Translation for the Visually Impaired: A Study of Kinetic Units, Implicature, and Coherence in Persian Audio Descriptions

Farnoush Ghadimi[1], Fatemeh Parham[2] iD

[2] M.A., Department of English Translation Studies, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran, Email: farnooshghadimi83@gmail.com

[1] *Corresponding author,* Assistant Professor, Department of English Translation Studies, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran, Email: parham@atu.ac.ir

Abstract

The present study aimed to examine kinetic units, implicature, and coherence in Persian audio descriptions. To this end, 10 Persian audio-described movies, totaling 1,004 minutes, were analyzed, and their audio descriptions (ADs) were documented, resulting in 3,485 records. These records were filtered for descriptions related to characters' movements, yielding 2,095 ADs, which were analyzed in three phases. First, the type of kinetic unit in each AD was identified. Second, the implicature of kinetic units was examined to determine whether it was included in the ADs or ignored. The final phase addressed coherence in ADs. The results revealed that gesture was the most frequent kinetic unit in Persian ADs, followed by manner and posture. Moreover, 96% of kinetic units lacked implicature, and only 2.5% of implicatures related to characters' movements were ignored in audio descriptions, which does not raise serious concerns about accessibility for blind and visually impaired audiences. Regarding coherence, 97% of Persian ADs were found to be coherent, highlighting the substantial accessibility they provide to their target audience. However, accessibility was not solely contingent upon linguistic and pragmatic factors. Technical challenges, such as background music overshadowing ADs, posed additional barriers to media accessibility.

Keywords: audio description, kinetic unit, implicature, coherence, media accessibility, Persian

1. Introduction

As societies increasingly rely on audiovisual content for entertainment, information, and education, the growing production of such content has prompted developers to ensure accessibility for audiences with visual impairments through audio description—a form of multimodal translation that conveys visual events through verbal narration to enhance accessibility. While audio description has been extensively studied internationally, it remains an emerging service in Iran, with only a few institutions involved. SEVINA (a Persian acronym for *cinema for the blind*) is a non-governmental organization pioneering audio description services in Iran. It commenced its activities in 2019 and provides audio descriptions for various movie and animation genres, catering to both young and adult audiences. Additionally, SEVINA offers live audio descriptions in cinemas for both children and adults. According to a 2021 study, aside from SEVINA, Iranian national radio channels also offer audio description services, whereas national television has taken no significant measures. However, the production of audio descriptions on radio remains limited to Iranian feature films (Khoshsaligheh & Shafiee, 2021).

Given the emerging nature of audio description in Iran, academic research in this area remains limited. This study aims to address this gap by examining audio description in the Iranian context, focusing on ten Persian movies audio-described for Persian-speaking individuals with visual impairments. First, it analyzes kinetic units in audio descriptions and explores their types. Subsequently, it investigates the implicature of these kinetic units to determine whether audio describers have adequately addressed them in Persian audio descriptions or overlooked them. The study then examines coherence, analyzing audio descriptions from both a discoursal and narratological perspective to determine their coherence or lack thereof. Finally, it considers the implications of these findings for accessibility.

2. Literature Review

2.1. Multimodality and Audio Description

For an extensive period, translation studies operated within a monomodal framework, largely disregarding texts accompanied by non-linguistic elements such as films, children's books, opera, theater, and comics, which were often analyzed without consideration of these elements. This narrow focus constrained the methodological, theoretical, and analytical scope of the discipline to the basic principles of language transfer. However, with the advent of multimedia, it became evident that texts encompassed more than just linguistic elements.

Pioneering works on paraverbal and multimodal translation, including Poyatos's (2002) exploration of nonverbal communication and Kress and van Leeuwen's (2001) seminal work on multimodal communication, laid the foundation for further discourse in the field. Subsequent contributions by Baldry and Thibault (2006), as well as Gambier (2006), expanded upon this

discourse, emphasizing the multimodal nature of texts and rejecting the notion of purely monomodal forms. Snell-Hornby (2006) focused on linguistic aspects but recognized the significance of investigating what she termed audiomedial texts. Kaindl (2013) directed attention to multimodal translation, while Kourdis (2015) highlighted semiotics as a vital tool in translation research, particularly for audiovisual texts, which inherently possess multimodal characteristics. Gottlieb (2018) defined language as "any communicative system operating through the combination of sensory signs" and text as "any combination of sensory signs conveying communicative intention" (p. 50). Building upon these definitions, he characterized translation as "any process, or product thereof, in which a text is replaced by another text reflecting, or inspired by, the original entity" (Gottlieb, 2018, p. 50).

Kaindl (2013, p. 257) argues that multimodal texts incorporate not only visual (images and graphics), acoustic (sounds and music), and linguistic elements but also texts that appear purely linguistic while featuring multimodal elements such as typography and layout. Kress and van Leeuwen (2001, p. 21) define mode as "semiotic resources which allow the simultaneous realization of discourses and types of (inter)action," emphasizing that modes do not function in isolation but rather in conjunction with other modes. Unlike single semiotic analysis, which isolates visual or musical signs, multimodal analysis operates on the premise that modes collaborate. This collaboration implies that each mode may serve multiple functions, working together to embody multimodality as a principle of text design. Pragmatic and culture-specific factors determine the functions of modes throughout the text, with modes operating in concert to shape the text's design (Kaindl, 2013). Stöck (2004) distinguishes between core modes and sub-modes, where core modes encompass language, image, sound, and music, while sub-modes constitute the components of core modes. Additionally, modes can be classified as core or peripheral depending on their realization through media, highlighting the relationship between mode and medium. The distinction between mode and medium, however, is not always clear (Kaindl, 2013, p. 259).

Mode and medium influence the process and product of translation, and the complexities inherent in defining these terms have led to the emergence of multiple definitions of multimodality in translation studies. Tercedor Sánchez (2010) defines multimodality as the transfer of information through different channels and signs, emphasizing the mixture of sensory channels, media, and modes. However, distinguishing between mode and medium is crucial due to the differing outcomes of text transfer. As Kaindl (2013, p. 261) observes, a clear distinction between media and mode is necessary, as changes in medial communication contexts and transfers trigger the use of different modes. In translation studies, the concept of medium encompasses both the performance form (e.g., opera, theater, comic) and the communication channel (e.g., TV, radio, writing). Expanding on Prunc's (2004) definition, Kaindl (2013) considers translation a cultural interaction that follows conventions to transfer texts modally and medially for a target group other than the original audience, defining intermedial translation as translation across media barriers, including

intracultural and transcultural contexts. Audio description falls under intermedial translation, as it translates motion pictures into verbal language.

Audio description (AD), aimed at enabling blind or visually impaired audiences to understand and enjoy visual content, is applied across various cultural mediums such as films, theater, museums, and visual arts. Braun (2011, p. 646) delineates different audio description techniques based on the source material's modality, whether monomodal, multimodal, or audiovisual. He states that audio description serves as a tool for enhancing media accessibility. Rai et al. (2010, p.3) define audio description as "an additional commentary between the dialogue of a film/television programme that tells the viewer what is happening on the screen so that he/she is able to keep up with the action." Holsanova (2016, p. 49) emphasizes audio description's role in enriching understanding and enjoyment of visual media for audiences with visual impairments. Reviers (2018, p. 23) underscores audio description as a form of media accessibility, providing verbal descriptions of visual elements to ensure audience comprehension. Mazur (2020, p. 228) defines audio description as a service facilitating visual content accessibility for the visually impaired, stressing coherence alongside accessibility.

2.2. Approaches to Study Audio Description

Studying audio description involves various approaches aimed at understanding its production, reception, and impact. Mazur (2020) outlines three primary approaches: text-based studies, reception studies, and experimental studies. Text-based studies encompass discourse analysis, corpus-based studies, and narratology, focusing on analyzing the linguistic and narrative aspects of audio description. Reception studies, on the other hand, investigate the effectiveness of description methods from the perspective of blind and visually impaired audiences. Methodologies borrowed from sociology, such as focus groups, in-depth interviews, and questionnaires, are commonly employed to assess reception. Additionally, more advanced methods, including those from psychophysiology, have been applied over time to delve deeper into audience responses (Mazur, 2020, p. 237). Experimental studies provide objective data on the creation and reception of audio description, utilizing techniques such as eye tracking and examining the cognitive processing of visual and auditory information (Mazur, 2020, p. 240).

Holsanova (2016, pp. 50-51) offers a complementary framework, delineating three perspectives on audio description: production, reception, and the 'meeting of minds'. The production perspective centers on the work of the audio describer, whose task is to make visual content accessible to blind and visually impaired audiences. By selecting pertinent information from visual scenes and conveying it through vivid descriptions, describers aim to evoke internal imagery in the minds of the audience. A key challenge lies in synchronizing verbal descriptions with the sounds, voices, and dialogues of the original content. The reception perspective shifts focus to the audience, emphasizing the need for a deeper understanding of how audio description is perceived

and experienced by users. Despite scant research in this area, insights into audience reception are crucial for optimizing description for target audiences. The 'meeting of minds' perspective highlights the interaction between describers and audience members, underscoring the importance of feedback to refine and tailor audio description (Holsanova, 2016, p. 57). While immediate feedback opportunities may be limited, online platforms and streaming services offer new avenues for user input and engagement. Through these multiple perspectives, researchers seek to advance our understanding of audio description and enhance its effectiveness in facilitating access to visual content for individuals with visual impairments. Some instances of such research at international level and in Iran are briefly reviewed in the Section 2.3.

2.3. Studies on Audio Description

In their exploration of audio description, Matamala and Ramael (2015) examined how films that rely heavily on visual effects rather than narrative are rendered in audio description for visually impaired and blind audiences. Their study, encompassing two diverse movies, *2012* and *Hero*, scrutinized contrasting genres—one a disaster film aimed at thrilling audiences with continuous action, and the other, a martial arts epic where effects serve an aesthetic function. Employing a narratology perspective, which falls within text-based studies, they underscored narrative components crucial for effective audio description, including "what is done, where/when and by whom", alongside considerations of space, time, and action treatment, as well as the presentation of events in terms of "perception, timing, and phrasing" (Matamala & Ramael, 2015, p. 64). Matamala and Ramael emphasized the necessity of understanding how authors construct stories and how audiences reconstruct them, particularly in the context of special effects films, where traditional narrative structures might be challenged. They noted that while such films often foreground visual and aural effects for their spectacular appeal, audio description faces the challenge of translating these effects into verbal descriptions that aid in reconstructing the narrative for visually impaired viewers (2015, p. 65).

Similarly, Braun (2007) advanced a discourse-based approach to audio description, investigating its complex cognitive-linguistic and intermodal activities. Emphasizing audio description's role in producing verbal discourse that describes essential visual and auditory elements of multimodal content, Braun posited that "audio description creates a new multimodal discourse integrating audio-described content, verbal, and auditory factors." Essential to this process is the establishment of coherence, both "locally within individual audio description sections and globally across scenes", ensuring that "the narrative retains its integrity across different modes of communication" (Braun, 2007, pp. 9–10). Drawing upon discourse analysis methodologies, Braun scrutinized the comprehension and production processes involved in audio description, shedding light on the intricate interplay between different semiotic modes and the knowledge, experience, and expectations of participants. Furthermore, Braun highlighted the importance of

"inferential models of communication" in understanding how verbal and visual representations in audio description elicit "explicit and implicit assumptions", thus enabling viewers to derive "rich meanings and interpretations akin to sighted audiences" (Braun, 2007, pp. 7–9). Ultimately, Braun's model underscores the nuanced decision-making processes involved in audio description production, as audio describers navigate between the explicit and implicit dimensions of verbal discourse to craft a coherent and meaningful experience for visually impaired audiences.

Some studies on audio description have adopted a reception study approach to understand audience preferences and experiences. Chmiel and Mazur (2022), for instance, explored the preferences of individuals with various vision dysfunctions regarding audio description, finding that "group differences were less pronounced than expected". They concluded that "audio description should aim for middle-of-the-road solutions to accommodate diverse audience needs." Similarly, Bardini (2020) investigated the film experience of blind and partially sighted viewers through audio description, contrasting conventional and interpretative audio description styles. Results indicated that "interpretative approaches, such as cinematic and narrative audio description, offered a more satisfying film experience by effectively conveying emotional nuances expressed through film language." These studies underscore the importance of understanding audience preferences and tailoring audio description accordingly to enhance accessibility and enjoyment for diverse viewers.

In contrast, Jankowska (2021) adopted an experimental approach to analyze the scripting process of audio description for films. Describers from Poland and Spain prepared audio descriptions for clips from their respective films, with data collection involving keylogging, think-aloud verbalizations, screen, and face recording. Qualitative analysis revealed that the scripting process resembles writing and translating, with three main processes and seven intertwining subprocesses identified. This investigation provided insights into decision-making processes in audio description scripting, highlighting its multifaceted nature.

There are other studies that have examined audio description with different focuses, shedding light on various aspects of its application and evolution. For instance, Brescia Zapata and Matamala (2020) conducted a descriptive analysis of audio description in violent scenes from Quentin Tarantino films, discerning trends in the treatment of violence in audio description.

Meanwhile, Fidyka and Matamala (2021) explored the implications of producing audio descriptions for 360° narrative videos, suggesting novel approaches to address the immersive nature of this emerging media format. Additionally, Greco and Romero-Fresco (2023) deliberated on universalist, user-centered, and proactive approaches in media accessibility, advocating for a nuanced understanding of accessibility paradigms and emphasizing the importance of diversity in addressing access needs. Furthermore, Hirvonen et al. (2023) provided insights into collaborative audio description practices, highlighting the roles of blind and sighted professionals in cooperative audio description production and underscoring the significance of user-centered translation and editing approaches. These studies collectively contribute to the ongoing discourse on audio

description, enriching our understanding of its applications, challenges, and potential avenues for advancement on the global stage.

Although audio description services are a very recent practice in Iran, there are a few academic studies in this area. For instance, Kavid (2022) examined the prioritization of characterization elements in audio descriptions for both children's and adults' movies in Iran. The study identified the priorities of characterization elements such as action, speech, environment, and external appearance, highlighting differences between audio descriptions for different age groups. Salehi (2021) conducted a comparative analysis of English and Persian audio descriptions, evaluating their compliance with Independent Television Commission (ITC) guidelines. The study found that both English and Persian audio descriptions showed a high level of compliance, although facial expressions were often overlooked, suggesting the need for a more systematic approach to their description. In another study, Salehi and Mousavi Razavi (2023) examined scene change indicators in English and Persian audio descriptions comparatively. They found that English ADs demonstrated a 10% higher degree of conformity to ITC guidelines. Boroon (2021) further explored the narrative constituents of time, space, and character to determine which was prioritized in English and Persian audio descriptions. She also examined the challenges associated with describing facial expressions in both languages, emphasizing the importance of developing local guidelines tailored to the Persian language and culture. Moreover, Sheikholeslami and Ehteshami (2023) analyzed the description of characters' facial expressions of emotion in Persian audio-described movies and found that negative emotions such as sadness and anger are depicted more frequently in audio descriptions.

Turning to the legal and practical aspects of media accessibility in Iran, Shokoohmand and Khoshsaligheh (2019) discuss the legal status of media accessibility and the satisfaction of the deaf and hard of hearing (DHH) community with existing services. Despite the absence of direct references in the Iranian Constitution to the rights of people with special needs, efforts such as subtitling and sign language interpreting have been made to enhance accessibility. However, these initiatives lack government regulation and fall short of meeting the expectations of the Iranian DHH community. Meanwhile, Shafiei and Khoshsaligheh (2021) provide an overview of global developments in audio description and assess the current status of audio description in Iran. Through document analysis and interviews with audio description practitioners, they reveal that audio description production in Iran is primarily undertaken by non-governmental organizations, with limited involvement from national mass media. The study underscores the need for further research and development in this emerging field, particularly in expanding audio description services beyond radio and into television.

3. Methodology

3.1. The Corpus

The corpus of the study included 10 Persian movies that were audio-described for blind or visually impaired audiences and distributed by Filimo, an Iranian digital streaming platform (*https://www.filimo.com/*). While regular movies on Filimo required a subscription for access, audio-described movies could be accessed free of charge, potentially to enhance accessibility for visually impaired viewers or to encourage inclusivity. These movies were originally in Persian, with Persian-language audio descriptions tailored for Persian-speaking audiences. All the movies were released in recent years, between 2012 and 2019, spanning various genres such as comedy, crime, drama, biography, and melodrama. Audience comments, feedback, and ratings on Filimo were also taken into account when selecting the movies, with all chosen films receiving ratings above four stars out of five. The average length of each movie was approximately 120 minutes, with the shortest being *Shirin's Palace* (81 minutes) and the longest being *When the Moon Was Full* (131 minutes). The movies selected for the study are listed in Table 1.

Table 1

Audio Described Movies

No.	Movie Title	Year
1	*Jahan Dance with Me*	2019
2	*6.5 per Meter*	2019
3	*When the Moon Was Full*	2019
4	*Takhti*	2019
5	*Don't Be Embarrassed*	2018
6	*The Singer*	2019
7	*Lovely Trash*	2013
8	*African Violet*	2019
9	*Shirin's Palace*	2019
10	*Eyes and Ears Closed*	2019

3.2. Data Collection

To collect the data, all 10 audio-described movies were thoroughly watched, totaling 1,004 minutes, and all audio descriptions were transcribed and documented, resulting in 3,485 records. Since the study focuses on kinetic units related only to characters, Vercauteren's (2012) three basic constituents of narrative—time, space, and character—were employed to determine which audio descriptions pertained to characters as opposed to time and space. Furthermore, to ensure the accurate selection of character-related audio descriptions, Prince's (2012, p.71) definition of character was applied, defining a character as someone or something "foregrounded at least once in the narrative rather than relegated to the background and made part of a general context or setting." This process identified 3,100 audio descriptions related to characters out of the total 3,485

records, with the remainder pertaining to time and space. In a subsequent filtering step, character-related audio descriptions were examined to identify those describing character movement, based on Poyatos's (2010) classification. In other words, audio descriptions containing kinetic units were marked, yielding 2,095 out of 3,100 records (see Table 2). These 2,095 audio descriptions, which included a character's kinetic unit, were documented along with corresponding screenshots to be analyzed in three phases.

Table 2

Audio Descriptions Statistics

No.	Movie Titles	Minutes	Total Number of ADs	Number of ADs about characters	Number of ADs about character's body movement
1	*Jahan Dance with Me*	93	222	185	141
2	*6.5 per Meter*	130	452	390	264
3	*When the Moon Was Full*	131	552	485	297
4	*Takhti*	105	407	342	238
5	*Don't Be Embarrassed*	89	330	327	230
6	*The Singer*	104	375	358	238
7	*Lovely Trash*	85	249	235	173
8	*African Violet*	93	327	287	192
9	*Shirin's Palace*	81	375	322	199
10	*Eyes and Ears Closed*	93	196	169	123
	Total	**1004**	**3485**	**3100**	**2095**

3.3. Data Analysis

The data analysis proceeded in three phases. In the initial phase, as described under the data collection process, the type of kinetic unit present in each of the 2,095 audio descriptions was identified. To this end, Poyatos's (2010) classification of kinetic units—namely gesture, manner, and posture—was employed. According to Poyatos, a gesture is "a conscious or unconscious body movement made mainly with the head, the face alone, or the limbs." Manner, "although similar to gesture, is a more or less dynamic body attitude that is mainly learned and socially codified." Thus, manners are closely related to gestures in definition, with the only difference being that they are socially loaded. He defines posture as "a conscious or unconscious general position of the body" (Poyatos, 2010, p. 375). At the end of this phase of analysis, the frequency of audio descriptions containing gesture, manner, and posture was determined.

The second and third phases of analysis relied on the framework proposed by Braun (2007, p. 10), which adopts a discourse analysis approach to the study of audio description. Braun's framework examines three constituents in audio descriptions: information processing, inference, and coherence. The present study focuses on the constituent of inference and implicature in the second phase of analysis and the constituent of coherence in the third phase of analysis.

Thus, the second phase of analysis examined whether a kinetic unit in the movie contained an implicature and, if so, whether the audio describer explained the implicature in the audio description or ignored it. To identify implicatures operationally, Gutt's (2000) relevance theory was employed. According to Gutt (2000), in an act of communication, explicatures are decoded, and implicatures are inferred. The communicator should provide stimuli for the audience to be able to infer what they mean. In the present research, kinetic units serve as the stimuli, and implicatures are propositions that can be understood by relying on the context, other audio descriptions, and movie dialogues. Put simply, each implicature was examined to ascertain whether it conveyed anything beyond the physical movement alone and, if so, what it conveyed.

The last phase of analysis addressed coherence in audio descriptions. According to Di Giovanni (2014, p. 72), cohesion relates to the lexicogrammatical systems, whereas coherence concerns the "structuring of information." In this study, the structuring of information is examined with regard to the three narrative constituents of character, time, and space. Due to the strict time constraints of audio descriptions, which cannot exceed certain lengths, they can only address one constituent at a time. Therefore, to operationally examine coherence, if an audio description addresses only one narrative constituent (time, space, or character), it is marked as coherent. If an audio description switches from one constituent to another, the coherence is disrupted, and that audio description is marked as incoherent. At the end of this phase, the frequency of coherent and incoherent audio descriptions was calculated.

4. Results

The results of all three phases of data analysis are presented in the following sections. Initially, a few examples, along with the findings regarding the types of kinetic units, are provided. Subsequently, instances of implicature analysis in audio descriptions are presented, accompanied by relevant statistics. The final section of results examines two cases of coherent and incoherent audio descriptions, followed by a presentation of statistics on coherence.

4.1. Analysis of Kinetic Units

Kinetic units, as defined by Poyatos (2002), are body movements driven by psychomuscular activity. These movements are categorized into three main types: gestures, manners, and postures. The first example provided below is taken from the movie *6.5 per Meter* at the timestamp 00:02:09. In this scene, a police officer named Hamid is depicted as chasing a boy who is accused of being a criminal. The action of the police officer running constitutes a body movement (see Figure 1).

Figure 1

Gesture – 6.5 per Meter

The audio description corresponding to this scene reads: "حمید به دنبالش توی کوچه پس‌کوچه‌های تنگ و باریک شهر می‌دوه" (literally, *"Hamid is chasing him in long and narrow alleys of the city"*). The audio description refers to the act of chasing, which includes movements of the arms and legs; thus, the kinetic unit in this audio description is marked as a gesture.

The second example is from *When the Moon Was Full*, at the timestamp 00:13:57. In this scene (see Figure 2), the two characters, Abdolhamid and Faezeh, are having their wedding. Abdolhamid, the groom, delighted to have married his dream girl, expresses his happiness by showering money on guests. In Iranian culture, this money is known as Shabash and is believed to bring luck and happiness. The corresponding audio description reads: "عبدالحمید شاباش رو سر مهمونا می‌پاشه" (literally, *"Abdolhamid is throwing shabash on guests"*), which refers to the act of showering money, a socially coded behavior, and is thus classified as an instance of manner.

Figure 2

Manner – When the Moon Was Full
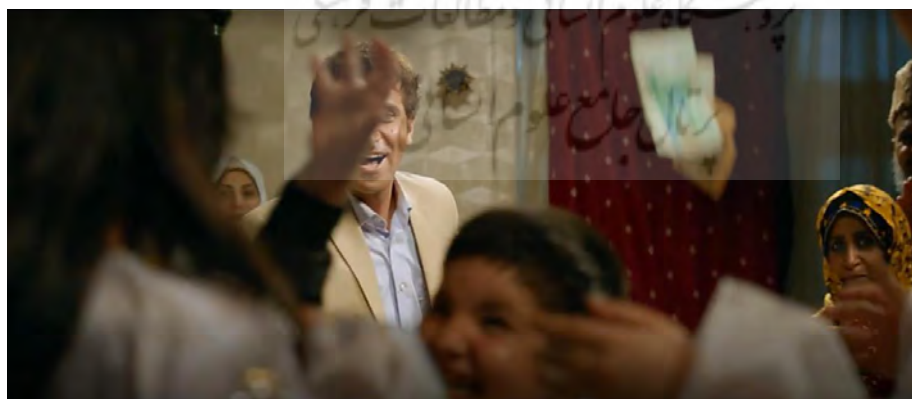
The third example in Figure 3 is a scene from *African Violet*, at 01:02:45, showing one of the characters, Shokouh, who has returned home after a night in custody and is listening to the messages left on the phone's answering machine. She is leaning on the wall, which is an instance of posture.

This posture is described in the corresponding audio description as: "شکوه به دیوار تکیه داده و به پیغام‌گیر گوش می‌دهد" (literally, *"Shokouh is leaning on the wall, listening to the answering machine"*).

Figure 3

Posture – African Violet

All 2,095 audio descriptions were assessed to identify the type of kinetic unit they contained. The results revealed that gesture was the most frequent kinetic unit in audio descriptions, with 1,809 instances, followed by posture with 230 instances and manner with 56 instances. These findings are summarized in Table 3 below.

Table 3

Kinetic Unit Types

No.	Movie Titles	Number of Gestures	Number of Postures	Number of Manners
1	*Jahan Dance with Me*	104	28	9
2	*6.5 per Meter*	239	23	2
3	*When the Moon Was Full*	249	37	11
4	*Takhti*	199	29	10
5	*Don't Be Embarrassed*	202	24	4
6	*The Singer*	203	24	9
7	*Lovely Trash*	165	4	5
8	*African Violet*	167	24	1
9	*Shirin's Palace*	177	19	4
10	*Eyes and Ears Closed*	104	18	1
	Total	1809 (86%)	230 (11%)	56 (3%)

4.2. Analysis of Implicatures

Analysis of implicatures in the second phase initially involved examining whether each individual audio description conveyed any implicit meaning beyond the explicit message through the kinetic unit. If an implicature was identified, the analysis then determined whether the audio describer included that implicature in the audio description or disregarded it. Two relevant examples are presented below.

Figure 4, from *The Singer*, at 00:35:34, illustrates an instance of a gesture, where the character Foad puts his finger to his nose. This gesture has an implicature and serves as a signal for

silence. Foad is talking to his sister, Ziba, on a video call while he is in his employer's bedroom, where she is asleep. Foad uses this shushing gesture to indicate to his sister to speak quietly, so as not to wake the employer. The audio description corresponding to this scene explicates the implicature and includes the meaning of the gesture in the audio description: "به زیبا اشاره می‌کنه که ساکت باشه" (literally, *"He gestures to Ziba to be quiet."*).

Figure 4

Implicature – The Singer

Figure 5 is a screenshot taken from *Eyes and Ears Closed*, at the timestamp 01:24:50. This scene illustrates two characters, Behrouz and Saeedeh, who are being held hostage on the second floor of a house and are discussing ways to escape. They are looking through the window, and Behrouz is pointing to a rope outside, which could serve as a means of escape. This act of pointing is an instance of a gesture and implies that they can use the rope to escape from the house. The corresponding audio description states: "به طناب جلوی پنجره اشاره می‌کنه" (literally, *"He points to the rope in front of the window"*), which only refers to the explicit meaning of the gesture and disregards its implicature. Therefore, visually impaired or blind audiences listening to this movie will not understand why he points to the rope.

Figure 5

Implicature – Eyes and Ears Closed

All kinetic units were examined to determine whether they conveyed any implicature. It was found that the kinetic units in 2,007 out of 2,095 audio descriptions did not have any implicature,

while the remaining 88 did. These 88 instances were further analyzed to determine whether their implicatures were explained in their corresponding audio descriptions or overlooked. The results revealed that only 38 (1.5%) audio descriptions included the implicature of the kinetic unit, while in 50 (2.5%) instances, the implicature was ignored. Table 4 outlines these findings.

Table 4

Analysis of Implicatures in Audio Descriptions

No.	Movie titles	Kinetic units with implicature		Kinetic units without implicature
		Implicature included in AD	Implicature ignored in AD	
1	*Jahan Dance with Me*	0	1	140
2	*6.5 per Meter*	3	10	251
3	*When the Moon Was Full*	1	11	285
4	*Takhti*	2	3	233
5	*Don't Be Embarrassed*	6	1	223
6	*The Singer*	3	5	230
7	*Lovely Trash*	5	7	161
8	*African Violet*	8	3	181
9	*Shirin's Palace*	7	6	186
10	*Eyes and Ears Closed*	3	3	117
	Total	38 (1.5%)	50 (2.5%)	2007 (96%)

4.3. Analysis of Coherence

The third phase of analysis addressed the coherence of audio descriptions. For an utterance to be coherent, it should make sense to the audience in a particular setting. Since the audience of audio-described content only has access to auditory and verbal channels, coherence is crucial. For an audio description to be coherent, it needs to focus on only one constituent—character, time, or space. If an audio description shifts from describing one constituent to another within a single description, coherence is disrupted.

Take, for instance, the scene from *African Violet*, at 00:19:03 (Figure 6). Shokouh is depicted preparing to leave the room after curling her hair and getting ready for bed. The corresponding audio description—"شکوه به طرف در اتاق می‌رود، از پنجره" (literally, *"Shokouh goes towards the door of the room; from window …"*)—begins by describing Shokouh's action but then abruptly switches to describing the space, leaving an unfinished sentence and disturbing the coherence.

Figure 6

Incoherent Audio Description – African Violet

Another example is from *The Singer* (00:11:03). Three characters are depicted in this scene (Figure 7): Ebrahim, the father, Foad, the son, and Ziba, the daughter. Foad is shaving his father's face, and Ziba is talking to him. The audio description reads: "فواد داره ریشای ابراهیم رو می‌تراشه" (literally, *"Foad is shaving Ebrahim's beard"*), which focuses only on one narrative constituent, i.e., character, and describes a gesture, thus maintaining coherence.

Figure 7

Coherent aAudio Description – The Singer

The examination of 2,095 audio descriptions in terms of coherence illustrated that 97.5% were coherent, while only 2.5%, or 53 instances, were found to be incoherent, transitioning between constituents. The proportion of incoherent audio descriptions is minimal, indicating the high quality of Persian audio descriptions in maintaining coherence. See Table 5 for details on the coherence analysis.

Table 5

Analysis of Coherence of Audio Descriptions

No.	Movie titles	Number of coherent audio descriptions	Number of incoherent audio descriptions
1	*Jahan Dance with Me*	140	1
2	*6.5 per Meter*	255	9
3	*When the Moon Was Full*	288	9
4	*Takhti*	233	5
5	*Don't Be Embarrassed*	223	7
6	*The Singer*	235	3
7	*Lovely Trash*	164	9
8	*African Violet*	182	10
9	*Shirin's Palace*	199	0
10	*Eyes and Ears Closed*	123	0
	Total	2042 (97.5%)	53 (2.5%)

The figures presented below illustrate the final outcome of this study. In Figure 8, it is evident that among 2,095 kinetic units analyzed in Persian audio descriptions, the majority (86%) were classified as gestures, followed by manners at 11%, while only a small percentage (3%) focused on characters' body postures. This distribution underscores the prevalence of movement among characters throughout the movies examined. Notably, almost every second of the movies featured some form of character movement, justifying the significant combined percentage of gestures and manners (97%). In contrast, postures, which primarily denote static stances, accounted for a mere 3% of the kinetic units analyzed.

The relatively small percentage of manners (11% compared to 86% for gestures) can be attributed to the fact that manners are body movements with social and cultural associations, and not all movements meet this criterion. This stark contrast underscores the prevalence and significance of gestures in Persian audio descriptions, highlighting their pivotal role in conveying character actions and emotions.

Moreover, the high percentage of gestures could reflect a broader trend in Persian cinema, where action-oriented narratives or character-driven movements take precedence over reflective, stationary moments. This may align with cultural preferences for expressive, active storytelling, which relies on visible actions to communicate character development and emotional shifts. The focus on movement over stillness also reflects a storytelling tradition that prioritizes action and dynamism in both visual and audio representations.
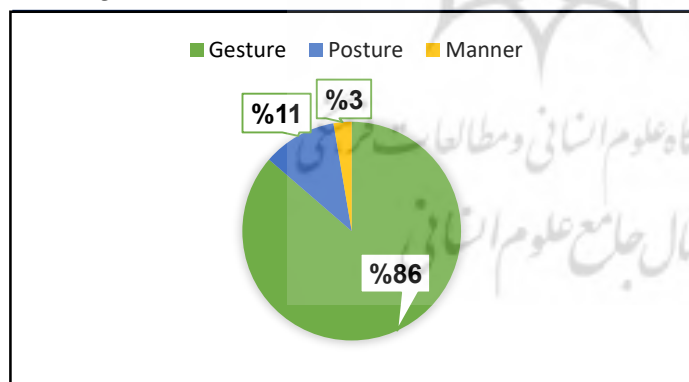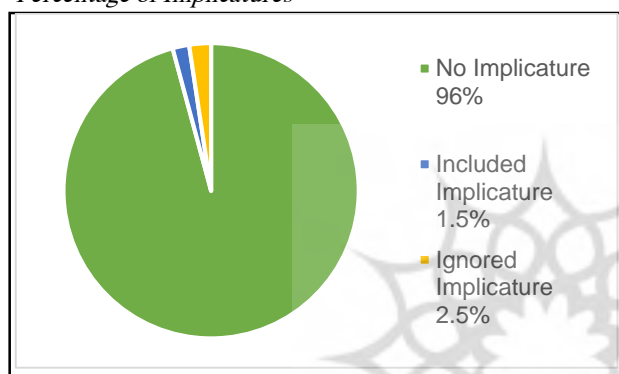
Figure 8

Percentage of Kinetic Units

Figure 9 displays the findings regarding the implicature of kinetic units, revealing that a significant majority (96%) of units lacked implicature. This could indicate that implicature is a low-risk feature in Persian audio descriptions. Only 2.5% of implicatures related to characters' movements were ignored in audio descriptions. While this very small percentage does not significantly affect the accessibility of the movies for blind and visually impaired audiences, it could suggest a potential area for improvement in the richness of the descriptions. Incorporating more

implicature could enhance audio descriptions by providing a deeper understanding of the social dynamics and emotional undercurrents present in characters' movements.

On the other hand, this minimal percentage of ignored implicatures suggests that Persian audio descriptions are well-optimized for providing an accurate representation of characters' actions, without significant omissions that would impede the understanding of key events and diminish accessibility. In fact, Persian audio descriptions are largely successful in maintaining the integrity of character movements and their social or emotional implications, ensuring that blind or visually impaired audiences receive as complete a representation of the film as possible.

**Figure 9**

Percentage of Implicatures

Finally, Figure 10 demonstrates the frequency of incoherent and coherent audio descriptions, with only 3% categorized as incoherent. The fact that 97% of Persian audio descriptions examined were found to be coherent again highlights the substantial accessibility afforded to their special audience.

Moreover, the minimal occurrence of incoherence suggests that the majority of the descriptions successfully align with the visual content, maintaining logical flow and consistency, which is crucial for accessibility. This coherence is particularly important in ensuring that blind or visually impaired audiences can follow the storyline without confusion or disruption. However, the presence of even a small percentage of incoherence might signal areas for improvement in ensuring uniformity and accuracy across all descriptions. Such instances of incoherence, though limited, could potentially disrupt the viewing experience for the audience.

Figure 10

Percentage of Coherent and Incoherent ADS
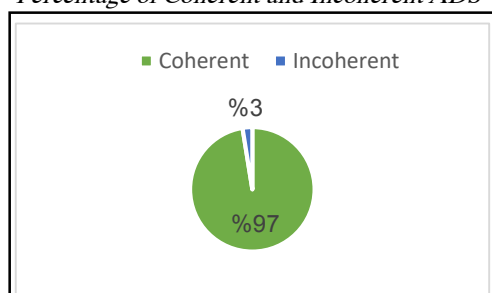
5. Conclusion

The findings of this study underscore the critical importance of linguistic and pragmatic features, such as implicature and coherence, in the creation of audio descriptions, as they significantly contribute to audience comprehension. As Reviers (2018) contends, audio description constitutes a form of media accessibility, serving the dual purpose of ensuring service availability and presenting information in a manner easily understandable to the intended audience. Particularly for blind and visually impaired audiences, who rely solely on auditory channels, audio descriptions play a pivotal role in translating essential visual cues into auditory form. Any omission or oversight of these cues can impede audience comprehension, emphasizing the significance of addressing implicatures in audio descriptions to maintain accessibility.

Similarly, coherence within audio descriptions is paramount for audience understanding. Disrupting coherence, such as through abrupt narrative shifts, can confuse listeners and hinder accessibility. This study underscores the intrinsic link between explaining implicatures and preserving coherence; 97.5% of audio descriptions had either no implicature or their implicatures were explained, and 97% were coherent, indicating that proper treatment of implicatures contributes to coherence and facilitates decoding of communicative cues for the audience.

However, the attainment of accessibility is not solely contingent upon linguistic and pragmatic factors. Technical challenges, such as overshadowing audio descriptions with background music or original dialogues, pose additional barriers to media accessibility, as observed in several instances in the present study. These issues not only disrupt coherence but also heighten the difficulty of comprehending information, directly opposing the aim of media accessibility to provide easily understandable content. Addressing these technical challenges is thus imperative to ensure seamless access to media content for individuals with visual impairments.

References

Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*. Equinox.

Bardini, F. (2020). Film language, film emotions and the experience of blind and partially sighted viewers: A reception study. *Journal of Specialized Translation, 33*.

Boroon, B. (2021). *A comparative study of narratological constituents in Persian and English audio-described movies* (Unpublished master's thesis). Allameh Tabataba'i University.

Braun, S. (2007). Audio description from a discourse perspective: A socially relevant framework for research and training. *Linguistica Antverpiensia New Series, 6*, 357–369.

Braun, S. (2011). Creating coherence in audio description. *Meta, 56*(3), 654–662.

Brescia Zapata, M., & Matamala, A. (2020). Violence in audio description: A descriptive study of three films by Quentin Tarantino. *TRANS: Revista de Traductología, 24*, 111–128. https://doi.org/10.24310/TRANS.2020.v0i24.9519

Chmiel, A., & Mazur, I. (2022). A homogenous or heterogeneous audience? Audio description preferences of persons with congenital blindness, non-congenital blindness, and low vision. *Perspectives: Studies in Translation Theory and Practice, 30*(3), 552–567. https://doi.org/10.1080/0907676X.2021.1913198

Di Giovanni, E. (2014). Audio description and textuality. *Parallèles, 26*, 69–83.

Fidyka, A., & Matamala, A. (2021). Retelling narrative in 360° videos: Implications for audio description. *Translation Studies, 14*(3), 298–312. https://doi.org/10.1080/14781700.2021.1888783

Gambier, Y. (2006). Multimodality and audiovisual translation. In M. Carrol, H. Gerzymisch-Arbogast, & S. Nauer (Eds.), *Audiovisual translation scenarios: Proceedings of the second MuTra Conference in Copenhagen* (pp. 1–5). Euroconferences.

Gottlieb, H. (2018). Semiotics and translation. In K. Malmkjær (Ed.), *The Routledge handbook of translation studies and linguistics* (pp. 45–63). Routledge.

Greco, G. M., & Romero-Fresco, P. (2023). Universalist, user-centred, and proactive approaches in media accessibility: The way forward. *Journal of Specialized Translation, 39*.

Gutt, E.-A. (2000). *Translation and relevance: Cognition and context*. St. Jerome Publishing.

Hirvonen, M., Hakola, M., & Klade, M. (2023). Co-translation, consultancy, and joint authorship: User-centred translation and editing in collaborative audio description. *Journal of Specialized Translation, 39*.

Holsanova, J. (2016). A cognitive approach to audio description. In A. Matamala & P. Orero (Eds.), *Researching audio description* (pp. 49–74). Palgrave Macmillan.

Jankowska, A. (2021). Audio describing films: A first look into the description process. *Journal of Specialized Translation, 36*.

Kaindl, K. (2013). Multimodality and translation. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 257–270). Routledge.

Kavid, V. (2022). *Priorities of characterization elements in audio description of movies for children and adults* (Unpublished master's thesis). Allameh Tabataba'i University.

Khoshsaligheh, M., & Shafiee, S. (2021). Audio description in Iran: The status quo. *Language and Translation Studies, 54*(2), 1–30. https://doi.org/10.22067/lts.v54i2.2101-1006(R2)

Kourdis, E. (2015). Semiotics of translation: An interdisciplinary approach to translation. In P. P. Trifonas (Ed.), *International handbook of semiotics* (pp. 303–320). Springer. https://doi.org/10.1007/978-94-017-9404-6_13

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

Matamala, A., & Remael, A. (2015). Audio description reloaded: An analysis of visual scenes in *2012* and *Hero*. *Translation Studies, 8*(1), 63–81.

Mazur, I. (2020). Audio description: Concepts, theories, and research approaches. In Ł. Bogucki & M. Decker (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility* (pp. 227–247). Palgrave Macmillan.

Poyatos, F. (2002). *Nonverbal communication across disciplines*. John Benjamins Publishing Company.

Poyatos, F. (2010). Gesture inventories: Fieldwork methodology and problems. In A. Kendon, J. Umiker-Sebeok, & T. A. Sebeok (Eds.), *Nonverbal communication, interaction, and gesture: Selections from SEMIOTICA* (pp. 371–401). De Gruyter.

Prince, G. (2012). *Narratology: The form and functioning of narrative*. De Gruyter.

Prunč, E. (2004). Zum Objektbereich der Translationswissenschaft. In I. Müller (Ed.), *Und sie bewegt sich doch ...: Translationswissenschaft in Ost und West*. Peter Lang.

Rai, S., Greening, J., & Petré, L. (2010). *A comparative study of audio description guidelines prevalent in different countries*. Royal National Institute of Blind People.

Reviers, N. (2018). Tracking multimodal cohesion in audio description: Examples from a Dutch audio description corpus. *Linguistica Antverpiensia, New Series–Themes in Translation Studies, 14*, 22–35.

Salehi, S. (2021). *Audio description of movies in English and in Persian: A comparative study* (Unpublished master's thesis). Allameh Tabataba'i University.

Salehi, S., & Mousavi Razavi, M. S. (2023). Indicators of scene changes in Persian and English audio descriptions: A comparative study. *International Journal of Linguistics and Translation Studies, 4*(4), 128–143.

Sheikholeslami, S., & Ehteshami, S. (2023). A study of characters' facial expressions of emotions: The case of Iranian audio-described films. *Language and Translation Studies, 56*(3), 205–232. https://doi.org/10.22067/lts.2023.84265.1217

Shokoohmand, F., & Khoshsaligheh, M. (2019). Audiovisual accessibility for the deaf and hard of hearing in Iran. *New Voices in Translation Studies, 21*, 62–92. https://doi.org/10.14456/nvts.2019.18

Snell-Hornby, M. (2006). *The turns of translation studies: New paradigms or shifting viewpoints?* John Benjamins Publishing.

Tercedor Sánchez, M. (2010). Translating web multimodalities: Towards inclusive web localization. *Tradumàtica, 8*, 1–8.

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. https://doi.org/10.6035/MonTI.2012.4.9