

CHAPTER I

INTRODUCTION

1.1 Background of the Study

Audiovisual Translation (AVT) is a specialized branch of translation studies concerned with the transfer of multimodal and multimedia content across languages and cultures. Pérez-González (2014) defines AVT as encompassing a range of practices, processes, and products in which auditory, visual, and textual elements are rendered into another language while preserving meaning, cultural nuances, and communicative intent.

Within AVT modes such as subtitling, dubbing, and voice-over, live voice-over stands out as a unique and underexplored modality. Unlike post-produced voice-over, where the translation is scripted and synchronized during post-production, live voice-over is performed in real time while the source audiovisual material plays. The translator overlays an oral rendition over the original audio, which remains partially audible, requiring precise coordination of timing and phrasing with visual content.

While live voice-over shares similarities with interpreting particularly in its real-time delivery, it is classified under Audiovisual Translation because the source material is not a live speaker but a fixed audiovisual text. The translator's task involves rendering the meaning of a recorded or scripted product that combines multiple modes (spoken audio, visuals, written text) into another language orally. This aligns with Multimodal Theory (Kress & van Leeuwen, 2001), which serves as the theoretical foundation of this study and provides a framework for analyzing how meaning is created through the interaction of various semiotic modes; auditory, visual, and textual.

In live voice-over, translators must integrate auditory (speech, sound effects), visual (gestures, facial expressions, setting), and textual (captions, visible text)

modes and render them in real time into the target language, maintaining accuracy and coherence. AVT is emerging as a powerful tool for bridging the gap between the source language and the target language. Gonzalez (2014) identifies AVT as one of the fastest-growing areas within translation studies, underscoring its relevance and applicability in diverse fields.

Live voice-over differs fundamentally from interpreting, as interpreting involves converting live spoken discourse from a person, while live voice-over translates from a fixed audiovisual text that may include narration, embedded subtitles, and on-screen written elements. Because it operates on pre-recorded multimodal content, live voice-over belongs firmly within the scope of AVT rather than interpreting studies (Chaume, 2012). The translator's oral output must maintain coherence with both the audio track and the visual content, ensuring the target audience can access the integrated multimodal message. Since this work directly involves translating audiovisual media rather than conversational discourse, it fits squarely within the scope of AVT.

Despite its practical importance in conferences, film festivals, news broadcasts, and live-streamed events, the real-time process of live voice-over has received relatively little scholarly attention compared to subtitling and pre-recorded dubbing. Much of the existing AVT research has concentrated on written text integration (subtitles) or pre-produced oral translations (lip-synced dubbing), leaving a gap in our understanding of how translators operate in time-pressured, unscripted voice-over environments.

In this research, the focus is deliberately placed on the process rather than audience reception. The target audience in this case, Indonesian listeners, is acknowledged only as the end recipient of the translation, but their learning outcomes, preferences, or comprehension are not the subject of analysis. Instead, the study examines the internal, operational aspects of live voice-over: How the translator listens, processes, and reformulates speech in real time and How visual

and auditory cues from the source influence linguistic choices. By isolating these components, this study aims to contribute to AVT scholarship by documenting and analyzing the step-by-step execution of live voice-over, specifically from English to Indonesian.

In an era where immediacy is increasingly valued and cross-cultural audiovisual exchange is constant, understanding the mechanisms of live voice-over is both academically and professionally significant. The findings from this research are expected to offer practical insights for translators, interpreters, and AVT professionals, while also expanding the theoretical application of Multimodal Theory in live, unscripted translation contexts. Therefore, this study focuses on analyzing the process of live voice-over from English to Indonesian within the framework of Audiovisual Translation, employing Multimodal Theory as its primary theoretical basis.

This research is motivated by that gap. In contemporary global communication, live voice-over plays a vital role in facilitating cross-linguistic access to audiovisual content. It is especially relevant in contexts where time constraints make pre-production impossible, such as translating breaking news broadcasts, live webinars, or spontaneous audiovisual presentations. The translator's role here is not only linguistic but also performative, as the voice-over must blend seamlessly with the original audio while maintaining clarity and coherence for the target audience.

Furthermore, recent technological developments, particularly AI-powered audiovisual tools, have begun to transform the way live translation is delivered. Real-time speech recognition, automated transcription, and instant subtitling tools are increasingly integrated into live events. These technologies both assist and challenge human translators: they can support comprehension and timing, but they also raise expectations for immediacy and accuracy. The evolving intersection of human skill and machine assistance adds another dimension to live voice-over

practice, making it an even more urgent subject for academic inquiry.

Poly English Indonesia, established in South Korea in 1999, has expanded its reach to Vietnam and Indonesia. As an edutech company, Poly leverages advanced AI technology to create personalized learning experiences. Its smart platform, POND, seamlessly connects classroom instruction with home-based, self-directed study, enabling students to preview, review, and reinforce their learning. In the context of Poly English Indonesia, an edutech institution integrating AI-assisted learning—live voice-over plays a role in enabling Teaching Assistants to translate audiovisual materials directly for students. Drawing from the researcher's firsthand experience as a Teaching Assistant at Poly English Indonesia inspired this thesis. This role provided valuable insights into the practical application of AVT, revealing its potential to help Teaching Assistants translate from English into Indonesian through live voice-over

This study, titled *"Analyzing the Process of Live Voice-Over in Audiovisual Translation: A Case Study at Poly English Indonesia,"* analyzing the process of live voice-over from English to Indonesian, as one of the AVT methods that expand the scope of translation studies, with the aim of contributing both to AVT theory and to the practical training of translators.

1.2 Statement of the Problem

How did the live voice-over process occur in the auditory, visual, and textual modes within the framework of Audiovisual Translation (AVT) from English into Indonesian?

1.3 Purpose of the Study

The purpose of this study is to examine how the live voice-over process occurs within the framework of Audiovisual Translation (AVT) from English into Indonesian by focusing on the auditory, visual, and textual modes. By analyzing these three modes, the study aims to highlight the multimodal nature of live voice-over and to provide insights into how meaning is constructed, transferred, and

received during the translation process. This research also seeks to contribute to the broader field of Translation Studies by offering a practical perspective on the application of multimodal theory to AVT, particularly in live classroom-based contexts.

1.4 Significance of the Study

Theoretically, this study contributes to the field of Audiovisual Translation (AVT) by offering a clearer understanding of live voice-over as one of its modes. It also applies multimodal theory by examining how auditory, visual, and textual modes interact within the process, thereby enriching the theoretical discussion of multimodality in translation studies.

Practically, this study benefits translators by providing insights into how multimodal elements—auditory, visual, and textual—can be managed in live voice-over. It serves as a practical reference for translators who wish to improve their competence in handling audiovisual materials and expand their skills in live revoicing practices.

1.5 Methodology of the Study

This research adopts a qualitative descriptive approach, which is commonly employed in translation and literary studies to explore complex processes, practices, and meanings that cannot be adequately captured through quantitative measurements. Qualitative research emphasizes an in-depth, contextual understanding of a phenomenon, focusing on interpretation and meaning rather than numerical generalizations (Creswell, 2014; Denzin & Lincoln, 2018).

In the context of this study, qualitative research allows the researcher to examine the dynamic and multimodal nature of live voice-over translation, particularly how translators interact with audiovisual cues in real time. By adopting this approach, the study provides a rich description of the strategies, decisions, and adaptations that emerge during live translation practices in an educational setting.

The source data of this study were obtained from teaching and learning activities in which the researcher participated as a Teaching Assistant at Poly English Indonesia. The data primarily consisted of live voice-over translation practices that were regularly carried out during classroom sessions. Each session lasted approximately one hour and thirty minutes and took place twice a week. Across these sessions, a total of 14 representative data samples were purposively selected to serve as the primary source of analysis.

The samples were chosen to reflect a variety of multimodal translation contexts, including the use of audiovisual flashcards, animated vocabulary videos, thematic charts, and traditional textbooks. This purposive selection ensured that the data captured a comprehensive range of live voice-over translation scenarios, highlighting how different types of multimodal input (visual, auditory, and textual) shaped the translation process. The data were gathered through the use of a participant observation approach, employing both recording and checklist techniques.

The procedures applied in this technique are outlined below;

1. First, during classroom sessions, the researcher performed live voice-over translations of audiovisual and textual teaching materials, directly overlaying oral Indonesian translation over English prompts, videos, or book passages. These translations were documented using audio and video recording tools, ensuring that the timing, synchronization, and interaction between voice-over and audiovisual materials were preserved for later analysis.
2. Second, field notes were taken during and immediately after each session to capture contextual details, such as classroom atmosphere, student reactions, and the situational use of audiovisual tools. These notes provided supplementary insights into how the live translations functioned within the pedagogical setting.

3. Third, the selection of 14 data samples was based on purposive sampling (Patton, 2015), a method that allows researchers to deliberately select cases that are most relevant to the research questions. This approach was particularly suitable for this study, as not all translation instances were equally rich in multimodal content. The chosen samples represented situations where multiple audiovisual cues were present (such as videos with subtitles and background music, or textbooks with diagrams), making them ideal for analyzing the translator's strategies in synchronizing voice-over with visual and auditory input.

For the methods and techniques of the data analysis process employed multimodal discourse analysis (MDA) as the primary framework (O'Halloran, 2011), combined with process analysis techniques as described by Miles, Huberman, and Saldaña (2014).

The analysis unfolded in several structured steps:

1. **Data Segmentation:** Each of the 14 recordings was carefully transcribed and segmented into units of analysis, corresponding to shifts in visuals, changes in spoken dialogue, or the introduction of new textual/diagrammatic content. This segmentation allowed for a fine-grained examination of how different multimodal cues triggered translation decisions.
2. **Coding of Modes:** Using multimodal discourse analysis, the researcher identified and coded different semiotic modes—auditory (spoken words, music, background sounds), visual (images, animations, gestures), and textual (subtitles, written passages in books). These codes formed the basis for understanding the interaction between modalities.

3. **Process Tracing:** Each segment was analyzed to trace the translator's process of aligning oral translation with multimodal input. Special attention was given to timing, synchronization, condensation, and expansion strategies that the translator used to adapt to audiovisual constraints.
4. **Thematic Analysis:** After coding, the researcher identified recurring patterns and themes across the 14 samples. For example, one theme involved the translator's use of images and flashcards as anchors for timing voice-over delivery, while another theme highlighted how background music or songs influenced decisions about pacing and emphasis.
5. **Interpretation and Meaning-Making:** Finally, the coded data and thematic patterns were interpreted in light of Audiovisual Translation (AVT) theories (Díaz Cintas & Remael, 2021; Gambier, 2003), providing theoretical grounding for the observed practices. This interpretive stage emphasized not only the mechanics of translation but also the broader communicative function of live voice-over as a multimodal practice.

By combining recordings, field notes, purposive sampling, and multimodal discourse analysis, the methodology ensured a rigorous and comprehensive examination of live voice-over translation within the AVT framework.

1.6 Scope and Limitation of the Study

This study is limited in scope to the analysis of live voice-over as a form of Audiovisual Translation (AVT) applied to the real-time rendering of English audiovisual materials into Indonesian. The focus of the research lies in the process of translation rather than the assessment of the final product, with particular attention to how auditory, visual, and textual modes interact within a multimodal framework during the live performance. The scope is further narrowed to

classroom-based contexts in which the researcher performed and observed live voice-over activities, resulting in a dataset of 14 recorded instances. Other AVT methods, such as subtitling, dubbing, and pre-recorded voice-over, are not part of this inquiry.

The limitations of the study arise from its design and data collection boundaries. Since the data are drawn exclusively from classroom-based activities, the findings may not fully represent the practices of professional translators working in broadcast media or commercial environments. The relatively small dataset of 14 samples also restricts the generalizability of the results, meaning that conclusions should be understood as exploratory rather than definitive.

Additionally, the absence of a product-oriented analysis means that this study does not evaluate the accuracy or reception of the translations, but instead emphasizes the cognitive and operational processes undertaken by the translator in real time. These limitations highlight both the practical focus of the research and its potential as a foundation for future studies that may explore wider contexts and comparative AVT practices.

1.7 Status of the Study

Previous studies in Audiovisual Translation (AVT) have predominantly examined subtitling and dubbing as the two most institutionalized and researched modes (Luyken et al., 1991; Díaz Cintas, 1997; Dries, 1995; Fawcett, 1996). These studies have shown how the interplay of linguistic transfer and technical constraints affects the translation of meaning in audiovisual products, especially in post-production environments. Research has also expanded into less explored modalities such as voice-over in documentary and television contexts (Franco, 1998, 2001; Pönniö, 1995; Orero, 2005), underlining its specific conventions, terminological issues, and pragmatic functions. However, while these works contribute to the categorization of AVT modes, they largely focus on pre-recorded or professionally produced media, leaving little room for the investigation of live or classroom-based

applications.

At the same time, studies on interpreting (Jensen & Jakobsen, 1998; Jensen, 1999) have explored the challenges of time pressure and real-time decision-making in oral translation, highlighting the cognitive demands on translators. Yet, interpreting research tends to focus on spontaneous oral discourse in political, legal, or conference contexts, without considering scenarios where translators orally render fixed audiovisual texts in real time. This creates a conceptual and methodological gap: AVT studies typically address pre-recorded products, while interpreting studies address live spontaneous speech. What remains underexplored is the hybrid space where real-time oral translation intersects with fixed audiovisual content—namely, live voice-over.

This study addresses that gap by analyzing live voice-over as a hybrid real-time mode that shares operational immediacy with interpreting but is fundamentally grounded in the translation of multimodal, pre-recorded audiovisual texts. By documenting how the translator synchronizes oral rendering with simultaneous auditory, visual, and textual cues, this research clarifies the distinction between live voice-over and interpreting: interpreting deals with spontaneous, unpredictable speech, whereas live voice-over processes stable audiovisual material requiring multimodal alignment.

The novelty of this study lies in its contribution to AVT scholarship by documenting real-time AVT practice outside of post-production environments and in examining the translator's cognitive and operational strategies without prior rehearsal or preparation. While earlier works on AVT have emphasized the product (finished subtitles, dubbing, or voice-over tracks), this study focuses on the process of translation, using multimodal discourse analysis to show how timing, segmentation, and coordination across modes are negotiated in live classroom-based settings. In doing so, it proposes live voice-over as a legitimate sub-mode of AVT that enriches theoretical discussions of multimodality, while also offering

practical insights into pedagogical and professional translation practices.

To the best of the researcher's knowledge, no prior study has systematically examined live voice-over as a real-time AVT mode in a classroom translation context. Thus, this research fills a clear gap by providing empirical and process-oriented insights into a practice that has been overlooked in both AVT and interpreting scholarship

1.8 Organization of the Study

Chapter I: Introduction

This chapter introduces the study by presenting the background of the research, the statement of the problem, the purpose of the study, the significance of the research, the methodology used, the scope and limitations, the status of the study, and an overview of the thesis structure.

Chapter II: Theoretical Framework

This chapter reviews key concepts on Audiovisual Translation (AVT) and AVT methods, specifically, the live voice-over modes, what tools that supports live voice-over modes as part of AVT, the chosen research approach, and the Multimodal Theory that forms the theoretical basis of this research.

Chapter III: Findings and Discussion

This chapter presents the findings and discussion of the study, including a process mapping of live voice-over translation and analyses derived from the collected chosen data.

Chapter IV: Conclusion and Suggestions

This chapter provides the conclusion and suggestions of the study by summarizing the main findings, highlighting the scholarly contributions, and offering recommendations for future research of Audiovisual Translation.