4. Synchronization or lip-sync: Read my lips

4.0 Preliminary discussion

(a) When you watch a dubbed movie, can you tell whether the lip-sync is good or bad? What happens when you notice that the actors on screen are not speaking the target language? Does it distract you from the film?
(b) Sometimes lip-sync constrains the translation to such an extent that dialogue writers have to rewrite the translation completely so that it fits into the screen actor’s or actress’s mouths. A pie can be turned into an omelette for the sake of lip-sync. What do you think of these changes?
(c) Who do you think should do the lip-syncing, the translator or the dialogue writer? Why?
(d) What happens in your country if a dubbing is not well lip-synced? In subtitling countries, what is the reaction to a badly synchronized dubbed cartoon?

Achieving synchronization is a question of practice and training, in which mastering the target language is crucial to a happy ending. Experienced translators and dialogue writers translate and synchronize at the same time, i.e. the first translation option usually fits into the screen characters’ mouths. But this is not the case with trainee translators and novice professionals facing this specific AVT challenge for the first time. It therefore seems appropriate to discuss the resources available to help translators deal with this aspect of AVT. These resources are translation strategies and techniques that explicitly refer to the cognitive processes the translator experiences (strategies) and the stylistic features s/he can make use of (techniques).

4.1 Synchronization: a key factor in dubbing

Synchronization (or lip-sync, also spelt lip-synch) is one of the key factors in audiovisual translation, especially in the context of dubbing. Although it is often considered as the distinguishing feature of dubbing, it is only one of several important aspects such as the reproduction of a prefabricated oral discourse or the translation problems posed by the interaction between image and word. If we take a look at the research on lip-sync, it is regarded as an urgent, vital issue, as can be seen from the wide range of publications on the subject (see Chauve 2004c).

Beyond doubt synchronization has a direct impact on the translation process and product, and as such, it should be given due consideration in Translation Studies, particularly in audiovisual translation courses. Synchronization puts all the translator’s creative skills to the test. It is precisely through practice in the learning environment that synchronization can help the trainee translator to move away from literal conceptions in translation and build up confidence in his
or her abilities to put forward alternatives that depart from the source text to focus on the function of the text and on the viewer, one of the essential features of audiovisual translation.

Dubbing is a well-known example of the invisibility of translation, an artistic and technical exercise that intentionally replaces the original dialogue track with a new track on which target language (TL) dialogue exchanges are recorded. In contrast to voice-over for example, the emphasis in dubbing lies in matching the translation to the silent mouths of the original actors. The result is that viewers watch and hear foreign actors speaking in the viewers' own language, a paradox which has been naturally accepted in all dubbing countries.

Let us now turn specifically to the translational aspects of synchronization. These concern the denominations and definition of synchronization, the scope of the subject (types of synchronization), relevant factors in its analysis and the techniques the translator commonly uses to solve synchronization-related problems.

### 4.1.1 Denominations

_Synchronization_ is the process of matching a target language translation to the screen actors' body and articulatory movements in a recording made in a dubbing studio. Although synchronization is the term used in academic circles to refer to this process (Fodor 1975; Mayoral et al. 1988; Whitman-Linsen 1992), other terms can also be found, such as _adjustment_ or _adaptation_ —although adaptation also includes creating or mirroring oral dialogues, i.e. trying to make target dialogues sound fresh and natural. The term _revoicing_ is also used synonymously, although this is really a synonym for dubbing, which also includes the process of intralingual post-synchronization. When revoicing is used to refer to dubbing or interlingual post-synchronization, it is an umbrella term that includes all dubbing types (dubbing itself, partial dubbing, narration, free commentary, etc.).

### 4.1.2 Definition

Luyken et al. (1991:73) provide an extensive definition of synchronization (_lip-sync dubbing_) as "the replacement of the original speech by a voice-track which is a faithful translation of the original speech and which attempts to reproduce the timing, phrasing and lip movements of the original." Agost (1999:59) defines synchronization (which she refers to as 'visual synchrony') as "the harmony between the visible articulatory speech movements and the sounds heard" (my translation).

Chaves (2000:114) explains it as synchronizing the translation with the lip movements. To achieve this, the dialogue writer substitutes the words that do not coincide phonetically with the screen actors' lip movements for others that do. Pauses, the start and end of the utterance, the openness of the vowel sounds and the presence of bilabials are all taken into account. The dialogue writer is also responsible for synchronizing the pace of the dubbing actor, at
times through modifications to the text received from the translator. According to Chaves, in summary, the dialogue writer is responsible for synchronization. Díaz Cintas (2001:41) indicates that synchronization is accomplished by “maintaining synchrony between the sounds of the language of the translation and the actors’ lip movements” (my translation).

This selection of recent definitions of synchronization shows how the idea of equivalence between the utterances in the source language and those in the target language is pursued according to phonetic articulation. We can therefore conclude that a thorough definition of the term should cover the following aspects:

- synchronization between the translated text utterances and the source text utterances
- synchronization between the translation and the screen actors’ body movements and
- synchronization between the translation and the screen actors’ articulatory movements (Chaume 2004c).

The following definition may thus be posited: Synchronization is one of the features of translation for dubbing that consists of matching the target language translation and the articulatory and body movements of the screen actors and actresses, and ensuring that the utterances and pauses in the translation match those of the source text.

4.1.3 Types of synchronization

The definition proposed in §4.1.2 leads us to establish the following types of synchronization (Chaume 2004c):

- phonetic or lip synchrony
- kinesthetic synchrony or body movement synchrony
- isochrony or synchrony between utterances and pauses

Lip or phonetic synchrony, referred to as phonetic synchrony by Fodor (1976:10 and 21-71), lip-sync by Luyken et al. (1991) and lip synchrony by Whitman-Linsen (1992:20), consists of adapting the translation to the articulatory movements of the on-screen characters, especially in close-ups and extreme close-ups, also called big close-ups. In order to achieve the reality effect and naturalize the product to make it appear less foreign and more familiar (Gori 1993), particular care should be taken in the translation to respect the open vowels and bilabial and labio-dental consonants pronounced on screen, as well as sentence endings.

Fodor’s (1976) comprehensive study includes much more detailed observations for phonetic synchrony. This author (Fodor 1976:54-57) advocates replacing bilabial consonants with bilabial consonants, labio-dental consonants with labio-dental consonants, and even rounded vowels with rounded vowels.
He also recommends that the dubbing actor should imitate the gestures of the screen actor in order to come as close as possible to the original as far as verbal mimicry is concerned. Fodor's study (1976:32-36) compares the mouth movements of various languages, inhaling and exhaling and head movements but, with the exception of close-ups, extreme close-ups or detailed lip shots, norms described in the professional contexts of European dubbing countries show that his approach is a bit exaggerated and his advice is not followed in professional practice (Chaume 2004c).

The synchronization of the translation with the actors' body movements is known as kinesic synchrony. Originally referred to by Fodor (1976:72) as character synchrony, Whitman-Linsen (1992:33) later employed the term kinetic synchrony. The translation must also correspond with the movements of the screen characters: a head shaking to indicate a negative cannot be accompanied by an affirmative "yes", and a character raising his hands to his head must make some kind of interjection to match the gesture.

The synchronization of the duration of the translation with the screen characters' utterances is known as isochrony (Whitman-Linsen 1992:28); i.e., when spoken, the translated dialogue must be exactly the same length as the time the screen actor or actress has his/her mouth open to utter the source text dialogues. Most criticisms of a badly dubbed film stem from isochrony deficiencies, since this is where the viewer is most likely to notice the fault. Instances in which the character's lips have closed at the end of an utterance but the viewer still hears the translated speech, or where an actor is obviously speaking while the viewer hears nothing, are frequent and justified grounds for criticism by both critics and the public.

The implementation of these three types of synchronization – lip-sync, kinesic synchrony and isochrony – is the result of a conscious agenda to domesticate the translated text, so that viewers are oblivious to the fact that what they are witnessing on screen is a translation. Good lip-sync will transmit what Metz calls the "impression of reality", a powerful phenomenon that essentially consists of the conjunction of images in motion, credible photography and icons, and credible dialogue and sounds. This impression, or tacit pact between the director of the film – and the scriptwriter, actors, and all the crew involved in a film – and the viewers, has been defined historically in terms of the use of realistic images and a realistic performance. Unfortunately, however, less attention has been paid to the features of oral discourse in dialogue exchanges that also contribute to this prefabricated realism (see Chapter 5).

4.1.4 What synchronization is not

Two further types of synchronization described in the literature do not strictly fall within the definition of synchronization: character synchrony and content synchrony. Character synchrony (Whitman-Linsen 1992) refers to the coincidence between the voices of the dubbing actors and what the audience expects the on-screen actors' voices to sound like; in general, a child actor cannot be dubbed
by an older male voice; a woman's voice must sound feminine; and the "baddie" must sound gruff and sinister. This opens up a fascinating debate, which would lead us into discussions about what is politically correct in terms of gender issues, or the criteria used for castings.

However, this issue is directly related to the dubbing actors' dramatizations, rather than a type of synchronization. As such, it falls outside the range of synchronization to which the translator or dialogue writer has access, and should not therefore be regarded as a type of synchronization as it does not directly affect translation operations or text re-writing. The type of language used by each character in the source text should provide sufficient indication of the character's idiosyncrasies for the translator to work with. The effects of dramatization are the exclusive concern of the dubbing actors and the dubbing director.

Other authors have discussed the existence of what is known as content synchrony (Mayoral et al. 1988), which refers to the semantic relation between the translation and what happens on screen (images and music). The term synchrony or synchronization in this context is misleading however, as these authors are referring not to synchrony but to coherence, a functional-systemic term. Translation must not only follow the written source text, but also the events on screen. In other words, it must be coherent with the communicative situation established on screen (context of situation). To achieve this, the translator has several cohesive links at his or her disposal (ellipsis, recurrence, substitution, conjunction, collocation, etc.), which help to produce a translation that is coherent with the on-screen action, but which do not fall within the area of synchronization (Chaume 2004c).

4.2 Kinesic synchrony

In certain circumstances, the meaning of kinesic signs (body language) must sometimes be made explicit, either because they need to be understood in order to follow the plot (a functional need), or because if they are not spotted, the viewer would be rudely awoken from the cinematographic reverie that comes of the conscious alliance of film and viewer. Kinesic signs are sometimes accompanied by (redundant) words which make their meaning explicit. At other times they appear alone, with no spoken word, caption, sound or other icon to explain them. In the first case, translators simply have to translate the words accompanying the kinesic sign. In the second case, however, they have to decide whether the sign needs to be made more explicit or not (Chaume 2008).

In the first case, when the sign has the same meaning in both cultures, translators use conventional words or phrases that would normally go with it. Viewers understand the sign, the gesture or the movement because it also belongs to their culture. Semiotic redundancy, typical in cinema, dispels any possible ambiguity. But it also forces translators to be respectful of this cohesive device since gestures will ultimately condition the words selected to explain them. In many Western cultures, nodding is the equivalent of giving one's assent to something previously said, while shaking one's head has the opposite meaning. In an old
dubbing into Catalan of the iconic TV series Mission Impossible (Bruce Geller et al. 1966-1973) the following example could not be missed:

Mr. Johnson: Then, won’t there be more holidays until next year?
Secretary: No, Mr Johnson.

The Catalan translation changes the point of view of the answer, and we hear the secretary saying, in Catalan, “Yes, Mr Johnson”, in the sense of acknowledging Mr Johnson’s inferences. The translation made sense because what the secretary meant was “Right, Mr Johnson, you understood what I said before”. A literal translation would have worked (“No, Mr Johnson, there won’t be more holidays”), but the modulation also worked (“Yes, Mr Johnson, you are right in what you are saying; you understood what I said”). While “no” answers the locutionary verb – “no, there won’t be more holidays” –, “yes” answers the illocutionary verb – “yes, you are right to think that there will be no more holidays”. The problem is that the secretary accompanies her affirmative answer with a very obvious shaking of her head. The translation thus provokes laughter and breaks the pact with the spectator discussed in §4.1.3 (Chaume 2008).

Translators, dialogue writers, dubbing directors and actors easily spot this obvious dischrony, making this type of mistake very infrequent. Nevertheless, it is very useful as an example to show that kinesic signs work as translation constraints. In most Western cultures, the audience expects a “no” to accompany a shaking head, or a “neither”, “nor”, or any other negative word or phrase. It is difficult to explain a mistake like this, since the translation passed through various hands before the final recording.

For our purposes, the following strategies and techniques can be used to produce a translation that camouflages the translator’s manipulation (Chaume 2008):

✓ Strategy:
  ➢ To be coherent with the actor’s body movements, in this particular case, with the shaking head.

✓ Techniques:
  ➢ Natural translation or coined equivalent in the TL. But when a natural translation does not work:
  ➢ Substitute the original words for other word classes (conjunctions, pronouns, etc. such as “neither”) and phrases with the same semantic content (“at all”, etc.). In cases in which a paralinguistic sign accompanies a body movement (a person about to fall, for instance), the translator will look for the coined equivalent that corresponds with that body movement. The English interjection “Oops!” can be rendered in Spanish or Catalan as ¡EPA! or even ¡EPI!, both of which respect the bilabial phoneme /p/ where required in case of close-ups and extreme close-ups.
  ➢ Repetition of the original ST, when the term has already penetrated
the target language or there is a similar equivalent. “Hey!” can be translated as *¡Hola!* in Spanish, or even the substandard *¡Ey!*, which also contains the open vowel for close-ups and extreme close-ups.

In a hierarchy of priorities, synchronies take precedence over a faithful rendering of the ST content. Complete substitution, including change of semantic meaning, would also be accepted when none of the proposed techniques works, and when the change of semantic meaning does not affect the overall meaning of the film, or the character’s personality.

4.3 Isochrony

Mouth articulation movements are directly related to two kinds of synchronization: isochrony and lip-sync proper. In dubbing, isochrony means equal duration of utterances, i.e. when spoken, the translated dialogue must be exactly the same length as the time the screen actor or actress has his/her mouth open to utter the ST dialogue exchanges. Criticism of a badly dubbed film, as we noted in 4.1.3, is generally due to isochrony deficiencies, as it is here that the viewer is most likely to notice the fault. Isochrony is thus what compels translators to fit the length of their translation to the length of the screen character’s utterance.

In subtitling, translators also aim to synchronize all subtitles with the screen characters’ utterances (a process called spotting), so that viewers can relate the subtitles to these utterances. Synchrony in subtitling, however, does not need to be as perfect as in dubbing: a subtitle may appear some frames before the screen character opens his/her mouth, and more commonly, a subtitle can remain on screen for some frames, or even seconds, after the character has closed his/her mouth (Chaume 2008).

If the rough translation for the dubbing – i.e. the first draft of the translation still without lip-sync – does not fit the screen character’s mouth movements, the dialogue writer will have to expand or reduce the translation. Ultimately, it is the task of the dubbing actors to fit the translation into the mute screen actors’ mouths in the dubbing studio.

Amplification and reduction techniques must be monitored by relevance theory (this theory argues that the audience will search for meaning in any given communication situation and having found meaning that fits their expectation of relevance, will stop processing, Gutt 1991) and by the conventions governing audiovisual genres in each culture and time period. In the case of both amplification and reduction in dubbing, several stylistic resources or translation techniques are available to help translators overcome problems posed by isochrony (Chaume 2008):

- **Strategy:**
  - To fit a translation into the duration of the screen characters’ utterances, matching their mouth articulation movements and their pauses and silences.
Techniques:

- In the case of amplification, expansion of the target text using translation techniques such as repetition, gloss, periphrasis, anacoluthon (when a sentence abruptly changes from one structure to another), paraphrase, [longer] synonyms, antonyms, hypernyms or general terms, and hyponyms or words or phrases whose semantic range is included within that of another word. Most of these techniques also confer a fresher and more oral touch to the translation, another of the agreed dubbing standards (see Chapter 1).

- In the case of reduction, ellipsis of performative verbs, modal verbs, interjections, markers of the phatic function (i.e., the use of language for the sake of interaction, which can be observed in greetings and casual discussions of the weather, for example), expressions performing purely social functions ("hello", "good morning", "yes", "no", "thanks", "sure", "certainly"), vocatives, surnames and proper names; omission of redundancies with the images; use of deictics instead of nouns and phrases, of all-purpose words like "thing" or "stuff", of (shorter) synonyms, antonyms, hypernyms, hyponyms, metaphors, and metonymy.

Mouth articulation and the duration of utterances are key factors in deciding how many syllables the translation should have. Although in daily professional practice dialogue writers do not usually count the number of syllables in both the source and the target texts (although it is recommended they do so), a good lip-sync can help achieve a truly remarkable dubbing. Respect for isochrony in dubbing follows the hidden agenda of keeping the suspension of disbelief and making the programme more realistic, credible and trueto-life (Chaume 2008).

### 4.4 Lip-sync

Another type of synchrony related to mouth articulation is known as phonetic synchrony or lip synchrony (see §4.1.3). The term lip-sync is used in real practice as a general term that embraces isochrony and phonetic synchrony. For academic purposes, however, we must distinguish between equal duration of utterances: isochrony, and mirroring certain phonemes in close-ups: phonetic synchrony. The difference between the two should be made clear, and in this book lip-sync is understood to mean phonetic synchrony.

Dubbing assistants, dialogue writers, dubbing directors, or translators performing this function have to analyze what types of shots they are dealing with. Traditionally, dubbing countries are only concerned with close-ups and extreme close-ups—shots showing only the character's face. Dialogue writers, then, adapt the translation to the articulatory movements of the on-screen characters, paying particular attention to ensure that the translation respects the open vowels and bilabial and labio-dental consonants pronounced on screen. Dubbing primarily consists of domesticking a foreign product to make it seem realistic, credible and in tune with the audience's expectations and experiences. The translation
should therefore contain an open vowel or a bilabial phoneme where the screen actor articulates an open vowel or a bilabial phoneme in the original. Fodor’s (1976) pioneering work goes into much more detail on phonetic synchrony, but as mentioned in §4.1.3, in real professional practice, lip-sync is only observed in close-ups and extreme or big close-ups (Chaume 2008).

In close-up shots, translators or dialogue writers match an open mouth or a closed mouth with open vowels or bilabials in the translation. Vowels or consonants, however, do not have to be identical. Open vowels in the ST — /i/, /æ/, /a:/, /e/ — must correspond with open vowels in the translation, but rotations are possible. This means that translations can use vowels other than those in the source text, in the same position. Exact correspondence between vowels is not required: an /æ/ can substitute an /e/ and vice versa, if necessary, although sometimes an exact match is preferable if possible. Even in the more demanding case of bilabials, words in the target language do not have to have the same consonants as the source: a /p/ can easily be replaced by an /m/, /b/ or even by the labiodentals /f/, or /v/. Phonetic articulations of close phonemes help the translator find excellent solutions that lead credibility to target dialogues in extreme close-ups with closed lips.

In these cases of close-up shots it should be noted that phonetic equivalence overrides semantic or even pragmatic equivalence: it is much more important to find a word with a bilabial consonant than to find a synonym or a similar word in the TL. Bilabial consonants or open vowels are easily interchangeable, thus opening up the possibilities for the translation and encouraging creativity. This practice can be compared to creating rhyme in poetry or songs. Translators must bear in mind that the function of the translation is to maintain the impression of reality discussed in §4.1.3, by matching open vowels and bilabials where screen characters in close-ups visibly open or close their mouths. Fortunately for translators, these instances are scarce in films. The strategies and techniques that can be used to overcome this problem are summarized below (Chaume 2008):

✓ Strategy:

- Selecting words in the TL containing the same or similar phonemes as those found in the ST, taking into account that this will be only required in close-ups, extreme close-ups or detailed mouth shots.

✓ Techniques:

- Repetition of the word or words in the source language, when the words are identical or very similar in both source and target languages: “football” and Spanish fútbol; “morning” and German morgen; “table” and French table.

- Change of word order (syntactic or informative – topic/comment – ) so that the word containing the marked phoneme coincides with another word in the TL containing similar or identical phonemes: “the parson had a house which belonged to...” can be replaced in French with la maison du clergé appartenaient à...[the house of the clergyman belonged to...], thus placing the bilabial /m/ of maison in the spot where the screen actor pronounces the /p/ of “parson”, and placing the bilabial /b/ of appartenaient where s/he says the /b/ of “belonged”.

4.5

4.5.1

Sync visu prior the e...
> **Substitution** of the target word, which is in principle the literal translation of the source word, for a synonym, antonym, hypernym, hyponym or any other stylistic resource that respects the original meaning. Thus, *The Pancake Day*, the title of a rap song repeated by the singer in a close-up in an American TV series, was translated into Spanish as *martes de carnaval* (Mardi Gras, Pancake Tuesday), fitting the Spanish /m/ to the English /p/. In some cases professional actors, dubbing directors or dialogue writers completely change the meaning of the ST to achieve phonetic equivalence, usually within the constraints of pragmatic equivalence, i.e. still unaffected characterization, plot, etc. This illustrates how phonetic equivalence is perceived as an absolute priority in professional practice.

> **Reduction or amplification** of the word, phrase or sentence, a technique which can be combined with those mentioned above.

> **Omission** of a word or sentence constituent or **addition** of a new element, techniques that in some cases might be considered to be translation errors in written translations, but are permitted in AVT, especially in dubbing.

### 4.5 Further discussion

#### 4.5.1 Considerations on genres and text types

Synchronization is not carried out to the same degree of precision in all audiovisual genres and in all dubbing types. For example, synchronization is not a priority in *documentaries*, which are usually *voiced-over* so the viewer can hear the original sound in the background at the same time with a few seconds delay. *Isochrony* is not strictly followed: the translation often comes in two or three seconds after the narrator or screen character has started to speak. A functional approach would consider documentaries as informative text types. In this case, convincing the viewer that the actor or narrator is speaking the target language is not as important as getting across certain information or ideas with the greatest possible respect for the source text. Thus the only synchrony that tends to be taken into account is isochrony, and even so, only partially, by ensuring that the spoken target text finishes at the same time as the source text, thereby avoiding any overlap with the following section.

The other three commonly dubbed audiovisual genres are texts with a predominantly expressive function: cartoons, television series and films. If the function of the translation is to convey emotions without sounding foreign or strange, and the viewer is to experience the events taking place on screen, then the role of synchronization is certainly much more relevant. However, there is an observable difference in how synchronization is used in each of the three genres.

Cartoons demand a minimal degree of synchrony (Katan 2010). Since the characters obviously do not speak, but rather seem to move their lips almost randomly without actually pronouncing the words, a precise phonetic adaptation
is not necessary, except in the case of extreme close-ups or detailed shots in which the character apparently utters an open vowel. A further relevant factor in any analysis of this genre is the viewer: child audiences are not demanding as far as synchronization is concerned and neither isochrony nor lip synchrony is strictly applied. In contrast, kinesic synchrony is important to children's cartoons, as the cartoon characters tend to use exaggerated gestures to capture the attention of their young viewers. These gestures should be accompanied by a coherent translation.

An obvious case of how genre affects synchrony is that of Japanese cartoons (anime) and comics (manga), which are aimed at different types of consumers. Adult-oriented cartoons are not so common in Europe, where the cartoon format is generally considered as implicitly designed for the children's market. When cartoons aimed at teenagers or adults in the source culture are shown to a child audience in the target culture (Shin-Chan, Yoshito Usui 1992, for example), synchronization can be more flexible, as it will depend on the target culture's conventions for the genre and the viewer, and not on the genre and viewer in the source culture.

A thorough application of all synchronization types is required for the television series genre. Although the degree of perfection is not as high as that demanded by the big screen, television series do include the three synchronization types in all their forms. Television series offer a magnificent apprenticeship for those interested in this area of translation, as all types of synchronization have to be applied, but the final result allows for a greater margin of error than in the dubbing of a film.

Last but not least, films demand a highly polished synchronization at all levels Without going into exceptions such as B movies or telefilms, the vast majority of films systematically require synchronization of the highest quality. Producers, distributors, and exhibitors are fully aware that the success or failure of a dubbed film depends on its synchronization. All the above-mentioned synchronization types (see §4.1.1, 4.1.2, and 4.1.3) are to be found in detail in this genre, from labial consonants and rounded or spread vowels, to pauses and syllables, and even facial movement synchronization with the on-screen characters (Chaume 2004c).

### 4.5.2 Considerations on language contact

Synchronization poses translation problems, but these are not the only challenges the translator has to deal with. Sometimes another translation problem arises simultaneously. For example, in some cases an icon (an image) accompanying the spoken words appears on screen. With no icons on screen the translator is freer to find a word that more or less relates to the situation and fits the on-screen actor's mouth. But with an on-screen icon related to the word pronounced in a close-up, for example, translation solutions are limited. This situation forces translators to call on all their translation resources, and apply every last ounce of their creative skills (Chaume 2004c).

Synchronization will be more accurate if the ST word or words posing a potential synchronization problem exist in the target language. Thus, the translation...
Synchronization or lip-sync: Read my lips

from English to most European languages of the word “mummy” in a close-up will be easy to translate in terms of lip synchronization, as very similar words exist in all these languages. Close languages can benefit as far as synchronization problems are concerned since they share many words coming from the same root.

Finally, there will obviously be variations in the norms and conventions that control the way synchronization is achieved in different target cultures. The degree of perfection in the application of the various synchronization types depends on the norms of each target culture, the viewer’s expectations, the role of tradition in the use of the different synchronization types and the audiovisual genre in question, among other factors.

4.5.3 Considerations on the translation brief

Of all the types of audiovisual translation, dubbing demands the most thorough synchronization. Other dubbing subtypes, such as partial dubbing or narration, do not prioritize synchronization to the same extent. As we have already seen, voice-over does not require an exact synchronization, neither in terms of isochrony nor lip synchrony, while simultaneous film interpretation completely bypasses even kinesic synchrony. Clearly, the effort involved in the oral translation of a film is quite sufficient on its own without the interpreter having to factor lip synchrony, kinesic synchrony and isochrony into the bargain.

A further relevant factor in synchronization is the client who commissions the translation. The cinema industry demands more careful attention to synchronization than do television companies, which in turn are more exacting than a company requiring a translation of a corporate video. While the cinema and television industries insist on thorough, careful synchronization, other clients are generally satisfied with a good translation.

The function of the target text also plays a part in how carefully synchronization is carried out in the translation. Equifunctional or heterofunctional translations (Nord 1997) demand different degrees of synchronization. If the function of the target text is not the same as that of the source text (heterofunctional translations), the translator will usually synchronize the text according to the conventions laid down by the target culture for that particular function. For example, in a programme about advertisements in other languages and from different cultures, the advertisements may be dubbed using voice-over, which requires practically no synchronization, or through a more relaxed dubbing style (or even narration or subtitling). However, if the same advertisement is introduced into the target culture in order to sell the product, the standard of synchronization required will be higher (for example, the case of Werther’s Original candies, dubbed into various European languages, see Chaume 2004c).

4.5.4 Considerations on the viewer

As we noted, child audiences are not as exacting as adults with synchronization quality is concerned. For this reason, and also because phonetic synchrony
does not have to be as precise since cartoon characters are not real and do not articulate real phonemes, a lower standard of synchronization quality is acceptable in the cartoon genre, both in lip synchrony and isochrony. Child audiences will not notice any delay, nor will they demand higher synchronization quality. By the same token, television series designed for young audiences also accept certain liberties in isochrony, as young audiences, although more aware than child audiences, do not consider synchronization quality as a top priority when judging a television series. Adult audiences, however, seem to demand greater perfection in synchronization quality, and consequently, television series and films aimed at adults are generally much more polished in this respect. Nonetheless, no empirical studies have been carried out to explore any differences in the standards demanded by different age or gender groups within the adult audience as a whole.

### 4.5.5 Factors relevant to synchronization

Not all synchronization types require the same degree of perfection in translation. Contrary to Fodor’s (1976) claims, lip synchrony does not require a source language bilabial consonant to be substituted for a target language bilabial consonant; rather, any labio-dental consonant will suffice. Open vowels can be replaced by any other open vowel as is evidenced by the numerous shots in which an /æ/ is substituted for an /e/, and even on occasions, with an /o/, and vice versa. Likewise, kinesic synchrony can also be flexible where rotations between interjections are concerned, except in affirmative or negative head movements, which do require a non-ambiguous solution. In the case of isochrony, one syllable before the screen actor opens his or her mouth, and even two syllables after he or she has closed are quite acceptable, as the effect will go practically unnoticed by the viewer.

The function performed by synchronization in the source text (sound post-synchronization, in this case) is also a relevant factor in translation. If sound post-synchronization has not been carefully respected in the source text, because it is not a priority in the text, then there is no reason why it should be respected in the target text, unless additional reasons prevail (client demands, target culture conventions, etc.). Thus, in advertising or publicity texts, in which source language post-synchronization has not been given priority, synchronization in the target language is not normally of prime concern (Chaume 2004c).

### 4.6 Exercises

The following exercises combine problems raised by isochrony, phonetic synchrony and kinesic synchrony, and become gradually more difficult. Isochrony can be practised in every single sentence of every single film. In other words, students can select any film, watch a scene, listen to the on-screen dialogue, and then produce a translation, read it aloud, and check whether the translation fits
5. The language of dubbing: a matter of compromise

5.0 Preliminary discussion

(a) When you watch a dubbed movie or cartoon, do certain expressions sound somewhat awkward and typical of dubbing? How do you react when on-screen actors don't sound quite right? Does it distract you from the film?
(b) What expressions are typical of dubbing in your country? If you live in a mainly subtitled country, what odd expressions are commonly heard in dubbed cartoons?
(c) Why do you think these expressions are still used in translations today?
(d) Do you suspect that taboo language is toned down in dubbings? Why?
(e) Does the language of dubbed cartoons sound natural to you?
(f) How would you give dialogues a natural flavour?

Writing fictional dialogues that sound natural and credible to the audience's ears is one of the major challenges in both scriptwriting and audiovisual translation, especially in the case of dubbing. This challenge does not lie so much in trying to imitate spontaneous conversations, but in selecting specific features of oral discourse that are widely accepted and recognized as such by the audience. Including those features in the translation should not interfere with the comprehension and interpretation of the dialogue, but must succeed in recreating a spontaneous-sounding conversation. Several academics have concluded that fictional dialogue is a combination of linguistic features used in both spoken and written texts (Rémal 2000 and 2008; Chaume 2004a; Pérez González 2007), and that both translators and scriptwriters should aim to achieve a balance between speech and writing (Baños-Piñero and Chaume 2009).

5.1 In search of oral discourse

5.1.1 A balance between planned and spontaneous speech

In their seminal 1978 article, Gregory and Carroll made the first systematic attempt to describe the linguistic register used in audiovisual texts. They argued that audiovisual texts present a peculiar mode of discourse – the language of audiovisual texts is "written to be spoken as if not written" (ibid.:42). Much has been written on the subject since then, and the language of audiovisual texts can now be said to present a combination of features deriving from both oral and written texts. The particular features that shape the audiovisual model of each language will, in the end, depend on cultural constraints. Each culture's norms explicitly or implicitly determine the linguistic and iconic appropriateness of an audiovisual text.
The critical role of fictional dialogues in audiovisual texts and their translations warrants a thorough study of the resources available to recreate spontaneous on-screen dialogue. The main purpose of this chapter is to analyze how the linguistic code is woven into translation for dubbing, focusing on what is specific to audiovisual texts and, therefore, to audiovisual translation. The multiple semiotic codes operating simultaneously in audiovisual texts are also taken into account. What we are ultimately dealing with are texts characterized by a strange kind of oral discourse, an orality which may seem spontaneous and natural, but which is actually planned or, as we might say, feigned, false, prefabricated (Chaume 2001) in terms of mode of discourse. This characteristic is common to most audiovisual fictional texts regardless of their origins (Baños-Piñero and Chaume 2009).

5.1.2 The notion of prefabricated orality

As stated in Baños-Piñero and Chaume 2009, prefabricated orality is common to most original and dubbed audiovisual programmes based on a script that is to be interpreted as if it had not been written, especially fictional texts. Various linguistic resources are available to help scriptwriters create believable dialogues that, despite their carefully planning, viewers recognize as true-to-life conversation. Depicting realism through dialogues is one of the keys to creating a successful audiovisual programme, according to professionals in the field (Comparato 1993; Toledo and Verde 2007), as well as scholars (Pavesi 2005; Whitman-Linsen 1992) and viewers.

Several factors seem to hamper scriptwriters’ creativity and actors’ improvisational skills, however. Firstly, scripts must comply with the constraints of audiovisual media, as well as with the conventions of the programme’s specific genre. Secondly, the production of audiovisual discourse is governed by norms that have been consolidated since the advent of cinema and television — such as the use of formulaic language (Pavesi 2008:93) — and that might differ according to the media in which the text is to be broadcast (television, cinema, Internet). The choice of particular linguistic features aimed at mirroring spoken speech will ultimately depend on what is considered acceptable in the system to which the audiovisual text belongs and on the varied factors which operate in that system (Karamitrogiou 2000:71-81).

This strange mode of discourse in audiovisual fictional programmes also has direct consequences for its translation: translators must be aware that the original script they receive from the translation agency or dubbing company has been written to convey the impression of spontaneous speech, and that in their translation they must take into account the multiple signifying codes that operate simultaneously in audiovisual texts. Audiovisual translators must therefore be skilled at imitating spontaneous-sounding conversation in the target language. In a sense, the audiovisual text translator is like a second scriptwriter, whose task is to transfer the exchanges on screen in such a way that they sound credible in the target language, and can thus be identified as true-to-life dialogues and easily understood by the target audience (Baños-Piñero and Chaume 2009).
On the surface, mirroring spontaneous conversation in dubbing might not seem too complicated as, unlike subtitling, spoken linguistic features of fictional dialogues are also transferred through speech in the target language. However, in addition to the above-mentioned restrictions to scriptwriters' creativity and actors' improvisational skills, and the constraints of dubbing as an audiovisual translation mode (lip synchrony, kinetic synchrony and isochrony, see Chapter 4), the tacit norms set by dubbing companies that enforce the linguistic and stylistic standardization of dubbed fictional dialogues must also be considered (Ávila 1997:25).

Concerns about the corruption of languages other than English through dubbing and subtitling is evidenced by the publication of style guides and books that attempt to regulate the correct use of target languages in the media. Several articles deal with the 'contamination' of languages other than English through the effects of audiovisual translation (see the work by Gottlieb, for example, 1997, 2005 and 2008). Given these concerns, it is paradoxical, to say the least, that very few publications are available to provide guidance for translators specializing in dubbing and subtitling in the major dubbing languages. One example is the style guide published by Televisió de Catalunya (1997) to control the quality of audiovisual programmes dubbed into Catalan: *Criteris lingüístics sobre traducció i dublaje* (Linguistic Criteria for Translation and Dubbing; see an updated online version http://esadisc.cat). The authors of this style guide argue that both translators and dialogue writers should imitate spontaneous speech using appropriate spoken linguistic features in Catalan, but that this must be accomplished without falling to observe the Catalan grammar rules and conventions (Televisió de Catalunya 1997:11). The differences between the discourses of domestic and dubbed audiovisual texts are also evidenced in these linguistic criteria, which suggest that some linguistic features of spoken Catalan should be exclusive to domestically produced programmes and should therefore not be used in audiovisual texts dubbed into Catalan (Baños-Piñero and Chaume 2009).

This statement brings us back to the key concept of this chapter: prefabricated orality in both original and dubbed audiovisual texts. It also invites us to reflect on what linguistic features set domestic programmes apart from dubbed productions and on what kind of norms govern the production of their discourses.

### 5.2 The language of dubbing: linguistic and translation issues

Translation oscillates between two poles: its adequacy to the source text and its acceptability within the target culture (see § 3.2). In the case of translation for dubbing, a major benchmark for good dubbing quality is that the target language sounds realistic, credible, and plausible; i.e. that it does not distract us from the storyline. Put another way, it must be acceptable according to the canonical standards of an audiovisual text translated into the target language. The threat to the breakdown of the tacit agreement between spectator and film must be overcome by achieving an oral register that can be defined as *false spontaneous*
(Marzà and Chaume 2009). This applies not only to dubbing and subtitling, but also to film production.

Nonetheless, academics and linguists have, until recent decades, neglected oral registers. Perhaps this has led television companies to publish guidelines for translators, as well as newsreaders and scriptwriters. These guidelines or style sheets explain how to achieve an acceptable non-spontaneous oral register, so that previously prepared written language (the script) sound as though it has not in fact been written. In the Spanish dubbing field, the following are worth highlighting: the Manual de Estilo de RTVE (RTVE Style Manual), written by Salvador Mendieta in 1993, and the more thorough Criteris lingüístics sobre traducció i doblatge (Linguistic Criteria for Translation and Dubbing), published by Televisió de Catalunya in 1997. The Catalan corporation proposes a set of norms to be followed when writing a translation for dubbing. This is a short summary of their proposal (1997:12-14):

<table>
<thead>
<tr>
<th>Avoiding:</th>
<th>Promoting:</th>
</tr>
</thead>
<tbody>
<tr>
<td>The passive voice</td>
<td>Juxtaposition and short sentences</td>
</tr>
<tr>
<td></td>
<td>Unfinished sentences</td>
</tr>
<tr>
<td></td>
<td>Topicalization</td>
</tr>
<tr>
<td></td>
<td>Ellipsis</td>
</tr>
<tr>
<td></td>
<td>Deixis</td>
</tr>
<tr>
<td></td>
<td>Clichés and set phrases</td>
</tr>
<tr>
<td></td>
<td>Discourse markers</td>
</tr>
<tr>
<td></td>
<td>Phatic markers</td>
</tr>
<tr>
<td></td>
<td>Onomatopoeia</td>
</tr>
<tr>
<td></td>
<td>Interjections</td>
</tr>
<tr>
<td></td>
<td>Puns</td>
</tr>
<tr>
<td></td>
<td>Informal register</td>
</tr>
</tbody>
</table>

Table 5.1. Summary of the style guide proposed by Televisió de Catalunya (1997) for dubbed audiovisual programmes

In reference to subtitling, Gottlieb also applies some of the factors mentioned in the Televisió de Catalunya guidelines:

- In spontaneous speech (which may be ‘artificial’, as in feature films) the subtitler will often find:
  1. Pauses, false starts, self-corrections and interruptions.
  2. Unfinished sentences and ‘grammatically unacceptable’ constructions.
  4. Overlapping speech, a feature very difficult to render in writing.

(Gottlieb 1997:112)
The proposal drawn up by the linguistic advisors of the Catalan corporation summarizes what is meant by false oral register in television and cinema; in short, what norms prevail in the production of credible life-like dialogues in the target culture, in this case Catalan, but needless to say, in all other target cultures. Once again, if the norms that shape the oral register in the target culture are not respected, the threshold of permissiveness is crossed, the norm is violated, and the tacit agreement between the spectator and the visual content is broken. Whitman-Linsen also regards the issue in the same way when she refers to an experiment conducted by Herbst on the recognition of oral register:

Artificiarity is one of the main faults pillared in denouncements of dubbed versions: the audience can hear that it is not an original. Dubbed language simply does not correspond to the way normal people talk. Herbst conducted an experiment in which he presented students with original and dubbed texts. The revealing findings indicated that the latter were clearly recognizable as such. No wonder the dubbing actors themselves take the brunt of criticism. (Whitman-Linsen 1992:118)

Here we are dealing with a prefabricated, artificial, non-spontaneous oral register; in other words, one which does not exactly imitate the spontaneous oral register, but echoes many of its characteristics (Cheume 2004a:167-186). As early as 1960 Caillé took the same line, particularly in reference to the translator's responsibility, who, if s/he did not produce a realistic oral text, later forced the actors into a false performance:

L'acteur, le comédien, ne peut donner sa mesure et avant tout 'jouer juste', que si le texte qu'il a à interpreter est lui-même naturel. (Caillé 1960:107)

Above all else, Caillé considers a realistic text in the right non-spontaneous oral register to be essential, even more so than lip-sync:

Pourvu qu'une phrase soit au rythme, qu'elle traduise l'original avec toute sa charge de sensibilité, de coëtre ou de tendresse, qu'elle garde la saveur, il n'est pas nécessaire, sauf dans certains cas de très gros plan, que toutes les labiales soient en place. [...] Si les voix des comédiens sont justes, si le texte doublé est juste, émouvant ou divertit, la part la gagnée. (Caillé 1960:107)

Duff also noted the peculiarity of this kind of discourse, and following the prevailing trend for prescriptiveness at that time, recommended translators not to use dubbed translationese. He identified two stages in translation: the first when writers 'translate' their thoughts into words, and the second when translators 'translate' from words into words, i.e. from one language to another:

[... concepts do not cover exactly the same fields of meaning in different languages. And the translator who imposes the concepts of one language
to another is no longer moving freely from one world to another but instead creating a third world – and a third language. (Duff 1989:10)

Nencioni (1976) spoke about “parlato-recitato”, i.e. recited spoken language, to distinguish between real oral discourse and planned discourse, a discourse which can be labelled “l'attributo di ‘recitato’, purché s'intenda nel senso della esecuzione di un parlato programmato” (1976:49), i.e. the attribute of ‘recited’, in the sense of performing a planned speech. In the same line, but in much greater detail, Pavesi baptized the language of dubbing as “a third norm” and focused on linguistic elements characteristic of this kind of language (Pavesi 1996). Melloni (2004) and Bruti and Perego (2005) also explore this issue, focusing their attention on some peculiar elements of dubbed discourse and subtitling discourse, respectively (Marzà and Chaume 2009). To date, Freddi and Pavesi (2009) is the most comprehensive volume on the peculiarities of the language of dubbing. Forchini (2012) shows a challenging experiment which confutes the claim that movie language is different from real oral language.

Although not all authors agree (Foder 1976; Kahane 1990-91), it would appear that most publications on dubbing confer greater importance on realistic dialogue than good lip-sync. If the intonation and the dialogues are credible and natural, the audience will be more tolerant of any unsynchronized lip movements that may appear in the dubbing. Whitman-Linsen sums this up in five sentences:

As long as a certain tolerance threshold is not overstepped in any of the different types, the illusion of authenticity can be successfully established. [...]. Most researchers and professional dubbers alike lend the greatest priority to a believable, convincing dialogue. [...] What matters is the impression, the credibility of the artistic work viewed as an integral whole. Ultimately, Caille claims, cinema is a factory of illusions. Dubbing attempts to give the illusion of an illusion. (Whitman-Linsen 1992:54-55)

To achieve realistic dialogues, certain more micro-textual norms must also be respected. In one of the most systematic studies on dubbing in the nineties, Goris (1993) summarizes his doctoral thesis, based on field work carried out on translation norms for dubbing in France. Following an analysis of various North American films dubbed into French, Goris (1993:169-190) discovers that during the translation process, non-standard fragments of the original text (i.e. slang, colloquial syntax) tend to be standardized and culturally adapted (from the conversion of measurements into the decimal system to the adoption of the French equivalent of foreign place names, and the use of synchrony as a primary strategy for naturalization). He also reveals the tendency to explain any ambiguous fragments of the original text, in order to produce a uniform, easily understood text that comes close to the receiver's culture. The norms Goris identifies are essentially repeated in the Spanish dubbing industry, as shown in the work of Ballester (2001) or Martí Ferrerol (2010).
Focusing on micro-textual norms, Chaume (2004a) proposed a model of analysis structured around four language levels (prosody, morphology, syntax and lexis), based on the comparison of written discourse, oral discourse and dubbese from a target-oriented perspective. This model was later validated and extended by Marzà et al. (2006) with a corpus of 900 minutes of dubbed cartoons, and then again later validated and extended by Baños-Piñero (2006). Their results offer a more detailed view of the specific position of dubbese along the written-spoken and high register-low register axes, showing that their defining features are unequally distributed among the four language levels.

The prosodic and phonetic level is oral by nature, but dubbing actors' professionally clear articulation and intonation place it high along the register line. Very few features, such as the expressive use of intonation or vowel liaisons – between two words liaison involves a follow-through between a final consonant and an initial vowel (though otherwise mute in liaison) or between a final vowel and an initial vowel thus transforming the first vowel into a semiconsonant — echo spontaneous oral speech. Similarly, the morphological level proves to be very standard: the limited use of verbal tenses and the extended use of derivation—the process of forming a new word on the basis of an existing word — are the only oral, spontaneous features found in that corpus (Marzà and Chaume 2009).

Syntactically, dubbese presents a balance between oral and written features. On the oral side, the sentence structure tends towards immediacy and simplicity, and the flexibility of syntax is exploited for pragmatic purposes, as in the case of topicalization—the tendency to place what is being talked about at the beginning of the sentence. In contrast, anything which could lead to a dispersion of information is usually avoided. This results in a syntactically direct, simple and compact language characterized by short sentences and a preference for juxtaposition. Another noteworthy syntactic and pragmatic feature is the strong interaction of dubbese with its context, expressed through conversational markers and deixis—the phenomenon by which understanding of certain words and phrases in an utterance requires contextual information (Marzà and Chaume 2009).

Finally, lexis is extremely open to oral and colloquial language. It is common in translation practice to bring together all the characteristic features of a given linguistic variety in this level, i.e., all substandard features in the four different levels of the source text are poured into the lexical level of the target text, to maintain the original substandard flavour of the film. The vocabulary used in dubbing is not varied, but it is very flexible, creative and cohesive, and therefore easily adapted to a given communicative context. However oral and colloquial, lexis is also restricted by a strong tendency to respect standard language, as a recurrent characteristic of dubbese on all levels (Marzà and Chaume 2009).

To summarize, dubbese is a culture-specific linguistic and stylistic model for dubbed texts, labelled by Pavesi as a third norm, similar, but not equal to real oral discourse and external production oral discourse (i.e., original target-culture films, sitcoms, etc.). Dialogue writing must comply with a series of demands, which include:
creating the effect of natural, credible and true-to-life dialogues
promoting a balance which avoids overacting and underacting when
dubbing actors perform the dialogues (i.e., avoiding cacophonies, etc.)
complying with lip-sync.

These demands, however, operate differently in real oral discourse, in-house
or outsourced production and dubbing, since target culture norms vary across
these three language production environments. If one had to place the mode of
discourse of these two production modes, i.e., in-house production programmes,
and dubbed programmes, on a continuum between the traditional written and
oral poles, the in-house production linguistic model would certainly be nearer
to the oral pole than would the language of dubbing:

```
Real oral discourse

written  spoken
---------------------------------------------------
                          X

In-house production oral discourse

written  spoken
---------------------------------------------------
                          X

Dubbese

written  spoken
---------------------------------------------------
                          X
```

Figure 5.1. The written-spoken continuum

Both in-house production oral discourse and dubbese strive to resemble real
oral discourse and use it as a benchmark, but dubbing is far more standardized
and strict than the language of in-house production (Baños 2006).

5.3. An analytical model for the study of prefabricated orality

In order to study the language of dubbing, we need an analytical framework to
guide us through the numerous features that language presents. This does not
mean that the same analytical framework should be used for all languages, or that
it has to be set in stone. Rather, the framework of analysis should be continually
nourished by new research findings. Most importantly, however, we must specify
the linguistic features we want to look for in our corpus, since otherwise we would
be overwhelmed by the infinite magnitude of the written and oral features we can
find in a screenplay (see Chapter 7). We may wish to look for phonetic features,
for syntactic misplacements or for semantic calques, for example. But it is difficult
to select one feature and not another, and in the end research projects on dubbing
tend to be limited to just one feature (such as interjections in dubbese; for