Communication across Borders: Translation & Interpreting; Vol. 05 No. 01 (2025)
ISSN (online): 2766-791X
ISSN (print): 2770-6354

Coherence and Cohesion in Multimodal Translation: A Case Study of Audio Description

Jie LÜ

School of English for International Business, Guangdong University of Foreign Studies, E-mail: 1833904290@qq.com

Abstract

This paper investigates coherence and cohesion in multimodal translation, with a particular focus on audio description (AD) as a representative case. While coherence and cohesion have been extensively studied in interlingual translation, less attention has been paid to their role in multimodal contexts, where meaning emerges from the interplay of linguistic, visual, and auditory modes. Drawing on theories of textual cohesion and multimodal discourse analysis, the study situates AD as a form of multimodal translation that must balance internal textual cohesion with cross-modal coherence. In audio description, cohesion primarily serves to create textual continuity through devices like reference and conjunction. Coherence, in contrast, is built upon a framework of narrative logic, clear temporal-spatial sequencing, and the strategic integration of dialogue and sound effects. The analysis highlights that cohesion secures the internal flow of the AD text, whereas coherence emerges from multimodal alignment and audience cognitive processing. These findings suggest that coherence in multimodal translation is distributed across semiotic resources and must be addressed holistically. The study concludes by emphasizing the theoretical and pedagogical implications of treating coherence as a central concern in multimodal translation.

Keywords: coherence, cohesion, multimodal translation, audio description, accessibility

1. Introduction

The study of textual coherence and cohesion has long been a cornerstone in translation studies and discourse analysis. Classical research has predominantly concentrated on single-modality, interlingual translation—especially written-to-written or speech-to-speech texts—where linguistic mechanisms such as reference, conjunction, ellipsis, substitution, and lexical repetition are the primary tools for ensuring textual connectivity (Halliday & Hasan, 1976). In such contexts, the translator's task has often been conceptualized as preserving semantic and pragmatic continuity across linguistic boundaries. While this language-centered paradigm has yielded valuable insights, it does not fully capture the complex semiotic reality of contemporary communicative practices, in which meaning is frequently constructed through multiple modes beyond language alone.

With the proliferation of audiovisual media, streaming platforms, and accessibility initiatives, translation increasingly involves multimodal texts, in which meaning emerges from the interplay of multiple semiotic resources such as images, sound, gesture, music, and language (Kress & van Leeuwen, 2001; Valdeón, 2024). In multimodal translation, coherence is no longer merely a matter of linking clauses and sentences but rather one of orchestrating cross-modal relations so that audiences perceive a continuous and meaningful whole (Stöckl, 2022).

Among multimodal practices, audio description (AD) constitutes a particularly compelling case. AD is a form of multimodal translation designed to render visual information accessible to blind and partially sighted audiences by converting visual elements into spoken descriptions (Maszerowska, Matamala, & Orero, 2014). Unlike subtitling or dubbing, AD operates under severe temporal constraints, as descriptions must be inserted into pauses between dialogues or sound effects, and it must remain unobtrusive while complementing the film's auditory environment. "In the case of audio descriptions (ADs), the text is written to be read and needs to be both linguistically cohesive within itself and cohesive with the visual content it describes, both in its original form and in its possible translated form"（Taylor, 2014, p.42). The AD script must therefore strike a delicate balance: it must maintain narrative coherence by clarifying character identities, spatial relations, and plot continuity, while also respecting multimodal cohesion with dialogue, sound, and music (Reviers, 2018). These demands render AD an exemplary site for investigating coherence in multimodal translation, since even minor disruptions in descriptive cohesion can lead to significant comprehension gaps for target audiences (Braun & Starr, 2020).

Despite its significance, coherence in AD has received comparatively limited theoretical attention. Existing AD guidelines emphasize clarity, concision, and neutrality (American Council of the Blind, 2010; Ofcom, 2019), yet they seldom provide a systematic account of how coherence operates across modes. Meanwhile, empirical studies in multimodal linguistics and translation have highlighted the need to extend the analytic toolkit beyond Halliday and Hasan's (1976) model, since multimodal coherence involves temporal synchronization, semiotic complementarity, and cross-modal referentiality that are

absent from monomodal texts (Stöckl, 2022; Reviers, 2018). This gap suggests that AD research not only contributes to accessibility studies but also has broader implications for refining multimodal translation theory. This paper therefore seeks to address the following questions: 1) How are coherence and cohesion operationalized and distinguished in AD as a multimodal translation practice? 2) What analytical frameworks can comprehensively capture the construction of coherence in AD?″

2. The Name and Nature of Multimodal Translation

The term multimodal translation has gained increasing traction in recent decades, reflecting the recognition that translation is no longer confined to the interlingual transfer of written texts but extends to meaning-making processes across multiple semiotic resources. Early translation theory often operated within a monomodal framework. This view, however, was challenged by key precursors. Roman Jakobson's (1959/2012) category of intersemiotic translation and Katharina Reiss's (1971/2000) concept of the audio-medial text type explicitly acknowledged the role of non-linguistic modes, thereby laying the groundwork for contemporary multimodal translation studies. Yet as scholars in semiotics and discourse studies have long emphasized, communication is inherently multimodal. Gestures, images, sound effects, typography, spatial layout, and color all contribute to meaning, and in contemporary media these semiotic modes frequently interact in complex ways (Kress & van Leeuwen, 2001).

Before proceeding with our analysis of multimodal translation, it is essential to clarify the conceptual landscape. The field encompasses several distinct yet overlapping research foci: translation in multimodal text, translation of multimodal text and multimodal translation. The following table clarifies the core differences between three key terms, which are often used interchangeably but represent distinct conceptual scopes:

Term	Translation in Multimodal Texts	Translation of Multimodal Texts	Multimodal Translation
Research Focus	The translation of the linguistic components within a multimodal context.	Treating the entire multimodal text as the unified unit for translation.	The process of translation is inherently multimodal, or the product is a new multimodal text.
Theoretical Perspective	Focuses on the strategies and challenges of translating verbal text, with other modes treated as a contextual frame.	Focuses on the coordinated translation and adaptation of multiple modes (text, image, sound) to recreate a coherent target text.	Views translation as a multimodal practice per se, often involving intersemiotic transposition across different sensory modes.
Example	Translating dialogue for subtitles in a film; translating the text labels in an infographic.	The localization of a website or a video game, where text, audio, and visual elements are adapted.	Audio Description (translating images into spoken words); creating a signed language performance of a written poem.

These distinctions are crucial for positioning current research. Studying translation in multimodal texts often aligns with traditional translation studies, expanded to a new context. Translation would not only be seen as a language and culture transfer, but also as a modal transfer (Kaindl, 2013, p.266 ）. Analyzing the translation of multimodal texts requires a holistic, coherence-driven approach. In contrast, investigating multimodal translation means engaging with the fundamental nature of translation as an intersemiotic process.

Against this backdrop, multimodal translation can be defined as the set of translational practices that operate within or across semiotic modes, in which the translator must manage not only interlingual correspondences but also intermodal relations. Gambier (2006, p.6) was among the first to articulate the importance of considering multimodality in translation studies, arguing that audiovisual translation (subtitling, dubbing, voice-over) already demonstrated that translation is never purely verbal but always semiotically hybrid.

The nature of multimodal translation thus differs from interlingual translation in at least three key respects. First, it foregrounds the semiotic ensemble rather than language alone: the task of the translator involves orchestrating relations between words, images, and sounds rather than simply rendering words into words. Second, it is inherently situated in

medium-specific constraints, as meaning is realized not only through linguistic systems but also through temporal synchronization, visual design, and acoustic layering. Third, it brings audience reception into sharper focus: in multimodal contexts, coherence and comprehensibility depend on how viewers integrate information across modalities, which makes reception studies indispensable as Braun (2011, p.648) states: "Coherence has been conceptualised as a process of linking ideas, taking place in the recipient's mind".

To illustrate these properties, consider audiovisual translation (AVT) more broadly. Subtitling requires condensing spoken dialogue into written form while aligning it temporally with visual and auditory cues; dubbing must synchronize spoken target-language utterances with lip movements and performance styles; and audio description transforms visual images into spoken language that fits into temporal gaps without disrupting the soundtrack. In each case, the translation process involves choices that affect the multimodal orchestration of the target text. Thus, multimodal translation is not a discrete subfield parallel to AVT but a conceptual framework that allows us to theorize translation practices that negotiate meaning across semiotic systems.

Importantly, debates about terminology persist. Some scholars prefer the term multidimensional translation (Gottlieb, 2005), emphasizing different semiotic dimensions; others employ multimodal mediation to capture a broader set of practices that extend beyond translation proper (O'Sullivan, 2013). Kress (2010) employs "transduction" to describe the movement of meaning across different semiotic systems, while using "transformation" to refer to the reorganization of meaning within a single mode. As O'Halloran (2011) argues, multimodal discourse analysis provides the methodological tools to study how meaning arises from the interplay of modes. And thus translation studies can draw on the above insights to account for how meaning is preserved, altered, or reconfigured in multimodal transfer.

In short, the name multimodal translation signals a shift in translation studies from language-centered to semiotically inclusive frameworks. Its nature is characterized by hybridity, orchestration, and reception-oriented complexity. Understanding its defining features provides the necessary foundation for analyzing coherence and cohesion in audio description, which epitomizes the challenges of multimodal translation.

3. Coherence and Cohesion in Audio Description as Multimodal Translation

To systematically deconstruct how AD achieves coherence, we propose a tripartite analytical framework that examines the practice from complementary angles: the product (functional), the process (operational), and the reception (cognitive). This framework allows us to move beyond descriptive accounts to a more mechanistic understanding of coherence construction.

Audio description is a typical multimodal translation as Gambier (2013, p.50) states: "it involves the reading of information describing what is going on on the screen (action, body language, facial expressions, costumes, etc.), information that is added to the soundtrack of the dialogue, or to the dubbing of the dialogue for a foreign film, with no interference from sound and music effects". Scholarly work in multimodality aims to analyze how meaning is constructed through both verbal and non-verbal channels, including visual and auditory elements, by systematically examining their available resources and semiotic capacities (Van Leeuwen, 2005). A key aspect of this process lies in the interplay within and across modes, which generates additional layers of meaning beyond what each mode can convey in isolation. This phenomenon has been conceptualized in various ways: Baldry and Thibault (2006) introduced the "resource integration principle," Royce (2007) termed it "intersemiotic complementarity," and Van Leeuwen (2005), within social semiotics, defined a closely related concept as "multimodal cohesion"—that is, "the integration and co-occurrence of different kinds of semiotic resources." Despite differences in terminology, these scholars concur that the integration of diverse modes constitutes the most critical mechanism for meaning-making in multimodal texts, including audiovisual forms, and that the full semiotic potential of each mode is realized only through such interaction. Tseng (2013) extends the cohesion into multimodal text by analyzing image, sound, verbal language, written language, camera movement, framing, colour, and many more operates.

Audio Description (AD) represents a complex multimodal translation practice that requires systematic analysis of its coherence mechanisms. This section examines AD through three complementary analytical frameworks: the functional perspective , the operational perspective and the cognitive perspective.

This tripartite framework is systematically derived from Audio Description's core nature as constrained translation. The functional perspective analyzes the AD product—what meanings (narrative, descriptive, emotional) are constructed. The operational perspective addresses the process—how describers negotiate meaning under temporal and selective constraints. The cognitive perspective centers on reception—to what effect visually impaired persons (VIPs) integrate descriptions into a coherent mental model. Together, they form a holistic cycle and provide a comprehensive understanding of how AD achieves coherence across different dimensions of the translation process.

3.1 Functional Framework: Meaning Construction in AD

The functional framework examines what AD accomplishes as a communicative act, focusing on the core meaning dimensions it must construct to ensure narrative comprehension and engagement. Narrative competence enables multimodal coherence not only in the different modes presented simultaneously, but also in their linear sequence (Meier, 2022, p.9). This statement reveals that narrative competence serves as the core mechanism for constructing multimodal coherence. It functions not only as an integrator of meaning spatially across simultaneously presented modalities but also as a logical connector temporally within linear sequences. Through narrative logic, fragmented, multimodal semiotic resources are woven into a semantically coherent and plot-fluid whole, thereby effectively conveying knowledge, emotions, and perspectives.

3.1.1 Narrative Coherence: Advancing Plot Comprehension

Narrative coherence in AD ensures the logical progression and comprehensibility of the story through strategic description of visual elements. A narratological approach helps identify〞What is narratologically most relevant,〞aiding describers in recreating the filmic experience for blind and partially sighted audiences (Vercauteren, 2012, p.6). This coherence is primarily achieved by maintaining plot connectivity and providing clear spatio-temporal orientation, which together guide the audience through the narrative.

1) Plot Connectivity

As Vercauteren (2012, p.1) notes, content selection lies at the heart of audio description research, focusing on 〞what should be described and how this should be done〞. AD maintains narrative flow by describing key actions and their sequences. For instance, rather than simply noting "a character picks up a book," effective AD might describe "her fingers tremble as she reaches for the leather-bound journal," establishing both action and emotional subtext. The description of causal relationships between visual events - such as showing how a character's discovery of a letter leads to their subsequent actions - ensures visually impaired persons （ VIPs ） can follow the plot's logical development without visual cues.

2) Spatio-temporal Orientation

AD constructs and maintains mental maps of story spaces through consistent spatial references. Descriptions like〞he moves from the dimly lit hallway into the brightly lit ballroom〞create clear spatial transitions. Temporal orientation is achieved through phrases that mark scene changes (“in the meeting room”) or duration ("throughout the night"), preventing temporal disorientation that could disrupt narrative comprehension. Spatio-temporal order is instrumental in structuring narrative events and securing story coherence. It functions as an organizing principle that defines the relations between events within a narrative topic, which is crucial for maintaining a meaningful story logic (Wildfeuer,2014, p.193).

3.1.2 Descriptive Coherence: Building Visual Representation

This dimension focuses on creating and maintaining consistent mental images of characters and environments. It serves as the foundational layer upon which narrative comprehension is built, translating visual cues into a stable mental framework for the VIPs.

1) Character Identification and Tracking

AD establishes character identities through distinctive visual features that remain consistent across scenes. A character introduced as "the tall man with a scar across his left cheek" should be subsequently referenced using these identifiable traits. Crucially, when a character's designation shifts—for instance, from a descriptive label like "the mysterious stranger" to a revealed name like "Mr. Darcy"—the AD must forge a clear logical link. This can be achieved through explicit verbal cues (e.g., "the stranger, who we now know is Mr. Darcy, nods") or through unambiguous situational context, ensuring VIPs perceive the continuity of identity without confusion. The description must also track character movements and spatial relationships, ensuring VIPs can follow who is present and where they are positioned in each scene.

2) Environment and Atmosphere Rendering

Visual atmospherics are translated through careful description of settings and lighting. Rather than simply listing objects, effective AD creates mood through descriptions like “cold moonlight filters through barred windows, casting long shadows across the dusty floor.” Color descriptions serve both identificatory and symbolic purposes, while lighting conditions are described for their emotional connotations rather than just their physical properties. The description of color in audio description is governed by the principle of narrative relevance rather than visual completeness. While color is a fundamental component of visual storytelling and should be included when it serves identificatory, symbolic, or mood-setting functions (Snyder, 2014), descriptors must be mindful of its potential ambiguity for blind-born audiences.

3.1.3 Emotional Coherence: Guiding Affective Response

Emotional coherence enables VIPs to connect with characters and narrative developments at an affective level.

1) Body Language and Expression Interpretation

AD translates non-verbal cues into emotional information through descriptions like “her shoulders slump in defeat”"or “his eyes widen in sudden recognition.” These interpretations must balance observable physical cues with narrative context, avoiding over-interpretation while providing sufficient emotional context for comprehension.

2) Stylized Description for Empathy

The linguistic style of description can enhance emotional engagement during key moments. During emotional climaxes, descriptions might adopt more evocative language or strategic pacing - slowing down for poignant moments or quickening during tense sequences. The vocal delivery itself, including tone and rhythm, works in concert with lexical choices to create empathetic resonance.

3.2 Operational Framework: Practical Challenges in AD

While the functional framework outlines what AD achieves, the operational framework reveals how these achievements are accomplished in practice, focusing on the concrete challenges and decision-making processes describers face.

3.2.1 Temporal Challenges: Creating within Constraints

The time-bound nature of AD presents fundamental challenges for coherence construction.

1) Information Density and Speech Rate Balance

Describers need to strike a balance between linguistic efficiency (to fit time constraints) and comprehensibility (for the audience), often employing syntactic compression strategies like nominalization ("his rapid exit" instead of “he exited rapidly”) to convey maximum information within limited time frames. The speech rate must allow for cognitive processing while fitting within available auditory gaps.

2) Anticipation and Delayed Description

Describers must constantly negotiate between describing immediate visuals and preparing for upcoming developments. Strategic choices about when to describe instantaneously versus when to delay description until a narrative lull require careful judgment about what information will be most coherently integrated at which moment.

3.2.2 Selective Challenges: Determining Content Priority

The necessity of selection from abundant visual information represents a core operational challenge.

1) Relevance Filtering

Describers employ hierarchical selection criteria, prioritizing information based on narrative significance, character development, and emotional impact. This involves distinguishing between essential plot elements, supplementary contextual information, and redundant visual details that can be omitted without compromising coherence.

2) Cultural Code Interpretation

Visual elements with cultural specificity require careful handling. AD must decide when to explicitly explain cultural references, when to approximate through culturally accessible analogues, and when to trust that the narrative context provides sufficient understanding, all while maintaining cultural authenticity and narrative flow.

3.2.3 Linguistic Challenges: Determining Expression

The translation from visual perception to verbal expression involves multiple linguistic decisions.

1) Precise Mapping from Image to Lexicon

Describers must select vocabulary that accurately captures visual qualities while remaining economically efficient. This involves choices between specificity and generality, such as deciding whether to describe a color as "red," "crimson," or "blood-red" based on narrative relevance and time constraints.

2) Objectivity and Implication Tension

Maintaining descriptive neutrality while providing sufficient narrative guidance requires careful balancing. Describers must avoid unjustified interpretation while still offering necessary contextual clues, such as describing a character's “nervous glance around the room” without labeling them “guilty”.

3.3 Cognitive Framework: Audience Processing of AD

While multimodal cohesion refers to observable cross-modal semiotic ties, multimodal coherence remains an interpretive construct that emerges in the audience's cognitive processing. The ultimate test of AD coherence lies in the audience's experience. The cognitive framework shifts perspective to examine how VIPs process and integrate AD information to construct meaningful narrative understanding.

3.3.1 Perceptual Coherence: Forming Basic Mental Representations

This level concerns how VIPs integrate auditory information to form unified perceptions.

1) Auditory and Verbal Integration

VIPs continuously synthesize AD with existing soundtrack elements, creating a coherent perceptual whole. Successful AD anticipates how descriptions will interact with musical cues, sound effects, and dialogue, ensuring they complement rather than compete with each other in the listener's perceptual field.

2) Spatial Mental Model Formation

AD constructs navigable mental spaces through consistent spatial language and vantage points. Descriptions that maintain stable spatial relationships and use consistent perspectival frameworks enable VIPs to build and update mental maps of the narrative environment throughout scene changes.

3.3.2 Cognitive Coherence: Enabling Inference and Understanding

This dimension addresses how VIPs process AD information to construct narrative meaning.

1) Guiding Causal Reasoning

AD provides visual evidence that enables VIPs to infer causal relationships. By describing crucial visual antecedents or consequences that might be absent from dialogue or sound, AD allows VIPs to reconstruct the logical chain of narrative events and character motivations.

2) Disambiguation

AD resolves potential uncertainties arising from ambiguous dialogue or sound by providing clarifying visual context. When dialogue references something visually present but auditorily invisible, or when sounds could have multiple interpretations, AD supplies the necessary visual information to ensure unambiguous comprehension.

3.3.3 Emotional Coherence: Achieving Immersion and Empathy

This final level examines how AD supports emotional engagement and sustained narrative immersion.

1) Emotional Signal Reception

VIPs extract emotional meaning from both the content and delivery of AD. The semantic content provides explicit emotional information, while paralinguistic features like timing, pitch, and rhythm convey implicit emotional cues that help VIPs align their emotional responses with narrative developments.

2) Maintaining Narrative Immersion

Coherent AD prevents disruptions to the "storyworld" experience by ensuring smooth transitions between description and original audio elements. Consistent character voices, uninterrupted emotional through-lines, and seamless integration with the soundtrack all contribute to maintaining the listener's sense of presence within the narrative.

Through the integrated application of these three dimensions, we can appreciate AD as both a technical practice and a cognitive interface that enables comprehensive access to audiovisual narratives. The functional perspective reveals what meanings AD constructs, the operational perspective shows how these meanings are technically achieved, and the cognitive perspective demonstrates how these meanings are ultimately realized in the listener's experience.

4. Conclusion

This study has examined coherence and cohesion in multimodal translation, with a particular focus on audio description (AD) as a paradigmatic case. Building on existing scholarship in translation studies and discourse analysis, the paper has argued that while cohesion in AD refers to the internal linguistic ties within the descriptive text, coherence extends beyond the verbal domain to encompass cross-modal integration and cognitive interpretation. Tseng et al. (2021) regard cohesion as the formal, multimodal textual structure that provides cues, and coherence as the viewer's constructed understanding, with multimodal cohesion playing a functional role in guiding attention and shaping coherent event interpretation. The combination of these two dimensions determines the accessibility, clarity, and interpretive richness of multimodal translation.

The review of multimodal translation theory demonstrated that the concept itself has evolved beyond interlingual frameworks, embracing intersemiotic and transmodal practices. In this context, AD stands out as a form of translation that inherently traverses modalities: it transforms visual cues into verbal narration while simultaneously synchronizing with existing auditory and narrative layers. This dual function reveals the centrality of coherence as an organizing principle. Cohesion ensures that the AD text is syntactically and lexically intelligible, while coherence ensures that this text harmonizes with the multimodal environment of film and television. Coherence in AD is maintained through strategies such as reference, lexical chains, conjunction, and parallel syntactic structures. At the same time, coherence relies on temporal ordering, causal logic, and spatial orientation, all of which must be negotiated across modalities. "Implicit or explicit 'sense-relation' exists between two or more signs of a different or same mode in a given text that helps the viewer to create a coherent textual semantic unit" (Reviers & Remael, 2015, p.54). Importantly, the AD script does not stand in isolation but interacts with dialogue, sound effects, and audience cognitive schemata.

This dual perspective reveals several broader implications for multimodal translation research and practice. First, cohesion and coherence must be understood as interdependent but distinct dimensions. Cohesion provides the textual scaffolding, while coherence ensures interpretive plausibility and emotional resonance. Second, coherence in multimodal contexts is distributed across semiotic resources. Unlike written texts, where coherence is constructed primarily through linguistic means, multimodal texts rely on alignment between modes. This insight confirms that multimodal translation is not merely linguistic transfer but an act of semiotic orchestration. Third, AD exemplifies how coherence can be audience-oriented, requiring translators to anticipate the inferential processes of visually impaired audiences, who construct meaning through partial input from multiple modalities.

The findings also raise methodological and pedagogical implications. For researchers, adopting multimodal discourse analysis frameworks allows for a systematic examination of coherence beyond linguistic cohesion. For practitioners, AD training should emphasize strategies that balance brevity with clarity, ensuring that cohesion at the textual level does not come at the expense of multimodal coherence. Furthermore, given the global rise of streaming media and accessibility initiatives, understanding coherence in AD contributes to the broader goal of inclusive communication, aligning with ethical imperatives in translation studies.

In conclusion, coherence and cohesion in multimodal translation represent not only linguistic concerns but also semiotic, cognitive, and ethical dimensions. AD, as a case of transmodal translation, foregrounds the necessity of addressing these dimensions holistically. Future research may further explore coherence across different genres and modalities—such as subtitling for the deaf and hard-of-hearing (SDH), sign language interpreting, or immersive media translation—to deepen our understanding of how coherence functions in increasingly complex multimodal landscapes. Ultimately, recognizing coherence as a central concern allows scholars and practitioners to refine both theoretical models and applied strategies, ensuring that multimodal translation continues to serve diverse audiences effectively.

Acknowledgement: The paper is funded and one of achievements of the 2024 Guangdong Philosophy and Social Science Planning Project "A Study on Coherence Construction in Audio Description from the Perspective of Multimodal Translation" (Project No.: GD24CWY02).[本文系 2024 年广东省哲学社会科学规划项目"多模态翻译视域下的口述影像连贯建构研究"（项目编号：GD24CWY02）资助成果之一。]

References

American Council of the Blind. (2010). *Audio description guidelines and best practices*. https://adp.acb.org/guidelines.html

Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimodal toolkit and coursebook with associated online course*. Equinox.

Braun, S. (2011). Creating coherence in audio description. *Meta*, 56(3), 645-662.

Braun, S., & Starr, K. (Eds.). (2020). *Innovation in audio description research*. Routledge.

Gambier, Y. (2006). Multimodality and audiovisual translation. In M. Carroll, H.

Gambier, Y. (2013). The position of audiovisual translation studies. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 45–59). Routledge.

Gottlieb, H. (2005). Multidimensional translation: Semantics turned semiotics. In *EU high-level scientific conference series: MuTra* (pp. 1–29).

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Jakobson, R. (2012). On linguistic aspects of translation. In L. Venuti (Ed.), *The translation studies reader* (3rd ed., pp. 126–131). Routledge. (Original work published 1959)

Kaindl, K. (2013). Multimodality and translation. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 257–269). Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.

Maszerowska, A., Matamala, A., & Orero, P. (Eds.). (2014). *Audio description: New perspectives illustrated*. John Benjamins. https://doi.org/10.1075/btl.112

Meier, S. (2022). Digital storytelling: A didactic approach to multimodal coherence. *Frontiers in Communication*, 7, 906268.

O'Halloran, K. L. (2011). Multimodal discourse analysis. In K. Hyland & B. Paltridge (Eds.), *The Continuum companion to discourse analysis* (pp. 120–137). Continuum.

O'Sullivan, C. (2013). Multimodality as challenge and resource for translation. *The Journal of Specialised Translation, 20*, 2–14.

Ofcom. (2019). *Guidelines on audio description*. https://www.ofcom.org.uk

Reiss, K. (2000). *Translation criticism: The potentials and limitations: Categories and criteria for translation quality assessment* (E. F. Rhodes, Trans.). Routledge; American Bible Society. (Original work published 1971)

Reviers, N. (2018). Tracking multimodal cohesion in audio description. *Linguistica Antverpiensia, New Series – Themes in Translation Studies, 17*, 176–198. https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/477

Reviers, N., & Remael, A. (2015). Recreating multimodal cohesion in audio description: A case study of audio subtitling in Dutch multilingual films. *New Voices in Translation Studies*, 13(1), 50-78.

Royce, T. D. (2007). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. D. Royce & W. L. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 63–109). Lawrence Erlbaum.

Snyder, J. (2014). *The visual made verbal: A comprehensive training manual and guide to the history and applications of audio description*. American Council of the Blind.

Stöckl, H. (2022). Multimodal coherence revisited: Notes on the move from cohesion to coherence in multimodal texts. *Frontiers in Communication, 7*, 900994. https://doi.org/10.3389/fcomm.2022.900994

Taylor, C. (2014). Textual cohesion. In A. Maszerowska, A. Matamala, & P. Orero (Eds.), *Audio description: New perspectives illustrated* (pp. 15–28). John Benjamins.

Tseng, C. (2013), Cohesion in Film: Tracking Film Elements, Palgrave Macmillan, Basingstoke.

Tseng, C. I., Laubrock, J., & Bateman, J. A. (2021). The impact of multimodal cohesion on attention and interpretation in film. *Discourse, context & media*, 44, 100544.

Valdeón, R. A. (2024). The translation of multimodal texts: challenges and theoretical approaches. Perspectives, 32(1), 1-13.

Van Leeuwen, T. (2005). *Introducing social semiotics*. Routledge.

Vercauteren, G. (2012). A narratological approach to content selection in audio description: Towards a strategy for the description of narratological time. *MonTI. Monografías de Traducción e Interpretación, 4*, 207–231.

Wildfeuer, J. (2014). *Film discourse interpretation: Towards a new paradigm for multimodal film analysis*. Routledge.