hanno chiamato oggi e hanno detto come fai a capire cosa includere nel tuo riepilogo Proctor and Gamble stava per eliminare il marchio delle spezie gialle e il mio amico di Widen Kennedy sapeva che il pubblico, composto principalmente da donne che prendono le decisioni di consumo in casa, aveva bisogno con cui parlare, quindi hanno ideato questa campagna e aveva una specie di personaggio di Keo Ree, Surfer Dude, che parlava direttamente al pubblico la lingua stessa sta evolvendo la lingua, l'inglese che usiamo oggi è diverso dall'inglese di le organizzazioni baywolf</p>	<p>Il sito web doveva essere attuale, reattivo e flessibile. Il marketing un tempo era guidato da gruppi di discussione, ma, come le particelle elementari negli esperimenti di Heisenberg sul principio di indeterminazione quando veniamo osservati cambiamo i nostri comportamenti La Pfizer ha eliminato la pubblicità del Viva Viagra quando si è resa conto che gli uomini cercavano su Google domande sulla DE (disfunzione erettile) che erano completamente diverse da quello che dicevano sui forum. volevano dimostrare quanto fosse potente hanno reclutato Clams Casino per la produzione e Vic Mena per fare da MC. Questo è stato il colpo di genio degli esperti di marketing. Hanno chiamato Us Today e hanno chiesto: "Come fate a capire cosa includere nel vostro riepilogo?" Procter and Gamble stava per eliminare il marchio Old Spice e il mio amico alla Wieden+Kennedy sapeva che bisognava comunicare con il loro pubblico, composto principalmente da donne che prendono le decisioni di consumo in casa. Così hanno ideato questa campagna con un personaggio come Keanu Reeves, un surfista che parlava direttamente al pubblico la lingua stessa sta cambiando la lingua. L'inglese che usiamo oggi è diverso da quello di Beowulf.</p>
--	--	--

It is worth noting that the misrecognition of entity names observed in the previous section led to a series of cascading errors in the Italian interlingual subtitles. In fact, most of the mistranscriptions of entity names were left unaltered in their translation. In one specific case (namely, the misrecognition of the name of the newspaper *USA Today* as "us today") led to a sentence whose meaning is totally different from what the presenter is actually saying (as it becomes "ci hanno chiamato oggi", meaning "they called us today", where "us" is an object pronoun and "today" is an adverb of time"). On the other hand, the post-edited version of the subtitles performed by the students presents an accurate rendering of the entity names (except just one case, namely Vic Mena, which was not corrected as Vic Mensa). Interestingly, the students added the meaning of the acronym DE (in the original "Ed"), by making use of a strategy of explicitation, presumably to clarify what it

refers to. Moreover, the students recognised the reference to the actor Keanu Reeves (which was not recognised by Google Voice and consequently by Google Translate), although the presenter clearly refers to the role he played in the film *Point Break* (“surfer dude”), which was released in 1991 before the students were born. Finally, as mentioned, the fact that the intralingual subtitles simply transcribe what the presenter says without punctuation and without intervening on certain features which are typical of spontaneous spoken language (such as hesitations and repetitions), gives rise to syntactic errors. This is especially clear in the last example, where, while correctly post-editing the name of the epic poem *Beowulf*, the students did not notice that the presenter was actually saying “Language itself is evolving. The language, the English that we use today, is different from the English of Beowulf”. Their first sentence, indeed, can be back translated as “Language itself is changing the language”, which is not correct.

4) Misrecognition of specialized terminology

As previously noted, a key goal of this study was to examine the specialized vocabulary used in the marketing field. The students were able to detect several instances where Google Voice had inaccurately transcribed words or phrases in the intralingual subtitles. In some situations, the students were unfamiliar with the terminology themselves, making the exercise beneficial not only for developing their listening abilities but also for expanding their vocabulary and deepening their understanding of concepts within this specialized area. The table below reports the interlingual subtitles of the passages that were analysed in the previous section, as well as the post-edited versions performed by the students:

Table 8.

Transcription of intralingual subtitles containing examples of misrecognition of specialized terminology, the corresponding interlingual (Italian) version and their post-edited version

Intralingual subtitles	Interlingual subtitles	Post-edited subtitles
if we're in a focused group of moms we're unlikely to admit to sharing junk food with our kids	se facciamo parte di un gruppo concentrato di mamme difficilmente ammetteremo di condividere cibo spazzatura con i nostri figli	se facciamo parte di un focus group di mamme difficilmente ammetteremo di condividere cibo spazzatura con i nostri figli
she actually drew it on a napkin in less than a minute and then she spent a whole year with Bor justifying her intuition	lo ha effettivamente disegnato su un tovagliolo in meno di un minuto e poi ha trascorso un anno intero con Bor giustificando la sua intuizione	In effetti lo ha disegnato su un tovagliolo in meno di un minuto e poi ha trascorso un anno intero con i consigli esecutivi per giustificare la sua intuizione
declining sales and Sheriff voice meant that they were losing out if you haven't seen a peso model in your marketing classes yet you will it sensor paid earn shared and own and it stands for your advertising which is scalable your external endorsements which are credible your shared media property like social which makes you authentic and your website which is where you define Who You Are	il calo delle vendite e la voce dello sceriffo significavano che stavano perdendo terreno. se non hai ancora visto un modello in peso nelle tue lezioni di marketing, lo farai con il sensore pagato, guadagnerai condiviso e possiedi e rappresenta la tua pubblicità scalabile, le tue sponsorizzazioni esterne che sono credibili, la tua proprietà multimediale condivisa come i social che ti rende autentico e il tuo sito web, dove definisci Chi sei	il calo delle vendite e del peso pubblicitario significava che stavano perdendo successo. se non avete ancora visto un modello PESO nei vostri corsi di marketing, lo vedrete. Questo viene condiviso ed è di proprietà e sta per la vostra pubblicità che è scalabile e le sponsorizzazioni esterne che sono credibili. Le vostre proprietà mediatiche condivise come social, che vi rende autentici, e il vostro sito web che è il luogo in cui definite chi siete.

we worked with 1 adla on a campaign called two days beat for HP	abbiamo lavorato con 1 adla su una campagna chiamata Two Days Beat for HP	Abbiamo lavorato con un'adla su una campagna chiamata "Due giorni di vita" per HP.
--	--	---

In this case as well, it is possible to note that some of the errors present in the intralingual subtitles led to cascading errors in the intralingual subtitles. First of all, in the first passage, the phrase “focused group” was translated literally as “gruppo concentrato”. The group of students who performed the post-editing version of the subtitles, however, identified the mistake and correctly translated it into “focus group”.

In the second passage, as seen in the section about intralingual subtitles, Google Voice misinterpreted the word “board” (meaning “committee”) and transcribed it as “Bor” (a non-existing word). Google Translate maintained the word in the interlingual subtitles (another case of cascading error). While the students who worked on this passage rendered it as “consigli esecutivi”, the correct terminology would actually be “consigli d’amministrazione”.

The third passage presents another case of a cascading error. Indeed, as Google Voice misinterpreted the phrase “share of voice” as “Sheriff voice”, Google Translate rendered it literally as “la voce dello sceriffo”, which makes the sentence completely meaningless. The students who performed the post-edited version of this passage translated the phrase as “calo del peso pubblicitario”, which is an acceptable translation, although Italian marketing discourse uses the English phrase¹⁰.

Moving on to the following passage, it is evident how the intralingual subtitles do not convey the exact meaning of what the speaker is referring to. First of all, “PESO model” is written in lower case letters (as in the intralingual subtitles), which makes it difficult to understand that he is referring to an acronym. Moreover, the Italian intralingual subtitles have inserted the preposition “in” (“un modello in peso”), which makes it even harder to realize that PESO is actually the name of this specific model, so it could be interpreted as the currency used in Southern American countries, such as Mexico and Argentina. Furthermore, the intralingual subtitles misrecognized the phrase “it stands for” (followed by the four words whose initials form the acronym: paid, earned, shared and owned) as “it sensor”, which causes a cascading error in the interlingual subtitles. Indeed, Google Translate renders the sentence as “con il sensore pagato, guadagnerai, condiviso e possiedi”, which makes no sense. Interestingly, whereas the group of students who had worked on the post-editing phase of the intralingual subtitles of this passage had managed to identify and render the meaning of the passage quite accurately (except for the tense of the verbs), the students who worked on the post-editing phase of the interlingual subtitles failed in this objective, so their Italian proposal of the passage presents many inaccuracies (“Questo viene condiviso ed è di proprietà” literally means “this is shared and it is owned”).

Finally, the last passage cannot be considered an efficient translation upon considering both the automatic interlingual translation and their post-edited version. Indeed, as observed in the previous section, ADLA is an initialism (where each letter is spelled out individually). By transcribing it with lower case letters the intralingual subtitles are not accurate, and such lack of accuracy was maintained both in the interlingual subtitles and their post-edited version.

6. Concluding remarks

The aim of this paper was to contribute to the field of foreign language teaching and learning by exploring the use of Audiovisual Translation (AVT). Specifically, it investigated how automatic subtitles can be employed in English as a Second Language (ESL) and English for Specific Purposes (ESP) classrooms. The objective was twofold: it aimed at enhancing students’ listening skills by analysing automatic intralingual subtitles provided by Speech Recognition Systems and detecting potential errors in the transcription of the spoken output, and it sought to improve the acquisition of specialized terminology in the specific field of marketing discourse, as the video that was analyzed focused on this topic.

The study has confirmed the fact that, although incorporating Automatic Speech Recognition and Machine Translation into platforms like YouTube has significantly expanded public access to information and entertainment, and despite significant advancements in these technologies in recent years, achieving perfect accuracy remains a challenging goal. In fact, although YouTube aspires to create a fully automated interlingual subtitling system capable of generating subtitles in various languages without human involvement, this ambitious target is proving to be quite complex at present. It is true that the primary purpose of these subtitles

¹⁰ <https://ahrefs.com/blog/it/share-of-voice/>

is to provide users with a basic understanding of the content uploaded to the platform, and they generally fulfil this role effectively. Nonetheless, accuracy continues to be a considerable concern.

It has to be noted that the study was based only on one video clip, and that the students had a very limited amount of time to work on it and perform the tasks they were required to carry out. Three lessons (lasting two hours each) are not sufficient to provide them with the necessary skills. Therefore, it would be necessary to dedicate more time both to the students' training in this field and to the time allotted to study the content of the material more in depth and obtain more significant results.

In spite of these limitations, the study has provided interesting insights. First of all, the students learned that analyzing automatic intralingual subtitling and detecting the errors they contain can be a useful exercise to practice, and hopefully enhance, their listening skills (something they did not have the chance to do by working only on written texts). Secondly, they had the opportunity to improve the specialized terminology related to one of their fields of study. Last but not least, they practiced their skills in post-editing texts, which is a necessary ability if they are to become mediators.

7. References

- Adams, Heather, & Díaz Cintas, Jorge (2022). Special issue: audiovisual translation in the foreign language classroom – Introduction. *Revista de lenguas para fines específicos*, 28(2), 10-15 <http://doi.org/10.20420/rife.2022.548>.
- Aliprandi, Carlo, Scudellari, Cristina, Gallucci, Isabella, Piccinini, Nicola, Raffaelli, Matteo, del Pozo, Arantza, Álvarez, Aitor, Arzelus, Haritz, Cassaca, Renato, Luis, Tiago, Neto, Joao, Mendes, Carlos, Paulo, Sérgio, & Viveiros, Marcio (2014). Automatic live subtitling: state of the art, expectations and current trends. *Proceedings of NAB Broadcast Engineering Conference: Papers on Advanced Media Technologies, Las Vegas*, Vol. 13.
- Baños, Rocio, & Sokoli, Stavrouls (2015). Learning foreign languages with ClipFlair: Using captioning and revoicing activities to increase students' motivation and engagement. In Borthwick, Kate, Corradini, Erika & Dickens, Alison (Eds.). *10 years of the LLAS E-learning symposium: case studies in good practice*, research-publishing.net. (pp. 203–213). ResearchPublishing.net
- Bolaños García-Escribano, Alejandro & Díaz-Cintas, Jorge (2023) Integrating post-editing into the subtitling classroom: What do subtitlers-to-be think? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 115–137.
- Bolaños García-Escribano, Alejandro, & Navarrete, Marga (2022). An action-oriented approach to didactic dubbing in foreign language education: Students as producers. *XLinguae*, April, doi:10.18355/XL.2022.15.02.08.
- Burchardt, Aljoshia, Lommel, Arle, Bywood, Lindsay, Harris, Kim & Popovic, Maja (2016). Machine translation quality in an audiovisual context. *Target*, 28(2), 206-221.
- Burgess, Jean & Green, Joshua (2018). *YouTube: online video and participatory culture. Second edition*. Polity Press.
- Bywood, Lindsay, Georgakopoulou, Panayota, & Etchegoyhen, Thierry (2017). Embracing the threat: machine translation as a solution for subtitling. *Perspectives Journal: Studies in Translatology*, 25(3), 492–508.
- Carreres, Ángeles (2006). Strange bedfellows: Translation and language teaching. The teaching of translation into L2 in modern languages degrees; uses and limitations. *Sixth Symposium on Translation, Terminology and Interpretation in Cuba and Canada*. Canadian Translators, Terminologists and Interpreters Council.
- Ciobanu, Dragos, & Secară, Alina (2019). Speech recognition and synthesis technologies in the translation workflow. In O'Hagan, Minako (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 91–106). Routledge.
- Cook, Guy. (2010). *Translation in Language Teaching: An Argument for Reassessment*. Oxford University Press.

- Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.
- Cronin, Michael (2003). *Translation and globalization*. Routledge.
- Danan, Martine (1992). Reversed subtitling and dual coding theory: new directions for foreign language instruction. *Language Learning*, 42(4), 497–527.
- Danan, Martine (2015). Subtitles for language learners: an overview of research and application. In Serban, Adriana, Matamala, Anna, & Lavaur, Jean Marc (Eds.), *Audiovisual translation in close-up: practical and theoretical approaches* (p. 41). Peter Lang.
- Díaz Cintas, Jorge, & Remael, Aline (2007). *Audiovisual translation: subtitling*. St. Jerome Publishing.
- Díaz Cintas, Jorge, & Fernández Cruz, Marco (2008). Using Subtitled Video Materials for Foreign Language Instruction, in Díaz Cintas, Jorge (Ed.) *The Didactics of Audiovisual Translation* (pp. 201–214). John Benjamins.
- d'Ydewalle, Géry., & Van de Poel, Marijke (1999). Incidental foreign–language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28(3), 227–244.
- Ellis, Brad (2011). Auto-Captions now available in Japanese. YouTube. <https://tinyurl.com/ycfjywyv>. Accessed July 2024.
- European Commission. (2013). *Translation and language learning: The European commission's study on translation and multilingualism*. Publications Office of the European Union.
- Ferrara, Maria Pia (2023). *Interlingual subtitles in TED Talks: Exploring the role and effectiveness of NMT*. Unpublished MA thesis. Università di Torino.
- Freitas, Connor, & Liu, Yudong (2017). Exploring the differences between human and machine translation. *WWU Honors Program Senior Projects*. 61. Retrieved from https://cedar.wvu.edu/cgi/viewcontent.cgi?article=1060&context=wwu_honors. Accessed July 2024.
- Gambier, Yves (2008). Recent developments and challenges in audiovisual translation research. In Chiaro, Delia, Heiss, Christine & Bucaria, Chiara (Eds.) *Between Text and Image* (pp. 11–33). John Benjamins.
- Ghia, Elisa (2012). *Subtitling matters: new perspectives on subtitling and foreign language learning*. Peter Lang.
- Harrenstien, Ken (2009). Automatic captions in YouTube [Blog post]. Retrieved from <https://googleblog.blogspot.com/2009/11/automatic-captions-in-youtube.html>. Accessed July 2024.
- Hussin, K. Khairiah, & Hishan, Sanil S. (2022). The implementation of brand awareness strategies to attract new customer to the company. *International Journal of Academic Research in Business and Social Sciences*, 12(1), 1534–1552.
- Incalcaterra McLoughlin, Laura, Lertola, Jennifer, & Talaván, Noa (Eds.). (2020). *Audiovisual Translation in Applied Linguistics: Educational Perspectives*. John Benjamins.
- Incalcaterra McLoughlin, Laura, Lertola, Jennifer, & Talaván, Noa (Eds.). (2018). *Audiovisual translation in applied linguistics: educational perspectives*. Special Issue of *Translation and Translanguaging in Multilingual Contexts*. John Benjamin. <https://doi.org/0.1075/ttmc.4.1>
- Jiménez-Crespo, Miguel (2019). Localisation research in translation studies: Expanding the limits or blurring the lines? In Dam, Helle V., Brøgger, Matilde Nisbeth & Zethsen, Karen Korning (Eds.), *Moving boundaries in translation studies* (pp. 26–44). Routledge.
- Karakanta, Alina, Negri, Matteo, & Turki, Marco (2020). Towards automatic subtitling: assessing the quality of old and new resources. *Italian Journal of Computational Linguistics*, 6(1), 63–76. <https://doi.org/10.4000/ijcol.649>
- Karakanta, Alina (2022). Experimental research in automatic subtitling. *Methodological Issues in Experimental Research in Audiovisual Translation and Media Accessibility*, 11(1), 89–112. <https://doi.org/10.1075/ts.21021.kar>.

- Kernez, Ross (2022). What is the PESO model in marketing? *Forbes*, November 9th, <https://www.forbes.com/sites/forbescommunicationscouncil/2022/11/09/what-is-the-peso-model-in-marketing/>, accessed July 2024.
- Kruger, Jan-Louis, Szarkowska, Agnieszka & Krejtz, Izabela (2015). Subtitles on the moving image: an overview of eye tracking studies. *Refractory: a journal of entertainment media*, 25, 1–14.
- Kruger, Jan-Louis, Hefer, Esté, & Matthew, Gordon (2013). Measuring the Impact of Subtitles on Cognitive Load: Eye Tracking and Dynamic Audiovisual Texts. *Proceedings of Eye Tracking South Africa, Cape Town, 29-31 August 2013*, 62–66.
- Laviosa, Sara (2014). Translation and language education: Pedagogic approaches explored. *The Interpreter and Translator Trainer*, 8(1), 1–16.
- Lee, Seryun (2021). An exploration of lingua-cultures on YouTube: translation and assemblages, *Social Media + Society*, 7(4), 1–12.
- Lertola, Jennifer (2019). *Audiovisual translation in the foreign language classroom: applications in the teaching of english and other foreign languages*. ResearchPublishing.net.
- McFadden, Cristopher (2025). A brief timeline of YouTube's history and its impact on the internet. *Interesting Engineering*. <https://tinyurl.com/pnu95rey>. Accessed July 2025.
- Navarrete, Marga, & Bolaños García-Escribano, Alejandro (2022). An action-oriented approach to didactic audio description in foreign language education. *Revista de lenguas para fines específicos* 28(2), 151–164 <https://doi.org/10.20420/rlfe.2022.556>
- O'Brien, Sharon (2022). How to deal with errors in machine translation: Post-editing. In Kenny Dorothy (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, (pp. 105–120). Language Science.
- O'Hagan, Minako (2009). Evolution of user-generated translation: fansubs, translation hacking and crowdsourcing. *The Journal of Internationalization and Localization*, 1(1), 94–121.
- Parini, Ilaria (forthcoming). Automatic generated intralingual and interlingual subtitling and the need for human intervention. In Rashid Yahiaoui (Ed.) *Transl-AI-tion 2.0: Embracing the AI revolution*. Peter Lang.
- Pujadas, Geòrgia & Muñoz, Carmen (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479–496.
- Smith, Chad E., Crocker, Samantha & Allman, Tamby (2017). Reading between the lines: Accessing information via YouTube's automatic captioning. *Online Learning, The official journal of OLC*, 21(1). 115–131.
- Soffritti, Marcello (2018). Multimodal corpora in audiovisual translation studies. In Pérez González, Luis (Ed.) *The Routledge Handbook of Audiovisual Translation* (pp. 334–349). Routledge.
- Sokoli, Stavroula, & Zabalbeascoa-Terrán, Patrick (2019). Audiovisual activities and multimodal resources for foreign language learning. In Herrero, Carmen & Vanderschelden, Isabelle (Eds.). *Using Film and Media in the Language Classroom* (pp. 170–187). Multilingual Matters.
- Talaván, Noa (2020). The didactic value of AVT in foreign language education. In Ł. Bogucki and M. Deckert (Eds.) *The Palgrave Handbook of Audiovisual Translation and Media Accessibility* (pp. 567–591). Palgrave.
- Tokusei, Hiroto (2010). The future will be captioned: improving accessibility on Youtube. YouTube. <https://tinyurl.com/bdac3dt4>. Accessed July 2024.
- Turovsky, Barak (2016). Ten years of Google Translate. Google Translate Blog. <https://tinyurl.com/23y7au5a>. Accessed July 2024.
- Vanderplank, Robert (1988). The value of teletext subtitling in language learning. *ELT Journal*, 42(4), 272–281.
- Vanderplank, Robert (1990). Paying attention to the words: practical and theoretical problems in watching television programmes with unilingual (CEEFAX) Subtitles. *System*, 18(2), 221–234.

- Vanderplank, Robert (2016). *Captioned media in foreign language learning and teaching: subtitles for the deaf and hard-of-hearing as tools for language learning*. Palgrave Macmillan.
- Varga, Cristina (2021). Online automatic subtitling platforms and machine translation: an analysis of quality in AVT. *Scientific Bulletin of the Politehnica University of Timișoara: Transactions on Modern Languages*, 20(1), 37–49.
- Volk, Martin (2008). The Automatic Translation of film subtitles. a machine translation success story? *Journal for Language Technology and Computational Linguistics*, 24(3), 113–125. <https://doi.org/10.21248/jlcl.24.2009.124>.
- Juang, Bing-Hwang & Rabiner Lawrence R. (2005). Automatic speech recognition—a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara* 1(67), 1–24.
- Zabalbeascoa Terrán, Patrick, Torres, Olga & Sokoli, Stavroula (2012). *Conceptual framework and pedagogical methodology. clipflair foreign language learning through interactive revoicing and captioning of clips*. Lifelong Learning Programme - Key Activity 2. Languages, Multilateral Project. D2.1. ClipFlair. <http://clipflair.net/wp-content/uploads/2014/06/D2.1ConceptualFramework.pdf>

Ilaria Parini, Università del Piemonte Orientale, Italia
 ilaria.parini@uniupo.it

- EN** | **Ilaria Parini** is senior lecturer in English language, linguistics and translation at the University of Eastern Piedmont, Italy. Her research interests include audiovisual translation, literary translation, translation of non-standard varieties, as well as manipulation and censorship in translation. She is a member of the European Association for Screen Translation, the European Society for Translation Studies, the Inter-University I-Land Centre (Identity, Language and Diversity), the Inter-University Research Centre on Metaphors (CIRM), AIA (Italian Association for the Study of English), and ESSE (European Society for the Study of English).
- ES** | **Ilaria Parini** es profesora contratada doctora de lengua, lingüística y traducción inglés en la Universidad del Piemonte Oriental, Italia. Sus áreas de investigación incluyen la traducción audiovisual, la traducción literaria, la traducción de variedades no estándar y la manipulación y censura en la traducción. Es miembro de la European Association for Screen Translation, la European Society for Translation Studies, el Centro Interuniversitario I-Land (Identidad, Lengua y Diversidad), el Centro Interuniversitario de Investigación sobre Metáforas (CIRM), la AIA (Italian Association for the Study of English) y la ESSE (European Society for the Study of English)
- IT** | **Ilaria Parini** è Ricercatrice Tenure Track di lingua, linguistica e traduzione inglese presso l'Università del Piemonte Orientale, Italia. I suoi ambiti di ricerca includono la traduzione audiovisiva, la traduzione letteraria, la traduzione di varietà non standard e la manipolazione e censura nella traduzione. È membro della European Association for Screen Translation, la European Society for Translation Studies, il Centro Interuniversitario I-Land (Identità, Lingua e Diversità), il Centro Interuniversitario di Ricerca sulle Metafore (CIRM), dell'AIA (Associazione Italiana di Anglistica) e della ESSE (European Society for the Study of English).