

CHAPTER II

LITERATURE REVIEW

This chapter presents a brief overview of relevant literature on audiovisual translation, focusing on its definition, professional roles, and significance in media and education. It also highlights previous studies related to the skills required in the field and the importance of introducing this profession to the students in the field of Applied Foreign Language.

2.1 Theoretical Framework

2.1.1 Audiovisual Media

Audiovisual media refers to instructional media that integrates visual and auditory components, allowing learners to use both their sense of sight and hearing simultaneously where this combination is considered effective in improving students' comprehension of learning materials (Affiati et al., 2021; Winarto et al., as cited in Manulang et al., 2023). Audiovisual media includes various forms such as videos, films, documentaries, and audio-slide presentations that deliver messages through moving images and sound at the same time. According to Utami et al. (2020), this type of media can present educational content as well as moral values in a more attractive and efficient way. Its use is regarded as a strategic method for creating diverse learning experiences, especially for learners who respond better to visual and auditory stimulation. Documentary videos, for instance, are capable of influencing public understanding and perceptions of social and cultural issues in a more emotional and interactive manner. Liao (2023) explains that audiovisual media not only functions as a means of delivering information but also has the ability to shape audience attitudes and opinions. Therefore, ethical considerations and the accuracy of the messages conveyed through audiovisual media are essential in public communication.

According to Guillot (2023), when Audiovisual Translation (AVT) first gained recognition during the mid-1990s, scholarly discussions primarily concentrated on positioning AVT within the field of Translation Studies and addressing the technical issues associated with different AVT modes, including subtitling, dubbing, voice-over, and audio description. As the discipline developed, the scope of AVT expanded beyond technical and linguistic concerns to include pragmatic and intercultural dimensions. AVT plays a role in shaping how interpersonal relationships, politeness strategies, and cultural identities are represented in translated audiovisual content. Translation decisions are influenced not only by technical constraints but also by the interaction of multiple linguistic and cultural systems. Consequently, AVT is guided not only by lexical equivalence but also by audience expectations and intended pragmatic outcomes. Research on the translation of swear words, for instance, has revealed that subtitling often involves mitigation strategies, although these do not necessarily produce negative effects on viewers. Such findings demonstrate that audience responses to AVT are multifaceted and highly dependent on context. Furthermore, Long (2023) highlights that each AVT mode should be examined individually because every type presents its own characteristics and challenges.

2.1.2 Audiovisual Translation

According to Chaume (2020, as cited in Paramita & Sampurna, 2025), audiovisual translation involves more than translating spoken words because translators must also consider visual elements, sound effects, and written texts that appear on the screen. This means that audiovisual translators are responsible not only for transferring meaning from the Source Language (SL) to the Target Language (TL), but also for ensuring that the translation fits the visual and audio components of the media. As a result, audiences can understand the message naturally and effectively. Since meaning in audiovisual products is communicated through various modes such as speech, images, music, gestures, facial expressions, and on-screen text, translators

need to pay attention to both linguistic and technical aspects throughout the translation process.

2.1.2.1 Subtitling

Díaz-Cintas and Remael (2021) describe subtitling as a form of translation in which spoken discourse is converted into written text and displayed on screen, usually at the bottom. Subtitles convey both verbal interactions between speakers and other forms of information presented visually, such as signs and written notices, as well as auditory information, including songs and voice-over narration. From a linguistic perspective, subtitling can be categorized into intralingual and interlingual forms. Intralingual subtitling, often referred to as Same-Language Subtitling (SLS), uses the same language as the original audiovisual content and is commonly applied in Subtitles for the Deaf and Hard-of-Hearing (SDH). In addition to accessibility purposes, intralingual subtitles are frequently utilized in language learning because they help viewers understand unfamiliar accents and pronunciations, expand vocabulary, improve listening skills, and increase cultural awareness through authentic audiovisual materials. In contrast, interlingual subtitling involves transferring dialogue from a Source Language (SL) into a Target Language (TL). This category may appear in monolingual, bilingual, or multilingual subtitle formats.

2.1.2.2 Dubbing

Cary (1960, as cited in Díaz Cintas & Remael, 2021) characterized dubbing as a *traduction totale* because of the extensive linguistic work involved in replacing the original dialogue of an audiovisual product with dialogue in the target language. The objective of dubbing is to create the impression that the characters naturally speak the audience's language. Chaume (2026) explains that dubbing is a multi-stage process carried out by several professionals within the audiovisual sector. The process starts when a client, such as a television broadcaster or streaming service, commissions a dubbing company to localize audiovisual content. A translator initially produces a translated version of the dialogue, which is subsequently adapted into a dubbing script

synchronized with lip movements, timing, and natural speech patterns. The adapted script is then performed by voice actors under the supervision of a dubbing director and sound engineer. Once recorded, the voices are combined with the original music and sound effects before undergoing quality control procedures to verify synchronization, translation accuracy, and overall sound quality prior to release.

2.1.2.3 Voice Over

According to Díaz-Cintas and Remael (2021), Voice Over (VO) is an audiovisual translation technique in which the original speaker's voice remains audible at a low volume while the translated version is spoken over it. Typically, viewers hear the original speech briefly before its volume is reduced to allow the target-language narration to become dominant. This method often involves dialogue being delivered in a neutral tone and is commonly used in documentaries, news broadcasts, and other nonfiction audiovisual productions because it preserves a close relationship with the source text. Voice-over remains a widely used practice in several Eastern European countries, particularly Russia and Poland.

2.1.2.4 Audio Description

Díaz-Cintas and Remael (2015, as cited in Díaz Cintas & Remael, 2021) define Audio Description (AD) as an accessibility service designed to communicate essential visual information through verbal narration, enabling blind and visually impaired audiences to understand audiovisual content more fully. Through audio description, audiences can follow storylines, recognize characters' physical appearance, facial expressions, body language, emotional states, and the settings in which events occur. The Prime Video Audio Description Style Guide (2024) outlines five key stages involved in producing audio description: analyzing visual content, writing the description script, determining appropriate timing or spotting, recording the narration, and completing the mixing and mastering process.

2.1.3 Audiovisual Translator

In the context of documentary translation, audiovisual translators play an important role in preserving factual information, cultural values, and audience understanding when adapting content into another language, they often encounter a number of linguistic and technical challenges. Gottlieb (2004, as cited in Paramita & Sampurna, 2025) explains that translators must work within several limitations, including subtitle timing, synchronization, reading speed, and restricted screen space. To overcome these challenges, translators may apply different translation strategies such as reduction, paraphrasing, equivalence, adaptation, or expansion in order to maintain readability while preserving the original message. These challenges become even greater in documentary productions, where translators frequently deal with fast-paced narration, culture-specific references, specialized terminology, and large amounts of information. Therefore, audiovisual translators need both strong language skills and technical competence to produce accurate and accessible translations for the target audience.

2.1.4 Video as an Educational Resource

Educational media are instruments designed to assist the development of students' knowledge and skills throughout the learning process. Alwahibee (2019) explained that one important approach associated with educational media is the concept of scaffolding, which originates from Lev Vygotsky's (1978) theory of the Zone of Proximal Development (ZPD). This concept suggests that learners are capable of reaching more advanced levels of understanding through structured assistance that is gradually withdrawn as students become more independent in their learning process, as stated by Nurhayati et al. (2016, cited in Purwasih & Rahmadhani, 2021). In educational practice, this form of support can be delivered through a variety of media, including educational videos that present information progressively and contextually.

According to Vygotsky (cited in Keith S. Taber, 2018), educators and educational systems should develop alternative methods to support students who encounter difficulties in conventional learning environments. In this regard, scaffolding functions as guidance that assists learners in progressing toward higher stages of development, as explained by Raymond (2000, cited in Alwahibee, 2019). Numerous educational tools have been created to support this process, including books, audio materials, images, and videos, all of which are intended to improve learning achievement. Purnaningsih (2017) argued that such educational tools help students comprehend and master learning materials more effectively.

Furthermore, Molder, Anderton, Howell, Goodwin, Behrman, and Ellefson (2025) applied the Cognitive Theory of Multimedia Learning (CTML) framework to examine the effectiveness of three forms of educational videos: animations, formal lecture videos, and documentary video. Their findings revealed that documentary video positively influence the development of students' science identity, perceptions of the speaker's credibility and relatability, attitudes toward science, and emotional engagement during learning. These results suggest that documentary-based educational media are effective not only in increasing student engagement but also in enhancing understanding of the material being taught. Therefore, these findings provide a strong foundation for research focused on developing documentary media as an educational tool to improve student engagement and comprehension in the study of Audiovisual Translation (AVT).

2.1.5 Documentary Project Production

Video is an audiovisual medium that communicates information through the integration of moving images, sound, and text arranged systematically to deliver particular messages. Within educational settings, video has long functioned as a supporting learning medium because it can represent the complexity of teaching and learning activities in a more realistic manner. According to Alsawie and Alghazo (2010)

as well as Star and Strickland (2008), cited in Yang and Lin (2025), video enables the capture of visual details and learning contexts that encourage deeper reflection and reasoning based on evidence. One form of video frequently utilized in education is documentary film, which places particular emphasis on presenting factual events and realities as learning resources. Nelson (2018, cited in Warden & Stirling, 2025) explained that documentary represents truth rather than the truth, since documentaries capture, organize, and present fragments of truthfulness derived from everyday life experiences. Furthermore, Reid and Sanders (2021) argued that documentary videos provide filmmakers with flexibility to explore topics in an open, personal, and constructive manner without being restricted by a rigid framework or structure. In *Introduction to Documentary*, Nichols (2024) categorized documentaries into seven modes: poetic, expository, observational, participatory, reflexive, performative, and interactive. The documentary developed in this research applies both the expository and participatory modes. The expository mode appears through narration and direct explanations concerning audiovisual interpretation, while the participatory mode is reflected in the interaction between the filmmaker and interviewees during the interview process. The integration of these two modes supports the educational purpose of the documentary while preserving the authenticity of its representation.

2.1.5.1 Pre Production

The pre-production stage is an essential early phase in documentary video production that emphasizes planning and preparation before filming begins. According to Rabiger and Hurbis-Cherrier (2020), pre-production involves several activities, including developing the video's concept and objectives, preparing scripts, selecting appropriate locations, organizing production schedules, forming the production team, and deciding the technical and visual approaches for the project. Careful preparation during this stage is highly important because it helps reduce both technical and conceptual problems that may arise during production. In documentary filmmaking, ideas must be supported by sufficient research to guarantee that the information

delivered is factual and accurate. Reid and Sanders (2021) further explained that, when producing observational documentaries, filmmakers can utilize the pre-production period to identify potential interview participants and prepare relevant interview questions. These questions are later developed into a storyline that serves as the foundation for the documentary's conceptual visualization and overall vision. They also emphasized the importance of transforming abstract ideas into concrete and achievable production goals. After the storyline has been formulated, filmmakers may proceed by designing storyboards, preparing shot lists, and planning camera placements. A properly organized pre-production process allows filmmakers to carry out production activities more efficiently while preserving both the creative value and informational quality of the documentary video.

2.1.5.2 Production

The production stage is the phase in which all plans prepared during pre-production are carried out in practice. At this stage, every visual and audio element is recorded according to the concepts that have previously been designed through the storyline, script, and storyboard. Reid and Sanders (2021), in *Documentary Making for Digital Humanists*, explained that the production stage centers on the practical aspects of filmmaking, including camera operation, shot composition, recording contextual footage, adjusting camera settings, conducting interviews, managing lighting, controlling camera movement, and recording audio. They also emphasized that both technical and creative considerations are essential to ensure that the recorded footage communicates meaning effectively and supports the intended message of the documentary.

2.1.5.2.1 Shooting

Shooting during the production stage concentrates on the process of capturing footage, activities, and interviews that serve as the main content of a documentary. According to Reid and Sanders (2021), effective shooting requires filmmakers to

understand several important aspects, including camera stabilization, shot composition, contextual footage, and camera settings. These elements are essential in helping filmmakers communicate ideas visually within the documentary narrative. Effective shooting techniques also enhance the realism and audience engagement of the documentary, influencing how viewers interpret and understand the information being presented.

1. Camera Stabilization

Reid and Sanders (2021) highlighted the importance of camera stabilization in producing clear and watchable footage. Stable shots can be achieved through the use of tripods or controlled handheld techniques. Handheld shooting is often used to create a sense of immediacy and realism, whereas tripods provide steadier and more formal visuals, especially during interview scenes. They also discussed several types of camera movements, including panning, tilting, handheld tracking, and dolly shots, which are commonly applied to follow actions, reveal information, and establish spatial continuity within a scene. Panning refers to horizontal camera movement from left to right, while tilting refers to vertical movement in an upward or downward direction.

2. Shot Composition

In visual composition, an important principle known as the “rule of thirds” is commonly applied in filmmaking. Reid and Sanders (2021) also included the use of head room, looking room, and framing techniques as part of creating balanced and visually effective compositions. Proper composition enables subjects to appear more natural within the frame while also guiding audience attention. The arrangement of visual elements significantly affects how audiences perceive a scene and helps communicate narrative meaning within the documentary.

Furthermore, different shot types and camera angles function as visual strategies for delivering information and establishing context in documentaries. Reid and Sanders discussed the importance of wide shots, mid-shots, and close-ups as essential elements of cinematic visual language, along with the role of camera angles. Wide shots are used to present environmental context, mid-shots focus on interaction and facial expression, while close-ups emphasize emotional details or important objects. Variations in shot size help create visual rhythm and contribute to structuring the documentary narrative. In addition, filmmakers may communicate impressions of strength, authority, or vulnerability by using low-angle or high-angle shots.

3. Camera Settings, Lenses, Focus, and Exposure

Reid and Sanders (2021) explained that filmmakers need to understand various camera settings, such as aperture, shutter speed, ISO, and frame rate, in order to produce visually effective footage. Exposure determines the amount of light that reaches the camera sensor, which directly affects image brightness and the visibility of details. Proper exposure is important to ensure that details in both bright and dark areas remain clearly visible. In addition, frame rate influences the smoothness of motion and the visual representation of movement, while lens selection affects perspective and spatial perception within the frame.

They recommended filming at 24 frames per second (fps) with a shutter speed of 1/50 or 1/48 to create natural cinematic motion. Lens focal length also plays an important role because it influences both magnification and the perception of space, where longer focal lengths produce stronger zoom effects and create a shallower sense of visual space. Furthermore, aperture controls both the depth of focus and the amount of light entering the camera. A smaller aperture setting allows more light to pass into the camera, making it particularly suitable for creating stylized footage with blurred backgrounds.

4. Contextual Footage

Reid and Sanders (2021) explained that filmmakers should plan the capture of contextual footage in addition to recording primary interviews or activity footage. Contextual footage helps audiences gain a better understanding of the environment, situation, and atmosphere surrounding the subject being presented. This type of footage also supports continuity during the editing process and enhances the documentary's visual storytelling by providing additional context and narrative depth.

2.1.5.2.2 Audio Recording

Audio is an essential component of documentary production because it conveys information, atmosphere, and emotional meaning within the narrative. Reid and Sanders (2021) explained that recording audio on location requires careful attention to environmental conditions, microphone placement, and the presence of unwanted noise. Clear audio quality is highly important because poor sound can reduce audience understanding even when the visual elements are presented effectively. They further stated that there are three important aspects involved in the audio recording process.

2.1.5.3 Post Production

Post-production is the final stage in the documentary filmmaking process, focusing on transforming all recorded footage into a complete documentary video. According to Reid and Sanders (2021), post-production involves arranging footage into a planned sequence and serves as a new creative phase in which the documentary can be reconstructed, further developed, or even shaped into a narrative form different from the original concept. Through the editing process, filmmakers are able to discover new meanings, modify the story structure, and adjust the rhythm of the film based on the available footage. They also explained that this stage requires flexibility and critical evaluation of the production outcomes. During post-production, scripts may be revised,

scene sequences may be rearranged, and footage that was not initially intended for inclusion can generate new intellectual or emotional meanings within the documentary.

2.1.5.3.1 Video Editing

Editing is the process of arranging and combining video and audio clips to create a coherent narrative. In their discussion of post-production, Reid and Sanders (2021) divided the editing workflow into three stages: rough cut, fine cut, and final cut editing. The first stage is the rough cut, which is assembled to provide a general overview of the documentary's narrative structure and flow. At this stage, the editing process mainly focuses on pacing, transitions, and identifying sections of footage that are not yet well organized.

Reid and Sanders (2021) also explained that structural problems may be significant enough to require revisions, the addition of new footage, or changes to certain sections of the documentary. This process requires filmmakers to be willing to remove scenes or sequences that do not contribute significantly to the visuals, meaning, or narrative of the film. Once the rough cut is completed, the editing process continues to the fine cut stage. This stage focuses on refining the editing rhythm, improving the timing between cuts, and ensuring audio and visual continuity throughout the documentary. During the fine cut stage, filmmakers begin ensuring that the narrative flows clearly and that audience engagement is maintained from beginning to end. This stage also includes developing the soundscape, selecting music, and refining other audio elements. Reid and Sanders (2021) explained that music and sound design strongly influence the emotional atmosphere of a documentary and can significantly strengthen the narrative impact of the video. The final stage of post-production is the final cut. According to Reid and Sanders (2021), this stage involves applying the finishing touches before the documentary is declared complete. All editing decisions are finalized, temporary materials are replaced with their final versions, and any remaining technical or narrative issues are resolved. Audio balancing is also completed

to ensure that dialogue remains clear while music and ambient sounds continue supporting the atmosphere without distracting from the documentary's main message. This final stage combines all visual and audio elements into a complete documentary ready for publication or public screening.

2.1.5.3.2 Producing Subtitle

Subtitle production is an important part of the post-production stage in audiovisual media. This process includes creating, editing, synchronizing, and reviewing subtitles to make sure that the translated text is clear for the audience and remains synchronized with the video's visual and audio elements. In subtitle production, subtitlers usually follow subtitle guidelines or style guides provided by broadcasting companies or subtitling agencies. For example, over-the-top (OTT) streaming platforms such as Netflix provide public access to their subtitle style guides in many languages through their official website. These guidelines contain technical and linguistic rules, including subtitle formatting, reading speed, punctuation, character limits, and translation consistency, to maintain subtitle quality across audiovisual productions. The use of subtitle guidelines is important because they help create consistent, readable, and professional subtitles that improve audience understanding and viewing comfort

Díaz-Cintas and Remael (2021), in *Subtitling: Concepts and Practices*, explained that the subtitling process consists of several stages. The process usually starts when a production company, distributor, television station, or streaming platform assigns a project to a language service provider (LSP) or subtitling agency. After that, the audiovisual material, dialogue list, and supporting documents are reviewed to check the completeness and accuracy of the information that will be translated. The next stage is spotting or timing, which involves deciding when subtitles should appear and disappear on the screen according to the dialogue, shot changes, and audiovisual rhythm. In some productions, spotting is done directly by the subtitler, while in modern

workflows it is often prepared by technicians through subtitle templates to support the translation process.

Once the template and video materials are ready, the subtitler begins translating the dialogue into the target language while also paying attention to other audio and visual elements such as songs, on-screen text, and background sounds. In modern subtitle production, subtitlers commonly use specialized software such as Aegisub and OOONA, which allow translation, synchronization, subtitle positioning, and subtitle simulation to be completed within one system. Some subtitling programs are also able to detect shot changes automatically, display audio waveforms for dialogue synchronization, calculate subtitle duration, monitor characters per second, and check spelling and grammar.

After the translation process is completed, the subtitles go through revision and quality control (QC) to ensure there are no translation mistakes, typing errors, timing problems, or terminology inconsistencies. At this stage, both the linguistic and technical quality of the subtitles are evaluated to make sure they follow the subtitle guidelines or style guides required by companies or distribution platforms. Once finalized, subtitle files can be exported into different formats such as .srt, .ass, and .vtt depending on the platform, and then distributed through television, film, DVD, or digital streaming services. The development of internet technology has also made file transfer easier through cloud platforms and File Transfer Protocol (FTP) systems.

2.2 Previous Studies

Previous study that supports the use of audiovisual media in education was conducted by Utami et al. (2020). The research examined the relationship between the creation of audiovisual media and students' learning motivation in an Educational Psychology course. The findings revealed a positive and significant relationship between students' involvement in audiovisual media production and their learning motivation. The study concluded that audiovisual media can enhance students'

engagement, creativity, and motivation in the learning process by combining visual and auditory elements that facilitate understanding and participation. The findings of Utami et al. (2020) provide important insights for the present study. While their research focused on the educational benefits of audiovisual media in classroom learning, it demonstrates the potential of audiovisual content as an effective medium for delivering information and fostering audience engagement. Inspired by this perspective, the present study utilizes a documentary video as an educational tool to introduce Audiovisual Translation (AVT) as a profession. By presenting authentic experiences, professional insights, and real-world practices of AVT practitioners, the documentary aims not only to inform Applied Foreign Language students about the profession but also to increase their awareness and interest in pursuing a career in the AVT field.

One of the main documentary references used in this study is “Salt in Their Veins” by Hencke (2025), published on YouTube. The documentary portrays the lives of the Bajo community and their close relationship with the sea as part of their identity and source of livelihood. It presents social and environmental realities authentically through an observational approach that allows audiences to understand the subjects’ lived experiences more deeply. In terms of presentation, the documentary mainly relies on strong visuals, natural sounds, and direct interviews with minimal use of dominant narration. The use of different camera techniques, such as wide shots to establish context and close-ups to capture emotional expressions, creates an immersive viewing experience. In addition, the documentary’s flowing storytelling style allows audiences to interpret meaning independently without excessive direction. In this research, Salt in Their Veins functions as a reference for immersive visual storytelling and the use of audiovisual elements as educational media. This approach inspires the development of a documentary that not only delivers information but also encourages emotional engagement, especially in introducing the profession of Audiovisual Translation (AVT). However, while the documentary focuses on portraying cultural identity and

community life, the present study applies similar cinematic techniques and ambient sound design to support educational content about AVT as a profession.

The second reference is “Secrets of Christ’s Tomb” by Strange (2025), available on the National Geographic YouTube channel. This documentary presents the restoration process of Jesus’ tomb in Jerusalem through an investigative and research-based approach. The documentary combines voice-over narration, expert interviews, visual reconstructions, and on-site documentation in its presentation. Its structured explanation and use of technical terms supported by visual illustrations help make complex information easier for general audiences to understand. The inclusion of on-screen text and subtitles also improves clarity and comprehension. This documentary is particularly relevant to this study because of its ability to present educational content in a structured and informative way through audiovisual media. Its approach also reflects the principles of audiovisual translation, especially in simplifying complex information to improve accessibility. Therefore, it inspires the researchers to create a documentary that introduces the AVT profession in a clear, systematic, and engaging manner. Nevertheless, unlike the present study, the documentary focuses on historical and archaeological topics rather than providing information about career opportunities and professional practices in AVT.

The final reference is “The Hidden Soul of Pekojan” by Kusuma, Sufitri, and Hermawan (2025), published through the Bahasa Asing Terapan YouTube channel. This documentary explores Kampung Pekojan in Semarang by presenting the community’s daily life, culture, and social atmosphere. Through an observational and interview-based approach, the documentary encourages audiences to closely follow the real experiences and cultural identity of the people living in the area. The combination of cinematic footage and interviews creates a natural and authentic storytelling experience. The interview sections become one of the documentary’s strongest elements because they allow the subjects to express their perspectives directly. This

approach is considered relevant to the development of a documentary video intended to introduce AVT as a profession through the direct experiences and insights of interview participants. While this documentary serves as a reference for footage selection and visual support for narration, it does not specifically aim to raise awareness of a professional field, which is the primary focus of this study.

Although these documentaries provide important references in visual storytelling, educational presentation, and interview-based documentary techniques, none of them specifically focus on introducing a professional field such as Audiovisual Translation as an educational medium. Most existing documentaries mainly discuss cultural, historical, or social topics. Therefore, this research aims to fill that gap by creating a documentary video designed to improve public understanding of Audiovisual Translation. The documentary combines informative explanations, cinematic storytelling, interviews, and subtitling elements to make the topic more accessible and engaging for audiences. Through this documentary, viewers are expected to gain a clearer understanding of the real experiences, challenges, and opportunities within the AVT profession, which is still relatively unfamiliar to the general public.