

Maija Hirvonen

**Sampling Similarity in Image and Language –
Figure and Ground in the Analysis of Filmic Audio
Description**

Abstract

Audio description can be defined as intermodal translation in which the visual representation (for example, of a film) is verbalised and spoken in order to facilitate and enhance reception by visually impaired audiences. By its very essence, audio description requires analysing the relation of language to non-linguistic, visual representation. The theory of Figure and Ground segregation has been developed for both visual perception and language to explain how we perceive “thing-like” figures and “substance-like” grounds in space. This segregation is reflected in language by coding certain elements as figures in reference to a more (static) ground. This paper addresses the Figure and Ground theory both in visual representation and in its linguistic translation. On the basis of theory-led sample analyses on a contemporary film and its different-language audio descriptions, this study presents evidence that the verbal representation can parallel the visual segregation of Figure and Ground. Furthermore, it discusses the application of the theoretical Figure and Ground characteristics and suggests some clarification to them.

1. Introduction¹

Audio description (AD) can be defined as a type of intermodal translation that substitutes for visual perception and enhances it by verbal, spoken descriptions (for example, see Cámara & Espasa 2011: 415; Hirvonen 2012: 21–22). For the blind and for others with a severe loss of sight, AD is a capacitating aid that renders the visual world accessible; for people with milder degrees of low vision, it supports visual perception. As AD aims at verbalising a range of visual and, occasionally, auditory phenomena, it can be applied in a variety of situations, such as film, theatre, television as well as art and museum exhibitions.² However, this verbalisation is conditioned by contextual and modal factors (see Hirvonen 2012: 23). In a film, both the dialogue and important sound effects restrict the time available for AD. The soundtrack itself must be taken into account in the verbalisation because sounds may also require a verbal description. Finally, the change from the visual to the linguistic mode means, for instance, that an iconic, naturalistic form of representation is conceptualised and abstracted. (Ibid.)

Regardless of the differences, both visual and verbal representations are presupposed to be perceived in terms of Figure and Ground segregation. F/G segregation originates from Gestalt psychology and explains how we organise space to accommodate figure/s and a ground (for example, see Koffka 1936). Figure is described as being smaller and perceptually more salient than Ground, which is used to define Figure. Furthermore, Ground is larger and less defined than Figure. This theory is also applied to explain the perception of film images (Bordwell 1985) and film sound (ibid.; Branigan 2010). In language, Figure and Ground have two different aspects. Figure may be understood as the extra-linguistic object and Ground as the extra-linguistic terrain of reference, or they can be

¹ Several people have contributed to this study. The idea of studying the variation between Figure and Ground first occurred to me while Paula Igareda and myself were analysing this data for other purposes. I am indebted to her for her help in data collection and transcription. I also extend my thanks to Lee Bye and Martina Wiemers for providing the English and German AD scripts for research purposes. Discussions within the Langram community of the Langnet doctoral programme have been illuminating. In particular, I would like to thank Jukka Mäkisalo and Liisa Tiittula for their support and feedback. Paula Igareda and Bernd Benecke have assisted me with the English translations. Finally, the three anonymous reviewers as well as the language reviser have greatly contributed to improve this paper.

² Examples of good introductions to AD are the edited books by Díaz Cintas, Orero and Remael (2007) and by Fix (2005).

understood in terms of foregrounded information (Figure) and backgrounded information (Ground) (Engberg-Pedersen 2011).

As AD involves translation from images to language, it provides data to compare F/G segregation intermodally. Moreover, AD provides a new context and research interest for the study of cognitive phenomena – in particular those that can be triggered and accessed both visually and verbally. Vandaele (2012: 96–97) maintains accordingly that the descriptive parameters developed within the framework of Cognitive Linguistics, such as the “figure-ground alignment”, can be used to describe the “mental imagery produced by narrative texts” in general and by AD in particular. The question therefore is whether the verbal description in AD renders a similar idea of spatial organisation as the film image.

The present article is a methodological study that applies the theories of F/G segregation to compare visual and verbal representation in both a film and its different-language audio descriptions. This analysis has two main objectives: The first is to test the theories on the analysis of film imagery and AD. The second is to compare the F/G segregation of the visual representation to its verbal translations in different languages. This orientation to research can lead to detecting interesting differences and parities between the visual and verbal representations concerning Figure and Ground. Furthermore, this study tests the explanatory power of the theories of F/G segregation and suggests a way to apply them. The data are from a mainstream feature film *Slumdog Millionaire* (Boyle & Tandan 2008) and from the audio descriptions of this film in three languages: German, English, and Spanish.

This article is structured as follows. After the theories of F/G segregation are surveyed in Section 2, these theories are applied to the analysis of film and AD in two sample cases in Section 3. The results of the analysis are summarised in Section 4, and the fifth and final section presents the conclusions of this study.

2. The Figure and Ground theories

In this section, I will outline the main ideas of F/G segregation in the cognitively oriented theory of psychology, film and language.

2.1 Figure and Ground in the visual perception and representation

Perception can be defined as a conscious awareness of something, be it thoughts or feelings or environment (Hatfield 2001). According to the cognitive theory of visual perception, a basic process in the visual perception of space is F/G segregation. More specifically, visual perception begins by identifying textures and objects in space, and the next stage involves discerning forms and grouping objects (Evans 2010: 29–31). It is at this point that the principle of Figure and Ground segregation becomes useful. As Evans (2010: 31) observes, this relates to the fact that

a fundamental way in which we segregate entities in our environment, thereby perceiving distinct objects and surfaces, comes from the our [sic] ability to perceive certain aspects of any given spatial scene as ‘standing out’ from other parts of the scene.

Even if F/G segregation seems to be an innate human ability, it occurs individually. In other words, each mind organises its visual environment potentially in different terms; hence, the optical illusion known as ‘Rubin’s vase’³ can be perceived differently depending on whether we perceive the faces or the vase as Figure (ibid.).

The Gestalt theory defines the aspects or perceptual differences that define the segregation of a visual scene into the categories of Figure and Ground. According to Evans (2010), this theory proposes the characteristics that are listed in Table 1.

Table 1. Figure and Ground characteristics in visual perception (Evans 2010: 32)

Figure	Ground
Appears to be thing-like	Appears to be substance-like
A contour appears at edge of figure’s shape	Relatively formless
Appears closer to the viewer, and in front of the ground	Appears further away and extends behind the figure
Appears more dominant	Less dominant
Better remembered	Less well remembered
More associations with meaningful shapes	Suggests fewer associations with meaningful shapes

³ For example, see Goldstein (2010/2007: 108) for a reproduction of Rubin’s vase.

Since Evans (2010) provides a summary and disregards more detailed explanations of the attributes of Figure and Ground, resorting to an original source of Gestalt psychology can be useful. Most of these characteristics are found in *Principles of Gestalt Psychology* by Kurt Koffka (1936). In Koffka (ibid.), a central feature is “duo formation”, which is described in the table above as the near-distance relation. Figure appears to be in front of Ground, which extends behind Figure (Koffka 1936: 178f.). Furthermore, the *thing-ness* of Figure is also asserted by the properties of solidness and shape, while Ground is “stuff”, loose and unshaped (ibid. 187; see also Köhler 1947). If we are more concerned with Figure than with Ground, as Koffka suggests (“where the interest lies, a figure is likely to arise”, ibid. 186), this may explain why Figure is better remembered and more easily attributed meaning. Concern can refer to memory, so that because some object is more easily remembered, such as the vase in the Rubin’s vase illusion, that object may be interpreted more readily as Figure (Goldstein 2010/2007: 108).

In the everyday scenes we perceive, what then may be conceived of as Figure and Ground? In a landscape such as a street, the sky is Ground while the houses, constituting a shape that stands in contrast to the sky, are Figures (Koffka 1936: 209; Köhler 1947: 186–187, 202). Similarly, for example, a pencil on a desk would appear as a well-marked part, as Figure, while “the desk appears as a relatively formless, featureless mass”, that is, Ground (Ehrenstein 2001: 11229). Moreover, Ehrenstein argues that Ground is not necessarily behind Figure: “For example, in looking through a window at a tree, the window screen appears as ground, but is clearly seen in front of the figure, the tree”. In addition, the F/G segregation of the visual field is a dynamic event rather than one that is static. The “multivalence of the stimulus field” means that objects and surfaces are definable as Figure or Ground depending on where one’s attention is directed. (Ibid.)

Bordwell (1985) adapts the F/G segregation to cinematic audiovisual representation. A central idea of this cognitively oriented theory of film narration is that spectators construct the story space and its components – “figures, objects, and fields” – on the basis of visual and auditory narrative cues (ibid. 113). Consequently, several visual cues in the shot space – the scenographic space delineated by the four frames of the camera – engage

spectators in F/G segregation. This account by Bordwell can be related to the Gestalt characteristics in the following ways:⁴

- *A contour appears at edge of figure's shape; Appears closer to the viewer (F) / Appears further away (G):* According to Bordwell (1985: 113), overlapping contours differentiate Figure(s) from Ground. This means that when one contour occludes another, we attribute the occluding edge to a near object (Figure) and the other edge to a distant one (another Figure, or the Ground) (ibid.). It is also possible to have more than one Figure on a scene (see Ehrenstein (2001) on the dynamicity of F/G segregation). With respect to near-distance relations, films are capable of furnishing various depth cues. Lighter, warmer, and intense colours seem closer than darker, cooler ones. Furthermore, the knowledge of perspective, that is, how straight lines behave in depth, helps to organise elements in space. Rougher and denser textures also stand out, whereas smoother and less dense textures recede. Bordwell summarises this as follows: "The more indistinct the surface, shape, color, or mass of an object is, the more distant we assume that object to be". (Bordwell 1985: 114.)
- *More associations with meaningful shapes / Better remembered / Appears more dominant:* The familiar size of objects, such as people, helps "decide what is nearer or farther away" (Bordwell 1985: 114). Furthermore, illumination suggests shapes and areas by highlighting and shadowing. For instance, backlight reinforces the Figure and Ground differences by suggesting planes. Some elements therefore seem to have a clearer shape (Figure), while others are more amorphous (Ground). By guiding our eyes to certain parts of space, light can render some aspects more dominant (Figure), whereas shadow obscures others (Ground). (Ibid.; see also Bordwell & Thompson 1990: 134.)

Another characteristic of Figure that is central in cinema, being *movies*, is movement. This is one of cinema's most important cues for object identification and spatial relations, creating a continuous flow of overlapping contours and "strengthening figure/ground hypotheses"

⁴ Kress & van Leeuwen (2006/1996) also describe how visual properties, such as placement in the foreground or background, sharpness and light, affect the "reading" of the (film) image.

(Bordwell 1985: 114). Yet another crucial factor when discussing contemporary films is sound; it segregates to Figure and Ground as well (Koffka 1936: 201; Bordwell 1985: 118–119). For instance, silence can be Ground, although it could be the opposite in a city, Figure (Koffka *ibid.* 201). In the sonic space, high-pitched tones tend to emerge as Figure from the lower Ground tones (Bordwell *ibid.*). Apart from the visual representation and in coordination with it, films build “on the relationship of sounds to one another – sonic figure and ground – [and] on the fluid relationship of sounds to an image” (Branigan 2010: 55).

While some characterisations of Figure and Ground that are proposed by the Gestalt theory and by its application to film are intuitively understood (‘thing/substance’, ‘shape/non-shape’, and ‘closer/more distant’), other aspects remain somewhat ambiguous. For instance, should we understand ‘dominant’ in terms of size, amount, intensity, or some other property? One answer from the filmic representation is that in terms of intensity, light and colour can be connected to dominance. Regarding the characteristics ‘better/less well remembered’ and ‘more/less associations with meaningful shapes’, familiarity seems to be an important aspect of Figure-ness, strengthening meaningfulness and recall.

2.2 Figure and Ground in the linguistic representation

In the linguistic mode, F/G segregation generally has two different meanings. Figure may be understood either as the extra-linguistic object that is referred to by the linguistic expression, or as the knowledge or information that is foregrounded. Similarly, Ground not only refers to the extra-linguistic terrain that is referred to, but may be understood in terms of knowledge or information that is backgrounded. Engberg-Pedersen (2011: 693) distinguishes three different usages of Figure/foregrounding:

1. “The centre of attention as a result of the context, which influences the choice of subject, e.g., *The bike in The bike is in front of the house.*” The prominent entity in the sentence is Figure.
2. “The centre of attention coded in the sentence as the asserted part, i.e., *is in front of the house.*” This suggests that the focus of the sentence is Figure.
3. “The centre of attention that the sentence brings about in our understanding of the represented situation, i.e., the view of the situation that is encoded in the sentence and that makes us

conceptualise the scene with the bike as the figure and the house as the ground in the Gestalt-psychological sense.” This points to the extralinguistic reference entity as Figure (and Ground).

In the present study, Figure and Ground in the linguistic representation are used in the meaning of extra-linguistic figures and grounds, reserving other notions, such as foregrounding and backgrounding, for the pragmatic domain of language.

In the cognitive linguistic framework, F/G segregation is considered to be a linguistic-conceptual phenomenon and is termed ‘figure-ground alignment/assignment’ (Langacker 1987; Talmy 2000). One instantiation of figure/ground alignment is the trajectory/landmark asymmetry in which elements are predicated in relation to each other so that a trajectory (figure) is “tracked” against the background of other elements (Langacker *ibid.* 231–232). This study adopts the account by Talmy of extra-linguistic objects and terrains of reference. According to this theory, Figure is “a moving or conceptually movable entity whose path, site, or orientation is conceived as a variable” and which therefore “needs anchoring”, whereas Ground is “a reference entity, one that has a stationary setting relative to a reference frame, with respect to which the Figure’s path, site, or orientation is characterized”; Ground “does the anchoring” (Talmy 2000: 312). Figure and Ground therefore refer to the extra-linguistic concept or referent as well as to its linguistic realisation. The same conclusion is made by Engberg-Pedersen (2011: 693): “Talmy here [in 2000, 2007] uses *Figure* and *Ground* both of the linguistic entities, i.e., the nominals and clause constituents, and of the referents in a described situation”.

A frequently cited example of F/G assignment in language is:

The bike (F) is near the house (G).
 The house (F) is near the bike (G).
 (Talmy 2000: 314.)

The first sentence specifies *the bike* as Figure, as a conceptually movable entity whose site is described with reference to *the house*, which therefore is the reference entity. The second sentence assigns *the house* as Figure and *the bike* as Ground; a situation that, as Talmy notes, does not “conform with the exigencies of the familiar world” because it is less familiar to conceive of ‘house’ as variable point and of ‘bike’ as its reference point. (Ibid.)

Talmy (2000: 315–316) lists a set of characteristics that define Figure and Ground. These are presented in Table 2.

Table 2. Figure and Ground characteristics according to Talmy (2000: 315–316)

	Figure	Ground
<i>Definitional characteristics</i>	Has unknown spatial (or temporal) properties to be determined	Acts as a reference entity, having known properties that can characterize the Figure's unknowns
<i>Associated characteristics</i>	more movable	more permanently located
	smaller	larger
	geometrically simpler (often pointlike) in its treatment	geometrically more complex in its treatment
	more recently on the scene/awareness	more familiar/expected
	of greater concern/relevance	of lesser concern/relevance
	less immediately perceivable	more immediately perceivable
	more salient, once perceived	more backgrounded, once Figure is perceived
	more dependent	more independent

The presentation by Talmy (2000: 315–316) of the Figure and Ground characteristics evokes similar questions as those that arose in the characteristics presented by Evans (2010). Talmy's formulation lacks illustration and explanation of some of the features. Certain characteristics even seem controversial and leave open questions. What exactly does 'perception' in "less/more immediately perceivable" refer to, and is the Ground feature of "more immediately perceivable" not in contradiction to the idea that Figure draws attention more easily and is, so being, more immediately perceivable?

Again, additional illustration of the characteristics can be detected in a field that applies F/G segregation to narration: cognitive poetics. Cognitive poetics draws from the cognitive linguistic tradition and considers F/G segregation to be a basic part of a narrative analysis (Stockwell 2002: 15). Since AD has traits of narrativity (Kruger 2010), cognitive poetics can be a useful tool for the analysis of F/G segregation in the audio descriptions. Indeed, some of the Figure characteristics proposed by Talmy (2000; see Table 2) find an equivalent in those suggested by Stockwell (2002: 15):

- *More movable*: Figure will “be moving in relation to the static ground” (Stockwell 2002: 15).
- *Of greater concern/relevance*: Figure will “be more detailed, better focused, brighter, or more attractive than the rest of the field” (Stockwell 2002: 15), if concern and relevance are defined in terms of attractiveness and focus of attention.

The remaining Figure characteristics in Stockwell (2002: 15) are comparable in distinct degrees to Talmy (2000) and to the Gestalt theory. For instance, in Stockwell’s terms, Figure will “be regarded as a self-contained object or feature in its own right, with well-defined edges separating it from the ground”, which seems to conform to two features from the Gestalt framework, namely “appears to be thing-like” and “a contour appears at edge of figure’s shape” (Evans 2010: 31–32). However, one contradictory feature is when Figure will “be on top of, or in front of, or above, or larger than the rest of the field that is then the ground” (Stockwell 2002: 15). Talmy (2000: 315–316), in contrast, assigns Groundness to a *larger* element. Another Figure feature from Stockwell, “be a part of the ground that has broken away, or emerges to become the figure”, is interesting because it seems to hint at the dynamic relations of Figure and Ground (Ehrenstein 2001), or that parts of Ground can become Figure. Cognitive poetics also links Figure and Ground to concrete narrative entities: characters are Figures and settings Grounds. For instance, characters “have boundaries summarized by their proper names” and “are likely to be the focus of the narrative”; they also move through different settings, that is, across Ground, and evolve psychological traits and perform wilful action (as opposed to attributive or existential action used to describe Ground). The tendency of focusing on characters appears to be due to our interest in tracking their experience in the story. (Stockwell 2002: 15–16.)

3. Testing the theories: Analysing the Figure and Ground in *Slumdog Millionaire* and in the audio descriptions in English, German and Spanish

This section explains the test analysis and presents the two sample cases. The main focus of the analysis is to discern whether the language of the audio descriptions and the extra-linguistic, visual mode of the film are similar in terms of F/G segregation and how the F/G theories may be

utilized as methodological tools. To address this aim, the analysis is a twofold process:

1. The theory of F/G segregation, developed by the Gestalt psychology, as well as its application in the cognitive approach to film (Bordwell 1985) are used in the analysis of the visual filmic representation. How this framework lends itself to the analysis of visual scenes is tested on two sequences of *Slumdog Millionaire* (Boyle & Tandan 2008).
2. The cognitive linguistic theory of F/G assignment by Talmy (2000) is adopted in the analysis of the linguistic representation in AD, and it is supplemented by insights from the cognitive poetics presented in Stockwell (2002). The suitability of the framework in the analysis of language is tested on three audio descriptions of the film sequences, including a UK-English version, a standard-German version and a Peninsular-Spanish version.

The film *Slumdog Millionaire* recounts the story of a boy, Jamal, who lives a difficult childhood with his brother Salim in the slums of Mumbai but then becomes a millionaire on the television show entitled “Who wants to be a millionaire?” and succeeds in rejoining his childhood friend and loved-one, Latika. This film has been audio described on DVD in three languages: English (UK), German (Germany) and Spanish (Spain). In the present study, two sequences from the film have been selected for analysis because they contain two different cases of F/G segregation. In Case 1, an element that is Figure in the first shot becomes Ground in the next one; in Case 2, something that first serves as Ground later becomes Figure (see Herman 1996: 563). These shifts illustrate a familiar situation in (film) narratives: the story action moves from a primary setting (for example, a street) to a secondary location, which itself is located in that setting (for example, a car on the street) (see Schubert 2009: 63).

In order to visualise the shot space, I provide black-and-white drawings of the film shots (see Shot protocols 1 and 2).⁵ Above the drawings, a text in COURIER CAPITALS describes the soundscape (the sound effects, dialogue, and music) in each shot. The plus symbol ‘+’ refers to a new sound, and the arrow symbol ‘→’ indicates the continuity of a sound between shots. The time code indicating the beginning and the end of the

⁵ The drawings are by Eero Tiittula.

sequence adheres to the original version of the film. In the linguistic analysis, the audio descriptions of the sequences are aligned in a table (see Transcriptions 1 and 2). The English, German, and Spanish versions are arranged from left to right, labelled as AD-EN, AD-DE, and AD-ES, respectively. The transcriptions are divided into cells to resemble the way in which they are heard during the film shots. Above the transcriptions, the comments on the soundscape occur in COURIER CAPITALS. The translations into English from the German and Spanish audio descriptions are provided in *italics* and are employed in the body text with single quotes unless the analysis requires the use of the original language. The passages that are the focus of interest in the transcriptions appear in bold font.

Although the filmic soundscape provides important cues for the narrative and also segregates to Figure and Ground, this soundscape is not analysed in depth in this study and it is beyond the scope of the present article. It should be mentioned, however, that the different AD versions of the film allow for distinct perceptions of the original sound. Firstly, the number and length of the descriptions vary (compare, for instance, the English and the Spanish version in Case 2/Shot 1), and, secondly, the volume of the describer's voice can be louder than the soundscape and prevent some of the softer film sounds from being heard. For instance, it may be difficult to distinguish the whirring sound in the background of the Spanish AD in Case 2/Shot 2, as one concentrates on the verbal description. During pauses in the AD, however, sounds stand out (see Hirvonen & Tiittula 2012: 393–394).

3.1 Sample case 1: Figure becomes Ground

In this case, a visual element that is Figure in the first shot becomes the Ground of the character action in the following shot. The sequence narrates an event in Jamal's childhood. Jamal, Salim and Latika are being transported from the poor conditions they have lived in, collecting waste in a rubbish dump, to a more prosperous life in an orphanage.

Shot protocol 1. “The bus sequence” (00:23:53–00:24:15 / *Slumdog Millionaire*)

BIRDS SINGING	→ BIRDS	→ A DISTANT DRONE
+ A DISTANT DRONE OF AN ENGINE	→ DRONE	+ KIDS CHEERING
	+ WHOOPS OF JOY	
	+ A WHIRR	



Shot 1

Shot 2

Shot 3

The first shot depicts a landscape of a forest and buildings from a distance. A road traverses the forest, and on the road is a small (yellow-coloured) object moving along it. It is a minibus (this is represented in the drawing by the small, rectangle-like figure in the middle). We may identify the bus as the same one that transported Jamal, Salim, Latika and other children from the rubbish dump in the previous sequence. In contrast, this landscape has not appeared previously in the film.

Moving in the landscape, the bus “appears to be thing-like” and has “contours” that form its square “shape”, which means that it can be “associated with a meaningful shape” (see Evans 2010: 32). This shape moves – in fact it is the only thing that seems to be mobile – and therefore attracts attention. As Bordwell (1985) observes, movement is a strong cue for Figure-ness. Movement reinforces the association with a meaningful shape since the nature of buses is that they move. Due to its light, yellow colour, the bus stands out from the landscape (see *ibid.*). Although the entire scene appears far away from our vantage point (the camera’s standpoint), the bus does seem to hold the Figure feature “in front of the ground” (Evans 2010: 32) because, as described above, it stands out from the scenery due to its physical qualities. Narratively, too, the bus receives Figure features. For example, based on the previous narration, it is “better remembered” (*ibid.*) than the landscape, which is a new element. When recognised as a bus – and moreover, *the* bus from the previous scene – the element becomes familiar and the focus of narrative attention, and it consequently receives a stronger Figure-ness (see Stockwell 2002). Being Figure, the bus also ought to appear “more dominant” than the landscape (Ground) (see Evans 2010: 32). Yet in this shot, the landscape-Ground is more dominant in terms of size or surface as it fills the image. On the other hand, the light colour contrast against the darker environment serves to

highlight the bus and thus renders it more dominant in terms of attention. Dominance may also be defined in terms of movement (Figure in a *movie* is mobile relative to the static Ground) or narrative weight (movies tell stories about people and their action). Moreover, the audible droning sound of an engine confirms acoustically the visible movement (see Fryer 2010: 207), thereby foregrounding the bus in the scene and strengthening its Figure-ness.

Shot 2 no longer depicts the bus as a “thing” in its whole but rather as “substance”, as horizontal and vertical structures in the background (which, in the present context, can be identified as walls and windows of a bus). The characters’ faces and upper bodies now fill the frame, and their action, facial expressions and body movement are in the foreground and attract attention. The backlight silhouettes the characters and reinforces their shape. At the same time, the space that is visible from the bus windows can also be regarded as Ground because it is an indistinct bright area (although some objects in it are recognisable later in the shot). F/G segregation thus seems to have a proportional hierarchy. In other words, in relation to the characters, the bus is Ground, but in relation to the bus, the outside space is Ground. This confirms the dynamicity of F/G segregation that is noted by Ehrenstein (2001). In addition, based on the previous narration and recognising familiar characters, we infer that the vantage point is now the interior of the bus. The visual closeness correlates with the soundscape as the droning of the engine has grown louder (see Fahlenbrach 2008: 96). The change in volume represents the perspectival change realistically and the continuation of the sound confirms that the bus is (still) moving and that the location of action has not changed (see Schubert 2009: 120). The Figure-ness of the children in Shot 2 is also enhanced by the point-like whoops of joy that poke out of the soundscape.

The third shot reiterates the Figure function for the bus, but the bus is one of several Figures (playing children) against Ground (a courtyard). The moving bus appears in the upper-right corner of the image, but is less distinguishable than in Shot 1 and less audible than in Shot 2. Accordingly, the scene entails various Figures and all are moving and may also attract narrative attention. For example, the bus appears as an old, familiar element, and the children as new, potentially relevant narrative entities.

Let us now turn to examine the three audio descriptions of this sequence.

Transcription 1. “The bus sequence”

	AD-EN (00:23:53)	AD-DE (00:23:29)	AD-ES (00:23:53)
Shot 1	BIRDS SINGING AND CHIRPING + A DISTANT DRONE OF AN ENGINE		
	(The boys grin and gulp down the drinks.) The minibus is driving through lush countryside.	Der Kleinbus tuckert eine Straße entlang, vorbei an grünen Bäumen und weiten Feldern. <i>The minibus is chugging along a street, passing green trees and extensive fields.</i>	Más tarde, los tres niños han montado en la furgoneta amarilla de los desconocidos. <i>Later, the three children have got on the yellow minibus of the strangers.</i>
Shot 2	→ BIRDS → DRONE + WHOOPS OF DELIGHT + A WHIRR		
	The bus is full of scruffy street kids, gazing out of the windows. The bus arrives at a large dilapidated residence...	Im Bus sitzen Jamal und Salim zwischen anderen Kindern. <i>In the bus, Jamal and Salim are sitting among other children.</i> Neugierig sehen sie aus dem Fenster. <i>With curiosity, they look out the window.</i>	Están sorprendidos y confiados ante la generosidad de los hombres. <i>(they) Are surprised and trustful due to the generosity by the men.</i>
Shot 3	→ A DISTANT DRONE + KIDS CHEERING		
	... where numerous children of all ages run around playing in the yard.	Eine Lichtung mit einem Gebäude, dem Waisenhaus. <i>A clearing with a building, the orphanage.</i> Kinder laufen umher. <i>Kids run around.</i>	La furgoneta llega a un poblado lleno de niños que juegan alegremente. <i>The minibus arrives at a settlement full of children who are playing joyfully.</i>

In the first description, *The minibus is driving through lush countryside* (AD-EN) and ‘**The minibus** is chugging along a street, past green trees and extensive fields’ (AD-DE), both the English and German audio descriptions treat BUS⁶ as Figure, that is, it is a moving entity (‘the minibus is chugging’; *the minibus is driving*) whose path (‘chugging along [...] passing...’; *driving through*) is a variable with reference to an entity that has a stationary setting, i.e. Ground (‘trees [...] fields’; *countryside*). (See Talmy 2000: 312.) The Spanish audio description, ‘Later, the three children have got **on the yellow minibus of the strangers**’, deviates from the two other AD versions. The Spanish description assigns BUS a Ground function by using a locative prepositional phrase (‘on the yellow minibus’), which serves as a reference entity for the action of the character-Figures (‘the three children have got on’), and thus anticipates the spatial composition of

⁶ Words that are written in capital letters refer to extra-linguistic referents.

the second shot, in which BUS is Ground. Moreover, ‘later’ marks an explicit temporal transition to a new scene (see Hirvonen 2012: 35).

In the second description, the English and German audio descriptions converge once again: *The bus is full of scruffy street kids, gazing out of the windows* (AD-EN) and ‘**In the bus**, Jamal and Salim are sitting among other children’ (AD-DE).⁷ By making the bus the first element, they continue with the familiar theme from the previous description (*the minibus* → ‘in the bus’/*the bus*). Even though BUS is Ground in both, it assumes different syntactic roles. In the AD-EN, it takes the subject role (*the bus is* [...]), and it is the head of a locative PP in the AD-DE (*im Bus*). The Spanish AD diverges again by describing the characters: ‘(they) Are surprised and trustful due to the generosity by the men’. According to Talmy (2000), this expression is a meta-Figure in that it describes a state of affairs or a property (ibid. 330–332): “Figure and Ground are the same objects (i.e., the Figure constitutes its own Ground)” in a self-referencing event of motion or stationariness (for example, ‘the balloon is round’). With a self-reference, the action in the story seems to halt as a state of affairs or as a property that is focused on (see Chatman 1978: 74), and the spatial attention is narrowed down to Figure (see ibid. 102 and Hirvonen & Tiittula 2012: 389). Yet the Ground function of BUS persists implicitly due to the continuity of the droning sound and to the *Prinzip der Raumkonstanz*: the location remains the same if no change is indicated (Schubert 2009: 119). Otherwise, the rise in volume implies that the vantage point is now closer (see Hirvonen & Tiittula 2012: 419). The audible whoops give voice to the characters in the scene, and the audio descriptions assign a Figure function to them.

With reference to the different linguistic representations of BUS in the AD-EN and AD-DE, Talmy (2000: 333) offers a similar example: “Smoke (F) slowly filled the room (G).”/“The room (G) slowly filled with smoke (F).” Talmy argues that the F/G assignment is retained even though the grammatical relations change because the distinction of the variable-point versus the reference-point persists. In the AD-EN, the subject in the utterance *the bus is full of scruffy street kids* functions as an anchor that determines the site of the *scruffy street kids*, whereas in the AD-DE, the locative PP (‘in the bus’) serves explicitly as a reference entity for the characters’ site (see Talmy 2000: 333). The difference arises from the

⁷ The sentences illustrate how the extra-linguistic Ground assignment (the bus/‘in the bus’) disagrees with the pragmatic Ground assignment (the bus/‘in the bus’ as the focal entity of the sentence) (see Engberg-Pedersen 2011: 693).

vantage point, that is, “where one places one’s mental eyes to look out over the rest of the scene in reference” (ibid.). While with the AD-DE solution it feels as if one is inside the bus, the AD-EN description has a more outsider aspect to it and seems to infer that the bus is visualised as a whole entity. Another difference can be detected in the character reference. While the AD-DE recognises the two characters as Jamal and Salim, thus enhancing their grade of familiarity, the AD-EN simply says *street kids*. Indeed, many Figure features comply with KIDS. For example, they are “conceptually movable” and “smaller” than BUS, and as characters are typically the focus of a narrative, they are also “of greater concern/relevance” (see Talmy 2000: 314–316; Stockwell 2002).

Moving on to the descriptions *gazing out of the windows* (AD-EN) and ‘With curiosity, they look **out of the window**’ (AD-DE), the English and German audio descriptions imply schematic coherence with regard to BUS by referring to a constituent part of buses, *window/s* (see Schubert 2009: 150–152; Hirvonen & Tiittula 2012: 404).⁸ The definite article *the* implies that the reference entity for the act of looking is (still) BUS, which receives a Ground function. Further cues for treating BUS as Ground are the locative adverb *out* (AD-EN) and the locative preposition *aus* (AD-DE); they encode the referent as a region (see Schubert 2009: 172). In contrast, the AD-ES offers the following description: ‘**The minibus** arrives at a settlement full of children who are playing joyfully’, orienting to ‘the minibus’ as Figure against ‘a settlement’ as Ground. The same occurs in the next description of the English AD: *The bus arrives at a large dilapidated residence where [...]*. Hence, the AD-EN and the AD-ES redefine the F/G assignment: ‘the minibus’/bus is now the moving, thing-like element – Figure – that *arrives* at a place, ‘a village full of [...]/a large dilapidated residence where [...], functioning as Ground.

3.2 Sample case 2: Ground becomes Figure

Case 2 is the beginning of another sequence from Jamal’s childhood in which a famous Indian actor visits Jamal’s slum in a helicopter. The first shots in the sequence show Jamal relieving himself in an outhouse, a wooden shack. The primary interest lies in the shack element that has different functions in terms of Figure and Ground. Case 2 presents a

⁸ Schematic coherence depends on the recipient recognising that windows are constituent parts of buses (see Schubert 2009: 154).

reverse situation of Case 1. This means that the place of action (Ground) of one shot becomes a “thing” (Figure) in the next. However, as we will see in the analysis, it is debatable as to what extent this thing is defined either as Figure or as an element of Ground.

Shot protocol 2. “The toilet sequence” (00:10:10–00:10:19 / *Slumdog Millionaire*)

A BUZZ OF A FLY	DISTANT BARKING OF A DOG	→ HELICOPTER
+ CLANKS	+ A HELICOPTER BUZZES	→ WHIRR
+ A SILENT HUM	+ A DISTANT WHIRR	+ DISTANT SINGING OF BIRDS
+ A DISTANT WHIRR		



Shot 1

Shot 2

Shot 3

Shot 1 begins by framing a metal bucket that is being lifted from the floor (this is not depicted in the drawing). The camera then tilts up to reveal the face of the lifter (as the drawing shows), who we recognise as Jamal. He is squatting in a narrow space surrounded by what seems to be timber walls. Yet the character in the foreground attracts attention immediately. A light entering from above highlights his upper body and face and defines his contours, so that we recognise not only a human shape, but the character himself. These properties attribute Figure features to the character. He also moves – his arms lift the metal bucket and his facial expression alters – and this movement is a further and significant cue for Figure. Conversely, the timber walls in the background have several Ground features. For instance, they remain static and appear more substance-like, having some form (resembling timber). The walls also extend behind the character who covers most of the frame and is therefore more dominant in size than what is visible from the timber. The walls seem more distant and in the background due to the darker colour and the brightly lit character in front. This character is likely to be better remembered at this point because he has featured in previous scenes, whereas the timber walls are seen for the first time. In short, the character is more familiar, propels the action and might therefore attract more attention than the wall; these are, again, Figure features.

The second shot depicts another character, Jamal’s brother Salim, seated on a chair outside the timber wall of a construction. Against the sky,

this construction is thing-like and has a shape, and, due to the timber wall, it can also be recognised from the previous shot. Hence, Ground in the first shot (timber walls) becomes Figure in the second shot (shack). However, the character and the chair also appear thing-like and their contours and shapes are well defined against the sunlight, so that in relation to the character, the Figure function of the shack may be questioned. The character appears better defined, is in front of the shack and mobile, and may be more readily associated with a meaningful shape (a human) than the shack. Moreover, the character's Figure-ness is enhanced by its function to propel action in a mainstream narrative.

Moving on to the third shot, the vantage point becomes significantly more distant than in the two previous shots: In the sequence, the view departs from the interior of the hut (Shot 1), shifts to its immediate exteriors (Shot 2) and shows the surrounding environment from the perspective of a bird's eye view (Shot 3).⁹ In this third shot, the dynamic nature of Figure and Ground prevails: though the shack now features more clearly as a "thing" in the landscape-Ground (there are three of them), it is also part of that landscape – is it therefore part of Ground rather than Figure? According to Stockwell (2002: 15), Figure can also be "part of the ground that has broken away, or emerges to become the figure". For example, the movement of a few characters in the scene is observable by their walking along the path in front of the shacks, whereas the shacks are stationary. Conceptually, the shacks, too, are mobile: their location could be changed. Other Figure characteristics apply as well. As mentioned above, the shacks are thing-like. They are also situated in front of the background consisting of a pond, vegetation and the sky, and have a clearly distinguishable shape. In fact, the shacks appear to be more defined than the characters due to the backlight. Being present in Shots 1 and 2, the shack is also remembered better than the other elements in Shot 3. However, character movement is also likely to attract attention.

Let us now examine the linguistic F/G assignment in the sequence.

⁹ The spatial construction of the sequence thus follows the principle of 'out of component parts' that is presented in Bordwell & Thompson (1990: 215), whereas the spatial composition in Case 1 follows the principle of 'analytical breakdown' (ibid.).

Transcription 2. “The toilet sequence”

	AD-EN (00:10:10)	AD-DE (00:09:45)	AD-ES (00:10:10)
Shot 1	A BUZZ OF A FLY + CLANKS + A SILENT HUM + A MELANCHOLIC TONE (?)		
	The five-year-old Jamal squats over a hole in a makeshift shanty toilet .	Im Slum. <i>In the slum.</i> Der siebenjährige Jamal hockt in einem kleinen Holzverschlag und blickt verträumt vor sich hin. <i>The seven-year-old Jamal squats in a small wooden shack and stares dreamily into space.</i>	Recuerda y vuelve al pasado de su infancia. <i>He remembers and goes back to his childhood.</i> Jamal, en cuclillas sobre un agujero, hace sus necesidades en el interior de una caseta de madera. <i>Jamal, squatted over a hole, relieves himself inside a wooden shack.</i>
Shot 2	DISTANT BARKING OF DOG + A HELICOPTER BUZZES + A DISTANT WHIRR		
	Salim sits guard outside .	Draußen wartet Salim. <i>Outside, Salim waits.</i>	Su hermano Salim está sentado frente a la puerta. <i>His brother Salim is sitting in front of the door.</i>
Shot 3	→ HELICOPTER → WHIRR + DISTANT SINGING OF BIRDS		
	A man rushes down towards the toilets .	Der Verschlag steht am Ende eines Stegs. <i>The shack stands at the end of a plank.</i>	Son retretes comunales. (they) Are public lavatories. No hay saneamiento alguno... <i>There is no sanitation whatsoever...</i>
(Shot 4)	(All are occupied.)	(Zu beiden Seiten liegen weitere Stege mit Verschlägen .) <i>On both sides, more planks with shacks.</i>	(... y las heces de todo el mundo caen a un pozo común desde los agujeros .) <i>and the faeces of everybody fall into a common pool from the holes.</i>

On the first reference to SHACK, all three audio descriptions refer to it with a locative PP that serves as reference entity for the subject’s (Jamal) action:¹⁰ *The five-year-old Jamal squats over a hole in a makeshift shanty toilet* (AD-EN), ‘The seven-year-old Jamal squats **in a small wooden shack** [...]’ (AD-DE), and ‘Jamal, squatted over a hole, relieves himself

¹⁰ Before referring to Jamal and the shack, the German AD furnishes a short description of the location *im Slum* ‘in the slum’ and thus renders a different amount of spatial information than is shown visually (a close-up of a character) (see Hirvonen 2012: 24, 29). The Spanish AD also begins by an explicit shift to a new scene: ‘He remembers and goes back to his childhood.’ This utterance describes Jamal remembering, an action which the sighted audience interprets from the look on his face, shown in the preceding shot, and from the flashback that follows the look.

inside a wooden shack' (AD-ES). The '(small) wooden shack'/*makeshift shanty toilet* can be assigned several Ground characteristics. This is Jamal's location and the reference entity for his squatting. It is also conceived of as more permanently located and larger than the character (the character is INSIDE of it).

Next, all three audio descriptions treat Salim as Figure in the sense that his whereabouts are determined in relation to an outside space. In other words, *Salim sits guard outside* (AD-EN), '**Outside**, Salim waits' (AD-DE), and 'His brother Salim is sitting **in front of the door**' (AD-ES). The locative adverb *outside* is contextually determined through its relation to the previously introduced place of action, that is, the shack/toilet, which makes SHACK Ground anew. The AD-ES locates Salim 'in front of the door', which activates the SHACK schema (shacks have doors) and seems to assign Figure and Ground more specifically than the English and German descriptions by implying "in front of the shack door". In contrast, 'outside' is more open-ended and potentially complex, so that one must ask what exactly this outside space consists of. In this regard, the AD-EN provides another kind of cue: 'to sit guard' implies an action that involves a vicinity of the actor towards the referent entity, which in this case is SHACK; the person who *sits guard* must be somewhere near the toilet. Therefore, the AD-DE seems the most implicit by both the locative expression ('outside') and the verb ('waits'), which does not imply proximity in the same way as *sits guard* does.

In continuation, the English version pauses but the German and Spanish audio descriptions proceed. In the AD-DE, SHACK converts to Figure ('**The shack** stands at the end of a plank') and is conceptualised as a subject whose location ('stands') is determined in reference to ('at the end of') another element ('a plank'). SHACK is also active and conceptually movable in that it *steht* 'stands'. Moreover, given that the slum is the overall Ground in the AD-DE (recall the first utterance 'in the slum'), SHACK is a smaller and geometrically simpler element. This is also valid for PLANK, as it is a new element and more recently on the scene/awareness. Thus, when framed against the larger spatial entity, the slum, both SHACK and PLANK can be conceived of as Figures. With the next utterance, 'On both sides, more planks **with shacks**', they merge into one Figure, PLANKS WITH SHACKS, consisting of "a multiplicity of points" (Talmy 2000: 312). The AD-ES also refers to SHACK and presents it as Figure but with a distinct approach. Instead of referring to the physical aspects of the scene, the Spanish AD predicates a functional property by '**(they) are public**

lavatories'. The verb *son* '(they) are' refers to SHACK anaphorically since the action that is carried out in it (Jamal relieving himself) relates to the function of lavatories. Instead of one, several SHACKS are identified as toilets. The utterance represents a meta-Figure that describes a property (ibid. 330–332), and the description of a property creates a pause in the story action (Chatman 1978: 74). The AD-ES sequence ends with 'There is no sanitation whatsoever and the faeces of everybody fall to a (common) pool **from the holes**'. The locative PP 'from the holes' evokes the SHACK schema again and assigns it a Ground function.

The English AD continues at the end of the sequence with utterances that extend beyond Shot 3 and these are discussed here because they assign F/G segregation to SHACK: *A man rushes down towards the toilets. All are occupied*. The locative PP *towards the toilets* features as the reference entity, Ground, for the man-Figure's action (*a man rushes down*).¹¹ The toilet referent is now in the plural, which implies the presence of more than one SHACK. In *all are occupied*, *all* refers anaphorically to *toilets*, yet the passive voice of the sentence seems tricky to interpret in terms of F/G assignment. One interpretation is that due to the passive form *are occupied*, *all* (toilets) are Ground with an embedded or implied Figure; another possibility is that the referent is a meta-Figure (*all are*) that is characterised by a state of affairs (*occupied*) (see Talmy 2000: 331–332). Nonetheless, assigning a meta-Figure function to *all* may present a problem because "being occupied" schematically entails an actor or actors – being occupied by someone – in which case the toilets would be the reference entity, that is, Ground.

4. Results

The analysis showed that the verbal representation of a visual scene can parallel the visual F/G segregation and that the different AD versions both converge and diverge in the verbal F/G segregation. The first shot in both Cases 1 and 2 seems to render the most unanimous segregation. For example, in Case 1, BUS was first assigned the Figure function in the visual composition as well as in the German and English audio descriptions; the Spanish audio description diverged from this order by not describing the first shot *per se*. For most of the time, the English and German audio

¹¹ A similar case can be found in Talmy (2000: 330): "The red leaf drifted toward the brown leaf".

descriptions proceed “shot by shot”, following the visual F/G segregation. The Spanish AD also reflects the spatial organisation of the imagery in terms of Figure and Ground, except for the first shot, but it does this asynchronously to the film shots. As a consequence, the narrative progresses differently in the AD-ES, resulting, for instance, in different combinations of sound and descriptions as in the AD-EN and AD-DE. In the first shot of Case 2, all three audio descriptions corroborate the Groundness of SHACK.

However, as the sequences of Case 1 and 2 continue, the linguistic F/G assignment becomes more varied. The audio descriptions do assign BUS a Ground function in the second shot of Case 1, but the English version refers to the Ground element as a grammatical subject (*the bus is full of...*). In the third shot of the sequence, the German AD excludes a reference to the bus completely (a description of the bus comes later). In Case 2, depending on what one relates SHACK to in the second and third shots, Figure and Ground may be assigned differently. With regard to the characters, it is Ground, but in relation with the landscape, it is Figure. Thus, the dynamic nature of F/G segregation plays a role here. After the second shot, the corresponding audio descriptions are indirect and refer to SHACK by an implicit schematic and contextual link (*outside/‘in front of the door’*). In the third shot, the linguistic F/G assignment in the audio descriptions differs even more: the AD-DE describes it as Figure (*‘the hut stands...’* and *‘...more huts with planks’*), the AD-EN refers to it as Ground in plural (*...towards the toilets*), and the AD-ES assigns it as meta-Figure (*‘(they) are public lavatories...’*). The passive voice in *all are occupied* is an interesting case since it seems to be more open-ended in terms of F/G segregation (a self-referencing Figure and/or the embedded Groundness of SHACK).

As for applying the Figure and Ground theories to the analysis, much of the F/G segregation seems to be definable by a few characteristics. For example, in the visual representation with features such as ‘thing-like versus substance-like’ and ‘appears closer/in front versus further away/behind,’ and in the linguistic representation with the characteristics ‘spatial properties to be determined versus a reference entity,’ ‘more movable versus more permanently located,’ and ‘smaller versus larger’. If, as Talmy (2000: 316) maintains, the definitional characteristics are determinative of Figure and Ground functions, then the question emerges whether and to what extent the rest of the features, in particular the “associated characteristics” (ibid. 315–316), should be applied in the

linguistic analysis. Case 1 touched on the challenge of defining dominance in the feature pair ‘appears more/less dominant.’ Dominance seems to be definable in terms of both narrative and physical properties. An element not being visually dominant but still encoded as Figure in the AD (recall the bus in the landscape in Case 1) implies that dominance can be understood in terms of narrative relevance. Or, is it that the other Figure features such as movement override dominance in this case? In addition, the analysis of the first case raised uncertainty in the characteristic ‘more recently on the scene/awareness / more familiar/expected’. Due to the thematic continuity – the familiarity based on the preceding narration – BUS seems “more familiar/expected”, which is a Ground feature, than “more recently on the scene/awareness”, which should characterise Figure. In contrast, the referents presented as Ground – *lush countryside* (AD-EN) and *eine Straße* ‘a street’ (AD-DE) – are new elements and therefore more recently focused on. What type of temporal frame is intended with ‘recentness’ remains obscure.

5. Conclusions

Since audio description refers to translating images into words, it requires studying the relation of language to non-linguistic, visual (and auditory) representation. The framework of Figure and Ground exists in the theory of both visual-perceptual and linguistic-cognitive representation and is therefore suitable for this kind of study. A limitation might be that the present study does not experiment on perception. The Gestalt theory is aimed at explaining how people perceive the environment, whereas here it is applied to model representation. In the present study, the theories of F/G segregation have been applied to the analysis in order to track Figure and Ground in the visual and linguistic data and, basing on the findings, to discuss intermodal differences and similarities of F/G segregation.

Segregating Figure and Ground in a narrative film seems to involve two types of characteristics: physical appearances and qualities such as shape, movement, lightness/colour and proportions, and functional properties such as narrative familiarity and dominance. Indeed, both bottom-up, such as object recognition, and top-down, such as narrative hypotheses, processes are in play while perceiving and interpreting films (Bordwell 1985; Vandaele 2012). In general, perception is governed by the principles of perceptual salience (for instance, the salience and intensity of the material) and subjective proximity (such as self-centredness) (see Wenz

1996: 278). These principles evoke thoughts that relate to the Figure and Ground characteristics. It seems that they can be divided along the dyad of perceptual–subjective, meaning that certain characteristics stem from perceptual salience while others relate to subjective preferences or orientations (though Wenz’s subjective proximity refers to the concrete position from which one perceives the things surrounding oneself). Thus, the following Figure features (from the Gestalt account) are defined by the salience of material properties: *appears to be thing-like; a contour appears at edge of figure’s shape; appears closer to the viewer and in front of the ground*. Other Figure features are determined more subjectively: *better remembered; more associations with meaningful shapes*. As the analysis demonstrates, the characteristic *appears more dominant* can have both a perceptual and a subjective basis. This also applies to the linguistic Figure (and Ground) characteristics. Thus, *has unknown spatial properties; more movable; smaller; geometrically simpler and less immediately perceivable* relate to perceptual salience, and *more recently on the scene/awareness and of greater concern/relevance* are more subjective. The characteristics that remain obscure are *more salient; once perceived* and *more dependent* and these could have both perceptual and subjective origins.

In a similar fashion, the verbalisation foregrounds certain aspects of the visual elements (Verhagen 2007: 50) that the present data illustrate with the references *kleinen Holzverschlag/caseta de madera* ‘(small) wooden shack’ in Case 2. The expression ‘(small) wooden shack’ refers to physical properties, size, material and type of the construction, whereas the expression *makeshift shanty toilet* profiles a function by the noun *toilet*. In addition, the attributes *makeshift* and *shanty* connote its possible location (slum, shanty town) as well as qualities (inferiority, poverty). As a matter of fact, this duality seems to parallel the two basic regularities people utilise when perceiving visual scenes. The first concerns physical regularities, when orienting to the physical properties of the environment, and the second is related to semantic regularities, when attending to the functions carried out in the scene (Goldstein 2010/2007: 115–117). One explanation for the superiority of sight over other senses in terms of spatial perception is offered by the concept “the gist of a scene”, that is, the rapid identification of the essential characteristics and regularities of a view (see *ibid.* 114). In AD, the way of perceiving the gist of a scene can be compensated for by naming and categorising the view (for example, *countryside*) (see Seiffert 2005: 77). However, by so doing, the linguistic representation seems to render Ground “thing-like”, while a “substance-

like” description (for example, *hazy green surfaces*) probably takes longer to utter. Related to this is a particular style of AD that aims at facilitating the following of the story, furnishing the viewers with clear cues about places of action or other narrative entities (see Ofcom 2000).

Although the results of the present study that point to the similarity and differences of conceptualisation are tentative, they are nonetheless intriguing. Regarding these differences, we may ask whether more differences arise when the visual source representation is more complex in terms of Figure and Ground.¹² This might suggest that the more ambiguous or dynamic the F/G relations are in the source material, the more varied the different versions of AD become. The linguistic-cultural framework of the data is Western, so further research is needed to determine whether greater divergence appears in other cultures, such as between the European and Asian AD, as cognitive differences between these have been observed (see Nisbett & Norenzayan 2002). Another issue is how languages prefer, or more easily employ, certain conceptual structures and what consequences this has on F/G segregation or other perceptual-cognitive aspects as well as on AD, in particular to the translation of the AD scripts between different languages. For instance, what shifts occur in the description of movement when a script is translated between “manner” and “path” languages such as English and Spanish (for the terms ‘manner’ and ‘path’, see Papafragou, Hulbert & Trueswell 2008)? Having said this, it is important to remember that AD does not necessarily reflect the perception that has taken place before the verbalisation since there may be a difference between a non-linguistic event perception and a linguistic conceptualisation of a scene. A study by Papafragou, Hulbert & Trueswell (2008) indicates that when people prepare to speak about what they see, they may allocate visual attention differently than when no verbal description is requested. This similarity in scene perception without verbalisation also appears cross-linguistically (*ibid.*). An intriguing conclusion regarding AD is that, between different-language descriptions, the “original” perception of the visual scene by the audio describers could be mutually similar but, due to undertaking the specific task of verbalisation, differences arise because of differences in the linguistic framework.

In conclusion, AD generates a range of topics to be studied from the perspective of the cognitive and psychological phenomena that are reflected in or constrained by language. As space continues to be the focus

¹² I would like to thank the reviewer for pointing out this issue.

of interest when studying the relation between language and thought (for instance, see Evans & Chilton 2010; Levinson 2003; Pütz & Dirven 1996), AD could be utilised as naturally occurring data of perception and verbalisation. Future research could involve experimental and observational studies on the perception, production and reception of the AD.¹³ Learning more about the effects that AD has in the minds of the audience and about how audio describers explain, discuss and share their conceptualisations of the visual or audiovisual source material would benefit not only science, but also the practice. Both fields would also be advanced by consistent, audience-involved and in-depth analyses of the perception and interpretation of the soundscape and of the describing voice.

Filmography

Slumdog Millionaire (2008) Directed by Danny Boyle & Loveleen Tandan, with audio descriptions in English (Pathé Distribution 2009), German (Hörfilm GmbH 2009) and Spanish (Navarra de Cine S.L. 2009).

References

- Bordwell, David (1985) *Narration in the Fiction Film*. London: Methuen.
- Bordwell, David & Thompson, Kristin (1990) *Film Art: An Introduction*. New York: McGraw-Hill.
- Branigan, Edward (2010) Soundtrack in mind. *Projections* 4 (1): 41–67.
- Cámara, Lidia & Espasa, Eva (2011) The audio description of scientific multimedia. *The translator* 17 (2): 415–437.
- Cabeza-Cáceres, Cristóbal (2013) *Audiodescripció i recepció*. Unpublished doctoral thesis. Universitat Autònoma de Barcelona.
- Chatman, Seymour (1978) *Story and Discourse*. Ithaca: Cornell University Press.
- Díaz Cintas, J.; Orero, P. & Remael, A. (eds.) (2007) *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*, Approaches to Translation Studies 30. Amsterdam: Rodopi.
- Ehrenstein, W. H. (2001) Perceptual organization. In Neil J. Smelser & Paul B. Baltes (eds.-in-chief), *International Encyclopedia of the Social & Behavioral Sciences*, pp. 11227–11231. Oxford: Pergamon.

¹³ A special issue of *Perspectives: Studies in Translatology* (Mazur & Kruger 2012) reports AD-related reception research with a cross-linguistic and -cultural rationale, and Cristóbal Cabeza-Cáceres' doctoral thesis at the Autonomous University of Barcelona analyses the reception of AD, more specifically, how the speed of speech, intonation and the explicitation of narration affect the comprehension of AD.

- Engberg-Pedersen, Elisabeth (2011) Cognitive foundations of topic-comment and foreground-background structures: Evidence from sign languages, cospeech gesture and homesign. *Cognitive Linguistics* 22 (4): 691–718.
- Evans, Vyvyan (2010) The perceptual basis of spatial representation. In Vyvyan Evans & Paul Chilton (eds.), *Language, Cognition and Space: The State of the Art and New Directions*, pp. 21–48. London/Oakville: Equinox.
- Evans, Vyvyan & Chilton, Paul (eds.) (2010) *Language, Cognition and Space: The State of the Art and New Directions*. London/Oakville: Equinox.
- Fahlenbrach, Kathrin (2008) Emotions in sound: Audiovisual metaphors in the sound design of narrative films. *Projections* 2 (2): 85–103.
- Fix, Ulla (ed.) (2005) *Hörfilm: Bildkompensation durch Sprache*, Philologische Studien und Quellen, Heft 189. Berlin: Erich Schmidt.
- Goldstein, Bruce E. (2010/2007) *Sensation and Perception*, eighth edition. Wadsworth: Cengage Learning.
- Fryer, Louise (2010) Audio description as audio drama: A practitioner's point of view. *Perspectives: Studies in Translatology* 18 (3): 205–213.
- Hatfield, Gary (2001) Perception: History of the concept. In Neil J. Smelser & Paul B. Baltes (eds.-in-chief), *International Encyclopedia of the Social & Behavioral Sciences*, pp. 11202–11205. Oxford: Pergamon.
- Herman, Vimala (1996) Space in dramatic discourse. In Martin Pütz & René Dirven (eds), pp. 553–569. Berlin/New York: Mouton de Gruyter.
- Hirvonen, Maija (2012) Contrasting visual and verbal cueing of space: Strategies and devices in the audio description of film. *New Voices in Translation Studies* 8: 21–43.
- Hirvonen, Maija & Tiittula, Liisa (2012) Verfahren der Hörbarmachung von Raum: Analyse einer Hörfilmsequenz. In Heiko Hausendorf, Lorenza Mondada & Reinhold Schmitt (eds.), *Raum als interaktive Ressource*, Studien zur Deutschen Sprache 62, pp. 381–427. Tübingen: Gunter Narr.
- Koffka, Kurt (1936) *Principles of Gestalt Psychology*, International Library of Psychology, Philosophy and Scientific Method. London: Kegan Paul, Trench, Trubner & co./New York: Harcourt, Brace and Company.
- Köhler, Wolfgang (1947) *Gestalt Psychology: an Introduction to New Concepts in Modern Psychology*. New York: Liveright Publishing Corporation.
- Kress, Gunther & van Leeuwen, Theo (2006/1996) *Reading Images: The Grammar of Visual Design*, second edition. New York: Routledge.
- Kruger, Jan-Louis (2010) Audio narration: Re-narrativising film. *Perspectives: Studies in Translatology* 18 (3): 231–249.
- Langacker, Ronald W. (1987) *Foundations of Cognitive Grammar, Volume 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Levinson, Stephen C. (2003) *Space in Language and Cognition, Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.
- Mazur, Iwona & Kruger, Jan-Louis (eds.) (2012) *Pear Stories and Audio Description: Language, Perception and Cognition across Cultures*. Special issue of *Perspectives: Studies in Translatology* 20 (1).

- Nisbett, Richard E. & Norenzayan, Ara (2002) Culture and cognition. In Douglas Medin (ed.) & Hal Pashler (ed.-in-chief), *Steven's Handbook of Experimental Psychology, Volume 2: Memory and Cognitive Processes*, third edition, pp. 561–597. New York: Wiley.
- Ofcom (2000). ITC Guidance On Standards for Audio Description. <http://stakeholders.ofcom.org.uk/broadcasting/guidance/other-guidance/tv_access_serv/archive/audio_description_stnds/> (10 September 2013).
- Papafragou, Anna; Hulbert, Justin & Trueswell, John (2008) Does language guide event perception? Evidence from eye movements. *Cognition* 108: 155–184.
- Pütz, Martin & Dirven, Rene (eds.) (1996) *The Construal of Space in Language and Thought*, Cognitive Linguistics Research 8. Berlin/New York: Mouton de Gruyter.
- Schubert, Christoph (2009) *Raumkonstitution durch Sprache: Blickführung, Bildschemata und Kohäsion in Deskriptionssequenzen englischer Texte*, Buchreihe der Anglia Zeitschrift für englische Philologie, 40. Band. Tübingen: Max Niemeyer.
- Seiffert, Anja (2005) Räumliches hören. Eine schemaorientierte Analyse der audiodeskriptiven Darstellung der Handlungsräume. In Ulla Fix (ed.), *Hörfilm: Bildkompensation durch Sprache*, Philologische Studien und Quellen, Heft 189, pp. 67–86. Berlin: Erich Schmid.
- Smelser, Neil J. & Baltes, Paul B. (eds.) (2001) *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon.
- Stockwell, Peter (2002) *Cognitive Poetics: An Introduction*. London/New York: Routledge.
- Talmy, Leonard (2000) *Toward a Cognitive Semantics, Volume 1: Concept Structuring Systems*. Massachusetts/London: MIT Press.
- Vandaele, Jeroen (2012) What meets the eye: Cognitive narratology for audio description. *Perspectives: Studies in Translatology* 20 (1): 87–102.
- Verhagen, Arie (2007) Construal and perspectivization. In Dirk Geeraert & Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, pp. 48–81. Oxford/New York: Oxford University Press.
- Wenz, Karin (1996) Iconicity in verbal descriptions of space. In Martin Pütz & René Dirven (eds.), pp. 269–286.

Contact information:

Maija Hirvonen
 Department of Modern Languages
 University of Helsinki
 PO Box 24
 00014 University of Helsinki
 e-mail: maija(dot)hirvonen(at)gmail(dot)com