

- The subsequent advent of televised broadcasting as a private commodity and its transformation into a mass medium of communication and entertainment provided new outlets for the dissemination of audiovisual translated texts. In keeping with this development in the global media ecology, terms such as ‘film and TV translation’ (Delabastita 1989) and ‘media translation’ (Eguíluz et al. 1994) came to feature more prominently in the specialized literature.
- More recently, the computerization of audiovisual texts, i.e. the popularization of electronic and/or digital media content to be viewed on the screen of a computer or a portable electronic device, has introduced some new terminology in the field. Terms such as ‘screen translation’ (Mason 1989, O’Connell 2007) and ‘multimedia translation’ (Gambier and Gottlieb 2001) are indicative of the extent to which audiovisual translation has grown far beyond its original disciplinary boundaries, bringing the investigation of ever new textual manifestations into its research remit.
- At different points during this period, specialists have also proposed terms that signal the medial restrictions (e.g. limited number of characters available to translate a text) under which audiovisual translators operate, e.g. ‘constrained translation’ (Titford 1982, Mayoral et al. 1988). From a complementary perspective, others have opted to foreground the transformative dimension of audiovisual mediation, as translators often need to distance themselves from the original text to cater for new audiences – e.g. Gambier’s (2003) ‘transadaptation’ – or preserve the immersive function of media content – e.g. Bernal Merino’s (2006) ‘transcreation’.

Amid this terminological inflation, Díaz Cintas and Remael have recently suggested that the term ‘audiovisual translation (AVT) has been gaining ground in recent years and is fast becoming the standard referent’ (2007: 11–12).

1.2 MODALITIES OF AUDIOVISUAL TRANSLATION

1.2.1 Subtitling

Widely regarded as an evolved version of the primitive intertitles (Section 2.2), subtitles are snippets of written text superimposed on visual

footage that convey a target language version of the source speech. Conventionally, each of the snippets into which the original speech – whether in the form of dialogue or narration – is divided for the purposes of translation has to be delivered in synchrony with the corresponding fragment of spoken language. In communities where several linguistic constituencies co-exist, ‘bilingual subtitles’ convey two language versions of the same source fragment, one in each line of the subtitle (Gambier 2003).

Subtitling has been defined as a ‘diasemiotic’ or ‘intermodal’ form of audiovisual translation (Gottlieb 1997: 95), as it involves the shift from a spoken to a written medium. Given that ‘people generally speak much faster than they read’ (O’Connell 1998: 67), this shift has important consequences for viewers’ experience of translated audiovisual texts. The empirical observation that, in order to match the temporalities of speaking and reading, subtitles can only accommodate 60 per cent of the source spoken text (de Linde and Kay 1999: 51) explains why they are normally worded as condensed, streamlined versions of the original dialogue (Karamitroglou 1998). These medial restrictions, that were first articulated in the 1930s (Section 2.3), dictate that the number of characters used in each subtitle should be commensurate with the duration of the corresponding speech unit and the reading speed of the target audience (Díaz Cintas and Remael 2007).

As discussed in Sections 2.4 and 4.5.2, the subtitlers’ adherence to industry validated standards – notably, the condensation and synthesis of the original spoken dialogue – has been shown to compromise the **interpersonal pragmatics** of subtitled dialogue. For example, audiovisual translators tend to disambiguate instances of conversational ambiguity or indirectness that may have played an important narrative role in the original text (Hatim and Mason 1997, Mason 2001). It has also been suggested that conforming to these constraints often proves detrimental to the dynamics of dramatic characterization envisaged by the creator of the original audiovisual text, particularly in the case of films or TV drama (Remael 2003). As a result, characters’ personalities may not be perceived by viewers as the film director had intended. It is therefore not surprising that independent film directors have raised criticisms against the creative limitations derived from these constraints (see Section 2.4). As illustrated in Box 1.2, deconstructing commercial subtitling standards represents an effective strategy for creators to signal their resistance to mainstream subtitling practices.

Box 1.2

As part of his study on ‘accented’ or diasporic films, Naficy (2004: 145–46) reviews the creative use of subtitles by Vietnamese film director Trinh T. Minh-ha.

Trinh’s *Surname Viêt Given Name Nam* uses superimposed titles and subtitles extensively, graphically, and critically. Their large numbers and varied contents and layout give this film a truly calligraphic accent. Throughout, subtitles consisting of the translation of the film’s dialogue and voice-over and of Vietnamese poetry and proverbs are displayed, as is customary, in the lower third of the screen. However, on many occasions, what the diegetic women say in Vietnamese or in heavily accented English is superimposed in different layouts, as blocks of English text on various regions of the film frame, including over the characters’ faces. These graphic titles, or what Trinh calls ‘visualized speech’, act as traditional subtitles by aiding spectator comprehension. However, they also serve other graphic, critical, and deconstructive functions . . . To these text-based complexities must be added Trinh’s filming style that in *Surname Viêt Given Name Nam*, like in her other films, violates many of the norms of cinematic realism as a critique of those norms. For example, in some sequences she places the subjects on the margins of the frame or decentres them by panning away from them. Close-up shots that would normally show the subjects’ full-face end up cutting off part of their faces. The film also subverts the accepted practices of lip-synching and title synchronization. Extra long or short duration titles draw attention to themselves and to the spectatorial readerly activities that are involved.

But the capacity to develop new subtitling conventions is not the sole privilege of text producers. Recent changes in the audiovisual landscape, notably the development of digitization techniques and new models of distribution and consumption of audiovisual products, have provided the impetus for the spread of transformational practices. Generally speaking, these new subtitling cultures aim to expose the expressive limitations of mainstream subtitling conventions. Fansubbing, namely the subtitling of television drama and films by networked fan communities, seeks to redress the shortage and cultural insensitivity of commercial translations (Section 3.3). Using a range of daring formal conventions, fansubbers produce and distribute their own subtitled versions through Internet-based channels and, in so doing, provide their fellow fans with a more ‘authentic’ and meaningful spectatorial experience. Subtitling is also being increasingly appropriated by groups of politically engaged amateurs to undermine the socio-economic structures that sustain global capitalism and/or to effect

social and political change (Section 7.3). As is also the case within fansubbing groups, members of politicized amateur cybercultures engage in complex negotiations of affinity: exploring their collective stance on aesthetic and political issues within their virtual communities brings them closer together.

In a separate development, technology-driven changes are increasing the need for subtitles and, by extension, creating a demand for a more 'efficient' approach to subtitling. This expansion is motivated by a number of reasons. For one thing, media content can now be accessed via a growing range of digital distribution channels, including video-sharing websites and on-demand Internet streaming media platforms. On the other hand, subtitling is now being used to mediate an expanding range of texts and genres. In addition to films and drama, subtitling is gaining popularity as the method of choice for the translation of utilitarian audiovisual texts, including promotional corporate films, videogames or documentaries. The fact that some of these screen-based texts make a relatively formulaic or routinized use of language represents an additional stimulus to automatize the production of subtitles.

Drawing on O'Hagan's (2003) pioneering work, Armstrong *et al.* (2006) have attempted to gauge the viability of Example-Based Machine Translation (EBMT) to the translation of DVD subtitles and bonus material for English-German and English-Japanese. Essentially, EBMT systems reuse examples of already existing translations, collected in the form of bilingual or parallel corpora, as a basis for the automated production of new translations. By combining real user evaluation studies with automatic metrics, Armstrong *et al.* concluded that 'the EBMT trained with homogeneous data [i.e. existing DVD subtitles] is likely to contribute to a higher translation quality than with heterogeneous data' (2006: 179). More recently, Volk *et al.* (2010: 53) have also presented a plausible rationale for the automation of subtitling in the Scandinavian context, where 'large amounts of TV subtitles are needed in Swedish, Danish and Norwegian'. Driven by the insight that subtitles are not always created independently for each language 'for efficiency reasons', the work of Volk *et al.* appraises the viability of automated subtitles generated by a statistical machine-translation system from Swedish to Danish and Norwegian. As was also the case with its EBMT counterpart, the statistical-machine translation system developed by Volk *et al.* is 'trained' by using large corpora of high-quality human translated subtitles in the different working languages, and assessed both by automatic metrics and real users.¹

The production and reception of 3D subtitles is another area of professional practice and scholarly research influenced by technological and industrial advances that is bound to attract considerable attention over the next years. According to Kozoulyaeu (n.d.), '3D subtitling calls for tectonic changes in all aspects of production and post-production of stereographic films and TV programmes', insofar as '[i]t requires new plots, new shooting approaches, new conventions and new workflows that will profoundly change the industry'. The basis for Kozoulyaeu's claim is that working conventions developed for the delivery of 2D subtitles back in the 1930s are incompatible with the immersive viewing experience that 3D movies are meant to deliver. A straightforward application of 2D conventions may, for example, cause subtitles to collide with objects located within the three-dimensional diegetic space. In order to enhance the spectatorial experience and reduce eye strain, 3D subtitles should be dynamic, i.e. able to follow the main point of interest as it moves around the scene. Decisions pertaining to the depth position of 3D subtitles are equally important; in this environment, the size of subtitles should be scaled and adjusted to the different positions that subtitles occupy within the three-dimensional space at each point of the scene. As '3D subtitles interact with the film (and game) environment in a much stronger fashion than 2D subtitles and become elements of the action and shots themselves', Kozoulyaeu (ibid.) argues, 3D subtitling constitutes 'a strong artistic undertaking'.

1.2.2 Revoicing

'Revoicing' is a generic term encompassing a range of spoken translation methods, including simultaneous interpreting, free commentary and narration (see Box 1.3) (Luyken et al. 1991, Baker and Hochel 1998). Although, technically speaking, voice-over and lip-synchronized dubbing are also types of revoicing, they are often dealt with and described separately.

In voiced-over or 'half-dubbed' (Gambier 2003: 173) material, the original and translated soundtracks co-exist and overlap. Typically, only the original voice track is audible at the beginning. A few seconds into the programme, however, the volume of the source dialogue or narration is lowered, while the translated voice track becomes acoustically salient until the end.² As the original sound remains vaguely audible throughout, this method of pre-recorded revoicing is widely considered to enhance the realism of voiced-over programmes (Luyken et al. 1991: 140). As a result, in

Box 1.3

Simultaneous interpreting is normally a live versioning method used to translate films or documentaries in settings where more elaborate forms of revoicing are not an option, typically due to time or funding constraints (e.g. film festivals). The interpreter, who may occasionally have to perform without accessing the relevant script, normally dubs the voices of the whole film cast (Lecuona Lerchundi 1994).

Free commentary, a revoicing technique performed on the spot by presenters or commentators broadcasting a high profile event, is 'clearly an adaptation [of the original voice track content] for a new audience, with additions, omissions, clarifications and comments' (Gambier 2003: 174). Delivered with a spontaneous tone, free commentary seldom aims to convey a faithful rendering of the original speech.

Narration is a live or pre-recorded form of oral transfer aiming 'to provide a summarized but faithful and carefully scripted rendition of the original speech'. Normally, 'its delivery is timed so that there is no clash with the visual syntax of the programme' (Pérez-González 2009: 16), as defined by the structural organization and the transitions between the constitutive units of the audiovisual text.

audiovisual markets dominated by subtitling and dubbing, voice-over is conventionally used for the translation of televised interviews, documentaries and other forms of audiovisual content where lip-synchronization is not of the essence.

Significantly, the fact that voice-over is 'a cheap alternative to dubbing' (O'Connell 2007: 124) explains why it is the audiovisual transfer method of choice for the translation of films in former Communist states and some countries in the Middle East and Asia (e.g. Iran or Thailand). The geographical dispersion of voice-over practices has resulted in a certain degree of variation as to the way in which this audiovisual transfer method is actually executed in each market. In Poland, for example, voice-over narrators or *lektors* 'provide a "whispering" translation, thus allowing the audience to hear the original voices to the greatest extent possible' (Szarkowska 2009: 187). Indeed, convention dictates that Polish (normally male) *lektors* should adopt a self-effacing presentational style and that no attempt should be made to mimic the prosodic and material features of the diegetic speech. By contrast, in Iran, where voice-over was originally used to revoice both Iranian and foreign films, the narrator's acting tends to be contagious and exuberant, as explained by Naficy (2003: 189):

Since Iranian movies were usually not filmed in sync during production, skillful [sic] and versatile voice-over artists dubbed the actors' voices in postproduction. It can be safely said that approximately two dozen voice actors dubbed the majority of Iranian and foreign films shown in the country. Audiences drew special pleasure from hearing John Wayne and Jerry Lewis use expressions that Iranian tough guys or comedians used. Such intertextuality hybridized the diegeses and the characters who inhabited them. Also, because each voice-over artist often dubbed the voices of a number of different characters, strange cross-over resonances and dissonances would be set up between voices and characters, which could serve to confuse the mirror-phase identification with individual characters.

The wide range of perceptual and technical issues at stake in the production of voiced-over versions, and the varying conventions associated with the execution of this audiovisual transfer technique in different locales are explored in depth by Franco *et al.* (2010).

In terms of future research, voice-over scholarship is likely to benefit from parallel developments in the study of audio description, for both transfer methods involve the incorporation of an additional voice track to the multimodal ensemble. As Woźniak (2012: 226) notes,

certain audio description principles could be applied successfully to voice-over to make it more unobtrusive and to reduce interference with the original soundtrack, while an in-depth analysis of . . . voice-over practices might have a positive impact on the study of audio subtitles and generally on media accessibility.

Ultimately, scholars in both fields are interested in understanding how narrators can best assist viewers in making narrative meaning from the film by themselves.

Once hailed as the supreme and most comprehensive form of translation (Cary 1969: 110), lip-synchronized dubbing involves 'the replacement of the original speech by a voice track which attempts to follow as closely as possible the timing, phrasing, and lip movement of the original dialogue' (Luyken *et al.* 1991: 311). When it was first developed in the 1930s (Section 2.3), the main priority for the film industry was producing target language dialogue that would perfectly fit the actors' lips, in what came to

be known as voice ‘doubling’ and/or ‘ghosting’ (Đurovičová 2003). For many years, this practice was surrounded by a certain mystique (Doane 1980). Intriguingly, voice was regarded as a revolving door between the diegetic and real worlds, on the one hand, and between the systems of values inscribed in the source and target cultures, on the other. To some extent, lip-synch dubbing still carries some of these connotations, as illustrated by the confusion surrounding Indonesia’s audiovisual translation policies in the late 1990s. In 1996, the Indonesian government announced that ‘foreign films on television should no longer be broadcast in their original language version with Indonesian summaries or subtitles but were to be dubbed into Indonesian’ (Boellstorff 2003: 235). However, only a few months later, President Soeharto reversed his decision on the grounds that

[d]ubbing can create gaps in family communication. It can ruin the self-image of family members as a result of adopting foreign values that are ‘Indonesianized’ [*diindonesiakan*] . . . This can cause feelings of becoming ‘another person’ to arise in family members, who are in actuality not foreigners . . . Whenever Indonesians view television, films, or other broadcasts where the original language has been changed into our national language, those Indonesians will think that the performances in those media constitute a part of themselves. As if the culture behind those performances is also the culture of our people.

(Ali 1997:341–42, quoted in Boellstorff *ibid.*)

Over the last two decades, a number of studies (Herbst 1994: 244–45, Chaume Varela 2004: 155–65, Chaume Varela 2012) have widened the scope of dubbing scholarship beyond the impact of lip synchrony constraints – e.g. the tendency of dubbed language to neutralize markers of sociolinguistic and cultural variation present in the source text (Pavesi 2005). In addition to the industry’s demand for equivalence between the original and translated speech, and for matching articulatory movements in the delivery of the source and target dialogue, today’s dubbing specialists are urged to be mindful of the film’s visual syntax, narrative pace, relevant medial restrictions and, more importantly, the multimodal distribution of meaning across different semiotics. This last aspect is explored in depth by Chaume Varela (2004), who argues that meaning is conveyed to viewers through the acoustic and visual channels along two clusters of semiotic codes. According to this specialist, three different codes or sign systems are realized in the

acoustic channel: spoken language (what is said), para-verbal signs (how it is said) and non-verbal acoustic signs (e.g. music, special effects, incidental sound, etc.). For their part, the visual channel is shaped by signs pertaining to the photographic code (e.g. colour, light and perspective), the iconographic code (e.g. symbols and icons) and the mobility code – involving the positioning (proxemics) and movements (kinesics) of the film characters. Ultimately, Chaume Varela foregrounds the importance of the non-verbal dimensions of film and illustrates how the matching of non-verbal semiotics in the original and target texts can make an important contribution to dramatic characterization and/or artistic integrity.

A large proportion of dubbing scholarship has shifted its emphasis away from the relationship of linguistic equivalence between the source and target text, focusing instead on the role and position of the target text as part of a dynamic system (translated fictional speech) of its own. Drawing on corpus-based models of translation (Sections 4.4.2 and 5.3.4), dubbing scholars have tried to gain a better understanding of the typicalities of dubbed speech and identify the defining features of ‘authentic’ or ‘spontaneous sounding’ dubbed dialogue (Romero-Fresco 2006, 2009; Pavesi 2009a, 2009b) vis-à-vis non-translated fictional speech. As Pavesi (2009a: 209) explains, these defining features are closely related to the structure of film discourse and ‘norms emerging both in simulated spoken [language] and film translation’.

As was also the case with subtitling, advances in digital communication technologies have led to the emergence of amateur dubbing or fandubbing subcultures (see Section 7.1). Beyond this fandom-driven development, experimental work on speech and image synthesis technologies has explored the feasibility of (i) transferring the physical qualities of actors’ voices across languages; and (ii) automating the production of audiovisual footage where actors are made to ‘mouth’ speech they did not actually utter (Box 1.4).

Box 1.4

Voxworks Technologies’ ReelVoice system: voice morphing (Bloom 2002)

Voxworks Technologies’ system creates . . . sound-alikes, solving a huge headache for those who create foreign-language tracks, have to cover up curse words or otherwise clean up and loop audio.

CEO Elio Zarmati says there’s no bigger hassle in dubbing than finding sound-alikes. Normally, a studio must find a sound-alike for each actor to

loop fill-in words for a film's TV-safe version. And when creating a foreign-language version, the sound-alike must also be fluent in that language, further narrowing the range of options. With Voxworks' ReelVoice, though, the voice-over talent need only be fluent in the translated language, and able to provide the actor's timing that distinguishes any talented performer. The technology takes care of the rest, 'morphing' the voice-over to sound like any or all of the original performers.

[. . .]

To work, the technology builds a library of phonemes, the little bits of words spoken in a performer's distinctive way, for both the star voice being morphed and for the voice-over actor. The technology then maps those two libraries of phonemes to each other.

Video Rewrite: visual morphing³

Video Rewrite uses existing footage to create automatically a new video of a person mouthing words that she did not speak in the original footage. This technique is useful in movie dubbing, for example, where the movie sequence can be modified to sync the actors' lip motions to the new soundtrack.

Video Rewrite automatically labels the phonemes in the training data [i.e. previous recordings of actors uttering certain sounds] and in the new audio track [and] reorders the mouth images in the training footage to match the phoneme sequence of the new audio track. When particular phonemes are unavailable in the training footage, Video Rewrite selects the closest approximations. The resulting sequence of mouth images is stitched into the background footage. This stitching process automatically corrects for differences in head position and orientation between the mouth images and the background footage. . . . The new video combines the dynamics of the original actor's articulations with the mannerisms and setting dictated by the background footage.

1.2.3 Assistive forms of audiovisual translation

Over the last two decades, audiovisual translation has contributed to enhancing the social integration of minorities all over the world. Freely available subtitled programmes, for instance, have long helped immigrants to develop their proficiency in the language spoken by their host communities. But the role of audiovisual translation as an empowerment tool has become particularly evident with the emergence and consolidation of relatively new forms of intersemiotic assistive mediation, such as subtitling for the hard of hearing or audio description, that aim to facilitate access to information and entertainment for sensory impaired members of the community. Capitalizing on the storage capacity of DVDs and Blu-ray discs,

media companies are able to release audiovisual products aimed at mainstream viewers, while simultaneously allowing sensory impaired audiences to access the media content in assistive mode – i.e. viewing the film in combination with dedicated audio or video tracks. Similarly, on-demand Internet streaming media platforms and most TV sets also provide viewers with access to subtitles and audio narrations.

Subtitling for the hard of hearing provides a text display of the characters' speech interspersed with written descriptions of sound features from the diegetic action that would otherwise not be accessible to Deaf viewers. This transfer of information from speech to written subtitles involves the deployment of dedicated mediation conventions in relation to the colour, spotting and positioning of the subtitles (Neves 2005). Subtitles for the hard of hearing were originally available only for films and pre-recorded broadcasts. However, the enactment of a growing number of legislative and regulatory provisions since 1980 (Remael 2007) has set minimum quota requirements for the provision of live subtitling by broadcasters and subtitling companies. Against this backdrop, the audiovisual industry has spearheaded the development of live subtitling practices.

Real-time subtitling for the deaf and hard of hearing, often known as *respeaking*, is

a technique in which a respeaker listens to the original sound of a live programme or event and *respeaks* it, including punctuation marks and some specific features for the deaf and hard of hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay.

(Romero-Fresco 2011: 1)

In addition to refining the workflow processes and the technological tools that shape and facilitate the *respeaking* process, a growing body of research is seeking to gain a better understanding of *respeaking*-specific translation techniques, e.g. the *respeakers*' reduction and omission of the original text (van Waes *et al.* 2013).

Audio description, a spoken account of those visual aspects of a film that play a role in conveying its plot, has become equally important in ensuring the accessibility of audiovisual products to the visually impaired. While transferring information from the visual to the acoustic channel – i.e. from images to the spoken narration that a voice delivers between the stretches of

spoken dialogue – the audio describer ‘engages in a delicate balancing exercise to establish what the needs of the spectator may be, and to ensure the audience is not overburdened with excessive information’ (Pérez-González 2009: 16). Ongoing work on audio description has significantly widened the range of theoretical and methodological frameworks informing translation studies research; its originality and rigour are thus foregrounded in Sections 4.3.2 and 5.2.

1.3 AUDIOVISUAL TRANSLATION AS AN ACADEMIC DISCIPLINE

The **institutionalization** and **academization** of audiovisual translation,⁴ two intertwined processes whose outcome is still uncertain and contested, have been arrested at different points by the dynamic nature of its object of study. The varying conceptualization of audiovisual texts in the literature is a good example of the obstacles found along the way:

- Consider the case of pioneering literature on audiovisual translation (Caillé 1960, Cary 1969), which predates the constitution of translation studies as a discipline in its own right. This body of literature was written at a time when films, the only consolidated form of mass entertainment, were presented as objects of artistic contemplation rather than simple cultural commodities. Audiovisual translation was thus conceptualized as an aesthetic form of mediation subject to a range of extra-linguistic constraints pertaining to the medium in which films are embedded.
- The consideration of audiovisual texts as hybrid textualities, influenced by both linguistic and extra-linguistic factors, is consistent with their treatment elsewhere – e.g. in early attempts to classify texts and formalize their functional characteristics as well as the translation methods typically associated with each text type. This is the case with the ‘Audio-mediale Texte’ category in Reiss’ (1971/2000) taxonomy, although ‘her definition seemed to refer more to fields such as advertising rather than film and documentary translation’ (Munday 2012: 268).
- As recently as the late 1980s and early 1990s, new efforts to chart the map of translation studies (Snell-Hornby 1988/1995, Bassnett 1980/1991) placed audiovisual texts in the orbit of literary translation. Bassnett, for example, argued that the fourth category of her taxonomy, ‘loosely called Translation and Poetics, includes the whole area of