Technologization of audiovisual translation

Panayota (Yota) Georgakopoulou

Since the turn of our century, the Audiovisual Translation (AVT) industry has seen a period of transformation as a result of rapid technological change. This is only natural, as technology and AVT have always gone hand-in-hand. The very existence of AVT is a by-product of developments in film, video and broadcasting technologies, and as such it continues to change and evolve with them. In our digital, globalized world, an industry international by definition, with professionals working in different languages around the world, can expect trends such as the cloud, the crowd and big data to play a pivotal part in shaping its future course.

The beginnings of audiovisual translation

AVT started out as an art form originally interlinked with the video material itself as filmmakers were experimenting with narrative in the early days of classical Hollywood cinema, e.g. intertitles in the silent film era (Thompson and Bordwell 1994: 39–40), or trying to cross the language barrier when sound made its appearance, e.g. by re-shooting a film in additional versions so as to cater for other languages (ibid.: 229–230). It became an add-on product to the film production process in the 1940s, when technology made it possible to separate the two and thus provide a financially viable way to translate film for foreign audiences. This led to the birth of subtitling and dubbing as the two main methods of audiovisual language transfer, which is still the case today.

AVT remained largely a manual craft through to the 1970s, with several professionals being involved in the process. Translators would work from a provided script to translate already spotted subtitles, or adapt film scripts for lip synchronization in the case of dubbing, but their skills were not considered sufficient for the more technical parts of the process. It was technicians and typists that transcribed the handwritten subtitles and cued them, while their insertion onto the film involved a variety of mechanical, thermal, photochemical and optical methods (see Ivarsson and Carroll 1998: 12–23).

The 1970s marked the first major milestone for the AVT industry, as the first open-captioned programmes were run on American TV. Line 21 of the broadcasting signal was reserved for the transmission of closed captions in the USA and teletext was invented in the UK, later to become an international standard adopted by the European Broadcasting Union (EBU).
Live closed-captioned broadcasts began in the 1980s, with the use of court stenographers, who were trained as stenocaptioners and utilized specially designed keyboards and short-forms to churn out captions at a speed similar to speech delivery in live broadcasts. Regular offline and live closed-captioning/teletext subtitling of television broadcasts for accessibility purposes became the norm from the 1980s onwards and its use increased as legislation was introduced in the decades that followed.

New technologies pushed the entertainment industry forward in the 1980s. The introduction of VHS in the late 1970s contributed to the rapid expansion of home video throughout the world, while the growth of cable and satellite channels in the 1980s boosted the volume of programming available and the need for AVT treatment. At the same time, the advent of the desktop computer revolutionized the translation process and, in particular, AVT. Dedicated subtitling software, originally DOS-based and linked to a TV monitor, caption generator and VHS player with jog shuttle, was developed, which allowed subtitlers to carry out the entire process from beginning to end, including the spotting, timing, translation of the dialogue and its adaptation to fit the length and time constraints of subtitles. This resulted in the definition of the job of a subtitler as a distinct profession in the market.

The digital revolution

**Word processing, subtitling software, digital video, Web 1.0**

The 1990s saw an increase in the pace of developments in the subtitling industry with concurrent technological advances. The rise in the use of Windows-type interfaces and word-processing programs was a turning point for the translation industry in its entirety, which evolved to the ‘localization’ industry as we know it today. Spell and grammar checkers were the first ‘tools’ translators ever used. Traditional translation agencies were transformed into Language Service Providers (LSPs), not simply translating, but localizing content, modifying it and adapting it to local requirements.

The subtitling industry also evolved as subtitling software became more sophisticated. More companies developing such software started appearing in the market to offer tools that simplified the subtitling task, mainly from a technical/timing aspect: FAB (Germany, 1996), SoftNI (USA, 1996), TitleVision (Denmark, 1996), CANVASs (Japan, 1997), Spot Software (Netherlands, 1997), Telesstream (USA, 1998), Starfish Technologies (UK, 2000) and EZTitles (Bulgaria, 2002) came to compete with pioneers such as SysMedia (UK, 1974), Screen Systems (UK, 1976), Softel (UK, 1983) and Cavena (Sweden, 1989). These companies remained the world leaders until today in terms of products that allow the creation, repurposing, encoding, transcoding, insertion and transmission of subtitles. SysMedia was acquired by Screen Systems in 2011 (Screen Systems 2011), while Softel was taken over by Miranda in 2013 (Grass Valley 2013)—which sadly discontinued Softel’s suite of products in 2018 (Grass Valley, n.d.).

Digital video formats in the 1990s made external VHS players used in subtitling workstations redundant. Films were digitized and stored on company servers, accessible from in-house subtitlers’ workstations, or later copied onto DVD-Rs to be shipped to freelance workers. By the turn of the century, Digital Betacam machines were also discarded as clients began moving to entirely tapeless workflows. Not only was there no need to have expensive software and hardware in order to deliver subtitles to end clients, it was no longer necessary to be located in close proximity to them either.

The same was true of the majority of the workforce that was responsible for the creation of the subtitles. As an international industry by default, it was only natural that translators
embraced technological advances and were counted among the early adopters of computers and dial-up modems. This led to an increase in homeworkers translating in text editors off-site and performing review and quality control only in dedicated subtitling workstations in the offices of subtitling localization providers. With the widespread adoption of the internet and the development of less costly subtitling software addressed to freelancers, this geographic proximity became even less of a necessity. At the same time, the introduction of basic keyword search on the web meant that translators no longer needed to rely on books and physical libraries for their research, making their geographical location almost irrelevant to their ability to offer their services to their clients.

**DVDs and the template workflow in subtitling**

The birth of DVD in the mid-1990s marked a new era for the AVT industry, in part because it boosted content volume further and made subtitling more prominent in traditional dubbing countries. It also posed a challenge for the industry, as it created demand for subtitling in a large number of languages within short turnaround times concurrently with two major issues for the Hollywood studios, namely piracy and costs. The result was the emergence of a new workflow for subtitling production, involving the use of already timed and spotted template files in the source audio language to be translated into all the target languages required in any film project (Georgakopoulou 2006), a workflow inspired by the early practices of Scandinavian subtitling companies which applied a similar methodology when working into more than one Scandinavian language simultaneously.

This separation of the act of translation from the more technical aspects of the craft of subtitling for the second time in AVT history had two important consequences. On the one hand, it solved the recruitment issue multilingual subtitling companies were faced with, as it automatically expanded the pool of freelance translation editors to include all available translators worldwide who, with little training in the principles of subtitling, could now become versed in this translation domain as well. On the other hand, an indirect consequence, the value of which would only become apparent much later, was the creation of a large volume of parallel subtitling corpora in multiple languages, which would eventually be used to further advance the technologization of the subtitling profession (Georgakopoulou 2012: 92). It also paved the way for the centralization and internationalization of the home entertainment AVT business in the hands of subtitling companies strategically located along the LA–London axis, in close proximity to their Hollywood client American and European headquarters (Carroll 2004).

**First applications of language technologies in the translation industry**

The 1990s were also the decade in which language technologies were introduced commercially on a large scale to enhance translators’ productivity and consequently reduce project turnaround times and costs. Prime examples were the use of continuous dictation software, as well as technologies that offered the ability to reuse text, such as Translation Memories (TMs) and Machine Translation (MT).

With research dating back to the 1970s (Hidden Markov Model), and the n-gram language model of the 1980s, the Automatic Speech Recognition (ASR) field saw rapid improvements, which led to commercially successful applications in the 1990s, including the launch of continuous speech dictation software, such as IBM’s ViaVoice (1997) and Dragon Systems’
NaturallySpeaking (1997), the latter being the most popular dictation package until recently. Dictation software has been used by professional translators as a means to increase their productivity. However, this is still primarily an individual choice rather than a workflow requirement enforced by LSPs. The practice dates back to the use of cassette tapes for dictating translation to then be transcribed, as a productivity enhancement tool by translators in the 1980s, before dictation software became available (Hendzel 2013).

The AVT industry took an interest in speech recognition as an alternative workflow for live captioning to replace the training- and cost-intensive workflows based on stenographers. Speech recognition was integrated in subtitling software and introduced to subtitlers via respeaking (or ‘voice writing’, as it is called in the States) around the turn of the century. The respeaking workflow consisted of professionals listening to the source audio and dictating it into a speech recognition engine, which transcribed it producing same language subtitles and captions. The practice was originally mocked for the incomprehensible, amusing or even insulting machine-made mistakes the use of this technology introduced (Daily Mail Reporter 2013). However, the improvements in speech recognition software, the enrichment of dictionaries used (including the use of crawling techniques to look for new named entities that appeared in the news for instance, so as to further enrich dictionaries), as well as the experience gained by respeakers themselves resulted in quality outputs similar to the ones achieved by stenographers, who are being phased out of the live subtitling profession. The practice of respeaking has recently also been implemented as an alternative way to produce real-time translation subtitles, which until now required the use of an interpreter in combination with a stenocaptioner, a process which resulted in high delays and costs, and was hence rarely practised. Interlingual respeaking workflows are slowly being implemented commercially (e.g. Red Bee Media Spain), while universities (e.g. Antwerp) are researching the possibility of cross-training interpreters into translation respeakers. The introduction of technology in live translation subtitling workflows and the resulting drastic reduction in cost will make it possible for real-time translation subtitling to become an everyday service in the years to come.

On the translation side, Computer Assisted Translation (CAT) tools became standardized and widespread, after being introduced in the workflows of LSPs specializing in the translation of technical texts. Shared glossaries evolved to fully blown terminology management systems, translation memories reduced the work of translators by providing the translation of previously translated chunks of text that to an extent matched the new source text, and quality assurance tools were integrated with text editors in translator workbenches. Rule-based machine translation (RBMT) was also commercially used for the first time, originally in controlled-language restricted-vocabulary domains, such as manuals, where the results were yielding significant accuracy to enhance translators’ productivity. The AVT industry remained largely detached from these developments until the following century, mainly on the assumption that they were inappropriate for implementation in the entertainment domain, which provided the bulk of the material undergoing interlingual subtitling. However, significant developments in recent years, both in terms of content types that require AVT, as well as in the overall volume of content to be translated, have paved the way for experimentation with such technologies that are yielding promising results.

The broadband revolution: trends and debates

The impact of broadband Internet since the turn of the century has been much greater on the AVT industry than one could have envisaged.
Outsourcing

Broadband Internet access catalyzed the transformation of the translation industry into the globalized and centralized localization industry we know today. As the industry drew increasingly more on freelance translators and editors located anywhere in the world, the business was eventually outsourced to low-cost territories with multilingual capabilities. In the AVT industry, the outsourcing trend began at the turn of the century, when USA and UK companies had to find cheaper ways of producing English subtitling and captioning work: the volume of such work was increasing rapidly, while the prices offered for such services were steadily declining. Respeaking was the method implemented to achieve this in live captioning workflows, while outsourcing was used for offline workflows. Countries like India, Malaysia and the Philippines were selected as the low-cost alternative for producing transcripts of English source video due to the availability of a cheap workforce and its familiarity with the English language. Such outsourcing solutions offered further positive side effects, providing USA- and UK-based companies with the scalability and flexibility to cater for increases in workload, as well as the possibility of a round-the-clock subtitling service by taking advantage of the time difference between Asia and America/Europe.

This outsourcing of a flourishing skilled profession (i.e. English intralingual subtitling) to non-specialists that were unaccustomed to the slang, regional variations and cultural references used in filmic material, and viewed subtitling/captioning as just another type of transcription, originally met with strong opposition, especially in Europe—where verbatim transcriptions of the dialogue were avoided even for hard of hearing purposes. An increase in mishearings and comprehension errors were reported and professional unions condemned the practice (Nakata Steffensen 2007; Alberge 2007), but eventually workflows were ironed out and quality control measures by native subtitlers were put in place to correct such mistakes. Multinational subtitling companies took further control of the workflow by establishing their own offices in outsourcing countries, so that they controlled their operations best and made them fit their needs.

Translation management systems and the cloud

The creation of global enterprises posed a new set of challenges in working across cultures and time zones, and harmonizing tools and processes. Such enterprises had a vested interest in investing in technology, so as to deliver efficiencies of scale, maximize control of their operations and maintain their market position. This meant that Translation Management Systems (TMSs), an indispensable tool of LSPs since the 1990s, were widely being adopted by the AVT industry as well a decade or so later. Development of such systems has been ongoing, so as to accurately track, streamline and automate as much of the procedural translation management work as possible and, in some cases, TMSs are custom-built with development taking place internally in the aforementioned organizations. TMSs have been used to centrally manage the assets to be translated, such as videos, subtitle files, etc., control the languages into which translation is meant to take place, manage and select the translators to work on projects, allocate work, submit completed files, track payments, etc. Translators originally worked off desktop editors, not linked to their client TMS, while use of translation technology tools, such as TMs and MT, was still non-existent.

The proliferation of video formats in the twenty-first century and the explosion of an ever larger volume of audiovisual content online created another peak in demand for translation and accessibility services, while the timeframes for the provision of such services kept
Technologization of audiovisual translation

shrinking. Enter the cloud: it provided cost reduction through automation and avoidance of effort duplication, scalability and, more importantly, real-time access to and sharing of project information. It is now possible for global enterprises to become bigger than ever before and offer continuous localization services, which they can achieve by shifting from server-based TMSs to cloud-based ones: online editors, dashboards, automated work allocation, elimination of project handovers and much shorter turnaround times as a result. AVT companies further benefit from greater media security by adopting online editors, as video is streamed instead of downloaded to translators’ workstations, while the download times disappear, speeding up the work of homeworkers. Additional efficiencies are made possible with the use of change management technology, tracking changes in source language content and mirroring them in all target language versions. Companies are now more inclined to integrate linguistic processing tools, such as terminology management tools and MT, to increase consistency and efficiency on the part of the translators. Increased quality control is made possible, metrics and performance indicators become available and clients have immediate access to the translated material for review and sign off. Such systems offer full project transparency, which has a hidden dynamic that could drastically change the visibility of translators in the supply chain of global LSPs, as they are now accessible by end clients who demand full disclosure of their vendors’ processes.

Audiovisual content explosion

Broadband Internet also had a direct impact on the amount of video content that could be made available to consumers, unleashing unprecedented volumes of material. Broadcast TV and Internet access was integrated through a single set-top box at the turn of the century through IP delivery over ADSL, and Video on Demand (VOD) was launched as a commercial service (Wikipedia 2015), with Netflix being the prime example. The generalization and growing ubiquity of on demand audiovisual services have caused a dramatic increase in video consumption, and changed the ways audiences consume such content. ‘Binge watching’ and ‘cord-cutting’ are terms that characterize audience behaviour of late, while many refer to the next generation of viewers as the ‘cord-nevers’, i.e. viewers that will watch all of their content online, never resorting to pay-TV subscriptions (Rick 2014).

Moreover, the use of a multitude of devices to watch content has resulted in an increase of cloud-based TV services and the broadcast industry is embracing online delivery of content (McDonald 2015). Devices such as tablets and smartphones are also used as companion devices to the TV; they enhance the viewing experience by making it more interactive and allow, for instance, real-time chat during broadcasts. What became known as the ‘second screen’ phenomenon is now used among other things to deliver AVT services in a personalized manner to one’s tablet or smartphone (e.g. the MovieReading App for synching subtitles and audio description tracks).

It is not only the term ‘broadcast’ that is being redefined in our days however. The term ‘audiovisual content’ has also taken on a whole new meaning, as communication itself in our century is becoming increasingly audiovisual. User Generated Content online is booming, largely promoted by social media sites such as Facebook, whereas 400 hours of audiovisual content are uploaded every minute on YouTube (Statista, n.d.), which boasts more than 1 billion users. Use of corporate video is also on the rise, as companies use video for their profiles, presentations and webinars, which they often share online. As a result, traditional LSPs are adding AVT services to their offering in order to satisfy their client needs, to the extent that support of subtitle formats is added in TM software used to ensure such subtitle
translations are consistent with the clients’ other printed and online material (Smith 2013). At the same time, the education sector is reinventing itself and Massive Open Online Courses (MOOCs) have become a global phenomenon, with universities expanding their repositories of video lectures and looking for ways to make them accessible. Coursera, a major MOOC provider, subtitles its courses in English and translates them online into more than 65 languages, by crowdsourcing translations through its Global Translation Community. In fact, translation is the platform’s main selling point, as the provision of translated subtitles leads to increases in course enrolments by up to 200–300 per cent (Open Education Europa 2014).

For the first time in subtitling history, everyone is trying their hand at subtitling, in order to make this tsunami of videos accessible to a wider audience. Desktop subtitling and captioning software is mushrooming and becoming affordable for all types of budgets (e.g. Annotation Edit, Subtitle Translation Wizard, Movie Captioner), while open source subtitling software is now also available (e.g. Aegisub, Gaupol, Gnome Subtitles, Jubler, Subtitle Edit, Subtitle Editor, Subtitle Workshop). In addition, online subtitle interfaces are making their appearance, either proprietary ones (e.g. iMediaTrans, Plint, Sfera, ZOOsubs) or commercially available (e.g. MediaWen, OOONA’s production tools), while companies like Amara, DotSub and Viki make use of online subtitle platforms to offer accessibility and translation services through a streamlined process of collaborative subtitling. While online subtitle platforms typically do not offer the standard of functionality found in high-end offline subtitle editors, they offer greater security to content owners via streaming videos, while simplifying the work of homeworkers. As such, we should expect their functionality to improve over time and eventually for them to completely take over from offline editors.

As the volume of content to be subtitled keeps increasing and the respective timeframes continue to shrink, companies are looking into new workflows in order to tackle their client demands, such as experimentation with respeaking for offline subtitle production, or the repurposing of live subtitle files into offline ones. At the same time, the notion of broadcast is redefined with Over The Top (OTT) offerings to a global audience, while non-entertainment video production is booming, creating new markets in which traditional subtitling/captioning enterprises compete against new types of companies offering alternative workflows based on technology and crowdsourcing.

**Crowdsourcing**

One of the most innovative trends that emerged as a side effect of moving to cloud-based enterprise offerings is crowdsourcing as the next level of corporate outsourcing. Not only have crowdsourcing models provided a levelling field for the supply and cost of labour, in the translation and AVT field, they have also provided solutions to issues such as speed, volume and resource availability, allowing the practice to flourish despite the attendant legality and quality issues.

**Fansubbing**

Crowdsourcing in AVT first appeared in the form of fansubbing, i.e. the creation of subtitle files for (popular) audiovisual productions by fans, in a fashion that breaks away from the established norms of subtitling practice. It started in the 1980s but has truly spread since the turn of the century, and it differs considerably from established forms of crowdsourcing in that the crowd maintains complete control over the translation process and the content to be translated, as opposed to ‘managed’ forms of crowdsourcing, where such control is maintained by
the originators of the work (European Commission 2012). The creativity of solutions employed in fansubbing, which can be seen as a hybrid between traditional subtitling, subtitling for the hard of hearing and video game localization (Díaz Cintas 2005), has the potential to question established norms in the subtitling profession and influence subtitle production in the future.

Additionally, the very fact that such solutions are primarily provided by non-professionals, and are thus prone to translation errors and a presumed lower quality level, is likely to strengthen the argument of ‘fit for purpose’ translation. In this sense, crowdsourcing can help pave the way towards establishing a scale of translation quality levels for clients to choose from, depending on their budgets and the intended use of their translated content. Taking this a step further, Gambier (2006) argues that this could also strengthen the reasoning regarding the application of language technologies such as machine translation in subtitling, by ‘viewing the output of machine translation programs . . . in a different light, in that they satisfy a certain number of users who are far from the illiterate but who do not need a polished, finely honed text’.

As the fansubbing phenomenon grows, the audience reception of amateur subtitles has been the focus of increasing academic research. Orrego-Carmona (2014, 2015) has shown that non-professional subtitles can provide a similar level of satisfaction to the audience as professional ones. The European Commission has also recognized the value of crowdsourcing by publishing a call for proposals (2015) focusing on innovative uses of crowdsourcing in order to strengthen the circulation of videos in lesser-used European languages through subtitling for VOD services, in cases where commercial subtitling services cannot be financially justified, as well as a call (2016) for the creation of an online tool to make subtitles easier to find and use.

**Enterprise crowdsourcing**

As opposed to fansubbing, what has come to be known as Enterprise Crowdsourcing in recent years (see Crow 2013; Deloitte Pixel n.d.; Lionbridge 2013; Matthews 2013) is an innovative approach to outsourcing by enterprises looking to benefit from ever ubiquitous and dependable Internet connections, the automation benefits of the cloud, and the billions of networked people around the globe. By putting to use their skills and availability, they are transforming them into an on-demand workforce that can complete from simple and repetitive to complex and creative tasks, for free or for a small fee. The benefits of such solutions are obvious: cost savings, scalability and better time-to-market through the simultaneous assignment of micro-tasks to a multitude of online workers, creating a dynamic for new business that is not viable with traditional models.

Solutions based on workflows powered by communities of amateurs are increasingly making their appearance in the translation market (Conyac, Gengo, Lionbridge Business Process Crowdsourcing, etc.). Amateur communities have attracted a lot of criticism, arising from concerns over their potential effect on the status of professional translation, the declining standards of translation quality and the ensuing blurring of the lines between professional and personal lives. By contrast, their positive impact is also duly noted when it comes to the strengthening of regional and minority languages (see European Commission 2012). In any case, crowdsourced translation companies are disrupting the marketplace, by offering much lower price points and higher throughput than traditional translation agencies. In some cases, they combine amateur crowd workers working alongside professional translators, who are mainly used to vet the amateur translators upon entry in the pool and edit their work so as to achieve higher quality output. By adding governance and a multi-layered quality control mechanism, Lionbridge (2014) coined the term ‘Business Process Crowdsourcing’ to refer
to a cross of crowdsourcing with traditional Business Process Outsourcing, offering enterprises managed crowdsourcing solutions that combine the elasticity of the crowd with the quality assurance, security and performance management businesses require.

Prominent examples of crowdsourcing from the AVT sector are Amara and DotSub, which host both amateur and professional solutions, while serving the needs of diverse communities, from non-profits, educators and governmental organizations, to entertainment, media and corporate clients, as well as individuals. Crowdsourced subtitling services are also used by large streaming sites, such as Viki, ‘the Hulu for foreign language content’ (Lunden 2013), which uses its fan base to collaboratively subtitle videos wiki-style, with segments edited and re-edited on the fly, and the ones judged by the community as the most-accurate prominently displayed under the video player (Upbin 2010, cited in Dwyer 2012: 222). Finally, Amazon Mechanical Turk, the most widely known crowd marketplace for thousands of human intelligence tasks, lists translations and captions among them.

Mature language technologies are often used in crowdsourcing scenarios in order to maximize and capitalize on the skills of pools of distributed workers. Companies that offer transcription and captioning services, for instance, may employ speech recognition technology to produce preliminary transcripts of videos that their crowd then post-edits (e.g. Cielo24). Similarly, crowdsourced translation companies may use machine translation as the first draft of the translation their crowd workers are asked to work on (e.g. Unbabel). Crowdsourcing has also been researched recently as an alternative to offering real-time captioning services, by employing an algorithm for combining and aligning partial captions created by non-expert workers, with initial results showing that such output outperforms experienced captioners and ASR systems (Lasecki et al. 2012; Lasecki and Bigham 2014; Kazemi et al. 2014), thus opening up new possibilities in real-time captioning creation.

The influence of technology in the era of Big Data

The latest buzzword in our decade is ‘big data’. It refers to data that is massive in volume, structured or unstructured, which, if processed, can help enterprises reach intelligent decisions. Subtitles are important metadata to help make videos searchable, so much so that, according to a Discovery Digital Networks study (Hammond 2014) conducted from January 2013 to April 2014, adding captions to videos led to increased viewings by 13.48 per cent, a lifetime increase of 7.42 per cent, significantly improved Search Engine Optimization (SEO) and also improved Return On Investment (ROI).

The use of statistical modelling in language technologies such as ASR and MT make them prime examples of technologies that harness the power of big data to yield useful results. Intralingual subtitles, together with the audio for which they provide timed transcripts, constitute big data used for the training of acoustic and language models that form the building blocks of ASR systems. When subtitle data are available in multiple languages for the same audio content, we talk of parallel data, which are of particular importance in the AVT field due to their applications in the training of Phrase-based Statistical Machine Translation (PBSMT) or Neural Machine Translation (NMT) systems. Subtitle files created with the template methodology are the ideal form of parallel data, already aligned at the subtitle level, and ready for use with very little effort invested in cleaning and pre-processing them. In other words, not only are subtitles important ‘data’ and ‘meta-data’, but they can also be used to improve the very tools that are used in their production; as such, they have appropriately been termed ‘meta-tools’ by Sánchez (2014).
Speech recognition

We have already seen how ASR has been used to transform the live captioning industry by implementing cost-effective workflows through the use of respeakers who, instead of typing, dictate the source language audio into specially designed subtitling software packages together with instructions relating to subtitle presentation and non-verbal audio elements that require descriptions for the benefit of hard of hearing audiences. Further progress has been made in ASR technology in the last 15 or so years since its first implementation in the AVT industry. Advances in the quality of ASR systems have made the training of respeakers faster and more affordable. The technology has also been making great strides towards speaker independent systems that require no training to the respeaker’s voice, bringing us closer to a future reality where live intralingual captioning can be completely automated.

The current accuracy levels of raw ASR output for entertainment material are still far from human accuracy, though the advances are remarkable when it comes to broadcast domains involving studio-quality sound, no overlapping speakers, and restricted vocabulary, such as weather forecasts and broadcast news. Pioneering applications of the use of ASR for the automated live captioning of the news have been made by Voice Interaction (Meinedo et al. 2003) in collaboration with RTP, the public Portuguese broadcaster, and CRIM (Brousseau et al. 2003) and the TVA network, a Canadian broadcaster. Efforts towards improving ASR output for broadcast material have continued in research projects. Most recently, the Quaero, SAVAS and EU-Bridge research projects focused on developing ASR technology in several European languages for the broadcasting industry, while the project transLectures dealt with providing automated timed transcripts in caption format for educational video lectures. All projects reported improvements on previously state-of-the-art technology and produced reduced word error rates.

As a result of the SAVAS project, three software packages were developed which address operational issues of subtitle production, such as speech classification (i.e. speech, music and jingle detection), identification of speaker turns and speakers, text normalization for segmenting running text to words and sentences and putting words in a standard format, subtitle formatting and editing, as well as subtitle generation according to broadcaster specifications (Aliprandi et al. 2014). S.Live! was developed for automated production of live intralingual subtitles, S.Scribe! for offline subtitle production and S.Respeak! as a remote respeaking system scalable from single use to collaborative subtitling, available both in offline and live modes, with a fast post-editing component (ibid.). Other subtitling software solutions with integrated speaker independent ASR systems have also started appearing in the market. The PerVoice Subtitling Workstation is one such software package offering an alternative way of working for live subtitle production, by combining the benefits of raw ASR output with a touch-screen interface for post-editing and the ability to switch to a respeaking mode as needed. VoxcribeCC is yet another solution which works for offline caption creation, prepopulating an editor with ASR output for the user to post-edit. There are also speaker independent ASR-powered systems, such as EnCaption, that are addressed to broadcasters and promise to do away with the costs of human captioners altogether, by linking the system to a station’s newsroom in order to access script information so as to build vocabularies and improve ASR performance. The Omnifluent Live Captioning Appliance is another such product available in a handful of languages, also addressed to broadcasters, which ensures high accuracy through customization of the ASR engines for speakers and language models based on domains, such as news, sports, etc. While none of these software packages have captured a large part of the market as yet and despite the fact that there are
significant disagreements over the span of time it will take for the intralingual subtitling market to be overhauled by machines, there is no question that speaker independent ASR has the power to truly disrupt this market.

Another innovative application of ASR in the intralingual subtitling market is its use for aligning available transcripts to audio, so as to produce automated time-coded caption files. Automatic identification of speaker turns and speakers is achieved with high precision, however poor text segmentation reduces the benefits achieved by time-stamping to the word level to make the resulting file of human captioner quality. There are arguments though for considerable savings in captioning effort to warrant the use of such tools, especially in the case of single-talking-head video material, and a number of platforms have appeared offering such services, e.g. 3PlayMedia, AudioAlign, eCaption, EZ-Sync, SyncWords, etc.

ASR is also used for media monitoring purposes, to check for the presence of captions to go with the right video in the right language, as well as their accurate synchronization. As such, it is increasingly used by broadcasters as an automated quality control (QC) tool, helping them maintain their quality standards while controlling their costs. The Nexidia QC tool (TVTechnology 2013), for instance, claims to automatically highlight and correct caption synchronization issues, and also offers an automated QC solution for audio description streams by analyzing the primary and secondary audio programs against each other and providing a video description coverage score.

These rapid developments in recent years show that ASR certainly has the potential to revolutionize the AVT industry further through improvements and innovations in its use. It is difficult to predict what the future holds for captioners and subtitlers. Perhaps their role will shift from text generation to QC and to the ‘arbiter of (human) taste and judgement’ (Maxwell 2014) or to content curation, enhancing caption data with useful metadata as content owners discover new ways to make value from their content (Padmore 2015).

**Machine translation**

As translation is growing from luxury to utility due to globalization, emerging markets, online content and technological advances (TAUS 2013), and considerations of quality are shifting from publishable quality to ‘fit for purpose’ (typically for content types characterized by low longevity and expedited timeframes in which translation needs to be produced, and also defined by the demographics of the target audience), the demand for automation in translation technology is growing. PBSMT, introduced into every household by Google through its Translate service (which as of May 2013 supported over 100 languages and served the needs of over 200 million users daily, according to Wikipedia), is yielding interesting results and translation enterprises have begun introducing it in their localization workflows, applying it in increasingly more types of material. The recent breakthrough in the field with NMT, introduced by Google and Systran in 2016 (Le and Schuster 2016, Systran 2016) and followed by Facebook, Microsoft and an increasing number of adopters, is creating a buzz in the MT research community as many new research avenues are opening up and further improvements in MT quality seem tangible.

Early experiments in applying MT in subtitle workflows consisted in developing systems that would produce machine-translated subtitles for broadcast. The first reported attempt was an English-Japanese system for the translation of part of the World News program of NHK’s satellite Channel 1 using the original news script (Sumiyoshi et al. 1995). Other such systems were developed later to translate closed captions. For over a decade now, TranslateTV (online)
has translated US English captions into LA Spanish using RBMT technology. The KANT system (Nyberg and Mitamura 1997) was developed for English into German translation of captions, using example-based MT; the ALTo system for English to Spanish and Portuguese (Popowich et al. 2000; Turcato et al. 2000); CaptionEye/KE for Korean to English (Yang et al. 2001) and CaptionEye/EK for English to Korean (Seo et al. 2001); and cTranie for multilingual closed caption translation from Korean to English, Chinese and Japanese for digital television (Yuh et al. 2006). The best known application of MT for automated subtitling is of course the integration of Google Translate in YouTube (Google 2009).

A number of European research projects have also attempted to tackle the issue of MT in offline subtitle workflows with the use of post-editing. MUSA and eTITLE were two such projects that took place after the turn of the century. MUSA integrated both a translation memory component and a machine translation engine (Piperidis et al. 2005), while eTITLE (Melero et al. 2006) experimented with the integration of multiple state-of-the-art language technologies (ASR, text compression, TM and MT) in tools aimed to aid subtitlers. The results of the eTITLE project showed improved quality output when TMs were integrated with MT (ibid.). The feasibility of translating subtitles aided by MT was also examined by an Irish project in 2006 (Armstrong et al. 2006a), followed by a preliminary study by O’Hagan (2003) on the use of TMs and MT in subtitling. Further work went into experimenting with the type of corpus that would yield the best results in terms of the MT output; it was shown that a corpus of homogeneous data had a significant impact on translation quality (Armstrong et al. 2006b).

These projects paved the way for more research on the subject. The most recent projects dealing with subtitles, funded by the European Commission and focusing mainly on European languages, are SUMAT, transLectures and TraMOOC. Both transLectures and TraMOOC deal with the machine translation of academic video lectures in the form of subtitles, so as to serve the needs of millions of students worldwide. SUMAT examines the application of machine translation in the subtitling industry and specifically in the entertainment domain. This involves arguably some of the hardest video materials to be translated by machines, as they draw on extralinguistic and contextual information, are characterized by multiple speaker changes, slang and register variation, and include a variety of genres. The project was carried out by a consortium of nine companies, four being subtitle agencies which provided high-quality in-domain parallel subtitle data from their archives, to be used as training data for the training of MT systems in seven bidirectional language pairs. A full year of human and automated evaluation was performed to assess the quality achieved by the MT and the post-editing effort required to turn the raw MT output into subtitles of publishable quality.

As this is the most extensive evaluation to date of MT output by translator end-users in a subtitling scenario of entertainment material, its results are of large significance in terms of the usefulness of MT and so as to inform corporate decisions about its application in subtitle workflows. The genres tested were drama films and series, documentaries, magazine programs and corporate talk shows, and the evaluators were professional subtitlers. They were asked to evaluate the quality of the MT indicating the post-editing effort that would be required to turn raw MT output to subtitles of publishable quality, and found over 50 per cent of the machine-translated subtitles to require little to no post-editing effort. Unsurprisingly, they were not as positive in their assessment of MT when confronted with the potential of having to use it in their daily work, possibly because of their preference towards translation work, or their lack of previous experience with post-editing, which might have made their experience more cumbersome leading them to underestimate the benefits of MT (Etchegoyhen et al. 2014: 47–49; 52–53).

One of the most important parts of the evaluation was the timed experiments performed to assess the productivity gain translators would have when post-editing MT output
versus translating the same material from a template file with no translation aid. The results varied considerably across different language pairs, even from translator to translator and from file to file within the same language pair, but the overall productivity result across all language pairs was a gain, with translators managing to produce approximately 40 per cent more subtitles than they would have in the same amount of time had they been translating from scratch (Etchegoyhen et al. 2014: 49–51). Post-editing effort in the MT of subtitles with respect to time was also investigated by Sousa et al. (2011) for the English–Brazilian Portuguese language pair, with results similar to the SUMAT findings. This is a significant finding in that it proves what many considered impossible: that MT can be of use in an open domain such as subtitles, even in the case of subtitles of varying genres of entertainment material.

Translator feedback during the SUMAT experiments indicated that the usefulness of the MT would be increased if combined with other CAT tools, such as TMs, and metrics indicating the expected MT quality, so that translators did not waste time dealing with MT output of substandard quality. Aside from the eTITLE project and O’Hagan’s early study (2003), the application of TMs in subtitling was also researched by Flanagan and Kenny (2007), Flanagan (2009) and Pérez Rojas (2014), while Mejías Moreno (2010) worked on creating a CAT tool for subtitlers called CATSUBS. Even where the research did not show particularly positive outcomes from the application of TMs in subtitling (e.g. O’Hagan 2003), functions inherent to TMs, such as the concordance search which looks up specific words or phrases in the TMs and displays their translations in context, were shown to be of use to translators. Such results strengthen the argument that the way forward for increased subtitle production lies with the integration of both CAT tools and MT in subtitle editors and stresses the importance of monetization of these assets that are often largely underused, hidden in the vast subtitle archives of client and LSP repositories.

Aside from the use of Google Translate for the translation of YouTube subtitles, there are other commercial applications of MT in subtitling workflows. Such use cases were reported among Scandinavian language pairs and English by Volk and Harder (2007), Volk (2008) and Volk et al. (2010). The interest of subtitling companies in MT seems to be on the increase as evidenced from questions received on the SUMAT experiments. It is no coincidence that we are already seeing the first integrations of machine translation in subtitle editors: Jubler (Nicholson and Musicman 2009) and Subtitle Translation Wizard (Castro n.d.) both use Google Translate, while Subtitle Edit integrates both Google Translate and Microsoft Translator. At the same time, subtitle formats and functionalities of subtitle editors are added to CAT tools such as Star TransitNXT.

A commercial application that combines both ASR and MT has also been described by Sawaf (2012). Omnifluent Media, part of AppTek’s suite of products, is a platform used for the creation of caption and subtitle files for live broadcast TV shows through the use of ASR and MT technologies, whilst keeping the human in the loop for quality assurance and optional post-editing purposes via a crowdsourcing workflow. A proof of concept of this system in the Arabic–English language pair reports results of throughput increase of transcribers and translators by a factor of 2 initially and 4 after training (ibid.). The product was implemented commercially in the Al Arabiya news channel in January 2014 (McLean 2014). Similarly, Omniscien’s Language Studio combines ASR and MT technologies for the purposes of semi-automating subtitle creation and translation. Language Studio has been used most recently by iflix (Larrieu 2018) to localize their catalogue of video material in 22 language pairs, reporting a 60 per cent productivity gain in the transcription step and 56 per cent in the translation step, while use of MT was also extended from subtitling to the process of translating scripts for dubbing.
Speech synthesis

The last language technology influencing the future of the AVT industry is speech synthesis, i.e. the automated rendering of Text To Speech (TTS). The use of speech synthesis has long been established as a means of providing accessibility to the blind and visually impaired in the form of spoken electronic program guides, or assistive technologies such as screen readers. It is now also slowly being applied in AVT domains that use audio as a deliverable, i.e. dubbing, voice-over narration, audio description and audio subtitling.

Audio Description (AD) is an accessibility service catering to the needs of the blind and partially sighted population in a number of countries and a growing requirement for broadcasters, though its high production costs have been a hindrance to its greater adoption. Its production involves two main parts in the workflow, the script writing and the recording process, both of which are aided today by the use of specially designed software packages developed by companies that typically develop subtitling software. Such software aids in the script preparation process, while it also offers a streamlined process for the recording of descriptions with no more than the use of a good microphone and a soundproof room for the recording to take place.

As the use of synthetic voice has been finding applications outside the disability market (e.g. in the gaming industry, in public transport, in interaction with mobile devices), and the quality of speech has been continuously improving, experimentation with the use of speech synthesis for AD purposes was inevitable. Reception studies conducted by Kobayashi et al. (2010) in Japan and the USA, Szarkowska (2011) in Poland, and Fernández-Torné and Matamala (2015) in Catalonia showed that TTS would be acceptable to blind and partially sighted patrons, even in cases where natural voices are still the preferred solution (ibid.). After all, according to Cryer and Home (2008), research studies suggest that blind and partially sighted people are likely to accept synthetic speech the more they are exposed to it and use it. Some may also prefer it if it helps them focus on the content of the text, while others would find it acceptable as a way of getting access to information faster, even if their preference lies with natural speech, (ibid.). A further factor to be taken into account when considering audience reception of synthetic speech is the maturity of the audience. In countries like the UK, where AD has been provided regularly with natural voices for over a decade now, it would probably prove harder for the audience to adjust and accept synthetic speech, as opposed to audiences that have not been conditioned to this type of service. The genre of the video material might also make a difference, as some productions might be more acceptable with ‘flat’ synthetic voices to sight impaired viewers than others.

Swiss TXT was one of the first companies to commission AD tracks with the use of text to speech. In an interview for the Languages and the Media 2012 conference, company representatives talked about their experience and stressed their belief that speech synthesis will become a standard component of AD in the years to come without being intrusive (Caruso and Linder 2012). SBS Discovery Media pioneered wide implementation of Speech Synthesis in their AD programming, and launched the SS AD service initially in Danish, Dutch, French, Italian, Polish and Swedish, and also in the UK in 2015 (Smith 2016). This has reportedly cut their AD costs by 56 per cent after initial equipment outlay (ibid.). Software providers, such as Cavena (Acapela 2014) and Screen Systems (2013) are also advertising integrated speech synthesis in their suite of products, offering it as an option to broadcasters both for AD and audio subtitling, the latter being used for additional accessibility purposes, mainly in countries where the dubbing practice is not widespread while subtitled foreign programming abounds.

The audio subtitling service in fact precedes the use of speech synthesis in AD, with the Dutch broadcaster NOS broadcasting audio subtitles with synthetic speech regularly since
2002 (Verboom et al. 2002), while the Swedish public broadcaster STV followed in 2005 (A-focus 2010, cited in Ljunglöf et al. 2012: 1). Audio subtitles are also in use in Finland since 2011 (European Federation of Hard of Hearing People 2011: 15) and currently offered for 100 per cent of Yle’s programming and 50 per cent for other channels (Viestintävirasto 2016). Audio subtitles were introduced in Denmark in 2012 (Thrane 2013), while there are reports on experiments with audio subtitles in the Czech Republic (see Hanzlicek et al. 2008; Matušek and Vít 2012). Audio subtitles are also broadcast in Estonia in two channels (ETV and ETV2) as of June 2013 (Mihkla et al. 2014:21).

Could speech synthesis improve to such an extent that it could be used for voice-over narration and lip-sync dubbing as well? Some claim it can. The Dubbitron system by Speech Morphing was advertised a few years ago to offer language dubbing for entertainment material at a fraction of the cost, although the solution did not fly. MediaWen are now offering automated voice-over and dubbing by integrating IBM Watson’s APIs in their STVHub platform. The company also partnered with XPERTEAM, an e-learning expert, to launch WBT VIDEO Learning, a collaborative video learning authoring software that makes use of language technologies such as speech recognition, translation memory, computer assisted translation and synthetic speech generation (Mediawen 2016). At the same time, Disney’s Research team in collaboration with the University of East Anglia are developing a program of ‘automated video redubbing strategies’, which finds word sequences that can plausibly match the actors’ speech motions, so speech redubbing is not obvious to the viewer (Choi 2015).

Conclusion

The second decade of the twenty-first century is defined by connectivity; information is accessible everywhere and it is increasingly audiovisual. TV is becoming yet another application on the Internet, with new OTT platforms announced almost every day (Huggers 2015), while the Federal Communications Commission is putting broadband policies in place. The AVT industry finds itself at a strategic inflection point, as both television and audiovisual content are being redefined, content volumes are growing exponentially, while the world is asking for access at all levels, immediately. Technology solutions are required that are able to transcribe, translate and index vast amounts of video content at the quality and speed that the market demands, and at costs it will support.

Cloud-based platforms are increasingly adopted by AVT enterprises, fansubbing is growing, subtitling is applied extensively to non-entertainment content, often with maximum flexibility as regards respect to the norms of the practice to date, while new entrants in the market feature managed crowdsourced solutions of amateurs and professionals. Meanwhile, experimentation with mature language technologies in the AVT market is flourishing and the promise to deliver solutions that make good use of valuable data and metadata hidden away in large content repositories is reinforced. Such developments have the potential to reshape subtitling workflows, client and LSP relationships, affect audience behaviour and challenge established norms, creating a paradigm shift in the industry. TAUS (2013) talks of a ‘convergence era where technologies, such as speech, search and others will continue to be combined with machine translation to create new solutions’. Image recognition is an example of another technology whose value could be maximized if combined with speech recognition in videos, for instance. Could it be that the unveiling of Skype Translator (Skype Team 2015) by Microsoft and the idea of speech translation, i.e. the combination of ASR, MT and TTS to automatically translate audio from one language to another, is a sign of things to come in the future of AVT?
Further reading


Georgakopoulou, P. and L. Bywood (2014) ‘MT in Subtitling and the Rising Profile of the Post-Editor’, Multilingual Computing, January/February: 24–28. This article presents the SUMAT case study and its main findings regarding the application of machine translation in subtitling, and makes an initial attempt at detailing the profile of the post-editor, which is expected to be the next job profile in demand in the subtitling industry.

Hearne, M. and A. Way (2011) ‘Statistical Machine Translation: A Guide for Linguists and Translators’, Language and Linguistics Compass 5(5): 205–226. This paper presents the basic principles underpinning Statistical Machine Translation, the dominant approach in machine translation research until recently, with multiple commercial applications, so that linguists better understand how their input is being used by researchers in building and improving SMT systems.

TAUS (2017) The Translation Industry in 2022. Available online: https://www.taus.net/think-tank/reports/event-reports/the-translation-industry-in-2022 [last access 20 December 2017]. A report outlining predictions (including the acceleration of automation) for the future of the translation industry over the following 5 years.

Related topics

2 History of audiovisual translation
3 Subtitling on the cusp of its futures
6 Subtitling for deaf and hard of hearing audiences: moving forward
7 Respeaking: subtitling through speech recognition
8 Audio description: evolving recommendations for usable, effective and enjoyable practices
10 Game localization: a critical overview and implications for audiovisual translation
20 Corpus-based audiovisual translation studies: ample room for development
21 Multimodal corpora in audiovisual translation studies
23 Audiovisual translation and audience reception

References


Technologization of audiovisual translation


Technologization of audiovisual translation


TAUS (2013) Translation Technology Landscape Report. Available online: https://www.scribd.com/fullscreen/134651594?access_key=key-2hr0yylsh8moed9i85yf&allow_share=true&escape=false&view_mode=scroll [last access 20 December 2017].


Sitography

3PlayMedia: http://www.3playmedia.com/ [last access 20 December 2017].
Aegisub: http://www.aegisub.org/ [last access 20 December 2017].
Amara: https://amara.org/en [last access 20 December 2017].
Amazon Mechanical Turk: https://www.mturk.com/mturk/welcome [last access 20 December 2017].
Annotation Edit: http://www.zeitanker.com/content/download/annotation_edit_demo/ [last access 20 December 2017].
AudioAlign: https://www.audioalign.com [last access 20 December 2017].
CANAvaS: http://canvass.co.jp/global/corporate/index.html [last access 20 December 2017].
Cavena: http://www.cavena.com [last access 20 December 2017].
Cielo24: https://cielo24.com/ [last access 20 December 2017].
Conyac: https://conyac.cc/en [last access 20 December 2017].
Coursera Community: http://www.coursera.community [last access 20 December 2017].
DotSub: https://dotsub.com [last access 20 December 2017].
Dubbitron: http://www.innuva.com/projects/dubbitron/ [last access 20 December 2017].
eCaption: http://optispeech.com/ecaption/ [last access 20 December 2017].
EnCaption: http://www.enco.com/products/encaption4.html [last access 20 December 2017].
eTITLE Project: http://cordis.europa.eu/project/rcn/78351_en.html [last access 20 December 2017].
FAB: http://www.fab-online.com/ [last access 20 December 2017].
EU-Bridge: https://www.eu-bridge.eu [last access 20 December 2017].
EZ-Sync: http://www.lnts.com/about-ez-sync [last access 20 December 2017].
EZTitles: http://www.eztitles.com [last access 20 December 2017].
Gaupol: https://otsaloma.io/gaupol/ [last access 20 December 2017].
Gengo: http://gengo.com [last access 20 December 2017].
Gnome Subtitles: http://gnome-subtitles.sourceforge.net/ [last access 20 December 2017].
iMediaTrans: http://www.imediatrans.com/ [last access 20 December 2017].
Jubler: http://www.jubler.org/ [last access 20 December 2017].
KANT Project: http://www.lti.cs.cmu.edu/Research/Kant/ [last access 20 December 2017].
MediaWen: http://mediawen.com/#solution [last access 20 December 2017].
Movie Captioner: http://www.synchrimedia.com/#movcaptioner [last access 20 December 2017].
MovieReading App: https://itunes.apple.com/us/app/moviereading/id460349347?mt=8 [last access 20 December 2017].
MUSA Project: http://sifnos.ilsp.gr/musa/ [last access 20 December 2017].
Nexidia: http://www.nexidia.com/ [last access 20 December 2017].
Omnifluent Media: http://omnifluent.com/ [last access 23 August 2017].
Omniscien’s Language Studio: https://omniscien.com/media-subtitling/ [last access 20 December 2017].
OOONA: http://ooona.net [last access 20 December 2017].
Plint: http://www.undertext.se/plint/ [last access 20 December 2017]
SAVAS: http://cordis.europa.eu/project/rcn/103572_en.html [last access 20 December 2017].
Screen Systems: https://subtitling.com/ [last access 20 December 2017].
Sfera: https://www.sferastudios.com [last access 20 December 2017].
SoftNI: https://softni.com/about-us/ [last access 20 December 2017].
Spot Software: http://www.spotsoftware.nl [last access 20 December 2017].
Starfish Technologies: www.starfish.tv [last access 20 December 2017].
Star TransitNXT: https://www.star-group.net/en/products/translation-and-localization.html [last access 20 December 2017].
Subtitle Edit: http://www.nikse.dk/subtitleedit/ [last access 20 December 2017].
Subtitle Editor: http://home.gna.org/subtitleeditor/ [last access 30 April 2015].
Subtitle Workshop: http://www.urusoft.net/products.php?cat=sw [last access 20 December 2017].
SUMAT: http://www.fp7-sumat-project.eu [last access 20 December 2017].
SyncWords: https://www.syncwords.com/ [last access 20 December 2017].
SysMedia (now part of Screen Systems): https://subtitling.com/screen-subtitling-acquires-sysmedia/ [last access 20 December 2017].
Telestream: http://www.telestream.net/captioning/overview.htm [last access 20 December 2017].
TitleVision: http://titlevision.dk/ [last access 20 December 2017].
TraMOOC: http://tramooe.eu [last access 20 December 2017].
TranslateTV: http://www.translatetv.com/ttv.php [last access 20 December 2017].
transLectures: http://www.translectures.eu/ [last access 20 December 2017].
Unbabel: https://unbabel.com [last access 20 December 2017].
Viki: http://www.viki.com [last access 20 December 2017].
VoxcriebCC: http://voxcribe.com/About%20Voxcribe.html [last access 20 December 2017].
ZOOsubs: http://www.zoodigital.com/platforms/zoosubs [last access 20 December 2017].