Nina Reviers, Aline Remael, Walter Daelemans

THE LANGUAGE OF AUDIO-DESCRIPTION IN DUTCH: RESULTS OF A CORPUS STUDY

1 Introduction

The European Union directive of 2007 on accessible audiovisual media services (2007/65/CE) was a landmark for the development of Audio Description (AD) in Europe. Five years later, AD is already firmly rooted as an access service for the visually impaired in several European countries, supported by academic research, formal standards and regulations (with the UK and Spain as pioneers). However, some countries are lagging behind. The Dutch speaking parts of Europe, Flanders and the Netherlands, are a case in point. The first experiments with Dutch AD for television date back to the 90s, but it is only in recent years that the service has professionalised. Since 2009 some 13 commercial DVDs with AD have been released and several cinema screenings with open or closed AD have been organised (mainly in the Netherlands). The Flemish public broadcaster VRT has committed itself to air one described TV-series a year, starting in 2012. Flanders also has a growing offer of AD in the theatre, at concerts and other live-events, including sports, and in museums (ADLAB)[1]. In brief, AD in Dutch has gotten off to a good start, but there are still barriers holding back its further development, among which the failure of the Flemish and Dutch governments to formulate enforceable policies and the lack of standards ensuring high quality AD-practice. Moreover, there is very little research specifically aimed at AD in Dutch. Even if the expertise developed elsewhere in Europe goes a long way in meeting the needs of Dutch describers, much more research focusing on its specific culture- and language-bound features is required.

This paper reports on a project that centres on AD for film and television in Flanders and the Netherlands. The focus is on the question of how visual cues are expressed in words, a process about which little is currently known, but an answer to which would constitute most useful input for the study and the practice of AD. More specifically, the project's aim is to further explore the hypothesis that describers use a specialised

language in AD, one that is shaped by its communicative function and a range of constraints linked to the multimodal nature of the text (Piety 2004, Salway 2007, Braun 2008).

When exploring the question "How to render what one sees in words?" with regard to AD, there is no way around the guidelines that have been published in different European countries so far. A recent comparative study of such guidelines by the Royal National Institute for the Blind in the UK (RNIB), attributes a large section to the language of AD and claims that all the guidelines are "in complete agreement in what they suggest", referring to the use of simple sentences, the present tense and vivid adverbs and adjectives, to give just a few examples (RNIB 2010: 5; Vercauteren 2007). However, we should be careful not to generalise and attribute the same features to "new" AD-languages, such as Dutch, too quickly. Moreover the guidelines in use today are often based on intuitive insights and practice-based conventions rather than academic research, and, last but not least, the guidelines leave many (more complex) questions unanswered (Vercauteren 2007, Braun 2008).

A number of articles have already addressed some of the more complex questions concerning the specificity of the language of AD, in terms of vocabulary, grammar and syntax. Fix (2005) and her team analyzed information structure and cohesion in the AD of one German Film. Salway (2007) was one of the first to take a quantitative approach to AD-language, identifying the most frequent words and collocations in a corpus of English ADs. Arma (2011), followed-up on his approach with an in-depth analysis of the use of adjectives in English AD. Jimenez (2007; 2012) reports on a large audiovisual corpus-project, focusing on several lexico-grammatical features of Spanish ADs. A smaller but equally interesting project is the one by Bourne (2007), in which the English and Spanish ADs of one film are compared, eliciting different linguistic choices made by the describers. More recent publications like Igareda & Matamala (2012) and Taylor & Mauro (2012) explore the question of how to render what one sees in words as well. They compare the descriptions of two groups of describers that have different training backgrounds or are from different language cultures, in order to find possible (linguistic) differences.

However, current literature is far from painting a complete picture, and research in this domain would benefit from empirical product-oriented research in different languages (see for example Braun 2008; 2011, Igareda & Matamala 2012). More specifically,

Salway (2007) states that a good starting point for complementing the existing studies would be to include parts-of-speech analysis, which is the approach this study intends to take. Our aim is to describe some of the idiosyncratic lexico-grammatical features in a collection of Dutch AD-scripts, using automated corpus analysis methods to identify the frequencies of the main parts-of-speech and compare the results to the findings of existing studies.

Section 2 presents the methodology of the project, followed by section 3 on the results of the corpus study. Section 4 discusses the results in light of the existing literature. We conclude the paper with some critical reflections and recommendations for further research.

2   Compiling & analyzing a corpus of Dutch ADs

The current paper follows the corpus linguistics approach proposed by Biber & Reppen (1998; 2002). This approach consists of describing the features of a special language or register in terms of statistically significant differences between the features in the corpus studied and a general language sample. The analysis consists of four steps.

First, we compiled a representative corpus. At the time of writing, there were about 30 AD-scripts of Dutch-spoken films, short films and TV-series in Flanders and the Netherlands. Since AD is still in its infancy in Dutch, only some of these scripts were written by trained professionals and released on DVD or aired on TV. A great deal of the material is experimental, was shown only to a select audience, developed for one-off screenings or written by students and researchers. Of this material, 17 scripts were selected, covering five film genres and representing the whole range of describers: professionals, students, researchers and amateurs. The 17 scripts add up to more than 71,000 words. Table 1 illustrates the composition of this sample of Dutch AD-scripts.

Table 1: The sample of Dutch AD-scripts

Genre	Film	Series	Short film	Total
Documentary			1	1
Drama	4		1	5
Humour	2	1		3
Historic drama	3			3
Action	1	4		5
Total	10	5	2	17

Upon completion of this first step, we proceeded to transcribe, digitise and parse the scripts. XML was chosen as the data representation format and the scripts were automatically provided with part-of-speech information using the FROG system.[2]

The third and the fourth step in the analysis consisted of counting the frequencies of the main parts-of-speech in the sample of Dutch AD-scripts and of comparing the results to a general language sample, in this case, parts of the Corpus Spoken Dutch (CGN).[3] Appropriate statistical tests were applied to count and compare the frequency scores. Firstly, the variety of the results across scripts was monitored by calculating the standard deviation for each part-of-speech. The standard deviation gives an estimate of how much each AD-script deviates from the average frequency score, in other words how dispersed the data are. The standard deviation in our collection of scripts was on the low side overall (max. 5%). This signifies that the results were generally clustered around the average and that the corpus has a high degree of consistency. Secondly, we tested whether the results from the Dutch AD-corpus were significantly different from

[2] (a)With regard to XML, we followed the TEI P5 guidelines. TEI stands for Text Encoding Initiative and their guidelines can be found on: http://www.tei-c.org/Guidelines/P5/ (last visited on 22/06/2012). (b) The Frog system is a memory-based morphosyntactic tagger and dependency parser for Dutch developed by the CLIPS research group of the University of Antwerp, Belgium (see Van de Bosch et al 2007). The current version of the software (Frog) will tokenise, tag, lemmatise, and morphologically segment word tokens in Dutch text files, and will assign a dependency graph to each sentence.

[3] Between 1998 and 2004 the Spoken Dutch Corpus (Corpus Gesproken Nederlands; CGN) was constructed. The project aimed to design a corpus that would constitute a plausible sample of contemporary standard Dutch as spoken in Flanders and the Netherlands. One third of the data were to be collected in Flanders, two thirds were to originate from the Netherlands. The entire corpus was transcribed orthographically, lemmatised and enriched with part-of-speech information (http://lands.let.ru.nl/cgn/ehome.htm).

the general language sample. We calculated the probability or p-value for each part-of-speech based on the Log-likelihood test.[4] Usually a p-value of 0.05 is taken as the critical value. The lower the p-value, the more significant the differences between the two samples are. For each part-of-speech analysed in our collection of scripts, the obtained p-value was at least less than 0.001, which means that the observed differences between the two samples are significant and did not occur by chance. Wherever relevant, more details on standard deviation and p-value per category are provided in section 3.

3   Findings

This section presents the results of the frequency counts of the main parts-of-speech in our collection of 17 AD-scripts and compares them to the frequency counts in the CGN. The analysis includes both open- and closed-class words. As illustrated by graph 1 below, there are more closed-class words than open-class words in the AD-corpus, respectively 57% and 43%. In the Corpus Spoken Dutch (CGN), we find the same relationship between open- and closed-class words, but in a different proportion: 70% closed-class words to 30% open-class words. The p-value (p<0.001) confirms that there are indeed significantly more open-class words in the AD-corpus. This can be explained by the higher frequency of nouns, adjectives and verbs. The following subsections first discuss nouns, adjectives and verbs in general, followed by the findings for specific verb forms, and round off with the results for certain closed-class words (adverbs, pronouns and conjunctions). Section 3 concludes with a preliminary overview of the most common words in the Dutch AD-corpus.

3.1   Nouns, adjectives & verbs

Graph 1 shows that within the group of open-class words, the nouns form the largest group, with 22%. Statistical tests confirm that this is significantly more than in the CGN. Verbs are the second largest group with 18% and also occur more often in the AD-corpus. It is interesting to note that in the AD-corpus there are more nouns than verbs, while in the CGN this is the other way around: verbs occur more often than nouns. We will discuss the verbs in more detail in section 3.2. The category of adjectives accounts for 3%. This too is more than in the CGN. Finally, the variety

---

[4] The Log-likelihood test is used to compare whether two models are correlated. Based on the log-likelihood ratio, one can calculate statistical significance (a p-value) (Rayson & Garside, 2000).

among scripts was higher for nouns (standard deviation = 5%) than for verbs and adjectives (standard deviation <2%), which signifies more consistency across scripts for the latter two categories.

Graph 1: Relative frequency scores of the open-class words in the Dutch AD-corpus
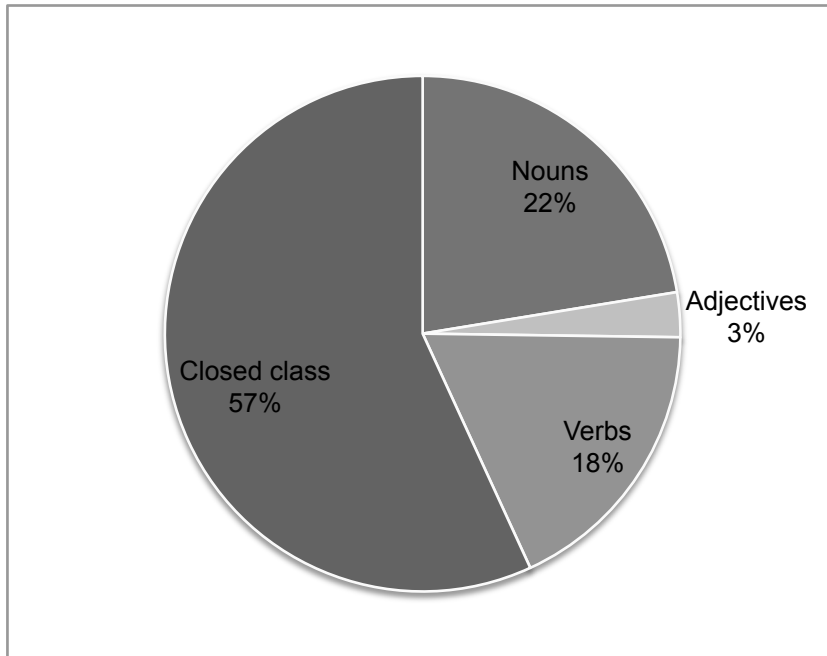
3.2 Specific verb forms

Graph 2 below, gives an overview of the different verb forms in more detail. What catches the eye immediately is that 80% of all verbs in the AD-corpus are finite verb forms, 79% of which are used in the present tense. The second largest category within the group of verbs, is the infinitives (12%), followed by the participles (8%). There are fewer present participles than past participles (respectively 2% and 6%). When compared to the general language sample, the most striking result is the preponderance of the finite verbs over the non-finite ones and the preponderance of the present tense over the past tense (and the complete lack of other tenses such as future forms). As a (logical) consequence, the other verb forms, namely the infinitives and the past participles have low frequencies in the AD-corpus and occur significantly less than in the CGN ($p<0.0001$). By contrast, the present participles stand out: they occur more in the AD-corpus than in the CGN.

Graph 2: Relative frequency scores of verb forms in the Dutch AD-corpus
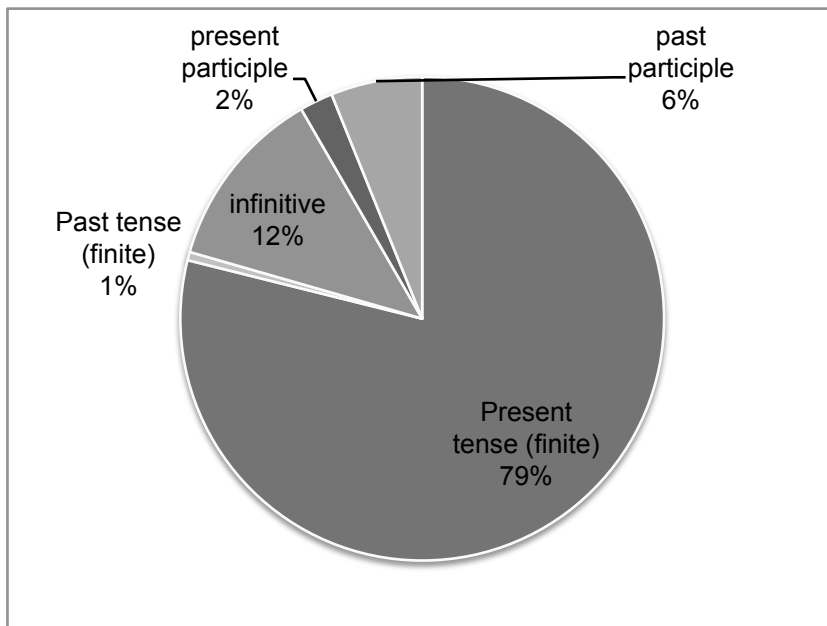
3.3 Closed class words

We have already established that there are fewer closed-class words in the AD-corpus than in the general language sample. Furthermore, the standard deviation for closed-class words is very low (1% or lower), indicating that the data are clustered around the average. What are the results for adverbs, personal, demonstrative and relative pronouns and conjunctions? Graph 3 below illustrates the frequencies of these categories. The category "other" in the graph represents the closed-class words that were not included in this study (such as articles and prepositions).
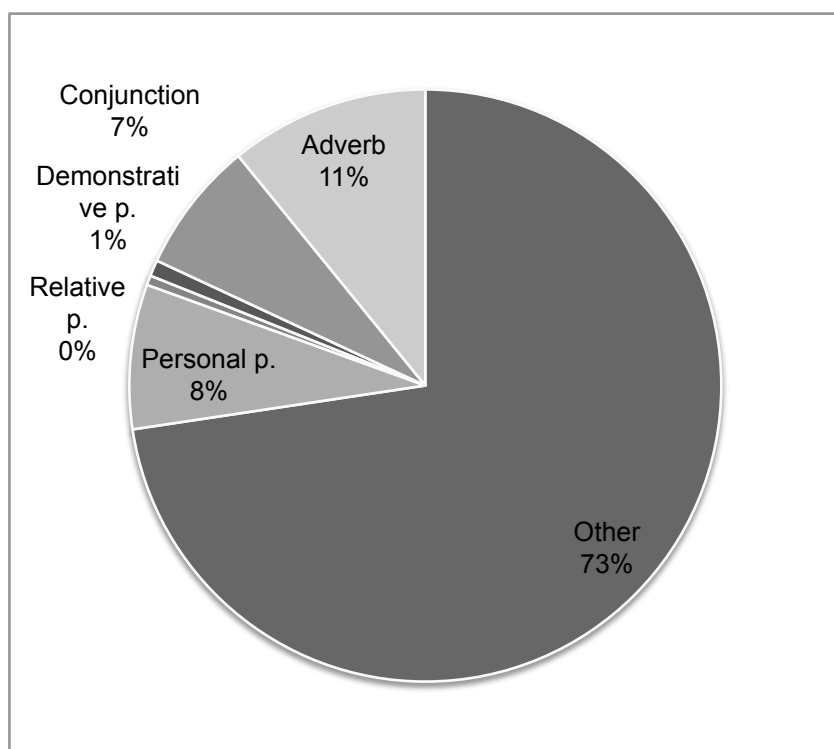
3.3.1   Adverbs

With regard to adverbs, the study shows that 11% of the closed-class words are adverbs. This means that in the AD-corpus, there are only twice as many adverbs as adjectives, while in the CGN, adverbs occur about three times as often as adjectives. Statistical tests confirm ($p<0.0001$) that adverbs indeed occur less often in the AD-corpus than in the general language sample.

3.3.2   Pronouns

16% of the closed-class words in the AD-corpus are pronouns, which is significantly less than in our general language sample ($p<0.0001$). We looked at three types of pronouns in more detail: personal pronouns, demonstrative pronouns and relative pronouns. Firstly, half of the pronouns are personal pronouns and virtually all of them (97%) are used in the 3rd person. Personal pronouns in the 1st and 2nd person barely occur in the AD-corpus. Secondly, the demonstrative and relative pronouns have low frequencies as well, respectively 1% and 0.3%.

The above frequencies hit extreme low and high scores and differ considerably from the frequency scores in the CGN. Firstly, the 3rd person pronoun occurs considerably more often in the CGN, whereas the 1st and 2nd person pronouns occur much less. The same

goes for the demonstrative pronouns. By contrast, relative pronouns are used more often in the AD-corpus than in the general language sample (p<0.001).

3.3.3 Conjunctions

This brings us to the last category: conjunctions. As graph 3 illustrates, there are 7% conjunctions in the group of closed-class words. Within this category, the coordinating conjunctions dominate (88%) over the subordinating conjunctions. Compared to the CGN, there are significantly fewer conjunctions in the AD-corpus: both the coordinating and the subordinating conjunctions occur less often (p<0.0001).

3.3.4 Common words in the Dutch AD-corpus

By way of conclusion, table 2 gives an overview of the 100 most frequent words in our collection of Dutch scripts, with their corresponding frequencies.[5]

---

[5] We filtered out proper names in the top 100.

Table 2: The 100 most frequent words in the AD-corpus

4862	de	390	er	177	tegen	131	houdt
2663	en	363	voor	171	heeft	127	laat
2622	een	358	komt	169	twee	125	even
2342	het	356	staat	167	water	124	hoofd
1751	in	335	om	162	vrouw	124	komen
1490	op	299	bij	159	wordt	124	we
1383	van	298	gaat	157	weer	123	naast
1339	zijn	289	weg	156	terug	122	gezicht
1337	hij	277	dan	156	trekt	119	hier
1220	haar	277	maar	153	ben	119	open
1194	ze	267	binnen	151	vader	118	elkaar
1062	naar	255	wat	147	deur	116	moeder
1026	met	245	als	146	pakt	113	na
872	aan	233	ziet	145	kijken	113	ogen
774	is	231	over	144	geeft	112	heb
759	kijkt	228	nog	144	ook	109	auto
618	hem	226	af	143	terwijl	109	verder
549	zich	219	zit	139	nu	109	zet
547	uit	212	hand	137	blijft	108	bed
540	te	204	neemt	136	buiten	108	mee
520	dat	191	man	136	onder	105	gaan
444	niet	186	achter	135	ligt	104	hen
411	die	180	toe	133	stapt	103	slaat
400	loopt	178	staan	132	draait	103	wil
393	door	178	zitten	131	blik	102	tot

The nouns in this list are: weg (which can mean "road", but which can also be the adverb "away"), hand (hand), man (man), water (water), vrouw (woman), vader (father), deur (door), blik (look), hoofd (head), gezicht (face), moeder (mother), ogen (eyes), auto (car) and bed (bed).

There are no less than 30 verb forms in the top 100 (23 different verbs). The 10 most frequent verbs are: zijn (to be), kijken (to look), staan (to stand), komen (to come), gaan (to go), lopen (to walk), zitten (to sit), zien (to see), nemen (to take), hebben (to have).

What is interesting to note is that there is not a single adjective in the top 100 of the most frequent words. The first adjectives appear in the top 250 with frequencies ranging from 99 to 30: andere (other), langzaam (slow), grote (big/large), witte (white), enkele (some), zwarte (black), donker (dark), vol (full), jonge (young), voorzichtig (careful), alle (all). Finally, three different conjunctions also occur in the top 100. A brief check revealed that these are virtually the only ones used in the entire corpus: en (and), maar (but), terwijl (while/as).

The type-token ratio (TTR) for our sample of AD-scripts, which reflects the diversity of words in the corpus, calculated on the basis of the above information, turned out to be on the low side, i.e. 9%, which indicates a high degree of word repetition.

4   Discussion: the "straitjacket" of AD

The objective of this paper was to describe the idiosyncratic features of the language of Dutch AD, on the basis of the frequency of parts-of-speech. The findings in section 3 shed some light on the nature of this language. Some of the results are in line with our expectations, based on the recommendations of the guidelines and the findings of previous studies (see introduction), whereas other results are more surprising and merit closer examination. Before studying some issues in more detail in the following subsections, a few general reflections on the nature of the language in our corpus are in order.

First of all, it is interesting to see that the language of Dutch AD is idiosyncratic on all the levels analysed, i.e. all the features are used significantly more or less often than in a general language sample, and this to a high degree of (statistical) significance. Secondly, the corpus displays a high consistency across scripts, with low standard deviations. This is especially noteworthy, considering the relative diversity of the material: different genres, describers from different backgrounds, describers from both Flanders and the Netherlands, etc. What is more, the describers did not follow specific guidelines and they applied different approaches. This might suggest that in terms of parts-of-speech, the nature and the constraints of the text type itself most influence the

choices made by the describers, more than convention, experience, genre or language variety.

Finally, the Dutch AD-corpus is characterised by extremes, both extreme low and extreme high scores. It contains virtually no past tenses, almost exclusively 3rd person pronouns, there is a clear dominance of finite verb forms, etc. This "restricted" language use is also reflected in the vocabulary. Even though the analysis of the common words in section 3.3.4 is only preliminary, and a more elaborate analysis is required to be able to draw reliable conclusions, it seems that the vocabulary is characterised by a high degree of repetition. Consider the following examples, which contain many of the frequent (finite) verbs identified in section 3.3.4 and repeat proper names and certain nouns excessively.

Example 1 (Iedereen Beroemd, 2000):[6]

Willy wordt wakker en het blad valt op de grond. Hij neemt op. --- Aan het kanaal wacht Jean in zijn auto. Willy komt aangereden. - Jean en Willy stappen op elkaar toe. --- Willy volgt Jean naar zijn auto. Door de ruit ziet Willy een vrouw liggen. Ze slaapt. --- Willy kijkt nog eens door de ruit. --- Jean gaat terug naar zijn auto. --- Jean gaat terug naar zijn auto. Hij gaat op de motorkap zitten.	*Willy wakes op and the page falls to the ground. He takes it. --- By the canal, Jean waits in his car. Willy comes by in his car. --- Jean and Willy walk towards each other. --- Willy follows Jean to his car. Through the window, he sees a woman lying down. She sleeps --- Willy looks through the window again. --- Jean goes back to his car. --- Jean goes back to his car. He goes and sits on the trunk.*

Example 2 (Meisjes, 2010):

Claire schrijft in het ruitjesschrift. Repetitie. Claire telt de maten op haar vingers. Claire amuseert zich. Lutgarde kijkt afkeurend.	*Clair writes in the notebook. Rehearsel. Clair counts the rhythm on her fingers. Clair has fun. Lutgarde looks disapprovingly.*

The following section looks at some of these issues in more detail. In addition, the results of the current project are linked to findings from previous studies, in order to try and catch a first glimpse of the reasons behind the specificity of the language of AD.

4.1 Preponderance of open-class words for descriptive language

In our sample of Dutch AD-scripts, there are more open-class words than in the sample of general language. Salway (2007: 155) found the same preponderance of open-class

---

[6] In the examples, the symbol "---" is used to indicate where the description is interupted by sound, music our dialogue.

words in his corpus of English ADs. He explains this feature as follows: describers describe "what they see" and therefore use concrete nouns and verbs. Furthermore, the guidelines advise scriptwriters to be descriptive and vivid, to focus on identifying characters and actions. A logical consequence of this approach seems to be the higher frequency of verbs (actions), nouns (identify characters/objects) and adjectives (be vivid and precise). Adjectives specifically have been highlighted in other studies as well, such as Igareda & Matamala (2012). Their study indicates that students trained in AD indeed use more and a wider array of adjectives than non-trained students. Arma (2011) studied the adjectives in more detail in an English corpus, identifying 1 adjective per 20 words.

Apart from the high frequency of open-class words, it is also worth re-considering the most common open-class words in the AD-corpus, described in section 3.3.4, since they seem to correspond to a large extent with the frequent words identified in Salway's (2007) corpus of English scripts. Nouns referring to characters and their body parts abound, as do nouns referring to objects, verbs referring to actions of looking and to movement. What is more, just like in our collection of scripts, there are few adjectives in Salway's top 100, which is dominated by closed-class words and high-frequency verbs. The few adjectives that do occur frequently in both the Dutch and Salway's (2007) and Arma's (2011) English corpus are: black/zwart, white/witte, dark/donker and young/jong.

4.2   Verb features

4.2.1   Present tense to describe events

The findings presented in this paper with regard to verbs also largely confirm our expectations: AD-scripts are written in the present tense. The guidelines advise describers to use the present tense and several product-oriented studies have confirmed this strategy as well (Chmiel 2012: 11-12, Taylor & Mauro 2012: 29).

However, past tenses do occur occasionally in all AD-scripts. Preliminary analysis of the tenses in context seems to indicate that the past tense is usually not used to describe actions that are happening or have just happened on screen (referring to the narrative), but are mainly used to link back to or specify characters (or actions) mentioned in previous descriptions. Consider the following examples.

Example 3 (Zwartboek, 2007):

Een zoeklicht springt aan terwijl een andere boot nadert. **SS-ers** *openen het vuur en Rob wordt in zijn borst geraakt. Rachel rent op hem af. De anderen, ook haar ouders, worden neergemaaid. --- Terwijl Rachel naar de pianist kijkt, ziet ze ineens dat het Franken is,* **de SS-officier** *die de slachting op de boot* **leidde**.	*A search light is switched on, while another boat is approaching.* **SS-officers** *open fire and Rob is hit in the chest. Rachel runs towards him. The others, including her parents, are mowed down. --- While Rachel watches the pianist, she suddenly notices that it is Franken,* **the SS-officer** *who* **led** *the slaughter on the boat.*

Example 4 (De storm, 2010):

Buiten loopt **een jonge man** *met muts op en een volle plunjezak over zijn schouder, naar de ingang van de feestzaal. Vlak voor de openstaande deur aarzelt hij, draait zich om en kijkt door het raam naar binnen. Hij ziet Julia dansen en plezier maken met haar vriendinnen. --- De kalende man en Koos,* **de jonge man** *die door het raam naar Julia* **stond** *te kijken, zijn in het donker druk bezig.*	*Outside*, **a young man** *with a hat and a full backpack over his shoulder, walks towards the entrance of the hall. In front of the open door he hesitates, he turns around and looks through the window. He sees Julia, dancing and having fun with her friends. --- De balding man and Koos,* **the young man** *who* **was** *looking through the window at Julia, are busy in the dark.*

4.2.2 Subordinating features

Our parts-of-speech approach to AD also revealed some interesting findings in terms of AD-syntax and subordination more specifically. We found that in 80% of the cases, Dutch describers prefer finite verb forms, indicating a preference for simple sentences. This in turn means lower frequency scores for the non-finite verb forms such as infinitives and past participles that occur in reduced subordinate clauses. All of this would appear to be in line with the guidelines' advice to use simple sentences.

Our results therefore confirm that in AD simple sentences are preferred, and subordinating sentences are infrequent. However, this seems to contradict one of our other findings, namely that relative pronouns (used to construct subordinating clauses as well) occur more often in ADs than in general language. The explanation for this apparent contradiction may reside in the fact that other means to make subordinate clauses (such as *te*-infinitives, past participles and conjunctions as we will see later in section 4.4.) occur much less often in the Dutch AD-corpus. It seems that on the rare occasion that a subordinate clause is used in AD, the use of a relative pronoun is

preferred over other subordinate structures. However, more in-depth syntactical analysis is needed to confirm this hypothesis.

Another surprising finding is that the frequency scores of the present participles in our collection of scripts exceed the scores in the CGN. Previous AD-studies do not comment on or explain verb features in much detail. However, a brief analysis of some of the present participles in context, reveals that they have two main functions. They are either used as adverbs specifying the action or to create reduced subordinate clauses indicating simultaneity (often in sentence-initial position). Simultaneity is indeed a challenge describers are confronted with: many actions happen simultaneously on screen in a film or TV-series and have to be rendered in linear sentences (Braun 2011). Consider the following examples. Sentence 6 is an example of a reduced subordinate clause, 5 and 7 of present participles used adverbially.

Example 5 (Koksijde Rescue: Windkracht 10, 2006).

| *Rick probeert Koen **al zwemmend** te bereiken.* | *Rick tries to reach Koen **swimming**;* |

Example 6 (Tirza, 2011):

| *Driftig **springend** trekt hij zijn broek uit.* | ***Jumping** wildly, he undoes his trousers.* |

Example 7 (Tirza, 2011):

| *Aan zijn tafeltje kijkt Jörgen **glimlachend** toe.* | *At his table, Jörgen looks on **smiling**.* |

4.3   Beware of personal & demonstrative pronouns

The reason why personal pronouns occur less often in the Dutch AD-corpus, as we have seen in section 3, can be found in the guidelines: (1) they discourage using personal pronouns to address the audience in order not to break the flow of the narrative and (2) they advise against overusing personal pronouns to refer to characters in order to avoid problems of textual cohesion. Describers therefore tend to repeat characters' names regularly (see for instance example 1 and 2 at the beginning of this section).

Furthermore, personal pronouns are almost exclusively used in the 3rd person. However, in some cases 1st and 2nd person pronouns do occur in the Dutch AD-corpus. The explanation is threefold. Firstly, some describers simply do not follow the guidelines and address the audience. Secondly, some (older) descriptions include the translation of foreign language dialogues in the AD (because audio-subtitling (AST) was not

available, yet). Finally, some scripts employ personal pronouns as a strategy for describing gestures, as in the following example.

Example 8 (Aanrijding in Moskou, 2008):

| *Johnny doet teken aan Werner, ga maar, Werner gebaart, nee jij eerst.* | *Johnny gestures to Werner, go ahead, Werner gestures, no, you go first.* |

With regard to the demonstrative pronouns, they are also infrequent in Dutch AD. Even though the guidelines and literature do not comment on them specifically, the reason for their limited occurrence may be the same as the one mentioned above in the discussion of personal pronouns: avoiding fuzzy cohesion. Indeed, a cursory analysis of these parts-of-speech in context reveals that in Dutch AD, demonstrative pronouns are used mostly to refer back to the previous sentence, whether that is a dialogue line or a description. Only rarely are they used to refer back to earlier descriptions, separated from the current one by dialogue and/or sound.

Example 9 (Karakter, 1997):

| *Jacoba laat de mannen binnen, en Dreverhaven stapt onderzoekend de woonkamer in. Hij loopt meteen door naar de aanpalende kamer waar Katadreuffes encyclopedieën liggen. Daar neemt hij een van de boeken op om ze te bekijken.* | *Jacoba lets the men in and Dreverhaven walks investigatively towards the living room. He walks straight to the adjacent **room** where Katadreuffes encyclopedia's are lying around. **There**, he takes one of the books and looks at it.* |

Example 10 (Zwartboek):

| *Rachel gaat, met alleen een truitje aan, naar de wc. Daar zit Ronnie te plassen.* | *Rachel, only wearing a sweater, goes to the **bathroom**. **There**, Ronnie is peeing.* |

4.4   Limited use of adverbs & conjunctions

According to the RNIB study (2010) adverbs are used in AD to describe emotions and actions. It might, therefore, seem surprising that adverbs have a lower frequency score in the Dutch AD-corpus than in the CGN. Several possible explanations can be found in the existing literature. Firstly, the guidelines are conflicting. On the one hand they recommend the use of adverbs for describing actions and emotions. On the other hand, describers are advised to give preference to specific verbs over verb plus adverb

combinations, because single verbs supposedly increase clarity and are shorter. For example the guidelines generally prefer "He stumbles" to "He walks clumsily". In his comparative study of a Spanish and an English AD, Bourne (2007) also finds that in the English AD, troponyms are preferred. [7] Secondly, the way in which temporal information is conveyed in AD may offer another explanation. As Salway (2007: 163) indicates, words that express temporal information (such as adverbs) are restricted in English AD, because the temporal information is usually embedded in the verb type (e.g. telic verbs), verb tense (present tense) and the order of speaking. But further (syntactical) analysis is required to confirm this. Finally, we should not forget that the discussion in section 4.2.2 demonstrated that present participles are often used adverbially as well, and present participles occur more often in AD than in general language.

This brings us to the last category: conjunctions. It seems hardly surprising that there are fewer conjunctions in the AD-corpus, since we have already established (see section 4.2.2) that simple sentences are more frequent and subordination does not occur very often. This is also confirmed in Kluckhohn's (2005) analysis of the AD of a German film. What is more, our analysis reveals that coordinating conjunctions (indicating simultaneity, see section 4.2.2 as well) are more frequent than subordinating conjunctions. Finally, the types of conjunctions used in our collection of scripts (en, maar, terwijl), correspond to the common conjunctions identified by Salway (2007) in his English corpus (and, but, while/as).

5 Closing remarks

The aim of this paper was to gain more insight into how Dutch audio-describers "put images into words". To this end, a corpus analysis was conducted, to identify the frequencies of the main parts-of-speech in a collection of Dutch AD-scripts. The results point to a highly idiosyncratic language. These results clearly support our initial hypothesis and lead us to say that there is a "language of audio description" that differs considerably from general language.

Several characteristics of the language of Dutch AD have been highlighted and we have found it to be consistent, repetitive and characterised by the restricted use of certain

---

[7] A troponym is a word that denotes a manner of doing something in more detail; "march" is a troponym of "walk".

linguistic features. What is more, initial analysis shows that there are clear similarities with AD in other languages, such as English, a finding that merits further examination. Briefly, the analyses have generated results that complement the guidelines as well as existing research and will hopefully contribute to painting a more complete picture of the language of AD and to yielding insights that can be of use to both (Dutch) AD-practitioners and students.

However, we need to express some reservations as well. The Dutch AD-corpus reflects the practice of a discipline that is still in full expansion, including work of students and initial experiments. There is certainly room for improvement in the current Dutch AD-practice, especially with regard to the "restrictiveness" of the language. In the experience of the author, beginners tend to be (over)cautious and more creative and varied approaches to AD come with experience. Therefore, research with more "mature" material than the current corpus, might produce more diverse or nuanced results. On a more general level, it has also been suggested that the current guidelines are too strict and that they should allow for more creative solutions in AD that adequately address the challenges posed by complex film language (Mälzer-Semlinger 2012). The results of this study do make one wonder whether the language of AD is indeed "forced into a straitjacket" and whether this is due to the guidelines or to the inherent constraints of the medium.

Another reservation is that the project presented here is a pilot project that would benefit from more in-depth follow-up research. Firstly, it is important to repeat the analysis in this paper with a larger corpus containing more recent scripts in order to confirm the initial results. Secondly, the analysis should be extended to other languages and more importantly to other linguistic features, to link the parts-of-speech analysis in this paper to the study of AD-vocabulary and syntax. Indeed, this study touched on some interesting findings in terms of syntax (such as subordinating features and the expression of simultaneity in AD) and vocabulary (the similarity between frequent AD words in Dutch and English) that can be the starting point of further research. Thirdly, we used the CGN as a sample of general language. And while the CGN is a reliable and well-tried corpus, it contains a mix of many different (spoken) text types. It would be interesting to compare the frequency counts of the Dutch AD-corpus to smaller and more specialised language samples (Arma (2011), for example, found that in terms of

adjectives, English AD shares many features with a corpus of fiction) in order to confirm whether the differences remain equally significant.

And last but not least, quantitative results like the ones presented in this paper must be combined with a qualitative analysis on the discourse level. In Section 4, we quoted a few examples to offer a better understanding of the context in which the linguistic features are used, but such an analysis must be conducted more systematically to be able to draw general conclusions. Only then will we gain insight into the complex reasoning that guides the choices of describers and the functioning of the language they use to express their choices.

Acknowledgments

References

Arma, S. 2011. *The language of filmic audio description: a corpus-based analysis of adjectives.* Phd of the Universita degli studi di Napoli Federico II.

Biber, D., et al. 1998. *Corpus linguistics: Investigating language structure and use.* Cambridge University press.

Braun, S., Orero P. 2008. "Audio description research: state of the art and beyond", *Translation Studies in the New Milleniu*m 6, 14-30.

Braun, S. 2011. "Creating Coherence in Audio Description", *Meta: Translator's Journal* 56/3, 645-662.

Bourne, J., Jimenez C. 2007. "From the visual to the verbal in two languages: a contrastive analysis of the audio description of The Hours in English and Spanish". In Diaz Cintas, J., Orero, P., Remael, A. (eds.) *Media for All: Subtitling for the Deaf, Audio Description and Sign Language.* (pp. 175-187). Amsterdam: Rodop

Fix, U. 2005. *Horfilm. Bildkompensation durch Sprache.* Berlin: Erich Schmidt Verlag.

Igareda, P., Matamala, A. 2012. "Variations on the Pear Tree experiment: different variables, new results?", *Perspectives: Studies in Translatology 20/1,* 103-123.

Jimenez, C (ed). 2007. *Traduccion y accesibilidad. Subtitulacion para Sordos y audiodescripcion para ciegos: nuevas modalidades de Traduccion Audiovisua*. Frankfurt: Peter Lang

Kluckhohn, K. 2005. "Informationsstrukturierung als Kompensationsstrategie: Audiodekription un Syntax". In Fix, U. (ed). *Horfilm. Bildkompensation durch Sprache*. Berlin: Erich Schmidt Verlag.

Mälzer-Semlinger, N. 2012."Narration or description: What should audio description "look" like?", in Perego, E. (ed.) *Emerging topics in translation: Audio description*. (pp. 29-36). Trieste, EUT Edizioni Università di Trieste.

Chmiel, A. 2012. "Towards common European audio description guidelines: results of the Pear Tree Project. *Perspectives: Studies in Translatology 20/1,* 5-23.

Piety, P. 2004. *The Language System of Audio Description: An Investigation as a Discursive Process*. Journal of Visual Impairment and Blindness.

Rai, S., Greening J., Petre,L. 2011. *A Comparative Study of Audio Description Guidelines Prevalent in Different Countries*. London.

Rayson, P., Garside, R. 2000. *Comparing corpora using frequency profiling*. In proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000). 1-8 October 2000, Hong Kong, pp. 1 - 6.

Reppen, R., Fitzmaurice, S. M., Biber, D. 2002. *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins.

Salway, A. 2007. "A corpus-based analysis of audio description". In Diaz Cintas, J., Orero, P., Remael, A. (eds.) *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*. (pp. 151-174). Amsterdam: Rodopi.

Taylor, C., Mauro, G. 2012. "The Pear Tree Project: a geographico-statistical and linguistic analysis", *Perspectives: Studies in Translatology 20/1,* 25-41.

Van den Bosch, A., Busser, G.J., Daelemans, W., Canisius, S. 2007. "An efficient memory-based morphosyntactic tagger and parser for Dutch". In van Eynde, F., Dirix, P., Schuurman, I., Vandeghinste, V. (eds.) *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting, Leuven, Belgium*, pp. 99-114.

Vercauteren, G. 2007. "Towards a European Guideline in Audio Description". ". In Diaz Cintas, J., Orero, P., Remael, A. (eds.) *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*. (pp.139-149). Amsterdam: Rodopi.

Filmography

Van Rompaey, C. (director). 2008. *Aanrijding in Moscou* (film).

Sombogaard, B. (director). 2010. *De Storm* (film).

Deruddere, D. (director). 2000. *Iedereen beroemd* (film).

Van Diem, M. (director). 1997. *Karakter* (film).

Enthoven, G. (director). 2010. *Meisjes* (film).

Van den Berg, R. (director). 2011. *Tirza* (film).

Herbots, H. (director). 2006. *Koksijde Rescue: Windracht 10* (film).

Verhoeven, P. (director). 2007. *Zwartboek* (film).

List of tables and graphs

Summary

In the 21st century, Audio Description (AD) has established itself as a quality access service for the blind and visually impaired in a whole range of European countries. Other countries, however, are lagging behind. The Dutch speaking parts of Europe, Flanders and the Netherlands, are a case in point. The aim of this project is to study AD in Dutch. More specifically, it explores the important underlying question of how visual cues are expressed in words. This is a process about which little is currently known (especially in Dutch), but an answer to which would benefit both the study and the practice of AD. The approach is to use automated corpus analysis techniques, to identify the frequency of the main parts-of-speech in a collection of 17 Dutch AD-scripts of films and series. The results of these frequency counts, are compared to frequency counts in a sample of general language. This comparison demonstrates that the language of AD in Dutch is idiosyncratic on several levels, and to a high degree of (statistical) significance. Some of its specific features include the preponderance of open-class words, the dominance of finite verb forms in the present tense, the inferior frequencies of conjunctions, pronouns and adverbs, to name just a few. The analysis in this paper confirms the hypothesis that there is indeed a "language of AD" that is characterised by very specific features. Moreover, the results of this study complement the findings of existing literature in painting a more complete picture of the language employed by describers. The paper concludes by explaining how the results generated by the current project can benefit future research.