Audio Description from a discourse perspective: a socially relevant framework for research and training

Sabine Braun
University of Surrey

The topic of this paper is Audio Description (AD) for blind and partially sighted people. I will outline a discourse-based approach to AD focussing on the role of mental modelling, local and global coherence, and different types of inferences (explicatures and implicatures). Applying these concepts to AD, I will discuss initial insights and outline questions for empirical research. My main aim is to show that a discourse-based approach to AD can provide an informed framework for research, training and practice.

1. Audio Description: from practice to research

AD in its present form is a rather recent development. It is a means of helping blind and partially sighted people to access audiovisual contents including film and theatre performances by using quiet moments in the original to provide a verbal description of the actions, scenery, body language, and other relevant details. While the practice of AD has its origins in theatre (Pfanstiehl & Pfanstiel 1985; Piper 1988), it has since been extended to cover a wide variety of settings from film and TV programmes to theatre, opera, museums and art galleries (see e.g. Benecke 2004; Holland forthcoming; Hyks 2005; Matamala 2005; Snyder 2005).

Published research on AD is still scarce. It has focussed almost exclusively on film and TV and has mainly addressed 'external' factors such as the overall technical conditions for the provision of AD, structural aspects of, and audience reaction to, audio described contents: The EU project AUDETEL (Independent Television Commission, 1992-95) investigated technical, artistic and economic issues associated with the provision of AD in the UK. Surveys among the target audience on their viewing habits and difficulties in following TV programmes, experimental viewing sessions and a critical analysis of existing materials revealed a wide range of AD styles and many differences in audience expectation (OfCom 2001; Pettitt *et al.* 1996; for further research on audience reception see Schmeidler & Kirchner 2001; on styles see Fels *et al.* 2006, Matamala 2005). The EPSRC project 'TIWO – Television in words' (University of Surrey, 2002-05) investigated AD texts as collateral texts in multimedia systems. Using quantitative methods, keywords and characteristic narrative structures were extracted from AD texts (Salway, forthcoming). In the project 'Hörfilme – compensation for pictures by words' (University of Leipzig, 2001-04), AD was analysed from a textlinguistic point of view, investigating e.g. how information unfolds in

AD texts and how links are created between AD text and film dialogue (Fix 2005). Piety (2004) suggested a systemic-functional framework for the analysis of AD. Other work has been of a more practical nature, including initiatives to produce guidelines for AD (Dosch & Benecke 1997; OfCom 2001; Orero 2005; Vercauteren 2007).

This work has provided valuable insights into the specific nature and structure of AD texts and many aspects of the production of AD. Much more work is currently underway. What has received less attention are the creative meaning-making *processes* involved in AD and its reception, *i.e.* in the comprehension of the audiovisual content and the production of the AD narrative[1] by the audio describer as well as the comprehension of the audio described content (the original verbal utterances and sound effects in conjunction with the AD narrative) by the target audience.

AD can be characterised as a complex cognitive-linguistic and intermodal mediation activity. Its aim is to produce verbal discourse (AD narrative) which describes the essential visual elements and other relevant elements (e.g. some sound effects which are difficult to interpret without access to visual information) of a multimodal discourse (*i.e.* the original audiovisual event containing verbal, auditory and visual elements). The outcome of the process of audio describing (AD narrative) forms part of a new multimodal discourse (*i.e.* the audio described content, containing verbal and auditory elements). It involves processes of discourse comprehension and production in which different semiotic modes interact with each another as well as with the individual knowledge, experience and expectations of those participating in the discourse.

Insights into these processes and interactions come from early work in psycholinguistics on the interaction of linguistic cues and knowledge (e.g. Clark & Clark 1977; Levelt & Flores D'Arcais 1978), work on visual perception and cognition (e.g. Marr 1982; Kosslyn & Osherson 1995) as well as the study of the mental representation and retrieval of knowledge (e.g. Shank & Abelson 1977; Johnson-Laird 1983) and mental imagery (e.g. Kosslyn *et al.* 2006). Approaches to discourse analysis which are concerned with information processing, achieving coherence in discourse and the role of inferencing are capable of bringing these insights together and can provide an understanding of how discourse is cognitively processed. While originally focussing on verbal language (Blakemore 1992; Brown & Yule 1983; Sperber & Wilson 1995; van Dijk & Kintsch 1983), discourse analysis is widening to take other semiotic modes into account (Kress & van Leeuwen 2001; Ventola *et al.* 2004). A discourse-based approach can, therefore, provide a theoretical and methodological framework for addressing the 'internal' factors of AD, *i.e.* the creative processes involved in the reception of the original audiovisual content, the production of the AD narrative and the reception of the audio described event.

In this paper I will outline such an approach to AD focussing on three dimensions: the role of mental modelling (section 2), the role of different types of inferences (explicatures and implicatures, section 3), and the

role of local and global coherence (section 4). Using examples from the film *The Hours* (Daldry 2002), I will apply these concepts to AD, discuss some initial insights and outline questions for further research.

2. A mental modelling approach to AD

The verbal discourse created in the process of audio describing is 'non-autonomous': it is intended to be processed in conjunction with the dialogue or speech and sound effects of the original audiovisual event. The blind audience uses the different elements available (original dialogue/speech, sound and AD narrative) to form a coherent whole. Likewise though, the audio describer does not perceive the different modes of the original separately. He forms a coherent whole of the elements available to him. One question which arises, therefore, is how the audio describer is able (or can be enabled) to identify and isolate from this coherent whole the information which is not available to the blind audience but essential in constructing a similar coherent whole.

    As a first step it seems useful to analyse what happens in the process of understanding verbal discourse, sound and images. Important clues come from mental modelling theory as well as research into visual perception and cognition. Johnson-Laird (1983: 141) has argued that we experience and interpret the world around us through our ability to construct mental models of reality (including fiction). In the comprehension of verbal discourse, addressees build up a mental model of the situation described, using the linguistic cues provided in the verbal utterances (bottom-up processing) as well as world knowledge, the perceived context of situation and previous utterances (top-down processing) (Brown & Yule 1983: 234). On the basis of this, addressees also create expectations about how the discourse continues. Research into visual perception/cognition reveals similar bottom-up and top-down processes for the comprehension of visual input (Dretske 1995), and for the purposes of this paper it will be assumed that mental models of audiovisual events are based on verbal, visual and auditory cues together with the knowledge and the associations that these evoke.

    An initial example from the film *The Hours* will be used here to illustrate this. The opening scene of the film is set in Sussex in 1941. After a brief shot of a river flowing though a countryside landscape, a close-up presents a woman's hands tying up the belt of a coat as if to get ready to go out. The subsequent shots reveal the woman wearing the coat and show her leaving her house and crossing her garden. Then we see part of a woman in a writing pose, in particular a close-up of a woman's right hand, part of her arm, her blouse and the top of what could be a Davenport desk. Using a dip pen, the woman writes something on a piece of paper. As the nib of the pen can be heard moving across the paper, a female voice-over starts reading: "Dearest, I feel certain that I go mad again. I feel we can't go through another of those terrible times and I shan't recover this time". As the voice-

over continues, shots alternate repeatedly between the woman in writing pose and the woman walking away from her house.

In this scene, visual cues (the woman holding the pen and the woman leaving her house) combine with an initial verbal cue (the voice-over), the viewer's knowledge (about typical uses of voice-over in feature films), further verbal cues (what the woman reads out) and an auditory cue (the nib) to create an assumption that the woman in writing pose is identical with the woman who leaves the house, that the voice belongs to her and that she is writing. Gradually we create an initial mental model of a woman writing a letter and then leaving her house. This is still highly hypothetical (what is being written could be a diary entry; the order of events is not entirely clear yet), but it is confirmed later (her husband is shown to find two envelopes on the mantelpiece).

As the voice-over continues to read "I begin to hear voices and can't concentrate. So, I'm doing what seems to be the best thing to do", the woman, who has arrived by the river bed, lifts up two large stones, puts them in her pockets and starts walking into the river. Most viewers will form the hypothesis that the woman is going to commit suicide from this visual cue, even before this can be construed quite clearly from the letter. It is confirmed shortly after by the image of the woman's body disappearing in the waters of the river. While the letter is simply undersigned 'Virginia', most viewers are likely to associate the scene with Virginia Woolf, through previous knowledge about the film or the writer (or both). Hence, at that point most viewers will have created a rather comprehensive mental model about the last moments of Virginia Woolf's life.[2] However, differences in knowledge about the writer, the reasons that led her to commit suicide, imaginativeness etc are likely to provide for considerable individual variation.

Mental models are the result of a complex bottom-up and top-down interaction of cues from a variety of sources. In the opening scene of The Hours, the comprehensive (or 'higher order') hypothesis that the scene is about the last moments in Virginia Woolf's life is construed on the basis of individual hypotheses (the woman writing a letter, the suicide and others) and assumptions of how these events hang together e.g. in space, time and intent (cf. Miller & Johnson-Laird 1976), which are in turn created on the basis of a range of individual cues. Equally important, these cues vary in their degree of reliability, and in their weight: some can be said to be decisive in forming a hypothesis, while others reinforce, confirm, complement or contradict emerging or established hypotheses. The image of the woman by the riverbed putting stones in her pocket and walking into the river certainly delivers the decisive cue for the suicide hypothesis. By contrast, the formation of the letter-writing hypothesis relies on so many cues that it seems more difficult to identify the one that is decisive. Interestingly though, the auditory cue of the sound of the nib does not seem to be more than complementary (for the sighted viewer!).

An example of cues which contradict initial hypotheses can be found in a later scene in the film, in which the second protagonist (Laura, who lives in Los Angeles in the 1950s) is introduced through a breakfast scene with her husband and her little son. While the conversation (about Laura's pregnancy) and the setting (a tidy family home) lead to the hypothesis that this is about a happy family, Laura's facial expressions, especially her last daunting gaze towards her husband as he is leaving, suggest otherwise.

Yet other information in any film scene (or text) may not immediately provide a cue for comprehension, but may be invested with meaning later, especially when the aim is to create suspense. Thus in the breakfast scene, Laura's little son wears shiny pyjamas with large all-over print. In a later scene, in which the third protagonist (Clarissa, living in New York in 2001) visits her long-term friend Richard, who is dying of AIDS, Richard wears a dressing gown made of similar material and with a similar print pattern. Thus the boy's pyjamas could suddenly be understood as a first hint that the boy is Richard in his childhood[3] – and hence create the expectation that the film will reveal something about Richard's life.[4] Further cues would then be processed in the light of this expectation.

In AD, the audio describer's task is to enable the blind audience to build up a mental model of each situation (and eventually of the whole film) that is similar to the mental models created by sighted viewers (or at least to the audio describer's mental model). A mental modelling approach makes it clear that the difficulty is far greater than 'just' translating visual images into words. One of the major problems seems to be the following: On the one hand, mental models appear to be 'holistic', *i.e.* it seems difficult to establish the precise contribution of each cue. At least, as Brown & Yule (1983: 250) have argued, such 'decomposition' is not how discourse is normally processed. On the other hand, audio describers need to be highly selective with regard to the cues they describe, because AD narrative has to fit into often short pauses between dialogue/speech, *i.e.* is compounded by timing constraints (see also Hyks 2005).

The process of audio describing, therefore, has to involve intense assessment and decision processes. In the overall flow of visual, verbal and auditory input, the audio describer needs to identify relevant cues and assess their contributions, *i.e.* evaluate them in terms of their relative weight (decisive, reinforcing, confirming, complementary, contradictory, etc.) and reliability at the point of occurrence. Could the sound of the nib in conjunction with the voice-over, for example, be identified reliably without any visual information? Furthermore, when the decisive cue appears to be visual, the audio describer needs to assess whether there are any redundant cues which would be strong enough as 'stand-alone' cues for the blind audience. Does the content of the letter, for instance, foreshadow the suicide as dramatically as the woman putting stones in her pockets?

A model of discourse processing which can provide guidance for the assessment and selection process would be highly relevant in the context of AD and in particular for beginning audio describers. Arnold & Whitney

(2005: 340) have pointed out that "our brain has a dynamic strategy that weights all the available cues according to their relative reliability when integrating information from multiple sources." It would be an interesting task for further research to investigate whether such strategies can be isolated and operationalised for the purposes of AD.

Through its focus on the different sources of comprehension and how these 'work together', a mental modelling approach seems to provide a valid starting point for this research. As emphasised by Brown & Yule (1983: 255), mental modelling theory has allowed for a richer and also more individual representation of situations than many other accounts of discourse processing. However, it has been less concerned with the question of precisely how different cues and knowledge are integrated into mental models, and how (only) relevant knowledge is activated in the creation of mental models (see also Harley 1995: 227-228). Insights into this come from inferential models of communication, which will be discussed in section 3.

3. Explicating and implicating in AD

Inferential models of communication have focussed on how individual utterances are processed and understood rather than on how addressees create the 'big picture'. One of the most comprehensive accounts of communication as an inferential process was suggested by Sperber & Wilson (1995), in the framework of Relevance Theory (RT). As with mental modelling theory, Sperber & Wilson's account cannot only be used to describe how *verbal* utterances are processed, but also to describe some aspects of *visual* processing which are highly relevant for AD. However, I will first outline the overall approach using a verbal utterance as an example.

Following earlier inferential models of communication (e.g. Grice 1975), Sperber & Wilson emphasise that in order to understand the meaning of an utterance, addressees normally need to identify its propositional (factual) content as well as the assumptions that the speaker wants to make manifest with an utterance. What sets RT apart from earlier approaches is a) the claim that deriving both the propositional content and the speaker's assumptions are highly inferential processes (Sperber & Wilson 1995: 183) and b) a distinction between assumptions which the speaker wants to communicate explicitly (*explicatures*)[5] and assumptions which the speaker encourages the addressee to make implicitly (*implicatures*) (*ibid*: 256).

The opening sentences of Virginia Woolf's letter ("Dearest, I feel certain that I go mad again. I feel we can't go through another of those terrible times …") can be used to demonstrate this. Firstly, the identification of the propositional content is difficult here because the phrase "those terrible times" is potentially ambiguous. On the basis of the preceding cotext ("I feel certain that I go mad again") it can be inferred that Virginia Woolf used it to refer to her personal situation rather than e.g. to England in

or shortly before 1941 – but it is difficult to establish this with certainty. This goes to show that even what an addressee perceives to be the 'factual' content of the utterance partially rests on inferences, *i.e.* on individual interpretation. Secondly, even after disambiguating the phrase, many questions about its meaning remain: What were those situations like? What was "terrible" about them? By referring to "those terrible times", Virginia Woolf may have wanted to remind her husband of her mental illness in general or of particular instances of her depression, or both. All of these would be explicatures, but again there is room for individual interpretation. Thirdly, depending on the sincerity assigned to the phrase "those terrible times", the implicatures could range from the assumption that Virginia Woolf is going to make a change in her life to the assumption that she is going to put an end to it.

The difference between explicatures and implicatures is a qualitative difference. Explicatures are basically construed from what is said, whereas implicatures go much beyond this (Blakemore 1992: 59). However, both explicatures and implicatures create meaning (*i.e.* have contextual/cognitive effects in Sperber & Wilson's terminology). Virginia Woolf could have written "My life has become unbearable". While the implicatures from this utterance would be very similar to those of the original, the overall meaning of the utterance would be different because of different explicatures.

Propositions and the assumptions derived from them can be seen as hypothetical mental representations that are integrated into mental models, even though this process does not seem to be straightforward [6]. Sperber & Wilson's essential claim that communication is highly inferential and relies on the activation of different types of assumptions accounts for the observation that hypotheses in mental models vary in terms of strength and reliability, as discussed in section 2. Another crucial claim of RT can account (better than mental modelling theory) for how addressees limit the scope of the assumptions and knowledge they activate: Sperber & Wilson (1995: 155-163) claim that all types of inferences are guided by one common underlying principle – the Principle of Relevance. According to this principle, addressees are entitled to believe that an utterance or cue is presented to them in the optimally relevant way. This enables them to use their processing capacities efficiently and in particular to stop processing an utterance once they have derived a meaning which they find sufficiently relevant. Utterances which require a high processing effort to reach this point, usually yield greater meaning effects (e.g. poetry).

Clearly, AD deals with multimodal discourse, and the essential 'candidate' for AD is the visual mode. It has been argued that visual input (as opposed to verbal utterances) does not lead to the identification of propositional representations, but – in most cases at least – to depictive representations (Kosslyn *et al.* 2006: ch. 1). As Kosslyn *et al.* demonstrate, the two types of representation make different types of information more easily accessible. However, there is no reason not to assume that inferences are

involved in the identification of both types of representation and that both are capable of encouraging explicit and implicit assumptions.

To illustrate this, I will return to part of the opening scene of *The Hours*. In Sperber & Wilson's terms, the close-up of a woman's hand together with the pen, the desktop and part of a woman's blouse provide the cues for identifying the 'factual' content of this shot. At the most basic level, this hinges on the successful recognition (viz. classification) of the individual items. Some of them may be ambiguous. The representation created on the basis of the assumption that the featured item is a pen (rather than e.g. a paint brush) would carry the *explicature* that there is a woman sitting at a desk writing something. This can be derived through one or more bridging inferences (e.g. 'hands belong to human bodies'; see Clark & Haviland 1974)[7]. The *implicatures* could include the assumptions that she is writing a letter or diary, or has something important to 'say'.

The AD narrative of *The Hours*[8] is particularly interesting with regard to the explicature-implicature distinction here. The above scene is described as follows: "She [the woman] crosses the lawn to a wicket gate at the end of the garden. Earlier, she sits writing." Using the phrase "she sits writing", the audio describer decided to verbalise the explicature, leaving it to the audience to draw implicatures regarding the purpose of the writing. He did not, however, describe the individual cues which led to the explicature. A description of a top of a Davenport desk and a woman's hand holding an antique-style dip pen may have created a rather 'poetic' effect, but it is crucial to understand the solution in the overall context. Firstly, there are, of course, the ever prevailing timing constraints. The above AD narrative was followed immediately by the beginning of the woman's voice-over. Therefore, something very short was required to convey the main action of writing, especially if the audio describer did not deem the sound of the nib of the pen to be a sufficiently strong cue for this action (which it may well not be in this case). Secondly, and equally importantly, the many shot changes in the opening scene require a large number of bridging inferences. Describing the individual cues instead of verbalising the explicatures would leave the blind audience with a rather heavy processing load, especially at the beginning of a film, where there are many other impressions to process. By contrast, describing the action of writing gives the audience instant access to the most relevant interpretation at explicature level. Similar considerations may have led the audio describer to indicate that the woman who "crosses the lawn" is identical with the woman who "sits writing" (through the anaphoric use of "she") and to provide a clear hint for the sequential order of the actions ("earlier").

The strategy of verbalising explicatures instead of the individual cues by which they are triggered can contribute to reducing the processing load of the audience and the timing problems typical of AD. At the same time, AD narrative which verbalises no more than the explicatures leaves scope for individual interpretation (*i.e.* for drawing implicatures). Holland (forthcoming) emphasises that AD should not just be taken to be a service ena-

bling blind people to access audiovisual content. It should provide them with the same freedom of interpretation that sighted people enjoy when engaging with audiovisual content and especially art work. If the point of art work is to create rich meanings and to stimulate assumptions, and if – as Sperber & Wilson argue – greater processing effort yields greater meaning effects, then the explicature-implicature distinction could provide useful guidance on what to describe and what to leave for the audience to infer.

The differences between propositional representations (created from verbal utterances) and depictive representations (created from visual input) would be an important topic for further research in the context of AD. Moreover, the question of *how* the explicatures of a visual image can be 're-created' in AD narrative needs to be investigated further. The final section will look into some of the issues which are relevant in this connection.

4. Local and global coherence in AD

The previous two sections should have made it clear that the challenges of AD start long before reaching the stage of producing the AD narrative, *i.e.* of putting selected visual cues (viz. their explicatures) into succinct and appropriate words. But when considering the actual creation of AD narrative, many more points need to be accounted for. One important point, which is characteristic of AD and other forms of Audiovisual and Multimedia Translation, is that the outcome of the translation process is not a stand-alone product. Thus, AD narrative is processed in conjunction with the verbal utterances and sound effects of the original, *i.e.* it forms part of a multimodal discourse. This places specific demands on the creation of coherence within and across individual AD sections, and between these and the other elements of the audio described event. The final section of this article will, therefore, focus on one further set of concepts from discourse analysis, which is capable of providing another starting point for further research and guidance – the creation of local and global coherence (Blakemore 1992; Brown & Yule 1983; van Dijk 1977).

Coherence can be characterised as connectivity in discourse. It is important to note that coherence is 'connectivity of content' rather than a connection between means of expression (Blakemore 1992: 85), and that it is created in the addressees' mind, based on their general assumption that utterances make sense in the context in which they appear (Brown & Yule 1983: 66). However, the selection of appropriate (verbal, auditory and visual) means of expression, which indicate how the parts of a discourse hang together, can certainly support the creation of coherence in the mind. Van Dijk (1977) distinguished two levels at which to achieve coherence, local and global coherence: In verbal discourse, *local* coherence is created between conjoined utterances (see also Blakemore 1992: 84-88), whereas *global* coherence emerges from an overall discourse topic and consistency of e.g. style, register and choice of expressions. In multimodal discourse

such as film, a similar distinction could be drawn, with local coherence being created within individual scenes and global coherence reaching out across scenes. Importantly though, in multimodal discourse coherence also needs to be achieved across different modes of communication.

Using one last example from the film *The Hours*, I will briefly exemplify this by analysing relevant AD sections from the breakfast scene (featuring the little boy, his mother Laura and his father) and the visiting scene (featuring Richard and Clarissa). In the breakfast scene, the father tells the little boy to eat his breakfast, reminding him that he needs to be strong because he will soon be the 'big boy' in the family (his wife Laura is visibly pregnant). As the father is leaving, one shot shows the boy looking up from his breakfast and, from a distance, the father pointing his finger at him and looking at him as if to say "remember what I told you". Yos (2005: 105) points out that conventional interpretations of gestures and facial expressions can often be verbalised more effectively than the actual behaviour (see also the difference made in section 3 between individual cues and the explicatures they trigger; the conventional interpretations of body language would be explicatures). In the AD narrative the father is described as pointing "his finger in a mind-you-eat-your-breakfast kind of way to the boy". The audio describer takes up a point made by the father in the original film dialogue, using similar means of expression. This creates *local* coherence between the audio description of the father's body language and the preceding film dialogue within this scene.

Later in the same scene, Laura reiterates to the boy that he should eat his breakfast. The subsequent action of the boy is audio described as follows: "keeping his mother in firm view, he obediently goes back to the kitchen area". In the scene featuring Clarissa and Richard, Clarissa tries to convince Richard to come to the party which she is organising to honour him (the same evening). After a slight argument, Clarissa firmly tells him that she will be back later to pick him up and asks him to be ready. Richard is described here as nodding "obediently" and standing "watching her as she goes to the door". Upon leaving, Clarissa is portrayed as pointing "an admonishing finger at him". Arguably the film director wanted to create some similarities between the two scenes to give the audience first hints about the identity of Richard and the little boy, and he used visual rather than verbal cues to do so [9]. The audio describer made sure that those gestures which are similar (the gestures expressing obedience and the finger pointing) are actually selected for description in both scenes, and he used similar vocabulary in each case. This way he supports the process of creating *global* coherence across the two scenes.

5. Conclusion

In this paper I have argued that approaches to discourse analysis concerned with information processing, coherence and the role of inferences provide a

useful framework for analysing the discourse processes involved in AD and its reception. The further study of AD requires further research into the role of the various modes of communication in the relevant comprehension and production processes. Regarding the audio describer's comprehension process, this calls for (empirical) research into the meaning-making potential of the verbal, visual and auditory modes and how they "complement and undercut one another in relation to the dramatic sequences as a whole" (Doloughan & Rogers 2005). With regard to the creation of AD narrative and audience reception, more research is required into the linguistic and communicative strategies for creating verbal discourse from visual input under the specific conditions of audio describing. This needs to include the verbal means of expression which are best suited to enable a blind audience to form a coherent mental model of an originally audiovisual event. A systematic analysis of existing materials (Piety 2004; Salway forthcoming) is an important step in this direction.

A discourse-based approach also forms a useful basis for AD training. Knowledge about the relevant discourse processes raises trainees' awareness of the challenges. This will help them to develop AD-specific comprehension and production strategies and enable them to cope with novel and unpredictable situations. This point is crucial in a society which is increasingly relying on audiovisual contents and in which the scope of AD, therefore, needs to be extended to cover a wide range of settings. A discourse-based approach, which foregrounds the complexities of AD as a multimodal mediation activity, promotes the creation of audio described contents which enable the audience to participate as fully in an audiovisual event as possible, and thus contributes to media accessibility. Outcomes of further research into AD also open up opportunities for research into related fields such as subtitling for the deaf and hard-of-hearing.

Bibliography

Arnold, Derek & Whitney, David (2005). "Adaptation and perceptual binding in sight and sound". In Clifford, Colin & Rhodes, Gillian (Eds), *Fitting the Mind to the World*. Oxford: OUP, 339-360.

Benecke, Bernd (2004). "Audio-Description". *Meta* 49(1), 78-80.

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

Pfanstiehl, Margaret & Pfanstiehl, Cody (1985). "The play's the thing - audio description in the theatre". *BJVI* 3(3), 91-92.

Brown, Gillian & Yule, George (1983). *Discourse Analysis*. Cambridge: CUP.

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

Clark, Herbert & Haviland, Susan (1974). "Psychological processes as linguistic explanation". In D. Cohen (ed), *Phenomena*. Washington: Hemisphere Publishing.

Dijk, Teun van (1977). *Text and Context London*: Longman.

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

Doloughan, Fiona & Margaret Rogers (2005). "To see things and texts: audiovisual translation, audiodescription and multimodality". Paper delivered at conference on *Multimodal Texts: Engaging Sign Systems*, Oct. 22-23, 2005 at the University of Portsmouth.

Dosch, Elmar & Benecke, Bernd (1997). *Wenn aus Bildern Worte werden – Durch Audio-Description zum Hörfilm.* München: Bayerischer Rundfunk.

Dretske, Fred (1995). "Meaningful perception". In Kosslyn, Stephen & Osherson, Daniel (Eds), *Visual Cognition: An Invitation to Cognitive Science* (vol. 2). Cambridge: MIT Press, 331-352.

Fels, Deborah; Udo, John Patrick; Diamond, Jonas & Diamond, Jeremy (2006). "Comparison of alternative narrative approaches to video description for animated comedy". *JVIB* 100: 5, 295-305.

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

Grice, Paul (1975). "Logic and Conversation." In Cole, Peter & Morgan, Jerry (Eds), *Syntax and Semantics. Vol. 3: Speech Acts.* New York: Academic Press, 41-58.

Harley, Trevor (1995). *The Psychology of Language. From Data to Theory.* Hove: Psychology Press.

Holland, Andrew (forthcoming). "For me Audio Description is so much more". In Anderman, Gunilla & Diaz Cintas, Jorge (Eds). *Audiovisual Translation: Language Transfer on the Screen.* London: Palgrave McMillan.

Hyks, Veronika (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, MA: Harvard University Press.

Kosslyn, Stephen; Thompson, William & Ganis, Giorgio (2006). *The Case for Mental Imagery.* Oxford: OUP.

Kress, Gunter & van Leeuwen, Theo (2001). *Multimodal Discourse.* London: Arnold.

Levelt, Willem & Flores D'Arcais, Giovanni (Eds) (1978). *Studies in the Perception of Language.* Chichester: Wiley & Sons.

Marr, David (1982). *Vision.* San Francisco: W.H. Freeman.

Matamala, A. (2005). "Live Audio Description in Catalonia". *Translating Today* 4, 9-11.

Miller, George & Johnson-Laird, Philip (1976). *Language and Perception.* Cambridge: CUP.

OfCom (2001). *ITC Guidance.* http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/audio_description/index.asp.html (accessed on 9 March 2007).

Orero, Pilar (2005). "Audio Description: Professional recognition, practice and standards in Spain". *Translation Watch Quarterly* 1, 7-18.

Pettitt, Bridget; Sharpe, Katharine & Steven Cooper (1996). "AUDETEL: Enhancing television for visually impaired people". *BJVI* 14(2), 48-52.

Piety, Philip (2004). "The language system of audio description: an investigation as a discursive process". *JVIB* 98: 8, 453-469.

Piper, Mark (1988). "Audio Description: pioneer's progress". *BJVI* 6(2), 75-76.

Pfanstiehl, Margaret & Pfanstiehl, Cody (1985). "The play's the thing - audio description in the theatre". *BJVI* 3(3), 91-92

Salway, Andrew (2007). "A Corpus-based Analysis of the Language of Audio Description." *Media for All. Subtitling for the Deaf, Audio Description.* Amsterdam: Rodopi.

Schmeidler, Emilie & Kirchner, Corinne (2001. "Adding audio description. Does it make a difference?" *JVIB* 95: 4, 197-212.

Shank, Roger & Abelson, Robert (1977). *Scripts,Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures.* Hillsdale, NJ: Erlbaum.

Snyder, J. (2005). "Audio Description. The visual made verbal across arts disciplines - across the globe". *Translating Today* 4, 15-17.

Sperber, Dan & Wilson, Deidre (1995²). *Relevance. Communication and Cognition.* Oxford: Blackwell.

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

Vercauteren, Gert (2007). "Towards a European Guideline for Audio Description." *Media for All. Subtitling for the Deaf, Audio Description.* Amsterdam: Rodopi.

Yos, Gabriele (2005). "Verknüpfungen von Audiodeskription und Filmdialog." In *Fix*, Ulla (Ed), 99-116.

Filmography

The Hours (2002).
USA/UK.
Dir. Stephen Daldry

---

1   I use this term in a loose sense to refer to what an audio describer produces.

2   Those who know and recognise Virginia Woolf's letter (which is authentic) are, of course, likely to build up a richer and more stable mental model. In principle, however, the process of activating knowledge on the basis of verbal and other cues is the same as the one described here.

3   This hint is very vague, illustrating the point that cues may vary in their reliability.

4   The expectation that the lives of the three women are intertwined (and that the film will reveal how) is established earlier in the film and may also be known to the audience through the blurb.

5   Sperber & Wilson (1995: 224) point out that in ordinary assertions the main explicature is the full-blown proposition, but also discuss other cases.

6   This process involves e.g. merging, strengthening and deleting individual propositions and assumptions (Sperber & Wilson 1995; Clark & Clark 1977; van Dijk & Kintsch 1982).

7   Blakemore (1992: 74-77) makes it clear that such inferences are part of the explicating process.

8   The film was audio described by IMS Media, London.

9   Another such cue is the similarity of the little boys' pyjamas and Richard's garments, as described in section 1. However, I will not discuss the audio describer's solution for this here.