Voice-over
Practice, research and future prospects
Anna Matamala

Introduction
Dubbing and subtitling have been the object of extensive research in audiovisual translation (AVT) (Chaume 2012, Díaz Cintas and Remael 2007) due to the popularity of these modalities in many countries. However, there is a third audiovisual transfer mode that has not been analyzed in such detail, but which nonetheless is extensively used in many audiovisual markets: voice-over. Sometimes termed the ‘ugly duckling’ of audiovisual translation (Orero 2006b), a ‘damsel in distress’ (Wozniak 2012: 211) or even an ‘orphan child’ (Bogucki 2013: 20), many academic and non-academic voices have drawn attention to the limitations of voice-over (Glaser 1991, Tomaszkiewicz 2006, Garcarz 2007), and have accounted for the fact that it continues to be used in some countries in terms of the low costs that it incurs. However, voice-over is a reality accepted by many audiences, and its academic study has increasingly captured the attention of translation scholars.

This chapter aims to define and categorize different varieties of voice-over, and see how this transfer mode is used in both fictional and non-fictional audiovisual genres. Beyond the realm of practice, attention will be paid to research issues in voice-over, focusing on synchronization constraints, aspects of linguistic and cultural mediation, manipulation and translator’s visibility, the technologization of voice-over practices, voice-over training and reception research.

Defining voice-over
The term ‘voice-over’ is used in translation studies, on the one hand, and film studies and the film industry, on the other (Franco 2001a). Within the field of translation studies, Díaz Cintas and Orero (2006: 477) define voice-over as a ‘technique in which a voice offering a translation in a given target language (TL) is heard simultaneously on top of the source language (SL) voice.’ The authors also indicate that ‘the volume [of the original programme] is reduced to a low level that can still be heard in the background when the translation is being read’ (ibid.). They also highlight that it is ‘common practice to allow the viewer to hear the original speech in the foreign language at the onset of the speech and to reduce subsequently
the volume of the original so that the translated speech can be inserted' (*ibid.*). This translation, according to the same authors, ‘usually finishes several seconds before the foreign language speech does, the sound of the original is raised again to a normal volume and the viewer can hear once more the original speech’ (*ibid.*).

On the other hand, within the field of film studies, Kuhn and Westwell (2012: 446–447) define voice-over quite differently, as the ‘voice of an offscreen narrator or a voice heard but not belonging to any character actually talking on screen’ (*ibid.*). They indicate that in ‘newsreels and documentary films a voice over will most commonly consist of a commentator (who may occasionally appear intermittently on screen) who provides third-person overview that orientates the viewer to what they are seeing (this kind of voice over is sometimes referred to as voice-of-God narration)’ (*ibid.*). As for fiction films, according to the same authors, ‘voice overs can take various forms’, and may be used to ‘convey the interior thoughts of a character seen on screen’ (*ibid.*). They also point out that ‘DVD releases are now often made available with directors, other creative players, or film scholars, providing a voice over that comments on the onscreen action’ (*ibid.*). ‘Voice-over’ is also used in the expression ‘voice-over artists’, which is widely used in the industry to refer to voice talents or voice actors reading commercials, audiobooks or dubbing movies, to give but three examples.

**Voice-over vis-à-vis other audiovisual transfer modes**

Voice-over has been conceptualized as a type of revoicing (Luyken *et al.* 1991: 80, Chaume 2004: 35), a type of dubbing (Dries 1995: 309), and even a type of interpreting (Gambier 1996: 9). This chapter endorses Pérez-González’s view (2014: 19) that whilst ‘technically speaking, voice-over and lip-synchronized dubbing are types of revoicing, they are often dealt with and described separately.’

Voice-over and simultaneous interpreting share the use of a superimposed voice to deliver the translation on top of its original counterpart. However, interpreting is generally produced live, without drawing on a written translation of the original content. By contrast, in voice-over, the translator works from pre-recorded material to create a written translation that is then generally read aloud by a voice talent. It must be acknowledged, though, that voice-over may have its origins in the simultaneous interpreting of films. As explained by Franco *et al.* (2010: 48), interpreters were hired to translate Western films in closed-door screenings and also in open film festivals in the former Soviet Union. Later on, the interpreters’ voices were recorded on tapes, and this is how voice-over was probably born. In fact, Burak (2011) suggests that pirate voice-over translation began to sweep the former Soviet Union when Russian *nouveau riches* hired them to watch the latest American movies. Translators generally worked at home and recorded their voice-over translations simultaneously while listening to the original soundtrack. One of the most famous translators was Andrey Gavrilov, and this is probably why voice-over is often known in Russia as Gavrilov translation (Burak 2011). This Gavrilov translation is said to have inspired voice-over in Poland (Holobut 2014), as will be elaborated below. However, other sources point towards commentators in silent films (Hendrykowski 1984), newsreels (Garcarz 2007) or even subtitling (Joanna Klimkiewicz, personal communication, in Szarkowska 2009) as the origins of the modality. Indeed, voice-over has been linked to subtitling by researchers such as Espasa (2004) in that both transfer modes imply a co-presence of two linguistic codes (the original and the translation) and in many language pairs they both entail condensing the original dialogues.
Voice-over and dubbing, on the other hand, share the fact that they are oral renderings of audiovisual content that has been prepared in advance. However, synchronization constraints differ in each case: in dubbing the translation has the same duration as (isochrony) and matches the lip movements of (lip synchronization) the original dialogue, whilst in voice-over this is not the case, as will be explained below. It is worth noting at this point that voice-over has been referred to by some authors as ‘partial dubbing’ or ‘half-dubbing’. Gambier (2004), for example, considers ‘demi-doublage’ (French for ‘half-dubbing’) a synonym for voice-over, whereas Chaume (2013: 108) considers half-dubbing a ‘type of voice-over.’ According to Chaume (2013: 108), in half-dubbing ‘a male reader reads the leading male’s dialogue in a film or series, a female reader reads the leading female’s dialogues, and sometimes a third voice reads the dialogues of other main characters in the film.’ Chaume explains that ‘[a]ttempts have been made to insert these target language dialogues into silences in the original film’, but these have just been ‘little more than experiments . . . and have had no significant impact.’ Chaume’s approach to half-dubbing has probably been inspired by Hendrickx’s short tentative paper on ‘partial dubbing’ (also termed by the author ‘concise synchronization’), where he proposes to ‘make full use of the silent passages in the original dialogues’, and suggests that the dialogue ‘may possibly even be replaced—partly or entirely—by a (shorter) description of what is going on, or it may simply be reproduced in indirect speech’ (Hendrickx 1984: 218). Hendrickx considers partial dubbing an easy and economical transfer mode that would involve ‘adding to the original soundtrack a spoken text giving the necessary information in the target language without providing a full translation of the dialogue’ (ibid: 217). However, Hendrickx acknowledges his proposal is subject to further testing, because many questions pertaining to the number of actors or the intonation to be used remain unanswered.

A trickier relationship is that of voice-over with two additional transfer modes that are usually bundled together under the category ‘revoicing’, i.e. narration and (free) commentary, which share the absence of lip synchronization (Pönniö 1995). Narration has been defined as ‘simply a kind of voice-over, where the translation has been summarized’ (Chaume 2012: 3) or as ‘an extended voice-over’ used in monologues (Luyken et al. 1991: 80), in contrast to voice-over, which is generally used for spontaneous speech (Gambier 1996). On the other hand, free commentary has been defined as ‘a variation of voice-over and dubbing, where a comedian manipulates the translation for humoristic purposes and adds jokes or funny comments’ (Chaume 2012: 4), generally in comedy or sports programmes in Europe (Chaume 2013: 110). Another definition of free commentary has been provided by Luyken et al. (1991: 80), who consider it as a means of adapting a programme for a new audience, totally replacing the original speech and only keeping synchronization with the image (Laine 1996). However, some authors have noticed that linking narration and free commentary to voice-over (either as a type of, or as a variation of the latter) can be confusing. First of all, voice-over in translation studies implies two superimposed voices, whilst very often in narration and in free commentary the original soundtrack disappears and only the target language voice is heard (Aleksonyte 1999: 6, Chaume 2013: 108). Secondly, ‘narration’ and ‘commentary’ seem more adequate terms to be used in film studies when referring to ‘speech sequences by invisible speakers over programme images’ (Franco et al. 2010: 40), rather than terms to be used when describing audiovisual transfer modes. This is why Franco et al. (2010) have proposed a new term, ‘off-screen dubbing’, to define a type of revoicing in which a translating voice replaces an off-screen voice from the original soundtrack. This off-screen dubbing—similarly to other types of AVT—can be either a faithful or a free version of the original, depending on the client’s requirements, and often coexists with voice-over.
in non-fictional genres. For instance, it is often the case that, in a documentary, off-screen dubbing is used for the narrator (meaning that only the target language voice, rather than the original narrator, is heard), while voice-over is used to designate the interviewees’ and other forms of spontaneous speech (meaning both voices overlap).

**Voice-over types**

Apart from situating voice-over within the taxonomy of audiovisual transfer modes and exploring its relationship with other transfer modes, some internal categorizations of this modality have been proposed.

Orero (2004) and Franco *et al.* (2010) examine how voice-over fits within the process of media content assembly and distinguish between voice-over for production and voice-over for post-production, depending on whether translators work from edited or non-edited content. In other words, whether they are given excerpts of audiovisual content that have not yet been converted into a full programme (voice-over for production) or they are given a fully-fledged audiovisual programme (voice-over for post-production). In the first scenario, the translator is often sent excerpts of audiovisual content (for instance, interviews), generally without a script or transcript, and has to deliver a written translation. Then, the excerpts are shaped into a full programme, and the relevant translation segments are voiced. In the second scenario, a finished product (for instance, a documentary), generally with a post-production script, is provided to the translator, who delivers a written document that will be used for the final recording in the target language. Although the final result in both scenarios is the same, a translation for voice-over, the process is different and has its own specificities in each of these varieties.

Other categorizations have been put forward in the literature. Grigaraviciute and Gottlieb (1999: 44) differentiate between first and third-person voice-over, the former being a direct voice-over and the latter a reported voice-over. The standard practice is to use first-person voice-over, meaning that the translation uses the same pronoun as the speakers in the original programme. For instance, if the speaker says ‘I think . . .’, the translation will keep the first person in the target language, making the translator more invisible. In a third-person or reported voice-over, the role of the mediator is more visible as the words of the speaker are reported in the third person. Examples of third-person voice-over have been provided by Franco (2000a: 238), who examines German versions of Brazilian documentaries in which the interviewees’ answers are frequently converted into indirect speech.

The number of voices featuring in the translated version could be another categorization criterion, as it enables scholars to differentiate between single-voice voice-over and multiple-voice voice-over. An instance of the former would be television voice-overs in Poland, where only one voice is used for all characters. An example of the latter would be documentaries voiced-over in Spain, where various voices, both male and female, are used to revoice the original speakers. And also the Lithuanian voice-over of TV films (Grigaraviciute and Gottlieb 1999), in which the common patterns is to use two actors, a male and a female for all male and female actors, respectively.

**Voice-over main features**

Our initial definition of voice-over outlined its most distinctive feature: the presence of a translating voice overlapping with a translated voice; in other words, a voice delivering a translation in overlap with the original voice. However, voice-over presents other defining features.
First of all, voice-over involves the observance of various types of synchronies. Inspired by existing classifications in dubbing (Chaume 2004), Orero (2006a) and Franco et al. (2010) differentiate four types of synchronies:

- ‘Voice-over isochrony’ designates the constraining effect that the length of the original speech has on that of the translated text—given that the translation usually begins some words after the original utterance and finishes some words before the latter ends. This allows the original words at the onset and at the end of each voice-over utterance to be heard, in an attempt to arguably enhance authenticity. In some instances, especially in fictional genres with fast-paced dialogues, reaching voice-over isochrony is not possible, and the original and the translation may finish approximately at the same time. It can also occur that the translation finishes later. In order to account for the various scenarios, Sepielak (2016a) proposes to differentiate between full isochrony (when at least one word is heard at the beginning and at the end of the utterance), initial isochrony (where at least one word is audible only at the beginning), and final isochrony (where at least one word is heard only at the end of the utterance).

- ‘Literal synchrony’ is used by authors such as Luyken et al. (1991: 141), who favour literal translation when the original voice is heard without any overlapping from the voice providing the translation. By doing so, members of the audience who understand the language can relate the translation to the original (Kauffmann 2004). Voice-over professionals, however, consider that this strategy should only be deployed when a literal transfer would not result in an unnatural translation.

- ‘Kinetic synchrony’ refers to translations that are synchronized with the body language of the characters on screen. This means that when a linguistic expression is linked to a certain gesture made by one of the characters, the translation should match this gesture to avoid inconsistencies between the verbal and the visual.

- Finally, ‘action synchrony’ involves the synchronization of the translation with the images on screen. The order of the elements in a sentence may differ in the original and in the translation, whether because of systemic differences between languages or because of the rephrasing that voice-over isochrony often entails. Still, words should be synchronized with the visuals they correspond to, thus avoiding a mismatch between what the translation states and what audiences see on screen.

A key aspect of voice-over is that, contrary to the norm in dubbing, lip synchronization is not retained. While dubbing generates the illusion that the screen actors speak the language of the translation, in voice-over the viewer is constantly confronted with a version in which original and translation coexist. A practical consequence is that voice-over is cheaper and faster to produce.

Secondly, voice-over generally implies the preparation of a written translation which is delivered orally in a pre-recorded format. Typically, translators are provided with an audiovisual file (with or without a script or transcript) and are required to deliver a written translation, following the formatting requirements of the client. In some instances it is expected that the translator will provide a perfectly synchronized translation, ready to be recorded. In other cases, an editor revises the translation to check whether it meets the required standards (Kotelecka 2006, Szarkowska 2009: 190). The last steps in the production of voiced-over content are the recording of the translated version by a voice-over narrator or various voice talents, and the final revision of the audiovisual output.
In some cases hybrid scenarios can be observed. For instance, the voice-over of interview excerpts can sometimes be outsourced to freelance interpreters working from home with their own recording software. In these scenarios they are not requested to provide the commissioner with a written transcript of the translated version, but with a sound file where the voice-over synchronies have been recorded. In order to produce that deliverable, interpreters usually watch the original a few times, take notes when necessary, and finally record a voice-over version that comes to an end before that of its original counterpart. This process does not unfold as per the voice-over workflow described above, and it does not fully match standard interpreting practices either (insofar as, in this particular scenario, the interpreter is given the chance to watch the video files a number of times). It should be noted, however, that the final output complies with voice-over conventions in terms of synchrony.

Thirdly, the coexistence of the original and the target language, the interplay of the written translation with its oral delivery, the synchronization constraints as well as the involvement of various agents in the process (from the translator to the voice talent) affect the text, which suffers various transformations along the process. From a linguistic point of view, a key feature is that the original text is often rephrased to fit in the space available, but also to make it more comprehensible to the audience. For instance, when the original version contains spontaneous colloquial speech characterized by hesitations, false starts, repetitions and discourse markers, the original is reformulated and many spontaneous features disappear for the sake of comprehensibility.

Finally, the final output is generally delivered by one or more voices, either male or female, depending on the country’s tradition, very often with a non-emphatic intonation. For instance, in Poland, filmic dialogue will always be read by a man, although in non-fictional genres both male and female voices can be heard, with a preference for female voice artists in cooking programmes and nature documentaries in some channels (Szarkowska 2009: 189). The intonation in Poland will always be flat, as discreet as possible so that the audience forgets the existence of the voice-over (Wozniak 2012: 215). In fact, as explained by Szarkowska (2009: 187), the term used by professional translators in Poland for voice-over is szeptanka, which literally means ‘whispering.’ Bogucki (2013: 20), on the other hand, notes that some professionals restrict the term ‘voice-over’ to non-fictional genres and prefer ‘lectoring’ to describe the practice in fiction films. By contrast, in countries such as Belarus, Bulgaria and the Ukraine, four or five actors imitating the emotions in the original version, providing a more emphatic intonation of voice-over for fictional genres, are used (Wozniak 2012: 215). Different voice talents with a non-emphatic pronunciation are also used in documentaries voiced-over in Spain, although prosody in reality shows is more and more emphatic.

Inextricably connected with delivery is the issue of accents in voice-over. Accents are generally standard, although foreign accents have been used in certain productions, especially in the UK (Fawcett 1996, Franco et al. 2010: 79). Díaz Cintas and Orero (2006: 478) state that sometimes, ‘if the person on screen speaks Spanish, the voice-over narrator will read the translation in English with a clear foreign accent, showing characteristic inflexions that are associated with a Hispanic person speaking English.’ However, they acknowledge that this practice ‘lends itself to debate’, as it could ‘be interpreted as a sign of the inability of foreign people to speak English correctly’ (ibid.: 478).

Voice-over practice

Although voice-over has also been used in radio broadcasting (Orero 2009: 133), the focus of this chapter is audiovisual content, where voice-over is deployed to mediate both fictional
Anna Matamala

and non-fictional content. Non-fictional genres include documentaries, interviews, commercials, among others, whilst fictional genres encompass films, TV series, or animation series, to mention a few examples. Each of these genres poses specific challenges that have been addressed in the literature. Orero (2004) discusses the difficulties of interview voice-over, whilst Matamala (2009a, 2009b, 2010), Franco (2000a) and Espasa (2004) have explored in some depth the challenges of translating documentaries using voice-over. Diaz-Vegas (2012) and Permanyer (2012), for their part, approach reality shows, and Matamala (2005) and Kotelecka (2006) deal with professional aspects of voice-over in Catalonia and Poland, respectively, and draw a picture of diverging workflows and processes.

Whether voice-over is used to translate fictional or non-fictional audiovisual content depends on each country’s tradition. Voice-over in fiction is limited to certain Eastern European countries such as Poland (on TV), Bulgaria, Russia and other former Soviet Union countries (Estonia, Latvia, Belarus, Lithuania) (Chaume 2013: 108). Voice-over in non-fiction is more widely used and encompasses not only voice-over countries but also traditionally dubbing and subtitling countries, as elaborated below:

• In Poland voice-over is used for both fictional and non-fictional content (Bogucki 2010) on television, although the new affordances of digitization allow some channels to choose between subtitling and voice-over. Szarkowska and Laskowska (2015) provide a summary of various studies on AVT trends in Poland, which show a preference for voice-over over subtitling. The results of a Canal Plus study undertaken in the 1990s showed that 52.2% of Poles preferred voice-over, with 8.1% of them preferring subtitles (Bogucki 2013). A BBC survey in the early 2000s revealed that 52% were in favour of voice-over, as opposed to 4.5% who were more partial to subtitles. A study about the Polish public broadcaster TVP found that 45% respondents were in favour of voice-over, with 45% in favour of dubbing and 4% opting for subtitling. A more recent poll by the same authors, however, hints at the emergence of new habits and views, with a strong preference for subtitles instead of voice-over (77.25% versus 6.88%, respectively)—thus challenging the widely held assumption that Poland is a stronghold of voice-over.

• In the Ukraine, voice-over translation is generally used for television broadcasting, except in programmes of Russian origin, which are either subtitled or left untranslated. Dubbing, though, is used for cinema releases, and DVDs usually contain Russian dubbing and Ukrainian subtitling (Stashkiv 2015). In Russia, voice-over is generally used for both fiction and non-fiction on television, cinema and DVDs, except for cinema films that may be dubbed (Burak 2011). Mixed scenarios can also be found: for instance, Krasovska (2004) explains that Latvian voice-over and Russian subtitles are combined in Latvian commercial TV channels.

• Older variations in voice-over practice have been documented in subtitling regions such as Scandinavia—as illustrated, for example, by the Swedish television versions of Pippi Longstocking and Emil from Lönneberga (now available on DVD with dubbing). Both productions were broadcast on Danish television with a voice-over where the ‘translating voice’ superimposed on the original track reported what was said, but also provided other cues that were not obvious from the picture (Pedersen 2010, Olaf Loom 2011).

• In Spain, traditionally a dubbing country, documentaries (Matamala 2009a) are generally translated using a combination of voice-over rendering the interviewee’s words and off-screen dubbing to convey the narrator’s words. Each voice in the original is voiced-over by a different actor of the same gender. Voice-over is also used in reality
Voice-over shows and in foreign language excerpts inserted in news programmes—although some TV news programmes, e.g. those aired by the public Catalan TV broadcaster—have recently moved to subtitling in the latter case.

- In subtitling countries such as Croatia, voice-over can also be combined with subtitling in non-fictional contents (K. Nikolic, personal communication, 07 September 2015).

Technological developments are likely to result in further variety and continue to shape viewers’ preferences in different ways. For instance in Poland, young audiences may be moving towards subtitling (Bogucki 2010: 8) or even dubbing, although voice-over remains the preferred option in translated TV programmes (Chaume 2012). In dubbing countries, the voice-over market is increasing, as popular hybrid genres such as reality shows are generally voiced-over. Finally, there is an increasing number of advertisements and infomercials shown in television shopping channels that are translated using voice-over (Chaume 2013: 118). All in all, firm boundaries between clearly delineated camps may be a thing of the past, since technologies seem to move towards a greater empowerment of audiences, who will hopefully be able to choose the audiovisual transfer mode that is more suited to their personal needs—irrespective of what the national traditions and collective preferences are.

Research issues in voice-over

In what follows, this chapter surveys a range of issues related to voice-over that have been addressed from a research perspective. They have been grouped in seven categories, although some studies could be included in more than one sub-section.

Translation and synchronization techniques

In many language pairs, translating for voice-over implies condensing the original text, omitting or rephrasing information. Various researchers have dealt with this topic in fictional genres, often comparing voice-over with other transfer modes. For instance, Aleksonyte (1999) contrasts the level of reduction in the Lithuanian voice-over of the Danish Films *Breaking the Waves* (10% of information lost) against its Danish subtitling (14% of information lost). Similarly, Grigaraviciute and Gottlieb (1999: 71) analyze the Lithuanian voicing over of the Danish production *Charlot & Charlotte* in terms of amount of information and semantic content transmitted. Their research shows that full translation is prevalent (71%) in voice-over, followed by reduction (19%) and omission (10%).

Working with the English-Polish language pair, Wozniak (2012) studies the condensation of the original texts in the voice-over of *Star Trek*, as well as the degree of audibility of the original soundtrack, combining the analysis of translation strategies and synchronization strategies. Wozniak proposes to transform the voice-over of feature films into a voice-in-between, to allow for a better access to the original soundtrack and to information in the translation. This would mean that the voice providing the translation would not be over the original but would be placed between different utterances. In Wozniak’s (2012: 216) words, ‘the principle of superimposition should be replaced with that of juxtaposition.’

Also dealing with Polish fictional genres, Sepielak and Matamala (2014) analyze film voice-overs in terms of synchrony and translation strategies. They observe that elements are very often condensed or omitted for the sake of voice-over isochrony, and this type of synchrony cannot be kept in very short utterances or fast-paced dialogues, with anticipation emerging as a useful strategy. Their corpus shows no evidence of recurrent translation.
strategies being deployed to reach literal synchrony; on a related note, only a few excerpts that illustrate kinetic synchrony are found in the corpus.

A more in-depth analysis of the topic is found in Sepielak (2014, 2016a), who focuses on both translation and synchronization techniques in voiced-over multilingual feature movies in Polish. Sepielak (2014) applies Gottlieb’s (1997) translation techniques to a corpus of four multilingual films voiced-over from English into Polish. Sepielak’s aim is to research how the multilingual elements in the corpus are transferred in the voiced-over version. Sepielak’s analysis shows that transfer, exposition, deletion and imitation are the most frequently used translation techniques. However, the researcher considers that imitation and exposition are the ones that most underline the multilingual aspect of the film. As indicated by Sepielak (2014: 269), imitation ‘is based on the assumption that spectators understand an L3 element which is simply included in the translated version and read out by the voice-artist.’ On the other hand, ‘exposition’, a term proposed by herself, refers to a special case of omission in which the multilingual element is not transferred into the target language but is still audible to the audience through the original soundtrack. In Sepielak’s words (2014: 269), exposition ‘assumes that spectators base their comprehension not only on the translated soundtrack but also on the original one.’ Her analysis is interesting in that it views the translation as part of a wide semiotic process that exceeds linguistic transfer. In other words, she does not only study how a unit in a source language is translated into the target language (Polish in this case) but also analyzes how the original soundtrack contributes to the process of meaning-making. In Sepielak (2016a), this scholar goes a step further and, using the same corpus, links the translation techniques identified in previous work to voice-over synchronization strategies. She analyzes which synchrony types are kept, and what translation techniques are used to that end.

**Linguistic aspects of voice-over**

The language features of voice-over have also attracted the attention of researchers, who have sometimes criticized the standardization that voice-over entails. Drawing on a set of documentaries voiced-over in French, Kauffmann (2004) notes that language variation often disappears in favour of linguistic standardization, a situation also observed by Franco (2000b: 228). Remael (1995, 2007) attributes this practice to language policies prioritizing the educational function of public broadcasters, and recognizes this standardization is also applicable to other transfer modes.

Language policies undoubtedly impact on the language of the translation, but it is also worth highlighting that this deletion of oral features may be a feature inherent to voice-over. As Franco *et al.* (2010: 74) explain, many spontaneous features such as hesitations, repetitions, false starts or syntactic anomalies disappear in the voiced-over version for two main reasons: on the one hand, to reach voice-over isochrony and, on the other, to prioritize comprehension. Whilst in other transfer modes such as dubbing an effort is made to recreate credible dialogues, in voice-over transmitting information in a comprehensible way is prioritized. And this often means deleting language features which are typical of spontaneous colloquial language.

Another language-based research strand in voice-over pertains to the study of slang in voiced-over films (Garcarz 2007). For instance, Holobut (2011) presents a qualitative analysis of Ben Stiller’s film *Zoolander* in which she compares how the Polish subtitled and the Polish voiced-over versions verbally portray the characters in the fashion world the film shows.
Holobut concludes that subtitles reconstruct lexical and phraseological idiosyncrasies of particular characters in the film and retain more metaphors and vulgarisms than the Polish voice-over, but they disregard pragmatic adequacy. By mixing slang expressions in complex grammatical structures typical of written language, characters appear to be inconsistent and “over-voiced” (in the sense of being at times excessively expressive). On the contrary, the voice-over version aims at maximal text reduction, avoids excessive interference with the original soundtrack, and disregards the stylistic idiosyncrasies of each character, but offers a more consistent portrayal of the fashion community. Characters therefore become ‘under-voiced’, ‘i.e. devoid of individual stylistic identity’, a strategy that combined with the reader’s interpretative competence is considered to be successful.

In a more recent study, Holobut (2014) approaches slang in Polish voice-over focusing on a feature-length pilot episode of *Miami Vice*. In this study, though, the researcher adopts a diachronic perspective and compares a 1989 version and a 2008 version of the same episode. Her analysis stresses interesting changes in the voice-over practice linked to cultural aspects, moving from a source-oriented approach in the former version to a target-oriented approach in the latter. The 1989 version is more literal, with few omissions, and slang is transferred by means of cultural equivalents, paraphrases and calques, which are combined with terminology from law and commerce, making the characters speak an awkward mix of slang and ‘bureaucratrese’. On the contrary, the 2008 version provides a free translation of the original message. Dialogues are reworked, and a more concise approach to the translation is taken, with slang terms being often omitted. When used, though, they sound natural and contribute to a better characterization of the characters. According to Holobut, these differences are due to diverging working flows: while in 1989 the translator provided a ‘raw translation’ that was reworked by an editor, in the 2008 version the professional translates and adapts at the same time. Holobut also attributes these changing practices to the fact that communist audiences needed assistance when facing foreign topics: explicitation was needed to make sure the viewers could appreciate the foreign reality. The 2008 version, on the contrary, addresses a post-communist society which is more familiar with American culture and fictional models. Therefore, the translation is more concise and consistent, and more independent stylistically from the original communicative patterns.

### Cultural aspects of voice-over

A recurrent topic in most audiovisual transfer modes, the translation of cultural references has also been an object of research in voice-over. Franco (2001b) focuses on documentaries and, using a corpus of documentaries about Brazil voiced-over into French and German, she addresses the issue of foreignization or adaptation. Franco observes that a greater degree of exoticism or foreignization is ‘almost inevitable in translated documentaries’ (Franco 2001b: 177) although she considers that a balance must be reached so that this foreign flavour ‘does not impair the target viewer’s comprehension of the whole information’ (*ibid*: 178). For her part, García Luque (2011) studies a science documentary translated from French into Spanish. Contrary to Franco’s analysis, García Luque observes a clear domestication of cultural references, which, according to the author, aims to increase the end-user acceptance of and engagement with the documentary.

Analyses encompassing various non-fictional genres and language pairs are needed to shed light on the most established practices, and hence understand whether they are linked to the original genre, target audience preferences, or broadcasters’ norms.
**Authenticity, manipulation and the translator's visibility**

Voice-over has often been said to contribute to the feeling of authenticity: the fact that the original is heard underneath has been claimed to create the illusion of reality. As Franco (2000a: 236) puts it, there is ‘at least some consensus about the implicit function of this mode which, like subtitling, provides a kind of “authenticity illusion” through the simultaneous presence of the original counterpart.’ This is enhanced when literal synchrony is kept, because viewers are assured that ‘what is being said is exactly what is being told’ (Luyken et al. 1991: 80). However, Mayoral (2001) voices a different view on this issue, noting that the coexistence of two auditory messages hinders comprehension.

Orero (2006c) explores the commonly held perception of voice-over as a form of translation that boosts authenticity by drawing on the concepts of simulacrum and hyper-reality. In the context of media studies, simulacrum is understood as a process in which significatory elements are combined to create a new reality. In the case of voice-over, the real is replaced by signs of the real and the effect is so powerful that it becomes a hyper-reality, i.e. ‘more real than the real.’ Orero illustrates the relationship between voice-over and the feeling of reality by analyzing aspects such as the delay effect (i.e. voice-over isochrony), the voice features, the format, and the visibility of the translator, a topic which she revisits in her later work. Darwish and Orero (2014), for example, reveal how content can be distorted through voice-over, that is, they show how a transfer mode associated with authenticity can in fact transfer information which does not correspond to the original. They also show how the voice-over translator may gain more visibility under certain circumstances. They illustrate their point by studying the translation of TV news through voice-over. More specifically, they use a real-life example from the BBC, namely a news item broadcast in 23 April 2006, in which a journalist delivers a news report followed by a Bin Laden audio tape. Whilst the voice of a male, supposedly Bin Laden speaking in Arabic, is heard on the background, the voice of a foreign man delivers the translation as a voice-over. The presence of the translator distinct from the journalist is not only made evident through the use of another voice, but also by the use of an on-screen caption which states ‘voice of translator’. Darwish and Orero interpret this strategy as an attempt on the part of journalists to detach themselves from sensitive content, and hence shift the responsibility to translators. However, both the physical presence of an overlapping voice with a foreign accent and the explicit acknowledgement of the voice-over translator may break the illusion of reality often attributed to voice-over. Darwish and Orero’s research also demonstrates that the original content of the news item under analysis does not match the voice-over. For instance, the BBC’s English voice-over reads ‘the enemy continues to murder our children, our women, the elderly, and destroy our homes’, whilst in the original version there is not explicit mention of ‘the enemy’, ‘our children’, ‘our women’ or ‘the elderly.’

Further manipulations of content in voice-over are also observed by Holobut (2012), who compares the voice-over version of the British series *The Saint* broadcast on Polish public television under the old regime and more recent ones. Holobut stresses the socio-cultural manipulation in the portrayals of Western reality and links her analysis to historical factors.

**The reception of voice-over**

Research on the reception of voice-over is still sparse. Polls on user preferences exist (see above), but empirical research in which users are confronted with voiced-over excerpts is almost non-existent. An exception is Sepielak (2016b), who compares the
reception of voiced-over and subtitled content in terms of content comprehension and language identification in multilingual movies. A total of 113 participants volunteered to participate in the experiment and were randomly assigned to two different groups: the first one watched a 15-minute excerpt of the movie Le Mépris (1963) voiced-over into Polish, and the second one watched the same excerpt in its subtitled version. Results show that content comprehension was higher in the subtitled condition, but when asked to identify the number of languages spoken by a character, volunteers watching the voice-over version performed better. More research is undoubtedly needed in this field. Bogucki (2010: 6), for instance, states that ‘dialogues involving several speakers are hard to follow when read by a single lector’, that is, by a single voice artist, an aspect that lends itself to reception research.

Technologies in voice-over

Rehm and Uszkoreit (2012: 38), in the Strategic Research Agenda for Multilingual Europe, mention ‘automatic voice-over’ as an important research topic, and highlight that ‘in 2020 we will see wide use of automatic subtitling and first successful examples of automatic voice over for a few languages.’ However, research so far has been limited, to the best of our knowledge, to a small-scale project: ALST (Matamala 2015). This project has researched the implementation of speech recognition, machine translation, and speech synthesis in voice-over. Each of these technologies is seen as a key element in a semi-automatized workflow that could be implemented, although not exclusively, in voice-over. The process would involve three key steps, always followed by human revision or post-editing: the first step would be the generation of a script semi-automatically; the second step would be the machine translation of the script into the target language, and the final step would be the text-to-speech voicing of the output.

Matamala et al. (2017) compare the time and self-reported effort involved in three situations: manually transcribing a non-fictional excerpt in English to be voiced-over, respeaking it, and post-editing a transcript automatically generated by a speech recognition system. Their experiment shows the potential of respeaking when creating a transcript for a text that has to be voiced-over, and the willingness of professionals to embrace new transcription methods. On the other hand, Ortiz-Boix and Matamala (2015) research the effort involved in two situations: translating non-fictional content for voice-over from English into Spanish versus post-editing non-fictional content that has been machine translated. Comparing the effort involved in both tasks means comparing temporal effort (time spent on the tasks), technical effort (keystroke, mouse movements and clicks for each tasks), and cognitive effort (pause to word ratio, average pause ratio), in all cases through keylogging data. Results prove that post-editing requires less effort than translating, although results are not always statistically significant. However, Ortiz-Boix and Matamala (2016) are not only interested in the process but also in the final product. This is why they also carry out a three-level evaluation of the translated and post-edited voice-overs. In the first stage, expert lecturers and professionals evaluate the written output generated by translators and post-editors. In the second, these excerpts are recorded in a dubbing studio, where the recording director and voice talents also assess their quality. Thirdly, end-users blindly assess both post-edited and translated audiovisual excerpts, once recorded. Results show no significant differences between human translations and post-edited machine translations in the first level assessments, whilst human translation performs slightly better in the second and third levels.
The didactics of voice-over

Translating for voice-over is a field of professional practice in many countries, hence training is needed in this transfer mode. Matamala (2008) and Franco et al. (2010: chapter 5) describe the curricular design of a course on voice-over offered by the Universitat Autònoma de Barcelona within the MA in Audiovisual Translation, both in its face-to-face and its online modes of delivery. Their work explores the course structure, contents, methodology, and assessment; they also reflect on a range of challenges, and provide sample exercises, with an emphasis on non-fiction. Similarly, Chmiel (2015) discusses how voice-over is taught in Poland, more specifically in the Postgraduate Programme in Audiovisual Translation offered by the Department of Translation Studies (Adam Mickiewicz University). The course contents and assessment model are described, highlighting the specificities of Polish fiction voice-over.

New topics, new methodological approaches

Research on voice-over has increased in the last years, but still more research is needed. Some of the innovative approaches taken in other audiovisual transfer modes need to find their place in voice-over. Research issues that would merit more investigation are elaborated on in this section:

- First of all, basic descriptive research is much needed, not only focusing on case studies but also dealing with larger corpora. We do not know how voice-over is implemented in certain countries and what translation and synchronization strategies are followed in different language pairs. And this knowledge is lacking not only synchronically but also diachronically, from a historical perspective.
- Secondly, we do not know how users react to various voice-over strategies. Various methods could be used to this end: from traditional questionnaires to more innovative tools such as eye trackers (already used in the analysis of subtitles) or equipment monitoring physiological reactions (heart rate, galvanic skin response, electroencephalography). Apart from seeing how the final output is understood or enjoyed by audiences, more information on the process of creating voice-over should be sought, including analysis of workflows and guidelines governing this practice and also investigations researching the translation process. This can be achieved, for instance, by planning experiments in which the on-screen and keyboard activity are recorded.
- Thirdly, technological developments are likely to play an important role in this field over the next years. Wider research projects in which various technologies are implemented in the process of voice-over could shed more light on the feasibility of new workflows. Alternative processes such as amateur or fan voice-over are also under-researched.

Finally, an area in which incipient research is present but which undoubtedly will increase in the coming years is the relationship of voice-over with accessibility (Jankowska et al. 2015, Szarkowska and Jankowska 2012) around two particular foci: the integration of voice-over with existing access modes such as audio description, and the similarities/differences between voice-over and audio subtitles.
Conclusion

Voice-over is an active field of professional practice in many countries; it is used to mediate various genres and audiences generally welcome it. Although the body of research on voice-over is much smaller than that focusing on other modes of AVT, the volume of quality research on voice-over has increased considerably in the last few years. Voice-over researchers should learn from what has been done in other fields, join efforts and develop ambitious research agendas that go beyond self-contained case studies. Generating knowledge about old, current and new practices will allow us to better understand voice-over and will positively impact on the whole field of AVT.

Summary

Voice-over is a pre-recorded transfer mode in which a voice delivering the translation is heard on top of the original voice. Voice-over is used for non-fictional genres in certain Western European countries, and for fictional genres in many others. Voice-over is not constrained by lip synchronization but observes other types of synchronies: voice-over isochrony, literal synchrony, kinetic synchrony, and action synchrony. Voice-over implies the preparation of a written translation in which the language is often rephrased. Depending on the country’s tradition and the genres, one or more voice talents deliver the voice-over, very often with a flat intonation. Research on voice-over has focused on translation and synchronization techniques, linguistic and cultural aspects, authenticity and manipulation, reception, technologies, and training, but new approaches are being developed.

Further reading

Franco, E., A. Matamala and P. Orero (2010) Voice-over Translation: An Overview, Bern: Peter Lang. This is the first academic book on voice-over, which provides both theoretical and practical insights into this transfer mode. It differentiates between voice-over for production and postproduction, and describes training experiences in the field. The book also includes the results of a global survey on voice-over.

Related topics

4 Investigating dubbing: learning from the past, looking to the future
8 Audio description: evolving recommendations for usable, effective and enjoyable practices

References

Aleksonyte, Z. (1999) Comparative Analysis of Subtitles and Voice-over in Danish and Lithuanian Respectively as Compared to English (based on the Danish film Breaking the Waves), Vilnius: Faculty of Philology, University of Vilnius.
Anna Matamala


**Filmography**


