

# **A method for analysing multimodal research material: audio description in focus**

**Maija Hirvonen**  
**University of Tampere**

**Liisa Tiittula**  
**University of Helsinki**

## **Abstract**

A modern text is seldom a stand-alone product consisting merely of writing, but a multimodal body of signs from different meaning-making systems. The multimodal context and the new forms of translation, in which the concept of 'text' goes beyond language, have become increasingly relevant in contemporary translation research and practice. In order to study multimodal context and intermodal forms of translation accurately, research methodology has to be adapted. The aim of this article is to present a systematic method for analyzing multimodal material, or 'text', from the point of view of translation analysis, and intermodal translation that takes place in this context. The method is applied to study audio description, a form of translation in which visual information is compensated by verbal descriptions in order to make a text more intelligible to blind and visually impaired audiences. As an illustration, the article looks at how space is constructed in a film and its audio description, and presents a model which allows it to expose and contrast the multiple modes of meaning-making that are at play: in this case, the visual, auditory and linguistic representations of space. The outcome is a systematic presentation of complex audiovisual material in a textual form. The method can be applied to cases whose research material is multimodal and/or which undertake the challenge of understanding intermodal translation.

Keywords: methodology, multimodality, translation analysis, audio description

## **1 Multimodality and translation**

Today, much of the material that is being translated or interpreted is multimodal, that is, it makes use of multiple semiotic resources such as language, sound and image. Think, for example, how often a translator or interpreter is encountered with a source text from the audiovisual media (television, cinema, internet, etc.). Besides this, the more traditional forms of meaning transfer can also be regarded as multimodal. Writing, and even more so publishing, have for long involved modes other than language in the creation of text: typographic means such as font type and size, and, in some cases, images have been in use lending meaning to the entirety of message (Schopp 2008; Kress & van Leeuwen 2006). Also spoken discourse, which constitutes the source material for interpreting, is highly multimodal. When people communicate, they do it not only verbally but also with non- and paraverbal means such as body language and vocal cues. It seems therefore utterly relevant in today's T/I research to be able to analyse source and target material that is multimodal (cf. Remael 2001).

Multimodality has become a watchword in other fields as well. As a general notion, it can be defined as “interdependence of semiotic resources in text” (Ventola et al. 2004: 1–2). Stöckl (2004: 9) brings the concept further by saying that

multimodal refers to communicative artefacts and processes which combine various sign systems (modes) and whose production and reception calls upon the communicators to semantically and formally interrelate all sign repertoires present.

Thus multimodality refers to relations established between different sign systems (or modes, or semiotic resources), and multimodal communication can be described as congruent meaning making based on the interactive presence of different meaning-making systems.

This article attempts to demonstrate a method for analysing multimodal material with an example that originates in filmic audio description. As a form of intermodal translation, audio description means verbalising visual information into a spoken form for the benefit of visually impaired people. In a film, audio description mediates, among others, characters’ appearances and nonverbal communication, sceneries and settings, and it appears as so-called “acoustic subtitles” (Poethe 2005) that are slotted in between dialogue. Audio description thus interacts with the film’s soundtrack constituting a multimodal, verbal-oral-aural text (Snyder 2005). Translating from one mode into another, from images into words, our interest in multimodality was aroused by the contemplation of intermodality. The article will, consequently, be looking at not only how different modes can be sorted out for multimodal analysis but also the way by which modes can be distinguished from the material for intermodal analysis.

Earlier accounts of multimodal analysis on (audiovisual) translation exist, some of which base their study in the multimodal transcription model proposed by Baldry (2000) (see e.g. Taylor 2003; also Mubenga 2009 discusses a multimodal approach to translation analysis). This model suggests a detailed transcription of each mode as they appear, and are active at a given moment: for example, in a half-a-minute video. The transcription allows, according to Baldry (2004: 84),

a TV advert to be reconstructed in terms of a Table containing a chronological sequence of frames, a technique that goes a long way to resolving the difficulties of taking linguistic, musical and pictorial modes into account.

Another example comes from Morgner & Pappert (2005) as a textual transcription of filmic audio description. The authors present audio description as part of the film in a table with separate columns for visual (detailing camera position, movement, etc. as well as action in each shot, i.e. in film’s smallest narrational unit, a single run delimited by a cut or blend) and auditory (e.g. dialogue and sounds) narration. Audio description is slotted within the dialogue so that the reader becomes aware of the interrelatedness between the two auditory resources. Similarly to Baldry’s (2000) model, this table can be read both horizontally and vertically, the former describing what happens simultaneously, and the latter representing the chronological order of things.

Our proposal relates to these two models in that we present the material also in a table and aim at a systematic representation of the semiotic resources of the source material. In continuation, our method is presented in three steps. First, we will explain our starting point, translation analysis, from which our multi- and intermodal deviations are shaped. The second step describes and exemplifies this framework to analysing translation as a transformation process between source and target text, and the third part demonstrates how to study translation as target text. We will also offer examples from our research to illustrate the use of the method. Finally, we will discuss the outcome and contemplate on prospective proposals for improvement.

## **2 Translation analysis as framework**

Translation and interpreting as intercultural, -linguistic, -modal and -textual activities provide for a diversified set of approaches and, thus, methodologies for research. Depending on the aim and interest, translational activity can be studied from the point of view of its process, participants, context, content, influence, and so forth. In compliance with these different terrains, it can be approached from different disciplines, each of which entail their own theories and methodologies: cultural, sociological, linguistic, computational, and other approaches offer important auxiliary methods for T/I's research questions. (For more discussion, see Duarte 2006.) Our framework is translation analysis, into which a suitable methodology may be added. English language enables to frame *translation* in two ways: as product (target text) and process (meaning transfer from source text to target text). Consequently, 'translation analysis' can refer to two kinds of contemplation:

- 1) **Target-text analysis** (TT analysis), where the study is conducted on the product(s). A translation is seen as a "stand-alone" product that can be analysed independently from its source text in its target context. The object of analysis can consist of several texts, for example, in the event that a text has been translated more than once or when the same text is translated into various languages.
- 2) **Source- and target-text analysis** (ST-TT analysis), where translation is seen as a transformation process of the source text into a target text. Therefore, features of the process are understood by comparing the two or more (in case of several translations) texts. The comparison will be made according to certain parameters; for instance, the differences and similarities in source and target texts can be described in order to find out how the translation relates to its original.

Hence, the notion of translation analysis gives a two-fold framework to approach the study of translation. Yet the frame needs methodological tools, and these can be retrieved from other, relevant disciplines. As the principal objects of interest in this article are multimodal texts, a natural filling comes from any field to which such text is attached: in our case, relevant supporting methods are film analysis (the source and target texts are filmic narratives) as well as discourse and conversation analysis (the target text, i.e. audio described film with acoustic elements, dialogue and audio description, constitutes a communicative event). With the combination of these

methods, we will now continue with demonstrating how filmic audio description can be analysed multi- and intermodally.

### 3 ST-TT analysis

This section presents the aforementioned variation of translation analysis with an objective to compare source text to target text(s). This will be illustrated with samples from actual research that investigates spatial representations inter- and multimodally in filmic audio description (Hirvonen 2010). Spatial representations refer to visual, auditory and linguistic portrayals of place/location (the “static” dimension of space) and action between characters (the “interactive” dimension of space). The samples come from the film *Der Untergang* (Germany, 2004) and its two audio descriptions in German and Spanish.

In what follows, we illustrate our table model and explain how it can be applied to multi- and intermodal ST-TT analyses. First, we present a general schema of the model (Table 1) and then illustrate how to apply it to a research problem, which is our case is space (Tables 2–3).

**Table 1.** A general schema

Visual mode (image)	Auditory mode (sounds and music)	Auditory mode (character speech)	Auditory mode (AD)
- shot size - camera movement - mise-en-scène - text on screen - nonverbal activities - ...	- sound effects - music - ...	- language - voice quality - ...	- language - voice quality - ...

In Table 1, the principal distinction has been done according to the two sensory channels at stake in filmic material, that is, between visual and auditory channels (modes). These are then further divided into core modes (cf. Stöckl 2004): ‘image’, ‘sounds and music’, and language (split into ‘character speech’ and ‘audio description’). Finally, the core modes can be assigned different constituents, depending on the aim: visually, for example, one might want to detail every shot according to its fine-grained elements such as ‘shot size’, ‘camera movement’ and ‘mise-en-scène’, or setting, which in turn requires further distinctions between lightning, sets, characters and their action (e.g. ‘nonverbal activities’), etc. The same applies to the auditory modes: sounds and

music can be described, language and vocal production transcribed. All in all, this part of the analysis depends heavily on the research question that one has set out to analyse.

Our example's inquiry concerns space and spatial representations in film and audio description. In particular, we could be interested in asking *how* space is represented in different modes and would, therefore, want to search for filmic/visual/auditory and linguistic *devices* that are in use in the research material. An example of how to represent different modes in the search of spatial devices is presented in Table 2, and the example is explained in the following paragraph.

**Table 2.** Annotating semiotic resources of ST and TT

Visual mode (image)	Auditory mode (sounds and music)	Auditory mode (character speech)	Auditory mode (AD)
C of typewriter on table in room, TILT-UP at Hi // MC of Tr, in bckgr: open door-another room > Tr looks up and down > sleeve (Hi's) into frame, OUT-FOCUS // MC of Hi fs //	rattling, footsteps,  paper rustling, character speech	(sound persp: close) Hi -> Tr "Nehmen Sie den Stenoblock"	<i>Im Lagerraum nimmt Traudl den Deckel von einer Schreibmaschine.</i>  <i>Traudl sieht ihn (Hi) mit großen Augen an.</i>

In the visual mode, we are describing cinematographic and scenic devices that allow for spatial meaning making, for example, shot sizes (C=close-up, MC=medium close-up), focus (OUT-FOCUS means that the object is out of focus, blurry), camera position ('fs'=from the side), and setting ('in bckgr'=in the background). Here, for example "MC of Hi fs" means that Hitler is filmed from the side in a medium close-up. The marking '//' stands for a shot change. Occasionally, grammatically incorrect expressions (e.g. "typewriter on table", leaving out the definite article) are used to save space in the table.

Auditory resources have been described and situated in relation to the visual narration so that they correspond to each other in the horizontal reading. Important spatial cues in the auditory mode are, for example, whether the sound comes from near or far (here, sound perspective is "close", implying that the character who speaks is near), and verbal references and structure in the linguistic mode. Here, we have found three instances in which space is represented. In the character speech, Hitler's order *Nehmen Sie den Stenoblock* [*take (you, formal) the writing pad*] reflects interactive space: Hitler is giving commands to his secretary, Traudl, and the characters are interpreted as being close to each other because there is no other indication, e.g. speech through telephone or shouting (the proximity to each other is also visible).

The two instances of audio description are full of spatial references and therefore transcribed in their entirety. Besides of lexis (i.a. *Lagerraum* [storeroom], *Deckel von einer Schreibmaschine* [cover of a typewriter]), information structure and syntax are important cues for spatial representation. Here, for instance, the topicalisation *Im Lagerraum nimmt Traudl...* [In the storeroom Traudl takes...] emphasises the physical location *Im Lagerraum* because in a sentence-initial place, a sentence element receives the greatest attention (Kluckhohn 2005). In cognitive linguistics' terms (e.g. Verhagen 2007) this could be interpreted so that the Ground (here the physical space, *Lagerraum*) is more salient than the normally prominent Figure (here the character and its action, *Traudl nimmt*).

In Table 3, elements that communicate space in the source and target text have been highlighted in red. Instead of analysing modal devices, we have now looked for cues that represent space. Next step is to describe the highlighted chunks. Here, the analyst has more information about the visual mode than the table reveals, because not everything has been verbally coded but can, however, be backtracked in the audiovisual material. In fact, we could call the description of the visual mode as 'notes': it does not render the visual narration in its totality but constitutes a body of cues about the visual-dynamic narration in the film, which might need to be revisited while fine-tuning one's analysis. Additionally, as audio description is based on an integral reading of the filmic text, taking into account both visual and auditory narration, we as analysts should consider the auditory meaning making as well (see also p. 7). Here, both the original, visual narration and its translation, audio description, occur in an auditory context that implies activity about furniture (rattling), characters (footsteps and speech) and paper (rustling).

**Table 3.** An intermodal, ST-TT analysis

Visual mode (image)	Auditory mode (sounds and music)	Auditory mode (dialogue)	Auditory mode (AD)
C of <b>typewriter on table in room</b> , TILT-UP at <b>Hi</b> // MC of <b>Tr</b> , in bckgr: open door- <b>another room</b> > <b>Tr looks up and down</b> > <b>sleeve (Hi's)</b> into frame, OUT-FOCUS // MC of <b>Hi</b> fs //	rattling, footsteps,  paper rustling, character speech	(sound persp: close) Hi -> Tr "Nehmen Sie den Stenoblock"	<b>Im Lagerraum nimmt Traudl den Deckel von einer Schreibmaschine.</b>          <b>Traudl sieht ihn (Hi) mit großen Augen an.</b>

Consequently, we may now describe and compare the spatial meaning making in both source and target text as follows (see Table 3). Implicit, interpreted meaning making is put in parentheses. The scene begins with a close-up of a typewriter, whose cover is being lifted up, on a table in a room (the room has to be imagined since only a wall and the table are explicitly shown). A tilt up from the typewriter then reveals Hitler in that room, too. There is rattling of furniture (are the characters moving it?) and (characters') footsteps can be heard. After that, there is a cut, and the next shot shows Traudl from waist up; behind her in the background, an open door that exhibits another room, Traudl looks up and down, paper (the note pad?) is rustling, etc. In this way, we are describing the audiovisual, dynamic narration, which can then be compared to the audio description that says: "In the storeroom Traudl takes off the cover of a typewriter." This is as much as can be described in this particular instance, because Hitler starts to speak (remember that audio description must not drown the dialogue). However, the auditory context provides more cues for interpretation: for instance, the rattling implies that there is furniture in the room (e.g. a table on which the typewriter is located). The auditory narration can also function as a cohesive and referential element for audio description if, for instance, the rattling originates from the typewriter or its cover. Sometimes, auditory narration is self-sufficient: here, Hitler's presence is not described but inferred from hearing him speak to the other character who he was with in another location in the previous scene, and who is now explicitly said to be in the room.

As one last point, one could briefly compare the devices used for communicating space in the beginning of the scene. As already explained (p. 7), the audio description topicalises the location "in the storeroom" and thus draws more attention to it than to other parts of the utterance. However, similar inferences cannot be made from the visual narration based on the cinematographic devices which, instead, seem to emphasise details and leave the room in the background. Visually, the space is revealed through a close-up of the typewriter and a medium close-up of Traudl, who then seems to be in the focus during the whole scene. Hence, what could account for the foregrounding of location in the audio description? One explanation could come from a strategic, deliberate or not, decision making that invites audiodescribers to explicitly indicate a scene change; something that can be immediately interpreted visually by seeing another kind of setting, and needs to be rendered in an equally effective way in the verbal mode (for example, the ITC guidance, 2000, for audio description in Britain suggests to give clear guidance on such occasions).

We have so far intended to illustrate how multimodal research material can be analysed on the basis of a table format and textual representation. The method has its advantages, some of which have already been mentioned, as well as shortcomings that need development. While leaving them for discussion in the final section, we would now like to propose yet another usage for the model.

#### **4 TT analysis**

It is possible to convert the table model into a tool for multimodal TT analysis, which enables both mono- and cross-linguistic studying of target text. That is, we can analyse

one or several translations and their interaction with other modes. In Table 4, audio description is related to the film’s auditory narration, and the visual mode is not included because the aim is to view audio described film as an independent text, a stand-alone product that is, in this case, received through the auditory channel. The table enables us to analyse, among other things, the parallel progression of sounds, character speech and audio description: for instance, as we hear footsteps and rattling (supposedly of furniture), the audio description renders *Im Lagerraum nimmt... Schreibmaschine*, which is followed by rustling of paper and Hitler’s line *Nehmen Sie den Stenoblock* as well as more character speech (in this case, Hitler’s).

**Table 4.** A TT analysis

Visual mode	Auditory mode (sounds and music)	Auditory mode (character speech)	Auditory mode (AD)
-	rattling, footsteps,  paper rustling, character speech	(sound persp: close) Hi -> Tr "Nehmen Sie den Stenoblock"	<i>Im Lagerraum nimmt Traudl den Deckel von einer Schreibmaschine.</i>  <i>Traudl sieht ihn (Hi) mit großen Augen an.</i>

Meanwhile, Table 5 shows how two different translations can be compared on the basis of the same table model.

**Table 5.** A TT analysis with cross-linguistic comparison

Auditory mode (sounds and music)	Auditory mode (character speech)	Auditory mode (AD-DE)	Auditory mode (AD-ES)
rattling, footsteps,  paper rustling, character speech	(sound persp: close) Hi -> Tr "Nehmen Sie den Stenoblock" / "Quiero que tome nota."	<i>Im Lagerraum nimmt Traudl den Deckel von einer Schreibmaschine.</i>  <i>Traudl sieht ihn (Hi) mit großen Augen an.</i>	<i>(Prepara la máquina) de escribir en el despacho.</i>

In Table 5, the column for the visual mode is intentionally left out and replaced by a fourth column for a second, parallel translation, which in our case belongs to the Spanish audio description (AD-ES). Positioning audio descriptions this way facilitates the comparison of the translations. Here, for example, the difference in length becomes apparent only by looking at the textual amount of verbal description in both columns. While the German audio description gives emphasis to the room (see p. 7), then describing Traudl's action in it (Traudl takes off the cover of a typewriter and, in a later instance, looks at Hitler with wide eyes; *mit großen Augen ansehen* is also a figure of speech for 'astonished'), the Spanish audio description tells that Traudl sets the typewriter ready (the chunk *prepara la máquina* in parentheses is spoken in the previous shot) in the office, without topicalising the room. The character speech in German and Spanish can be contrasted as well; here both of them represent the interactive space between the two characters (see p. 7), the Spanish utterance saying: *Quiero que tome nota* [I want you (formal) to take notes].

In this section, our aim has been to demonstrate how the table model created in the ST-TT analysis can be applied to analyse target text as a stand-alone product. It is now time to evaluate the method and suggest improvements for further analyses.

## 5 Evaluation of the method

As by far not the only model for analysis, the article has tried to illustrate one way to "unpack" a complex multimodal text into signifying entities, i.e. communication modes, for the aim of translation analysis. It has sought to present a model for transcribing and describing multimodal material and exemplified its use in translation analysis. This final section discusses questions and shortcomings of the model that both types of analysis have raised relating to general matters such as representation of the material and degree of interpretiveness.

In this model, multimodal material is represented as a table that allows a vertical and horizontal viewing of meaning-making progression. However, the tables presented here are missing a column on the left, which should render the time code and enable the reader to keep track of the time. Furthermore, it should be stressed that while we categorise certain modes and submodes, other authors may make different distinctions, and the level of precision may be different; the categorisation depends on what is studied and sought after. For example, Baldry (2004) includes kinesic action in his table, which in our model could be situated as a subcategory of visual mode/nonverbal activities.

Another significant point to make regarding the representation of the material pertains to the question of authenticity/interpretiveness. When presenting research that has been carried out with this method, one could reproduce the visual mode in the table through still images in order to offer readers a reliable and minimally manipulated access to the source text. Nevertheless, publishing original material is often not possible because of copyright issues, and researchers must confine themselves to a textual representation. However, while representing the material in a textual form, the interpretative nature of

the method becomes more apparent. To render visual material with words already is an interpretation, and a justification for using this method can be found in the same argument: all analysing requires interpretation, and representing visual (or auditory, for that matter) communication in a textual form actually simultaneously exhibits the analysis. A weak point is that it seems practically impossible to render all visual information verbally (for instance in a film, all aspects from shades of light to positions and posture of characters would require an almost interminable description). Therefore, one solution is to selectively describe the aspects that are relevant for each analysis (e.g. Mubenga 2009 presents visual cues textually, rendering merely the information that is needed to support the argument). In the end, in representing audiovisual, dynamic material “on paper”, the observer always comes short of full reproduction, because the visual and auditory narration must be stopped in order to be analysed.

On the whole, multimodal analysis may be applied to any kind of situation in which different modes interact to analyse the integral meaning making and the capacities of distinct modes in the process. Doing multimodal analysis is time-consuming and cognitively demanding because so many aspects have to be taken into account and viewed critically. If the material is modally very rich, it seems feasible to analyse only small parts at a time. Therefore, a successful analysis has a clear focus or phenomenon that it sets out to study in a particular context. We have tried to illustrate this by choosing spatial elements as focus. As much as there is more substance to study on the matter, this article cannot venture much more in this direction and must leave more profound analyses for further research.

Finally, both TT and ST-TT analyses seem to have their advantages and shortcomings regarding the translator’s two objectives: loyalty to the source text’s style and content as well as functionality in the target situation (Nord 1991). While comparing source and target texts in the ST-TT analysis, the researcher can make arguments about loyalty but will have problems with ‘observer’s effects’. For example in the case of audio description, sighted analysts will find it difficult to distance themselves from the visual material, because once they have seen the image, it is impossible to start the linguistic analysis from scratch, that is, without the visual representation in mind. In fact, it seems worthwhile for any kind of translation analysis to begin with TT analysis, because this way the researcher will have a more independent grasp of the translation as a product in itself, hindering the impact of source text to the interpretation of target text. When studying very distinct core modes such as image and language, one will also have to know the capacities of each mode in order to make judgements about equivalence. Meanwhile in the TT analysis, the researcher can make some arguments about functionality but will have problems with taking into account every aspect of a contingent receiver situation. In our case, first of all, the majority of the audience has some sight albeit to distinct degrees, and it seems very challenging to imagine every type of situation in which audio described films are received. Secondly, one may analyse the linguistic-auditory message, and make hypotheses about the interpretations that may be constructed, but, basing on theories, do we really reach the audience’s mind? Here again, we need other kind of research and supporting methodologies to come along and contribute to the analysis.

## Research material

*Der Untergang / El hundimiento*. Dir. Oliver Hirschbiegel. Constantin Film, 2004. [Audiodescribed in German by Bayerischer Rundfunk. DVD. Highlight, 2005; audiodescribed in Spanish by ONCE, 2006]

## Works cited

Baldry, Anthony 2000. English in a Visual Society: Comparative and Historical Dimensions in Multimodality and Multimediality. In Anthony Baldry (ed.)

*Multimodality and Multimediality in the Distance Learning Age*. Campobasso: Palladino Editore, 42–89.

Baldry, Anthony P. 2004. Phase and transition, type and instance: patterns in media text as seen through a multimodal concordancer. In O'Halloran, Kay L. (ed.) *Multimodal Discourse Analysis*. London: Continuum Int., 83–108.

Duarte, Joan F. (ed.) 2006. *Translation Studies at the Interface of Disciplines*. Amsterdam: John Benjamins.

Hirvonen, Maija 2010. "Translating space: inter- and multimodal analyses of spatial representations in filmic audiodescription". Paper presented at *I International Symposium for Young Researchers on Translation*, Barcelona, Spain, 23 June 2010.

ITC Guidance on Standards for Audio Description 2000. [online] Available at: [http://www.ofcom.org.uk/static/archive/itc/itc\\_publications/codes\\_guidance/audio\\_description/index.asp.html](http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/audio_description/index.asp.html) [accessed at 23 August 2010].

Kluckhohn, Kim 2005. Informationsstrukturierung als Kompensationsstrategie – Audiodeskription und Syntax. In Fix, Ulla (ed.) *Hörfilm*. Berlin: Schmidt, 49–66.

Kress, Gunther & Theo van Leeuwen 2006 [1996]. *Reading images*. London/NY: Routledge.

Morgner, Henrike & Steffen Pappert 2005. Darstellung des Untersuchungsmaterials: Sequenzprotokoll, Einstellungsanalyse und Transkription. In Fix, Ulla (ed.) *Hörfilm*. Berlin: Schmidt, 13–32.

Mubenga, Kajingulu Somwe 2009. Towards a Multimodal Pragmatic Analysis of Film Discourse in Audiovisual Translation. *Meta* 54:3, 466–484.

Nord, Christiane 1991. *Textanalyse und Übersetzen*. 2nd ed. Heidelberg: Julius Groos.

Poethe, Hannelore 2005. Audiodeskription: Entstehung einer Textsorte. In Fix, Ulla (ed.) *Hörfilm*. Berlin: Schmidt, 33–48.

Remael, Aline 2001. Some thoughts on the study of multimodal and multimedia translation. In Gambier, Yves (ed.) *(Multi) Media Translation. Concepts, practices and research*. Philadelphia: John Benjamins, 13–22.

Schopp, Jürgen 2008. In Gutenbergs Fußstapfen: Translatio typographica Zum Verhältnis von Typografie und Translation. *Meta* 53:1, 167–183.

Snyder, Joel 2005. Audio Description. The visual made verbal across arts disciplines - across the globe. *Translating Today* 4, 15–17.

Stöckl, Hartmut 2004. In between modes – Language and image in printed media. In Ventola, Eija, Cassidy Charles & Martin Kaltenbacher (eds) *Perspectives on Multimodality*. Amsterdam: John Benjamins, 9–30.

Taylor, Christopher J. 2003. Multimodal Transcription in the Analysis, Translation and Subtitling of Italian Films. *The Translator*. 9:2, 191–205.

Ventola, Eija, Cassidy Charles & Martin Kaltenbacher 2004. Introduction. In Ventola, Eija, Cassidy Charles & Martin Kaltenbacher (eds) *Perspectives on Multimodality*. Amsterdam: John Benjamins, 1–6.

Verhagen, Ariel 2007. Construal and perspectivization. In Geeraerts, Dirk & Hubert Cuyckens (eds) *Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 48–81.