A Corpus-based Analysis of Audio Description

Andrew Salway

Submitted to be considered for inclusion in Selected Proceedings of Media for All

Abstract

This paper presents the beginning of a corpus-based investigation into the language used for audio description. The automated analysis of audio description scripts for 91 films was successful in characterising some idiosyncratic features of what appears to be a special language. Our investigation also began to create an empirically-grounded overview and classification of the main kinds of information provided by audio description. The existence of a special language is explained in part by the fact that audio description is produced by trained professionals following established guidelines, and its idiosyncrasies are explained by considering its communicative function – in particular that it is being used to tell a story. Encouraged by the relatively high degree of regularity observed in the corpus, we go on to speculate about the application of language technologies for 'assisted audio description' and for repurposing audio description as a basis for indexing digital video archives.

1. Introduction

Audio description is crucial to improvements in media accessibility for blind and visually impaired people. At risk of oversimplification, it can be said that whereas subtitles improve media accessibility by letting audiences read what they cannot hear, audio description lets audiences hear an account of what they cannot see. Recent social, legal and technical developments mean that the amount of audio description being produced globally is rising

steeply, and this offers exciting and important challenges for research in audiovisual translation. The production of audio description fits well within Orero's definition of audiovisual translation which encompasses "all translations – or multisemiotic transfer – for production or postproduction in any media or format, and also the new areas of media accessibility" (Orero 2004, p. viii). More specifically, audio description is a kind of intersemiotic translation (Gambier 2004) involving a transfer from visual to verbal information. It is important to note though that source texts such as television programmes and films are complex mixes of codes carried by audio and visual channels, so that audio description, acting as a surrogate for the visuals, must interact appropriately with the existing dialogue and sound (Ballester Casado, in press).

This paper investigates the language used in audio description of films, overviews the kinds of information audio description provides and aims to test the hypothesis that audio describers use a special language that is shaped by the communicative needs placed on audio description. So-called functional explanations of language registers, such as special languages, investigate the mapping between the communicative needs of language users and the prevalence of idiosyncratic linguistic features (Hoffman 1984). Our approach is to use automated corpus analysis techniques to identify idiosyncratic linguistic features in a collection of audio description scripts. By way of background information, Section 2 attempts to provide a snapshot of the rapidly developing real-world state of audio description professional practice, uptake, teaching and academic research. In particular, factors that may impact on the language used by audio describers are highlighted. Section 3 presents and discusses results from the analysis of a corpus of audio description scripts for 91 films. The discussion concentrates on relating the idiosyncratic features observed to the communicative needs of the users of audio description, and on giving an overview of the

different kinds of information conveyed by audio description. Encouraged by the relatively high degree of regularity observed in the language of audio description, Section 4 speculates about the application of language technologies for 'assisted audio description' and for repurposing audio description as a basis for indexing digital video archives. Section 5 closes the paper by noting the need and opportunity for diverse research strategies to understand more about the fascinating processes by which the visual is made verbal.

2. Background

Audio description is a description of visual information delivered via an audio channel. Among other things, it helps blind and visually impaired viewers to enjoy films and television programmes, including dramas, comedies, soap operas, documentaries, and sometimes live news broadcasts. The practice of audio description originated in US theatre in the 1980's, and these days it is becoming available in museums and art galleries, and at sporting events as well as on television and in cinemas. During the last few years, several factors have contributed to an ongoing rapid growth in activity around audio description for television and cinema[1]. Most importantly, lobbying by associations of the blind and visually impaired has led to legislation and regulation requiring broadcasters to provide audio description. There have also been important technical developments to facilitate the delivery of audio description with television broadcasts, in cinemas and on VHS and DVD releases. Given the fast pace of developments in all aspects of the global audio description scene, I cannot claim that the review here is in any way complete, but I will highlight some major points of development and some issues relevant to audiovisual translation practice and research. I must also acknowledge that, due to the limits of my knowledge, the review has a British and Anglophone bias.

[1] Also known as audiodescription, described video and video description.

In the gaps between existing speech, audio description for films provides enough information about what is depicted on-screen for the audience to follow the story being told by the film. Audio describers must ensure that the information they provide interacts appropriately with the dialogue and other sounds. Most importantly, the description is restricted by the amount of time available in between the dialogue. Audio description is scripted before it is recorded: the following is an excerpt from the audio description script for *The English Patient*, produced by the Royal National Institute of the Blind (RNIB) – the timecodes, shown as [minutes:seconds], indicate when the description is spoken.

[11:43] Hanna passes Jan some banknotes
[11:55] Laughing, Jan falls back into her seat
[12:01] An explosion on the road ahead
[12:08] The jeep has hit a mine

In effect, audio description acts as a surrogate for the visual components used to tell stories with film, be they factual or fictive. It might be then that the language of audio description is shaped in part by its narrative function. Narrative is usually defined as involving chains of events in cause-effect relationships occurring in space and time. Bordwell and Thompson, discussing narrative in films, append to this definition a statement about how the agents of cause-effect are characters with goals, beliefs and emotion (Bordwell and Thompson 1997, pp. 90-96). We could expect then that audio description will concentrate on providing sufficient information so that, along with the dialogue, audiences are able to understand who is doing what, where and why. For example it may be important to know the emotional reaction of a character to an event that has just happened because this explains their subsequent actions.

Audio description for television and cinema is produced by trained professionals who use dedicated software systems and typically follow established guidelines, be they in-house,

national or international. It may take 50-60 hours and many viewings to produce audio description for a 2-hour feature film, whereas a 30-minute soap opera which is almost full of dialogue and has familiar scenes may take only 90 minutes to describe. The production of audio description can be a collaborative process with teams of describers working on the description of a single film or television series. Experienced describers typically edit the audio description scripts of others before they are recorded.

In the UK, audio description for television has been driven by legislation, now under the auspices of the government's Office of Communications (OFCOM). The 2003 Communications Act requires terrestrial, satellite and cable television broadcasters to provide audio description for 10% of their broadcasts by the 10th year of their license; the Royal National Institute of the Blind (RNIB) is campaigning for this level to be 50%. Audio described television can now be received with commercially available set-top boxes and with PC TV-receivers. Some film distributors have been proactive in providing audio description and now most major film releases in the UK are audio described. Over 160 cinemas have facilities to provide audio description in at least one of their screens and it is included on some DVD releases. The RNIB has produced and makes available a library of over 150 audio described films on VHS tapes. For the latest information about audio description availability in the UK, as well as advice for producing audio description, see the RNIB's webpage[2]. Major producers of audio description for television and cinema in the UK include ITFC, BBC Broadcast and IMS. With regards to the global picture, it should be noted that audio description was trialled in Japan by NTV in 1983, in Spain by Televisión de Cataluña in the late 1980's and since 1990 it has been made available in the US by WGBH. There have also been developments in Canada, France, Germany, Australia and Ireland. For

[2] www.rnib.org.uk/xpedio/groups/public/documents/publicwebsite/public_audiodescription.hcsp

links to global audio description information, see the Audio Description International webpage[3]. For information about US activity see the webpages of the American Foundation for the Blind, the National Center for Accessible Media, Audio Description Associates and National Captioning Institute; for Canadian activity – the Canadian Network for Inclusive Cultural Exchange and AudioVision Canada; and for Australian activity – the Australian Caption Centre.

Audio description for television and film in the UK was pioneered in the early 1990s by the AUDETEL project (Lodge and Slater 1992, Lodge, Green and Nunn 1994), which brought together broadcasters, technology manufacturers and organizations representing the interests of the blind and visually impaired. The issues addressed by the project included the requirements of the target audience (what should be described and in what style), and the logistics of producing and delivering audio description. Through questionnaires, focus groups and user trials, a set of guidelines for audio description was produced. These form the basis of the current guidance on standards for audio description provided by OFCOM[4]. As well as giving advice for scripting, editing and recording audio description, and suggesting strategies for working with different kinds of material, these guidelines also give specific advice about what to describe and how. They recommend use of the present tense, avoidance of potentially ambiguous pronouns, use of adjectives only where they convey relevant information, and use of adverbs to enhance the description of an action, e.g. 'she stamps her right foot impatiently' – but it is stressed that care must be taken not to use subjective adjectives and adverbs. The guidelines also note using verbs that refer to specific ways of doing something can be an effective way of enhancing a description, for example rather than just saying 'she walks' a describer could say 'she swaggers / lopes / tiptoes / marches'. Three issues that recur whenever audio describers discuss the question

[3] www.adinternational.org
[4] www.ofcom.org.uk/tv/ifi/guidance/tv_access_serv/audio_description_stnds/

of best practice are: (i) how to ensure that description strikes the right balance between frustrating the audience with insufficient information to follow the story, and patronizing them by spelling out obvious inferences; (ii) how to maintain objectivity in audio description; and, (iii) how to voice audio description appropriately. Of course, these are problematic questions, the answers to which vary according to the audience's needs and cultural experience, and according to the kind of material being described.

Looking to the future there is the potential for audio description to benefit the whole television and film audience, for example allowing people to enjoy a film on an audio CD or on a device with a small visual display, like a mobile phone. If this was realized it would lead to a huge increase in demand for audio description. There is also the possibility that audio description will come to be seen as a resource for indexing television programmes and films in digital libraries, a possibility discussed in Section 4, and this too would increase demand. Given the growing surge of interest in the production of audio description, it is timely that recent initiatives have included it as a topic within audiovisual translation education and research. There are at least four University programmes in audiovisual translation in which audio description features prominently – in the UK at the University of Surrey and the University of Roehampton, and in Spain at the Universitat Autònoma de Barcelona and the Universidad de Granada. A number of recent conferences in the area of audiovisual translation have included sessions for audio description – 'In So Many Words' (London 2004), 'Languages and the Media??' (????), 'Media for All' (Barecelona 2005), However at these events audio description still feels like a minority interest compared to subtitling and dubbing.

The first research in audio description for television and cinema, such as the AUDETEL project, naturally focussed on eliciting the requirements of blind and visually impaired people,

developing good practice and solving technical, economic and legal challenges for the delivery of audio description. Since then, only two papers have been published that analyse samples of audio description to investigate what information audio description actually provides. Working in the context of media accessibility Piety adapted spoken discourse analysis techniques in his analysis of the audio description of four films, totalling 23,000 words (Piety 2004). He proposed a set of concepts to analyse the structural and functional properties of audio description which included: four structural elements – insertions, utterances, representations and words; and, seven types of information – appearance, action, position, reading, indexical, viewpoint and state. Piety went on to discuss how these concepts could be used to compare and evaluate audio description practices. Working in a rather different context, Turner (1998) was interested in the potential to reuse audio description scripts as a basis for automatically indexing television programmes and films in digital video collections. His analysis focussed on two issues that would determine this potential: (i) how well aligned audio description is with the visual content; and, (ii) what aspects of visual content are described. He analysed the audio description and accompanying visual content in 27 minutes of each of a television documentary, a television drama and a feature film. For this analysis, 15 types of information conveyed by audio description were defined: physical description of characters, facial and corporal expressions, clothing, occupation and roles of the characters, attitudes of the characters, spatial relationships between characters, movement of the characters, setting, temporal indicators, indicators of proportions, décor, lighting, action, appearance of titles, text included in image. While the ideas proposed by Piety and by Turner both provide good starting points in elucidating what audio description does and how it works, it may be that they are limited by the relatively small samples of audio description available to them.

The research presented in the current paper was carried out as part of the 3-year TIWO (Television in Words) research project at the University of Surrey. The overall aim of the project was to develop a computational understanding of storytelling in multimedia contexts, with a focus on the processes of audio description. The research team at Surrey worked in conjunction with a project Round Table comprising senior representatives of organisations that produce audio description (ITFC and BBC), a manufacturer of technology for the production and delivery of audio description (Softel) and the Royal National Institute of the Blind. There were three main strands to the research: first, a corpus of audio description scripts was gathered and analysed to investigate the language used in audio description; second, ideas for increasing the role of information technologies in the production of audio description were developed; and third, systems that repurposed audio description as a basis for retrieving and browsing video data were prototyped and evaluated. Project reports and research publications are available from the TIWO webpage[5].

3. Analysing a Corpus of Audio Description

To summarise and recap from Section 2, some factors that may impact on the language used for audio description of films are:

- audio description for films refers to a restricted domain of discourse, which is what can be seen on-screen;
- it fulfils a communicative function which is to provide enough information for the audience to follow the story told by a film, without patronizing the audience by spelling out obvious information that could be inferred;

[5] www.computing.surrey.ac.uk/personal/pg/A.Salway/tiwo/TIWO.htm

- audio description must provide information about events in cause-effect relationships occurring in space and time, and about the characters involved in the events and their emotional states
- audio description must combine with existing dialogue which, among other things, means it must be concise; and,
- these requirements are embodied in the training and guidelines followed by the professionals responsible for producing audio description.

The use of language, by trained professionals, to communicate about a restricted field of discourse and for a specific purpose, normally results in a *special language* characterised by a preponderance of linguistic features that are idiosyncratic in comparison with everyday general language. Here we investigate the hypothesis that the language used for audio description is a special language, in other words that there exists a 'language of audio description'. Following the kind of corpus linguistics approach described by Biber, Conrad and Reppen (1998), the analysis presented here seeks to identify and describe a special language in terms of statistically significant differences between linguistic features in a corpus of audio description scripts and a general language sample. We also overview the kinds of information that are typically provided by audio description.

Crucial to such an approach is the compilation of a representative corpus, i.e. a collection of texts that adequately represents the language under analysis. The corpus analysed here is limited to British English audio description for films, but within those boundaries we sought to select audio description scripts that would cover the different ways films are described. In consultation with two senior audio describers we defined nine film genres, not so much in terms of the

properties of the films, but in terms of how the describers thought the language used for audio description would vary. The TIWO Audio Description Corpus includes complete audio description scripts from all these genres which were acquired from three major producers of audio description: Royal National Institute of the Blind, ITFC and BBC Broadcast. The corpus comprises 91 audio description scripts totalling 618,859 words, Table 1. Our analysis of the corpus was automated using the text analysis package *System Quirk*[6].

Table 1 Composition of the TIWO Audio Description Corpus (September 2005)

Genre	***Number of films***
Action	10
Children's (animation)	8
Children's (live action)	4
Comedy	13
Dark	8
Period Drama	9
Romantic	11
Thriller	7
Miscellaneous	21
TOTAL	91

3.2 Unusually Frequent Words

Let us begin by examining the most frequent words in our audio description corpus, Table 2. As in all corpora, both general and special language, the very most frequent words are grammatical words like *the*, *in*, *and*, *a*, etc. The first sign of a contrast with general language is the relative preponderance of non-grammatical words within the top 100: in general language corpora, like the British National Corpus, there tend to be 2-3 such words in the top 100. Many of the top 100 words in the TIWO corpus look like concrete nouns and verbs that refer to material processes. Right away, these words give an impression of what is commonly referred to by the audio description (the numbers in brackets are the words' frequencies in the TIWO corpus), i.e.

[6] www.computing.surrey.ac.uk/SystemQ/

characters and their body parts – *man* (1491), *head* (1268), *face* (1145), *eyes* (1111), *hand* (1089), *hands* (814), *men* (742), *woman* (573) as well as character names; actions – *looks* (2482), *turns* (1400), *takes* (1074), *walks* (986), *goes* (728), *stands* (721), *steps* (711), *smiles* (689), *stares* (659), *puts* (649), *watches* (646), *opens* (623), *looking* (620); and, objects and scenes – *door* (1913), *room* (1099), *car* (922), *window* (685), *table* (635), *water* (616), *bed* (577), *house* (574).

Table 2 The 300 most frequent words in the TIWO Audio Description Corpus (Sept. 2005).

Position in frequency-ordered wordlist	***The 300 most frequent words in the Surrey Audio Description Corpus***
1-100	the, in, a, and, out, of, to, his, he, her, 's, on, at, she, up, with, as, is, him, it, into, down, back, from, looks, they, over, then, by, door, off, through, are, man, you, them, turns, for, away, an, head, one, i, their, face, eyes, room, now, hand, towards, takes, around, walks, who, two, car, behind, that, sits, across, hands, other, but, look, 't, white, pulls, men, goes, front, stands, side, steps, smiles, onto, tom, window, outside, open, inside, stares, has, puts, another, watches, table, all, round, opens, looking, water, along, again, way, bed, house, woman, black, music, later
101-200	runs, gets, john, under, young, stops, this, sitting, slowly, comes, some, red, david, change, picks, still, moves, floor, light, dark, sees, against, next, go, holds, small, its, which, after, just, no, beside, starts, little, me, long, where, right, hair, past, himself, arms, before, gun, top, wearing, what, not, there, wall, arm, boat, heads, each, street, mike, smile, large, nods, scene, more, feet, be, see, howard, standing, about, jack, can, get, like, above, road, come, have, girl, forward, blue, leans, ground, glances, gives, mouth, holding, watch, beat, grabs, horse, we, night, george, leaves, falls, grace, reaches, here, so, phone, close, boy
201-300	willard, desk, jim, harry, was, when, watching, shot, moment, glass, crowd, gerry, makes, drops, old, do, air, closes, lifts, end, throws, grey, shakes, stand, your, shoulder, green, left, while, danny, hat, lights, passes, run, appears, been, annie, office, book, three, follows, high, bob, between, officer, people, wooden, bag, lady, stairs, chair, suddenly, home, paper, seat, climbs, postman, kiss, something, pushes, others, kisses, body, will, both, don, take, dormer, lying, tries, robin, oh, my, walk, together, sky, fire, police, soldiers, if, guard, tears, move, lies, jacket, staring, michael, cole, leaving, bedroom, ellen, johnny, mr, corridor, carrying, father, stop, frank, almasy

NOTES: Words that are clearly 'artefacts' of how audio description scripts are written and formatted have been removed, e.g. a describer's notes that are not spoken when the audio description is recorded. Other words, like *in* and *out* have had their ranking boosted because they are used in formatting, but they are also frequently spoken. Some audio description scripts include fragments of film dialogue (used by the describers to cue the recording of the audio description) – these are included, but we believe do not have a significant effect on results.

Of course, when analysing these words out of their contexts we must not be too hasty to assume anything about their usage: many of them are, in general language at least, highly polysemous, members of several word classes, and often used in phrases. In the next section we will analyse some of these words in their contexts, but for now, note that most of the non-grammatical words in positions 101-300 also fit comfortably into the same groupings as above:

- Characters and their body parts: *man, head, face, eyes, hand, hands, men, woman, hair, arms, arm, feet, girl, mouth, boy, crowd, shoulder, officer, people, lady, body, police, soldiers, father*
- Actions: *looks, turns, takes, walks, goes, stands, steps, smiles, stares, puts, watches, opens, looking, runs, sitting, comes, picks, sees, holds, wearing, smile, nods, standing, leans, glances, gives, holding, watch, beat, grabs, leaves, falls, reaches, watching, drops, closes, lifts, throws, shakes, passes, run, follows, climbs, kiss, pushes, kisses, walk, lies, staring, carrying*
- Objects and scenes: *door, room, car, window, table, water, bed, house, floor, gun, boat, street, road, ground, horse, phone, desk, hat, office, book, bag, stairs, chair, seat, sky, fire, jacket, bedroom, corridor*

Frequency alone draws an interesting set of words to our attention and gives us some insight into the vocabulary of audio description. However, we may also be interested in words, that whilst not highly frequent in the audio description corpus are nevertheless being used in audio description much more than they are in general language. A statistic proposed by (Ahmad and Rogers 2001) identifies words that, whilst not necessarily among the most frequent in a corpus are nevertheless appearing in the corpus at much higher rates than they do in general language. The statistic is sometime referred to as the SL/GL ratio (where SL = special language and GL = general language) and it is calculated by diving the *relative frequency* of a word in the special language corpus by its relative frequency in a general language corpus. Relative frequency is calculated by dividing the frequency of a word in a corpus by the total number of words (tokens) in the corpus. An SL/GL ratio equal to 1 means that a word is being used 'normally', i.e. as relatively often in the special language corpus as in the general language sample. In contrast, an SL/GL ratio of say 50 means that a word is being used relatively 50 times more often in the

special language. Words with high SL/GL ratios are being used with idiosyncratic frequency in a special language corpus and therefore demand closer inspection. Table 3 shows a selection of words with high SL/GL ratios in the TIWO corpus: here the general language sample is the British National Corpus. In an attempt to bring to attention words that are used across more than a few audio description scripts only words with a frequency in the corpus of at least 30 are included – nevertheless some 'rogue' words appear, for example *llama* and *periscope* make it into the list because they are very frequently referred to in just one film.

Table 3 Words with SL/GL ratio > 25 and Frequency > 30 in the Surrey Audio Description Corpus (Sept. 2005).

SL/GL Ratio	
>100	saunters, hurries, stares, shoves, clambers, straightens, gazes, kneels, scrambles, leans, glares, nods, periscope, strolls, crouches, tosses, blinks, trots, frowns, hurls, clunk, grabs, pulls, llama, watches, smashes
50-100	unlocks, hauls, staggers, heaves, minion, stumbles, shakes, wipes, hesitates, pats, haired, lowers, pushes, wanders, crawls, grins, glances, flings, picks, flicks, slaps, hugs, smiles, sniffs, glides, scarecrow, sits, slams, rubs, pours, squeezes, diner, postman, spins, shuts, salutes, drags
25-50	rips, walks, climbs, closes, sips, strides, slumps, gallops, flashback, leaps, knocks, throws, fades, stirs, rushes, kisses, tugs, creeps, jumps, dives, shrugs, crashes, lifts, turns, licks, opens, silhouetted, elevator, pauses, swings, sighs, bounces, stops, dials, swims, bangs, presses, slips, removes

NOTES: The words are ordered by their SL/GL ratios. Words with frequency > 100 are underlined. Character names which typically have high SL/GL ratios have been excluded.

Table 3 includes some words that we saw already in Table 2. However, the SL/GL statistic brings to attention other words that characterise the language used in audio description, many of which refer to actions:

saunters, hurries, shoves, clambers, straightens, gazes, kneels, scrambles, glares, strolls, crouches, tosses, blinks, trots, frowns, hurls, smashes, unlocks, hauls, staggers, heaves, stumbles, wipes, hesitates, pats, lowers, wanders, crawls, grins, flings, flicks, slaps, hugs, sniffs, glides, slams, rubs, pours, squeezes, spins, shuts, salutes, drags, rips, sips, strides, slumps, gallops, leaps, knocks, stirs, rushes, tugs, creeps, jumps, dives, shrugs, crashes, licks, pauses, swings, sighs, bounces, swims, bangs, presses, slips, removes

3.2 Concordances of Some Unusually Frequent Words

By examining concordances of some of the frequently occurring words identified in the previous section we learn more about how they are being used in audio description and we can begin to characterise the kinds of information that audio description tends to provide. The focus here is on commonly occurring phrases that include the frequent words. These have been grouped, in a preliminary attempt at a classification, according to how they give information about:

- characters' appearances;
- characters' focus of attention;
- characters' interpersonal interactions;
- changes of location of characters and objects; and,
- characters' emotional states.

It must be emphasised that what is presented here is very much a preliminary analysis of what is vast quantity of concordance data – there are over 27,000 instances of the 29 top 100 non-grammatical words alone.

Often when a character first comes on screen in a film they are introduced in the audio description with a relatively simple description of their appearance, for example their age, clothing or some distinctive feature. Some common phrases are: *woman in* and *man in* followed by an item of clothing, or an age; *woman / man wearing* followed by an item of clothing; and *woman / man with* followed by a distinctive physical feature. In some cases *man* and *woman* are preceded by an adjective relating to the character's appearance, e.g. *young*, *old*, *elderly*, *short*, *tall*, *burly*, *bearded*, *burly*:

a man in a white T-shirt leans towards Jim
he sees a man and a woman in a red suit walk by
the door of a low-rise brick apartment building opens and a woman in her thirties steps out

a dark-haired man with a moustache stands at the door
an old woman with a pointed nose and wild, white hair stands in a gloomy room

The words *looks* and *looking* are often used in phrases like *looking at*, *looks at*, *looks up at*, *looks down at*, *looks around*, etc. These all tend to provide information about a character's current focus of attention, consider for example some concordances of *looks at* which appears 487 times in the corpus:

Corelli looks at his men
Samuel looks at the blue and black picture
Iris looks at John curiously as he puts down his cup
he picks up a striped scarf and looks at her
he looks at a painting on the wall
Memphis looks at the five young men
he looks at Harry thoughtfully
Lucy looks at him for a moment, then kisses him

Other words appear to be used in a similar way to *looks*, to indicate a character's focus of attention but are more specific about how the manner in which the character is looking – *stares at*, *glances at* and *gazes at*. Phrases formed with the words *eyes* and *head* also indicate focus of attention:

young Parker keeps walking, his eyes fixed on Mary Jane
keeping her eyes fixed on Ben, she walks over
Willard opens his eyes and absently regards the fan
slowly she turns her head to face the door
she sighs then turns her head sharply and looks behind her for a long time

A preliminary examination of the concordances of *looks at* and the like suggests that they are used when a character's attention is on another character or object. In contrast, *watches as* seems to be used to indicate a character's attention to something happening:

the girl watches closely as he carefully puts the needle on the record
David watches through his binoculars as Kirgo puts the coin in a flight bag

Many instances of characters' attention of focus involve them focussing on other characters; there is some overlap then with our proposed category of "characters' interactions". However, certain phrases like *turns to*, *shakes hands, sits next to, their eyes meet, puts hand on, gazes into each others eyes* and *turns away* are perhaps more indicative of key moments in an interaction:

the captain turns to Gatlin
she turns to Drosoula, who glares at her
Pelagia moves forward into the room, then turns to look back at him
they shake hands, Diane nods once quickly, smiling
Stitch sits next to an elderly lady and takes her hand
their eyes meet for a moment, then she turns to close the curtain
Jules puts her hand on Ellen's shoulder

Other key information about how an interaction between characters is proceeding is commonly given with phrases like *smiles at*, *shakes his / her head* and *nods* with an adverb:

Luc smiles at her conspiratorially
as Ellen smiles at her father, her eyes moisten
Tess shakes her head and swallows nervously
Ellen looks steadily at him and shakes her head
Michael nods, tentatively
Prince John nods, approvingly
Annie nods and Tom grins to himself

For the audience it is crucial to keep track of the location, and presence / absence on-screen of characters and key objects. This broad class of information is expressed with quite a large number of frequently occurring phrases in audio description. Characters' changes of location, typically within a scene, are expressed with phrases including *goes to / into / off / out*, *walks away / off / out / over to* and *steps towards / into / onto*. These actions may be preceded by a character standing up – *stands up* occurs frequently in the corpus, or completed when a character *sits down*. The opening and closing of doors are frequently described – *opens / closes the door*, *door opens / closes*, as a characters moving *through the door*, and are often connected with characters entering and leaving scenes; similarly when a vehicle *pulls up* or *drives off*. Objects

change location, and sometimes possession, when one character *hands* something to another a character, or when a character *picks up* or *pulls out* an object, or *puts* an object somewhere:

> quietly she gets out of bed and goes to the window
> Luca hangs his head solemnly and steps towards a bench with Mary
> Corelli looks at him coldly, then turns and walks away
> he stands up and goes back into the main room
> Gogan steps in through the door
> the next day, a white car pulls up at the house
> she picks up a jar from the kitchen table
> she hands him a rucksack

As well as keeping track of where characters are, what they look like and who they are interacting with, in order to properly appreciate a story an understand must understand something about characters' emotional states. Of course in some instances it will be possible for the audience to infer this either because of what has happened to a character, or from the dialogue. However, our corpus analysis reveals some commonly recurring phrases that seem to be used to convey characters' emotional states. The most straightforward way this is described is by saying a character *looks* or *is looking* followed by an adjective such as *confused, shocked, surprised, thoughtful, troubled, uneasy, annoyed, puzzled, concerned, dejected*:

> Mrs Mills looks confused, then recovers herself
> Rebecca looks blissfully happy, Samuel doesn't
> Noelle's looking faintly embarrassed and puts her hand over her lips

The words *smiles*, *stares* and *walks* all occur very frequently in the corpus and can be modified with adverbs to indicate an emotional reaction to the events affecting a character – *smiles contentedly / fondly / happily / sadly / shyly / wryly*, *stares blankly / coldly / curiously / proudly / uncertainly / in confusion / in disbelief*, *walks briskly / calmly / slowly / stiffly*. Less frequent than *smiles* are *frowns* and *grins*. In a different sense of the word, *looks* is also modified with adverbs – *looks anxiously / nervously / desperately*.

> Vianne closes the window and turns to Anouk who smiles contentedly
> her smile fades and she stares blankly upwards

he stops walking and stares in confusion at the mass of demonstrators

Actions involving characters' heads, faces and eyes also give information about their emotional states:

Billy's young face breaks into a wide smile
she leans back and her face crumples in despair
Thurman strides across to the dock, his head held high
Ellen's eyes fill with tears and she smiles sadly at Kate

3.3 The Expression of Temporal Information in Audio Description

A key aspect of the definition of narrative is that it involves events organised in a temporal sequence and so it may be interesting to consider how audio description conveys temporal information. Of course the written audio description script contains timecodes which synchronise the speaking of utterances with the film – in effect the order of speaking is the major source of temporal information for the audience – if one event is described before another then it is assumed to have happened (in film time) before the other. Audio description guidelines specify the use of present tense and this seems to be used in the vast majority of audio description utterances: the simple present tense appears to be used interchangeably with the present continuous. Occasionally, the present perfect is used when only the end-state of an event is depicted on-screen, or to identify unnamed characters:

Pike has killed Grueller, but another vampire is after him, as Buffy stakes her opponent
the short man who has been pursuing Darby meets with Sneller

There are however other indicators of temporal information that seem to be favoured by audio describers. The words *stops, starts, begins* and *finishes* all occur relatively much more frequently that in general language as evidenced by their SL/GL values, Table 4. They are often used to indicate the end points of events:

Jerry skips for a while, then stops and stares at the rope

Lynn stops fiddling with the wall temperature gauge
he walks back to the party and starts dancing
slowly she starts to climb the stone staircase
she discovers it is a letter and begins reading
the guard moves again, and Griff begins to climb the ladder
Hester finishes decoding the Kestrel signals with the Enigma machine

Table 4 Words for expressing temporal information that are unusually frequent in the Surrey Audio Description Corpus (Sept. 2005).

Word	***Frequency***	***SL / GL Ratio***
stops	519	28.9
starts	426	9.3
begins	140	3.8
finishes	31	5.0

The use in audio description of other words that can indicate when events happen was reported by Salway and Tomadaki (2002); that analysis was based on a corpus of audio description for 12 films comprising 70,852 words. A first analysis of the current TIWO corpus supports the previous findings. Although words to express temporal information did not appear unusually frequently compared with general language, and in fact words like *before* and *after* were relatively infrequent, they did appear to be used in restricted ways compared with general language. The words *then* (2086) and *now* (1090) when used were redundant as far as adding information about temporal ordering, because the order of events is implicit in the ordering of the utterances. However, *then* often implies the completion of the first event and *now* suggests a contrast with a previous state or event:

he watches for a moment then continues walking
the major throws the towel to the ground, then dives in after Mss Bentley
the businessman is now rather less smug
the terrified prisoners are now running as they go about their work
Boulton walks into the common room, now wearing a Royal Navy lieutenant commanders uniform

Both *as* (3651) and *while* (257) were used to express the simultaneity of two events, but only *as* implies a causal connection between the two events:

the gang take aim as the brothers run up a ladder

he jumps as the troll gives a snort
Lady, Tramp and Jock all giggle, as the puppies play
Buddy follows Ray Bob onto the bus while a couple of musicians bid farewell
Sully hauls himself back inside, while Boo starts hitting Randall with a baseball bat

The timing of one event relative to another was also expressed with *when* (290) and *until* (109). The inclusion of one event within another can be expressed by using a non-finite verb with a subordinate clause:

she walks off then pauses when she realises Jack is lagging behind
Ives looks tense, and when they move off, he's holding his arms
the others watch until Ned interrupts
they attack it with their rifle butts until the door panels smash in and they gain access
coughing, Mary gives the medicine to Tom

The word *again* (593) was generally used to indicate a second instance of an event within a scene of a film:

bending forward, he kisses her again and goes
a rhino-like dinosaur knocks Aladar over again as it plods slowly past

To give information about when events occur in "story time" rather than in "film time", times of day like *morning*, *evening*, *dusk* and *dawn* are sometimes used in the audio description to introduce a new scene, as is *later*. Words referring to specific festival days, months, seasons and years were infrequent. This is perhaps because a lot of information about such things is conveyed in films by costumes, props and lighting and the audio description will concentrate on describing these and allow the audience to make inferences.

3.4 Conclusions and Ongoing Work

The analysis presented here offers several pieces of evidence to support the idea that there is a special 'language of audio description', reflecting a systematic approach to the production of

audio description. There is a substantial set of non-grammatical words that appeared unusually frequently in the audio description corpus. These included a large number of concrete nouns and verbs referring to material processes which reflects the aim of audio describers to only describe what can be seen on screen. The need for concision in descriptions may account for the high occurrence of XXX, i.e. verbs that express a particular manner of doing something, such as *saunters, hurries* and *strolls*. The usage of grammatical features and words to express temporal information was restricted compared to general language, because for the most part audio description describes something happening in the present time. Further analysis, for example of spatial information, and analysis with part-of-speech tagging and lemmatization is required to strengthen the case for a language of audio description. It is also important now to analyse corpora of audio description produced in different countries, where different guidelines and practices apply, and in different languages.

The concordance analysis only skimmed the surface of a vast quantity of data but by concentrating on the most frequent words and phrases it began to give an empirically-grounded overview of the main kinds of information provided by the audio description. That said, there is a great deal of scope for much more detailed investigations to describe more precisely the way in which frequent words and phrases are used, and to refine the classification of what kinds of information audio description provides. A colleague on the TIWO project at the University of Surrey, Andrew Vassiliou, will shortly complete his doctoral research that has focussed on a more systematic and in-depth analysis of concordances from the TIWO corpus in order to describe a local grammar of audio description.

A by-product of analysing the kinds of information commonly provided by audio description is that we also learn something about what events commonly happen in films. Consider, for

example, phrases that describe characters looking at each other and at key objects, phrases that indicate characters changing location and phrases that describe characters' expressions of emotions. It could be argued that the preponderance of the actions described by these phrases makes them important narrative elements for story-telling in films; this idea is elaborated in (Salway, Vassiliou and Ahmad 2005). For this, and other reasons, the analysis of audio description could be of interest to narratologists. Narrative is a multi-faceted phenomenon studied by philosophers, literature and film scholars, linguists, cognitive scientists and computer scientists (Herman 2002) and the study of narrative explores what the media-independent features of stories are, how different kinds of media can convey (the same) stories, and how stories are understood (Ryan 2004). Because audio description has to concentrate on the essential information required to tell a story, its analysis could help to answer questions about what are essential narrative elements and about how stories can be told in combinations of different media. It is also interesting to compare the ways in which audio description gives information about the characters and events of films with other kinds of text that serve similar functions, for example screenplays and plot summaries, and perhaps even novels where available. Vassiliou's work, mentioned previously, includes analysis of a corpus of film screenplays. Another colleague on the TIWO project at Surrey, Eleftheria Tomadaki, has been comparing how audio description and plot summaries refer to the same events in films – she will present the results in her doctoral thesis due in Autumn 2005.

4. Audio Description and Information Technology

In anticipation of a sharp increase in the demand for audio description, and the need for organisations to deliver to tight deadlines whilst maintaining, if not improving, the quality of

their descriptions, it is important to consider the possibilities for information technology to assist in the production of audio description. A rather looming backdrop to the discussion here is of course the limited achievements of automatic translation systems in the past, and the observation that video data is generally considered to be less computationally-tractable than text data. To a computer, moving images consist of pixels of 32,000,000 different colours, organised into frames of say 800x600 pixels, with typically 24 or 25 frames per second. The mapping from this mass of pixels to the meanings inherent in the moving image for a human viewer has been termed the 'semantic gap' (Smeulders et al 2000). The state-of-the-art in computer vision technology is a long way from successfully recognising large sets of objects and actions in relatively unconstrained video data such as films[7]. This rules out, for the meantime, the possibility of anything like fully automating the generation of audio description from video data. On a positive note, the relatively systematic nature of the language of audio description observed in Section 3 is encouraging for the application of language technologies to assist in the production of audio description. Furthermore, the problem of the semantic gap creates a great potential added-value for audio description to be reused as a basis for indexing moving images in digital archives. It can be hoped that if added-value is demonstrated then broadcasters and film distributors would have even greater incentive to produce audio description.

4.1 Some Possibilities for 'Assisted Audio Description'

This section reviews requirements and a list of desired functionality for software to assist in the production of audio description, and discusses the opportunities and feasibility of applying state-of-the-art language and multimedia technologies; this list was elicited during the TIWO project

[7] Perhaps the most recent advances in this area have been realised in the Cognitive Vision Systems project - http://cogvisys.iaks.uni-karlsruhe.de/

from the Round Table. Currently available software, such as Softel's *Adept* system, assists in the preparation of audio description scripts, the recording of audio description and its synchronization with programme and film material. It presents the material on screen alongside a window that marks time-coded speech-free intervals into which the describer can type their descriptions. Once the script has been completed and reviewed, it is spoken, recorded and synchronized with the video data by the time-codes. In considering further functionality that could be provided by information technology it is crucial to remember that any gain in efficiency (speed) of producing audio description, is instantly negated by any lowering of quality which points to human editors always having the final word. The ideas for further functionality are discussed here in order of how technologically feasible I think they are.

A relatively simple addition to current software would be online access to lexical resources such as specialist terminologies and thesauri to help describers find the best words for their descriptions, and to avoid too much repetition. A system could also be proactive in providing links to relevant information on the web, such as pages related to the subject matter of a documentary being described, or online plot summaries and scripts of a film – caution to maintain focus on what is being described. They could also be provided with access to previous descriptions in order to see how certain things have been described before, either for inspiration for their description or to ensure consistency with previous descriptions within a TV series or film (remember team work), e.g. retrieve all scenes with a certain character.

A key aspect of guidelines for maintaining the quality of audio description is the specification of preferences for certain linguistic features including vocabulary and grammar. This raises the possibility for automatic style-checking whether to give feedback to novice describers, to assist editors or as a tool for organizations such as the RNIB that monitor the output of organizations

producing audio description. For example, it could be feasible for a system to check for vocabulary inappropriate to the genre or period of a film, confusing use of pronouns, the over-repetition of vocabulary items and the use of inappropriate tenses and overly complex utterances. For novice describers the feedback could be given as they work, for editors it would be given on a script by script basis, and monitoring organizations may wish to analyse corpora representative of an organization's output to get an impression of how guidelines are being followed.

For some material, like films, pre-existing descriptions of on-screen scenes, characters and actions like film screenplays are available which suggests the idea of creating a first draft audio description automatically. This can be conceived of as a text summarization task, for which there is already an established body of research and technology. On average, screenplays contain about three times as many words as the equivalent audio description, but a corpus-based analysis reported in Salway, Vassiliou and Ahmad (2005) suggests that both text types concentrate on providing the same kinds of information – though of course screenplays also contain dialogue. As a follow-up to the TIWO project we are currently working with the audio description team at BBC Broadcast to develop and evaluate technology for creating a first draft audio description automatically from a screenplay. Our focus is on algorithms to identify the most important things to describe, and then to edit utterances from the screenplays to fit the time available for description and to ensure an appropriate style of audio description. If successful such algorithms might contribute to technology for near live audio description so a describer could speak their description during a first viewing of material, and a system with speech recognition could automatically transcribe and edit what they say. This would also require the adaptation and application of existing multimedia techniques to detect scene boundaries and dialogue-free intervals in video data.

The final set of opportunities and challenges for information technology to be considered here relate to the adaptation of audio description. It is recognized that it is not possible to produce a single audio description that would meet the needs and preferences of all audience members, so perhaps in the future it would be possible to customize descriptions to better suit individuals. Technological developments such as DVD, and its possible successors like BluRay, offer new ways to deliver audio description for example by allowing users to choose alternative description tracks, or accessing more detailed descriptions whilst pausing a film. Alternative tracks might take the form of alternative vocabularies for different age groups, or different quantities of descriptive and interpretive information to suit personal preferences. The cost of producing multiple versions would be prohibitive, but perhaps language technology can be used to automatically adapt a 'master' description either at the time of production (giving the option for human editing) or at the point of delivery giving full control to the audience. Any kind of adaptation would benefit from audio description scripts being marked-up, for example using XML, to make explicit to the machine the relative importance of, and interdependencies between, audio description utterances and dialogue. Such mark-up may also be a starting point for a kind of template-based translation of audio description into different languages, allowing the translator to see quickly what needs to be described. Perhaps the relatively restricted nature of the language of audio description makes automatic translation into a first draft feasible. The hardest adaption-related challenges come from the prospect of programme and film material being edited for reuse after the audio description has been produced. Ideally the original description would be reused but after editing the space available for description may alter, visual content may change, and the original speaker of the description may not be available to record changes.

4.2 Using Audio Description to Increase Machine's Understanding of Films

Whilst improving media accessibility will always be the primary motivation for producing audio description, the coming years may see interest in it for another reason too. As it becomes easier to store and share vast quantities of video data, so the need increases for technologies that can assist in the indexing and retrieval of television programmes and films from digital libraries. In the past archives such as that of the BBC have been for in-house use only, but the advent of the web creates the demand and opportunity to open them up for public access. A minimal requirement is to store production details such as Title, Director and Genre with every programme and film. More useful though is shot- or scene-level indexing whereby keywords are associated with shots and scenes, enabling users to retrieve precise intervals of video data that match their queries, for example "find me all scenes showing a woman on a horse". Creating such indexing manually is prohibitively expensive in many cases, and the challenge of the semantic gap limits the scope for machines to generate keywords by analysis of the pixels in the video data. An increasingly popular strategy is to automate video indexing by exploiting so-called collateral text associated with video material. For example the research prototype *Informedia*[8] and the publically-available *Google Video*[9] use keywords from closed captions to index television programmes. Closed captions can work well for some kinds of programmes, such as news broadcasts, where there is a reasonably close correspondence between the spoken words of presenters and what is depicted in the moving image. However, that correspondence is often more ad-hoc, especially in other kinds of programmes and films, and so the advent of audio description seems especially timely as an alternative source of indexing information.

[8] www.informedia.cs.cmu.edu
[9] http://video.google.com/

By definition audio description includes a verbal account of the most important aspects of the visual information. Developing an idea first proposed by Turner (1998), in TIWO we demonstrated promising results for keyword-based video indexing using audio description: in ongoing work with the BBC Archive we are evaluating the use of audio description for indexing television programmes. The TIWO project also explored new ways of accessing films stored in digital libraries, beyond the kind of keyword-based retrieval described above. Imagine in the future asking your computer to find you a film with a similar kind of story to the one you just watched, or asking it to show you a summary of a film containing the key parts of the story. Such functionality would require the machine to in some sense understand narrative aspects of the film. In TIWO we found that certain narrative-related information could be extracted relatively simply from audio description scripts, for example information about the emotions being experienced by characters at different points in time (Salway and Graham 2003).

Figure 1 shows plots for two different films that were generated automatically from their audio description: the technique was based on mapping the occurrences of about 600 emotion tokens (keywords in the audio description) to one of 22 emotion types. The plot shows each time an emotion token occurs in the audio description; compared to human judgments this technique achieved about 65% accuracy. The extracted patterns of emotion are suggestive of some key aspects of the films' stories. For example in the plot for *Captain Corelli's Mandolin* a high density of positive emotion tokens appear 15-20 minutes into the film, e.g. JOY and LIKE, corresponding to Pelagia's betrothal to Madras. The negative emotion tokens which immediately follow are associated with the invasion of their island. The cluster of positive emotions between 68-74 minutes occurs during scenes depicting the growing relationship between Pelagia and Corelli. The group of FEAR, DISTRESS and SELF-REPROACH tokens between 92-95

minutes maps to a scene in which German soldiers are disarming their former Italian allies, during which a number of Italians are gunned down. Towards the end of the film, the last DISTRESS token appears, followed by a RELIEF token when Pelagia discovers her father has survived an earthquake, and a LIKE token as Pelagia and Corelli are reunited. Our more recent work is attempting to extend the range of narrative-related information we can extract from audio description in order to generate richer machine-level representations of stories (Salway, Vassiliou and Ahmad 2005).

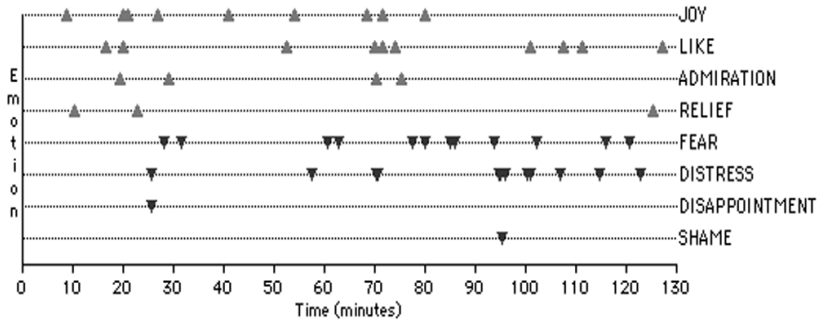

Figure 1a. A plot of emotion tokens found in audio description for Madden's *Captain Corelli's Mandolin*. From Salway and Graham (2003).

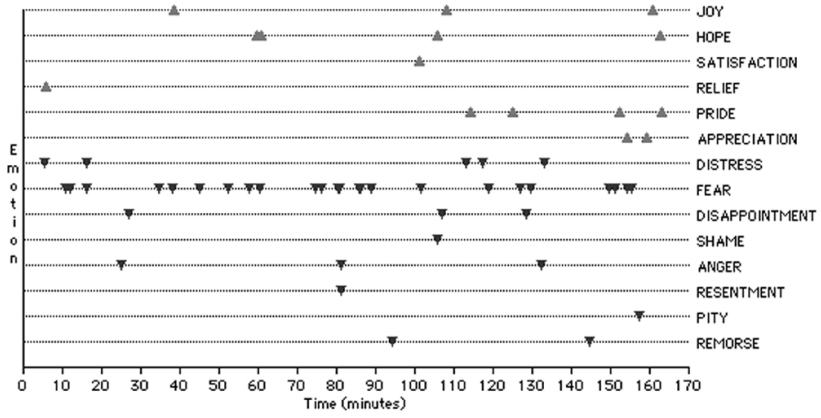

Figure 1b. A plot of emotion tokens found in audio description for Costner's *The Postman*. From Salway and Graham (2003).

5. Closing Remarks

The relationships between the visual and the verbal, between vision and language and between image and text have fascinated scholars in many disciplines for many centuries. Audio description provides a novel and tangible scenario for exploring these relationships, and it offers immediate, important and exciting challenges for research in audiovisual translation. The automated corpus-based approach reported in this paper was successful in characterising some idiosyncratic features of what appears to be a special language of audio description. These features can be explained by considering the communicative needs of the users of audio description. Furthermore, our investigation began to create an empirically-grounded overview and classification of the main kinds of information provided by audio description. The degree of systematicity observed in the audio description scripts should be encouraging to those who produce guidelines for audio description and those who train audio describers. It is also encouraging for those seeking to develop language technologies for 'assisted audio description' and for repurposing audio description to index digital video archives.

Of course, corpus-based approaches must be complemented by other ways to understand the processes of audio description. Future investigations would do well to follow the advice of Piety (2004) and analyse audio description with respect to the visual and audio content of the films it describes. There is of course a trade-off between the wholly automated large-scale analysis of text corpora and the more in-depth, but painstaking analysis of text alongside video. Although considering subtitling in particular, the argument and suggestions made by Remael (2004) for attending to the narrative functioning of film dialogue in audiovisual translation is highly pertinent to the study of audio description. The need for interdisciplinary input into audiovisual or multimedia translation research was articulated by Gambier and Gottleib (2001, p. xii).

Synergies with three of the subject fields that they mentioned have been hinted at in the current paper, namely film studies, information science and computer science. Synergies with narratology and with it cognitive science and artificial intelligence. If investigating processes of audio description, rather than the artefacts, then consider cognitive / language production studies such as Chafe (1980, 1994), and Ericsson and Simon (19XX)

Acknowledgements

This research was conducted as part of the Television in Words (TIWO) project – funded by the UK's Engineering and Physical Sciences Research Council (EPSRC). I have enjoyed and learnt from discussions about the language of audio description with two TIWO colleagues – Andrew Vassiliou and Eleftheria Tomadaki, their analyses which are going much further than the work presented here are mentioned in Section 3.4. Early versions of this paper were presented in the form of a seminar at Universidad de Granada, and a talk at the 'Media for All' conference in Barcelona – I am grateful to all those of offered comments, questions and suggestions.

Throughout the project, the TIWO Round Table was an invaluable source of audio description expertise and provided texts for the TIWO corpus. I give many thanks to, in alphabetical order, Marie Campbell (BBC Broadcast), Martin Davies (BBC Broadcast), Garry Duguid (ITFC), Denise Evans (Royal National Institute of the Blind), Claude le Guyader (ITFC), James O'Hara (ITFC), Dave Harris (ITFC), Jane Westrop (BBC), Jill Whitehead (Royal National Institute of the Blind) and Russ Wood (Softel). I also learnt about audio description in conversations with Joan Greening (RNIB), Veronika Hyks (IMS) and Joel Snyder (Audio Description Associates).

References

Ahmad, Khurshid and Rogers, Margaret (2001) The Analysis of Text Corpora for the Creation of Advanced Terminology Databases. In S. E. Wright and G. Budin, *The Handbook of Terminology Management*. Amsterdam: John Benjamins.

Ballester Casado, Ana (in press) La audiodescripcisn: apuntes sobre el estado de la cuestisn y las perspectivas de investigacisn. *Cadernos de Traducao*.

Biber, Douglas Conrad, Susan and Reppen, Randi (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Bordwell, David and Thompson, Kristin (1997) *Film Art: An Introduction*, 5th edition. McGraw-Hill.

Gambier, Yves and Gottlieb, Henrik (2001) Multimedia, Multilingua: Multiple Challenges. In Yves Gambier and Henrik Gottlieb (eds.) *(Multi)Media Translation: Concepts, Practices, and Research* (pp. viii-xx) Amsterdam/Philadelphia: John Benjamins.

Herman, D. (2002), *Story Logic: problems and possibilities of narrative*, Uni. of Nebraska Press.

Hoffman, Lothar (1984) Seven Roads to LSP. *Fachsprache: international journal of LSP*, 6 (1-2), 28-38.

Lodge, N. K. and Slater, J. N. (1992) Helping Blind People to Watch Television – the AUDETEL Project. In *Proceedings International Broadcasting Convention*, IEE conference 358, Amsterdam, July 1992, (pp. 86-91).

Lodge, N. K. Green, N. W. and Nunn, J. P. (1994) Audetel, Audio Described Television – the Launch of National Trials. In *Proceedings International Broadcasting Convention*, IEE conference 397, Amsterdam, September 1994, (pp. 140-145).

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

Piety, Philip (2004) The Language System of Audio Description: an Investigation as a Discursive Process. *Journal of Visually Impairment and Blindness* 98(8), pp. 453-469.

Remael, Aline (2004) A Place for Film Dialogue Analysis in Subtitling Courses. In Pilar Orero (ed.) *Topics in Audiovisual Translation* (pp. 103-126) Amsterdam/Philadelphia: John Benjamins.

Ryan, Marie-Laure (2004 ed.), *Narrative Across Media: the languages of storytelling*. University of Nebraska Press.

Salway, Andrew and Tomadaki, Eleftheria (2002) Temporal Information in Collateral Texts for Indexing Moving Images. In Setzer, Andrea and Gaizauskas, Robert (eds.) *Procs. of LREC 2002 Workshop on Annotation Standards for Temporal Information in Natural Language*, pp. 36-43.

Salway, Andrew and Graham, Mike (2003) Extracting Information about Emotions in Films. *Procs. 11th ACM Conference on Multimedia 2003*, 4th-6th Nov. 2003, pp. 299-302.

Salway, Andrew Vassiliou, Andrew and Ahmad, Khurshid (2005) What Happens in Films? In *Proceedings of IEEE ICME 2005: International Conference on Multimedia and Expo* IEEE electronic publication.

Smeulders, A. W. M.et al. (2000), 'Content-based image retrieval: the end of the early years'**,** *IEEE Trans. PAMI*, 22 (12), 1349 - 1380.

Turner, James (1998) Some Characteristics of Audio Description and the Corresponding Moving Image. *Proceedings of 61st ASIS Annual Meeting*, vol. 35, pp. 108-117. Medford, NJ: Information Today.