Multimodal Representation and Intermodal Similarity
Cues of Space in the Audio Description of Film

Maija Hirvonen

ACADEMIC DISSERTATION
to be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki, in Auditorium PII, on the 15 November 2014 at 10 o'clock.

The articles have been included in the paperback version with the kind permission from their respective publishers.
ISBN 978-951-51-0368-0 (paperback)
ISBN 978-951-51-0369-7 (PDF)
http://ethesis.helsinki.fi
Unigrafia, Helsinki 2014

Abstract

The present dissertation analyses the representation of space in filmic audio description. The main objective of this study is to shed light on two critical challenges for audio description: (1) how the filmic space becomes audible in audio-described film through spoken language, sound effects, and music, and (2) how the visual representation of space in film can be cued by the linguistic mode in an audio description.

This dissertation consists of four articles and a thesis summary. The first article focusses on the multimodal representation of space through auditory cues. The other three articles explore intermodal similarity, which involves the question of how language reflects visual representation. To explain the complex phenomenon of translating images into words, this study applies the theoretical and analytical tools from translation studies, film studies and cognitive linguistics, and also adopts a cognitive orientation to explain both the filmic and linguistic representations as being cognitive representations that are constructed through visual, auditory, and linguistic cues.

This research first establishes that the auditory multimodality of an audio-described film creates a variant of the multimodality of the audiovisual film that entails dynamic constellations, perspectives and foci. Secondly, the analysis presents evidence for the varied potential of the linguistic mode in terms of representing space. Thirdly, this study defines strategies for intermodal similarity in terms of the corresponding linguistic signs for the filmic cues of representation and narration. The results of this dissertation may be valuable in justifying the consistency and standards for audio description as well as in developing other, more far-reaching usages of audio description as a mode of transforming information from one system of representation into another.

Acknowledgements

Like a little stream that collects drops of water along its way to becoming a river, I have learned and continue to learn from so many people. First and foremost, I express my most heartfelt gratitude to my main supervisor, Professor Liisa Tiittula. Thank you for guiding me along the way with your wisdom, for asking the right questions, for the balance of praise and critique, and for being the teacher who sees the potential in their students even before they are aware of it themselves. I look forward to the many years of working together to come!

I also owe my sincere gratitude to my second supervisor, Professor Henry Bacon, who showed interest in the topic of filmic audio description and guided me with inspiring viewpoints on cinema. I would also like to thank the two pre-examiners of this thesis, Professor Aline Remael and Dr. Sabine Braun, for the constructive feedback and suggestions for revision in the thesis summary, as well as all the editors of the publications and the anonymous reviewers who contributed greatly to the quality of the research articles. Furthermore, my warmest thanks go to the language revisers and proofreaders of the research articles and the thesis summary, in particular to Kathleen Moore for your flexibility.

During the span of this doctoral life, I have had the privilege of working in different places and collaborating with different people and professions. I would like to thank the *LANGNET* Doctoral Programme in Language Studies for the foothold they have provided in the research community and for their multifaceted perspectives into the study of language. Among my mentors are Pentti Haddington and Jukka Mäkisalo and, in particular, the translation studies sub-programme leaders Merja Koskela, Kaisa Koskinen, Nina Pilke and Liisa Tiittula – I will always remember our research seminars as the source of inspiring ideas and a resource for supportive feedback.

My warmest thanks also go to the *TransMedia Catalonia* research group at the Autonomous University of Barcelona, led by Dr. Pilar Orero, and to *Institut für Deutsche Sprache* in Mannheim, especially the research group of multimodal interaction led by Dr. Reinhold Schmitt. These communities and researchers greeted me with great hospitality, showed curiosity towards my work and improved it through enthusiastic discussions. Furthermore, I am grateful for the senior and peer support and feedback received along the way. Thank you Anna M., Anna V., Anukaisa, Cristóbal, Jan-Louis, Leena, Mari, Maarit, Marta, Meri, Minia, Nazaret, Paula, Tia and Tuija! Finally, I wish to express my gratitude and pride for having been able to share my research with users and practitioners of audio description. I am grateful to Anu Aaltonen & Hannele Antikainen and to Bernd Benecke & Elmar Dosch for introducing me to the art of audio describing in Finland and Germany. I am indebted to the

non-profit organisation *Näkövammaisten kulttuuripalvelu ry*. Being part of the commission that develops audio description in Finland has constantly reminded me of the world outside academia and of the actual applications that research could and should have.

I am grateful to the following institutions for believing in this topic, for providing an office, and for financing my thesis: the University of Tampere, the Kone Foundation, the *LANGNET* Doctoral Programme in Language Studies, the German Academic Exchange Service, the Finnish Academy of Science, as well as the Department of Modern Languages and the Translation Studies and Terminology Research Group at the University of Helsinki.

The preparation of this thesis would not have been possible without the enduring and kind support of my "home team". My dear family and friends have helped me on uncountable occasions and in numerous ways, offering to babysit, listening to my work-related agonies and, last but not least, reminding me of the important world beyond work. Thank you for sharing your life with me. The past years have been a time of great grief and great joy. I dedicate this dissertation to my late mother and sister, who left this life much too early, as well as to my son and my husband, for whom I am grateful every day. Never would have made it without you, Ari!

It seems an impossible task to mention on this occasion all of the inspiring people that I have met over the past seven years – you know who you are and what you are worth! This doctoral journey has, so I believe, shown me my calling in science and society. As one of the pre-examiners put it, a PhD study is to be considered ongoing research, not the end but the beginning of a career.

Lempäälä, 17th of October 2014
Maija Hirvonen

Contents

Abstract **3**
Acknowledgements **4**
Contents **6**
List of original papers **7**
Part I **9**
1 Research design **9**
1.1 Research problem **10**
1.2 Objectives **12**
1.3 Data **14**
1.4 Methodology **15**
1.4.1 Data transcription and presentation 16
1.4.2 Multimodal and intermodal analyses 17
2 Research environment **19**
2.1 Background to audio description **19**
2.1.1 Audio description as an assistive service 19
2.1.2 Audio description as translation 23
2.1.3 Audio description in the filmic context 24
2.2 Research on audio description **25**
2.2.1 Overview of the state of the art 25
2.2.2 Introducing research on filmic audio description 26
2.3 Cognitive orientation to film and language **29**
2.3.1 Cues and schemata in the filmic representation 30
2.3.2 Cues and schemata in the study of audio description 31
Part II **34**
3 Results **34**
3.1 Multimodal representation of space in audio-described film **34**
3.1.1 Interplay of the auditory modes 35
3.1.2 Cues of space in speech 36
3.2 Intermodal similarity between visual and linguistic cues **37**
4 Discussion **41**
4.1 Theoretical discussion **41**
4.1.1 Filmic audio description 41
4.1.2 Similarity and divergence between image and language 43
4.2 Methodological discussion and evaluation of research **45**
4.3 Implications for practice and future research **47**
References **50**
Film data **50**
Other films **50**
Literature and online sources **50**

List of original papers

ARTICLE 1: Verfahren der Hörbarmachung von Raum: Analyse einer Hörfilmsequenz [To Make Space Audible: Analysis of a Sequence of an Audio-Described Film].

Published in: Hausendorf, Heiko, Mondada, Lorenza & Schmitt, Reinhold (Eds.) (2012), *Raum als interaktive Ressource*. (Studien zur Deutschen Sprache 62). Tübingen: Narr, 381-427. Reprinted in the paperback version with the kind permission of Institut für Deutsche Sprache, Mannheim.

Commentary on Article 1:
This article was co-authored. A report on the contributions of the authors is provided in a separate attachment.

ARTICLE 2: Contrasting Visual and Verbal Cueing of Space - Strategies and Devices in the Audio Description of Film.

Published in: *New Voices in Translation Studies 8* (2012), 21-43. Available at: http://www.iatis.org/index.php/publications/new-voices-in-translation-studies/item/-488-current-issue8-2012.

ARTICLE 3: Perspektivierungsstrategien und -mittel kontrastiv: Die Verbalisierung der Figurenperspektive in der deutschen und finnischen Audiodeskription [A Contrastive Study of Perspectivation Strategies and Linguistic Devices: Verbalising Character's Perspective in German and Finnish Audio Description].

Published in: *trans-kom: Zeitschrift für Translationswissenschaft und Fachkommunikation 6(1)* (2013), 8-38. Available at: http://www.trans-kom.eu/ihv_06_01_2013.html.

Commentary on Article 3:
The reference to Kuusi (2011: 78) on p. 25 is not completely accurate. Through a re-examination of Kuusi's work, I find that she connects a lack of cohesion with a greater implictiness or ambivalance of interpretation. In other words, Kuusi does not directly assert that this would imply a change in perspective, as I infer in the article. The break in cohesion requires more interpretation from the receiver in order to produce coherence in the story (Kuusi ibid.). The implication for my analysis is that the implicitness of the link between a glance and its stimulus in the 'experienced character perspective' is further corroborated. The

lack of cohesion between utterances does, however, shift the focus of attention from one thing to another and this can ultimately also imply a change in the perspective (for instance, the situation is viewed from a different angle).

Example 5, on page 24, is missing a transcription of the musical tone that begins simultaneously with the audio description on line 8.

In the references to Gutenberg, the year of publication is incorrect. Instead of 2001, it should read 2000.

ARTICLE 4: Sampling Similarity in Image and Language – Figure and Ground in the Analysis of Filmic Audio Description.

Published in: *SKY Journal of Linguistics 26* (2013), 87-115. Available at: http://www.linguistics.fi/julkaisut/sky2013.shtml.

Commentary on Article 4:
During the publication process of this article, some of the final changes were not made before publishing, and this resulted in a minor inaccuracy with regard to the last audio descriptive utterances in the analysis of Case 2. The last paragraph of the audio description analysis (p. 107) incorrectly states 'In continuation, the English version pauses, but the German and Spanish audio descriptions proceed', although the correct statement is that all the audio descriptions continue but with distinct references to the 'shack' element.

Part I

1 Research design

This study is an academic dissertation for a PhD degree in German translation and consists of four previously published scientific articles, of which one is co-authored, and of the present summary, which includes an introduction, the main results, and a discussion. This thesis defines how filmic space is represented by the different resources of communication, or the three modes that are operative in the audio description of film: the visual, auditory, and verbal modes. In addition, this dissertation reports on intermodal similarity, that is, how spatial representations in the visual mode are cued by special representations in the verbal mode.

The research articles that comprise this dissertation have been published in the fields of translation studies and linguistics. In addition to these fields, the research should be of interest to scholars in film studies and multimodality as well as to practitioners in audio description, audiovisual translation and accessibility. Three of the publications have been published in refereed journals (*New Voices in Translation Studies*, *trans-kom* and *SKY Journal of linguistics*) and one has appeared in an edited book (*Raum als interaktive Ressource*, which is a volume in the series *Studien zur Deutschen Sprache*). Two of the articles are in English and two are in German. However, to address an international readership, this thesis summary has been written in English.

The four publications are the following:

Article (1). (co-authored with Tiittula, Liisa). Verfahren der Hörbarmachung von Raum. Analyse einer Hörfilmsequenz. In Heiko Hausendorf, Lorenza Mondada & Reinhold Schmitt (Eds.), *Raum als interaktive Ressource*. Tübingen: Günter Narr (2012), 381-427.

Article (2). Contrasting Visual and Verbal Cueing of Space – Strategies and Devices in the Audio Description of Film. *New Voices in Translation Studies 8* (2012), 21-43.

Article (3). Perspektivierungsstrategien und -mittel kontrastiv: Die Verbalisierung der Figurenperspektive in der deutschen und finnischen Audiodeskription. *trans-kom: Zeitschrift für Translationswissenschaft und Fachkommunikation 6(1)* (2013), 8-38.

Article (4). Sampling Similarity in Image and Language – Figure and Ground in the Analysis of Filmic Audio Description. *SKY Journal of Linguistics 26* (2013), 87-115.

As the publications have different readerships and as the research subject is assumed to be new to many readers, some repetition necessarily occurs in the articles, but this mainly concerns the introductions to audio description. All of the articles involve the analysis of real data, that is, films and audio descriptions, and make use of a qualitative methodology to understand and explain the phenomenon in focus, which is the representation of space in audio-described film. Yet each article explores a different research question, and the analyses in each article exhibit distinct data samples.

This summary is structured as follows: Part I introduces the research by describing its general design in terms of the research problem, the objectives, the data and the methodological approach. Part I then reviews the practical and scientific environment in which the research is conducted as well as the environment it contributes to. Following the publication of these research articles, the second part (Part II) presents the principal results of the research in relation to the research problem that was introduced in Part I. Finally, the third part (Part III) discusses the significance of the research in terms of the theoretical, methodological, and practical input. Furthermore, this latter part offers an evaluation of the research design as well as recommendations for future research.

1.1 Research problem

When I first became interested in audio description in 2007, the question that initially came to mind concerned how images are translated into language. This has continued to be a persistent line of inquiry throughout this study.

The research subject of this dissertation is audio description. This can be defined as a form of intermodal translation (Braun 2008, Hirvonen 2012) in which the visual channel of perception is changed to an auditory one and in which visually represented information is verbalised and spoken. Audio description serves to offer access that aims at improving the integration of the blind and visually impaired with the culture and communication that is visual and audiovisual. Accessibility in general terms refers to overcoming different types of barriers, such as those that are sensory, linguistic or physical, to enable people with impaired hearing to grasp auditory information and to likewise help the visually impaired to comprehend visual cues. A central part of contemporary culture consists of the audiovisual narratives in cinema and television programmes. These narratives are part of people's everyday life as sources of enjoyment and as topics of conversation. For this reason, it is relevant to enhance and to enable access to films and other audiovisual narration through audio description.

This dissertation analyses two critical challenges in audio description. The first is the

interactive constitution of communication and meaning through multiple modes in audiovisual narratives. The second challenge is the shift or transformation of one mode, the visual, to another, the verbal.

Hence, this analysis explores the problem of *multimodal representation,* as it entails the question of how different modes, that is, communicative resources that involve a channel of perception and a system of representation, interactively and often simultaneously furnish narrative information. Since the multimodality of filmic narration consists of images, language, music and sounds received through hearing and vision, one must ask what remains of this multitude in audio-described film. In other words, the question here is how the interplay of sound effects, music, spoken language and voices in the forms of dialogue, voice over, and audio description, constitutes multimodal representations. Another research question relates to the *intermodal similarity* that potentially exists between the visual representation in film and the linguistic, verbal-oral representation in audio description. The visual representations in film arise in many ways, including through the photographic moving images. These images depict characters, objects and places, graphic illustrations, such as film company logos, as well as texts in different forms, such as subtitles and credits.

In the present study, the focus of multimodal and intermodal inquiries is the representation of space. This focus is motivated by the assumption that space appears to be a highly visual concept and is a central element in film narration. In many ways, sight seems to outrank hearing in interpreting spatial information, and people with sight perceive space rapidly and efficiently. For instance, when determining the location of an object, the visual cues, encompassing such factors as the size of the object and its relation to other objects, offer more accurate information than auditory cues. As an example, let us imagine that a person hears the wind whisper through the moving branches of trees. Through vision, he or she may count the number of trees that are moving, whereas that counting would be more inexact by merely listening to the sounds.

The concept of space used here is the film-narratological sense of scenographic space or the 'imaginary space of narration', the world in which the film narration suggests that the events of a story occur (Bordwell 1985: 113). The present research analyses various aspects of this imaginary space of film, including the spatial composition of the shot and the point of view. Cinema has a variety of means and techniques that can be used to represent space. All in all, cinema can imitate our viewing of the real-world space by being 'literal' or analogous to the objects, dimensions and relations in the real world (Chatman 1978: 96-97). Unlike theatrical representation, cinema can draw the images closer to the viewer through cinematographic means, such as by using close-ups, so that our experience of characters' cognitive and

emotional states becomes intensified (Prince 1993: 24). Another constitutive factor of the spatial representation in film is the point of view from which space is presented.

A further motivation for focusing on space is that language seems to be less efficient than vision in representing space in detail. For example, by merely looking at a scene, we easily notice the exact configuration of it – the types of trees in the landscape and how the branches of the trees are illuminated against the sky. In fact, we perceive the characteristics of the trees visually even if we do not recognise their species. While language works on the basis of abstract concepts, visual representation provides unique features (see Grodal 1997; Landau et al. 2010; Schwarz 1992). Further evidence of the visual efficiency in space perception is how we rapidly comprehend 'the gist of a scene' and identify the scene as a basic configuration, such as a cityscape or the countryside (Goldstein 2007/2010: 114). It is interesting to note that language can be used to preserve some of this efficiency in translating images.

1.2 Objectives

The main question that is analysed in this study concerns how the different modes in an audio-described film cue or carry information on space. This study is an attempt to answer this question by analysing the potential of the auditory, visual and linguistic modes in representing space and by determining how to cue similar information on space with linguistic-propositional representations rather than with visual-pictorial representations (see Braun 2007: 9; for the notion of 'cue', see Chapter 2.3). The different foci of the research articles are presented in Table 1 on the next page.

Table 1: Overview of the research articles

Articles	**Approach**		**Objectives**
	Intermodal	**Multimodal**	
(1) *Verfahren der Hörbarmachung von Raum: Analyse einer Hörfilmsequenz*		X	Exploration of the multimodality in audio-described film
(2) *Contrasting Visual and Verbal Cueing of Space - Strategies and Devices in the Audio Description of Film*	X		Exploration of the visual and verbal cues of space; definition of the similarity of strategies in film and audio description
(3) *Perspektivierungsstrategien und -mittel kontrastiv: Die Verbalisierung der Figurenperspektive in der deutschen und finnischen Audiodeskription*	X		Definition of the linguistic perspectivation strategies for the filmic point of view shot
(4) *Sampling Similarity in Image and Language: Figure and Ground in the Analysis of Filmic Audio Description*	X		Exploration of the intermodal similarity between visual and linguistic representation through a shared theory

Article 1 presents the first stage of this study by analysing how the spatial dimension of film narration becomes audible through speech, sound effects, and music. The dimensions of filmic space that were defined in Article 1 served as a basis for the spatial categories in the other articles. Article 2 explores intermodality, comparing the visual representation of the filmic space to the linguistic representation in the audio description, and the main objective has been to determine the similarity between the two modes. Inspired by the strategies formulated in Article 2, Article 3 was designed to define strategies for the verbal representation of a particular point of view in film. Article 4 explains intermodal similarity by discerning how a theory that models both visual perception and verbal representation can be applied to an analysis of audio description.

Instead of focusing on the differences that translation necessarily produces – audio description even more so by changing the mode – and given the dissimilarities between visual and linguistic representation, this dissertation explicitly concentrates on the similarity between the filmic imagery and audio descriptions while also arguing for its existence. I adopt the term 'similarity' (Chesterman 1996, 2007) to denote a possibility of sameness between the source and the target material. In other words, translation involves both sameness and difference in the sense that the same source material may produce different translations (by different translators, in different times, etc.), which may still be regarded as being similar to the source. Thus, I have decided not to use the concept of equivalence in the context of audio description because, as I mention in Article 2, equivalence is, traditionally at least, centred on the area of interlingual translation and seems too narrow in scope to describe a translation that is intermodal.

In addition to the main goals mentioned above, the methodological significance of this dissertation is that it proposes and discusses methods that can be adopted to analyse intermodality and multimodality. The focus being on the representation of space, this study contributes to a contemporary interest in the research on audio description (Seiffert 2005; Tanis Polat 2013), as well as the ongoing interest in linguistics (for instance, Evans & Chilton 2010; Pütz & Dirven 1996). Moreover, the results of this study will hopefully inspire practical work in audio description, for example, by generating new and justified ideas for translation strategies and solutions.

1.3 Data

The research corpus comprises eight full-length, contemporary narrative films that were used as research data in the different articles. These films are listed in Table 2 on the next page.

Table 2: Research corpus

Film	Year	Director	Language	AD	Article n°
Der Untergang	2004	Oliver Hirschbiegel	German	German	1; 2; 3
El hundimiento			Spanish (dubbed)	Spanish	2
Dancer in the Dark	2000	Lars von Trier	English	English	2
Dancer in the Dark			German (dubbed)	German	2
Tauno Tukevan sota	2010	Heidi Köngäs	Finnish	Finnish	3
Slumdog Millionaire	2008	Danny Boyle, Loveleen Tandan	English, Hindi	English	4
Slumdog Millionär			German (dubbed)	German	4
Slumdog Millionaire			Spanish (dubbed)	Spanish	4

Except for the film *Tauno Tukevan sota* [Tauno Tukeva's war], which is only in Finnish and with a Finnish audio description[1], the corpus includes both the original version of each film with an audio description and a version that is dubbed, or two that are dubbed and audio-described. All the films can be classified as drama. The languages analysed are English, Finnish, German, and Spanish. The multilingual corpus enables the comparison of different-language audio descriptions, and consequently, of the different translation solutions of the same filmic source material.

Table 2 shows that the four articles access different parts of the corpus. For example, Article 1 analyses the first two minutes of the audio-described German version of *Der Untergang* because the research method was both time- and space-consuming (see Chapter 1.4). Articles 2 and 3 focus on a selection of sequences from the films *Der Untergang* / *El hundimiento*, *Dancer in the Dark*, and *Tauno Tukevan sota*, and their audio descriptions (see the articles for information on the selection). Finally, Article 4 examined two sequences from *Slumdog Millionaire* and the corresponding English, German and Spanish audio descriptions. These sequences were selected in the preliminary analysis of the films and audio descriptions.[2]

1.4 Methodology

This research was for the most part qualitative in that it relied on theory and observation to describe and explain the representation of space in the data (Article 2 also reports on

[1] To my knowledge, Finnish films that have an audio description in a language other than Finnish are currently not available.

[2] The preview of *Slumdog Millionaire* was conducted in collaboration with Paula Igareda (see Hirvonen & Igareda 2011).

quantitative findings). The articles display a methodological process that began with a data-driven approach in Article 1 and ended in a theory-driven study in Article 4. Article 1 analysed a film sequence step-by-step as the film proceeded, describing the way in which different resources provide spatial cues. In Articles 2, 3 and 4, the data were first observed in a preliminary, data-driven analysis to determine relevant foci of interest, but the actual analyses illustrated in the articles departed from preconceived theoretical categories (the film techniques of shot distance in Article 2, and point of view shot in Article 3) or a theoretical framework (Figure and Ground segregation in Article 4).

1.4.1 Data transcription and presentation

In order to conduct analyses and to disseminate the research in printed articles, I presented multimodal and non-written data in a visual and/or textual form that can be examined. This data presentation required some form of intermodal translation or transformation, resulting in some elements of the original information being transformed or even omitted. (See Flewitt et al. 2009/2011: 45, 51.) Having said this, I need to emphasise that the original communicative situation in which the data exists – film and audio-described film – was accessed during the analysis to check both the accuracy of the data transcription and the reliability of the analysis.

The data presentation involved primarily textual transcriptions and other visual representations of the data. These included the shot protocols of films, transcriptions of audio descriptions and audio-described film, as well as audio description scripts. Furthermore, some of the non-linguistic material was represented linguistically, whereas the auditory material was presented visually. The sounds, music and the vocal features of speech were verbally described or expressed in symbols or written down (for instance, Article 1 presents music as note symbols). Some auditory qualities were also represented by symbols, such as the intensity of music being reflected through a denser succession of the note symbols, and a higher volume through the use of capital letters in the textual description. As concerns the linguistic data, various systems and levels of specificity were adopted for the transcription, and the reader will notice that the data presentation varies among the research articles. For example, Articles 1, 2 and 3 adopt the GAT system (Selting et al. 1998) to represent speech, whereas Article 4 renders the verbal dimension of speech orthographically. The underlying reason for using the different systems is the difference between their analytical foci. In other words, in presenting audio description as speech, the focus was not merely on the verbal dimension, but also on the vocal aspects of spoken language that contribute to meaning-making such as pausing, intonation, emphasis and other prosodic features. Where the vocal dimension is excluded from the transcription, the analysis focussed on the verbal or conceptual level of language (Article 4). Throughout these studies the conversation analytical approach is applied in the presentation and in the sequential treatment of the data (see also

section 1.4.2).

Beside the textual form, the pictorial visualisation was provided in the form of the sketches and drawings of film shots in combination with text (Articles 2 and 4), and the abstract graphics of the potential 'imaginary' space in mind (Article 1). As a consequence, the moving images became static, and the realistic, photographic scenes were reproduced in a graphic form, so that the reader would have access to 'frozen' instances rather than to the dynamic process of film narration and to less elaborated representations (black-and-white drawings instead of colour photographs). These still images were accompanied by the verbal descriptions of the respective film sequences in the body text. Although a more authentic view of the film data is obtained by using still shots, I chose to use drawings to represent shots, as many publications require authorisation from the copyright holder in order to publish original images. Obtaining these rights would have been prohibitively expensive and time-consuming. In Article 2, the sketches aim at reproducing the shot's essential characteristics, which in the respective analytical context was the spatial composition in terms of shot size. Article 4, on the other hand, provides realistic, but black-and-white drawings made by an artist in an attempt to illustrate the complexity of the spatial organisation in a shot. Again, it should be noted that I used authentic data, that is, I watched and listened to the films during most of the data analysis, and the graphic illustrations were created for disseminating the research in the printed publications.

1.4.2 Multimodal and intermodal analyses

To understand both the intermodality and multimodality in the film audio description, the present study applied analytical tools from different disciplines to establish means of observing and contrasting the representations in different modes (see Remael 2010).

To analyse the multimodal construction of space, a multimodal interaction analysis was applied in Article 1. The audio-described film was treated as a stand-alone product rather than being compared to the source material. Multimodal interaction analysis orients to all communication modes as being equally potentially relevant for the participants of communication and, rooted in conversation analysis, it focusses on the sequential constitution of interaction (Hausendorf et al. 2012: 9, 12). In Article 1, the speech, music, and sound effects were transcribed and analysed step-by-step as they were audible in the audio-described film. With this analysis, it was also possible to reconstruct the gradual development of the scenographic space, which is imaginable through the audible resources, and to illustrate it graphically. A sufficient amount of information was obtained from the first two minutes of the audio-described film to yield data for the entire scope of the article. In addition to Article 1, multimodality was also considered to some extent in the following studies because the

intermodal analysis was supplemented by examining information from the filmic soundscape.

Articles 2 and 3 move the analytical focus from multimodality to intermodality and the analytical perspective from reception to translation. The key question is how the representation of space that occurs in one mode (the visual) can be achieved by means of another mode (the linguistic). Article 2 analysed shot distance by adopting film analysis and audio descriptions by using both conversational and text analyses. Article 3 also departed from the idea of determining ways to recreate a certain representation of space in an audio description, and its analysis centred on language and the comparison between two different-language audio descriptions. However, the analytical background for this article was intermodal, meaning that a film analysis preceded the linguistic analysis in which I studied the representations of the characters' subjective perspective (for point-of-view shots and shots depicting mental states, see Branigan 1984).

Article 4 deviates from the first three articles in that it exemplifies a theory-based analysis and that it concentrates on describing and evaluating the use of a theoretical framework in the analysis of film and audio description data. The framework adopted is Figure and Ground segregation, which explains the perception of visual space (reviewed in Evans 2010) as well as the linguistic representation of spatial entities (Talmy 2000). In Article 4, I conducted a sample analysis of the visual and linguistic data by following the analytical categories from the theories and then compared the results of the two analyses, looking for similarities and discrepancies in the visual and linguistic representations.

This thesis involved a comparison between the different-language versions of audio description as a method to further explore intermodality, that is, to contrast parallel audio descriptions with the same visual representation. As a consequence, this dissertation explores language as a representational system (involving, among other things, the semantic and structural levels of expression) and does not conduct cross-linguistic analyses of languages per se, no matter how intriguing and important these may be (for a discussion on this topic, see Chapter 5.3).

2 Research environment

This chapter describes the practical and scientific context in which the present dissertation is conducted. In particular, it reviews the development of audio description in Finland because the Finnish field is less known by the international scholarly community.

2.1 Background to audio description

2.1.1 Audio description as an assistive service

Audio description has been developed to improve the accessibility of the visually impaired to the visual and audiovisual culture and communication. This verbal description supports the participation and integration of visually impaired persons by rendering audiovisual and visual information more understandable through language and hearing. The principal user[3] group of audio description consists of people who have different types and degrees of sight loss that cannot be corrected by ordinary eyeglasses. Some of these users are congenitally visually impaired, meaning that they have had reduced sight from birth, while others have become impaired only later in life. The causes for a loss of sight include ocular and other diseases as well as accidents. In Europe, sight loss occurring in adulthood typically relates to advanced age. (European Blind Union n.d.) Those with reduced sight also vary in terms of their visual perception. For instance, tunnel vision is a condition in which sight is reduced to the central area of the field of vision, and peripheral vision is lost. In some cases, on the other hand, central vision may be lost while peripheral vision remains. Other sight problems include blurred or patchy vision, resulting in dark spots in the field of vision. Overall, being blind often refers to a person perceiving at least some degree of light, as only approximately 4% of the blind have no light perception whatsoever. (Royal National Institute of Blind People 2013.)

With these distinct types and degrees of sight loss in mind, the function of audio description can be defined as being twofold. For the blind and for those with severe loss of sight, audio description is a capacitating aid, substituting for visual perception, whereas for people with milder degrees of low vision, audio description supports their own visual perception.

[3] In contrast to the varied terminology in my research articles, this thesis summary applies the term 'user' (in place of 'recipient' or 'audience') to refer to those who benefit from audio description and use it, be it in a more passive form (such as watching a film, in which case the communication in unidirectional), or by assuming a more active role (such as using audio description in a lesson and as part of interaction).

Navarrete Moreno (1997: 70) provides details of this from the standpoint of film by stating that audio description constitutes a supplementary aid for those who are already familiar with cinematographic art, and introduces this art form to those who have thus far been unable to appreciate it. In addition to its cultural relevance, audio description presents socially relevant knowledge by describing nonverbal, visual communication such as habitual gestures (Hernández-Bartolomé & Mendiluce-Cabrera 2004: 266).

Various situations, texts, and discourses can benefit from audio description (see Hirvonen 2013c for a review in Finnish). For example, audio description can transmit the visual aspects of communication and the physical environment, such as the appearances of the guests and the venue of a wedding ceremony. Audio description can also support learning by verbalising both the lecturer's body language and the educational material. In addition, verbal descriptions can convey the visual information of natural or cultural environments such as landscapes, the flora and fauna in a national park, or the architecture in a city. Audio description can likewise verbalise visual media and art forms, including works of art, advertisements, and museum artefacts. It can also complement the auditory perception in appreciating audiovisual media and art, which refers not only films, but also to television programmes and advertisements as well as to the performing arts such as theatre, opera and dance. As media and art are increasingly placed online, audio description is also found and used in digitalised museum collections[4] and in films and television programmes that are watched or downloaded online (see ADLAB 2012: 59). Another factor is that the newest technological developments affect the use of audio description in live performances. The traditional method of transmitting the description live has in some places been replaced by a pre-recorded audio description that is downloaded in hand-held devices and received through them.

Audio description generally conveys the visual aspects of objects, characters or people, settings, and action. When necessary, the description clarifies auditory information, such as the sources of sound, and reads aloud textual information. Audio description can occur in real time, as in a theatre play, or it can be recorded as a sound track for a film. For both these formats, however, an audio description script needs to be prepared beforehand. By comparison, in spontaneous interaction, the audio description likewise reflects this spontaneity. In some situations, users can influence the content and style of an audio description by intervening with questions or instructions. This occurs, for example, when the

[4] Audio-described artefacts at the British Museum in English (https://www.britishmuseum.org/explore/online_tours/museum_and_exhibition/audio_description_tour/audio_description.aspx) (accessed 6 October 2014), and at the Finnish National Gallery in Finnish (http://www.ateneum.fi/en/node/1076) (accessed 6 October 2014).

audio description of an exhibition is conducted face-to-face with a small group of visitors. The verbal-vocal communication of an audio description can be supported by other means, such as by drawing on a user's body, or by touching the objects that are being described.

The audio description technique was initiated in the United States in the 1970s and 1980s.[5] The initial stages in developing the audio descriptive technique in the 1970s were created by Gregory Frazier in his research on the adaptation of film for a blind and visually impaired audience. A few years later, Margaret and Cody Pfanstiehl launched an ongoing service of theatre audio description that became popular throughout the US in the 1980s. This service subsequently expanded to different areas, ranging from television programmes to museums. (Benecke 2014: 11–12.) In the 1990s, audio description was adopted in Europe (Hernánzed-Bartolomé & Mendiluce-Cabrera 2004: 267), where the audio description of film and television became firmly established particularly in Spain, Britain and Germany. Some countries adopted audio description as a regular practice in television and theatre at an early stage, whereas others are currently struggling to implement widespread and consistent usage. In Germany, for instance, the Bavarian Broadcasting began regular transmissions of audio-described films in 1997 (Benecke 2014: 13).

In Finland, the Cultural Service for the Visually Impaired, henceforth referred to as CSVI, began to audio describe events in the 1980s (Aaltonen 2007).[6] The first genres that were audio-described in Finland during the 1980s and 1990s were theatre, sport, fine arts and cinema. By 2010, several museum exhibitions, theatre performances, and some films had been audio-described. By the time this dissertation went to publication, four films have been broadcast on Finnish television (*Varpuset* in 2005, *Virginie* in 2009, *Tauno Tukevan sota* in 2010, and *Puolin ja toisin* in 2013), and two films provide an audio description on DVDs (*Postia pappi Jaakobille* in 2009 and *Risto Räppääjä ja polkupyörävaras* in 2010). In addition, a vast number of other Finnish films have been audio-described in the film clubs of the visually impaired, in which case the audio description has been transmitted live. Overall, the Finnish audio description has developed under 'environmental description' or simply 'description' (in Finnish, *kuvailu*) (Lahtinen et al. 2009; Lahtinen & Palmer 2012), which refers to the inter-sensorial communication in which information is conveyed between distinct senses (Lahtinen & Palmer 2012: 107). Audio description, or 'voice description'

[5] There have also been other forms of verbal description developed for a general audience, such as the radio narrators who described movies in the 1930s (Orero 2007; Pujol & Orero 2007), but they can be thought of as 'precursors' of audio description. Furthermore, a type of spoken discourse known as 'description in conversation' (*Beschreiben im Gespräch*, Stutterheim & Kohlmann 2001) shares features with audio description.

[6] An overview of the Finnish situation can be found (in Finnish) on the website of CSVI (see http://www.kulttuuripalvelu.fi/?id=151 (accessed 6 October 2014)) and in Aaltonen (2007).

(*äänikuvailu*), refers to the verbal-vocal translation of visual information and is therefore a subcategory of description, along with the various types of description in spoken, written and sign languages, haptics, and other methods (ibid. 109).

Today, media accessibility is advocated by the European Union, and this puts pressure on national authorities to introduce and enforce legislation and regulation that concern audio description (see Mazur & Chmiel 2012: 5). The Audiovisual Media Services Directive, which was regulated by the European Union in 2007 and amended in 2010[7], states the following:

> The right of persons with a disability and of the elderly to participate and be integrated in the social and cultural life of the Union is inextricably linked to the **provision of accessible audiovisual media services**. The means to achieve accessibility should include, but need not be limited to, sign language, subtitling, **audio-description** and easily understandable menu navigation. (AVMSD-2010/13/EU, Article 46; emphasis added.)

While the Directive is apparently being implemented in many EU countries, with the regulation of legislation relating to the accessibility of television broadcasting, the results of this implementation vary. For example, Britain has a fixed quota for the provision of audio-described television programmes (Greening & Rolph 2007: 128), while Germany (see Weißbach 2013: 354) and Finland (see below) have the legislation, but without a fixed quota for audio description. Despite this disparity, German as well as British television channels provide audio-described programmes regularly, and an extensive selection of audio-described cinema is available on commercial DVDs. Furthermore, recent modifications in the financing system of the German broadcasting and film industry are likely to contribute to an increase in the provision of audio description (Benecke 2014: 13). In Finland, although the provision of audio-described programmes is rare, the Audiovisual Media Services Directive is reflected in the legislation. For instance, an amendment to the law enacted in 2010 instructs the Finnish public service broadcasting company, Yleisradio (YLE), to render its programmes accessible to hearing and visually impaired users. A review of the wording of this law, however, reveals these users only have partial accessibility. The amendment orders that Finnish- and Swedish-speaking television programmes must be supplied with subtitling, and other programmes must provide a commentary or a service that makes the programme's subtitles audible (FINLEX 733/2010[8]; my translation from the following Finnish statement: 'Suomen- tai ruotsinkielisiin televisio-ohjelmiin on liitettävä tekstitys sekä muihin ohjelmiin selostus tai palvelu, jossa

[7] The Directive, AVMSD-2010/13/EU, can be found at the EUR-Lex portal: http://eur-lex.europa.eu (accessed 6 October 2014).

[8] The law amendment can be found at: http://www.finlex.fi/fi/laki/alkup/2010/20100733?search%5Btype%5D=pika&search%5Bpika%5D=televisio (accessed 6 October 2014).

tekstitetyn ohjelman teksti muutetaan ääneksi.' Hence, the two services for the visually impaired, the commentary (or audio description) and audio subtitling, are presented as alternatives. This may, in turn, lead to broadcasters considering themselves to be obliged to provide only one of the services and not both. This, in fact, seems to be the case because audio subtitling is widely offered, but audio description is all but non-existent on television. While the Audiovisual Media Services Directive states that media services should be made accessible gradually, the Finnish interpretation of this seems to be that, during the first stage, the legally required services of accessibility are the subtitling for the hard-of-hearing and deaf viewers, and the audio subtitling for the visually impaired.

2.1.2 Audio description as translation

Even though audio description can be regarded as peripheral translation (see Chesterman 2004: 73), it is currently established as a type of audiovisual translation (see the entries 'Audiovisual translation' in the *Handbook of Translation Studies* by Remael 2010 and in the *Routledge Encyclopedia of Translation Studies* by Pérez González 2009). Audiovisual translation evokes the multimodal nature of the text or discourse to be translated. Multimodality means that more than one type of sign system and channel are present, which implicates the 'constrained' nature of audiovisual translation; the multimodality of communication sets certain boundaries for the translation, such as the need to compress or paraphrase the original dialogue in subtitling (Remael 2010). Audiovisual translation has become an umbrella notion for 'multisemiotic transfer', which according to (Orero 2004: viii), comprises all translations 'for production or postproduction in any media or format, and also the new areas of media accessibility'.

In addition, audio description is defined as intersemiotic or intermodal translation (see also 'multidimensional translation' in Gerzymisch-Arbogast (2005) and 'partial translation' in Benecke (2014)). The attributes 'intersemiotic' and 'intermodal' pinpoint the change of mode that occurs in audio description. In his renowned classification, Jakobson (1959) asserted that, in addition to inter- and intralingual translation, there is intersemiotic translation or transmutation from one system of signs into another. The importance of this classification lies in recognising that translation does not occur merely between languages (Snell-Hornby 2006: 21). Braun (2008: 15-16) combines the definitions by other researchers and characterises audio description as intermodal translation. In the present study, I also adopt the term 'intermodal' to define audio description as translation because the term entails the idea of 'change of mode' or 'between modes'. Mode, on the other hand, is used to refer to both 'channel' (vision, hearing) and to the 'semiotic system' (images, language) (cf. Muntigl 2004: 40–41; Stöckl 2004b: 11), whereas the term 'intersemiotic' refers only to the change between systems of meaning making.

2.1.3 Audio description in the filmic context

The audio description of a film is typically an extra narration that is scripted, pre-recorded and edited so that it is rendered in the dialogue-free slots of a film. Thus, filmic audio description[9] is a type of 'audio-text', a text that is written to be spoken out loud (Gutenberg 2000), and it exists both in a written and a spoken format.

What, then, does an audio description convey from the filmic source material? In reviewing a set of audio description guidelines, Vercauteren (2007) lists the following:

- *Images*: Information is given with regard to where the story is taking place, what is happening and when, and who are acting and how. (Vercauteren 2007: 142.) This can contain references to characters (such as characters' appearances and emotional states), actions as well as objects and locations (Salway 2007: 155-156).
- *Sounds*: When sound effects are difficult to identify, they can be referred to in the audio description. Silence is also a factor: on one hand, it renders space for audio description, but on the other, not all silence should be filled with description. (Vercauteren 2007: 143.)
- *Text*: Text that appears in the film image includes background information of the film (logos, opening titles, cast lists and credits) as well as story information, such as text in the setting (like road signs or advertisement) and subtitled speech. (Ibid.)

Cinematic discourse – how films express stories by auditory and visual means (Chatman 1978: 26) – sets constraints on audio description and poses certain challenges (see also Braun 2008: 16–18). The auditory context constrains audio description in the sense that the film soundtrack designates the time available for description (for instance, the location and length of the slots in which audio description can be uttered). Furthermore, the technical process of recording an audio description affects the quality of the end product, including its intelligibility. As concerns intermodality, three aspects present challenges. The first challenge is to select the information that is to be verbalised from the holistic film image that offers a variety of visible details. The second challenge is that the visually simultaneously presented information must be made linear in the linguistic representation. The third challenge is that the realistic or iconic representation of filmic imagery must be rendered abstract and conceptual in the verbalisation. The present work aims at contributing to the dilemma of intermodality by proposing ways to meet these challenges. The auditory context of film is

[9] A note on terminology: 'filmic audio description' is used in the present analysis to refer to the audio description of cinema. Alternatively, 'film audio description' can be used (see Jimenez Hurtado & Soler Gallego 2013).

also analysed, not as a constraint, but as an anchor that 'lightens the audio describer's workload since a great deal of information that is visual is also comprehensible through the sound' (Benecke 2014: 9).

2.2 Research on audio description

Audio description has been the focus of intensive investigation in translation studies since the turn of the century. The main objective of this chapter is to review some of the literature, first (2.2.1) by surveying briefly the general trends, and then (2.2.2) by explaining how the previous studies on filmic audio description have inspired the present work. The overview in 2.2.1 is based on my review article in Finnish (Hirvonen 2013c) and on the doctoral dissertation in Catalan by Cabeza-Cáceres (2013).

2.2.1 Overview of the state of the art

There is a large volume of published studies on the practice, functions and characteristics of audio description (see Cabeza-Cáceres 2013: 58f.). These include reports on the practical implementation in different countries or continents, such as Hernández-Bartolomé & Mendiluce-Cabrera (2004) on Spain, Greening & Rolph (2007) on Britain, Seibel (2007) on Germany, and Orero (2007) on Europe. A considerable number of guidelines for the preparation and delivery of audio description have also been published (for instance, Ofcom (2000) for Britain, Dosch & Benecke (2004) for Germany, and the AENOR standard (2005) for Spain). These are presented in Vercauteren (2007) in a review of the general aspects of audio description. At present, guidelines continue to evoke interest, and research is currently being undertaken to explore the possibility of creating standardised international guidelines for audio description (see Mazur & Taylor 2013). The different genres of audio description are also discussed in the literature. Apart from the audio description of film, which is explored in the next section (2.2.2), studies such as Corral & Lladó (2011) report on the audio description of opera, Márquez Linares (2007) as well as Quereda Herrera (2007) on educational interaction, De Coster & Mühleis (2007) on art, and Cámara & Espasa (2011) on the audio description of scientific multimedia.

The didactics of audio description are also analysed in the literature (see Cabeza-Cáceres 2013: 49f.). For example, Remael & Vercauteren (2007) address the issue of teaching students what elements to choose from a film to an audio description. Scholars (Orero 2005) and practitioners alike (Hyks 2005; Navarrete Moreno 1997; Snyder 2008) discuss the competencies of an audio describer. In addition, audio description can serve as an educational tool, enhancing, for example, the linguistic capacities of blind children (Palomo López 2010).

A number of studies also analyse the language of audio description. For instance, corpus studies on audio description scripts familiarise us with the aspects that are linguistic (Jiménez Hurtado 2007; Salway 2007) and multimodal (Jimenez Hurtado & Soler Gallego 2013). Another method used to examine the linguistic elements of audio description is contrastive translation analysis in which parallel audio descriptions in distinct languages are analysed and their differences are highlighted (Bourne & Jiménez 2007). The same method is used to study the translation of culturally specific items in different-language audio descriptions (Matamala & Rami 2009). The issue of translating audio descriptions scripts from one language into another has produced preliminary analyses (Bourne & Jiménez 2007; Remael & Vercauteren 2010) and a PhD dissertation in Poland (see Jankowska & Szarkowska 2013). Some studies incorporate the framework of narratology (Kruger 2010) and discourse studies (Braun 2007) and these orientations provide analytical tools for the study of audio description (see Cabeza-Cáceres 2013: 73–76).

Finally, the reception of audio description continues to draw attention and it has been studied with respect to different genres such as the theatre, opera and film (Cabeza-Cáceres 2013: 76f.). Szarkowska (2011) and Szarkowska & Jankowska (2012) test the use of speech synthesis as a substitute for human voice in delivering audio description. Apart from the more traditional methods of reception research – surveys and interviews – audio description research also employs more novel tools such as eye tracking, which is used to study aspects such as the relation between viewer activity on screen and audio description (Orero & Vilaró 2012). An additional method is proposed in the doctoral dissertation by Cabeza-Cáceres (2013), who applies experimental research using different parameters to test the reception of film audio description.

A central lesson learned from this review seems to be that to adequately explain and describe audio description, the application of distinct disciplines and theoretical frameworks is required. For example, analysing audio description from the perspective of narratology focusses on those characteristics of audio description that are related to narration. At the same time, using analytical categories from film studies or linguistics highlight other systems of meaning making. Highlighting differences and mixing these approaches can also result in understanding the similarities between the different meaning-making systems, such as between the audiovisual and verbal representation.

2.2.2 Introducing research on filmic audio description

Films have specific features and audio description must take them into account (see Cabeza-Cáceres 2013: 62f.). These features include opening credits, which are analysed in Matamala & Orero (2011), as well as characters (Fresno 2012) and their nonverbal communication

(Igareda 2011). Vercauteren (2012) discusses strategies for filmic audio description in terms of the narrative dimension of time. Another important feature is the original soundscape of the film – the sound effects, music and dialogue. Fryer (2010) has analysed the functions of sound in audio drama and argues that understanding how sounds acquire their meaning in audio narratives can be beneficial for audio description because sounds and other effects render a feeling of realism or evoke emotions in the hearer. Moreover, the potential relationships between film sound and audio description are discussed in Remael (2012), and the role of music is examined in Igareda (2012). Let us now turn our attention to review the research on the filmic audio description that is directly related to the aspects that this thesis attempts to examine.

This dissertation adapts some of the analytical and theoretical frameworks of other scholars to conduct systematic analyses of filmic audio description. Firstly, one objective of this study is to adopt a broader, more comprehensive perspective on the issue of *Räumliches hören* [to hear space] (Seiffert 2005). This is accomplished by elaborating on the (mainly) linguistic analysis that is conducted in Seiffert, and by examining the multimodal constitution of space through the verbal-auditory resources that are effective in audio-described film (Article 1).

Secondly, this analysis examines space from the viewpoint of filmic representation (Articles 2 and 3). Hence, this approach is source-text oriented. The issue of determining means to reflect the visual representation linguistically in audio description is addressed by different authors in Fix (2005). These authors illustrate a range of possibilities and foci of interest that arise in when analysing filmic audio description, as well as provide analytical tools for the present research (see Seiffert 2005 and Kluckhohn 2005 in the next paragraph). However, while these studies do not demonstrate filmic representation by disseminating it non-linguistically, I have decided to focus on the audiovisual aspects more closely by the pictorial data presentation and the careful account of the analysed examples. This type of approach to both visual and verbal data constitutes the basis of intermodal analysis. An example of an intermodal comparison of filmic and linguistic representation is Pérez Payá (2007). Her study explores the 'language of audio description' from the perspective of the 'language of the cameras' and in terms of three levels of narration. The first level is of the *mise-en-scène* or 'setting the scene', which includes narrative elements such as characters and places, and these can be compensated for by an audio description that has an adequate vocabulary (Pérez Payá 2007: 87). On the second, shot level, narrative elements are emphasised differently, and this can be imitated in audio description by describing the larger elements before the smaller ones, and the mobile elements before the static ones (ibid. 88). Finally, at the editing level, audio description can reflect the narrative construction by indicating aspects such as the temporal relations with temporal expressions such as 'later' and 'on the next day' (ibid. 89).

The linguistic techniques that are used to substitute for the visual means of representation are reported in Seiffert (2005) and Kluckhohn (2005). Seiffert (2005) observes that one explanation for how the linguistic mode can function simultaneously as the image is the use of keywords to trigger schematic knowledge in the receivers' minds. An example of this is the identification of the essential properties of space (ibid. 69, 84). Furthermore, schematic knowledge contains stereotypical representations or models in long-term memory that relate to situations and actions (for instance, see Schwarz 1992: 84–86). By analysing the audio description of the space for action (*Handlungsraum*), Seiffert demonstrates how distinct linguistic expressions have distinct potentials in terms of triggering schemata. For instance, the *Büro* [office] potentially evokes a schema that involves typical objects that belong to an 'office' (a computer and desk) as well as the typical action that occurs in an office (office work and meetings). By comparison, *Raum* [room and space] is semantically more abstract and therefore less effective in evoking schemata (Seiffert 2005: 76–78). In addition to audio description, information on space is provided through film dialogue (for example, when an object is explicitly mentioned in it) and through sound effects (such as a telephone ringing) (see ibid. 72).

Another relevant aspect is the structure of language. One question is how the linear organisation of the linguistic mode corresponds to the filmic composition of space. Kluckhohn (2005) presents one departure point by analysing the syntactic expression of audio description and arguing that it can compensate for the visual arrangement of a film. Kluckhohn focusses on three areas of syntactic manipulation – information structure, specificity, and word order – and all of these relate to how elements are represented as either 'known' or 'new' to the discourse and how they are anchored to the recipient's knowledge and to each other. Since a key property of filmic audio description is condensation, that is, expressing elements in a concise manner (see also Hyks 2005: 6), Kluckhohn poses the important question of how filmic imagery can be transformed into short texts that are as expressive and informative as possible (Kluckhohn 2005: 49). To illustrate, audio description can encode concepts variously as either specific or general. An element can have a definite form (for instance, *der Fotograf* [the photographer]) because it has been mentioned previously in the discourse, but also because the information that makes it familiar can be auditory (for example, the sound of a camera releasing can be heard) (ibid. 57). Another example is related to the manipulation of 'standard' word order, such as topicalisation, that can be used to emphasise a character or to signal a change of scene so that the sentence-initial position has an adverb that is used to describe a facial expression or a locative prepositional phrase to indicate a location (ibid. 61).

Thirdly, this dissertation focusses on comparing the visual and verbal means of representation that are relevant to filmic audio description. As an example, Braun (2007) introduces two

approaches that audio description can have in relation to filmic discourse. By adopting discourse analysis and the relevance theory proposed by Sperber & Wilson (1995), Braun (ibid. 6) contemplates the difference between using explicatures ('assumptions which the speaker wants to communicate explicitly') and implicatures ('assumptions which the speaker encourages the addressee to make implicitly', that is, to be deduced from the verbalisation). Braun illustrates this difference by examining the opening scene of the film *The Hours* (2002), which entails a close-up of a woman's blouse and of her hand holding a pen as she uses it to write something on a desktop (Braun 2007: 3). These elements 'provide the cues for identifying the 'factual' content of this shot', an action that depends on recognising the individual items (pen, desktop, woman or woman's blouse, and writing). The explicature of the scene could express that a woman is sitting at a desk writing something, whereas an implicature of the scene goes on to interpret the cues, such as 'she is writing a letter or diary'. Braun analyses the scene further by noting that although the English audio description verbalises the explicature by stating *she sits writing*, 'leaving it to the audience to draw implicatures regarding the purpose of the writing', it nevertheless omits the individual cues that prompt this explicature. (Ibid. 8.) The connection between Braun's contribution and the results of this dissertation is discussed further in Chapter 5.1.1.

Finally, it should be noted that while the present research has been in progress, other scholars have reported similar research interests in audio description. For example, Tanis Polat (2013) investigates the concept of space by comparing the representation of space in German and Turkish audio descriptions. In addition, Remael (2012) explores the relevance of the film sound in audio-described film.

2.3 Cognitive orientation to film and language

As this thesis explores the similarity between visual and linguistic representations, a relevant framework is offered by the cognitively oriented theories of film and language. As Seiffert (2005) demonstrates, schema theory can explain how the audience of audio-described films can mentally reconstruct filmic space through appropriate schemata, which are cognitive representations that are triggered by keywords from the audio description (in addition to the cues from the filmic soundscape). Schemata are also central to the cognitive, or constructivist theory of film (Bordwell 1985, 2010). According to this orientation, spectators are considered as coming 'armed and active to the task' in their perception of a narrative film and they employ schemata to make sense of the story on the basis of visual, auditory and narrative cues that are furnished by the plot and film style. (Bordwell 1985: 38–39, 50.)

2.3.1 Cues and schemata in the filmic representation

In cinema, space is cued in many ways. Firstly, space is depicted within a frame with four borders that define the amount of space that is visible; this is the 'shot space' (Bordwell 1985: 113). By comparison, the sense of space that spectators have is often larger. Thus, shot space implicates 'off-screen space' beyond the frame by virtue of the whole-part relationship. An example of this would be the parts of a desk and a chair serving as cues for an office room, with a room being a cue for a whole building, and a building in turn being a cue for a neighbourhood or an entire city. This spatial map is constructed because we assume that the space continues beyond the frame (see Bordwell & Thompson 1979/1990: 173). Sometimes this off-screen space becomes concrete through sound, as in the traffic noises on a street that are audible in a scene that is set inside a building. As a result, a 'sonic space' is constructed, with foregrounding and backgrounded sound events (Bordwell 1985: 118–119). Finally, the changes between the shot space and off-screen space is achieved in editing by joining shots together, by controlling the length of the individual shots (Monaco 1977/2009: 242), and by coordinating the separate shots with each other (Bordwell & Thompson 1979/1990: 207) This 'editing space' (Bordwell 1985: 117) makes filmic representation dynamic in the sense that off-screen space may become shot space, or vice versa.

Bordwell (1985) proposed that films function on the basis of schemata that are triggered by cues. Schemata are recurrent networks that summarise multisensory and multimodal knowledge and past experiences, providing a framework for the acquisition, interpretation, and retrieval of new information (Carlston & Mae 2001: 13526–13527)[10], exemplified by the inference-making and hypothesising that occurs in film narration (Bordwell 1985: 37). For instance, applying a prototype schema of a doctor's office produces hypotheses about the typical décor and the action that takes place in that office. Cues, on the other hand, can be understood as information that perceivers use to construct a narrative (ibid. 35). In other words, making sense of space requires interpreting visual cues as objects and scenes so that the spectator can construct 'a cognitive map' (see ibid. 117). In addition to this data-driven (or bottom-up) process, however, concept-driven (top-down) processes operate as well, and the perception in the concept-driven process is guided by prior knowledge and experience (Whitney 2001: 13525, see also Bordwell 1985: 101). Schemata therefore provide 'defaults for filling in missing information' (Carlston & Mae 2001: 13528) and schematic knowledge is summoned when a setting is constructed through the identification of typical elements (for instance, a monument alluding to a particular city) and when, in order to interpret a sound, spectators make inferences on the likely source of the sound. According to Goldstein

[10] Although the literature has distinct terms for the mental representations of concepts, events and activities, this dissertation uses 'schema' to refer to all of these (see Whitney 2001).

(2007/2010: 11), the role of concept-driven processing increases as input becomes complex. This means that when we perceive real-world scenes (or analogies of these, such as film scenes), 'our knowledge of how things usually appear in that environment can play an important role in determining what we perceive' (Goldstein 2007/2010: 11).

Hence, although film viewing is based on general cognitive principles, cues are interpreted according to one's prior knowledge and experience, that is, in addition to one's social and cultural background (see Bordwell 1985: 32, Bacon 2005: 8f.). If we consider colours and the types of connotations or associations that we harbour of them, they depend, at least in part, on the cultural context (see Monaco 1977/2009: 194). Film viewing can similarly be understood as involving general cognitive principles, in that it mimics the perception of the real world, but also as resorting to an aesthetic system of a given culture (Bacon 2005: 10). In short, films both present and narrate. On the one hand, they transmit visual, auditory and vocal stimuli; on the other, they furnish cues that are based on narrative conventions provided by the cinematographic medium and stylistic conventions (Bordwell 1985: 101). These conventions comprise different film techniques such as the use of camera in implying glances between characters or the use of blurry focus and unstable camera when depicting drunkenness (see Branigan 1984: 98). While many conventions resemble or imitate prototypical real-life experiences, some visual effects are more clearly 'representational codes' that do not have a direct correlate in real-life experience (for instance, for wipe effects in which the existing frame is wiped out by a new image, see Prince 1993: 23).

2.3.2 Cues and schemata in the study of audio description

The notion of cue is relevant both to audio description and to communication in general.[11] In discussing filmic audio description from a discourse point of view, Braun (2007) observes that 'verbal, visual and auditory cues', together with prior knowledge and the associations they evoke, are the basis of the mental models that spectators create on the basis of an audiovisual discourse (ibid. 3). In a similar fashion, Remael & Vercauteren (2007: 73) attempts 'to show how a better insight into film narration and a better perception of the many visual clues used by the filmmaker might help audio describers decide on what information to prioritize'. Here 'clue' basically refers to a 'cue' relating to the information that spectators 'pick up, link up and interpret, inferring information that is not explicitly given' (ibid. 77 following Branigan 1992). In addition to the aforementioned studies, a similar position is

[11] People use cues, or clues, in communication and in linguistic interaction (Brown & Yule 1983: 4, 190; Gumperz 1982: 131), including translation (Gutt 1991: 127). The use of cues indicates that the communicative or narrative information expressed by the speaker/writer is often incomplete and must be actively processed by the hearer/reader in order to (re)construct the message.

discussed in Vandaele (2012) as a 'trigger'. Thus, making a 'cue' relevant to audio description seems to involve the assumption that audio description enables an active participation in story construction and is therefore neither necessarily, nor merely, a ready-made interpretation of the film's visuals. Remael & Vercauteren (2007) observe the following:

> All audio describers can therefore hope to do is provide clues that will allow their special target audience to construct a story that makes sense and offers them satisfaction. In order to accomplish this, the describer must allow the blind or visually impaired public to make the necessary inferences from the *plot* of the film, in order to construct its entire *story*. (Remael & Vercauteren 2007: 78.)

Audio description can therefore entail communication between two minds. The first is the mind of the viewer or the audio describer(s) who verbalises visual cues. The second is the mind of the audio description user who constructs a cognitive map and a story on the basis of these verbalisations and of the narrative and auditory cues of the film. Within this network of minds, language, or audio description, is a 'vehicle for vision' (the expression is from Bordwell 1985: 8), a mode used to express and mediate the cues from the original work. Consequently, audio description users can do nothing but trust the describer's (=speaker's) ability to intelligently and sensitively translate what they see into what they speak; this is a state of affairs that dictates translation in general (for instance, see Bassnett 2011: 22).

The main objective of the present dissertation is to analyse how the linguistic representation in audio description cues the extra-linguistic world of film. Appropriate conceptual tools for this endeavour are offered by the cognitive orientation to linguistics that analyses language 'as reflections of general conceptual organization, categorization principles, processing mechanisms, and experiential and environmental influences' (Geeraerts & Cuyckens 2007: 3). In this regard, one key aspect of language is schematisation and categorisation that resort to schemata in representing the extra-linguistic experience in words and structures. Schematisation orients to language as consisting of conventionalised forms and meanings that are based on schematised usage events. Categorisation, on the other hand, involves interpreting this experience with respect to previously existing structures. (Langacker 2008: 3.1.) Of the essence here is the specificity of words and expressions (from superordinate to basic and subordinate levels, ibid.; see also Seiffert 2005) and the context in which they occur. These factors trigger different types of schema and present the visual information or usage event with more or less precision.

Concerning the organisation of the extra-linguistic world through language, a premise I adopt here relates to the notion of the iconicity of language. In the light of intermodal similarity, linguistic iconicity is highly promising. Although the linguistic sign is generally considered to

be symbolic (for instance, the word *dog* as representation does not capture the form of real dogs, in the same manner as a picture of a dog), the principle of iconicity covers a set of aspects in which language can be iconic to the extra-linguistic world (see van Langendonck 2007). To illustrate, onomatopoetic words have phonetic forms that resemble the sound of the referent (for instance, the singing of birds is verbalised as *chirp*). While some types of iconicity have been criticised as having other motivations than iconic representation (such as frequency, see Haspelmath 2008), one type that has received more approval and that is relevant to audio description is the 'iconicity of sequence'. This means that the sequence of the linguistic references matches the sequence of the experiences they conceptualise (ibid. 3, see also Langendonck 2007: 407). As a result, word order can be iconic in the sense that it reflects the temporal order of perceiving the events (ibid.), such as the order in which film shots present space. It is necessary to note, however, that besides the order of perception, other global and local motivations may also determine the linguistic form, such as relevant discourse aspects and the theme-rheme construction.

The third framework applied in this thesis to explore the similarity between image and language is the theory of Figure and Ground.[12] This theory explains how the organisation of space in both the language and the visual (as well as auditory) perception is at one point determined by the assignment of the spatial field into separable, thing-like Figures and into a more formless and substance-like Ground. Whereas the Figure and Ground theory is specifically focussed on in Article 4, schematicity, categorisation, and sequential iconicity are repeatedly applied in the articles. However, many additional aspects of the cognitive orientation to language have yet to be analysed (see Vandaele's list of possible analytical categories, 2012: 96–97).

[12] The theory of Figure and Ground is also applied by Schubert (2009) to explore the linguistic representation of space in descriptive texts, and, more recently, by Mäkisalo & Lehtinen (2014) to explain shifts in interlingual translation.

Part II

3 Results

This third chapter summarises the main results from the four research articles. The results are presented in terms of the research problem and the two challenges of audio description. These challenges concern the question of the multimodal representation of space in audio-described film as well as the question of intermodal similarity between the visuals in a film and the linguistic representation in the audio description. In this chapter, 'cues' refer both to the semantic-referential level (the type and parts of space that are referred to) and to the structural level (how these references are formed and organised). The core question in this dissertation pertains to how language is used to cue visual filmic space. This means that 'language' is understood here as the form of speech because the vocal level has proven to be relevant. Furthermore, in contrast to Part I, this second part refers to the articles by using a bibliographic reference (author/s, year, pages) in order facilitate the search in the original studies.

3.1 Multimodal representation of space in audio-described film

In Hirvonen & Tiittula (2012), we first investigated the interplay of the auditory modes and how they constitute a multimodal representation of space. In addition, we attempted to determine the dimensions of space and how they are cued by the different modes. The analysis revealed the presence of the following auditory modes:

- Music, both as a jingle and as film music.
- Speech and voice acting by film characters, voice-over, and the audio describer (or the voice talent narrating the audio description).
- Sound effects.

The cues that occur in these modes can be considered as resources for the constitution of the imaginary space. They therefore display a potential that the users employ individually during their film viewing. These cues carry information on the different dimensions of space, including the setting or the physical environment as well as the space between and around characters in which these act and interact (Hirvonen & Tiittula 2012: 424). Furthermore, these cues reflect varied foci of interest for the space (ibid. 425). As they are acoustic, the auditory modes have certain general properties that influence the constitution of space. The perception of spatial depth is affected by factors such as volume and how sound travels in space, creating

near-distance relations and information that is foregrounded and backgrounded (see ibid. 419 and Hirvonen 2013b: 100).

Another general property is that, in addition to auditory modes representing objects, locations and actions and their relations, these modes carry meanings that are connotative, associative and emotive. One example would be film music, which affects the constitution of the story by establishing the continuity between scenes and by contributing to formulating hypotheses about the storyline (Hirvonen & Tiittula 2012: 419–420). Speech, on the other hand, is multifaceted in the sense that it has an iconic, bodily level of expression for aspects of communication such as actions and moods through voice, as well as a symbolic level for aspects, or extra-linguistic referents, that it denotes (ibid. 420–421).

3.1.1 Interplay of the auditory modes

The verbal and sonic dimensions support and complement each other in audio-described films. The result is that just as audio description can confirm or steer the interpretation of a sound, a sound can make a verbalisation more concrete and lively. Sound and voice also furnish iconic cues for aspects such as the character action, for instance, whereas a verbalisation identifies this action. This can be illustrated by a sequence from Hirvonen & Tiittula (2012: 411) in which voice acting (a grunt), a sound (a short creak) and an audio description (a verbal description in German of a character stumbling down, *und stolpert* [and stumbles]) are heard consecutively. To express this conversely, the verbal-symbolic dimension specifies otherwise ambiguous acoustic and vocal cues (ibid. 423).

Indeed, auditory modes are organised both simultaneously and consecutively. When several resources are audible simultaneously, audio description may surface as the most prominent sound, relegating sound effects and music to the background. Nonetheless, the simultaneity of the modes points to a multitude of spatial elements and presents them as relevant. Some aspect of the setting might merely be audible through sound (a rustle of leaves on the ground), while another might be described verbally (*der Schlagbaum geht hoch* [the barrier (of the gate) is lifted up] (Hirvonen & Tiittula 2012: 406–407). Furthermore, when one mode pauses, the other modes receive more attention; for instance, a pause in audio description foregrounds a momentary sound event (ibid. 411). Another example is continuous music that produces a continuity between audio descriptive utterances, and a continuous sound effect implies spatial continuity (the ‘rustling’ sound maintains an impression of movement in a ‘forest’ setting, ibid. 413). However, sounds can also be discontinuous and consequently imply a change in the spatial composition, such as the change from a distant view to a close-up (Hirvonen 2013b: 102). However, a change from one sound to another may also imply a change of scene (Hirvonen 2012: 33-34).

3.1.2 Cues of space in speech

As is evident from the analysis of the first two minutes of *Der Untergang* (Hirvonen & Tiittula 2012), speech and voice can be central elements in audio-described film and originate from a variety of sources (including the audio describer, film characters or actors, and voice-over). Distinct sources can be identified by their voices through intonation and voice quality. In fact, one purpose of describing the film credits and the film company logo at the beginning of the film is to introduce the voice and speech style of the audio describer to the receiver (ibid. 388). The acoustic voice properties also provide cues as to the location of the speakers, whether they are near or far with respect to the point of audition (ibid. 399–400). Furthermore, a clear voice quality foregrounds the audio describing voice and supports the listener's access to it and, as a consequence, also to the visual narration of the film (ibid. 421).

On the referential level, audio description serves to cue the different dimensions of space by using concepts that denote places, action, people, and objects as well as their qualities. Moreover, different conceptualisations present different aspects of the referents. For instance, the verbalisations *in einem kleinem Holzverschlag* [in a small wooden shack] and *in a makeshift shanty toilet* describe the same location, but while the German description denotes its physical appearance, a 'wooden shack', the English description profiles its function, a 'toilet' (Hirvonen 2013b: 111). In addition, concepts encode different levels of specificity. Let us compare, for instance, the descriptions *Himmler y Fegelein llegan a uno de los salones de la cancillería* [Himmler and Fegelein arrive at one of the halls of the chancellery] and *viele Militärs in einem Saal* [many officers/military people in a hall] (Hirvonen 2012: 32–33). The German audio description introduces the characters and the setting within a general category (officers or military people, and a hall). In contrast, the Spanish description identifies the characters and the setting in terms of their proper names (Himmler and Fegelein) or specificity (the chancellery).

Since concepts activate a network of knowledge in the user's mind, audio description can verbalise space by referring to some constituent part of it. This means that keywords, such as *Nachtschränkchen* [night table], can be used, and by belonging to a schema of 'bedroom', these cue a more complex and larger spatial setting, a 'bedroom location' (Hirvonen 2012: 34). Another means of cueing space by using schemata is to create schematic coherence. In this way, references can be introduced as 'known' elements without having been explicitly referred to previously, because they are linked by a common schema (for examples, *eine Schranke* → *der Schlagbaum* [a gate → the barrier], see Hirvonen & Tiittula 2012: 423, and *im Bus* → *aus dem Fenster* [in the bus → out of the window]; *the bus* → *out of the windows*, see Hirvonen 2013b: 101–103).

Another relevant factor concerns how audio description structures and linearizes information because the structural level affects the interpretations of the conceptualisations. For instance, the syntactic level of language assigns different functions to extra-linguistic entities, such as adverbs and prepositions that express relations and constellations (Hirvonen & Tiittula 2012: 422). As Hirvonen (2012) demonstrates, audio description can emphasise the setting, the interactional space or a detail in space by foregrounding these linguistically (see ibid. 28ff.). For example, topicalisation can be used for this type of emphasis. Topicalisation can also signal a change of scene by changing the theme (ibid. 38) or, conversely, create thematic cohesion between utterances and cue a continuity of space. To illustrate, the same element can be encoded alternately as both an object and a location: *der Kleinbus tuckert eine Straße entlang* → *im Bus sitzen Jamal und Salim* [the minibus is chugging along a street → in the bus, Jamal and Salim are sitting] (Hirvonen 2013b: 101–102). By contrast, the lack of cohesion or a change in topic in the audio description can cue a change in perspective (Hirvonen 2013a: 26). The vocal dimension of speech entails multiple cues as well. Pauses in speech give rhythm to the verbalisation and can also mark a change. Furthermore, the elements of space referred to in the audio descriptive utterances can be linked to each other through an intonation cue that signals continuity. (Hirvonen & Tiittula 2012: 411, 420–421.)

3.2 Intermodal similarity between visual and linguistic cues

Following the other focus of this thesis, this section describes the main results concerning the intermodal similarity that can occur in filmic audio description, that is, how the visually cued space is represented linguistically in audio description.[13] The objective of defining similarity was analysed by posing specific questions in three articles, and these questions concern the linguistic cueing of certain film techniques (shot distance in Hirvonen 2012 and the point-of-view shot in Hirvonen 2013a) as well as the linguistic representation of the composition of space in terms of Figure and Ground segregation (Hirvonen 2013b).

The spatial coverage of the shot, which is determined by the distance of the camera to the object filmed, can be reflected in the 'referential coverage' in the audio description. This means that the verbalisation refers to those elements that are visible in the shot. (Hirvonen 2012: 39.) Based on the different types of shot distance, I defined three strategies for cueing space: the setting-driven strategy, the interaction-driven strategy and the detail-driven strategy. If, for instance, the physical environment in which the story takes place is made focal by a long shot, the audio description can refer to the space by referring to a location (*im*

[13] I will concentrate here predominantly on the 'sameness' of the film and audio description, discussing their differences in the next chapter.

verwüsteten Garten hinter der alten Reichskanzlei [in the devastated garden behind the old Reich Chancellery], Hirvonen 2012: 29). When some property of space, such as the lack of illumination, dominates the image, this can be reflected by describing the property before revealing further aspects (*es dämmert* [it is getting light] or *in the gray light of dawn*, ibid. 30). Furthermore, as the interactional space in which characters act is made focal by medium-length shots, the audio description can first make reference to characters and their actions and only then expand the view to the setting (see the example of Himmler and Fegelein in 3.1.2). Similarly, audio description can focus on a detail in space. For instance, an object or a facial expression that is shown as a close-up can be translated into audio description by topicalising a reference to the object (*auf einem Nachtschränkchen* [on a night table], Hirvonen 2012: 34) or by topicalising a reference to the facial expression (*beklommen nähert sich Traudl…* [anxious, Traudl approaches…], ibid. 35). Framing the space through referential coverage may also be relevant in the situations that cue a transition from the general, neutral point of view in a film to a character's subjective perception or imagination. Here, audio description can focus on the character's facial expressions, eye movements or other actions that serve as visual cues of a transition to the character's point of view (Hirvonen 2013a: 29).

By verbalising the distinct aspects of a space by utterances that are demarcated by pauses, audio description can cue different points of interest in the film space, such as distinct locations and actions. In other words, audio description offers different points of view to the story world in the same way as film shots show different views. For instance, Hirvonen (2013b: 99–100) reports how one shot shows a minibus in its entirety, but the next shot narrows the space to a view of the windows and walls from the bus interior. While many visual cues of the film as well as schemata operate to inform the viewer of spatial continuity, the spatial continuity can also be cued in the audio description due to schematic coherence: *the minibus → the bus is full of… out of the windows* and *der Kleinbus → im Bus… aus dem Fenster* [in the bus… out of the window] (Hirvonen 2013b: 101–103).

Changes of the point of interest also occur in point-of-view shots as well as in the audio descriptions of these sequences (Hirvonen 2013a). Despite the name, the point-of-view shot consists of several shots that display a character's gaze and the object of that gaze, and this usually occurs in consecutive shots that are associated by various cues. Moreover, point-of-view shots represent a character's perspective somewhat overtly, and certain cues seem to be affirmative, while other shots entail fewer cues. This also occurs in audio description, in that the descriptions of the gaze and the object can be linked more or less to explicitness (Hirvonen 2013a: 28–30). The strategy[14] of the 'experienced character's perspective' presents the references to the gaze and the object, or stimulus, in syntactically separate and

[14] At present, I would rather refer to these as the 'techniques' of translation than as 'strategies'.

independent utterances, which makes the link between the perception and the stimulus less explicit. For example, *im Trichter öffnet Peter die Augen und kuckt sich verwirrt um* [in the crater, peter opens his eyes and looks around in confusion] and *eine Hand und ein Kopf ragen neben ihm aus der Erde* [a hand and a head stick out from the ground next to him] are two separate utterances that are demarcated by a pause. In addition, these utterances are syntactically independent, separate entities, although the deictic expression 'next to him' links the utterances thematically (see Hirvonen 2013a: 25–27). By contrast, the glance and the stimulus are syntactically bound in the 'narrated character's perspective'. An example is the following: *sie entdeckt blühende Märzenbecher* [she notices snowflake flowers in bloom] (ibid. 20). Here it is explicitly stated that the stimulus (snowflake flowers in bloom) is the object of the glance (she notices). Structurally, therefore, the syntactic independence of the 'experienced character's perspective' reflects the structure of the point-of-view shot by constituting separate utterances for glance and stimulus. Hence, the original, visually implicit cue shifts to the audio description as likewise being implicit, so that users may or may not interpret the character perceiving that stimulus. However, the subsequent audio description may confirm the perception by referring to the character's reaction, which in fact occurred in the Peter example by an utterance that topicalises an adverb of an emotional state, *geschockt springt Peter auf* [shocked, Peter jumps up]. (See ibid. 30.) It also transpired that the choice of expression that refers to the glance has consequences for how the perceiving character is represented. This involves whether he or she either observes volitionally, or rather perceives passively, and whether or not the perception itself is confirmed (compare, for instance, the semantics of the verbs 'look' and 'see') (ibid. 21–22).

Finally, audio description can also parallel a film's visuals in terms of the composition of the shot space as Figure and Ground (Hirvonen 2013b). This means that in both modes, the space segregates into 'thing-like' Figures and a 'substance-like' Ground. To study this segregation, I analysed two different cases of spatial composition. The first is a case in which the visual element, a minibus, is Figure in the first shot and it subsequently becomes Ground of the character action in the next, and a case in which the opposite occurs, Ground (a shack) becomes Figure (shacks in a landscape) (ibid. 98ff.). Regarding the 'minibus' element in the first case, considerable similarity was detected between the visual composition and the English and German audio descriptions. The Spanish audio description, which was the third parallel text, diverged more (for instance, by verbalising different elements from those in the English and German audio descriptions). (Hirvonen 2013b: 101.) In the second case, all three audio descriptions orient to the 'shack' element similarly as the film: first as Ground and then as Figure (ibid. 106–108). Nevertheless, it was noted that as the film scene develops, the linguistic Figure and Ground assignments become more varied, at least in terms of the linguistic devices used (for instance, the different syntactic functions of the 'bus' in *im Bus sitzen Jamal und Salim…* [in the bus, Jamal and Salim are sitting…], which is a locative

prepositional phrase, and ***the bus is*** *full of scruffy street kids*, which presents it in a subject role) (ibid. 102).

4 Discussion

This final chapter presents a discussion of the research reported in the previous chapters. The results from this study offer new insight into the strategies of filmic audio description as well as into the idea of similarity in audio description. Moreover, these findings have important implications for the discussion of the similarity and divergence between image and language. Even though language and image are different modes with distinct capacities for communicating information, the 'sameness' between the two also merits attention. To conclude, this chapter offers an evaluation of the present research and proposes ideas for future directions in audio description research.

4.1 Theoretical discussion

4.1.1 Filmic audio description

On the basis of the results and in the context of source-text oriented translation, one could argue for adopting a 'discourse-faithful' strategy to audio describe film narratives (for film as story and discourse, see Chatman 1978: 26). Discourse-faithfulness is defined in this context as the attempt to conserve the film-specific forms of representation, such as the dynamic variation in spatial foci and points of view, and the implicitness of these, in audio description in order to facilitate access to the experience of filmic art for a visually impaired audience (see Navarrete Moreno 1997). Furthermore, because films recycle, that is, they repeatedly employ specific techniques to represent and narrate, this recycling is relevant to the discussion of the intermodal similarity that reflects these techniques. This means that consistent film techniques could result in a consistent use of audio description techniques (see Remael & Vercauteren 2007: 85), with the advantage that the blind and partially sighted filmgoers could appreciate different styles and learn about filmic discourse in a systematic way. This dissertation has therefore demonstrated that it is possible to match audio descriptions with certain film techniques, in particular with shot distance and point-of-view shot (see also Jimenez Hurtado & Soler Gallego 2013: 588, 591 for a discussion of corresponding intersemiotic translation strategy regarding camera movement, close-ups and the shot-reverse-shot structure). Other forms of audiovisual media and narratives, such as television programmes and video, could also benefit from this consistency. In addition, a correspondence between the visual source material and the verbal-vocal target material can be established in terms of the following factors: the representation of shot space, editing space, and off-screen space, the use of spatial continuity or discontinuity as well as in the alteration of points of view and spatial dimensions. I therefore argue that in addition to what should be verbalised being relevant (cf. Vandaele 2012: 88), how visual input is verbalised is also

significant for shifting a filmic-visual representation into an audio description. The findings of this dissertation also demonstrate that the multimodality of the auditory target text of the audio-described film, with numerous points of interests, creates a multimodal variant of the audiovisual source text (see Remael 2012: 266).

The cross-linguistic observations in the present study revealed a discrepancy in the different versions of audio description, such as the naming of referents between the German and Spanish audio descriptions (Hirvonen 2012) as well as the narrative progression and the selection of the point of interest between the Spanish and the English and German audio descriptions (Hirvonen 2013b). Nonetheless, this study indicates that this discrepancy does not depend on the linguistic system, but rather on differences in style and cultural preferences. For instance, the distinct linguistic devices listed in Hirvonen (2012: 38) could be used by any of the languages in question, and the strategies defined in Hirvonen (2012) and (2013a) were found in all the analysed language versions (however, one language-dependent factor affecting audio description is presented in 4.3). As a consequence, such differences in style and preferences are likely to affect a possible standardisation of audio description or the application of this in practice.

Further challenges persist because the filmic visual discourse communicates information on a number of levels: the aesthetic-graphic, the referential-denotative, the realistic and the narrative (Bordwell 1985: 102; see also my discussions in Hirvonen 2012: 40 and Hirvonen 2013b: 111)). In addition, filmic imagery is particularly effective in denoting and offering 'a close approximation of reality' and in communicating 'precise information about physical realities (Monaco 1977/2009: 179). Audio description, in contrast, can verbalise the individual visual cues or the interpretation of cues rather explicitly (Braun 2007: 8). Although linguistic verbalisation always requires some level of interpretation, the concept of cue is significant and worth considering because it emphasises the importance of mediating some of the discursive cues and causes that prompt the receivers to formulate narrative hypotheses and to envision the narratives (Vandaele 2012: 89; see also Bordwell 1985).

However, it is also important to bear in mind that the style of verbalising cues may lead to fragmentation and this might risk making the reception of the audio description unpleasant or even impossible (Kruger 2010: 245). In the light of discourse-faithfulness, however, creating fragmented views can be interpreted as being functional because like film shots, they select slices from the 'implicitly continuous' world to show us (see Bordwell & Thompson 1979/1990: 173). In addition, providing fragmented views (for instance, the variation between neutral and subjective points of view) can be precisely the feature that is unique to the audio description of film and, hence, makes it 'filmic', as opposed to other types of audio description (see Hirvonen 2013a: 32).

Yet audio description is not solely responsible for mediating the cues that are used to construct the film narrative, but is supported by, or supports, the filmic soundtrack. Overall, the findings of this dissertation corroborate the functions of film music that have been reviewed in Igareda (2012: 238–239) as well as with the functions of film sound reviewed in Remael (2012: 261–262). To provide some examples, music accompanies and supports action. Music also provides rhythm for narrative structures, links scenes, and plays the characters' thoughts. The filmic soundtrack also suggests a mood for the scene, heightens or diminishes realism and ambiguity, either draws attention to detail or away from it (creating a near-distance effect and perspectival changes) and creates acoustic space. Even though this thesis predominately considers the sonic representation of space that is iconic to the real world, Remael (2012: 261–262) aptly reminds us that film sounds are also staged, not always realistic, and are usually post-produced. As a consequence, they may potentially be difficult for the visually impaired audience to understand (ibid.). Furthermore, sounds sometimes occur independently of the audio description in the sense that, instead of drawing implicatures for the sound effect, audio description can merely rely on describing the visuals (see Hirvonen 2013b: 106–108) and let the narrative progression of the film provide the answers to the asynchrony (see Remael 2012: 265).

4.1.2 Similarity and divergence between image and language

A widely recognised challenge for translating images into words is that images are iconic in the sense that the impressions from picture viewing are similar to those received by perceiving the real world visually (Stöckl 2004: 17). While film images also represent unique items – such as a *certain type of* man, woman, dog, house, and so forth – visual percepts must be abstracted from this uniqueness in order to be effectively communicated through the linguistic mode (Grodal 1997: 22). Furthermore, verbalisation must make a selection of the visual whole and conceptualise its aspects with roles and functions (see Stutterheim & Klein 2002). Thus, 'the concept of CAR, is a schematisation derived by generalising across many different sorts of specific (episodic) experiences relating to automobiles in order to form a single representation' (Evans 2010: 22). Conceptualisation therefore involves a classification of items and necessarily profiles some aspect from the referent (for instance, see Langacker 2008). If we recall the example with the referents *toilet* and *wooden shack*, they highlight either the physical aspect or the functional aspect of the same object.

Images pose another challenge because they are holistic, providing multiple points of interest, and viewers can scan through them in different ways. This becomes, however, more restricted in film narration, which steers and manipulates perception so that the viewing is experienced through cues that are visual, auditory and narrative, and these include the lighting, shot

length, perspective, camera movement, and a number of other factors. In this sense, language can be iconic for the sequence in which the film organises and relates the elements, even though the objects themselves are not iconic. As a consequence, the sequential organising of the information in audio description can be matched with the presumed or intended order in which the filmic space is perceived (see also Seiffert 2005: 84). But at the same time, the selection from the visual whole involves selecting a frame of reference in the verbalisation (see also Landau et al. 2010: 56). The same image may give various options for this, as was evident in the 'garden'/'bunker' example (Hirvonen 2012: 29). In the description of a setting that a long shot depicts, the Spanish frame of reference in audio description is 'outside the bunker', whereas in the German audio description, the frame of reference is 'in the devastated garden behind the old chancellery'. The Spanish audio description leaves the environment around the bunker implicit, whereas the German description verbalises that exact environment and does not explicate the bunker element (although this can be inferred).

Furthermore, while verbalisation and therefore audio description determine which aspects of the film scene are accessed explicitly, the visual mode allows us to appreciate various points of view. For instance, we can admire beautiful colours and forms on the aesthetic level, expressed as a meadow of flowers on the referential and realistic levels, and/or as the object of a gaze on the narrative level. The challenge of audio description is fitting the possibility of interpreting on all these levels into (one short) verbalisation. Of course, not all the levels must be expressed by the verbal description because many cues are auditory (such as the realistic effect through sound, see Fryer 2010) and narrative (such as the point-of-view shot as an interpretation of particular sequence). Furthermore, the narrative context of film viewing and our desire to follow the story often accentuate the referential level rather than the pictorial qualities or the visual communication as a picture (Bordwell 1985: 102; see also my discussion in Hirvonen 2013b: 111–112).

The 'garden'/'bunker' example illustrates a situation of divergent similarity in translation in which 'an original entity may give rise to more than one additional similar entity' (Chesterman 1996: 163):

$$A \rightarrow A', A''$$

In the example mentioned, both descriptions of the shot space are similar to the source material, but they are simultaneously different. For instance, the shot depicts both an area 'outside the bunker' as well as 'the devastated garden behind the old chancellery'. Thus, one source text has produced two different translations that nevertheless share traits with the source. This notion of divergent similarity is useful to describe audio description overall because the visuals of film provide multiple cues to be verbalised, and two different 'selections of cues' may also reflect the original, albeit in different ways (see also Matamala

& Rami 2009). In fact, the present study has presented many findings that involve divergent similarity. For example, while being similar to the source material on the structural level (in terms of the spatial dimension and emphasis that is cued by the shot distance), the audio descriptions diverge at the referential level as to which aspects they profile by the concepts they use.

4.2 Methodological discussion and evaluation of research

This dissertation adopted a multimodal, intermodal and cross-linguistic methodology to approach the representation of space in filmic audio description. Moreover, the research problem was analysed from the analytical perspectives of both translation and reception. The translation perspective was applied to the source material and the translation (film and audio description) to compare their intermodal similarity (how language corresponds to images). The translation perspective was theory-driven and it examined how certain source text features have been translated, but at the same time, it directed the analysis to focus on certain aspects, while other, relevant factors may have gone unnoticed. For example, regarding the analysis in Hirvonen (2013a), one might ask whether the findings would have been different had I departed from the reception perspective and analysed the audio description or the audio-described films without the pre-defined film-narrative categories: What type of instances of perspective would have become relevant? Additionally, the reception perspective was adopted to observe audio-described film as a stand-alone product that provides multimodal cues for the product users. In the reception perspective, we are left with what is provided to the recipient without guidance from the source text, but the transformation of information that occurs in the translation process remains hidden. The general disadvantage of studying the specific features in complex texts is that there is less focus on the complexity of the film narrative (see Hirvonen 2012: 40 and Hirvonen 2013a: 31). The advantage, however, is that this type of analysis makes a complex source text more systematic and enables us to observe larger sets of data to evaluate whether this particular system is reflected in the translation.

The presentation and transcription of the data slightly complicated the dissemination of the research because the filmic material had to be represented as still images and the multimodal, audio-described film as transcriptions. These representations and transcriptions had different forms in the articles (for instance, compare the still images in Hirvonen 2012 and 2013b, or the transcriptions in Hirvonen & Tiittula 2012 and Hirvonen 2013b). As for the underlying rationale for using the different forms, which is explained in Chapter 1.4, let me just note here that it is up to the analyst to decide on how to represent the multimodal data because 'the complex simultaneity of different modes and their different structure and materiality' continue to produce unanimity in multimodal transcription overall (Flewitt et al. 2009/2011:

46). However, a consistent form of data representation and transcription would have made the comparison of the different data more straightforward.

As for the film images, acquiring copyright permissions for still shots can be time-consuming and expensive; it is compulsory for many publishers to acquire rights for work by third parties. In this regard, it would be advantageous if pictorial material could be referred to with the same ease as we currently cite verbal material. Furthermore, I had to compromise between the complexity and the amount of the data. For instance, in Hirvonen & Tiittula (2012), a detailed transcription was made of the auditory multimodality, and only two minutes of data were analysed. These data were monolingual, but the length of the data transcription increases when multilingual data are presented, which is evident in Hirvonen (2013b). The analysis of parallel audio descriptions necessitates the translation of audio description samples into the language of the research report. As a consequence, the number of words and the length of the paper increase. Nonetheless, the translations themselves also provided valuable evidence to address the research objectives because the translations raised questions on cross-linguistic matters and on the possible problems that arise from the interlingual translation of audio description scripts (see Hirvonen 2013a: 31). Furthermore, while creating a shot protocol from scratch is more time-consuming than using ready-made scripts, the researcher becomes rather familiar with the data when de-constructing and reconstructing it. Another advantage of creating a shot protocol and the matrix format is that they connect the information of the different modes and therefore represent the simultaneity and multimodality of film (see also Flewitt et al. 2009/2011: 47). Relying solely on the script is risky because the final product often differs from it, but the advantage is that the researcher can scan through the contents and the structure of the audio description or film without the need of reconstructing the material. Furthermore, both scripts and transcriptions can serve as 'raw material' and these can serve as the basis for creating more elaborated transcriptions. Finally, while there are different means to illustrate visual shots and the 'imaginary space' (see my articles and Tanis Polat 2013), a direct way of visualising sounds seems to be more difficult to achieve because sound is a continuum and formless. Nevertheless, one could try to illustrate it through curves or other visualisation techniques (for instance, from acoustics and phonetics). The technical setup also influences the analysis of the filmic soundtrack. In my experience with the present data, some of the softer sounds were barely audible when listening to the film through computer loudspeakers as opposed to headphones. In the analyses presented in this study, my main goal was to be the best possible listener, attempting to note down every sound.

The analyst's dilemma also prevails in this dissertation. Any methodology that needs to account for the multimodal meaning making in film has to take into consideration the researchers' subjectivity (see Jimenez Hurtado & Soler Gallego 2013: 582-583, 591). Toward this end, my objective was to analyse the data as transparently as possible and to present the

observations and the reasoning in a detailed manner so that the argumentation can be followed by anyone who reads the analysis. Moreover, my adaptation of the Gestalt-psychological theory of Figure and Ground segregation in Hirvonen (2013b) can be problematic because the theory originally models a very basic visual perception of static images, whereas I applied it to the analysis of moving images (I was unable to find research on Figure and Ground segregation in moving images). In addition, Figure and Ground segregation seems to be activated in the data-driven process rather than in the concept-driven one, which is what perceivers tend to employ when interpreting real-life or film scenes (see Goldstein 2007/2010: 11). The difficulty in applying some of the more cognitively oriented features of Figure and Ground to my analysis may stem from this methodological issue.

While attempting to understand the potential of the different modes, it is important to note that the outcome of the study focusses on the linguistic mode. The departure point was the visual techniques of film narration, but a more detailed analysis of the variety of the visual cues of, for instance, shot space could have been conducted. While arguing for the means of language in cueing the visually represented space, the present study does not permit direct assumptions concerning the interpretation of these cues by individual users. Whereas narration necessarily involves establishing meaning at a personal level, my analyses cannot account for the individual judgements by the users. For instance, based on the data in Hirvonen & Tiittula (2012), while it is not possible to determine which of the cues are perceived and understood by different users, a range of cues can be distinguished and an explanation can be formulated on their possible functions in the narrative context of film. On the other hand, filmmakers utilise film techniques in an attempt to create certain effects. In conclusion, my main objective has been to demonstrate the meaning-making potential of these modes and to build on previous research and existing theory by explaining the points of similarity between film and audio description.

4.3 Implications for practice and future research

The methodology introduced in these articles and the data collected for the analyses could be further developed and utilised in future research. Central issues that have yet to be examined systematically include a cross-linguistic analysis of the data as well as an analysis of the prosodic features of speech. Moreover, reception research is another resource that might provide more knowledge on how non-sighted or partially sighted users employ the filmic soundtrack (also see Remael 2012: 272–273). While Hirvonen & Tiittula (2012) note some ways in which the voice, with its prosodic and extra-linguistic properties, cues filmic space, this issue merits more comprehensive examination. Although Cabeza-Cáceres (2013) reported that the comprehension of film is not significantly influenced by intonation, it is however

likely that this affects the enjoyment of the film and divides audio description users 'between those liking and those rejecting uniform and emphatic intonations' (ibid. 331). Other prosodic features such as variation in the pitch level can be assumed to affect the perception of audio description for the same reason that it does in other forms of spoken discourse, and therefore also influence the reconstruction of the story by indicating relationships between utterances (for instance, see Wiklund 2014). One question for future research concerns how prosody contributes to the verbal opening strategies in Hirvonen (2012). With regard to speech in film dialogue, on the other hand, further research might explore the amount of information that can be obtained from dialogue, as voice renders different types of information (for instance, the geographical origin, nationality, and mood of speakers, see Bose 2010: 29).

While this dissertation has focussed on the language system in general and has detected general linguistic devices that are applicable to a range of languages, the analysis of audio description poses interesting problems in terms of cross-linguistic research. These include which meanings are language-specific and must be expressed in that linguistic system (Jakobson 1959), and the implications of those meanings in the context of audio description. One such language-dependent factor is the encoding of movement. Some languages, such as English, seem to prefer encoding the manner of the movement in the verb, whereas others, such as Spanish, express the path and direction of the movement (for instance, see Landau et al. 2010: 55, 62ff. and the example below). The corpus consulted for this study provides data for this type of research. In fact, this cross-linguistic difference can be observed in the Spanish parallel of the German audio description data used by Hirvonen & Tiittula (2012). While the movement of the characters is verbalised in the German version by the verbs expressing manner (*hasten… eilt*, both meaning 'to hurry'), the Spanish audio description uses verbs expressing a path (*cruzan… atraviesa*, both denoting 'to traverse/go through'). An implication of this is the possibility that, taking large corpora of German, English and Spanish audio descriptions, we find a significant discrepancy in the verbalisation of movement.

This need for understanding the cross-linguistic and cross-cultural factors in audio description is in agreement with other recent studies (for instance, see the *Pear Tree Project*, referred in Mazur & Chmiel 2012, and the ADLAB project[15]). The results of these inquiries have implications for the practice by identifying the possible cultural and linguistic characteristics of audio description that may become an issue when audio description scripts are translated from one language into another (see Remael & Vercauteren 2010). Toward this aim, comparative research also needs to be conducted on the different types of audio description processes (preparing audio description from scratch versus translating audio description

[15] See the aims and objectives of the ADLAB Project (Audio Description: Lifelong Access for the Blind): http://www.adlabproject.eu/?page_id=44 (accessed 16 October 2014).

scripts) to determine more specifically what the advantages and disadvantages are. In this respect, we might also be interested in analysing the intersubjective work that is being undertaken by audio description teams, including those who are sighted as well as those who are non-sighted, to determine how factors such as their different cultural and linguistic backgrounds affect their verbalising and conversing about the non-linguistic world. This type of research could create new knowledge of Figure and Ground segregation and other aspects of scene analysis.

The results of this dissertation may prove to be valuable in developing standardised audio description. If internationally accepted and applied standards and guidelines on audio description become reality, these agreements could be justified by the discourse-faithful strategy. Furthermore, standardised audio description, which should increase the efficiency of the audio description process, might further encourage the introduction of audio-described content in the mass media, such as the television and the internet, in countries where audio description remains a marginal service of translation. In the event of an increase in audio-described content, further research could explore the relevance of audio description as a technique of transforming the information between two different modes of representation, such as the use of audio description as a basis for the textual tagging of audiovisual and visual materials in digital environments (see Salway 2007: 168–169).

Hence, one issue in need of further analysis is the development of standardised audio description that involves the cinematic, rather conventionalised techniques of representation and narration in cinema. Devising a method to mediate these in audio description would be a significant contribution not only to the conventional use of audio description, but also to the development of a text-based access to visual and audiovisual information. Another benefit to devising this method is that different linguistic forms could be matched with different forms of visual representation so that users would be able to identify individual forms of representation and even genres (see Hirvonen 2013a: 33–34; Jimenez Hurtado & Soler Gallego 2013: 580). Having said this, difficulties may arise because languages offer a wide array of possibilities for cueing the same thing (for example, the different linguistic devices that cue a change of scene). In this sense, the present research has also demonstrated how audio description, as the verbalisation of the visual, provides interesting issues and data for linguistic investigation.

References

Film data

Dancer in the Dark; Denmark 2000. Directed by Lars von Trier. With audio description in English (IADA Ltd. 2000) and German (DBSV / Projekt Hörfilm 2001).

Der Untergang/El hundimiento; Germany 2004. Directed by Oliver Hirschbiegel. With audio description in German (Bayerischer Rundfunk 2005) and Spanish (ONCE 2006).

Slumdog Millionaire/Slumdog Millionär; UK 2008. Directed by Danny Boyle & Loveleen Tandan. With audio description in English (Pathé Distribution 2009), German (Hörfilm GmbH 2009) and Spanish (Navarra de Cine S.L. 2009).

Tauno Tukevan sota; Finland 2010. Directed by Heidi Köngäs. With audio description in Finnish (Yleisradio 2010).

Other films

Postia pappi Jaakobille/Letters to Father Jacob; Finland 2009. Directed by Klaus Härö.

Puolin ja toisin/Middle of the Road; Finland 2013. Directed by Matti Ijäs.

Risto Räppääjä ja polkupyörävaras/ Ricky Rapper and the Bicycle Thief; Finland 2010. Directed by Mari Rantasila.

The Hours; USA 2002. Directed by Stephen Daldry.

Varpuset; Finland 2005. Directed by Heidi Köngäs.

Virginie; Finland 2009. Directed by Heidi Köngäs.

Literature and online sources

Aaltonen, Anu (2007). *Tietopaketti kuvailutulkkauksesta* [Information about audio description]. Retrieved from http://www.kulttuuriakaikille.info/tietopaketit_ja_oppaat_kaikki_tietopaketit_ja_oppaat.

ADLAB (2012). Report on user needs assessment. Report no. 1, ADLAB (Audio Description: Lifelong Access to the Blind) project. Retrieved from http://www.adlabproject/eu.

AENOR (2005). *Norma UNE: 153020. Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*. Madrid: AENOR.

Agost, Rosa, Orero, Pilar & Giovanni, Elena di (Eds.) (2012). *Multidisciplinarity in Audiovisual Translation (Monografías de Traducción e Interpretación MonTI 4)*. Retrieved from http://dti.ua.es/es/documentos/monti/monti-4-indice-y-enlaces.pdf.

Bacon, Henry (2005). Synthesizing approaches in film theory. *The Journal of Moving Image Studies 4*. Retrieved from http://www.avila.edu/journal/index1.htm.

Bassnett, Susan (2011). *Reflections on Translation*. Bristol: Multilingual Matters.

Benecke, Bernd (2014). *Audiodeskription als partielle Translation. Modelle und Methode.* Berlin: LIT.

Bordwell, David (1985). *Narration in the Fiction Film*. London: Methuen.

Bordwell, David (2010). The Part-Time Cognitivist: A View from Film Studies. *Projections 4(2)*, 1-18.

Bordwell, David & Thompson, Kristin (1979/1990). *Film art: an introduction (3rd ed.)*. New York: McGraw-Hill.

Bose, Ines (2010). Stimmlich-artikulatorischer Ausdruck und Sprache. In Arnulf Deppermann & Angelika Linke (Eds.), *Sprache intermedial. Stimme und Schrift, Bild und Ton*. Berlin/NY: Gruyter, 29-68.

Bourne, Julian & Jiménez, Catalina (2007). From the visual to the verbal in two languages: a contrastive analysis of the audio description of The Hours in English and Spanish. In Jorge Diaz-Cintas et al. (Eds.), 175-187.

Branigan, Edward (1984). *Point of view in the cinema. A theory of narration and subjectivity in classical film*. Berlin/New York/Amsterdam: Mouton.

Branigan, Edward (1992). *Narrative comprehension and film*. London: Routledge.

Braun, Sabine (2007). Audio Description from a discourse perspective: a socially relevant framework for research and training. *Linguistica Antverpiensia 6*, 357-369. Retrieved from https://lans-tts.uantwerpen.be/index.php/LANS-TTS/issue/view/10.

Braun, Sabine (2008). Audiodescription Research: State of the Art and Beyond. *Translation Studies in the New Millennium. An Internation Journal of Translation and Interpreting 6*, 14-30.

Brown, Gillian & Yule, George (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

Cabeza-Cáceres, Cristóbal (2013). *Audiodescripció i recepció*. Unpublished doctoral thesis. Universitat Autònoma de Barcelona.

Cámara, Lidia & Espasa, Eva (2011). The Audio Description of Scientific Multimedia. *The Translator 17:2*, 415-437.

Carlston, Donal E. & Mae, Lynda (2001). Social psychology of schemas. In *International Encyclopaedia of the Social & Behavioral Sciences*. Elsevier Science, 13526-13530.

Chatman, Seymour (1978). *Story and Discourse. Narrative Structure in Fiction and Film.* Ithaca/London: Cornell University Press.

Chesterman, Andrew (1996). On similarity. *Target 8(1)*, 159-164.

Chesterman, Andrew (2004). Where is similarity? In Stefano Arduini & Robert Hodgson (Eds.), *Similarity and Difference in Translation*. Rimini: Guaraldi, 63-75.

Chesterman, Andrew (2007). Similarity Analysis and the Translation Profile. *Belgian Journal of Linguistics 21*, 53-66.

Corral, Anna & Lladó, Ramon (2011). Opera multimodal translation: Audio describing Karol Szymanowski's Krol Roger for the Liceu theatre, Barcelona. *JoSTrans - The Journal of Specialised Translation 15*, 163-179.

De Coster, Karin & Mühleis, Volkmar (2007). Intersensorial translation: visual art made up by words. In Jorge Díaz Cintas et al. (Eds.), 189-200.

Díaz Cintas, Jorge, Orero, Pilar & Remael, Aline (2007). *Media for all: subtitling for the deaf, audio description and sign language.* Amsterdam: Rodopi.

Dosch, Elmar & Benecke, Bernd (2004). *Wenn aus Bildern Worte werden. Durch Audio-Description zum Hörfilm (3. Auflage)*. Munich: Bayerischer Rundfunk.

European Blind Union (n. d.). Facts, figures and definitions concerning blindness and sight loss. Retrieved from http://www.euroblind.org/resources/information/nr/215.

Evans, Vyvyan (2010). The perceptual basis of spatial representation. In Vyvyan Evans & Paul Chilton (Eds.), 21-48.

Evans, Vyvyan & Chilton, Paul (Eds.) (2010). *Language, cognition and space. The state of the art and new directions*. London/Oakville: Equinox.

Fix, Ulla (2005). *Einleitung*. In Ulla Fix (Ed.), 7-11.

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

Flewitt, Rosie, Hampel, Regine, Hauck, Mirjam & Lancaster, Lesley (2009/2011). What are multimodal data and transcription? In Jewitt, Carey (Ed.), *The Routledge Handbook of Multimodal Analysis (paperback ed.)*, 40-53.

Fryer, Louise (2010). Audio description as audio drama - a practitioner's point of view. *Perspectives: Studies in Translatology 18:3*, 205-213.

Geeraerts, Dirk & Cuyckens, Hubert (2007). Introducing cognitive linguistics. In Dirk Geeraerts & Hubert Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford/New York: Oxford University Press, 3-21.

Gerzymisch-Arbogast, Heidrun (2005). Multidimensionale Translation: Ein Blick in die Zukunft. In Felix Mayer (Ed.), *20 Jahre Transforum. Koordinierung von Praxis und Lehre des Dolmetschens und Übersetzens*. Hildesheim/Zürich/New York: Georg Olms, 23-30.

Goldstein, Bruce E. (2007/2010). *Sensation and perception (8th ed.)*. Belmont, CA: Wadsworth Cengage Learning.

Greening, Joan & Rolph, Deborah (2007). Accessibility: raising awareness of audio description in the UK. In Díaz Cintas et al. (Eds.), 127-138.

Grodal, Torben (1997). *Moving Pictures. A New Theory of Film Genres, Feelings, and Cognition.* Oxford: Oxford University Press.

Gumperz, John J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

Gutenberg, Norbert (2000). Mündlich realisierte schriftkonstituierte Textsorten (mrskT). In Klaus Brinker, Gerd Antos, Wolfgang Heinemann & Sven F. Sager (Eds.), *Text- und Gesprächslinguistik / Linguistics of Text and Conversation (Halbbd. 1/Vol. 1)*. Berlin: Gruyter, 574-582.

Gutt, Ernst-August (1991). *Translation and Relevance. Cognition and Context.* Oxford/Cambridge: Blackwell.

Haspelmath, Martin (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics 19(1)*, 1-33.

Hausendorf, Heiko, Mondada, Lorenza & Schmitt, Reinhold (2012). Raum als interaktive Ressource: Eine Explikation. In Hausendorf, Heiko, Mondada, Lorenza & Schmitt, Reinhold (Eds.), *Raum als interaktive Ressource*. Tübingen: Narr, 7-36.

Hernández-Bartolomé, Ana I. & Mendiluce-Cabrera, Gustavo (2004). *Audesc*: Translating Images into Words for Spanish Visually Impaired People. *META 49(2)*, 264-277.

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 (2013a). Perspektivierungsstrategien und -mittel kontrastiv: Die Verbalisierung der Figurenperspektive in der deutschen und finnischen Audiodeskription. *trans-kom: Zeitschrift für Translationswissenschaft und Fachkommunikation 6(1)*, 8-38.

Hirvonen, Maija (2013b). Sampling Similarity in Image and Language – Figure and Ground in the Analysis of Filmic Audio Description. *SKY Journal of Linguistics* 26, 87-115.

Hirvonen, Maija (2013c). Katsaus kuvailutulkkaukseen – visuaalisen tiedon saavuttaminen puheen ja kielen kautta [Review of Audio Description – Making Visual Information Accessible Through Speech and Language]. *Puhe ja kieli 33(3)*, 91-106.

Hirvonen, Maija & Igareda, Paula (2011, June). Verbalizing space: Verbal cues of setting in the English, German and Spanish audio descriptions of Slumdog Millionaire. Paper presented at the conference *Media for All 4. Audiovisual Translation: Taking Stock*, London, UK.

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*. Tübingen: Narr, 381-427.

Hyks, Veronika (2005). Audio Description and Translation: Two related but different skills. *Translating Today 4*, 6-8.

Igareda, Paula (2012). Lyrics against images: music and audio description. In Rosa Agost et al. (Eds.), 233-254.

Jakobson, Roman (1959). On linguistic aspects of translation. Reprinted in Lawrence Venuti (Ed.) (2000), *The Translation Studies Reader*. London: Routledge, 113-118.

Jankowska, Anna & Szarkowska, Agnieszka (2013). Translation as an alternative method of creating audio description scripts. In Lucja Biel & Kyriaki Kourouni (Eds.), *EST Newsletter 43*.

Jiménez, Catalina (2007). Una gramática local del guión audiodescrito. Desde la semántica a la pragmática de un nuevo tipo de traducción. In Catalina Jiménez Hurtado (Ed.), 55-80.

Jiménez Hurtado, Catalina (Ed.) (2007). *Traducción y accesibilidad.* Frankfurt am Main: Peter Lang.

Jimenez Hurtado, Catalina & Soler Gallego, Silvia (2013). Multimodality, translation and accessibility: a corpus-based study of audio description. *Perspectives: Studies in Translatology 21(4)*, 577-594.

Kluckhohn, Kim (2005). *Informationsstrukturierung als Kompensationsstrategie – Audiodeskription und Syntax.* In Ulla Fix (Ed.), 49-65.

Kruger, Jan-Louis (2010). Audio narration: re-narrativising film. *Perspectives: Studies in Translatology 18(3),* 231-249.

Lahtinen, Riitta & Palmer, Russ (2012). Environmental description. In Elisa Perego (Ed.), *Emerging topics in translation: Audio description*. Trieste: Edizione Università di Trieste, 105-114. Retrieved from http://hdl.handle.net/10077/6356.

Lahtinen, Riitta, Palmer, Russ & Lahtinen, Merja (2009). *Aisti kuvailu* [Sense description]. Helsinki.

Landau, Barbara, Dessaleng, Banchiamlack, Goldberg, Ariel Micah (2010). Language and space: momentary interactions. In Paul Chilton & Vyvyan Evans (Eds.), 51-77.

Langacker, Ronald (2008). *Cognitive Grammar: A Basic Introduction*. Oxford University Press. Retrieved from Oxford Scholarship Online (www.oxfordscholarship.com).

Langendonck, Willy van (2007). Iconicity. In Dirk Geeraerts & Hubert Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford/New York: Oxford University Press, 394-418.

Mäkisalo, Jukka & Lehtinen, Marjatta (2014, April). Kääntämisen kognitiivinen kuvaus: hahmo ja tausta [The cognitive description of translation: Figure and ground]. Poster presented at the conference *XII Symposium on Translation and Interpreting*, Tampere, Finland.

Márquez Linares, Irene (2007). *Chuchotage* para ciegos: un susurro ensayado. In Catalina Jiménez Hurtado (Ed.), 209-227.

Matamala, Anna & Orero, Pilar (2011). Opening credit sequences. Audio describing films within films. *International Journal of Translation 23(2)*, 35-58.

Matamala, Anna & Rami, Naila (2009). Análisis comparativo de la audiodescripción española y alemana de "Good-bye Lenin". *Hermenéus, Revista de Traducción e Interpretación 11*, 1-13.

Mazur, Iwona & Chmiel, Agnieszka (2012). Towards common European audio description guidelines: results of the Pear Tree Project. *Perspectives: Studies in Translatology*, *20(1)*, 5-23.

Mazur, Iwona & Taylor, Christopher (2013). Audio Description: Lifelong Access For The Blind (Adlab). In Lucja Biel & Kyriaki Kourouni (Eds.), *EST Newsletter 43*.

Monaco, James (1977/2009). *How to Read a Film (4th ed.)*. Oxford/New York: Oxford University Press.

Muntigl, Peter (2004). Modelling multiple semiotic systems. The case of gesture and speech. In Eija Ventola, Charles Cassily & Martin Kaltenbacher (Eds.), *Perspectives on Multimodality*. Amsterdam/Philadelphia: John Benjamins, 31-50.

Navarrete Moreno, Francisco Javier (1997). Sistema *AUDESC*: El arte de hablar en imágenes. *Integración 23*, 70–75.

Ofcom (2000). ITC Guidance On Standards for Audio Description. Retrieved from http://stakeholders.ofcom.org.uk/broadcasting/guidance/other-guidance/tv_access_serv/archive/audio_description_stnds/.

Orero, Pilar (2004). Audiovisual translation: A new dynamic umbrella. In Pilar Orero (Ed.): *Topics in audiovisual translation.* Amsterdam/Philadelphia: John Benjamins, vii-xiii.

Orero, Pilar (2005). Audio Description: Professional Recognition, Practice and Standards in Spain. *Translation Watch Quarterly 1*, 7-17.

Orero, Pilar (2007). Sampling audio description in Europe. In Jorge Diaz-Cintas et al. (Eds.), 111-125.

Orero, Pilar (2007). Pioneering Audio Description: An Interview with Jorge Arandes. *JoSTrans: Journal of Specialised Translation 7*, 179-189. Retrieved from http://www.jostrans.org/issue07/issue07_toc.php.

Orero, Pilar & Vilaró, Anna (2012). Eye tracking analysis of minor details in films for audio description. In Rosa Agost et al. (Eds.), 295-320.

Palomo López, Alicia (2010). The benefits of audio description for blind children. In Jorge Díaz Cintas, Anna Matamala & Josélia Neves (Eds.), *New Insights into Audiovisual Translation and Media Accessibility: Media for All 2*. Amsterdam/New York: Rodopi, 212-225.

Pérez González, Luis (2009). Audiovisual translation. In Mona Baker & Gabriela Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies (2nd ed.)*. Taylor & Francis e-Library.

Pérez Payá, María (2007). La audiodescripción: traduciendo el lenguaje de las cámaras. In Catalina Jiménez Hurtado (Ed.), 81-91.

Prince, Stephen (1993). The Discourse of Pictures: Iconicity and Film Studies. *Film Quarterly 47(1)*, 16-28.

Pujol, Joaquim & Orero, Pilar (2007). Audio Description Precursors: Ekphrasis, Film Narrators and Radio Journalists. *Translation Watch Quarterly 3(2)*, 49-60.

Pütz, Martin & Dirven, Rene (Eds.) (1996). The construal of space in language and thought. Berlin/New York: Mouton de Gruyter.

Remael, Aline (2010). Audiovisual translation. In *Handbook of Translation Studies Online (Vol. 1)*. John Benjamins.

Remael, Aline (2012). For the use of sound. Film sound analysis for audio-description: Some key issues. In Rosa Agost et al (Eds.), 255-276.

Remael, Aline & Vercauteren, Gert (2007). Audio describing the exposition phase of films. Teachings students what to choose. *TRANS: Revista de traductología 11*, 73-93.

Remael, Aline & Vercauteren, Gert (2010). The translation of recorded audio description from English into Dutch. *Perspectives: Studies in Translatology 18(3)*, 155-171.

Royal National Institute of Blind People (2013). *Sight Problems. Changing the way we think about blindness*. Retrieved from http://www.rnib.org.uk/sites/default/files/sight_problems_guide.pdf.

Salway, Andrew (2007). A corpus-based analysis of audio description. In Díaz Cintas et al. (Eds.), 151-174.

Schubert, Christoph (2009). *Raumkonstitution durch Sprache*. Tübingen: Max Niemeyer.

Schwarz, Monika (1992). *Kognitive Semantiktheorie und neuropsychologische Realität.* Tübingen: Max Niemeyer.

Seibel, Claudia (2007). La audiodescripción en Alemania. In Catalina Jiménez Hurtado (Ed.), 167-178.

Seiffert, Anja (2005). *Räumliches Hören. Eine schemaorientierte Analyse der audiodeskriptiven Darstellung der Handlungsräume.* In Ulla Fix (Ed.), 67-86.

Selting, Margaret, Auer, Peter, Barden, Birgit, Bergmann, Jörg, Couper-Kuhlen, Elizabeth, Günthner, Susanne, Meier, Christoph, Quasthoff, Uta, Schlobinski, Peter & Uhmann, Susanne (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Lebende Sprachen 173*, 91-122.

Snell-Hornby, Mary (2006). *The turns of translation studies: New paradigms or shifting viewpoints?* Amsterdam/Philadelphia: John Benjamins.

Snyder, Joel (2008). Audio description: The visual made verbal. In Jorge Díaz Cintas (Ed.), *The Didactics of Audiovisual Translation*. Amsterdam/Philadelphia: John Benjamins, 191-198.

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

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

Stutterheim, Christiane von & Klein, Wolfgang (2002). Quaestio and L-perspectivation. In Carl F. Graumann & Werner Kallmeyer (Eds.), *Perspective and Perspectivation in Discourse*. Amsterdam/Philadelphia: John Benjamins, 59-88.

Stutterheim, Christiane von & Kohlmann, Ute (2001). Beschreiben im Gespräch. In Klaus Brinker, Gerd Antos, Wolfgang Heinemann & Sven F. Sager (Eds.), *Text- und Gesprächslinguistik / Linguistics of Text and Conversation (Halbbd. 2 / Vol. 1)*. Berlin: de Gruyter, 1279-1292.

Szarkowska, Agnieska 2011. Text-to-speech audio description: towards wider availability of AD. *JoSTrans: The Journal of Specialised Translation 15*, 142-162. Retrieved from http://www.jostrans.org/issue15/issue15_toc.php.

Szarkowska, Agnieska & Jankowska, Anna (2012). Text-to-speech audio description of voiced-over films. A case study of audio described *Volver* in Polish. In Elisa Perego (Ed.), *Emerging topics in translation: Audio description*. Trieste: Edizione Università di Trieste, 81-98. Retrieved from http://hdl.handle.net/10077/6356.

Talmy, Leonard (2000). *Toward a cognitive semantics (vol. 1): Concept structuring systems*. Massachusetts/London: MIT Press.

Tanis Polat, Nilgin (2013). *Raum im (Hör-)Film. Zur Wahrnehmung und Repräsentation von räumlichen Informationen in deutschen und türkischen Audiodeskriptionstexten.* Berlin: Frank & Timme.

Vandaele, Jeroen (2012). What meets the eye. Cognitive narratology for audio description, *Perspectives: Studies in Translatology 20(1)*, 87-102.

Vercauteren, Gert (2007). Towards a European guideline for audio description. In Díaz Cintas et al. (Eds.), 139-149.

Vercauteren, Gert (2012). Narratological approach to content selection in audio description. Towards a strategy for the description of narratological time. In Agost et al. (Eds.), 207-231.

Weißbach, Marleen (2012). Audiodeskription und Hörfilme. Eine kontrastive Analyse der deutschen und englischen Audiodeskription des Filmes *Brokeback Mountain*. In Panier, Anna, Brons, Kathleen, Wisniewski, Annika & Weißbach, Marleen: *Filmübersetzung. Probleme bei Synchronisation, Untertitelung, Audiodeskription.* Frankfurt am Main: Peter Lang, 347-409.

Whitney, Paul (2001). Schemas, frames, and scripts in cognitive psychology. In *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier Science, 13522-13526.

Wiklund, Mari (2014). The realization of pitch reset in Finnish print interpreting data. *Text & Talk 34(4)*, 491–520.