ΜΕΤΑ

Creating Coherence in Audio Description

Sabine Braun

Article abstract

As an emerging form of intermodal translation, audio description (AD) raises many new questions for Translation Studies and related disciplines. This paper will investigate the question of how the coherence of a multimodal source text such as a film can be re-created in audio description. Coherence in film characteristically emerges from links within and across different modes of expression (e.g., links between visual images, image-sound links and image-dialogue links). Audio describing a film is therefore not simply a matter of substituting visual images with verbal descriptions. It involves 'translating' some of these links into other appropriate types of links. Against this backdrop, this paper aims to examine the means available for the re-creation of coherence in an audio described version of a film, and the problems arising. To this end, the paper will take a fresh look at coherence, outlining a model of coherence which embraces verbal and multimodal texts and which highlights the important role of both source text author (viz., audio describer as translator) and target text recipients in creating coherence. This model will then be applied to a case study focussing on the re-creation of various types of intramodal and intermodal relations in AD.

Creating Coherence in Audio Description

SABINE BRAUN
University of Surrey, Guildford, United Kingdom
s.braun@surrey.ac.uk

RÉSUMÉ

En tant que forme émergente de traduction intermodale, l'audiodescription pose nombre de nouvelles questions pour les études de traduction et les disciplines qui y sont liées. Le présent article examine si la cohérence d'un texte source multimodal tel qu'un film peut être recréée dans l'audiodescription. La cohérence filmique résulte généralement de liens dans et à travers différents modes d'expression (par exemple les liens entre les images visuelles, entre les images et le son, ou entre les images et le dialogue). L'audiodescription d'un film n'est donc pas le simple fait de remplacer des images visuelles par des descriptions verbales. Elle nécessite la « traduction » de certains de ces liens par d'autres types de liens. Dans ce contexte, le présent article examine les moyens disponibles pour la recréation de cohérence dans un film en audiodescription, et les problèmes qui se posent alors. À cet effet, l'article revisite le concept de cohérence et propose un modèle de cohérence qui englobe des textes verbaux et multimodaux et qui accentue le rôle important de l'auteur du texte source (à savoir l'audiodescripteur comme traducteur) et des destinataires du texte cible en créant de la cohérence. Ce modèle sera ensuite appliqué à une étude de cas qui analyse la recréation de divers types de relations intramodales et intermodales dans l'audiodescription.

ABSTRACT

As an emerging form of intermodal translation, audio description (AD) raises many new questions for Translation Studies and related disciplines. This paper will investigate the question of how the coherence of a multimodal source text such as a film can be re-created in audio description. Coherence in film characteristically emerges from links within and across different modes of expression (e.g., links between visual images, image-sound links and image-dialogue links). Audio describing a film is therefore not simply a matter of substituting visual images with verbal descriptions. It involves 'translating' some of these links into other appropriate types of links. Against this backdrop, this paper aims to examine the means available for the re-creation of coherence in an audio described version of a film, and the problems arising. To this end, the paper will take a fresh look at coherence, outlining a model of coherence which embraces verbal and multimodal texts and which highlights the important role of both source text author (viz., audio describer as translator) and target text recipients in creating coherence. This model will then be applied to a case study focussing on the re-creation of various types of intramodal and intermodal relations in AD.

MOTS-CLÉS/KEYWORDS

audiodescription, traduction audiovisuelle, traduction intermodale, cohérence, multimodalité
audio description, audiovisual translation, intermodal translation, coherence, multimodality

1. Introduction

Over the last decade, audio description (AD) has firmly established itself as a form of Audiovisual Translation, aiming to enable visually impaired people to understand and enjoy visual content. Different AD techniques have emerged depending on whether the source material is *monomodal* (e.g., a still image) or *multimodal*, i.e., performance-based material (theatre, opera and dance) or audiovisual material (films and TV programmes). In AD for film, on which this paper will focus, moving images and sounds which are ambiguous or incomprehensible without visual cues are translated into a verbal description. In relation to a film, AD is part of the post-production process. The description is 'inserted' into the finished audiovisual product (or sometimes even into a 'post-product' such as a dubbed or subtitled film). This has several implications:

At a *technical level*, it means the description *is added to* a film as a second verbal track. To avoid overlap with the primary verbal track (dialogue and/or a narration) and with essential auditory elements (sound effects, music), the description is normally delivered in chunks of no more than a few seconds, fitted into silent moments;

At the *level of meaning*, the implication is that the description *refers to* the film. AD texts are not intended to be stand-alone texts. They fall into Reiss's category of 'multi-medial texts.' Being "part of a larger whole" (Reiss 1981: 126), they are created and processed in conjunction with those filmic elements that remain accessible for visually impaired audiences, i.e., the dialogue, possibly a narrator, sounds and music (Braun 2007).[1]

At a *presentational level*, the verbal description *replaces* information which was originally conveyed visually. Compared to the audiovisual source, the number of modes involved in the audio described version is therefore reduced, the mix and weight of modes changes, and more information is conveyed verbally. As Yos (2005: 115) points out, this entails a more linear presentation of information because the simultaneous presentation of dialogue and visual images is transformed into alternating sequences of dialogue and AD.

Despite these differences between an audiovisual and audio described version of a film, an audio described version is a multimodal text just as its audiovisual source. In any multimodal text, different modes of expression intertwine to contribute to meaning jointly. Crucially, therefore, the recipients of any multimodal text need to make multiple intra- and intermodal links to create coherence, i.e., the general impression of a continuity of sense in a text (Beaugrande and Dressler 1981: 84). However, it can be assumed that the changes in the mix of modes and the way information is presented, especially the tendency towards linearization in AD, have implications for the creation of coherence.

Against this backdrop, this paper aims to examine how the creation of coherence in AD can be conceptualized. This requires an analysis of how coherence emerges in audiovisual source texts and how the audio describer can support its re-creation in the audio described version. In section 2, key approaches to coherence will be reviewed, drawing on insights from the two disciplines that have mainly dealt with it, linguistics and multimodality research. Section 3 will first suggest an extended model of coherence which can capture the common basis of, and the differences between, verbal and multimodal coherence and then highlight the implications for

creating coherence in AD. In section 4, a small-scale case study will be presented in which this model was applied to analyze different types of coherence in audio described film footage. The wider aim is to provide initial answers to the question of why some AD solutions work better than others and to raise questions for future research into AD.

2. Approaches to coherence

As stated above, coherence can be described as the impression of continuity of sense and connectivity in a text, and perhaps in our perception of the world at large. Ever since Halliday and Hasan's (1976) seminal work, coherence has often been analyzed from a semantic point of view, as a product of textual cohesion. Guided by Halliday's (1985) views on the text-organising function of language and Halliday and Hasan's (1976) model of text as a semantic unit that is 'bound together' by more than grammatical structure, work in this tradition has emphasized the role of lexico-grammatical cues on the text surface ('cohesive ties') in the recipient's recognition of the semantic relations underlying a text. Cohesive ties have been regarded as crucial in the development of a continuity of sense (see Tanskanen 2006 for a recent account). This approach has also been adopted in multimodality research, leading to a discussion of cross-modal links in multimodal texts in terms of 'intersemiotic cohesion' (e.g., O'Halloran 2004 for page-based multimodal texts; Baumgarten 2008 and Chaume 2004 for films; Ventola, Charles *et al.* 2004 for a range of text genres).

However, continuing linguistic research has demonstrated that coherence is in fact a much more complex concept (e.g., Blakemore 1992; Beaugrande and Dressler 1981; Brown and Yule 1983; Bublitz, Lenk *et al.* 1999; Gernsbacher and Givón 1995) and has moved away "from reducing coherence to a product of (formally represented) cohesion and/or semantically established connectivity" (Bublitz 1999: 1) to a view that:

- links between textual entities are not necessarily "made explicit in the text, that is, they are not activated directly by expressions of the surface" (Beaugrande and Dressler 1981: 4);
- text recipients "will supply as many relations as are needed to make sense out of the text as it stands" (Beaugrande and Dressler 1981: 4);
- "formal cohesion will not guarantee […] textual coherence" (Brown and Yule 1983: 197).

Hence, coherence has been conceptualised as a process of linking ideas, taking place in the recipient's mind. This represents a shift from coherence as a semantic concept to coherence as a pragmatic concept, i.e., "an interpretive notion, which is intrinsically indeterminate because it is relative to participants ascribing their understanding to what they hear" (Bublitz and Lenk 1999: 154). This approach embraces the contribution of 'cohesive ties,' while addressing the broader question of how text recipients are able to create connectivity when processing a text.

A pragmatic notion of coherence is more difficult to pin down than a semantic, cohesion-centred notion, and there is a danger of sliding into an 'anything goes' view of coherence (Edmondson 1999). However, Bublitz and Lenk (1999: 154) stress that "[t]hough not given in the text, i.e., not a text-inherent and invariant property at all, coherence nevertheless 'comes out' of the text." The key to reconciling these two attributes of coherence – 'text-basedness' but lack of 'text-inherentness' – can be said to lie

in the very nature of text comprehension, i.e., in how people go about constructing meaningful discourses from texts (Widdowson 2007). As Edmondson (1999) contends, models of coherence therefore need to be based on a sound model of discourse.

Such models are available for verbal discourse through work from discourse analytical, pragmatic, narratological and cognitive perspectives, presenting alternative, but not incompatible accounts of how coherence emerges in the process of discourse construction. Generally, they conceptualize discourse construction as a text recipient's formation of a mental representation of the text (a 'mental model,' 'text world,' 'story world'), a process in which the linguistic cues provided in the text are complemented with information from other sources to make the representation coherent (Brown and Yule 1983; van Dijk and Kintsch 1983; Herman 2002; Johnson-Laird 1983; 2006). While some models of discourse processing focus on the role of background knowledge (schemata or scenarios) – activated through cues in the text – as the recipient's major source for retrieving additional information (Sanford and Garrod 1981; Sanford and Moxey 1995; Shank and Abelson 1977), other models emphasize the role of inferencing processes for enriching textual information with necessary, plausible or possible additional information where required (e.g., Clark and Clark 1977; Blakemore 1992; Sperber and Wilson 1995). By contrast, such approaches are conspicuously absent from multimodality research, which has been dominated by work in a systemic functional tradition (e.g., Kress and van Leeuwen 2001).

In the following section, approaches from (linguistic) Discourse Analysis, Pragmatics and the Cognitive Sciences will be used to outline an extended model of coherence that can be used to conceptualize the creation of coherence in both verbal and multimodal texts and that can serve to analyse the processes of coherence creation in AD. The aim is not to provide a comprehensive account of verbal and multimodal meaning-making but to focus on those aspects which have a bearing on the modelling of coherence.

3. An extended model of coherence

Drawing on the discourse-based models of coherence introduced above, this section will first consider how coherence emerges in verbal texts (3.1) and multimodal texts (3.2), and then discuss the implications for the 'recreation' of coherence in AD (3.3). Some remarks on the limits of coherence (3.4) will conclude this part of the paper.

3.1. Coherence: spreading activation of knowledge and inferencing

The model outlined here assumes that the construction of a discourse and the creation of coherence on the basis of a *verbal* text can be assumed to rest upon two pillars. First, text recipients use initial textual cues to activate "background knowledge structures which contain defaults for the situation" the text is assumed to be about (Sanford and Moxey 1995: 169). The activated knowledge provides a frame of reference for linking and interpreting textual information. In addition, it can be assumed that an initial knowledge scenario is expanded by further textual cues ('spreading activation,' Beaugrande and Dressler 1981: 88) and that recipients operate within the activated scenario unless the text indicates otherwise ('assumption of normality,' Brown and Yule 1983: 62). Second, recipients also rely on inferencing

processes to derive assumptions about how textual information is linked, enabling them to deal with information that is not specified in the activated knowledge scenario. In contrast to other accounts of discourse construction (see section 2), the assumption made here is that both the activation of knowledge and the generation of inferences contribute to shaping a mental model of the text at hand.

The respective weight of knowledge activation and inferencing can be assumed to vary. Familiarity with the topic or genre of a text enables a recipient to activate a rich knowledge base 'at once' and to create a wide range of links between textual entities without much recourse to explicit textual cues or inferences. It may indeed enable him/her to create expectations about how the text proceeds (Brown and Yule 1983: 235). By contrast, a lack of knowledge requires 'close reading,' i.e., increases the dependence on textual cues and inferences to identify links between textual entities.

Regarding *multimodal* texts, it can be assumed that they are processed by the same principles as verbal texts. The difference is, of course, that recipients of multimodal texts use cues from different modes of expression to activate knowledge and/or to draw inferences, creating links within and across different modes of expression. This will be discussed in a separate section, before moving on to coherence and AD.

3.2. Coherence in multimodal texts

To illustrate some of the aspects involved in forming a coherent understanding of a multimodal text, this section will use the opening scene of the film *Girl with a Pearl Earring*[2], which will also be used in the case study in section 4. Based on a novel by Tracy Chevalier and set in the household of 17th-century painter Jan Vermeer van Delft, the film tells the story of Vermeer's (fictitious) maid Griet and how she comes to be the artist's model for the painting after which the film is named. In the opening scene, which shows Griet taking her leave from her parents to join the Vermeer household, the mother says, "their food may be strange to your stomach" and she urges Griet to "keep clear of their Catholic prayers." These remarks leave Griet visibly shocked.

In terms of *processing of textual cues*, the mother's verbal reference to the Vermeers' Catholic prayers along with Griet's visual reaction will enable some viewers to activate a more or less rich base of knowledge relating to the religious divide of the Netherlands in the 17th century (in the aftermath of Spanish rule and reformation). This provides a framework into which other visual and verbal signs can be mapped. Viewers without this knowledge may find it more difficult to link the mother's verbal utterance and Griet's visual reply, and their initial mental model of the situation will be poorer, but they will at least be able to infer that there are issues of religious denomination. Irrespective of prior knowledge, Griet's visible distress gives rise to the inference that her family is not Catholic, and the mother's remark about food is likely to generate the additional inference that the religious divide also entailed differences in dietary habits.

What cannot be inferred from this remark is why the mother refers to food at that moment. The utterance is in fact accompanied by the mother's passing Griet a small bundle, something wrapped up in a cloth, and it is a good example of some of the complexities of *intermodal linking of textual cues*. It would of course be possible to analyze the link in terms of intersemiotic cohesion (the bundle as a visual synonym

for the verbal reference to food) as an explanation for coherence. The model outlined here, however, assumes that a coherent understanding of this scene comes about through a set of interdependent assumptions on the part of the recipient. These include the 'global' assumptions about the two families' religious dominations and their dietary habits as well as 'local' assumptions about the visual action of passing the bundle and the verbal utterance. On the one hand, the recipient can assume that the content of the bundle has to do with food precisely because of what the mother says. In other words, the mother's utterance constrains the understanding of the visual action, although there is still room for interpretation. The bundle could contain food as well as a remedy for sickness. On the other hand, either of these assumptions enables the viewer (*viz.* audio describer) to interpret the mother's verbal utterance in a coherent way, as a rationale for the visually conveyed action.

An analysis of this scene merely in terms of intersemiotic cohesion, i.e., in terms of potential links an outside analyst may be able to identify, would miss much of the interaction between the visual and verbal cues, knowledge and assumptions involved in deriving a coherent interpretation of this scene.[3] Using the more dynamic model outlined above, section 3.3 will take a closer look at the implications for AD.

3.3. Coherence in audio description: intermodal and intramodal linking

In the context of translation, Baker (1992), Blum-Kulka (1986), Hatim and Mason (1990) and others have recognized that coherence is not dependent on the presence of formal cohesive devices but that a text author and also a translator can support a recipient's creation of coherence through appropriate choices of expression. However, they have also emphasized that the means available and required to support coherence tend to be "language-specific or text-specific" (Hatim and Mason 1990: 195). If we take AD texts to be a specific text type (Salway 2007; Bourne and Jiménez-Hurtado 2007), Hatim and Mason's observation is a good starting point for discussing coherence in AD.

As was pointed out in section 1, AD texts are specific because a) they are only part of what the target audience actually processes, replacing some elements of the audiovisual source text while leaving others unchanged, b) they refer to the unchanged elements (dialogue, sound) in various ways and c) they are delivered in short chunks. As a consequence, the translation process an audio describer engages in involves identifying and recreating a variety of intermodal and intramodal links.

With regard to *intermodal links* in the filmic source text, AD involves the translation of intermodal links between images, sound and dialogue into a) *inter*modal links between sound and AD, and b) *intra*modal verbal links between dialogue and AD, as required. The scene discussed in section 3.2, for example, is based on complex verbal-visual links. Failure to audio describe them would leave blind audiences with an incoherent piece of dialogue. At the same time, as the model of coherence outlined above highlights, and as is the case with any translation, there is a double interpretive 'filter' in this. Both the identification of actual links in the audiovisual source and the degree to which they are made explicit in the AD text are matters of a describer's judgement, based not only on his/her knowledge, inferences and preferences but also on his/her (meta-) awareness of different possible interpretations. Whilst a reference to the bundle would be crucial in the description of the above scene, the describer

needs to recognize the different possible interpretations of the bundle to avoid a description which has a bias towards one of them. Similar points could be made about the intermodal links between sounds. This will be considered in the case study (see section 4.1).

Apart from the intermodal links in the audiovisual source, the audio describer also has to identify *intramodal visual links* and recreate them as (explicit or implicit) intramodal verbal links. One of the difficulties with this arises from the natural connectivity inherent in the visual mode, i.e., the fact that items which appear together in a visual image are easily assumed to be connected. This stands in opposition to the sequential nature of the verbal mode, in which entities are introduced sequentially, with the result that links may have to be made more explicit. Another difficulty is that a film (as opposed to a still image) 'refers' to a set of items repeatedly, so that co-reference links have to be indicated in the verbal AD text. The 'chunked' delivery of AD texts further increases the need to do so. Thus the sequential nature of the verbal mode and the chunked nature of AD texts create specific requirements for intramodal linking within and across AD passages. They are, however, partially counteracted by the timing constraints for AD, which necessitate succinct descriptions and may inhibit the use of explicit links.

Yet another difficulty is that film normally makes use of a variety of editing techniques. The simplest of these are 'continuity editing' (Philipps 2000: 39) to save time by omitting what is obvious and 'shot/reverse-shot' editing (e.g., alternating shots of speakers in a dialogue; Philipps 2000: 42) to create a particular point of view and to solve the problem of presenting a three-dimensional 'reality' in a two-dimensional medium. Across the edits, the viewer tries to create a continuum of time, space/place and actions or events, and it is the illusion of this continuum that eventually needs to be carried across by an AD text to support the blind recipient in the creation of a coherent mental model.

The extent to which AD can draw on the wide range of devices normally available in verbal communication to support coherence has yet to be investigated. This needs to include classic cohesive ties and information structure as well as means of expression to support referential identification, knowledge activation, and much more. The case study in section 4 is intended to be an initial exploration of some of these means. Before this, however, some final points regarding the limits of coherence are in order.

3.4 Partial and disturbed coherence

The model outlined in this paper has emphasized the role of a text recipient in *recognising* textual referents and any links that might hold between them. Differences between the recipients in this respect help to account for the intersubjective differences in the perceived strength of the coherence of a text. However, there are of course cases in which many recipients would agree that coherence is difficult to recognize or achieve.

Acknowledging this observation, Bublitz and Lenk make a useful distinction between 'partial coherence' and 'disturbed coherence.' As they point out, "if we accept that it is not texts that have meaning, force or coherence, but rather speakers and hearers who ascribe meaning, force and coherence *to* a text, we may safely argue that

coherence is *always* only partial coherence" (Bublitz and Lenk 1999: 155; emphasis in the original). By contrast, "[f]or the hearer, the coherence of a text is disturbed when he is unable to make it coherent but assumes that it could be made coherent because he has no reason to believe otherwise" (Bublitz and Lenk 1999: 172).[4] It may not be possible to define sharp boundaries between these two notions, but this does not render them invalid. Coherence is relational, individual and a matter of degree.

Disturbance may be the result of a lack of knowledge on the part of the recipient. It may, however, also be the result of insufficient grounding by the author, i.e., insufficient or inappropriate reference to, and linking of, textual entities, preventing recipients from deriving inferences that could otherwise be derived irrespective of prior knowledge. The case study presented in section 4 analyzes further extracts from the audio described version of the film *Girl with a Pearl Earring* with regard to the verbal means of expression the audio describer used to support the target recipients in the creation of coherence. The model outlined above will be used as a framework for explaining why some of the chosen examples can be said to be cases of 'disturbed' coherence.

4. Coherence in the audio described version of *Girl with a Pearl Earring*

In the film scene chosen for this case study, the new maid Griet is introduced to her task of cleaning Vermeer's painting studio, a 'sacrosanct' place in the Vermeers' household, which the other family members rarely enter. The visual actions are slow-paced, and there is little talk. This highlights the 'sanctity' of the studio and gives the sound and the sound-image relation a prominent role in this scene. It also leaves ample time for description, which makes the typical timing constraints for AD less of a problem and helps to focus on the core issues of coherence. The present section considers the re-creation of intermodal coherence between sounds and images (4.1) and dialogue and images (4.2), before turning to the re-creation of intramodal visual coherence in the AD text (4.3). The assessment of the descriptions is based on a self-experiment of listening to the audio described version and on general feedback from blind people on the AD of this film.

4.1. *From sound-image coherence to coherence between sound and AD*

Sound is often taken for granted in AD and has received little attention in AD research (but see Remael 2007; van der Heijden 2007). The very beginning of the chosen scene, which has no dialogue, is a good example of what sounds contribute to filmic meaning, mainly by 'working together' with the images to create the natural audio-and-visual continuum that is so characteristic of film. Vermeer's wife Caterina, her daughter and Griet walk through the Vermeers' house, heading for the painter's studio. They pass a chirping parrot, and their steps can be heard as they climb a wooden staircase. Caterina carries a large bunch of keys, which is rattling with her every move.

Such sounds are perhaps the most basic type of film sound. They are real-life sounds, familiar to most people and in principle identifiable by visually impaired audiences. As Remael (2007) has stressed, however, even such sounds only become meaningful in the context of a particular film when they can be associated with a

source. Seeing – and thus knowing – that it is Caterina who carries the keys helps sighted viewers to make sense of the unfolding action (Caterina will show something to Griet), to derive inferences about what happens next (the women are heading for a secluded place) and to link this information with the painting studio later on (what does it mean that it is locked). A crucial function of AD is therefore to ensure that blind recipients get to know the sources of relevant sounds.

The AD script for this sequence is shown in example 1. All examples used in this paper are based on shots, which are numbered consecutively. A brief summary of what happens in the relevant shot(s) is given on the left-hand side. The boxes on the right-hand side provide the audio description and, where relevant, the sounds (in capitals) and dialogue (in italics). This is not to suggest that films are perceived on a shot-by-shot basis, nor that shots and AD should be aligned. (In example 1, for instance, the AD of what happens in shot 1 runs into shot 2.) It is merely intended to enable the reader to follow the story closely enough to assess the AD.[5]

EXAMPLE 1

Shot description	Audio description / [SOUND]
(1) Vermeer's pregnant wife Caterina enters the hallway of the house, carrying a large bunch of keys. She is followed by Griet and by her daughter Cornelia. Griet carries a bucket of water and a mop. As the women cross the hallway to reach the stairs, they pass a parrot on a perch.	(1) Griet follows with a bucket and a mop as Caterina walks through the house. [STEPS] Caterina, who is pregnant, carries keys. [KEYS RATTLE] Cornelia, the solemn daughter, follows. (2) They pass a parrot on a perch. [A BIRD CHIRPS]

Before focussing on how the sound-image coherence in this sequence is recreated in the description, it should perhaps be noted that some aspects of this description are problematic, especially the information structure (who follows whom) and some of the lexical choices (e.g., *keys* rather than *a bunch of keys*). While this also affects coherence, a detailed discussion of these problems would be beyond the scope of this paper.

As an overall pattern, the description supports the identification of sounds by referring to the sources emitting them, i.e., the parrot, the keys and the steps (via inference from *walk*), and – where applicable – by referring to the agents and actions causing them (*Caterina* and *the women*; *carry* and *walk, follow, pass*). Hence, there is no direct description of the sounds (such as *a bunch of keys rattles, a parrot on a perch chirps*). It is left to the audience to infer that the sounds they hear come from the sources and agents mentioned in the AD. What is also interesting is that the describer seems to have given priority to multifunctional descriptors. The chosen descriptors help the recipients to activate the relevant knowledge scenario (WALKING THROUGH A HOUSE),[6] to draw inferences about the unfolding action and to identify important sounds. The use of the verbs *walk, follow, pass*, for example, supports the activation of the knowledge scenario but also helps to identify the sound of the steps. Likewise, the information that Griet carries a bucket enables inferences about the purpose of the women's action while also explaining the intensity of the steps. Further support for identifying the sounds comes from the timing of the description. The relevant sounds can be heard in short pauses of the slow-paced AD.

The above-mentioned problems with information structure and lexical choices in this example notwithstanding, it seems likely that blind recipients can recreate the

intermodal links between sounds and images in the audiovisual source as intermodal links between sounds and AD in the audio described version. Arguably, the inferences arising from *being told* that Caterina carries keys while they can be heard rattling are similar to the inferences arising from *seeing* Caterina carrying a large bunch of keys and hearing them rattling. It is, however, likely that blind recipients, being aware that an audio description is by necessity a selective description of visual cues, create an additional inference to derive what Sperber and Wilson (1995) have termed an 'implicated premise' (i.e., "if Caterina carries keys and no other rattling objects are described, then the rattle must be from these keys"). In other words, the cognitive effort required by blind audiences to create coherence may be higher in such cases than that of sighted audiences.

The quasi-simultaneous presentation of AD and sound as in example 1 is not always possible or appropriate. In many cases, the only option is to insert a description before or after the sound to which it refers. Examples 2 and 3 suggest that successful linking of the AD and the sound is subject to sensitive constraints in these cases. In these two examples, the women have arrived at the studio, which is dark and only reveals some dim-lit objects as the camera pans around, including a wooden mannequin. Griet opens the shutters, creating a rattling sound. Later, she is alone in the studio and begins her cleaning task, exploring the objects in the studio and producing a variety of sounds.

EXAMPLE 2

Shot description	Audio description / *Dialogue* / [SOUND]
(10) Caterina and Cornelia stand in the doorway of the studio, looking inside the studio. (11) The camera pans through the dark room. The window shutters are closed. Griet starts pushing up the shutters towards the ceiling, using a long pole. (12) Back to Caterina and Cornelia in the doorway.	(10) *Caterina: Well, open the shutters, you can't work in the dark.* (11) [SHUTTERS RATTLE] (12) Griet pushes up the shutters towards the ceiling.

In example 2, the opening of the shutters is referred to in the dialogue before the corresponding sound is heard. This may raise the question whether the link between the sound and the visual action of opening the shutters needs to be recreated in the AD at all. Given that there is ample time, the description in (12) seems useful as it provides a complementary cue for identifying the sound. The actual reference to the sound follows the same pattern as above, i.e., reference is made to the agent and action producing the sound (*Griet pushes open*) and the source emitting it (*the shutters*). The description thus also contributes to continuity between dialogue and action, confirming that it is indeed Griet who opens the shutters. What links the description to the sound even more closely is that it is delivered immediately after the sound is heard.

By contrast, the sound of dusting at the mannequin's clothing in example 3 is described *before* it is heard, but there is a gap between the description and the sound in which even another sound can be heard (Griet's steps while walking around the mannequin). Because of this and in the absence of any other information which could support the interpretation of the sound, the sequence is likely to cause some disruption. At least, it may require more 'implicated premises,' serving as interim steps

towards an interpretation of the sound. Thus, the cognitive processing load increases, and if it reaches a stage where a recipient's overall processing capacity is exceeded, it may become impossible to make the link between the sound and the description.

EXAMPLE 3

Shot description	Audio description / [SOUND]
(17) In the studio: Griet wipes the frame of a painting with a cloth. (18) In the back of the room, Griet curiously looks at a wooden mannequin in a silk taft dress. She walks around it. Then she begins to brush its shoulder with a cloth.	(17) She wipes the frames of pictures (18) and dusts at the clothing of the wooden mannequin. [STEPS, THEN CLOTH BRUSHING OVER THE SILK TAFT DRESS]

These remarks on linking sound and AD may suffice to highlight the need for further, systematic investigation of the patterns of sound description in AD (how are sounds referred to, how are the descriptions timed) and for research into their effectiveness (in terms of achieving coherence) and efficiency (in terms of cognitive processing load required for achieving it). This also needs to take into account different types of film sound, as discussed by Remael (2007).

4.2. From dialogue-image coherence to coherence between dialogue and AD

The important role of sound in film notwithstanding, it is certainly the relation between the dialogue (or narration) and the visual mode that is central to film. As Baumgarten (2008: 10) puts it, the "functional combination of verbal and visual information […] is the defining characteristic of film texts." She adds that "visual and verbal information do *not* simply co-exist in a film text but that they are internally *related* to each other in specific ways" (Baumgarten 2008: 11; emphasis in the original). As the 'bundle' example (section 3.2) has shown, such links go far beyond the cases that could easily be pinned down as 'intermodal cohesion.' Equally importantly, formal cohesion is never sufficient for creating coherence (see section 3.1).

One explicit manifestation of the visual-verbal relation in film are the numerous references that film characters normally make to visible objects and to other characters. In example 4, Griet is distracted from her cleaning by a painting on an easel. While she is contemplating it, Vermeer's mother appears in the studio, addresses her and refers to the painting verbally.

EXAMPLE 4

Shot description	Audio description / *Dialogue*
(22) Close-up of a painting on an easel, showing a woman who is posing and dressed like the wooden mannequin.	(22) The mannequin's pose reflects a painting of a dark woman at a window.
(23) Close-up of Griet's face against the backdrop of a curtain which divides the studio.	(23) Griet gazes at it in awe.
(24) Close-up of the painting.	(24) A curtain is swept back by an older woman in a black dress, wide,
(25) Griet looks on, at the painting.	(25) white ruff and severely

(26) A severe-looking older woman appears from behind the curtain as she sweeps it back. She wears a black dress with a white ruff and a widow's cap.	(26) drawn back hair in a widow's cuff [sic].
(27) Griet turns and looks at the woman in shock.	(27) She looks
(28) The older woman, Maria, looks at Griet.	(28) down on Griet. *Maria: You're not the first to forget your manners in front of his paintings.*
(29) Griet, still in shock, curtsies.	(29) Griet curtsies. *Maria: Tell me, girl,*
(30) Close-up of Maria	*do you think it finished?*

Maria's reference to the painting using the pronoun *it* in (30) is interesting, because the successful identification of this reference relies specifically on a recipient's prolonged access to visual input. The painting appears first in (22). After this, the audiovisual source text keeps the painting salient through Griet's continuing gaze, allowing a sighted recipient to link Maria's question to the painting quite easily. In the audio described version, the verbal references to the painting in (22) and (23) may not achieve the same salience. The problem in the audio described version is therefore a rather substantial lapse of time between the introduction of the painting in (22) and Maria's utterance in (30). This is compounded by the semantic obscurity and syntactic ambiguity of the description in (22).[7] To what extent the generic reference to Vermeer's paintings in (28) is helpful and whether or not a blind recipient will indeed be able to link Maria's utterance in (30) to the painting can only be established empirically. This is why detailed reception studies are urgently required for AD. One point does, however, emerge from this discussion and it supports the observation made in example 3. Rather than just the formal presence of a potentially meaningful cue in the AD text, it is the specific circumstances of its delivery (timing) and presentation (clarity) that are crucial for the creation of coherence.

Another type of verbal-visual relation is the range of links between verbal utterances and visual signs produced by speakers or other characters, including visual signs produced in response to verbal utterances. For instance, Griet frequently curtsies to members of the Vermeer family in response to instructions she receives or as a general sign of respect. This is illustrated in (29) in example 4. The curtsies constitute the second part of what Conversation Analysis (Sacks, Schegloff *et al.* 1974) has termed 'adjacency pairs.' Because the first part of any such pair normally creates expectations about the second part, Griet's curtsies are highly expectable in a 17th century context. In example 4, this is reinforced through Maria's allusion to Griet's manners. The verbal 'translation' of the curtsies thus ensures continuity, indicating to a blind audience that Griet behaves as expected.

In contrast to this, the visual signs produced by a speaker, i.e., gestures or facial expressions accompanying speech, normally have to be filtered much more. Their description is often difficult to fit in (creating timing problems in the delivery) and can in fact interrupt the flow of the dialogue (creating a problem of presentation) and thus disrupt continuity.

What is different again are the visual signs indicating who is addressed and who speaks next. Such signs normally occur at the beginning or end of a speaker's turn and are crucial pointers missed by audiences without access to the visual mode. An

illustration of such signs is presented in example 5. In this sequence, the three women – Caterina, Griet and Cornelia – have arrived on the landing leading to the painting studio.[8]

EXAMPLE 5

Shot description	Audio description / *Dialogue*
(2) The three women stand on the landing, looking into a passageway. Caterina is in the foreground and turns towards Griet to talk to her. Caterina briefly looks at Griet and Cornelia, then focuses on Griet. Caterina turns her head towards the passageway. Griet steps forward into the passageway.	(2) Caterina looks into the room and hesitates. *Caterina: I… my husband does…* She looks at Griet and Cornelia. *Caterina: Go in.* *Go in.* Griet steps timidly into the passageway…

In the audiovisual source, Caterina is identified as the speaker through her body movements (she turns towards Griet), supported by the camera perspective (she is in the foreground). Apart from that, she can also be seen speaking. In the audio described version, the visual focus on Caterina is translated into a verbal description (*Caterina looks into the room and hesitates*) which offers several links to Caterina's subsequent utterance. To begin with, the adjacency of the description and the utterance suggests a connection. This is why when time is limited, "the describer may only be able to mention the name of a person" (Hyks 2005: 7) and still be able to support the creation of coherence. In example 5, there is sufficient time for a complete sentence, and the connection between the description and the utterance is reinforced through the structure of this sentence, in which Caterina is the subject. Moreover, the verb *hesitate* links the description to Caterina's way of speaking. After hearing this description and the utterance *I… my husband does…*, blind audiences are likely to understand *hesitates* as 'speaks with hesitation' and to assign the utterance to Caterina precisely because of its hesitant presentation. As the 'bundle' example (see section 3.2), this shows again how important it is to understand coherence as a reciprocal relation in which the interpretation of one cue depends on another.

In the description following Caterina's utterance (*She looks at Griet and Cornelia*), the proform *she* keeps the focus of attention on Caterina as a speaker, while the verb *looks at*, which is frequently used in AD "to provide information about a character's focus of attention" (Salway 2007: 160), indicates whom Caterina is going to address, contextualizing her instruction *go in*. The only problem is the lack of accuracy here. The description suggests that Caterina addresses Griet and Cornelia, whilst the context makes it clear that the imperative is directed at Griet alone. The problem is not that the audience would be unable to work this out. It is that the activated knowledge scenario – Griet, the new maid, being introduced to her task – creates a strong expectation of Griet being the only addressee. The mention of Cornelia, running counter to this expectation, therefore binds processing capacities unnecessarily.

It could, of course, be argued for most of example 5 that blind recipients would be able to discern who speaks, who is addressed and who is to speak next, using voice recognition and inferencing abilities, so that a description would not be necessary to

create coherence. However, a description of the visual signs relating to speakers, addressees and speaker changes reduces the cognitive load that would be required for voice recognition and predictive inferences.

The very last part of example 5, the description of Griet's stepping forward, is a slightly different case. It is once again a description of a reaction to a verbal utterance, similar to the description of the curtsy in example 4, but the description of Griet's stepping forward also provides information about the spatial environment, which is another important aspect of coherence. This will be discussed in section 4.3.

4.3. From intramodal visual coherence to intramodal coherence in the AD text

The previous sections have focussed on links between individual sounds, images and verbal descriptions. Ultimately, however, discourse connectivity emerges from a recipient's attempt to "build a coherent picture of the series of events being described and [to] fit the events together" (Brown and Yule 1983: 197). The verbal AD text needs to support this process against the odds of having to comply with more or less extensive timing constraints while having to provide a sequential account of events that often take place simultaneously in the audiovisual source. Furthermore, the AD text needs to be delivered in 'chunks' alternating with dialogue while trying to recreate the cinematic 'illusion of continuity' (see section 3.3).

Translating the wealth of simultaneously presented visual signs into a sequential yet succinct verbal account leaves the audio describer not only with often difficult decisions about what to describe. It also raises questions about whether and how a set of simultaneously presented visual elements should be linked in the verbal description and about the order in which the selected elements should be presented. In view of the timing constraints in AD, a desirable order is clearly one which minimizes the necessity for explicit linking, 'saving' words wherever possible, while maximizing support for creating coherence. From a broader point of view, however, the sequential order of events is only one dimension in the mental model a viewer tries to construct from the filmic surface. As was pointed out earlier, what a viewer ultimately tries to retrieve, and what the AD text needs to help a blind recipient to recreate, is a continuum of time, place and actions or events. One final example, repeating the very beginning of the chosen film scene, will be used to illustrate some of the difficulties for AD in this respect. The chosen sequence makes use of continuity editing (see section 3.3) between (1) and (2), and of a series of shots that show the spatial environment from different perspectives (from 2 onwards).

EXAMPLE 6

Shot description	Audio description
(1) Vermeer's pregnant wife Caterina enters the hallway of the house, carrying a large bunch of keys. She is followed by Griet and by her daughter Cornelia. Griet carries a bucket of water and a mop. As the women cross the hallway to reach the stairs, they pass a parrot on a perch.	(1) Griet follows with a bucket and a mop as Caterina walks through the house. Caterina, who is pregnant, carries keys. Cornelia, the solemn daughter, follows.

(2) The three women stand on the landing, looking into a passageway. Griet steps into the passageway.	(2) They pass a parrot on a perch. They arrive on the landing. Caterina looks into the room and hesitates. […] Griet steps timidly into the passageway which leads to the room.
(3) The passageway ahead. At its end, a door half open.	(3) She slowly opens the door into the darkness.
(4) Griet opens the door further and passes through the doorway.	

Shots such as (1) and (2) can be connected easily by a sighted viewer. The visibility of the women in both shots, the activated knowledge scenario WALKING THROUGH A HOUSE and perhaps also an overall 'assumption of normality' (i.e., everything is assumed to be as expected unless indicated otherwise; Brown and Yule 1983: 62) allow a sighted viewer to infer that the women must have climbed the stairs to get to the landing. The AD seems to build on this uncontroversial inference. Going beyond what is actually seen in (2), i.e., the women *standing* on the landing, the AD describes them as *arriving on the landing*. According to Vendler's (1967) classification, *arrive* is a 'telic' verb, indicating as it does the end point of an action, here of the open-ended action of *walking through the house*. The use of *arrive* therefore captures some of the women's movement and helps blind audiences to create a link to their prior action of crossing the hallway.

The other shots in this example visualize the space around the women and are equally easy to connect for a sighted viewer. The treatment of these shots in the AD text is, however, rather problematic. From (2), the women standing on the landing and looking on, the viewer is entitled to infer that they are looking into something like a room. But whilst the visual source text makes it clear in (3) that the women look into the passageway which leads to the studio, the AD in (2) insists that they look into a room (or even *the* room) before describing Griet's action in (2) as *stepping into the passageway which leads to the room* and explaining in (3) that Griet *opens the door*. The latter entails semantically that the door was previously closed. The description may leave the recipient wondering how the women were able to look into the studio when it was secluded by a passageway and when the door was closed. To avoid such incoherence, it might be better for the AD follow an internal logic, i.e., in this case to describe the spatial environment as this sequence of shots emerges as a whole.

Example 6 has illustrated some problems of recreating visual coherence across a small number of shots and within individual AD sections. What is equally important is coherence across AD sections which alternate with dialogue. A crucial point in this case is to ensure that recurring visual entities are referred to consistently. In the scene chosen for this case study, this creates further problems. The wooden mannequin in the studio, for example, is referred to as *a life-sized wooden figure* in shot (7) and as *the wooden mannequin* in shot (18). Given the lapse of time between the two mentions, the less than straightforward lexical relation between the two expressions does not provide the strongest possible support for recognizing the co-reference link. This example is investigated in Braun (in preparation) in connection with problems of referential identification in AD. What is interesting to note here is that the problems in this example are reminiscent of a number of problems discussed in this present case study. Thus, the inconsistencies in the two references is similar to

the inconsistencies in the spatial description in example 6, and a lapse of time also contributes to the problems in example 3 (Griet's dusting a mannequin's shoulder) and example 4 (Maria's reference to the unfinished painting).

5. Conclusion

This paper has explored some aspects of recreating coherence in AD, as part of a wider attempt to analyze and describe the intermodal translation processes taking place in AD. The focus was on intermodal and intramodal linking, based on a model of coherence which highlights the crucial role of a text recipient in recognizing explicit and implicit links in a text to construct a coherent discourse.

The case study has highlighted timeliness, precision and consistency in the descriptions as being important pre-requisites for a blind recipient's recognition of any potential links within an AD text or between the description and the other accessible elements of an audio described film. An equally important point, emerging from the discussion in section 4.3, is that it may sometimes make little sense for AD to try and give a step-by-step account of what is seen. Bearing in mind that AD has to be selective, the crucial point is that it 'tells a story,' i.e., produces an internal logic within the AD text and within the audio described version as a whole, rather than delivering isolated 'reports' about selected visual elements. On the face of it, this may be reminiscent of discussions about the appropriate degree of intervention and the role of the audio describer, i.e., the question of whether s/he merely fills in gaps or adopts responsibility for the audio described version as a whole (see Yeung 2007: 241). It may also remind us of discussions about whether AD is 'merely' an access service or a form of (narrative) art. Ultimately, however, the need to 'tell a story' is linked to achieving coherence, and this must be the goal of AD in any case, irrespective of how AD is perceived or how someone chooses to describe.

In connection with this, further research needs to investigate the effectiveness and efficiency of different types of description, i.e., the conditions under which coherence is most likely to emerge and the cognitive processing effort required. A case in point is the question of whether AD is more efficient when it supports the prospective activation of knowledge and mental modelling rather than the retrospective generation of inferences.

It is clear that the detailed analyses presented in section 4 are only possible for small data samples, which would seem to limit the value and validity of any such case study. This is compounded by the fact that the questions revolving around coherence in AD are only a small subset of the many research questions arising for AD as a relatively new form of intermodal translation (for an outline of these, see Braun 2008). However, they are useful from an explanatory and pedagogic point of view, because in the end, it is such analyses that will help to reveal why some AD solutions are likely to be more effective and efficient than others in supporting the creation of coherence.

NOTES

1. Partially sighted people process the description in conjunction with the accessible visual cues as well.
2. *Girl with a Pearl Earring* (2003): Directed by Peter WEBBER. Amsterdam: Archer Street Productions.
3. The same could be said for an analysis of verbal-visual links in terms of Barthes' (1977) concepts of 'anchorage' and 'relay.'

4. They further distinguish this from 'coherence impairment,' i.e., cases where there is a lowered or no expectation of coherence (e.g., discourse of small children, schizophrenic patients, some forms of art).
5. A similar approach was adopted by the team of Ulla Fix at Leipzig University in their project of analyzing an audio described film (Fix 2005, especially Morgner and Pappert 2005).
6. The house was introduced into the AD prior to this scene. Its presentation as 'known' (*the* house) is therefore natural and unproblematic here.
7. It would seem that the mannequin's pose *is reflected in* the painting rather than vice versa. Moreover, the utterance is structurally ambiguous, since *at the window* can refer to the woman or the painting.
8. The problematic expression *the room* in the AD of (2) is discussed in Braun (in preparation). It leaves the audience wondering which room is meant.

REFERENCES

Baker, Mona (1992): *In Other Words. A Textbook on Translation*. London: Routledge.

Barthes, Roland (1977): *Image, Music, Text*. New York: Hill.

Baumgarten, Nicole (2008): Yeah, that's it!: Verbal reference to visual information in film texts and film translations. *Meta*. 53(1):6-25.

Beaugrande, Robert and Dressler, Wolfgang (1981): *Introduction to Text Linguistics*. London: Longman.

Blakemore, Diane (1992): *Understanding Utterances*. Oxford: Blackwell.

Blum-Kulka, Shoshana (1986): Shifts of cohesion and coherence in translation. *In*: Juliane House and Shoshana Blum-Kulka, eds. *Interlingual and intercultural communication: discourse and cognition in translation*. Tübingen: Narr, 17-35.

Bourne, Julian and Jiménez-Hurtado, Catalina (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*: Jorge Díaz Cintas, Pilar Orero and Aline Remael, eds., *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam/New York: Rodopi, 175-188.

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

Braun, Sabine (2008): Audio description research: state of the art and beyond. *Translation Studies in the New Millennium*. 6:14-30.

Braun, Sabine (in preparation): Inferential identification in audio description. *In*: Sabine Braun and Margaret Rogers, eds. *Audio description for visually impaired people: Towards an interdisciplinary research agenda*.

Brown, Gillian and Yule, George (1983): *Discourse Analysis*. Cambridge: Cambridge University Press.

Bublitz, Wolfram (1999): Introduction: views of coherence. *In*: Wolfram Bublitz, Uta Lenk and Eija Ventola, eds. *Coherence in Spoken and Written Discourse: How to Create it and How to Describe it*. Amsterdam/Philadelphia: John Benjamins, 1-7.

Bublitz, Wolfram and Lenk, Uta (1999): Disturbed coherence: 'Fill me in.' *In*: Wolfram Bublitz, Uta Lenk and Eija Ventola, eds. *Coherence in Spoken and Written Discourse: How to Create it and How to Describe it*. Amsterdam/Philadelphia: John Benjamins, 153-174.

Bublitz, Wolfram, Lenk, Uta and Ventola, Eija, eds. (1999): *Coherence in Spoken and Written Discourse: How to Create it and How to Describe it*. Amsterdam/Philadelphia: John Benjamins.

Chaume, Frederic (2004): *Cine y Traducción*. Madrid: Cátedra.

Clark, Herbert and Clark, Eve (1977): *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich.

Edmondson, Willis (1999): If coherence is achieved, then where doth meaning lie? *In*: Wolfram Bublitz, Uta Lenk and Eija Ventola, eds. *Coherence in Spoken and Written Discourse*. Amsterdam/Philadelphia: John Benjamins, 251-265.

Fix, Ulla, ed. (2005): *Hörfilm. Bildkompensation durch Sprache*. Berlin: Erich Schmidt.

Gernsbacher, Morton and Givón, Talmy, eds. (1995): *Coherence in Spontaneous Text*. Amsterdam: Benjamins.

Halliday, Michael (1985): *An Introduction to Functional Grammar*. London: Arnold.

Halliday, Michael and Hasan, Ruqaiya (1976): *Cohesion in English*. London: Longman.

Hatim, Basil and Mason, Ian (1990): *Discourse and the Translator*. London: Longman.

Herman, David (2002). *Story Logic*. Lincoln: University of Nebraska Press.

Hyks, Veronikka (2005): Audio description and translation. Two related but different skills. *Translating Today*. 4:6-8.

Johnson-Laird, Philip (1983): *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge: Harvard University Press.

Johnson-Laird, Philip (2006): *How We Reason*. Oxford: Oxford University Press.

Kress, Gunter and van Leeuwen, Theo (2001): *Multimodal discourse*. London: Arnold.

Morgner, Henrike and Pappert, Steffen (2005): Darstellung des Untersuchungsmaterials: Sequenzprotokoll, Einstellungsanalyse und Transkription. *In*: Ulla Fix, ed. *Hörfilm. Bildkompensation durch Sprache*. Berlin: Erich Schmidt, 13-32.

O'Halloran, Kay, ed. (2004): *Multimodal Discourse Analysis*. London: Continuum.

Philipps, Patrick (2000): *Understanding Film Texts*. London: British Film Institute.

Reiss, Katharina (1981): Type, kind and individuality of text: Decision making in translation. (Translated by Susan Kitron) *Poetics Today*. 2(4):121-131.

Remael, Aline (2007): Audio description: a first attempt at investigating the challenges posed by film sound. Paper delivered at the seminar on *Audio description: towards an interdisciplinary research agenda*. University of Surrey, 28-29 June 2007.

Sacks, Harvey, Schegloff, Emanuel and Jefferson, Gail (1974): A simplest systematics for the organization of turn taking in conversation. *Language*. 50(4):696-735.

Salway, Andrew (2007): A Corpus-based analysis of the language of audio description. *In*: Jorge Díaz Cintas, Pilar Orero and Aline Remael, eds. (2007): *Media for all: Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam/New York: Rodopi, 151-174.

Sanford, Anthony and Garrod, Simon (1981): *Understanding Written Language: Explorations of Comprehension Beyond the Sentence*. Chichester: John Wiley/Sons.

Sanford, Anthony and Moxey, Linda (1995): Aspects of coherence in written language: a psychological perspective. *In*: Morton Gernsbacher and Talmy Givón, eds., *Coherence in Spontaneous Text*. Amsterdam: Benjamins, 161-214.

Shank, Roger and Abelson, Roger (1977): *Scripts, plans, goals and understanding*. Hillsdale: Erlbaum.

Sperber, Dan and Wilson, Deidre (1995): *Relevance. Communication and Cognition*. Oxford: Blackwell.

Tanskanen, Sanna-Kaisa (2006): *Collaborating towards Coherence: Lexical Cohesion in English Discourse*. Amsterdam: Benjamins.

Van der Heijden, Mereijn (2007). *Making film and television accessible to the blind and visually impaired*. MA dissertation. Utrecht School of the Arts. Visited on 27 October 2010, <http://www.audiodescriptie.com/downloads/Making%20film%20and%20television%20accessible%20to%20the%20blind%20and%20visually%20impaired%20-%20M.%20van%20der%20Heijden.pdf>.

van Dijk, Teun and Kintsch, Walter (1983): *Strategies of Discourse Comprehension*. New York: Academic Press.

Vendler, Zeno (1967): Verbs and times. *In*: Zeno Vendler, ed. *Linguistics in philosophy*. Ithaca: Cornell University Press, 97-121.

Ventola, Eija, Charles, Cassily and Kaltenbacher, Martin, eds. (2004): *Perspectives on Multimodality*. Amsterdam: Benjamins.

Widdowson, Henry (2007): *Discourse Analysis*. Oxford: Oxford University Press.

Yeung, Jessica (2007): Audio description in the Chinese world. *In*: Jorge Díaz Cintas, Pilar Orero and Aline Remael, eds. (2007): *Media for all: Subtitling for the Deaf, Audio Description and Sign Language*. Amsterdam/New York: Rodopi, 231-244.

Yos, Gabriele (2005): Verknüpfungen von Audiodeskription und Filmdialog. *In*: Ull Fix, ed. *Hörfilm. Bildkompensation durch Sprache*. Berlin: Erich Schmidt, 99-116.