1. Introduction

Audio description (AD) is a verbal commentary of visual content and as such it inherently involves language. The language of AD is governed by a number of factors, which ultimately contribute to its unique nature. Firstly, since AD is usually part of a multimodal text (film, TV, theatre, etc.) and its primary role is to replace the visual mode, it does not function as an independent text, but in conjunction with information conveyed by the other semiotic modes, especially the auditory one: sounds, dialogues and music. This means that cohesion must be ensured between the different semiotic channels. Secondly, save for descriptions of museum exhibitions, AD is not a running text, as it is normally inserted between dialogues and other meaningful sounds (so AD will also have to be cohesive internally, to ensure continuity despite the fragmentation). This is usually done under severe time constraints, as the gaps between dialogues are usually quite short, therefore AD needs to be as concise and meaningful as possible. This has direct implications for both the syntactical and vocabulary choices. Thirdly, AD involves relaying information expressed by a complex non-linear image, which is perceived holistically, by means of words that are strung together in a linear fashion. This is relevant for content selection, on the one hand, and for the actual AD phrasing, on the other. Lastly, AD is a text that is written to be spoken so the rhythm is important, cacophonous sounds are to be avoided and it has to “flow” with the dialogues. All of these constraints have an impact on both the linguistic and textual aspects of AD, which will be the focus of this chapter.

The chapter starts with an overview of the main research approaches to the study of AD as a linguistic and textual construct (section 2), which is followed by an analysis of the syntactic and lexical aspects (section 3). Section 4 discusses textual aspects with a special focus on coherence and cohesion, while in section 5 a comprehensive AD matrix is proposed, which incorporates the issues discussed in previous sections.

2. Research methods

One of the key approaches in investigating the language of AD is through corpus linguistics, which allows for examining (large) bodies of “real world” text samples in order to uncover...
recurring linguistic patterns. In AD the first corpus-based investigation was conducted by Piety (2004) who studied a small corpus of audio descriptions of four films, which altogether contained over 23,000 words. His investigation was aimed at definitions of structural and functional components of audio description and the results “clearly indicate that the language used in audio description is not entirely the same as spoken or written language” (p. 457). Along similar lines, Salway (2007) hypothesises that there exists “a language of audio description” (p. 155), that is a language used for a specific purpose and characterised by some idiosyncratic linguistic features in comparison with general language. In order to test the hypothesis, he examines an AD corpus created as part of the Television in Words (TIWO) project. It includes English AD scripts from 91 films (618,859 words) representing different genres (such as period dramas, thrillers, comedies, children’s). Salway identified the most common words in the corpus, which in fact reflect the special language of AD. He then compared the use of words in the special corpus (the language of AD) with the British National Corpus (representative of general language use) in order to determine the most striking differences between the two. Arma (2011), on the other hand, created a corpus using 69 AD film scripts from the TIWO corpus and focuses in her analysis on the adjectival patterns in AD. In another study, Arma (2012) looked contrastively at English and Italian AD scripts for the film Chocolat (Brown & Hallström, 2000) in terms of sentence structure, relevance and objectivity. A contrastive analysis of English and Spanish ADs of The Hours (Rudin et al., 2002) was conducted by Bourne and Jiménez Hurtado (2007), where the researchers examined both linguistic and extra-linguistic differences (e.g., film content included in AD) between the two scripts and the related ramifications for translating AD scripts from English into Spanish.

The corpus-based studies mentioned earlier were all based on monomodal corpora (which included just text). Jiménez Hurtado and Seibel (2012) took AD corpus-based research a step further and created a multimodal and multisemiotic AD corpus as part of the TRACCE project. The multimodal corpus, defined as “a computer-based collection of language and communication related material drawing on more than one sensory modality” (Allwood, 2008: 208), contains around 300 Spanish-language films along with their ADs as well as 50 audio described films in English, French and German. The films have been segmented into meaning units, which comprise the AD and the corresponding audiovisual content. The segments have been tagged at three levels: narratology, cinematography and grammar, making it possible to determine equivalence patterns between the three conceptual levels. For instance, one may find a particular type of camerawork and see how it has been reflected in the ADs, or look for a specific language pattern and establish what kind of events it describes (Jiménez Hurtado & Seibel, 2012: 413; also see Jiménez Hurtado & Soler Gallego, 2013).

Reviers (2017) built a multimodal corpus of Dutch AD that included 39 titles (both films and TV series) totalling 154,570 words of AD text and 3,070 minutes’ worth of video. The AD text was annotated and aligned with the AD recordings, images and soundtrack using a multimodal concordancer. The corpus was created to conduct a frequency analysis of the lexicogrammatical features of the language of Dutch AD and also to facilitate the multimodal analysis of the AD text.

Other vantage points from which researchers looked at linguistic and textual aspects of AD are discourse analysis and pragmatics. These approaches have been mainly applied to the study of cohesion and coherence in AD. For example, Braun (2007) looked at AD from a discourse perspective by examining the English AD of The Hours (Rudin et al., 2002) in terms of building mental models, making inferences and ensuring local and global coherence in AD. In a later article, she expanded on these notions by analysing portions of the English AD of Girl with a Pearl Earring (Paterson et al., 2003) to see whether and how cohesion and
coherence were ensured in the AD (Braun, 2011). Taylor (2014), on the other hand, used multimodal transcription (Baldry & Thibault, 2006) and phasal analysis (Malcolm, 2010) in order to identify the cohesive links in a scene from Inglourious Basterds (Bender & Tarantino, 2009). Baldry and Thibault’s multimodal transcription (2006) was also used by Reviers (2018) in her analysis of selected fragments from the Dutch AD corpus (discussed earlier) for the purpose of studying multimodal coherence in AD.

And finally, some AD scholars looked at AD from the point of view of text linguistics and text typologies. Reviers (2017), for instance, sets her Dutch AD corpus-analysis within the context of Systemic Functional Linguistics (SFL) (Halliday, 1994) and applies such SFL categories as field, tenor and mode as well as context of situation to the study of AD. Mazur (2020), on the other hand, draws upon Reiss’s (1977 [1989]: 113–114) typology of informative, expressive and operative text types and proposes a classification of audiovisual (AV) text types for AD purposes, which is the basis of a functionally-oriented text analysis aimed at determining a text’s functional priorities and the related AD strategies.

Selected results of the studies mentioned previously will be presented while discussing individual linguistic and textual aspects of AD.

3. Linguistic aspects

As already mentioned in section 2, the language of AD is characterised by a number of idiosyncratic features. In this section we discuss those features both with respect to syntax and lexicon.

3.1 Syntax

Since AD relates what is visible on the screen at any given moment, “there should be no words indicating conditions, past or future states, or any number of other valid language constructs that do not reflect the immediate reporting required for audio description” (Piety, 2004: 26). Therefore, AD typically uses the present tense to convey the events as they happen. Usually the simple present is used for one-off events and the present continuous for actions happening over a stretch of time. The -ing form of the verb is sometimes used as a modifier or participle (“Holding a cup of coffee, Anna speaks on the phone”). Occasionally, the present perfect tense may be used if what is to be emphasised is the sequence of two events. Interestingly, even flashback scenes are described in the present tense, usually after announcing “In a flashback”, as “using the past tense is thought to remind people not only of what they have missed, but also of the fact that they have missed it” (Fryer, 2016: 64).

Generally, sentence structure corresponds to the images the AD describes, so in screen AD sentences or their parts will most likely be determined by shot types and scene cuts. This means that if we have a fairly fast-paced show with the corresponding changing shots, the AD language should reflect that by means of short and simple sentences. If, on the other hand, scenes are shot using “slower”, more lingering shot types, such as panning or the aerial view, the AD may then use longer and more complex sentences for illustrative purposes to set the scene or to describe a landscape. The clauses will then be separated by commas or conjunctions, most likely coordinating ones such as “and”, “while” or “as”, if we assume simultaneity of the presented events, as in the following sentence: “as she rolls herself a cigarette, a young man spots her from the back of the space”. The positioning of “as” at sentence-initial position is deliberate, since placing it after the comma (“She rolls herself a cigarette, as a young man spots her from the back of the space”) could imply that she rolls the cigarette because the man
notices her. What is more, the use of “as” implies that the two actions happen simultaneously (Fryer, 2016: 70).

3.2 Cross-linguistic differences

When discussing syntax, one cannot disregard cross-linguistic differences. When it comes to English and Italian, Arma (2012) notes that while there are few secondary clauses in the English AD of *Chocolat*, its Italian counterpart “is characterised by many embedded secondary clauses, mostly noun phrases or temporal/causal clauses, typical of the Italian written language” (p. 45). For English and Spanish ADs of *The Hours*, on the other hand, Bourne and Jiménez Hurtado (2007) observe that the English AD abounds in present and past participle clauses, whereas the Spanish AD shows preference for “a more linear style, achieved through coordinated structures and/or the use of independent sentences” (p. 182).

Another syntax-related issue that could be looked into is word order. Some languages, such as English, have a fairly strict Subject-Verb-Object (SVO) order which has to be preserved in AD for it to sound grammatical. Other languages (such as Polish), allow for a non-SVO order, making it possible to play around with AD language structure to a much greater extent and actually reflect what is visible on the screen much better. For instance, if in a film we first have an image of an oar that is pushing away a boat from the shore and only then do we see the person on the boat, you can start the sentence with the object, rather than the subject and say: “wiosłem odbija łódź od brzegu” [“With an oar he pushes the boat away from the shore”], a sentence which would not sound natural in English, even though it could be said and, for the purposes of AD, might be encouraged.

Another difference may come from some languages being inflective (again, Polish for example) and some not (e.g., English), which will make AD much less flexible (and longer) in the latter case. For instance, information that in English would need three words to express may need just one word in Polish, such as “She has left” can be expressed with just one word in Polish using an inflected form of the verb “leave” (“Wyszła”, which is a complete sentence). The sentence includes information on both the gender of the doer and the tense. The fact that in Polish the subject need not be explicitly expressed (“implied subject” as in Italian or Spanish) makes room for other flexible AD solutions. For instance, if a character performs a series of actions one by one, which are interspersed with utterances, in Polish AD you can separate clauses reflecting each action with a comma and indicate with intonation that the sentence will continue, for instance: “Anna wchodzi do domu [utterance], zamyka drzwi [utterance], podchodzi do okna [utterance], otwiera je [utterance] i zamyka”. This sentence flows well in Polish and closely reflects the on-screen actions and while it would be grammatically correct to have such a sentence in English (“Anna enters the house [utterance], closes the door [utterance], goes to the window [utterance], opens it [utterance] and closes”) it would not sound natural, as none of the clauses would work as stand-alone sentences due to the lack of explicit subject, so pronouns or repetition of nouns would be required. Also, some of the commas would need to be replaced with a conjunction such as “and” for better coordination.

Another cross-linguistic difference relevant for AD are elliptical clauses, which are much more common in Polish than English, especially when it comes to verbless sentences used very often in Polish AD (e.g., “Na łóżku Anna” [“On the bed Anna”]), as they are dynamic and time-effective. Such clauses are unacceptable in English and to sound grammatical they would have to be supplemented with a verb and preferably have a reversed order too (“Anna lies on the bed”). This, on the other hand, has direct implications for the theme-rheme structure in AD.
In a sentence, theme includes information that is assumed to be known to the audience and precedes rheme in which new information is provided. For instance, in the previous example, in the Polish AD the theme would be the bed and the rheme (new information) who is lying on it (“Anna”). If we change the word order in the English AD, we change the information sequencing. Therefore, languages with a more flexible word order allow more possibilities for rendering filmic language in AD, since the individual words, phrases and clauses can reflect more closely what is conveyed by the different shot types, editing techniques and montage.

### 3.3 Parts of speech

Following her corpus-based study of Dutch ADs, Reviers (2017) concluded that “the language of AD turns out to be different at many levels with a statistically significant number of parts of speech being used highly idiosyncratically” (p. 263). Salway (2007: 154) found that in comparison with general language there is a preponderance of lexical words in the AD corpus, which is also confirmed by Reviers’ findings for Dutch (2017: 264). Since AD usually provides the information about when and where the action is taking place, who is doing what and potentially how, it is no wonder that nouns, verbs, adjectives and adverbs as well as adverbial clauses play an important role in the language of AD. Subsequently is an overview of how those parts of speech are used in AD.

#### 3.3.1 Nouns and pronouns

Nouns are mainly used to signify on-screen characters as well as objects that are important for plot development. In his corpus-based study, Salway (2007) found that concrete nouns are especially common in AD (when compared to general language) and they are often related to characters and their body parts (e.g., man, head, face, eyes, hand), characters’ names as well as objects and scenes (e.g., door, room, car, window, table, water, bed) (p. 154). Nouns should be specific so that they can evoke a clear image for the target audience, so it is better to say “bull terrier” instead of “dog” or “spruce” instead of “tree”. Fryer (2016: 60), however, warns against the use of highly specialised vocabulary, such as “mesa”, a type of a flat rock formation resembling a table (hence the name from Spanish), which may be too specialised for those who do not know Spanish (or are not particularly knowledgeable about geology). In such a case, however, a short explanation could be given, if the type of the rock formation is essential for the programme and time allows for it.

Another thing to consider when using nouns is whether they are homonyms or even homophones (i.e., words that sound the same but mean different things), as AD is a text written to be spoken. As noted by Fryer (2016: 62), when a character picks up a file they could manicure their nails, polish wood or take out a document. Usually the context will help the audience decide on the right meaning, however extra caution should be paid to make sure that it will certainly be the case.

It is important to consistently name characters in a film so that the target audience can properly identify them. The ITC Guidance also advises repeating proper names often so that it is clear to the audience who is doing or saying what (Independent Television Commission [ITC], 2000: 20). Nouns can be replaced by pronouns, which are shorter and therefore more economical, however caution should be exercised to ensure that there is no ambiguity as to who or what the pronoun refers to. Since AD should not only be coherent with the dialogue, but also internally coherent, it is assumed that a pronoun refers to the last person mentioned in the AD, rather than the last person speaking. So, for example, if there are two (or more)
men in a scene, it is better to repeat the proper name or identify a character otherwise (e.g., “the bearded man”), rather than refer to him as a “he”, to make sure that the viewers are not confused as to what the antecedent to the pronoun is. This approach has been confirmed by research. For instance, Reviers (2017) found that Dutch AD is characterised by a lower frequency of pronouns as compared to general language, on the one hand and “a high repetition of the same words”, on the other (p. 264).

3.3.2 Verbs

Just as with nouns, verbs used in AD should be specific in order to conjure up a concrete image in the AD users’ minds. For example, it is better to say “skips” rather than “walks” or “walks joyfully”, as the verb “to skip” is more semantically loaded than “to walk”, on the one hand and more economical in terms of time when compared with “to walk joyfully”. In fact, Fryer (2016) reports that in one of the London-based AD units the AD scripting software has been configured in such a way so as to reject the word “walk” in order to make audio describers look for more precise or imaginative verbs expressing movement. She warns however against the use of verbs that are too specific and as such may not be known to the viewers. She gives the example of the verb “suspire” (meaning to collapse with exhaustion) as used in the AD of the sitcom Friends (“Phoebe suspires onto the sofa”). The verb precisely describes the melodramatic nature of the character’s action, but its meaning may be lost on the target audience who may therefore need more time to process the description and lose a portion of the show. Moreover, the verb belongs to a higher register which is not in tune with the rather informal nature of the programme.

The use of specific verbs is confirmed by AD research. For example, Salway (2007: 154) found that the English AD corpus abounds in concrete verbs of movement and gaze (such as “strolls” instead of a more general “walks” and “glares” instead of “looks”). This, however, may be language specific, as Bourne and Jiménez Hurtado (2007) found that in Spanish the use of a verb plus adverb or adverbial phrase is more common than using a specific verb (p. 179). In Salway’s study, verbs denoting manner of movement and gaze were more common in AD than in general language, which should come as no surprise, as the first type helps capture the actions of the characters, whereas the second type reflects the character’s focus of attention and reactions. They may also be indicative of the filmic language used, as in the first case we can assume the characters to be shown in full view, while in the second a close-up is more likely to be used.

3.3.3 Adjectives

Another word class essential in AD are adjectives, which “play a vital role, since they are meant to convey carefully selected features of a visual content to people who cannot see” (Arma, 2011: 284). Arma (2011) identified the most common classes of adjectives, which included dimension, colour, as well as position/direction/order. Just as with nouns and verbs, adjectives too should be specific so that they are easy to imagine by the target audience. For example, “a young girl” could be anywhere between four and twenty-four years of age, so it is better to mention the approximate age of the girl (e.g., “a seven-year-old girl”) or even add some supporting visual information, such as “a seven-year-old girl missing her two front teeth” (Fryer, 2016: 62). What is more, subjective or evaluative descriptions should be generally avoided, so that the audience can make the judgment themselves. So, instead of saying that a sofa is hideous, it is better to describe it (e.g., “a moth-eaten sofa”) (ITC: 20).
same goes for describing characters. Saying that someone is beautiful or ugly will not help AD users imagine a given character, so it would be more effective to mention a feature that makes them appear ugly or beautiful (e.g., “full red lips”). Though in some cases, where time is of the essence and the level of attractiveness of a character is relevant for the plot, using one evaluative adjective may be more effective than trying to describe them in greater detail. Fryer (2016) recommends caution when using long strings of adjectives “as this means the listener is constantly forced to revise the image they may (or may not) be building in their mind’s eye” (p. 60).

Another point to consider when it comes to adjectives is the use of colours. Generally, most AD guidelines advise that, where relevant, colours should be mentioned in description (e.g., ITC, 2000: 21; American Council of the Blind, 2009: 31). It should be noted that most blind persons lost their sight later in life and have visual memory of colours and even congenitally blind persons have some mental conceptualisations of colours, as they are “members of the same speech communities as are sighted persons” (Piety, 2004: 29). They may have some colour associations, like green grass or blue sky, or may know colours from general language use or books, for example. Colours are also part of many phraseological phrases and as such evoke some associations (e.g., “green with envy”). They may also be linked to cultural concepts, for instance white used at christenings or weddings or black used at funerals and thus associated with the feelings of innocence or mourning, respectively. Moreover, directors use colours consciously in their films to communicate meaning. For instance, purple may signal that a character will die, whereas red stands for anger, passion or romance (Bellantoni, 2005). In an AD reception study, 50 visually impaired respondents were asked if AD should include colours. The answers were recorded on a seven-point Likert scale and the average result was 6.2, which is “yes” (only 2% of all the respondents gave a negative answer to the question). It is important to add that as many as 70% of the respondents were blind (including 38% congenitally blind) (Chmiel & Mazur, 2014: 265). The results confirm that even persons who have never seen colours find them meaningful and expect them to be included in AD.

3.3.4 Adverbs

Adverbs are usually employed to qualify an action performed by a character, so they typically accompany verbs. To save precious time, whenever possible adverbs should be avoided and replaced with precise verbs that capture the manner of doing something. For example, instead of saying “He eats the soup greedily” it is better to say: “he devours the soup” (also see section 3.3.2 on verbs). In cases where adverbs need to be used, it is important that they are concrete and help imagine how an action is actually performed. This means that vague adverbs such as “clearly”, “characteristically” or “suitably” should not be used and replaced with more descriptive adverbs such as “anxiously”, “cautiously” or “jovially” (ITC, 2000: 21). So a description such as “She smiles at him mischievously” will be more telling than “She smiles at him characteristically”.

When it comes to adverbs of time, “now” is sometimes used to signal the change of a scene (“Now outside”) or a change in the look of a character (“Now naked”), whereas “then” may be used to stress the completion of an event (“He watches for a moment then continues walking”) (Salway, 2007: 166; Fryer, 2016: 64). Salway notes, however, that the use of “then” and “now” is mostly redundant, as the sequence of events is implied by the order of utterances, while Fryer adds that the changes happening on the screen can be signalled by the use of the right stress and intonation.
3.3.5 Similes and metaphors

Similes can be a useful tool for visualising descriptions, since they refer to something persons with sight loss (PSL) may know or may have experienced and as such may find relatable. For instance, a description “He gasps for air like a fish out of water” uses a concept that PSL may be familiar with and so they can get a better idea what the facial expression actually looks like. What is more, such descriptions are more time-effective than describing the manner in which these actions are performed. However, similes should be used with caution and we should always ask ourselves whether the image built on their basis in the minds of AD recipients is the intended one. In the already mentioned AD reception study by Chmiel and Mazur (2014) respondents were asked to rate the effectiveness of selected similes in terms of their visualisation using a seven-point Likert scale. One of the examples was an oft-quoted simile “a building as tall as ten elephants placed one on top of the other”, which the respondents rated as a three, meaning that the description is “rather not” effective. In the follow-up interviews the respondents reported that instead of the building they saw elephants in their mind’s eye (Chmiel & Mazur, 2014: 271; cf. American Council of the Blind, 2009: 9).

As already noted, concrete nouns and verbs are far more frequent in AD than in regular language, so a metaphor can be especially useful for bridging the gap between the abstract and the concrete, since the purpose of a metaphor “is to construe abstract concepts in terms of concrete ones” (Feng & O’Halloran, 2012: 2069). Consider the following description: “the lightning hits the water. A web of blue zigzags illuminates the roaring ocean”. “A web of blue zigzags” is a vivid metaphor that is surely likely to help the audience visualise the rather spectacular physical phenomenon. Metaphors are usually more descriptive and time-efficient than similes, as in the previous example a simile would take up more space: “the lightning hits the water. It looks like a web of blue zigzags illuminating the roaring ocean”. What is more, as noted by Fryer (2016: 69), while “a simile flags up the description as an interpretation, suggesting room for doubt; a metaphor brooks no discussion”. Metaphors may be a powerful tool in the hands of the describer, but should not be overused so as not to overwhelm the viewer with too many figurative descriptions (see Secchi in this volume). Also, they will be more suited for longer, illustrative descriptions, rather than ones that refer to simple, everyday activities which abound in most ADs.

4. Textual aspects

As already mentioned in the Introduction, AD does not function on its own as it is a part of a complex multimodal text and must properly interact with the other semiotic modalities (images, dialogues, sounds, music) to make a coherent whole for the target audience. In this section the key issues of cohesion and coherence are discussed. This is followed by an overview of how text types and text functions may influence both syntactic and lexical choices in AD.

4.1 Cohesion and coherence

When it comes to textual aspects of AD two key concepts that emerge are those of cohesion and coherence. Braun (2011), for example, investigates how film coherence can be recreated in AD. She extends the notion of textual coherence, understood as “the general impression of a continuity of sense in a text” (Beaugrande & Dressler, 1981: 84), to film and posits that it does not suffice to simply “translate” images into words, but also to recreate cohesive links which
in a film appear within and between the different modes of expression: the images, images and sounds and images and dialogues. Reviers (2018) posits that while the visual aspect of multimodal texts has received quite a bit of attention in AD research, the aural aspect (especially music and sounds) has been largely neglected. She then proposes that the semiotics of sound should be studied more carefully for AD purposes, as “AD text is a predominantly aural text type in which dialogues, music and sound effects jointly create meaning” (p. 22). The audio describer thus needs to recreate sound-image coherence in the original film as coherence between AD and sound as well as dialogue-image coherence as one between AD and dialogues. What is more, intramodal visual coherence is to be reproduced as intramodal coherence within the AD text (Braun, 2011: 650).

Along similar lines, Taylor (2014) points out that AD “needs to be both linguistically cohesive within itself and cohesive with the visual content it describes” (p. 42). In order to determine the cohesive links to be recreated in the AD Taylor proposes a frame-by-frame or shot-by-shot multimodal transcription of an audiovisual text (Baldry & Thibault, 2006), which includes elements such as the visual image, kinetic action and soundtrack. Such an analysis helps determine how the different elements in a film create meaning together and what is “most relevant visually when placed in context with the verbal and sound modes” (Taylor, 2014: 49). A drawback of this method is that it is very laborious and can be useful only for analysing shorter stretches of a film. The researcher thus also proposes phasal analysis (Malcolm, 2010), whereby a film or its portion is divided into phases (such as scenes), which can be continuous or discontinuous (i.e., interspersed with other scenes), in which case we deal with a macrophase divided into subphases. In the analysis the phases are described in terms of the visual image they convey, the verbal component as well as the type of transition between them (e.g., a cut or a flashback). Such an analysis helps determine the links between the different subphases of a macrophase, which can then be rendered in AD.

Drawing on approaches from discourse analysis, pragmatics and cognitive science, Braun differentiates between coherence understood as a semantic concept and coherence seen as a pragmatic construct. While the former concept is the metaphorical “glue that holds the text together” (Taylor, 2014: 41) manifested through cohesive ties at a text’s surface level (see Halliday & Ruqaiya, 1976), the latter is “an interpretive notion, which is intrinsically indeterminate because it is relative to participants ascribing their understanding to what they hear” (Bublitz & Lenk, 1999: 154). In verbal discourse the latter concept is linked to the emergence of a mental model, whereby linguistic cues in the text are combined with the recipient’s information from other sources, such as tapping into background knowledge or schemata (e.g., Brown & Yule, 1983), or making inferences to form a coherent whole of the text. Braun (2011) extrapolates these findings to multimodal texts and makes the assumption that recipients of such texts use “cues from different modes of expression to activate knowledge and/or to draw inferences, creating links within and across different modes of expression” (p. 649). This makes coherence hard to recognise or achieve and causes it to be “relational, individual and a matter of degree” (p. 652), or in other words “partial” or “disturbed” (Bublitz & Lenk, 1999: 155).

Braun (2011) uses selected scenes from the film Girl with a Pearl Earring to illustrate the potential problems with coherence in AD, which may result from lack of coherence between sound and AD, dialogue and AD and intramodal coherence in the AD text. In an earlier article, Braun (2007) also discusses local and global coherence in a film (and AD). While local coherence refers to coherence within one scene, global coherence pertains to coherence across different scenes. The latter could be construed in terms of what Dávila-Montes and Orero (2014: 98) describe as “visual links between consecutive scenes or as anaphoric/cataphoric references
in the relation of scenes that are chronologically distant within movies”. Such visual links should be properly recreated in AD without giving away too much too soon.

### 4.2 Text types and genres

Another area to consider when discussing textual aspects of AD are text types and genres. According to the ITC Guidance, “One of the most important rules of feature film description is to be true to the film in mood and style” (ITC, 2000: 24). We could extend this rule to cover other audiovisual genres as well. Description should thus be stylistically in tune with a programme and also appropriate for its target audience. For instance, in a wildlife documentary we can expect longer, more illustrative sentences and more specialised vocabulary to name the different wildlife species. On the other hand, in a show for children, we will likely have simplified syntax and vocabulary, suited to the age range of the target audience.

Mazur (2020) has developed a functional model of AD, which involves identifying functional priorities of the to-be-described programme, which then guide the audio describer’s choice of appropriate AD strategies (both at the macro- and micro level). This is facilitated by contextual, macro- and microanalyses of the source text (ST) and is based on a proposed classification of audiovisual text types, such as: informative, narrative, expressive, persuasive and entertaining. For example, in a cooking show the predominant function will be the informative one, as the target audience should be able to recreate the presented recipes in their homes, but the show may also have some secondary functions, such as the expressive (e.g., interesting camerawork) or the narrative (telling a story). Therefore, while the macro strategy will be to relay any relevant recipe-related information in the AD (i.e., content preservation), some of the micro strategies may be oriented towards conveying the film language (e.g., using specific language structures or filmic terms) and retelling the story (e.g., using action verbs). This will ensure that all of the three identified text functions will be made available to AD users as well.

### 5. The textual and linguistic audio description matrix (TLADM)

The issues discussed previously have been put together in a matrix in order to systematise the study of linguistic and textual aspects of AD (Figure 6.1). At the top of the pyramid there are...
text types and genres which will impact both the lexical as well as syntactical and grammatical AD choices. All of the corners of the pyramid interact with each other to create cohesion of AD text both intra- and intermodally as well as its coherence for the target audience.

Such a matrix can be a useful tool to analyse the linguistic and/or textual aspects of AD in a variety of ways. For instance, it can be used to analyse just one AD or a number of ADs in a contrastive manner, such as by comparing and contrasting different ADs prepared for the same film (to determine possible AD styles), text type or genre (to see if they will exhibit any characteristic features in comparison with general AD language). The matrix can also be used for cross-linguistic analyses between two or more languages, for instance, by comparing ADs drafted for the same film(s) in different languages to discover any syntactic and/or lexical features specific to individual languages. The matrix can be applied in its entirety or partially (e.g., the focus can just be on the lexical choices or even on one part of speech, such as verbs or adverbs). It may also be a helpful “orientation tool” in analysing larger-scale electronic AD corpora.

The following elements may fall under each of the four categories. Please note that the lists are not exhaustive and could be expanded depending on the purposes of a given study. Also, due to space limitations they are not elaborated on, though most of them have been discussed in greater detail already.

1) text types and genres: what text type does the described film or programme belong to? What is the genre? Is there a correlation between text type/genre and the type of language used (e.g., short sentences replete with action verbs in a fast-paced action film)?

2) syntax and grammar: what is the average sentence length? Are there predominantly simple or compound sentences? If the study is a cross-linguistic one, are there differences in word order (e.g., SVO vs. non-SVO)? What about the average sentence length and elliptical/verbless sentences? If you compare AD(s) in a language which has articles (such as English) with AD(s) in one that does not (such as Polish), how is the information conveyed through articles (both definite and indefinite) in the English AD expressed in the Polish AD?

3) lexical choices: what does the analysis reveal about the different speech parts? Is there any specialised vocabulary? Are lexical choices in sync with the genre (e.g., for a period drama, does the AD use period-specific vocabulary, for example, for the costumes or weaponry)? In a cross-linguistic study is there a greater variety of verbs or nouns in one language version than the other? Does one language version use more adjectives or adverbs? These could be easily determined in an electronic corpus by performing part-of-speech tagging and counting word types;

4) cohesion and coherence: whether and how cohesion is achieved intramodally within the AD text itself (e.g., do all the pronouns in an AD have clear antecedents), between AD and dialogue (is it always obvious who is speaking) as well as intermodally between AD and sounds (e.g., are all relevant sounds clearly identified)? Can we assume that the AD is coherent (in the sense of being comprehensible, see Braun, 2011) for the target audience (given the text type and the presumed knowledge the target audience has)? In a cross-linguistic it could be examined through what means coherence has been achieved in AD drafted in different languages. For example, one could look into cultural references and see whether and how they have been described differently in different language versions, given the differences in cultural backgrounds of the target audiences.
These are just some examples of issues that could be looked into when it comes to textual and linguistic aspects of AD. Research configurations both as regards analysed aspects as well as language combinations are practically limitless.

6. Future directions

Given that AD is a special use of language that is characterised by its own idiosyncratic features, it may come as a surprise that to date rather few systematic descriptive studies (and especially ones involving large corpora) have been conducted with a view to examining its unique nature. This could be due to the fact that as AD was gaining ground as an AVT modality, it was more imperative to focus on the principles of its drafting and delivery as well as its reception by the target audience. So the results of some AD studies, even if related to language, were largely prescriptive, rather than descriptive and often resulted in some form of guidelines, recommendations or indications of “how things are done or should be done” (see e.g., the deliverables of such large-scale projects as ADLAB and ADLAB PRO). However, as AD is slowly coming of age and more and more of it is being produced in different languages, we can expect researchers to become interested in examining the features of this special language type. One way to do that would be to create large-scale electronic corpora of AD. These could be monomodal or multimodal, as well as monolingual or bi- or even multilingual. The latter type could be especially useful in determining the similarities and differences between textual and linguistic features of AD drafted in different languages, which could then have implications for the feasibility of translating AD scripts. The uncovered differences between different AD languages could be used in the training of prospective translators-cum-describers to ensure that their translated ADs conform to the norms and style of ADs that have been originally drafted in the target language. Creation of large multimodal corpora, on the other hand, could be a significant step forward in the automation of the AD process (see Salway, 2007; Braun & Starr in this volume).

Another area of research that could gain in significance is AD language processing by blind and visually impaired audiences, especially as regards the effectiveness of different types of linguistic choices when it comes to plot comprehension or building mental models of described scenes. Such studies should however be based on some objective measures such as comprehension or recall tests, rather than preferences (see e.g., Mazur & Chmiel, 2016: 115–116). This, in turn, may be linked to the study of cohesion and coherence in AD, such as “the conditions under which coherence is most likely to emerge and the cognitive processing effort required” (Braun, 2011: 660).

Notes

1 In the chapter I use some examples from Polish ADs discussed in Chmiel and Mazur (2014). All translations are mine (IM).
2 For details see www.adlabproject.eu/.

7. Further reading

8. References


