Audiovisual content analysis in the translation process

Elena V. Aleksandrova[1]

Nella A. Trofimova[2]

Svetlana Y. Rubtsova[3]

Irina R. Pulekha[4]

Ekaterina V. Tulina[5]

Maria V. Artamonova[6]

[1]Candidate of Philology, Associate Professor at the Foreign Languages Department, Institute of Linguistics, Murmansk Arctic State University, 15 Egorova street, Murmansk, 183038, Russia,

[2]Dr of Philology, Professor at the Department of German, St. Petersburg state university, 7-11 Universitetskaya nab., 199034, St. Petersburg, Russia

[3]PhD, Ass Professor at the Department of LSP in economics and law, St Petersburg University, Universitetskaya nab. 7/9/11, 199034, Saint Petersburg, Russia

[4]Candidate of Philology, Associate Professor at the Department of Linguistics and Translation, Nosov Magnitogorsk State Technical University, 38 Lenin Avenue, 455000, Chelyabinsk Region, Magnitogorsk, Russia

[5]PhD, Associate Professor at the Department of Linguistics and Translation, Nosov Magnitogorsk State Technical University, 38 Lenin ave., 455000, Magnitogorsk, Russia

[6]PhD, Associate Professor at the Department of Linguistics and Translation, Nosov Magnitogorsk State Technical University, 38 Lenina ave., 455000, Magnitogorsk, Russia,

Audiovisual content analysis in the translation process

Elena V. Aleksandrova[1], Nella A. Trofimova[2], Svetlana Y. Rubtsova[3], Irina R. Pulekha[4], Ekaterina V. Tulina[5],Maria V. Artamonova[6]

[1]Candidate of Philology, Associate Professor at the Foreign Languages Department, Institute of Linguistics, Murmansk Arctic State University, 15 Egorova street, Murmansk, 183038, Russia,

[2]Dr of Philology, Professor at the Department of German, St. Petersburg state university, 7-11 Universitetskaya nab., 199034, St. Petersburg, Russia

[3]PhD, Ass Professor at the Department of LSP in economics and law, St Petersburg University, Universitetskaya nab. 7/9/11, 199034, Saint Petersburg, Russia

[4]Candidate of Philology, Associate Professor at the Department of Linguistics and Translation, Nosov Magnitogorsk State Technical University, 38 Lenin Avenue, 455000, Chelyabinsk Region, Magnitogorsk, Russia

[5]PhD, Associate Professor at the Department of Linguistics and Translation, Nosov Magnitogorsk State Technical University, 38 Lenin ave., 455000, Magnitogorsk, Russia

[6]PhD, Associate Professor at the Department of Linguistics and Translation, Nosov Magnitogorsk State Technical University, 38 Lenina ave., 455000, Magnitogorsk, Russia,

Email: aevea@mail.ru[1], nelart@mail.ru[2], rubtsova_svetlana@mail.ru[3], irinapulekha@mail.ru[4], tulina78@yandex.ru[5], tulinaev@magtu.ru[5], maria_art@inbox.ru[6] , m.artamonova@magtu.ru[6]

ABSTRACT

The article presents a comprehensive approach to the process of audiovisual translation that includes application of multimodal analysis of semiotic codes present in audiovisual productions. The article dwells on how the proposed approach can be applied to analyzing audiovisual productions for different types of audiovisual translation. Due to its multimodal nature, an audiovisual production is understood by the authors as an audiovisual text that combines image, sound and verbal means, that is, different modes conveying meaning. The means of conveying meaning in an audiovisual production include the visual non-verbal elements, visual verbal elements as well as audio non-verbal and verbal elements. The priority of these means of meaning transfer and their interaction in meaning generation differ significantly depending on the genre of audiovisual productions and the specifics of the process of its creation.

Keywords: Audiovisual content analysis, audiovisual translation, semiotic code, multimodal transcription, interlingual translatio, intralingual translation, intersemiotic translation, voice-over, dubbing, subtitling, audio description.

1. INTRODUCTION

The emergence of audiovisual translation is inextricably linked to the development of arts, and in particular to the history of world cinema. For many years, scientists from various fields of scientific knowledge have been studying cinema. Though expressing different opinions, they agree on one thing: as any art form, cinema has its own specific language, the elements of which form a harmonious whole that is actualized in a particular type of audiovisual content.

Audiovisual productions include cinematographic productions as well as all types of content expressed by means similar to cinematographic production (television and video films and other similar works), regardless of the method of their original or subsequent fixation. It is traditionally believed that audiovisual productions in their narrow sense are not a literal reflection of reality, but its figurative interpretation on the screen. According to this approach, it is customary to divide cinema into four types:

- scientific;
- documentary;
- feature;
- animation.

Recently, there has appeared a tendency to divide audiovisual productions as follows:

- non-fiction (filming of authentic events and people - newsreels, science films, documentaries, etc.);
- fiction (reality is acted out by actors under the direction of a director: comedies, dramas, tragedies, etc.);

- animation ("animating" inanimate objects - drawn animation, shadow animation, combined animation and other types of "animation" of narrative objects).

At the same time, audiovisual productions in the broader meaning can be presented also by a wide variety of content that conveys meaning by means of audio and visual channels.

Cinema, despite its relatively short existence, has been the most popular art form for several decades. Every year production companies generate a huge amount of audiovisual content. The viewers get access to a huge number of movies, TV series, cartoons, commercials and other audiovisual products, the translation of which is often not of high quality. This can be explained by several factors: the complexity of the audiovisual translation process, the emergence of "fan studios" and non-professional amateur translators and the lack of unified theoretical foundations of this type of translation.

The concept of audiovisual translation itself appeared relatively recently. K. Reiss defines audiovisual translation as a type of translation aimed at translating multimodal, multimedia texts and their transfer to another culture, in order to have the same effect on the recipient of the translation as on the recipient of the original multimedia text [Reiss, 1981].

At the same time, as defined by Jorge Diaz-Cintas, audiovisual translation is the translation of materials in which the verbal sphere is supplemented by elements from other communication channels [Cintas, 2006: 13]. Thus, an audiovisual translator faces the task of processing and transferring not only the semantic context, but also extra-verbal integral visual and semantic elements. Hence, when working with audiovisual translation, one needs to pay attention to a huge number of nuances that require a special approach of the translator. The task of the translator is not only to translate the text, but also to transfer its extra-linguistic features and the multitude of semiotic codes present in the audiovisual content.

Audiovisual translation nowadays is considered to include several varieties: subtitling, subtitling for the deaf and hearing impaired, closed captioning, dubbed translation, voice-over translation, audio description. Within the framework of such approach, audio description is understood as intersemiotic translation (in other terms - intermodal, crossmodal), that is, the description translates visual images into verbal text. Obviously, this type of translation differs significantly from other types of translation. In translation studies, it is accepted that audiovisual translation is a limited type of translation due to various restrictions imposed on the text by technical requirements, performance conditions, customer recommendations. For example, in a dubbed translation, a character's facial expressions should be taken into account if the character is shown on the screen in close-up, which can lead to a significant difference between the translated text and the source text. In subtitling, the limitations are determined by the reading speed and the maximum number of characters in the subtitle line.

A person who has no limitations in the perception of the surrounding world, i.e. a person with normal sight and hearing, who perceives information without cognitive limitations, can potentially receive 14 semiotic codes in full: the linguistic code (dialogues, monologues, reading); the paralinguistic code (speech manner, delivery, intonation, accents); the literary and theatre code (plot, rhythm, dramatic development) is transmitted through the audio channel via the verbal sign system. Through the non-verbal sign system, the audio channel transfers special sound effects (noise, rumble, special effects); music and paralinguistic (voice quality, voice volume, silence, pauses) codes. The visual channel becomes a means of conveying the graphic code (subtitles, intertitles, letters, menus, headlines, street names) of the verbal sign system. The numerous codes of the non-verbal sign system conveyed through the visual channel for receiving information include iconographic (images, emoticons); photographic (lighting, perspective, light); scenographic (visual signs); film (shooting, framing, cameraman techniques); kinesic (gestures, poses, facial expressions, looks); proxemic (movements, use of space, distance between characters); dress codes (costumes, hair styles, makeup) [Gambier 2013].

It can be presented in the following form:

Table 1:Semitics codes within an audiovisual production

	Audio channel	Visual channel
Verbal elements	linguistic code (dialogues, monologues, reading, comments, voices off); paralinguistic code (speech manner, delivery, intonation, accents); literary and theatre code (plot, rhythm, dramatic development, drama progression, sequences, narrative);	graphic code (subtitles, intertitles, letters, menus, headlines, street names);
Non-verbal elements	special sound effects (noise, rumble, special effects); musical code; paralinguistic (voice quality, voice volume, silence, pauses);	iconographic code (images, emoticons); photographic code (lighting, perspective, light); scenographic code (visual environment signs); film code (shooting, framing, cutting and editing, genre conventions);

		kinesic code (gestures, postures, facial expressions, gazes, manners; proxemic code (movements, use of space, distance between characters); dress code (costumes, hairstyles, makeup).

The ratio of these elements varied considerably in different periods of cinema development. For example, during the period of silent cinema the noises 'came to life' in the viewer's perception due to dynamic screen scenery (waves rolling on the shore, tree crowns bowing to the ground under the onslaught of the wind, etc.). The dominant position also belonged to the music accompanied the screen narrative. Speech was represented by intertitles, and noises and silence were expressed by means of associations.

With the advent of sound recording, the ratio of these elements changed dramatically in favor of the verbal component. The so-called "sound expansion" began as the elements of the verbal (dialogues, monologues, voice-over, intonation, accents, etc.) and non-verbal (voice quality and volume, special effects, pauses and silences) sign systems began to gain ground in audiovisual productions competing with visual elements.

2. Hypothesis

Taking into consideration the complex nature of any type of audiovisual content, the quality of audiovisual translation depends on whether the analysis of the content was carried out with reference to the type of the translation to be performed or not.

The variety of translation types (voiceover, subtitling, recast, lipsync, SDH, audio description, live captioning, sign language interpreting etc) stipulates the approach to the audiovisual content analysis. Subtitling, for example, requires considering the pragmatics of the verbal text, its pace and channel (on screen/ off screen) as well as other elements of the graphic code combined with the analysis of the frame aimed at choosing the proper position of the subtitle on the screen. Another approach is required for the vertical video subtitling as the plot and shooting of vertical videos is often structured around different film narration building techniques. At the same time, subtitles for the deaf-and-hard of hearing (SDH) require, first and foremost, the analysis of noise, sound, music and paralinguistic codes (rumbling, special effects, voice volume, silence, voice quality, etc).

Thus, translation type-specific audiovisual content analysis based on multimodal semiotics is the integral part of audiovisual translation process.

3. Methods

The proposed type-specific audiovisual content analysis involves the comprehensive study of semiotic codes present in the audiovisual production with reference to its production type (feature or documentary film, TV series, cartoon, commercials, live performance, etc) and the type of translation (voiceover, subtitling, recast, lipsync, SDH, audio description, live captioning, sign language interpreting).

4. Main body

A detailed discussion of all the principles underlying film narrative is beyond the scope of the study of language theory, but in general we can say that the creation of an audiovisual production is a three-stage process in which the authors or creators (director, cameraman) combine elements from the two main blocks of film narrative.

In the first stage, the author makes the decides on the following issues:

1) which characters to include in the narrative and what actions they take;
2) in what spatial and temporal conditions these actions will take place.

At the second stage, the author of the audiovisual production decides how the story of the production will be told, which includes determining the following components of the process:

- the order in which the actions will be presented - they may be presented in chronological order, the chronological order can be changed for various reasons (flashbacks to explain a particular character trait, a delay in the action taking place to create tension, two different action lines presented in parallel or intertwined);
- frequency of actions - the director may decide to show actions more than once (e.g., from the perspective of different characters) or not to show them at all, e.g., to get the audience interested;
- duration of actions - whether the actions will be presented at their normal speed, accelerated (e.g., to omit irrelevant parts of the narrative or to create a comic effect), or at a slower pace (e.g., to create tension);
- the characteristics of the characters - their physical and their mental characteristics, and their behavior;
- the spatial and temporal framework of the narrative and its details.

It can be concluded that the creator of an audiovisual production determines at what time and in what sequence the actions in a given film narrative will take place. This information is important for the analysis of the audiovisual content when performing the translation.

In the first two stages, the narrative is essentially an abstract construction, and only in the third stage does the author decide exactly how they will present it, which involves using various filming techniques to determine what is shown (e.g., a close-up to depict a character's emotional reaction to an event or a panoramic shot to represent a landscape), how it is shown (e.g., a certain camera angle to show the superiority of one character over another) and what the relationship is between the various shots. All of these techniques taken together, used to bring together different shots and scenes, create a coherent and consistent whole. Not only do they serve a narrative function, they are also used to define the style of the audiovisual production and ensure its coherence.

In other words, the creation of an audiovisual production is a complex process involving various steps and offering almost endless possibilities. Translating audiovisual productions requires a thorough analysis not only of its verbal elements as well as its genre, the specific style of the director, the techniques of filming, diegetic and non-diegetic sounds, etc.

The process of audience perception of an audiovisual production is often referred to as audience reconstruction of the film narrative [Remael et al 2019]. Viewers are presented with some specific narrative and must interpret it in order to arrive at the original, abstract (chronological) construction with which the author began. This is partly an individual process, depending on the knowledge and experience of each recipient, but to a large extent this reconstruction is a more or less universal process that all audiovisual viewers have mastered to a greater or lesser extent. Viewers reconstruct narratives according to general principles, and a deeper understanding of these principles can help audiovisual translators decide what principles to apply to the process of translation so that the result (translated content) serves the same purpose as the original production. Central to this perception are the semiotic codes that tell the story of the audiovisual production - when interpreting it, viewers combine the codes (the characters and the spatial and temporal conditions in which they take place, shooting and framing techniques, etc). As the story progresses, viewers continually update this mental model by adding new information to it, confirming what has already happened or that they have drawn the correct conclusions from the information received, and modifying existing information or assumptions. Hence, a similar mental model has to be created by means of audiovisual translation for any given translation type in any given language with reference to the pragmatics of the translation - depending on what purpose it is supposed to serve (comprehensive accessibility for audio described productions or semantic comprehension of corporate instructional video, SDH for feature films or recast for animation productions).

Voice-over translation is a type of translation in which the translated speech of the voice actors is heard over the original soundtrack of the production. Often the translation is about 25% of the volume of the English track and 75% of the volume of the Russian track. This type of translation is the most popular among TV channels, documentaries and movies not aimed to be released in movie theaters, as it is much cheaper than dubbing and can be done in a shorter time.

With voice-over translation the film code plays an important role in the analysis: dialogues of characters should be in the foreground, music and noises - in the background, at the same time with the original dialogues, the music and noises are in the in the background all of which bear the emotionality of the film. A. Matamala [Matamala 2009] highlights the following aspects of performing voice-over translation with reference to documentary productions:

- adapting the length of the text (voice-over isochrony),
- synchronizing text and body language (kinetic synchrony),
- synchronizing text and visuals (action synchrony),
- rephrasing the language to create a comprehensible discourse which nonetheless retains the register.

Another type of audiovisual translation is translation for two-dimensional subtitling. With two-dimensional subtitling, the main problem for the translator is the rapid growth of external constraints such as the need to fit the translation into a certain number of characters and lines that are due to international standards of reading speed and subtitle display on the screen; correlation of subtitle change to the change of plans in the frame (that leads, again, to the vital role of analyzing the film code of the production as the display of the subtitles on the screen is directly connected with the shooting and framing peculiarities of the audiovisual production).

When subtitling, a number of physiological peculiarities of information perception on the part of the viewer and technical peculiarities of audio and video content must be taken into account. The following restrictions are imposed on the subtitling process that are directly connected with the precedence of certain semiotic codes of the audiovisual production:

- subtitles should be placed at the bottom of the screen (except for Korean, Japanese and Chinese languages in which subtitles can be placed on the side of the screen);
- the length of the subtitles must be within two lines or they will overlap the image;
- the average length of one line has a limit of characters (different for each language);
- subtitles have to be synchronized with the characters' speech - they should disappear and appear along with the character's line;
- significant words are often italicize intonationally as well as off-screen speech is.

The above mentioned restrictions require special attention to be paid, apart from the linguistic and graphic codes, to the paralinguistic (both verbal and non-verbal) and scenographic codes.

In order to interpret the meaning of the multimodal elements of an audiovisual text, one can resort to multimodal analysis. It was first carried out by E. Baldry in 2005, when he took a step forward in the study of multimodality and performed an in-depth multimodal analysis of car and movie advertisements. The multimodal concordancer MCA used by Baldry allows for the film to be divided into several parts that can then be analyzed, signed with detailed comments regarding the semiotic codes used in the multimodal text to create meaning. If the translator has identified the existing connections between images and words, they will be able to sift out the material for translation. This is especially necessary for subtitling or dubbing an audiovisual production due to the prevailing role of timing for these translation types.

The multimodal transcription method consists in dividing the multimodal text into separate frames arranged in a table, which also describes the details of the internal content of the frame. As a results, a table can be made with four columns:

Table 2:Multimodal transcription for subtitling purposes

frame	visual image + kinetic actions	audio	subtitle translation

The visual image + kinetic action column describes what the viewer sees in the frame. The audio column notes the background noise, music, dialogue and spoken language. The last column shows the subtitle translation while the frame is presented as a video image.

Multimodal transcription can also be a useful tool for determining what means are used to create meaning in a multimodal text, and whether something other than words can convey that meaning. After all, if an audiovisual text uses other semiotic codes to create meaning, the translation of the verbal content in subtitling can be cut down in order to fit the translated text into the time frame.

Similarly, multimodal transcription method can also be applied to dubbing, which requires not only timing synchronization, but also lip synchronization. Usually dubbing refers to the substitution of original speech with speech in the target culture's translated language. The implementation of translation for dubbing without a holistic understanding of the external references of the audiovisual production often leads to translation failure. Translation for full dubbing (lip-sync) is a type of audiovisual translation that implies a complete replacement of the foreign speech for the purpose of subsequent broadcasting in countries that do not speak the language of the production's original. The main task of the translator in this type of dubbing is to provide synchronization of lip movements and phonetic image of the translated text. When translating for dubbing the translator re-synthesizes the text on the basis of parallel semantic streams, transforms the semantic whole of the text and the image in the situation of another language and another culture. In dubbing, the voices of the original actors are not heard, and the voices of the dubbing actors are synchronized with the lip movement, facial expressions, etc. The main function of the dubbed translation is to create the impression that the film sounds in the viewer's native language which requires a thorough analysis of the majority of the visual codes in combination with the verbal elements.

Lip synchronization is not required for the whole of audiovisual production, in most cases it is necessary to synchronize the text (the precise articulation of the characters) only for close-up scenes, when there is an opportunity to follow the face and facial expressions of the characters. Such meticulous work does not have to be done when the screen is showing a general plan or when the viewer cannot see the faces of the characters. A modified table of multimodal transcription can be a useful tool for analyzing the close-up frame and the verbal elements:

Table 3: Multimodal transcription for dubbing (lip synchronization)

frame	visual image + kinetic actions	audio	subtitle translation

In addition to interlingual audiovisual translation types there are also intralingual translation and intersemiotic ones that can greatly benefit from applying comprehensive multimodal analysis approach.

The intralingual and intersemiotic types of audiovisual translation may include:

- subtitling for people with hearing impairment;
- sign language interpretation for people with hearing impairment;
- audio description;
- audio introduction;
- live subtitling/ captioning;
- subtitling in opera and theater.

5. CONCLUSIONS

Given the multimodal nature inherent in audiovisual productions, translators have the task of providing information appropriate to the type of audiovisual translation so that viewers can fully perceive the narrative presented in the audiovisual content. This task, in turn, consists of two parts:

- analyzing the audiovisual production that serves as the source text in order to determine what exactly is inherent in the film narrative and what principles and techniques were used to do so;
- determining which elements of the narrative are prioritized and how the limitations of a particular type of audiovisual translation can affect the quality of the transfer of these elements.

The better audiovisual translators understand how a film narrative is constructed in an audiovisual production and how viewers reconstruct it, the better prepared they will be to create a quality translation.

Conflict of Interests

The authors confirm that the presented data does not contain any conflicts of interest.

REFERENCES

1. Baldry, A. A Multimodal Approach to Text Studies in English. / A. Baldry. — Campobasso : Palladino Editore, 2005.
2. Baldry, A. Multimodal Transcription and Text Analysis. / A. Baldry, P. Thibault. — London : Equinox, 2006.
3. Cintas, J. D. Introduction: the didactics of audiovisual translation / J. D. Cintas // The Didactics of Audiovisual Translation. – John Benjamins Publishing Co., 2008. – p. 1 – 18.
4. Gambier, Y. Screen Transadaptation: Perception and reception / Y. Gambier // Screen Translation: Special Issue of The Translator. – 2003. – V. 9. –№ 2.– P. 171 – 189
5. Gambier Y. The position of audiovisual translation studies // The Routledge handbook of translation studies. New York: Routledge, 2013, pp. 45–59.
6. Matamala, A. Main challanges in the translation of documentaries / A. Matamala // New trends in audiovisual translation. – Multilingual Matters. – 2009. – p. 109 –120.
7. Matamala, A. Terminological challenges in the translation of science documentaries: A case-study / A. Matamala // Across Languages and Cultures. – 2010. – Vol. 11. № 2. – p. 255 – 272.
8. Metz, Ch. Film language: a semiotics of the cinema / Christian Metz: translated by Michael Taylor.New York: Oxford University Press, 1974. 286 p.
9. Orero, P. Voice-over in audiovisual translation / P. Orero // Audiovisual Translation. – London: Palgrave Macmillan, 2009. – P. 130 – 139.
10. Reiss, K. Type, kind and individuality of text: Decision making in translation trans. Susan Kitron. Poetics Today 2:4. 1981. 121-131.
11. Remael, A. Audiovisual translation / A. Remael // Handbook of translation studies. – 2010. – V. 1 – P. 12 – 17
12. Szarkowska, A., Jankowska, A. Text-to-speech audio description of voice-over films. A case study of audio described Volver in Polish / A. Szarkowska, A. Jankowska // Innovation in Audio Description Research. – EUT Edizioni Universita di Trieste, 2012. – P. 81 – 98.
13. Thibault P. The multimodal transcription of a television advertisement: Theory and practice // Multimodality and Multimediality in the Distance Learning Age. – Campobasso : Palladino Editore, 2000. – P. 311–385.