

Audio description between MT Translation and Recreation: An Interview Study for the Language Pair English-German

Merle Sauter¹

Ekaterina Lapshinova-Koltunski¹

Sylvia Jaki²

¹University of Hildesheim

²KU Leuven

¹sauter, lapshinovakoltun@uni-hildesheim.de

²sylvia.jaki@kuleuven.be

Abstract

This study examines the machine translation of audio descriptions (AD) as an alternative to producing new AD for audiovisual formats in a foreign language. To assess acceptance and comprehensibility among German users, a small-scale survey is conducted with blind and visually impaired participants, examining key AD strategies, such as character description and naming, facial expressions and gestures, and spatio-temporal settings. Participants are shown machine-translated English AD, and original human German AD in comparison, and are questioned on these aspects. Findings suggest that AD translation is feasible for the German audience as the vast majority of the machine-translated stimuli are rated as helpful and understandable by the test persons. However, further studies are needed on machine translation, production costs, as well as larger-scale user studies.

1 Introduction

The study of audio description (hereafter: AD) as intersemiotic translation is intriguing due to its distinctive characteristics: its multimodality, its accessibility requirements, and its heterogeneous target audience. As a product of interlingual translation, however, AD has been less present so far within audiovisual translation studies. Its status as an interlingual translation of an intersemiotic translation makes it an especially interesting and multi-faceted object of study.

As an accessibility service for blind and visually impaired audiences, AD is a complex multimodal text form that transforms visual into verbal information. In many countries, including Germany, public broadcasters are legally required to provide AD as part of their accessibility services. However, producing AD remains a time-consuming and cost-intensive process (López Vera, 2006, 5). With the growing international film and TV market, not only are audiovisual products translated and dubbed for the respective audience, but they also need to be described in the target language. A common practice is not to translate the existing AD of the original, but to recreate the target-language AD from scratch. On the one hand, there might be legal restrictions or limits to the availability of source language AD scripts. On the other, it seems reasonable to assume that translating existing AD scripts might be a means of saving time, personnel and money. Instead of exploring legal limitations or the possibilities and quality limits of the translation of existing AD, this paper aims to put the target group's perspective first. It aims to shed light on the question whether AD translations are perceived to be acceptable and comprehensible, bearing in mind that they would be translated from AD scripts that followed possibly different AD guidelines. Like the work by Jankowska (2015), who lets a target group compare a recreated and a translated AD version, this study assumes a similarly positive acceptance and outcome with a German target group. To put this hypothesis to the test, a German target group is shown two different AD versions. The main aim of the interview that follows is to investigate the acceptability and comprehensibility of AD machine-translated from an original English AD into German and synchronised with the German dubbed version.

2 Theoretical Background

2.1 Audio description

AD can be considered a complex and challenging text type to translate due to its intermodal nature and its heterogeneous target audience. AD makes visual media accessible to blind and partially sighted audiences by verbally describing relevant visual elements so that the content can be understood through sound. The resulting text combines the spoken description with the existing acoustic elements of the audiovisual product. As a text, AD is always dependent on the film and aims to create a listening experience that approximates the visual experience as closely as possible (Poethe, 2005).

The target audience for AD is highly diverse, particularly with regard to the degree and type of visual impairment. Different forms of vision loss are associated with different perceptual experiences and world knowledge and need to be taken into consideration in AD production (Remael and Vercauteren, 2015, 15). For example, people who lose their sight later in life often retain concepts of space and colour that differ from those of individuals who have been blind since birth (Jüngst, 2020, 178). In addition, the average age of AD users is relatively high, as common eye diseases such as age-related macular degeneration and glaucoma occur more frequently in older age as reported by DBSV (2025, 6ff.).

Additionally, the intermodal nature of AD, the transfer of visual information into a verbal description, also poses various challenges. First, the film as source text is an audiovisual, multimodal product in which both the verbal-auditory and the visual communication channels must be considered (Braun, 2008, 3). Second, the resulting auditory description must interact with other elements of the target product, such as dialogue, music, and sound effects. Third, the target text reflects the describer's interpretation of the source material, selecting the most relevant of the many visual elements that are presented simultaneously, and rendering them sequentially during the few pauses in dialogue (ibid.; Hirvonen and Tiittula, 2012, 383).

In this sense, AD can be understood as a form of intersemiotic translation, following Jakobson's concept of translation, which distinguishes between interlingual, intralingual, and intersemiotic translation (1959). An intersemiotic translation is generally understood as a transfer between differ-

ent sign systems: in the case of AD, from a visual sign system to a verbal one (Korycińska-Wegner, 2014; Jung and Nuñez, 2018; Jaki et al., 2024). It is therefore fair to state that AD has its place in translation studies. As a consequence, it seems reasonable to apply theoretical paradigms developed for other forms of translation (Vercauteren, 2014; Mazur, 2020).

One approach that can be extended to multimodal text and intersemiotic translation is the function-oriented approach by Reiß and Vermeer: Prioritizing communicative equivalence, AD as a target text should fulfil the same communicative function for its audience as the original AD does for its audience (Jung and Nuñez, 2018, 173).

When describing a dubbed film originating from another culture, the source context must also be culturally contextualised for the target audience. Some scholars suggest that, ideally, an AD script would first be translated by a professional translator and then revised for AD by a professional audio describer (Polat, 2013, 120). The concept of audio description translation (ADT) essentially combines these two roles.

In practice, however, the skill sets of translators and audio describers do not fully overlap (Jankowska, 2015, 49). The intersemiotic transfer of AD goes beyond purely linguistic or cultural definitions of translation (López Vera, 2006). While describers mediate between different sign systems, they often lack intercultural mediation competences, and the needs of blind or visually impaired audiences can differ from those of the target culture in translation (Jankowska et al., 2017). AD translation therefore demands expertise in audiovisual translation, AD-specific challenges, *and* AD quality requirements (Remael and Vercauteren, 2010, 5). The resulting text must be stylistically and culturally appropriate without appearing translated (Weißbach, 2012; Jankowska et al., 2017) while still preserving relevant cultural elements from the source that might otherwise be lost in a target-language AD written from scratch (Jankowska, 2024; López Vera, 2006). With ADT, countries with lesser AD experience would be able to benefit from the accessibility skills and knowledge of another country. They may have fewer trained audio describers but more professional translators, as they often import audiovisual content (Remael and Vercauteren, 2010, 157). ADT can therefore improve quality and expand the

availability of accessible media (Jankowska, 2024; Jankowska, 2015; Korycińska-Wegner, 2014).

In addition to cultural differences, linguistic factors also play a role. As in other forms of translation, these constraints can be addressed through standard translation strategies, and minor adjustments in narration speed may help fit the script into the available pauses (Jankowska, 2015; Herrador Molina, 2006).

2.2 AD guidelines

To simplify the handling of this complex text type and semiotic transfer, most countries have their own set of guidelines. The UK was among the first to introduce official AD guidelines for broadcasting (the ITC Guidance) in 2000, later updated by Ofcom, most recently in 2024. In Germany, initial guidelines were created by Benecke and Dosch in 1994 and revised in 2004 (2004). Since then, major broadcasters and organizations in German-speaking countries have agreed on shared principles for producing audio description (NDR, 2019). The UK Ofcom and German NDR guidelines share core principles, such as prioritizing dialogue over AD or providing basic information on where, who and when, but differ in their key areas. A major difference concerns character names, linked to character description. The Ofcom guidelines place stronger emphasis on inclusive descriptions of diversity, while the NDR guidelines more readily mention features like hair and skin colour. Both offer general advice on locations, focusing on plot relevance. Overall, the NDR guidelines are more detailed on practical issues, whereas Ofcom remains more general. Beyond guideline differences some linguistic contrasts between English and German AD are notable. British AD is often described as more detailed, using semantically rich verbs and compound adjectives (Remael and Vercauteren, 2010, 4), while German AD favours shorter sentences, often with more descriptive adjectives, especially for characters. These differences stem from language structure rather than guidelines (Poethe, 2005, 44).

In the interviews, similar to Weißbach's (2012) comparative analysis of English and German AD, we wanted to focus our investigation on three specific AD aspects crucial for understanding visually narrated stories: character description, gestures and facial expressions, and spatio-temporal settings. This section will briefly discuss the approaches taken in the existing guidelines and in

AD theory.

Character description. Characters' physical and psychological traits, behaviour, and actions are central to narratives (Remael and Vercauteren, 2015; Mazur, 2015). As mentioned above, one of the most prominent differences between the German NDR guidelines and the British Ofcom guidelines concerns the timing of name introduction. Ofcom recommends describing a character's key traits as early as possible and introducing the character's name unless it must remain unknown for plot-related reasons (Ofcom, 2011). In contrast, the NDR guidelines emphasise early description of protagonists, including appearance, age, clothing, body language, and facial expressions. Benecke recommends introducing a character's name only when it is mentioned in the film dialogue (Benecke, 2014, 20), arguing that naming characters immediately upon their first appearance might reveal information not yet available to sighted audiences and therefore not provide an equivalent viewing experience. Until the name appears in the dialogue, the character should be identified by a temporary descriptive label, such as "the blond man". Regarding visual appearance, both guidelines emphasise describing key attributes such as age, clothing, body language, and facial features early in the film (Fresno, 2016). Descriptions should remain concise to avoid overwhelming the audience and to improve memorability (Fresno et al., 2016, 72).

Gestures and facial expressions. Non-verbal cues convey emotions, reactions, and interpersonal dynamics, contributing to character development and narrative coherence (Mazur, 2014; Igareda, 2011; Hirvonen and Tiittula, 2012). They must be verbalised in a limited time and presented sequentially, even when occurring simultaneously (Hirvonen, 2013). While Ofcom recommends including body language and facial expressions, NDR advises caution to avoid misinterpretation. Scholars remain divided whether AD should remain strictly objective (example (1-a)) or allow some interpretation (example (1-b)) to facilitate comprehension (Chmiel and Mazur, 2012).

- (1) a. *She raises her eyebrows.*
b. *She looks surprised.*

Although AD generally aims to remain neutral and avoid subjective interpretation, purely objec-

tive descriptions may not always enable audiences, especially those blind from birth, to infer the intended emotion. Consequently, many suggest combining both strategies when necessary, especially in view of time constraints and cultural differences in gesture interpretation (Mazur, 2014; Igareda, 2011).

Spatio-temporal settings. While guidelines recommend indicating where and when the action takes place and introducing unfamiliar locations, they offer limited practical instruction on how this should be done (NDR). Temporal cues in AD include both broader story periods and local indicators such as season, time of day, or aging of characters, often conveyed visually or auditorily (Vercauteren, 2012). AD must also indicate filmic devices such as slow motion, flashbacks, or non-chronological sequences. Spatial cues mark scene transitions and highlight relevant features, supporting narrative structure and atmosphere (Benecke, 2014; Remael and Vercauteren, 2015; Polat, 2013). In AD, such transitions are frequently indicated through brief descriptions or elliptical phrases (Weißbach, 2012, 371). Descriptions focus on relevant features, size, furnishings, landscapes, or atmosphere, and reflect the perspective created by cinematography and staging. AD must translate multidimensional visual space into linear language, balancing conciseness, selectivity, and clarity while enabling audiences to construct coherent mental images (Seiffert, 2005; Hirvonen and Tiittula, 2012; Vercauteren, 2014).

3 Related work

Even among sighted viewers, viewing experiences are never exactly the same, which is why AD production was traditionally carried out by a team of three, two sighted and one blind describer (Benecke, 2014, 14). However, developments in the rapidly evolving media landscape have made the thorough and analytical work of such a team increasingly difficult to sustain. As a result, the three-author model has largely disappeared (Schruhl, 2020, 777). In the UK, AD production by a single describer had already become common practice by 2014 (*ibid.*) These industry changes, combined with financial constraints and increasing time pressure in production and dubbing, highlight the need for faster and more cost-effective approaches to AD production (Kurch, 2019; Schruhl, 2020).

Still, with a few exceptions, most broadcasters produce AD in-house (Korycińska-Wegner, 2014, 66), and so do most Streaming services. Romero-Muñoz observes the same phenomenon, with the exception of Apple TV, where the English AD of synchronised products is also translated (Romero-Muñoz, 2024, 248). This contrasts sharply with common subtitling practices, where translation based on templates is widely used and leads to significant savings in time and cost, as the laborious processes of segmentation, timing, and subtitle positioning have already been completed (Jankowska, 2024, 157f). Given this advantage, it seems reasonable to consider whether a similar strategy can be applied to AD. Jankowska (2024) suggests that translating AD, even without the use of templates, could be a viable approach, provided that the resulting translation achieves a quality comparable to that of a newly created AD.

Empirical evidence on the time efficiency of ADT remains limited, which has contributed to ongoing skepticism regarding its productivity (Remael and Vercauteren, 2010). Early reflections by Hyks (2005) question the viability of ADT, suggesting that translating and culturally adapting an existing script might be more demanding than creating one from scratch; however, this claim has not empirically been substantiated (Jankowska, 2015).

Initial attempts to test this assumption produce mixed results. López Vera (2006) compare the time required for translating versus newly creating AD scripts and found only marginal differences, tentatively suggesting potential economic benefits of ADT. Herrador Molina (2006), incorporating audience feedback for the first time, demonstrates that while linguistic differences pose challenges, these can generally be addressed through translation strategies, and reception among visually impaired audiences was largely positive. Contrastive studies (Bourne and Jiménez Hurtado, 2007; Remael and Vercauteren, 2010) further highlight language-pair-specific difficulties as well as cultural differences and divergent national AD guidelines, pointing to the need for adaptation rather than direct transfer.

More robust empirical data are provided by Jankowska (2015), who find that newly created AD scripts require significantly more time than translated ones, up to three times longer in student samples and approximately twice as long com-

pared to professionals' estimates. Audience testing indicate generally positive reception of translated AD, with a majority preference for the translated versions.

Recent research has also explored machine translation (MT) and neural machine translation (NMT) in AD production. While MT combined with post-editing may reduce production time (Fernández-Torné and Matamala, 2016), quality assessments remain inconclusive. Lüthi (2024) examines AI-assisted translation of AD scripts and identify syntactic structures as a major source of error. This emphasises the need for further reception studies.

Overall, the literature suggests that ADT holds potential for increased efficiency, yet questions of quality, cultural adaptation, and audience acceptance remain open for systematic investigation.

In general, previous studies examine the time factor of AD translation (ADT) compared to writing new AD (ADN), conduct translation-oriented comparisons, and in some cases survey audiences, generally reporting positive feedback. While some studies investigate MT in ADT, and others linguistic differences between AD versions, only a few address audience acceptance. Since, at least to our knowledge, no study has yet examined ADT for the English–German language pair or consulted a German blind and partially sighted audience on their perspective and expectations with respect to ADT, this paper aims to fill that gap.

4 Data and Methodology

4.1 Research Design

Based on Jankowska's (2015) findings, it can be assumed that audiences are generally receptive to ADT. To test a comparable hypothesis, the present study adopts a similar research design with several modifications. Participants are shown a set of short clips from the same series, with each clip presented twice: once with a translated audio description (ADT) and once with a newly created German audio description (ADN). The ADT consists of a German translation of the English BBC AD produced in accordance with Ofcom guidelines, whereas the ADN is the German AD created by ZDF following German public broadcasting guidelines. After viewing both versions, participants are asked about comprehensibility, acceptance, and personal preference.

In addition, the study focuses on specific

challenges and strategies in AD. Following Weißbach's (Weißbach, 2012) comparative analysis of English and German AD, three aspects are examined: (A) character introduction and naming, (B) facial expressions and gestures, and (C) spatio-temporal settings. These aspects are selected because they represent key narrative elements and allow for different descriptive approaches.

4.2 Data

The object of analysis is the British mystery thriller series *A Good Girl's Guide to Murder* (2024). In the series, the main protagonist Pippa (or Pip), a girl in her last school year, decides to investigate an unsolved killing of a local school girl. The series was produced for BBC Three and the German broadcaster ZDFneo and was selected because the AD guidelines of public broadcasters are publicly available and therefore comparable, unlike those of most streaming services. The German version is a newly written description based on NDR guidelines and independent of the English AD created according to Ofcom guidelines. As the series was released only in 2024 (UK: July; Germany: November), it is likely to be unfamiliar to the participants.

Five clips of 55 to 90 seconds are selected from several episodes because they illustrate one or more of the three analytical aspects of this study. The clips also display clearly distinguishable AD strategies, such as different levels of detail in character descriptions or varying terminology for indicating flashbacks, which allows the study to more effectively assess whether the English AD would function for a German audience if translated. If both AD versions are stylistically very similar, a translation would likely go unnoticed and easily be accepted. As no official AD scripts are available, the English and German descriptions are manually transcribed and compared before the final selection, revealing both similarities and differences. The selected clips are also checked against the relevant guidelines (Ofcom and NDR) to ensure they do not contradict them. However, because these guidelines leave room for interpretation, individual production choices remain likely.

The translated AD version (ADT) used in the study is created with the audio description tool *Frazier*, a software by the German company VIDEO TO VOICE. It serves to synchronise the AD-segments with the gaps in the dialogue during the production process and provides AI-generated

voices to read out the text. It also includes a MT feature for 22 languages based on DeepL. To avoid subjective translation decisions and produce the ADT efficiently, we use this MT for a first translation. The previously transcribed English AD is thus machine-translated segment by segment within Frazier and then lightly post-edited to correct major errors. This approach is chosen to prevent the evaluation from being distorted by obvious translation mistakes that arise due to the segmented translation, while still reflecting a realistic professional workflow. Overall, the MT output provides a solid translation. Out of 564 translated words, 21 are adjusted. Most of the issues addressed during post-editing are related to contextual inconsistencies between segments, such as the incoherent use of pronouns, unidiomatic terms or mistakes caused by lack of context in individual segments. For example, the MT of *piece sign, Friedenszeichen*, is linguistically correct, but is changed to the more contextually adequate translation *Peace-Zeichen* during Post-Editing, and *Sie blättert durch Fotos in ihrem Netz* is edited to *Sie scrollt durch die Fotos in ihrem Profil* for the translation of *She scrolls through photos on her grid*, as *scrollen* is a better collocate for online environments than *blättern*.

4.3 Interview

We interview seven participants, a sample size that allows for in-depth qualitative analysis while still capturing a range of perspectives. The participants have been contacted via regional associations for the blind and have expressed interest in partaking in a study about AD. To cover a wider age range, special effort is taken to source participants under 40, so that the participants in the final set are between 30 and 74 years old; two are female and five male. All are native speakers of German and report previous experience with AD.

The demographic section of the interview reveals the following distribution of characteristics among the participants: Two participants (P5, P6) are congenitally blind, while the others are early blind (P1, P3) or late blind (P2, P4, P7). All participants report such severe visual impairment that they can at most distinguish light from dark on a screen and therefore rely on AD. All participants regularly watch public-service television; most use AD several times per week. All participants report listening exclusively to German (both the film audio and the AD). However, they note that this also

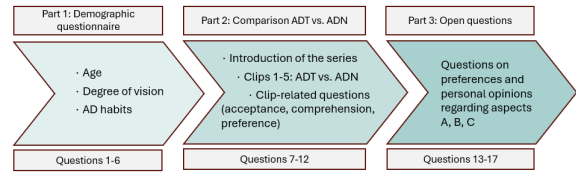


Figure 1: Interview guideline

includes German-dubbed productions originally in English. None of the participants is familiar with the series *A Good Girl's Guide to Murder*.

The interview guide (see translation of the originally German guide in the appendix), developed with reference to Jankowska's (2015) questionnaire, consists of three parts: (1) a short demographic questionnaire with six standardised questions, (2) the presentation of five clips, each shown in two AD versions, followed by related questions, and (3) four to five open questions addressing participants' general preferences regarding AD (see Fig. 1). Participants have not been informed about the research question and aim of the study nor the source of the presented AD, as this could compromise the internal validity of the study.

The first part of the interview explores the participants' age, the degree of their visual impairment using predefined categories (congenitally blind, later blind, visually impaired or other). It adds a series of questions addressing their familiarity with AD and their viewing habits. The second part of the interview briefly introduces the series, after which participants are shown five short scenes, each presented in two AD versions, one after another. Rather than randomising the order of presentation, the translated BBC version (ADT) is consistently shown first. This decision is made to collect an unbiased, initial evaluation regarding the clarity and comprehensibility of the translated AD, without any prior exposure to the alternative version (ADN).

To assess the general acceptance of ADT, it is followed by a brief (perceived) comprehension question, as illustrated in example (2-a). Thereafter, the second version (the ADN) is shown and these three questions are asked each time, respectively, as illustrated in example (2-b-d). The question in (1-b) aims to determine whether participants perceived any differences at all. While an affirmative answer is expected, the question mainly serves to prompt comparison and reflection. Questions in (1-c) and (1-d) address (perceive) com-

prehension and personal preference when comparing the two versions. The closed structure encourages participants to take a clear position (Albert and Marx, 2010, 10), while the follow-up *Why?* prompts them to recall specific details and explain their reasoning.

- (2)
 - a. *Did you find the description in this clip helpful? Did you feel you understood the scene well?*
 - b. *Did you notice any differences between the two versions?*
→ *If yes: Which ones?*
 - c. *Which version did you understand better?*
→ *Why?*
 - d. *Which version did you like better?*
→ *Why?*

We acknowledge that the order in which the versions are presented can influence subjective evaluation. Due to the recency effect, participants often remember the second version, the newly created AD, more clearly (Brosius et al., 2012). However, the comprehension question after the first version is intended to prompt a short pause and encourage participants to reflect on and recall the content before moving on.

For each scene a few more questions are added that prompt participants to compare ADT and ADN to address character description, description of facial expression and gestures, or spatio-temporal settings, such as *In which version could you imagine the described characters more clearly?*, *In which version were the characters' emotions and unspoken thoughts clearer to you?*, or *In which version did you better understand where and when the scene took place?* In the third part of the interview, open and more general questions are asked to further explore individual perspectives on AD aspects.

4.4 Interview data analysis

The qualitative data analysis is prepared, structured and carried out with the software MAXQDA, following Kuckartz and Rädike's (2024) flexible and explorative six-step method of interview analysis. The demographic information of the first part of the interview is handled as document variables to allow for comparisons. Categories are developed following the interview structure, e.g. grouping questions into code sets that explore character

description. They are further refined to subcategories that represent more detailed aspects, such as naming or appearance. After coding the interviews according to the categories, the categories are analysed one after the other with the question in mind: Which characteristics (of aspect A, B or C) have ensured comprehensibility and preference in this case, and what conclusions can be drawn from this for the general acceptance of ADT?

5 Results

5.1 Acceptability

After each ADT version, participants are asked whether they find the AD helpful and whether they understand the scene. Thus, the ADT's general acceptance and usefulness are evaluated before the comparison of the ADT and the ADN (see Table 1). In almost all cases the answer is positive. Only in Clip 3, where many characters and names are introduced and an Instagram page is described, do three of the seven participants report that the AD is *very confusing* (P1, P2) or at least *difficult* (P7) and *not easy* (P2). In the other four clips, the ADT is perceived as helpful by all participants.

responses	abs	in %
Accepted	32	91%
Not accepted	3	9%
TOTAL	35	100%

Table 1: Acceptance of ADT: 5 Clips, 7 participants.

responses	abs	in %
ADT	14	40%
ADN	21	60%
TOTAL	35	100%

Table 2: Preference for ADT vs. ADN: 5 Clips, 7 participants.

When compared with the ADN (see Table 2), participants often develop a preference. Questions about perceived differences encouraged direct comparison and help them judge which version they understand better.

Comprehensibility (illustrated in Table 3) varies depending on the clip. In some cases, participants notice the differences but find both versions comprehensible; in others, one version is clearly easier to follow. Out of 35 responses regarding comprehensibility (seven participants commenting on five clips), 49% (17 responses) rate the ADN as more

responses	abs	in %
ADT more comprehensible	7	20%
ADN more comprehensible	17	49%
Both equally comprehensible	11	31%
TOTAL	35	100%

Table 3: Comprehensibility of ADT: 5 Clips, 7 participants.

comprehensible. In 31% of cases (11 responses), both versions are considered equally comprehensible, while 20% (7 responses) favour the ADT. Asked about their preference, however, in 60% of cases participants choose the ZDF version (ADN), whereas in 40% the ADT is preferred. However, the participants are most likely used to the ZDF guidelines and thus develop a tendency to prefer what sounds familiar (Romero-Fresco, 2020, 18).

The explanations given for comprehension and preference often overlap. General reasons include a clear structure of the AD text (e.g., pauses that allowed listeners to process what they had heard), as well as the simplicity and length of the descriptions. More specific arguments relate to character description, facial expressions and gestures, and/or spatio-temporal settings. Alongside general elements such as structure, simplicity, and precision of the AD text, the implementation of these three aspects play a key role in comprehension and acceptance.

5.2 Specific features

Character description prove important for the participants' mental representation of characters. In Clip 1, the ADN briefly describes Pip's outfit ("brown shorts and a blue-green sweater vest"), while this information is absent from the ADT. Several participants immediately notice this difference and associate the description with improved comprehension: Most state that the ADN give them a clearer image of the character. Opinions differ on the level of detail required. Some participants value selective clothing details because they often allow a deeper understanding of the character's personality, whereas others considered excessive information (e.g., shoe colour) distracting and prefer the ADT for its brevity and action-oriented precision. One participant associates the mention of shorts with a summer setting. In Clip 2, both AD versions describe characters in a photograph but differ in how they convey ethnic and physical features. The ADT provides more detailed descriptions, as in example (3), whereas the ADN is more

concise.

- (3) ADN: *Naomi hat olivfarbene Haut, dunkle Augenbrauen und rot gefärbtes Haar. (Ofcom original: Naomi has olive skin, dark eyebrows and dyed red hair.)*

Four of seven participants prefer the ADT for its more detailed character descriptions. Some also admit the ADT presents ethnic differences more naturally. However, the relevance of colour descriptions varies: Congenitally blind participants note that colours themselves convey little meaning, although they might still be useful for people who lost their sight later in life.

The importance of names becomes particularly evident in Clip 3, where multiple characters appear in photographs. The ADT immediately lists several names (*Jake Lawrence, Max Hastings*), which some participants find confusing. Five of seven participants therefore prefer the ADN, where characters are first described (e.g., *the blond boy*) and then explicitly linked to their names, making identification easier. However, one congenitally blind participant prefers the ADT precisely because it relies more on names and less on visual attributes such as hair colour. When asked about preferences regarding the timing of name references, opinions are mixed. Some participants prefer names to be introduced early in the AD and to be repeated consistently to maintain orientation, while others favour introducing names only once they appear in the dialogue, arguing that this better reflects the structure of the film without giving away more than the sighted viewers know. Regarding aspect (A), overall, the ADT is positively received in several instances, particularly where it provides natural descriptions of ethnicity and detailed character information. However, the ADN is preferred for its clearer linking of names and attributes and for inserting concise appearance descriptions into the narrative.

Facial expressions and gestures. We analyse which forms of describing non-verbal behaviour contribute to the acceptance and comprehension of the AD and how the ADT differ from the ADN. In the flashback scene in Clip 4, emotional tension is conveyed largely through non-verbal cues rather than dialogue. While both AD versions describe the scene, the ADN provides more detailed descriptions of facial expressions and gestures, as in example (4-a), whereas the ADT remains more

concise as illustrated in example (4-b).

- (4) a. ADN: *In der Schule fährt sich Sal angespannt mit der Zunge über die Lippen. Unsicher mustert Pip ihn. (back translation EN: Sal licks his lips nervously. Pip looks at him uncertainly.)*
- b. ADT: *In der Rückblende spricht Sal Pip ruhig an. (Ofcom original: In the flashback, Sal addresses Pip calmly.)*

All seven participants report that the ADN conveys the characters' emotions and the scene's atmosphere more clearly, even though one person prefers the ADT for its more precise and factual description. Several explain that the additional details make the tension and urgency more apparent. A similar pattern emerges in Clip 5, where Pip confronts her mother. Participants note that the ADN's more detailed descriptions of gestures and facial expressions (illustrated in example (5)) make the characters' emotions, the dynamic of the scene and the strained relationship easier to understand.

- (5) a. ADT: *Pip schaut auf ihr Frühstück hinunter. (Ofcom original: Pip looks down at her breakfast.)*
- b. ADN: *Nachdenklich senkt Pip den Blick auf ihren Teller, hebt ihr Brot an und lässt es wieder fallen. (back translation EN: Pip looks down at her plate thoughtfully, picks up her piece of bread and then drops it again.)*

These non-verbal details are even interpreted as adding narrative meaning, for example, the dropped toast is seen as an expression of frustration or helplessness, while the mother's avoidance suggests she might be hiding something.

However, preferences differ regarding the degree of interpretative description. Some participants prefer the ADT's more concise approach over the more detailed ADN, arguing that tone of voice and dialogue already convey much of the emotional information. Clip 4 evoke the question of how interpretative AD should be: The ADT describes a gesture interpretatively (*she signals her to be quiet*), whereas the ADN describes the physical movement more objectively (*she puts a finger to her lips*). Most participants prefer the more objective description, as it provides clearer information and allowed listeners to interpret the gesture

themselves. Others favour the interpretative solution for efficiency or clarity, particularly when gestures might not be universally understood.

Spatio-temporal settings. In Clip 1, where Pip walks through a forest, both AD versions describe the surrounding landscape, although the ADN provides more detail. While three participants prefer the ADT for its brevity and precision, all report that the ADN enables a clearer mental image of the setting because it includes more environmental details. However, while some value the richer description because it conveys atmosphere, others consider such details unnecessary if they are not relevant to the action. One participant also notes that the shorter ADT leaves more room for background music, allowing the scene's tense atmosphere to emerge more clearly. In Clip 2, with a scene that takes place inside, differences concern both environmental detail and spatial orientation. The ADT includes a more detailed description of decoration on the wall, as in example (6-a), while the ADN refers to it more briefly, see example (6-b).

- (6) a. ADT: *hängt ein dekoratives lila Laken an einer der Wände, auf dem ein großer rosa Kreis und kleinere bunte Formen zu sehen sind. (Ofcom original: a decorative purple sheet hanging on one of the walls, featuring a large pink circle and smaller colourful shapes.)*
- b. ADN: *An einer Wand hängt ein großes, gemustertes Tuch. (back translation EN: On one of the walls there's a large patterned cloth.)*

Four participants prefer the ADT in this case because the additional detail makes the room easier to imagine. Additionally, spatial orientation is highlighted as important. Some participants find the ADN clearer because it explicitly mentions elements such as the bed (example (7-a)), whereas others prefer the ADT's description of the character's position while moving through the room (example (7-b)).

- (7) a. ADN: *Pip umrundet ein Bett und beugt sich zu einer Pinnwand mit Fotos. (back translation EN: Pip walks around a bed and leans toward a pinboard.)*

- b. ADT: *Pip geht zu einer Kommode hinüber, auf der eine Pinnwand steht.* (Ofcom original: *Pip walks over to a dresser which has a pinboard sitting on top of it.*)

Several participants suggest that a combination of both approaches, clear spatial positioning and sufficient environmental detail, would be ideal. Clip 4 illustrates differences in description of temporal markers with the scene alternating between the present and a flashback. The ADT explicitly signals these shifts using markers such as *in the flashback* and *in the present*, whereas the ADN indicates the temporal change indirectly through contextual cues (e.g., referring to the character's age or environmental settings). While both versions are generally understood, three participants prefer the ADT because the explicit markers make the transition between time levels clearer. Others find contextual cues equally helpful.

6 Conclusion and Discussion

This study tests the acceptance and comprehension of an audio description machine-translated from English (ADT).

The interviews reveal a generally positive acceptance. For the overwhelming majority of cases, participants report that the ADT is helpful and understandable when considered on its own, supporting the hypothesis that ADT can be beneficial for a German blind and visually impaired audience, which is in line with earlier findings by Herrador Molina (2006) and Jankowska (2015). However, this does not mean that the ADT is more comprehensible or more often preferred than the ADN: Compared with the native German AD, participants find the ADN more comprehensible than ADT.

Preferences show a similar distribution, with the ADN preferred in the majority of cases. The ADN therefore yields slightly better results, although the overall acceptance of the ADT remains high. Despite these reservations, this study demonstrates that translated AD can be well accepted by German audiences. Concerns that language-pair-specific translation challenges (Bourne and Jiménez Hurtado, 2007) might hinder AD translation or acceptance are not confirmed.

For character description, the results indicate that even brief descriptions of appearance, such as clothing, can significantly improve the audience's

mental representation of characters. Participants generally agree that person descriptions enhance understanding, although opinions differed regarding the level of detail required. When it comes to the timing and structure of name references, the results partly support the concept of interim character identification proposed by Benecke for German AD guidelines (i.e., description first, naming second). Facial expressions and gestures prove crucial for understanding mood and emotional atmosphere.

The interviews also highlight the debate between interpretative and objective descriptions, without offering a one-fits-all-result. Most participants favour objective descriptions of gestures, although a combination of objective description and brief explanation is also suggested as a possible compromise.

For spatio-temporal settings, the results indicate that detailed environmental descriptions and detailed, clear spatial orientation can enhance mental imagery, although some preferred shorter AD to allow space for music and sound effects. Explicit temporal markers facilitate comprehension in scenes with shifts between time levels.

Overall, the interview confirms that preferences are individual, highlighting the heterogeneity of the target group. Some participants consistently prefer concise AD, while others value richer descriptions.

Differences can also be linked to the onset of visual impairment: Congenitally or early blind participants consider colour descriptions less meaningful, whereas late-blind participants appreciate more detailed visual information. Due to the small sample size, however, no clear correlation between type of blindness and preference for ADT or ADN can be established, but the differences between individual preferences of congenitally and late-blind persons regarding AD are worth investigating further. Despite this limitation, this study provides first insights into the preferences and needs of German AD users.

References

- Albert, R. and N. Marx. 2010. *Empirisches Arbeiten in Linguistik und Sprachforschung*. Narr Francke Attempto Verlag GmbH, Tübingen.
- Benecke, B. and E. Dosch. 2004. *Wenn aus Bildern Worte werden – Durch Audiodeskription zum Hörfilm*. Bayerischer Rundfunk, München.

- Benecke, B. 2014. *Audiodeskription als partielle Translation. Modell und Methode*. LIT, Berlin.
- Bourne, J. and C. Jiménez Hurtado. 2007. From the visual to the verbal in two languages: a contrastive analysis of the audio description of *The Hours* in english and spanish. In Cintas, J. D., P. Orero, and A. Remael, editors, *Media for All. Subtitling for the Deaf, Audio Description, and Sign Language*, pages 175–187. John Benjamins, Amsterdam/New York.
- Braun, S. 2008. Audio description research: State of the art and beyond. *Translation Studies in the New Millennium*, 6:14–30.
- Brosius, H.-B., A. Haas, and F. Koschel. 2012. *Methoden der empirischen Kommunikationsforschung. Eine Einführung*. Springer VS, Wiesbaden, 6., erweiterte und aktualisierte auflage edition.
- Chmiel, A. and I. Mazur. 2012. Ad reception research: Some methodological considerations. In Perego, E., editor, *Emerging topics in translation: Audio description*, pages 57–80. EUT Edizioni Università di Trieste, Trieste.
- Deutscher Blinden- und Sehbehindertenverband e.V. (DBSV). 2025. Sehen und sehverlust in deutschland. woche des sehens.
- Fernández-Torné, A. and A. Matamala. 2016. Machine translation in audio description? comparing creation, translation and post-editing efforts. *SKASE Journal of Translation and Interpretation*, 9(1):64–87.
- Fresno, N., J. Castellà, and O. Soler-Vilageliu. 2016. ‘what should i say?’ tentative criteria to prioritize information in the audio description of characters. In Matamala, A. and P. Orero, editors, *Researching Audio description*, pages 143–168. Springer Nature, London.
- Fresno, N. 2016. Carving characters in the mind. a theoretical approach to the reception of characters in audio described films. *Hermeneus*, 18:59–92.
- Herrador Molina, D. 2006. La traducción de guiones de audiodescripción del inglés al español: Una investigación empírica. Master’s thesis, Universidad de Granada.
- Hirvonen, M. and L. Tiittula. 2012. Verfahren der hörbarmachung von raum: Analyse einer hörfilmsequenz. In Hausendorf, H., L. Mondada, and R. Schmitt, editors, *Raum als interaktive Ressource*, pages 381–427. Narr Francke Attempto, Tübingen.
- Hirvonen, M. 2013. Perspektivierungsstrategien und -mittel kontrastiv. die verbalisierung der figurenperspektive in der deutschen und finnischen audiodeskription. *trans-kom*, 6(1):8–38.
- Hyks, V. 2005. Audio description and translation. two related but different skills. *Translating Today*, 4:6–8.
- Igareda, P. 2011. The audio description of emotions and gestures in spanish-spoken films. In Serban, A., A. Matamala, and J.-M. Lavour, editors, *Audiovisual Translation in Close-up: Practical and Theoretical Approaches*, pages 223–238. Peter Lang, Bern.
- Jaki, S., M. Bolz, and S. Röther. 2024. Ki-technologien in der audiovisuellen translation. *trans-kom*, 17(2):320–342.
- Jakobson, R. 1959. On linguistic aspects of translation. In Brower, A. R., editor, *On Translation*, pages 232–239. Harvard University Press, Cambridge, MA and London, England.
- Jankowska, A., M. Milc, and L. Fryer. 2017. Translating audio description scripts. . . into english. *SKASE Journal of Translation and Interpretation*, 10(2):2–16.
- Jankowska, A. 2015. *Translating Audio Description Script: Translation as a New Strategy of Creating Audio Description*. Peter Lang, Frankfurt am Main.
- Jankowska, A. 2024. ‘guidance but not the ultimate commandment’: Pros and cons of using pivot templates in the audio description production workflow. *The Journal of Specialised Translation*, 42(42):155–173.
- Jung, L. and A. J. C. Nuñez. 2018. Intersemiotisches Übersetzen: Anmerkungen zur audiodeskription aus einer übersetzungswissenschaftlichen perspektive. *Glottodidactica (Poznań)*, 45(2):169–183.
- Jüngst, H. E. 2020. *Audiovisuelles Übersetzen: Ein Lehr- und Arbeitsbuch*. Narr Francke Attempto, Berlin, 2., überarbeitete und erweiterte auflage edition.
- Korycińska-Wegner, A. 2014. Audiodeskriptionsübersetzung als alternative methode der filmbeschreibung. *Studia Germanica Posnaniensia*, pages 65–84.
- Kuckartz, U. and S. Rädike. 2024. *Fokussierte Interviewanalyse mit MAXQDA. Schritt für Schritt*. Springer VS, Wiesbaden, 2. auflage edition.
- Kurch, A. 2019. Produktionsprozesse der hörgeschädigten-untertitelung und audiodeskription: Potenziale teilautomatisierter prozessbeschleunigung mittels (sprach-)technologien. In Maaß, C. and I. Rink, editors, *Handbuch Barrierefreie Kommunikation*, pages 437–454. Frank & Timme, Berlin.
- López Vera, J. F. 2006. Translating audio description scripts: The way forward? tentative first stage project results. In Carroll, M., H. Gerzymisch-Arbogast, and S. Nauert, editors, *Audiovisual Translation Scenarios: Proceedings of the Marie Curie Euro-conferences MuTra*. Peter Lang, Berlin.

- Lüthi, N. 2024. Ki-gestützte Übersetzung von audiodeskription. eine quantitative und qualitative untersuchung der sendung Passe-moi les jumelles. Master's thesis, ZHAW Zürcher Hochschule für Angewandte Wissenschaften, Winterthur. Graduate Papers in Applied Linguistics 26.
- Mazur, I. 2014. *Gestures and facial expressions in audio description*. John Benjamins, Amsterdam/Philadelphia.
- Mazur, I. 2015. Characters and action. In Remael, A., N. Revers, and G. Vercauteren, editors, *Pictures painted in words: ADLAB Audio Description guidelines*, pages 19–25. Università di Trieste, Trieste.
- Mazur, I. 2020. A functional approach to audio description. *Journal of Audiovisual Translation*, 3(1):226–245.
- NDR. 2019. Vorgaben für audiodeskriptionen. Accessed 03 Oct 2025.
- Ofcom. 2011. Ofcom's guidelines on providing television and on-demand access services. Accessed 03 Oct 2025.
- Poethe, H. 2005. Audiodeskription– entstehung und wesen einer textsorte. In Fix, U., editor, *Hörfilm: Bildkompensation durch Sprache*, pages 33–48. Erich Schmidt Verlag, Berlin.
- Polat, N. T. 2013. *Raum im (Hör-)Film. Zur Wahrnehmung und Repräsentation von räumlichen Informationen in deutschen und türkischen Audiodeskriptionstexten*. Frank & Timme, Berlin.
- Remael, A. and G. Vercauteren. 2010. The translation of recorded audio description from english into dutch. *Perspectives*, 18(3):155–171.
- Remael, A. and G. Vercauteren. 2015. Pictures painted in words: Adlab audio description guidelines. Università di Trieste.
- Romero-Fresco, P. 2020. The dubbing effect: an eye-tracking study on how viewers make dubbing work. *The Journal of Specialised Translation*, 33:17–40.
- Romero-Muñoz, A. 2024. La traducción de audiodescripciones. situación en las plataformas de streaming y propuestas didácticas para la 11 y 12. In Martínez-Roig, R., C. Pulido-Montes, and B. Mateu Luján, editors, *Nuevas investigaciones educativas para definir la enseñanza y el aprendizaje*. Octaedro Universidad.
- Schruhl, S. 2020. Audiodeskription in der praxis – bericht einer blinden rezipientin und hörfilmautorin. In Maaß, C. and I. Rink, editors, *Handbuch Barrierefreie Kommunikation*, pages 771–780. Frank & Timme, Berlin.
- Seiffert, A. 2005. Räumliches hören. eine schemaorientierte analyse der audiodeskriptiven darstellung der handlungsräume. In Fix, U., editor, *Hörfilm. Bildkompensation durch Sprache*, pages 87–98. Erich Schmidt Verlag, Berlin.
- Vercauteren, G. 2012. A narratological approach to content selection in audio description. towards a strategy for the description of narratological time. In Agost, R., E. Di Giovanni, and P. Orero, editors, *Multidisciplinarity in Audiovisual Translation*, volume 4 of *Monografías de Traducción e Interpretación*, pages 207–231. Università di Trieste, Trieste.
- Vercauteren, G. 2014. A translational and narratological approach to audio describing narrative characters. *Target*, 27(2):71–90.
- Weißbach, M. 2012. Audiodeskription und hörfilme. eine kontrastive analyse der deutschen und englischen audiodeskription des films 'brokeback mountain'. In Panier, A., K. Brons, A. Wisniewski, and M. Weißbach, editors, *Filmübersetzung. Probleme bei Synchronisation, Untertitelung, Audiodeskription*, pages 343–409. Peter Lang, Frankfurt am Main.

7 Appendix: Interview Guide

Part I: Demographic Questionnaire

1. Age – How old are you?

2. Assessment of Vision – How would you describe your visual impairment?

- ☐ congenitally blind ☐ late blind ☐ visual impairment ☐ other

3. Viewing & AD habits

- 3.1. **How often do you watch film or TV with AD?** ☐ daily ☐ once a week ☐ multiple times a week ☐ once a month ☐ multiple times a month ☐ less often ☐ never

- 3.2. **What do you watch most?** ☐ TV shows ☐ Films on TV ☐ Series on TV ☐ Films on streaming services ☐ Series on streaming services ☐ Other

- 3.3. **What AD do you use?** ☐ The AD provided by TV or streaming services ☐ The AD provided by the GRETA-App ☐ Other

- 3.4. **In what language do you watch/listen to films?** ☐ German films with German AD ☐ Foreign films (dubbed in German) with German AD ☐ Foreign films (untranslated) with German AD ☐ Foreign films (untranslated) with original-language AD

Part II: Comparison of ADT and ADN

Introduction

The clips I'm about to show you are from the series *A Good Girl's Guide to Murder*. It's based on a young-adult crime novel by an English author. It's about Pippa (or Pip), a sixth-form student who reopens a five-year-old murder case for a school project. Back then, Andie was allegedly murdered by her boyfriend Sal. But Pip doesn't believe it was Sal, and sets out to investigate the matter together with his brother Ravi. The series is available in German with audio description on the ZDF Mediathek.

1. Do you know the series?

I'm going to play two versions of the same scene for you. Please listen carefully to both of them. Afterwards, I will ask you a few questions. Don't worry, I won't be testing your knowledge of the content. Nevertheless, try to take as much in as possible and capture your impressions.

Clip 1 – Pip is out walking her dog after having had an unsettling encounter

2. Clip 1

- a) (after Clip a, ADT) Did you find the description of this clip helpful, and do you think you understood the scene well?
- b) (after Clip b, ADN) Did you notice differences between both versions?
If so, which?
- c) Which version did you understand better?
Why?
- d) Which version did you like better?
Why?
- e) In which version did you get a better idea of the surroundings? Why?
- f) In which version did you get a better idea of the mood and the atmosphere of the scene?
- g) (In which version did you get a better idea of the main character?)

Clip 2 – Pip is at her friend Cara's house

3. Clip 2

- a) (after Clip a, ADT) Did you find the description of this clip helpful, and do you think you understood the scene well?

- b) (after Clip b, ADN) Did you notice differences between both versions?
If so, which?
- c) Which version did you understand better?
Why?
- d) Which version did you like better?
Why?
- e) In which version did you get a better understanding of the room and the photo board? Why?
- f) In which version did you get a better idea of the characters? Why?

Clip 3 – Pip collects all her research on her wall at home

4. Clip 3

- a) (after Clip a, ADT) Did you find the description of this clip helpful, and do you think you understood the scene well?
- b) (after Clip b, ADN) Did you notice differences between both versions?
If so, which?
- c) Which version did you understand better?
Why?
- d) Which version did you like better?
Why?
- e) In which version did you get a better understanding of the characters? Why?
- f) In which version did you find the description of the social-media profile more understandable? Why?

Clip 4 – Pip just had a fight with her best friend Cara

5. Clip 4

- a) (after Clip a, ADT) Did you find the description of this clip helpful, and do you think you understood the scene well?
- b) (after Clip b, ADN) Did you notice differences between both versions?
If so, which?
- c) Which version did you understand better?
Why?
- d) Which version did you like better?
Why?
- e) In which version did you get a better understanding of where the scene took place?
- f) In which version were the characters' feelings or unspoken thoughts clearer to you?
Why?

Clip 5 – Pip’s parents (Victor and Lianne) in a kitchen scene

6. Clip 5

- a) (after Clip a, ADT) Did you find the description of this clip helpful, and do you think you understood the scene well?
- b) (after Clip b, ADN) Did you notice differences between both versions?
If so, which?
- c) Which version did you understand better?
Why?
- d) Which version did you like better?
Why?
- e) In which version were the characters’ feelings or unspoken thoughts clearer to you?
Why?

Part III: Open Questions

Aspect A

- 1. What kind of information is specifically important to you when describing the looks of a person?
- 2. How do you feel about the name of a person being mentioned in the description before they are identified by name in the film?

Aspect B

- 3. Which do you prefer: a detailed, factual description of facial expressions and gestures (e.g. “Andie looks at Pip and puts one finger to her lips”) or a more interpretative description that translates the gesture to what is meant (e.g. “Andie signals her to be quiet”)?

Aspect C

- 4. Is it important to you that light and atmosphere are described as well, if time and space allow it?
- 5. Is there anything else on audio description that you would like to add?