

VISUAL INTERPRETATION OF FILM TRANSLATION

THÉRÈSE ENG*

<https://orcid.org/0009-0001-4071-4863>

STOCKHOLM UNIVERSITY

Abstract: Which references are considered necessary for understanding and empathy in visual interpretation of translated feature films? This is the starting point for this article on audiovisual translation and visual interpretation. Visual interpretation is a scientifically relatively unexplored field of research that can be linked to both cognitive science, semiotics, and audiovisual translation. Just over a decade ago, there was little or no research into visual interpretation in Sweden or the Nordic countries. The first Swedish research initiatives started in the form of workshops in sight interpretation organized by Jana Holsanova, Mats Andrén and Cecilia Wadensjö (2010-2014) and resulted in a report on sight interpretation (Holsanova et al. 2016). The task of the visual interpreter is to select and describe relevant information such as events, environments, people, characters and their appearance, facial expressions, gestures, and body movements in television programs, cinema, or theater performances by giving verbal descriptions of visual scenes to evoke vivid mental images and audience empathy. Visual interpretation should contribute to our understanding and convey impressions and mood. It is a so-called intermodal translation, because the visual interpreter transfers content from image to words (Jakobson 1959; Reviere 2017). Through the language, those who listen should be able to follow along in the action. But they should not only know what is happening, but also be able to laugh at the same time as everyone else, understand why a certain sound occurs when it is heard and know who is doing what. It is thus about a completion of what is missing in the multimodal interaction (Holsanova 2020: 4). According to professional visual interpreters, the aim is to use a neutral voice to be clear, concise, and descriptive, so that the target group can imagine what something looks like with the help of internal images. In today's rapid technological development, we also want to reflect on the opportunities and challenges of automated visual interpretation and translation, using a ChatGPT.

Keywords: visual interpretation, audiovisual translation, intermodal translation, artificial intelligence

1. Introduction

In a world where audiovisual content plays a significant role, accessibility is essential to include everyone in that world of entertainment. People with visual impairments

* therese.eng@su.se



face challenges when trying to enjoy movies and TV shows. Visual interpretation, (hereafter VI), and film translation have a key role in making these media accessible to all.

What is visual interpretation? Choose a film or theatre performance or why not a live event you have never seen before. Close your eyes. You hear the dialogue, the music, the songs and all the sounds in the room. You take part in everything that is clearly stated. But you miss all the nuances, everything that is said “between the lines” and in all the shifts in facial expressions. You miss the starry sky that illustrates that it is night, the glances between two people that are the only thing that reveals their love, that a person walking on a sidewalk repeatedly looks over his shoulder and thus signals a fear that is never expressed. Everyone else around you marvel at the scenery, the clothes, the choreography, which they also comment on without you being able to participate in it all. That is where the visual interpreter comes in. The interpreter's task is to tell briefly and concisely what is not being said. Place the one who cannot see in time and space, describe gestures and facial expressions, concretize the unspoken. Simply “lend their eyes” to help them put together the puzzle of the said and the unsaid, illustrated and implied that is the core of a film, theater, or other arrangement.

The task of the interpreter is to select and describe relevant information (events, environments, people, characters and their appearance, facial expressions, gestures, and body movements) in television programs, cinema, or theatre performances by giving verbal descriptions of visual scenes to evoke vivid mental images and audience empathy. VI should contribute to understanding and convey impressions and mood. It is a so-called intermodal translation, because the visual interpreter transfers content from image to words (Jakobson 1959; Reviers 2017).

Through the language, people listening should be able to follow along in the action. But they should not only know what is happening, but also be able to laugh at the same time as everyone else, understand why a certain sound occurs when it is heard and know who is doing what. It is thus about a completion of what is missing in the multimodal interaction (Holsanova 2020: 4).

2. Research Field

VI is a scientifically relatively unexplored field of research that can be linked interdisciplinarily to both cognitive science, semiotics, and audio-visual translation. Just over a decade ago, there was little or no research into VI in Sweden or the Nordic countries. The first Swedish research initiatives started in the form of workshops in sight interpretation organized by Jana Holsanova, Mats Andrén and Cecilia Wadensjö (2010-2014) and resulted in a report on sight interpretation (Holsanova et al. 2016).

Another important aspect is the future research's contribution to the training of visual interpreters (translating and non-translating) as professionals in both the VI industry and the film industry (editors, sound and image editors, screenwriters, etc.), which in turn has consequences for its users.

3. Premises of Visual Interpretation

VI is an art form that requires a balance between conveying visual and auditory elements while maintaining the rhythms and pace of the film. Holsanova (2020a) emphasizes the importance of integrating both the visual and auditory content of sight interpretation to create a complete and engaging experience for people with visual impairments.

Other important aspects that are also one of the subtitling translator's work premises are time and space: the text strip(s) in the subtitle is visible for a certain time when we have time to read it and each text strip has a limited space for letters and spaces. Synchronization of the subtitle and the dialogue is for a film translator a constantly ongoing working condition.

The time aspect is also important for an interpreter, but the other way around, because the interpretation does not take place during the film dialogue itself, but in the breaks in between. Just as the subtitler (usually) cannot translate all the dialogue, the interpreter cannot describe everything that is seen, but must choose what is necessary and relevant to the plot.

According to professional interpreters, it can be tempting to interpret exactly everything in every break in the film, but too much information can be just as bad as too little information, if there is silence for too long, the user may wonder if something has gone wrong. The aim is to use a neutral voice to be clear, concise, and descriptive, so that the target group can imagine what something looks like with the help of internal images. (Stjärnfeldt 2020: 4).

4. Required References According to Interpreters

Which references are considered necessary for understanding and empathy in VI of translated feature films? Cultural references in translation have long been an area of interest for translation researchers. In the case of audiovisual translation, in the case of non-interpreted feature film, we have visual pathways for the audience, but in the case of a combination with translated film, what cultural references are necessary for understanding and empathy and what choices are made here?

To answer that question, we must examine the work being done by professional interpreters.

One of them, specialized in film and theatre interpretation, Emmeli Stjärnfeldt¹ is basically an actress and has worked for over 13 years in the profession as a visual interpreter. She is now active in Malmö, Sweden and started working for the company Syntolkning Nu just over 10 years ago, one of the largest visual interpretation companies in Sweden. To work as a visual interpreter there is an acting training course, but many of them come from the theatre world, mainly actors, but also journalists appear as acting interpreters (Stjärnfeldt 2021: 1,4). In the case of Stjärnfeldt. a

¹ Interviews were taken from Emmeli Stjärnfeldt, professional visual interpreter (mostly for syntolkning.se) on April 1, 2021 and August 27, 2023.

recruiter from Syntolkning Nu had seen her in an English performance at one of the theatres in Gothenburg and needed someone who could interpret a play from English to Swedish. Stjärnfeldt also writes visual interpretation scripts for films and describes that “what you need is to put things into words, in a way that is not judgmental, but at the same time transparent.” She continues: “You try to give as much information as possible, no, by the way, not as much information as possible, but the important information.”

VI is, just as the word reveals, a question of interpretation. No one perceives impressions in the same way as anyone else, everything is filtered through our personal experiences, our background, our associations. Therefore, no one will interpret the same way as you.

Although sight interpretation has existed for a while, research in the area has not been around for a long time. It is therefore important to know that the information we have about visual interpretation today may be updated and changed in the future.

Most often in VI, you cannot say exactly how something should be done. VI requires dexterity. What works for one film may not work for another. It is particularly important to work with feedback and hear what the users themselves think about the product delivered. The easiest way to understand what you need to think about when interpreting is to familiarize yourself with what it is like for the users. Closing your eyes and playing a scene that is not visually interpreted creates a series of question marks. Who is talking? What is she wearing? What do the people do? Are we indoors or outdoors? Is the sun shining? Is it raining? How many are in the room? Have we changed places now? This is what we want answers to in a visual interpretation.

Although an interpretation is always personal, their aim is to avoid personal opinions about a place or a person. Interpreting with a neutral voice is good, it means that you as the interpreter do not need to dramatize what you say. Above all, the language must be clear, concise, and descriptive, so that the target group can imagine what something looks like with the help of internal images.

What should be interpreted? You interpret all relevant information so that the user has the same understanding as the rest of the audience. You must also convey atmosphere with words. But what is important in one context may not be important in another. The atmosphere in a commercial is possibly more important than the atmosphere in an informational film. It is up to you to decide which information is important for the message of the film to get out.

Sometimes it becomes a matter of prioritization if time is short. What does the scene or film say? Is it clear to everyone? How can we make it clear? For example, events, environment, clothes, season, sound, weather, graphic signs, audio descriptions, facial expressions, appearance, furniture, gestures, pictures, who is talking, and colors can be valuable information.

When should you interpret? Timing is important in sight interpretation. The visual interpretation must match what is shown in the image, at the same time as it is shown. Sometimes, if something is funny or we know we will run out of time later, the interpretation must happen a bit earlier, so that everyone can laugh at the same time, for example. Do not interpret when there is dialogue or narration unless it is necessary. If the lyrics of a song are important, do not interpret then either.

Already in the planning stage and script writing of the film, you can advantageously think about making room for visual interpretation. Otherwise, there is a risk that the interpretation must be printed in a short time, which neither benefits the user nor the interpreter. You can make room for the visual interpretation, for example, by letting clips be a little longer in the picture. Or to include the visual interpretation information in the film's natural sound with 'verbalization.'

It can be tempting to interpret exactly everything and in every break in the film, but too much information can be just as heavy as too little information. However, if there is silence for too long, the user may wonder if something has gone wrong. If you need to repeat a person's name that has already been introduced, it may be enough to say the person's name before they speak.

Use a concise description. Often the interpretation must be done in a brief time. If you have not planned for the visual interpretation in the planning stage, it can therefore become a challenge. Filtering the language not only saves you time, but it also makes the interpretation stand out from the other voices. One of the challenges of visual interpretation is to describe concisely yet meaningfully.

"Eva is sitting facing us at her teak desk and typing on her computer. She has wavy blonde hair down to her shoulders, side legs, blue eyes, sharp black glasses, pink lipstick, purple T-shirt, and a short silver necklace around her neck. She has a sign on her T-shirt that says Malmö city. She has a red flower in a white vase in front of her on the desk. The rain pours down on the window behind her."

This describes the picture before us. But there are more effective and clearer ways to convey the image. Maybe not all the details will fit, depending on the space of time.

"The office. The rain patters against the window. Eva, facing us, is sitting at her desk, typing on her computer. She has wavy blond hair down to her shoulders, black glasses, big smile, purple t-shirt with a sign that says Malmö city."

While we want to be tight-lipped, we would rather tell you what we see than how we experience it. When interpreting faces and expressions, interpret what you see, do not censor but do not use offensive language either. This detailed description might work:

"Ulla-Stina 87 years old, curly short gray hair, bright blue intense gaze, high cheekbones, curved nose, pink cat-eyeglasses on the tip of the nose, red lipstick, face covered with fine wrinkles."

If we make an informational film, we might want to shorten it and focus on what stands out a little:

"Ulla-Stina 87 years old, curly gray hair, pink glasses, red lipstick."

If there is an opportunity to interpret facial expressions and body language without evaluating, that is also positive. Rather:

“Theo clenches his fist tightly under the table.”

Than:

“Theo gets mad at Andrea.”

However, it depends entirely on how much information we get in terms of sound, maybe the anger can already be heard in the voice? If there are no clues in the voice and the information must be conveyed quickly, we may have to be overly clear.

The way you say something can also give many clues - pauses, intonations, emphasis. Speak clearly when extra clarity is required.

What should not be interpreted?

1. Telling whether we are in close-up or half-shot is not relevant.
2. When interpreting diagrams, information is the most important thing, not what type of diagram it is or colors etc.
3. The visual interpretation must not mislead, confuse, or distract.
4. Avoid jargon.

What and when should you interpret?

1. Do not interpret during dialogues or when other important information is conveyed, this also applies during singing with important text.
2. Read relevant graphics, but also audio descriptions are notable features.
3. Describe things as they happen or possibly briefly in advance.
4. Pauses, music, sounds must be able to come through for the sake of the atmosphere, do not interpret in every small pause.
5. Also, do not wait too long to interpret, then the user may wonder if something is wrong.
6. Plan for visual interpretation already when you plan the film, leave breaks where the visual interpreter has time to tell the necessary information.
7. Remember that you can include visual interpretation information in the natural sound of the film, for example in what the speaker reads out or when an interviewee introduces themselves (verbalization)
8. If it is a female voiceover, it can be good to have a male voiceover for the sake of clarity and vice versa.
9. If you are making an informational film, the information is the most important.
10. Rehearse your voiceover script several times before recording.
11. Animated films are interpreted in the same way as other films.
12. Avoid the term “we see.”
13. Describe from the viewer’s perspective, think about the audience you are targeting.
14. Describe in the present tense.
15. Keep it simple.

There is a lot that does not get interpreted, just because there is no time. As an interpreter, you must be prepared to delete things because the dialogue is what is central in a film. For an interpreter, the more knowledge they have of a country, the better they are as visual interpreters of a foreign film. The greater and deeper the knowledge of the subject in the film, the better they will be as visual interpreters. The aim must always be to put the film experience at the highest possible level for the target audience.

5. Possibilities with Automated Visual Interpretation and Translation

The rapid technological development has opened the door for automated VI and translation. We will discuss the potential and challenges of using advanced AI models like ChatGPT to automate sight interpretation and translation and compare this to human expertise².

These developments offer both advantages and challenges in making audiovisual content accessible to people with visual impairments.

Now let's quickly ascertain numerous advantages with the newest technology:

- **Rapidity and Efficiency:** Automated systems can generate VI quickly and at scale. This allows for faster availability to a wide audience, especially when it comes to releasing new movies or TV shows.
- **Cost efficiency:** By using automated systems, producers and platforms can potentially reduce the cost of producing dubbed and translated versions of their content. This can contribute to increased availability in the long term.
- **Adaptability:** Automated systems can be flexible enough to adapt to different target groups and their needs. This can be particularly useful when offering visual interpretation and translation in different languages and with various levels of detail.

But we see also a wide range of challenges, and we will state some of them:

- **Cultural Nuances:** Translating and interpreting cultural nuances is a complex task. For example, irony, sarcasm, and subtle references can easily be lost in translation if the system does not have the ability to understand deep cultural concepts.
- **Lack of Human Understanding:** Automated systems lack the human understanding and contextual awareness required to account for emotional nuances and subtleties.
- **Voice and Tone:** Choosing the right voice and tone to be appropriate is a complex task. Automated systems can have difficulty achieving the right balance and conveying the right feeling.

² Two examples are provided through the link, one where a synthetic voice interprets and another with a human interpreter: <https://www.syntolkcentralen.se/riktig-eller-syntetisk-rost/>.

- **Technical Accessibility:** Although automated sight interpretations and translations may be easy to generate, technical barriers may arise in making them accessible to users.

6. Conclusion

While automated systems have the potential to make audiovisual content more accessible than ever before, it is important to remember that they cannot replace the human insight and creativity of professional visual interpreters and translators. The future should aim to integrate these automated tools to complement human work to create the best possible VI and translation experience.

Despite the progress made in VI and translation, there are still areas that often do not receive the attention they deserve due to their complexity. These areas represent limits and challenges that interpreters and translators face when trying to make audiovisual content accessible to all.

References

- Holsanova, Jana. 2020b. "Uncovering Scientific and Multimodal Literacy through Audio Description." *Journal of Visual Literacy*, Special Issue 39(3). DOI: 10.1080/1051144X.2020.1826219
- Holsanova, Jana. 2020a. Att beskriva det som syns men inte hörs. Om syntolkning. *Humanetten*, 44 (2020): 125-146. DOI: 10.15626/hn.20204406
- Holsanova, Jana, Johansson, Roger, and Lyberg-Åhlander, Viveka. 2020. "How the Blind Audiences Receive and Experience Audio Descriptions of Visual Events." - a project presentation. *Book of Extended Abstracts*. 3rd Swiss Conference on Barrier-free Communication, 39-41.
- Holsanova, Jana. 2016. "Cognitive Approach to Audio Description." In *Researching Audio Description: New Approaches* edited by A. Matamala. & P. Orero, 49-73. London: Palgrave Macmillan.
- Holsanova, Jana, Wadensjö, Cecilia, and Andrén, Mats (Eds.). 2016. "Syntolkning – forskning och praktik." *Lund University Cognitive Studies 166 / Myndigheten för tillgängliga medier*, rapport nr. 4.
- Holsanova, Jana, Hildén, Anita, Salmson, Maria, and Kesen Tundell, Veronica. 2015. "Syntolkning och uppläst text - en studie om hur användarna vill ha det." In *Med riktlinjer för audiovisuella medier*. Stockholm: Tundell Salmson Lär.
- Pedersen, Jan. 2016. "Översättningsvetenskaplig forskning om syntolkning," i Holsanova, Jana Andrén, Mats & Wadensjö, Cecilia (eds.), "Syntolkning – forskning och praktik." *Lund University Cognitive Studies 166/Myndigheten för tillgängliga medier*, rapport nr. 4: 53-59.
- Remael, Aline, Reviere, Nina, and Vandekerckhove, Reinhild. 2016. "Audiovisual Translation. Theoretical and Methodological Challenges: Some Research Trends." In *Audiovisual Translation. Theoretical and Methodological Challenges*, special issue, Target 28(2): 248-260, edited by Y. Gambier & S. Ramos Pinto.

- Remael, Aline, and Reviere, Nina. 2015. "Recreating Multimodal Cohesion in Audio-Description: The Case of Audio-Subtitling in Dutch Multilingual Films." *New Voices in Translation Studies* 13: 50–78.
- Reviere, Nina. 2018. *A Corpus-Based Study into the Linguistic Features of a New, Multimodal Text Type* [PhD thesis]. University of Antwerp. Summary.
- Reviere, Nina. 2017. *Audio Description in Dutch. A Corpus-Based Study into the Linguistic Features of a New, Multimodal Text Type* [PhD thesis]. University of Antwerp.
- Stärnfeldt, Emmeli. 2020. *Handbok för syntolkning Malmö stad*. Stadskontoret Malmö.
- Vercauteren, Gert. 2016. *A Narratological Approach to Content Selection in Audio Description* [PhD thesis]. University of Antwerp.
- Riktlinjer för undertextning i Sverige - Medietextarna, 2020. Accessed August 25, 2023. [https://portal.research.lu.se/portal/sv/projects/audio-description--research-and-practices\(043b472d-afcd-4b8f-a829-73236270c740\).html](https://portal.research.lu.se/portal/sv/projects/audio-description--research-and-practices(043b472d-afcd-4b8f-a829-73236270c740).html)

Conflict of Interests

The author declares no ethical issues or conflicts of interest in this research.

Ethical Standards

The author affirms this research did not involve human subjects.