This is a postprint version of the following article

Matamala, Anna (2019) "The VIW corpus: multimodal corpus linguistics for audio description analysis". *RESLA. Revista Española de Lingüística Aplicada,* 32:2, 515-542.

Link to journal: https://www.benjamins.com/catalog/resla

Link to published article: https://benjamins.com/catalog/resla.17001.mat

Doi: https://doi.org/10.1075/resla.17001.mat

The VIW project: multimodal corpus linguistics for audio description analysis

Abstract

Following an overview of corpus linguistics in audiovisual translation, and more specifically in audio description, this article presents the VIW project and its resulting corpus. It describes the compilation and annotation processes, highlighting the main challenges found. The article also presents the web application that has been developed, explaining in detail various data visualisation and search possibilities.

Keywords: audiovisual translation, accessibility, audio description, corpus

Resumen

Después de una panorámica general sobre la lingüística de corpus en traducción audiovisual, y más específicamente en el ámbito de la audiodescripción, el artículo presenta el proyecto VIW y el corpus que se ha desarrollado. Se describen los procesos de compilación y anotación del corpus, destacando los principales retos que se han encontrado. El artículo también presenta la aplicación web desarrollada durante el proyecto, que permite varias visualizaciones de los datos así como múltiples posibilidades de búsquedas.

Palabras clave: traducción audiovisual, accesibilidad, audiodescripción, corpus

The VIW project: multimodal corpus linguistics for audio description analysis

Anna Matamala

The analysis of audiovisual translations has been tackled in audiovisual translation (AVT) studies from various perspectives. However, research using a corpus linguistics approach is relatively recent, ranging from bigger multimedia corpora (Heiss and Soffritti, 2008; Jiménez Hurtado and Seibel, 2012) to smaller *ad hoc* corpora (Matamala, 2009). Despite their relevance, a recurrent problem in many of these studies has been dealing with copyright issues, which has made it difficult to share existing corpora in open access. Taking this situation into account, and inspired by Chafe's Pear Tree project (1980) and its posterior implementation in audio description (Mazur and Kruger, 2012), the VIW project was created, aiming to provide the basis of a multilingual and multimodal corpus of audio descriptions which would be freely available to the scientific community.

Audio description (AD) is an intersemiotic transfer mode in which visual content is translated into words (Maszerowska et al., 2014). These words are read aloud and integrated into the audiovisual content soundtrack so that people who cannot access the visuals can enjoy and understand the content only through the audio channel. Research on audio description has increased in recent years, generally within the context of AVT studies (Braun 2008). It has focused on a myriad of aspects, from descriptive works to reception-based or technologically-oriented research (Matamala and Orero, 2016). However, corpus-based approaches have been scarce (Salway, 2007; Jiménez Hurtado and Seibel, 2010) and open access materials are currently not available.

This paper aims to present the VIW corpus, a corpus of audio descriptions developed within a one-year project (October 2015-September 2016) funded by the Spanish Ministerio de Economía y Competitividad under "Europa Excelencia" funding scheme (reference code FFI2015-62522-ERC). Firstly, the article briefly portrays the situation of corpus studies in AVT research, centering its attention on previous work on audio description. It then presents the VIW corpus, providing a description of the short film and the audio descriptions available at the moment, as well as the annotation protocols that have been followed. Finally, the article illustrates various search possibilities offered by the web application and points to future research directions.

1. Corpus studies and audiovisual translation

Corpus linguistics has been used to study audiovisual translations with diverging approaches in terms of scope and data processing. A prototypical example is the Forlì Corpus of Screen Translation (Heiss and Soffritti, 2008; Valentini, 2008, 2013), an electronic database of original films and TV series, and their dubbed and subtitled versions in different languages (Chinese, Dutch, French, German, and Italian).

Focusing on dubbing, the Pavia Corpus of Film Dialogue was created to analyse the language of dubbing (Freddi and Pavesi, 2009). Both American and British films and their dubbed versions into Italian were compiled in a corpus that totals 117,956 words in English and 111,865 in Italian. Many investigations have been carried out with this corpus, such as research on formulaic language (Freddi, 2009) or on reference to third persons (Pavesi, 2009). Other corpora have been developed bearing specific research interests in mind: Matamala (2005) compiled an audiovisual corpus of sitcoms which included a monolingual subcorpus of sitcoms originally created in Catalan

(18,222 words) and a bilingual parallel corpus of sitcoms in English dubbed into Catalan (9,222 words in Catalan, and 9,498 in English). This corpus allowed the author to research the translation of interjections in dubbing (Matamala, 2009). In a later investigation, Matamala (2010) used what Baños et al. (2013) have termed a "draft corpus", i.e. a corpus of preliminary versions of translation, to analyse the changes translations undergo during the process of dialogue synchronisation.

A comparison of fictional dialogue in original and dubbed sitcoms, in this case in Spanish, was also carried out by Romero-Fresco (2009), who used a parallel corpus of transcripts of the American TV series *Friends* and their dubbed versions in Spanish (approx. 300,000 words), a comparable corpus of the Spanish sitcom *Siete Vidas* (approx. 300,000 words) and the spontaneous speech section of the Spanish corpus CREA, created by the Real Academia Española. Similarly, Baños (2014) also developed a corpus based on the same series (16,136 words for *Siete Vidas*, 13,592 words for *Friends*) to analyse the prefabricated orality of Spanish dubbing (Baños-Piñero and Chaume, 2009). In Italy, Bonsignori et al. (2011) used a corpus of films dubbed into Italian to analyse formulaic language, greetings and leave-takings (Bonsignori et al., 2012). Also in Italian, and with a focus on movie language, Forchini (2012) created the American Movie Corpus (AMC), a corpus of 204,636 words in both American English and Italian. Unfortunately, to the best of our knowledge, none of the previous corpora have been made available publicly.

Regarding subtitles, the Veiga corpus is a multimedia corpus of English>Galician subtitles (Sotelo Dios and Gómez Guinovart, 2012) which allows for various types of searches. It contains almost 300,000 words, 167,909 in English and 126,805 in Galician, from 24 films.

Other examples of corpora compiled for specific investigations related to various aspects of subtitling are the following: Pedersen (2011) used the Scandinavian subtitling corpus to analyse extralinguistic cultural references, whilst Mattson (2009) observed the subtitling of discourse particles in an *ad hoc* corpus in Swedish. On the other hand, Rica (2014) has been working on the Corpus of Bilingual English Spanish Subtitles, CORSUBIL, which has been used to analyse discourse markers. Still in the field of subtitling, other interesting resources are the ESIST corpus, which consists of 48 subtitled versions of three short segments developed within the Comparative Subtitling Project (www.esist.org/comparative-subtitling-project/). And the corpus compiled by Tirkkonen-Condit and Mäkisalo (2007) from the Finnish Broadcasting Company subtitle files, totalling more than 100 million words, which has been analysed by the authors to identify, for instance, cohesive devices. A final example is the multilingual corpus created by Mouka et al. (2012), including five films in English with English, Greek, and Spanish subtitles, which was used to carry out an analysis of racist discourse.

A different approach has been taken in corpora of subtitles that do not have linguistic analyses in mind but the creation of a parallel corpus that can assist in machine translation, such as SUMAT (Bywood et al. 2013) or the Open Subtitle corpus (Tiedemann 2007). They both include textual elements (subtitles) and not visual components.

Baños et al. (2013) acknowledge many of the previously mentioned corpora and consider corpus linguistics applied to audiovisual translation in greater depth in a special issue which features various studies. As they rightly point out, corpus linguistics allows researchers to "capture the distinctive features and patterns of translated texts" and "[g]eneralisations can thus be made on more solid ground not only because of the vast amount of data, but also because computer software makes it possible to detect

patterns that would be difficult to identify through manual analysis" (Baños et al., 2013, p. 483). The same authors acknowledge that in translation studies both parallel corpora and comparable corpora can prove useful, but in the field of AVT studies, multimedia corpora could also be even more important. Content which includes visual, audio, verbal and non-verbal elements can only be thoroughly researched when all elements are properly integrated.

1.1. Audio descriptions and corpus studies

Regarding audio description (AD), few corpora have been developed to date. TIWO (Television into Words) was a project led by Andrew Salway at the University of Surrey between 2002 and 2005. The project aimed "to develop a computational understanding of storytelling in multimedia contexts, with a focus on the processes of AD" (Salway, 2007, p. 153). In order to fulfil this aim, 91 audio description scripts in British English were collected from three producers of audio descriptions. The corpus was made up of 618,859 words and allowed the researcher to carry out an in-depth analysis of the language of audio description in English (Salway, 2007). Additionally, Salway suggested some ideas for assisted audio description, and how to re-use AD for keyword-based video indexing. Part of the TIWO project, namely 69 AD film scripts, was also analysed by Arma (2011), who focused on adjectives at the textual level using Antconc. It is worth mentioning that Arma states that "since the project is funded no longer and the scripts still belong to the broadcasters, even though the TIWO team had all authorizations required, a special authorization has been requested to use the scripts for research purposes only" (Arma, 2011, p. 287).

TRACCE (Jiménez Hurtado and Seibel 2011), a project at the University of Granada between 2006 and 2009 led by Catalina Jiménez Hurtado, gathered a corpus of 300 films audio described in Spanish by the association for the blind ONCE, plus 50 films in German, English, and French. A three-level multimodal annotation system was created, considering film narrative, camera language, and recurrent grammatical structures in the AD. A specific tool was also developed, but unfortunately neither the corpus nor the tool are now publicly available.

Other projects, such as the Pear Tree Project (Mazur and Kruger, 2012), worked with a remarkable number of materials, but did not incorporate them in a corpus. Indeed, most researchers in AD focus on case studies and do not use corpus processing tools. Reviers is an exception: with corpus linguistic tools she aims to demonstrate that describers use a specialised language, "one that is shaped by its communicative function and a range of constraints linked to the multimodal nature of the text" (Reviers et al., 2015, p. 168). Reviers describes some of the idiosyncratic lexico-grammatical features of the AD language in a corpus of Dutch AD scripts, implementing corpus analysis methods to calculate the frequencies of the main parts of speech. The corpus is made up of 17 AD scripts of Dutch-language films, short films and TV series in Flanders and the Netherlands, covering five film genres and including ADs by professionals, students, researchers and amateurs, and totalling more than 71,000 words. The scripts have been tagged using the FROG system, which provides part of speech information in Dutch.

Finally, outside AVT studies, a corpus of audio descriptions was created by Rohrbach et al. (2015): they consider this dataset of movie descriptions an interesting data source for computer vision research. They gathered a parallel corpus of over 68,000 sentences and video snippets from 98 movies, and used it to benchmark different approaches for semi-automatically generating audio descriptions. The MPII Movie

Description dataset (MPII-MD) provides transcribed and aligned AD and script data sentences.

1.2. From textual analysis to multimodal analysis: challenges and limitations

Despite the multimodal and multilingual nature of audiovisual translation, research has very often focused on linguistic aspects, and multimodality has taken a secondary role. Many of the previously mentioned investigations have given their attention to textual features, an undoubtedly significant aspect which nonetheless fails to account for the complexity and richness of the audiovisual text as a whole. In the field of corpus linguistics, Bednarek (2015) is one of the researchers who advocates the creation of multimodal corpora and resources. However, the development of multimodal spoken corpora is still in its infancy and the challenges of creating and exploiting multimodal corpora are enormous. As Baldry and O'Halloran (2010, p. 202) put it, we "stand on the threshold of an exciting era in which experimental research into automatic and semi-automatic corpus-based annotation and detection of multimodal genres is likely to lead to new applications and new search and retrieval techniques". The development of multimodal corpora should be subjected to similar considerations to monomodal corpus development, especially concerning sampling, representativeness and size. However, multimodal corpora present specific limitations given the time and effort involved in compiling them (Adolphs and Carter, 2013, p. 178). This is why multimodal corpora vary in their characteristics, and can be classified taking into account various aspects (Knight, 2011):

a. design and infrastructure, namely what type of data are included and how they were

collected, compiled, and annotated. Knight states that most multimodal projects use multimodal corpora tools;

b. size and scope, in other words, the amount of data and its variation. According to Knight, a small number of corpora extend beyond a few thousand words and they generally contain a few hours of video and a limited number of words. This is due to the fact that many multimodal corpora provide a detailed visual annotation and require a manual transcription of the speech content;
c. naturalness, i.e. the degree of authenticity of the data. Knight acknowledges that making up corpora with naturalistic language and authentic materials is a challenge;
d. availability and (re)usability, aspects tightly related to access rights. Knight highlights that privacy and copyright restrictions make most corpora closed projects and not publicly available.

In the field of AVT, Valentini (2013, p. 543) stresses some of the challenges posed in building corpora:

a. the need to analyse verbal information whilst considering the audio and visual components,
b. the need to define specific segmentation criteria,
c. the need to devise a methodology "that allows researchers to quantitatively measure relevant aspects of the multimedia text – linguistic, cultural, pragmatic and semiotic – and to compare the results obtained from the exploration of the verbal with the results obtained from the analysis of the non-verbal, as well as the results obtained from the association of the two"

The description of the project in the following sections will allow us to see how our corpus should be defined according to Knight's classification and how Valentini's challenges were addressed.

2. The VIW project

VIW aimed to develop a multimodal and multilingual corpus of audio descriptions departing from a single stimulus. The ultimate aim was to create a corpus of materials that would allow comparison of audio descriptions of the same visuals into one language but also across languages and cultures. The rationale behind the project was to offer the research community all materials with an open access policy, hence content with copyright was to be avoided. This challenge was overcome by commissioning a short film exclusively for the project and signing copyright agreements with all project contributors, from the film director to the audio description providers. This has allowed us to offer all materials through a Creative Commons licence CC-BY-NC-SA on the project website (pagines.uab.cat/viw) and the UAB's open access repository (ddd.uab.cat/record/147267), solving one of the recurrent problems put forward by Knight (2011) and explained in the previous section.

2.1. Corpus description: the short film

The corpus is built upon a short film, *What happens while…*, created by Catalan film director Núria Nia specifically for the project. To guarantee that the short film would include some of the main challenges in AD, a literature review and experts' discussion allowed us to highlight the key aspects in any AD, namely: characters and actions

(including gestures and facial expressions), spatio-temporal settings, film language, sound effects and silence, text on screen, and intertextual references.

Taking into account the previous analysis, the instructions given to the film director were to create a short film with a standard narrative structure, various actions taking place, and at least four characters communicating in English, except for one, who would speak in another language. The reason for this was that subtitling had to be included in the original film. The director was also requested to include at least three different spatio-temporal settings and to incorporate some text on screen, plus credits. Additionally, a specific instruction was to incorporate at least one sound that could not be easily identified by the audience, and to portray silent passages for artistic purposes. The film director was also told that the film would be audio described, meaning that some segments without speech should be included. A length of between 12 and 15 minutes was requested, as it was considered that this would allow for a variety of future experimental studies. If it was any shorter, aspects such as engagement or presence would be difficult to measure. If it was longer, experimental sessions with these materials would be too long and, therefore, more difficult to arrange. The result of these instructions, as mentioned above, is *What happens while…*, a 14-minute short film by Barcelona-based director Núria Nia.

The film director also provided a 'making of' track in the form of a director's commentary, which is also available in open access. This additional material, alongside the technical script, offers a glimpse of how the director conceived the product and what elements are considered to be more important to her.

Once the short film was finished, it was then dubbed into Catalan and Spanish in a Barcelona-based professional dubbing studio, following high quality professional

standards. The same dubbing actors were used for both the Catalan and Spanish versions.

The film deals with how different characters – James, a businessman; Rick, a retiree, and Jess, a student – envisage time. They are all shown on the phone, talking about how busy they are or about how they have too much time. All of them hear a noise and are led to the same place, where they meet. A disembodied voice greets them and asks them if they want to stop time, and this leads to a discussion among the different characters about the concept of time. In the end, they all agree that they would not like to stop time but rather enjoy the time they are given. A final character, Zoe, is shown at the end, also very busy, and the noise is heard again, before the end credits appear. The characters are physically different in terms of race, complexion, and age, and they all speak English in the original short film, except for Zoe, who speaks French and is subtitled in English. In this regard the director followed the instruction to include at least 4 characters.

Concerning spatio-temporal settings, the action takes place at different locations, at different moments in what is presumably the same day: a promenade is shown at the beginning, before the film title appears on screen. A beach is the main location of the scene where James is presented. Rick's action is set on a park, whilst Jess is shown initially in a flat. They all converge in a mountain clearing, but before that some of them are shown on a street or on a mountain walking towards their target. Zoe is shown on a rooftop at the end of the short film. Overall the film includes more than the three settings the director was instructed to include as a minimum.

It is also interesting to notice the presence of text on screen. The instruction given was vague and only requested credits and "some on-screen text" to be included. The result is an opening title, end credits, plus text on screen presenting each of the

characters and subtitles translating Zoe's words. Non-diegetic text on screen is also shown on a mobile phone screen.

One of the requirements, as explained above, was the inclusion of a sound that could not be easy to identify by the audience. The film director achieves this by including a mysterious cricket-like sound motif that leads the characters to a clearing. Although no specific instructions were given, the film also includes non-diegetic sound effects such as sound shot transitions, dramatic sounds, and diegetic sound effects such as phones ringing.

It is our belief that the film will allow for a wide variety of analyses on aspects such as the audio description of characters, of spatio-temporal settings or of text on screen, among others, thereby fulfilling the project aims and allowing comparisons in future investigations of how different audio describers convey some certain elements in film construction. However,the fact that the film is a product created specifically for the project may raise some interesting questions on the use of authentic or prefabricated materials for corpus research. The arguments which have compelled us to adopt this approach are the following:

Firstly, our interest lies in analysing the audio descriptions, and the professional ones have been commissioned and produced in real-life industry environments, not in a controlled lab situation, as explained in the next sub-section. Therefore, even if the short film has been created following some previous rules, the audio descriptions have been created following standard practices.

Secondly, it could be argued that an existing film could have been used for the same purposes. In this case an initial search on video websites proved the difficulty of finding one which was self-contained, of a certain length, copyright free and including a wide variety of audio description challenges.

Thirdly, although the film director was given some instructions, they were very general and the researcher was not involved in the script writing or in the actual recording. It could be argued that making the director aware that the film would be audio described influenced the artistic result. Moreoever, it could be said inserting audio description units in non-fabricated films may prove more challenging due to the absence of distribution of silent passages. Nevertheless, we follow the accessible filmmaking approach (Romero-Fresco 2013), and believe that this should not be viewed as a problem but as the standard rule when making films.

2.2. Corpus descriptions: the audio descriptions

Departing from the short film, either in its original or dubbed version, audio descriptions were commissioned to professionals, who were requested to follow the usual professional standards and were paid their standard fees. They were asked to deliver an .mp4 file containing the final mix, a time-coded script, and the sound files, and were given two weeks to do the job.

On the other hand, students were contacted to contribute voluntarily to the project as part of an experiment approved by the UAB's ethics committee. In this case, only the written script was requested, not the recording.

At the end of the one-year project, as of 30 September 2016, the corpus contains 47 audio descriptions, with a total of 32,417 words according to the web application countings, and is subdivided into the following sub-corpora:

a. A corpus of 10 professional audio descriptions in English (WHW-EN-Pr), including both the text and the audiovisual file: 6,799 words.

The corpus includes audio descriptions in British English such as the ones provided by BTI Studios, Deluxe, Ericsson, Mind's Eye, SDI Media, but audio descriptions by professionals from Canada (Sarah Mennell), the USA (Audio Description Associates, Bridge Multimedia), Australia (Ericsson) and New Zealand (Able) were also obtained.

b. A corpus of 10 professional audio descriptions in Catalan (WHW-CA-Pr), including both the text and the audiovisual file: 6,888 words.

The service providers were Access Friendly, Aptent, Descriptik, LB, Multisignes, Narratio, Plurals, SDI, Sonidos and Subtil.

c. A corpus of 10 professional audio descriptions in Spanish (WHW-ES-Pr), including both the text and the audiovisual file: 6,191 words.

In this case the companies providing audio descriptions in Spanish were Aptent, Aristia, CEIAF, Edsol Producciones, Ericsson, Kaleidoscope, Navarra de Cine, SDI Media, Sonidos and Trágora.

d. A corpus of 7 audio descriptions in Catalan, made by students (WHW-CA-St), including only the text file: 7,354 words.

The students, who were completing their MA in Audiovisual Translation at the Universitat Autònoma de Barcelona, were Judit Altadill, Elvira Arderius, Sara Beneito, Sara Bonjoch, Sandra Colomer, Magdalena Juan, and Laura Mor.

e. A corpus of 10 audio descriptions in Spanish, made by students (WHW-ES-St), including only the text file: 5,185 words.

The students providing the audio descriptions in Spanish were also from the same MA as those delivering them in Catalan. The list is the following: Aina Castro, Virginia de la Fuente, Isabel García Arias, José A. Jiménez, Carmen Marco, Bárbara Martín del Río, Antonio Mateo, Raquel Palacios, Marina Roldán, and Jennifer Rubio.

Table 1 summarises the data for each sub-corpus.

<INSERT TABLE 1 HERE>

AD	**Versions**	**Words**
English, professionals	10	6,799
Catalan, professionals	10	6,888
Spanish, professionals	10	6,191
Catalan, students	7	7,354
Spanish, students	10	5,185
Total:	47	32,417

Table 1. AD sub-corpora and number of words

3. Corpus processing: segmentation and annotation

Many annotation and analysis tools exist for multimodal corpora, but for this specific project ELAN (Sloetjes and Wittenburg, 2008) was chosen because it allows annotations to be linked to the video file easily and is a very robust and powerful tool to carry out searches. The resulting annotation file is an XML file conforming to the .eaf format.

The protocol for inputting data into ELAN in the VIW project was defined as follows:

1. First, data (movies with their corresponding AD transcripts) are loaded into ELAN. AD transcripts are collected in tabular files where each line contains an AD unit and

each AD unit is assigned an initial and final time code. This was especially challenging with the professional group as different formats were provided by professionals, sometimes including wrong time codes that had to be manually corrected. When later in the project AD were gathered from students, specific instructions were created to solve this initial problem: students were requested to use Subtitle Workshop and to deliver a time-coded text document, which made data processing much easier.

2. As a second step, transcripts are sent to natural language processing (NLP) tools, which produce linguistic annotations.
3. Next, these linguistic annotations are loaded into ELAN.
4. Once all the annotations are in ELAN, the tool is used to query and export data.
5. Finally, a web app, as will be presented later, allows the data to be browsed and provides different visualizations.

Two types of annotation levels (or tiers, in ELAN's terminology) were created: linguistic and filmic tiers.

3.1. Linguistic tiers

For practical reasons, linguistic tiers were split into two top tiers: the AD-unit tier and the Credits tier. Although the AD of credits is part of the AD, a preliminary analysis showed that their presence in such a short film impacted enormously the final results, hence it was decided to exclude the credits from the analysis and group them in a specific category which will be processed at a later stage.

As for the AD-unit tier, AD units were considered to be units separated by pauses longer than one second. The only exception to this rule was when the AD unit was cut to accommodate part of the film dialogue and resumed after this speech: in this case, the AD was not split. This type of segmentation coincided with the segmentation provided by most AD providers.

Each AD-unit was also further split into smaller parts, namely sentences, chunks, and tokens. For sentences and chunks, ELAN's tokeniser was used. For tokens, NLP tools were used both for segmenting and annotating, as the tokenisation produced by ELAN and by NLP tools did not match. The token-level annotation took into account parts of speech, lemmas, and semantic values. The Stanford (for English) and the Freeling (for Catalan and Spanish) parsers were used to annotate the tokens linguistically, using the Pympi library to import and merge the annotations into the .eaf files. Additionally, the RDF version of the Multilingual Central Repository was used to semantically annotate verbs, adjectives, and some nouns and adverbs. Semantic tagging was not thoroughly developed in the project due to time constraints and to the shortage of resources, hence its implementation was a preliminary and pragmatic one. Semantic tags were taken from the Suggested Upper Merged Ontology (SUMO), and only a selection were annotated, even if not always with the same fine-grained criteria.

Many of them were encoded with top-level tags in the ontology such as Process for verbs or Objects for nouns. Concerning verbs, it was decided to focus on: (a) verbs used to describe the spatial aspects of the scene; (b) verbs used to communicate and what would be termed 'sensorial' verbs, and (c) verbs used to describe the characters both physically and psychologically. Regarding adjectives, our interest lay in those describing the mood of the characters and dealing with hearing and sight. For nouns, the

focus was on the following subset: Location, BodyPart, StateOfMind, Time, Object, Human, Clothing. And for verbs Time and Location.

A manual error check was also carried out on Freeling and Stanford files before merging them with the .eaf files. A manual revision was also needed in the semantic annotation process, as the process did not include semantic disambiguation. This took longer than expected as the number of mistakes made by automatic process exceeded our expectations.

Although not developed at this stage, an annotation level called AD-focus was established at AD-unit, sentence, and chunk level. This can be used in the future to include additional annotations related to the audio description: for instance, to annotate the segments where specific features such as characters or settings are described.

3.2. Filmic tiers

Filmic tiers were used to carry out the visual tagging taking into account relevant elements in film construction. After a literature review and a working session with the film director, which proved highly useful in understanding the film construction, the following tiers were created for visual tagging.

First, the Scene tier refers to the setting where the scene takes places, but it also considers black screens and final credits where no action takes place. The coding used was an "S" followed by a number and a description of the location. In the short film under analysis, action takes places in a promenade, a beach, a street, a mountain, a park, a flat, a clearing, and a rooftop.

Second, the Short tier annotates the visuals according to what the camera is showing, taking the human body as a measure. Bordwell and Thompson's (2008, p.191)

categorisation was used to define extreme long shots, long shots, medium long shots, medium shots, medium close-ups, close-ups, extreme close-ups, and detail shots, which were coded using their acronyms (for instance, ECU for extreme close-up). Additionally, when text without a human body appeared the word “Text” was used as a code, and “BlackScreen” was used to tag black screens with no action nor characters. When one shot evolves towards another, a combination of shots was used (for instance, “ELS-LS”), which can be seen as an alternative way to indicate camera movements, an aspect of film construction that was not coded explicitly at this stage of the project.

Third, the Sound tier reflects the various sounds that can be heard and which can overlap. The categories considered are: speech (i.e. verbal language spoken by the characters), paralinguistic elements (i.e. non-verbal sounds made by the characters such as coughs, sneezes, etc.), music, sound motif (i.e. a recurrent cricket-like sound that is present throughout the movie), non-diegetic sound effects such as sound shot transitions or dramatic sounds, and diegetic sound effects, such as a mobile phone ringing. In the two last instances, only the most relevant ones were tagged. Moreover, it was often the case that two sound categories overlapped, and this was indicated in the annotation.

Fourth, the Character tier annotates the character or characters shown on screen: extra/s, James, Rick, Jess, and Zoe. A “Null” tag was created for when no characters are present on screen.

Finally, the Text tier accounts for all verbal elements printed on the screen, such as title, chyron (i.e. text on-screen which is not part of the action and provides information about character names, location, time, etc.), subtitle, credits, and also mobile text, a diegetic text which is highly relevant for this short film and merited a specific tag.

All these visual tags were carried out in a unified way for all three versions (English, Spanish, Catalan) since the visuals are exactly the same. The only exception was the Sound tier, which was annotated independently for each version as the dubbing incorporated some minor changes.

To sum up, the corpus consists of a single short movie, in three different languages, which has been annotated according to filmic criteria and contains a set of different annotated versions which vary according to language and provider. These versions can be seen as making up a comparable corpus annotated against the same timeline, as represented in Figure 1.

<INSERT FIGURE 1 HERE>

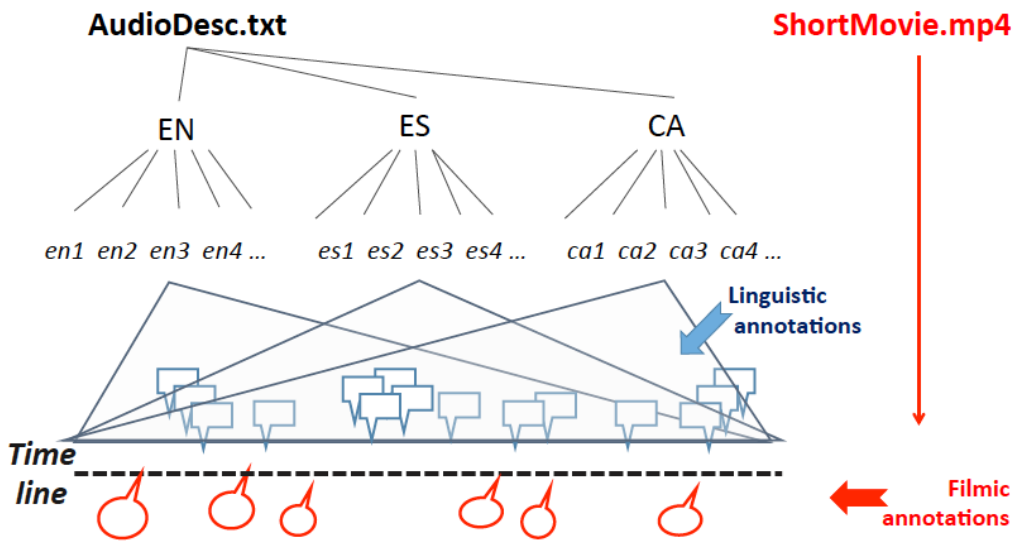

Figure 1. Corpus structure

4. Corpus exploitation: web app and visualisations

The fact that the corpus is annotated at different levels allows for a wide range of analyses, which may run on a particular file or on a set of files, at one level of annotation or at different levels. A web application was deployed using Symfony and a hosting service offered by the Universitat Autònoma de Barcelona. All codes and data are available at GitHub. The web app was designed as a data browser and visualisation service, containing no database. All data are located in the root data directly and were exported from the ELAN tool. Different controllers are used to take the data files and display them using chart APIs, mostly Google chart tools.

The web application gives access to source data, and also provides some graphical visualisations. In other words, the corpus provides the raw materials that can be imported into ELAN and further analysed. Moreover, all linguistic annotations are supplied as CQP files so that they can be analysed using the powerful text processing tool CQPweb. Finally, the web app also provides some already pre-established analyses and data visualisations, as explained below.

4.1. Visualisations at corpus and subcorpus level

One of the functionalities of the web app allows two AD files to be selected and plotted on the timeline, as in the density graph displayed in Figure 2 which compares, as an example, the ADs of two providers (BTI and Able) in the timeline. The horizontal line indicates the minute in the short film in which a certain audio description unit is inserted, whilst the vertical line shows the duration of the audio description at a certain

point on the timeline. This allows us to compare where each company positions their audio description units and how long they last.

<INSERT FIGURE 2 HERE>

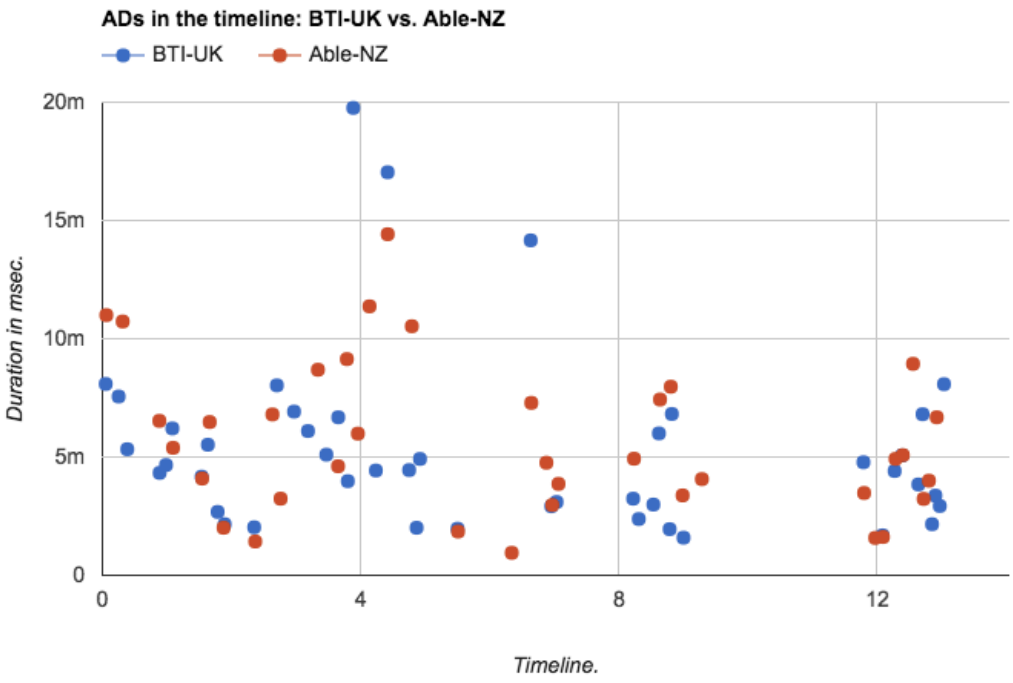

Figure 2. Comparison of the duration of two audio descriptions on the timeline

The web app also includes a browse facility (transmediacatalonia.uab.cat/web/search) that allows searches by sub-corpus and language (English, Catalan, Spanish), but also by level of expertise (professional/student), provider, and area (Australian, Canadian, New Zealand, UK, and US, for English). It also allows various downloads, search possibilities, and visualisations at subcorpora level:

a. Filmic annotations: the .eaf file (to be imported in ELAN) and the html version of the filmic annotations are available. The latter presents the five types of visual tags aligned against the same timeline. The researcher can click on a specific item (for instance, speech) and go directly to that excerpt in the video, showing the truly multimodal nature of the corpus. Figure 3 shows an excerpt of filmic annotations, which include in this case scene, shot, sound, and character on the top row and the time code related to each tag on the bottom row. For instance, in the excerpt presented in Figure 3 we can read that the cricket-like sound that becomes the sound motif is heard, then a detail shot of a cup of coffee is shown and the scene at the park begins, with a close-up of Rick making some paralinguistic sounds.

<ADD FIGURE 3 HERE>

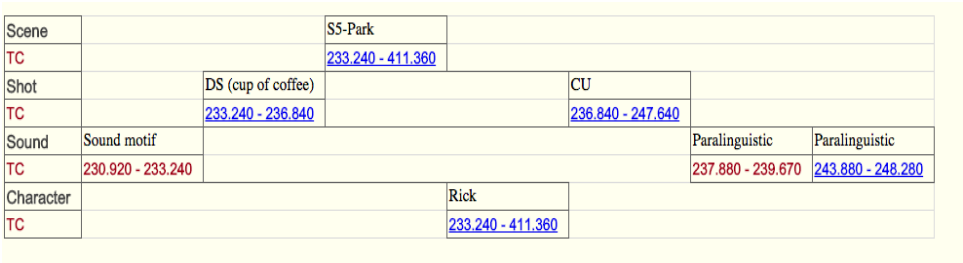

Figure 3. Excerpt of visual tagging represented in the .html file

b. Simple string search: allows a word to be searched in the whole subcorpus, and the results show this word within the context of the AD while providing a link to the corresponding video file.

c. AD units, sentence and word counts provide figures on: the number of AD units, number of sentences, number of words, mean, median and mode number of words

per AD unit, as well as minimum, maximum and range of words per AD unit. The same data are provided per sentence. Two graphs allow these to be visualised: on the one hand, the number of sentences per AD unit and the number of words per AD unit and, on the other, the number of words per sentence.

d. AD distribution on the timeline, allows visualisation, both in a compact view or an expanded view, of the various ADs in the subcorpus along the timeline, including the annotations for the five filmic tiers at the bottom of the graph. Thanks to this graph, the distribution of the AD proposed by the different providers can be compared (transmediacatalonia.uab.cat/web/hits/timeline/WHW-EN-Pr). Its expanded version allows a viewing of all the ADs in the corpus split into units along the timeline (transmediacatalonia.uab.cat/web/hits/timelinejs/WHW-EN-Pr). This is an excellent way to contrast not only the content but also the positioning of the ADs.

e. Verb distribution in the timeline allows a verbal semantic class to be selected and seen in the timeline ordered by frequency. For instance, when selecting verbs of Body Motion, a total of 158 are found, of which 54 are different. The distribution along the timeline is presented as in Figure 4. Verbs, represented by blue dots, are positioned on the vertical axis depending on their frequency in the subcorpus. For instance, the verb “sit” has 18 occurrences, so it is found at the top of the graph. The horizontal axis adds another layer of information by indicating the moment in which the verb is used along the film timeline. Additionally, when clicking on each blue dot, the verb is shown on screen. Although not shown in Figure 4, the web application includes an additional circular graph showing the percentage for each verb (for instance, 11.4% for the verb “sit”).

<INSERT FIGURE 4 HERE>

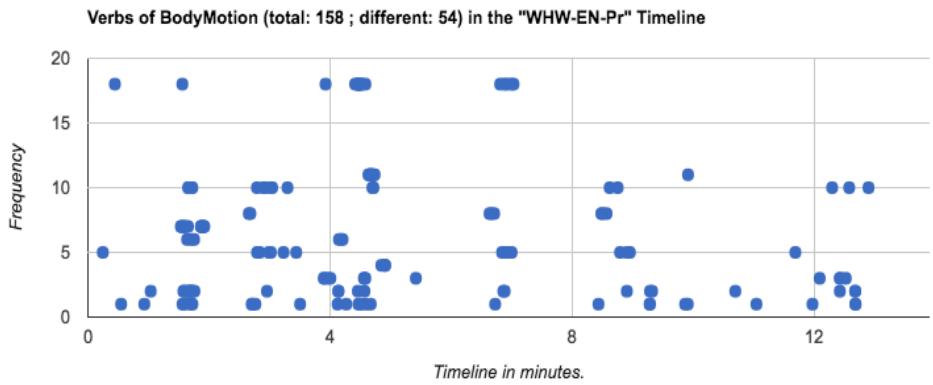

Figure 4. Distribution of verbs of Body Motion along the timeline

f. AD similarity computed using the Ted Pedersen's Text-Similarity module is also provided. It measures the similarity of two documents based on the number of shared words by the lengths of the files. A dynamic table showing the results between the various ADs in the subcorpus is also included in the web application (transmediacatalonia.uab.cat/web/similarity/WHW-EN-Pr).

g. Word frequency by part of speech is provided in various formats: as a downloadable .csv file, as a circular graph, and as a dynamic table in which information can be organised by lemma, part of speech and frequency. A filter by frequency can also be applied.

h. Word frequency by provider: a graph shows the number of verbs, nouns, adjectives and adverbs per provider in bar charts, including both the number of verbs and the number of unique verbs. The same information is provided as a downloadable .csv file.

i. Although semantic tagging is preliminary, semantic data and visualisations are provided for verbs, nouns, adjectives, and adverbs. More specifically, a dynamic table including lemma, semantic class and frequency is shown, next to a circular graph and a functionality that allows data to be filtered by frequency.

4.2. Visualisations for each audio description file

For each specific provider, the following data, searches, and visualisations are provided:

a. An html version of the .eaf file, in which the video is shown at the top and the AD is split into AD units, where different tags are shown, as in Figure 5.

<INSERT FIGURE 5 HERE>

BTI-UK/BTI-UK.eaf

AD-unit	In the distance, a couple walk their dog along an urban beach. They pass a jogger running in the opposite direction. The dog stops to sniff at something on the sand.															
Tokens	In	the	distance	,	a	couple	walk	their	dog	along	an	urban	beach	.	They	pass
PoS	IN	DT	NN	,	DT	NN	VBP	PRP$	NN	IN	DT	JJ	NN	.	PRP	VBP
Lemma	in	the	distance	,	a	couple	walk_along	their	dog	along	an	urban	beach	.	they	pass
Semantic	-	-	Location	-	-	Human	Walking	-	Animal	-	-	A-PositionalAttribute	Location	-	-	Process
TC	3.070 - 11.160															

Tokens	a	jogger	running	in	the	opposite	direction	.	The	dog	stops	to	sniff	at	something	on
PoS	DT	NN	VBG	IN	DT	JJ	NN	.	DT	NN	VBZ	TO	VB	IN	NN	IN
Lemma	a	jogger	run	in	the	opposite	direction	.	the	dog	stop	to	sniff	at	something	on
Semantic	-	Human	Walking	-	-	A-SubjectiveAssessmentAttribute	Location	-	-	Animal	Motion	-	Smelling	-	-	-

Tokens	the	sand	.
PoS	DT	NN	.
Lemma	the	sand	.
Semantic	-	Location	-

Figure 5. Sample html visualisation of an audio description unit

Figure 5 shows the first audio description unit provided by BTI Studios. The visualization includes an embedded video player where the film can be shown. The top line includes the AD unit, which is then divided into tokens, part of speech tags, lemmas, semantic tags, and the time code. In fact, when clicking on any time code on the html visualisation, the video plays exactly the audio description unit it relates to.

b. The .eaf file, which can be imported into ELAN to carry out further searches.

c. A simple string search, which shows all the contexts in which the searched word is found in the AD.

d. AD units, sentence and words counts: the same information offered for the subcorpus (see above) is now given per provider.

e. Word frequency by part of speech: again, the same information offered for the subcorpus is now given per provider.

f. AD duration in the timeline: the duration of ADs in seconds is plotted along the timeline in minutes, and numerical data are provided, either independently or merged with filmic annotations showing new scenes. For instance, in Figure 6 the red dots indicate the beginning of a new scene. When clicking on each red dot, a filmic tag is shown on screen (for example, S9-clearing). The blue dots indicate where audio descriptions units are inserted in relation to the timeline shown on the horizontal axis, and their duration is depicted on the vertical axis. Moreover, when clicking on each blue dot, the actual duration of the audio description unit that each blue dot represents appears on screen.

<INSERT FIGURE 6 HERE>

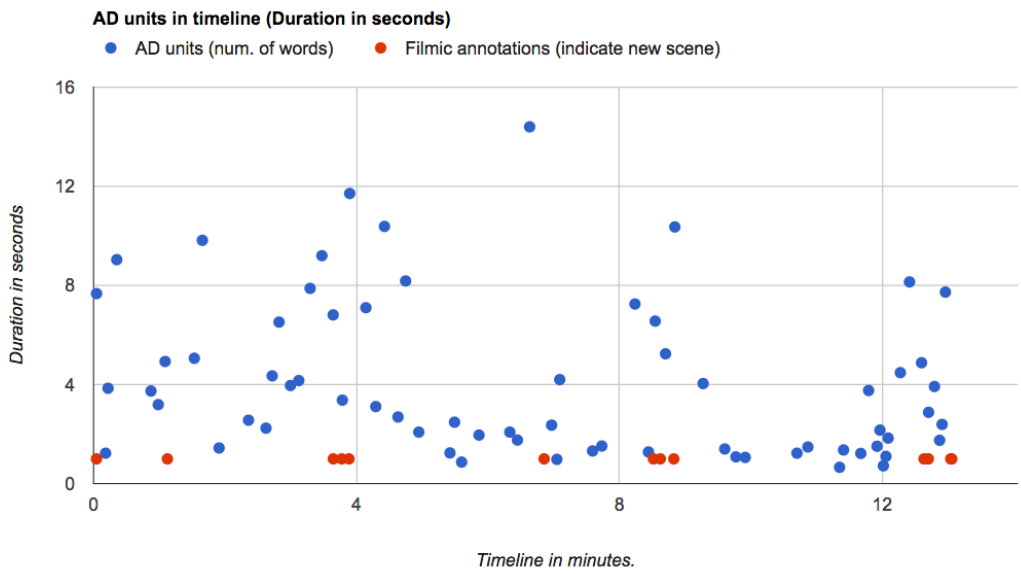

Figure 6. Duration and distribution of audio description units

g. AD length in the timeline: the number of words in the AD is plotted along the timeline in minutes, and numerical data are provided, again either independently or together with scene tags. For instance, Figure 7 shows the same filmic annotations as Figure 5, but the information provided by the blue dots relates to the number of words each AD unit contains. Similar to the other figures, this information is shown not only by the position of the dots on the vertical axis but also by clicking on the dot.

<INSERT FIGURE 7 HERE>

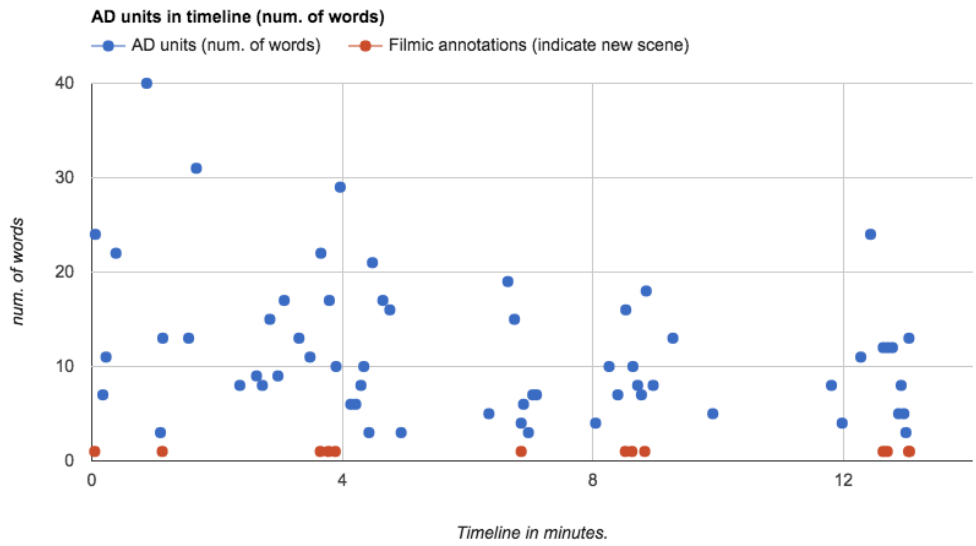

Figure 7. Number of words per AD unit

h. AD density in the timeline is shown in a visualisation that considers the characters per second of the ADs, both independently or together with scene tags.

i. Special focus has been put on verbs, providing the frequency in bar charts for the 20 most frequent verbs next to mean data in the subcorpus, and also the semantic class frequency for the 20 most frequent verbs.

j. Concerning the preliminary semantic tagging applied, the same data provided for the subcorpora are given per provider, with the same types of visualisations.

5. A first quantitative sample analysis

All the previous visualisations, as well as the .eaf files that can be imported in ELAN, allow for a myriad of analyses: individually per provider, within a subcorpus and between different subcorpora. Interesting investigations that could be carried out in this corpus include how professionals converge or differ in their audio descriptions, either intralinguistically or interlinguistically, but also the distance between professionals and students (see Matamala, forthcoming). Similarly, information from the two different annotation levels (filmic tiers and linguistic tiers) could be crossed to analyse how linguistic features related to images. These analyses will be the object of forthcoming papers and are beyond the scope of this paper. In this article the focus has been on describing the tool and contextualising it within the AVT studies tradition. However, a first global analysis comparing the three professional subcorpora will be provided focusing on numerical data related to the number of units, sentences, and words, and the parts of speech found in the professional subcorpora, in order to provide some additional data. Table 2 presents the number of AD units, sentences and words in each subcorpora, following the web application countings.

<INSERT TABLE 2 HERE>

Number of…	**WHW-EN-Pr**	**WHW-ES-Pr**	**WHW-CA-Pr**
AD units	439	495	468

Sentences	731	757	757
Words	6799	6191	6888
#words in AD unit			
mean	15.48	12.50	14.71
median	12	10	12
maximum	83	106	75

Table 2. Numerical data for the professional corpora

The data show a similar distribution of AD units in the various languages, ranging from 439 in English to 495 in Spanish. When distributed equally among providers, the difference between languages is very low, with a maximum difference of six AD units between English and Spanish. This shows a similar segmentation of the AD units based on the same visual input.

Regarding the number of sentences, the figure is exactly the same in Spanish and Catalan, but lower in English, although when distributed equally among the ten providers the difference is very small.

Finally, concerning the number of words, the Catalan subcorpus is the one with the highest number followed by the English subcorpus and the Spanish one. The AD units in English and Catalan have approximately the same number of words per AD unit (median = 12), while the Spanish usually have fewer (median = 10). However, the Spanish version is the one with the maximum number of words per AD unit (106), compared to the Catalan (75) and English (83) versions.

Table 3 shows the 20 most frequent lemmas in each professional subcorpus considering nouns, adjectives, verbs and adverbs, with their back-translation into English.

<INSERT TABLE 3 HERE>

WHW-EN-Pr	**WHW-ES-Pr**	**WHW-CA-Pr**
James (N), 86	Mirar (V), 100 (to look)	Mirar (V), 98 (to look)
Rick (N), 80	James (N), 54	James (N), 68
Be (V), 73	Rick (N), 54	Rick (N), 65
Then (R), 57	Caminar (V), 50 (to walk)	Haver (V), 48 (auxiliary verb)
Phone (N), 53	Móvil (N), 44 (cell phone)	Deixar (V), 49 (to leave)
Jess (N), 52	Playa (N), 41 (beach)	Jess (N), 46
Look (V), 48	Estar (V), 41 (to be)	Caminar (V), 42 (to walk)
Cup (N), 42	Alrededor (N), 38 (surroundings)	Got (N), 42 (glass/cup)
Coffee (N), 34	Dejar (V), 38 (to let)	Mà (N), 30 (hand)
Sound (N), 33	Hablar (V), 37 (to talk)	Mòbil (N), 39 (cell phone)
Sand (N), 32	No (R), 35 (no)	Home (N), 38 (man)
Beach (N), 31	Vaso (N), 33 (glass/cup)	Fer (V), 37 (to do)
Wear (V), 30	Jess (N), 32	Aturar (V), 35 (to stop)
Man (N), 29	Tener (V), 30	No (R), 34
Head (N), 28	Hombre (N), 29 (man)	Parlar (V), 34 (to talk)
Talk (V), 27	Llevar (V), 27 (to wear)	Buscar (V), 32 (to search)
Eye (N), 26	Banco (N), 26 (bench)	Posar (V), 30 (to put)
Walk (V), 26	Haber (V), 26 (auxiliary verb)	Sorra (N), 28 (sand)
Hair (N), 24	Mano (N), 25 (hand)	Blanc (Adj), 27 (white)
Hand (N), 24	Sonido (N), 25) (sound)	Voltant (N), 26 (surroundings)

Table 3. Twenty most frequent lemmas

In English, the most frequent lemmas include 14 nouns, 5 verbs and 1 time adverb ("then"), while in Catalan and Spanish the trend is slightly different, in line with a higher verbalisation in both languages: 11 nouns in Spanish and 9 in Catalan versus 8 verbs in Spanish and 9 in Catalan. Both Spanish and Catalan include one negation adverb ("no") among this list, but no time adverbs.

In terms of nouns, among the most frequent lemmas, all languages share the names of the main characters ("James", "Rick", "Jess"), general subject nouns ("man/home/hombre"), relevant filmic objects ("phone/móvil/mòbil", "cup/vaso/got") and body parts ("hand/mano/mà"). However, English seems to make more frequent use of words related to body parts, such as "head", "eye" and "hair". "Sand" is found in both English and Catalan ("sorra"), but not in Spanish, while "alrededor/voltant" ("surroundings") are found only in Spanish and Catalan. Nouns which are only present in one language in the list of the most frequent are "coffee" in English and "sonido" ("sound") in Spanish.

Regarding verbs, for all three languages the list includes "look/mirar/mirar", "walk/caminar/caminar", and "talk/hablar/parlar". "To be/estar" and "wear/llevar" are shared by English and Spanish, while "dejar/deixar", "haber/haver" and "poner/posar" are shared by Spanish and Catalan. As far as adjectives are concerned, only one makes it to the list: "blanc" (literally, "white") in Catalan.

When considering each part of speech separately, it is observed that the 5 most frequent verbs, nouns, adjectives and adverbs are those shown in Table 4.

<INSERT TABLE 4 HERE>

	WHW-EN-Pr	**WHW-ES-Pr**	**WHW-CA-Pr**

Adjective	White 21	Blanco 20 (white)	Blanc 27 (white)
	Black 22	Negro 19 (black)	Negre 14 (black)
	Sandy 14	Mismo 10 (same)	Dret 10 (standing)
	Grey 14	Pensativo 10 (thoughtful)	Alt 8 (tall)
	Long13	Próximo 9 (near)	Interior 8 (interior)
Noun	James 86	James 54	James 68
	Rick 80	Rick 54	Rick 65
	Phone 53	Móvil 44 (cell phone)	Jess 46
	Jess 52	Playa 41 (beach)	Got 42 (cup/glass)
	Cup 42	Alrededor 38 (surroundings)	Mà 40 (hand)
Verb	Be 73	Mirar 100 (look)	Mirar 98 (look)
	Look 48	Caminar 50 (walk)	Haver 48 (auxiliary verb)
	Wear 30	Estar 41 (be)	Deixar 49 (leave, let)
	Talk 27	Dejar 38 (leave, let)	Caminar 42 (walk)
	Walk 26	Hablar 37 (talk)	Fer 37 (to do)
Adverb	Then 57	No 35 (no)	No 29 (no)
	Now 23	Después 13 (after)	Enlaire 13 (up)
	Up 15	Claro 12 (sure)	Ara 12 (now)
	Back 14	Más 10 (more)	A banda i banda 12 (all around)
	Not 12	Ahora 9 (now)	Ja 10 (already)
	Again 12	Ya 7 (already)	Encara 9 (still)

Table 4. Frequency by part of speech

The most frequent adjectives in English are mostly related to colour (“black”, “white”, “grey”), distance (“long”) and quality (“sandy”). The Catalan and Spanish counterparts

for “black” and “white” are also among the most frequent adjectives, but other adjectives related to states of mind (“pensativo”) and used to refer to something already mentioned (“mismo”) are included in the Spanish list. In Catalan, the list includes adjectives used in our corpus for physical description (“alt”, “dret”, “interior”). Due to the limited size of the corpus, adjectives which are oral rendering of written captions such as “próximo” make it to the list.

Regarding nouns, names of characters are included in all languages, next to the objects which are shown in close-up in the film: a cup and a mobile phone, two nouns that also appear in one of the other subcorpora. The nouns “playa” (“beach”) and “alrededor” (“surroundings”) are extensively used in Spanish, whilst in Catalan the preference is for “mà” (“hand”).

As far as verbs are concerned, all three languages share “look” and “walk”, two of the main actions in the short film. “To be” is used in both English and Spanish, probably as an auxiliary, but not in Catalan, where it is not so frequent and its usage is often regarded as a calque from English. “Dejar/deixar” is used in both Catalan and Spanish, and “talk/hablar” is used in both English and Spanish.

Finally, the analysis of the most frequent adverbs shows how “not/no/no” is present in all three subcorpora. Time adverbs such as “now/ahora/ara” is also found in all three subcorpora, while “then/después” only appears among the most frequent in Spanish and English. Location adverbs such as “up/enlaire” only appear in English and Catalan, and there are further specificities linked to each subcorpora. These data provide just a preliminary overview of the many possible analyses that can help characterise the language of audio description.

6. Conclusions

This article has presented an overview of corpus linguistics and AVT, as well as an innovative multimodal multilingual corpus which is offered in open access to the research community. Reviewing the challenges expressed by Knight (2011) and Valentini (2013) discussed above, one could reach the following conclusions.

Regarding design and infrastructure, the corpus resorted to a multimodal corpora tool – ELAN, which proved a very robust and powerful tool to carry out analyses. It allowed the definition of specific segmentation and annotation criteria which allowed quantitative measurement of various aspects related to the language of AD and their relationship to the visuals. In fact, the corpus has provided tools to analyse the verbal language while considering the audio and visual elements, which have also been tagged.

In terms of size and scope, the corpus is limited, as with many multimodal corpora, but the fact that the written content does not rely on a manual transcription and that the audio descriptions could be viewed as visual tagging opens up many future possibilities in terms of project expansion and sustainability.

Concerning naturalness, the corpus commissioned professional audio descriptions to service providers who were paid their regular fees. This allowed us to overcome copyright issues linked to existing AD but can also be seen as a limitation, since expanding the corpus following the same approach will be costly.

Regarding availability and reusability, this is one of the strengths of the project, as all data are publicly available and can be reused by researchers. In this regard, the project can be seen as a prototypical example of what publicly funded research should be, since sharing data with other researchers will allow the project to be expanded and experiments to be replicated.

All in all, and despite its limitations, the project provides a wealth of data that can be analysed using the web app visualisations or importing the publicly available files into various tools to extract trends and patterns. Much effort has been put into this project, but still more effort will be required in the future to exploit all the possibilities the VIW corpus provides.

Acknowledgments

The research described in this paper has been funded by project Visuals into Words (FFI2015-62522-ERC). The corpus analysis is also part of the project "Nuevos Enfoques sobre Accessibilidad" (NEA, FFI2015-64038-P, MINECO/FEDER, UE). TransMedia Catalonia is a research group funded by the Catalan government (2017SGR113).

References

Adolphs, S. & Carter, R. (2013). *Spoken Corpus Linguistics. From Monomodal to Multimodal.* London and New York: Routledge.

Arma, V. (2011). *The language of filmic audio descriptions: a corpus-based analysis of adjectives*. PhD dissertation. Naples: Università degli Studi di Napoli Federico II.

Baldry, A. & O'Halloran, K. (2010). The annotation of multimodal corpus of university websites: an illustration of multimodal corpus linguistics. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 177-2010). Bern: Peter Lang.

Baños, R. (2014). Orality markers in Spanish native and dubbed sitcoms: Pretended spontaneity and prefabricated orality. *Meta: journal des traducteurs/Meta: Translators' Journal*, 59(2), 406-435.

Baños, R., Bruti, S. & Zanotti, S. (2013). Corpus linguistics and AVT: in search of an integrated approach. *Perspectives. Studies in Translatology*, 21(4), 483-490.

Baños-Piñero, R. & Chaume, F. (2009). Prefabricated orality: a challenge in audiovisual translation research. In M. Giorgio Marrano, G. Nadiani & C. Rundle (Eds.), *Intralinea Special Issue: The* Translation *of Dialects in Multimedia.* Available online: www.intralinea.org./specials/article/1714 (last access 20 September 2016)

Bednarek. M. (2015). Corpus-assisted multimodal discourse analysis of television and film narratives. In P. Baker & T. McEnery (Eds.), *Corpora and discourse studies* (pp. 63-87). Basingstoke/New York: Palgrave Macmillan.

Bonsignori, V., Bruti, S. & Masi, S. (2011). Formulae across languages: English greetings, leave-takings and good wishes in dubbed Italian. In A. Serban, A. Matamala & J.-M. Lavaur (Eds.), *Audiovisual translation in close-up. Practical and theoretical approaches* (pp. 23-44). Bern: Peter Lang.

Bonsignori, V., Bruti, S. & Masi, S. (2012). Exploring greetings and leave-takings in original and dubbed Language. In A. Remael, P. Orero & M. Carroll (Eds.), *Audiovisual translation and media accessibility at the crossroads – Media for All 3* (pp. 357-379). Amsterdam, New York: Rodopi.

Braun, S. (2013). Audio description research: State of the art and beyond. *Translation Studies in the New Millennium*, 6, 14-30.

Bywood, L., Volk, M., Fishel, M. & Georgakopoulou, P. (2013). Parallel subtitle corpora and their applications in machine translation and translatology. *Perspectives. Studies in Translatology*, 21(4), 595-610.

Chafe, W. (Ed.). (1980). *The Pear Stories*. Norwood: Ablex.

Forchini, P. (2010). *Movie language revisited: evidence from multi-dimensional analysis and corpora.* Bern: Peter Lang.

Forchini, P. (2010). 'Well, uh no. I mean, you know'. Discourse markers in movie conversation. In L. Bogucki & K. Kredens (Eds.), *Perspectives on audiovisual translation* (pp. 45-59). Bern: Peter Lang.

Freddi, M. & Pavesi, M. (Eds.). (2009). *Analysing audiovisual dialogue: linguistic and translational insights*. Bologna: Clueb.

Freddi, M. (2009). The phraseology of contemporary filmic speech: Formulaic language and translation. In M. Freddi & M. Pavesi (Eds.), *Analysing audiovisual dialogue: Linguistic and translational insights* (pp. 101–123). Bologna: Clueb.

Heiss, C., & Soffritti, M. (2008). Forlixt 1 – The Forlì Corpus of Screen Translation.

Exploring microstructures. In D. Chiaro, C. Heiss, & C. Bucaria (Eds.), *Between text and image. Updating research in screen translation* (pp. 51–62). Amsterdam and Philadelphia: John Benjamins.

Jiménez Hurtado, C. & Seibel, C. (2011). Multisemiotic and multimodal corpus analysis of audio descriptions. In A. Remael, P. Orero & M. Carroll (Eds.), *Audiovisual translation and media accessibility at the crossroads* (pp. 409-425). Amsterdam: Rodopi.

Jiménez, C., & Seibel, C. (2012). Multisemiotic and multimodal corpus analysis in audiodescription: TRACCE. In A. Remael, P. Orero, & M. Carroll (Eds.), *Audiovisual translation and media accessibility at the crossroads* (pp. 409–425). Amsterdam and New York: Rodopi.

Knight, D. (2011). The future of multimodal corpora. *Revista Brasileira de Linguistica Aplicada*, 11(2), 491-415.

Lison, P. & Tiedemann, J. (2016). Opensubtitles2016: extracting large parallel corpora from movie and TV subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. Portoroz, Slovenia.

Maszerowska, A., Matamala, A. & Orero, P. (Eds.). (2015). *Audio Description. New perspectives illustrated*. Amsterdam: Benjamins.

Matamala. A. (2005). *Les interjeccions en un corpus audiovisual. Descripció i representació lexicogràfica* [Interjections in an audiovisual corpus. Description and lexicographical representation]. Barcelona: Publicacions de l'IULA.

Matamala, A. (2009). Interjections in original and dubbed sitcoms: a comparison. *Meta: journal des traducteurs/Meta: Translators' Journal*, 54(3), 485-502.

Matamala, A. (2010). Translations for dubbing as dynamic texts: strategies in film synchronisation. *Babel*, 56(2), 101-118.

Matamala, A. (forthcoming) One short film, different audio descriptions. Analysing the language of audio descriptions created by students and professionals. *Onomázein*, 41.

Matamala, A. & Orero, P. (2016). *Researching audio description. New approaches*. London: Palgrave.

Mattson, J. (2009). *The subtitling of discourse particles. A corpus-based study of well, you know, I mean, and like, and their Swedish translations in ten American films*. Unpublished PhD thesis. Gothemburg, Weden: University of Gothemburg.

Mazur, I. & Kruger, J.-L. (2012) Pear Stories and audio description: language, perception and cognition across cultures. *Perspectives. Studies in Translatology*, 20(1), 1-3.

Mouka, E.; Saridakis, I. & Fotopoulou, A. (2012). Racism goes to the movies: a corpus-driven study of cross-linguistic racist discourse annotation and translation analysis. In C. Fantinuoli & F. Zanettin (Eds.), *New directions in corpus-based translation* (pp. 31-61). Berlin: Language Science Press.

Pavesi, M. (2009). Dubbing English into Italian: a closer look at the translation of spoken language. In J. Díaz-Cintas (Ed.), *New trends in audiovisual translation* (pp. 197-209). Bristol: Multilingual Matters.

Pedersen, J. (2011). *Subtitling norms for television*. Amsterdam: Benjamins.

Reviers, N., Remael, A. & Daelemans, W. (2015). The language of Audio Description in Dutch: Results of a corpus study. In A. Jankowska & A. Szarkowska (Eds.), *New Points of View on Audiovisual Translation and Accessibility* (pp. 167-189*)*. Bern: Peter Lang.

Rica, J.P. (2014). La traducción de marcadores discursives (DM) inglés-español en los subtítulos de películas: un estudio de corpus. *Journal of specialised translation*, 21, 177-199.

Rohrbach, A., Rohrbach, M., Tandon, N. & Schiele, B. (2015). A data set for movie description. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR 2015), 3202-3212

Romero-Fresco, P. (2009). Naturalness in the Spanish dubbing language: a case of not-so-close friends. *Meta: journal des traducteurs/Meta: Translators' Journal*, 54(1), 49-72.

Romero-Fresco, P. (2013). Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking. Jostrans. Journals of Specialised Translation, 20, 201-223.

Salway, A. (2007). A corpus-based analysis of audio description. In J. Díaz-Cintas, P. Orero & A. Remael (Eds.), *Media for all. Subtitling for the deaf, audio description, and sign language* (pp. 151-174). Amsterdam: Rodopi.

Sotelo Dios, P., & Gómez Guinovart, X. (2012). A multimedia parallel corpus of English-Galician film subtitling. In A. Simões, R. Queirós & D. da Cruz (Eds.), *Proceedings of the 1st symposium on languages, applications and technologies*, SLATE 2012 (pp. 255–266). Schloss-Dagstuhl: OASIcs.

Tiedemann, J. (2007). Building a multilingual parallel subtitle corpus. In P. Dirix, I. Schuurman, V. Vandeghinste, & F. Van Eynde (Eds.), *Proceedings of the 17th meeting of computational linguistics in the Netherlands* (pp. 147-162). Utrecht: Utrecht University.

Tiedemann, J. (2009) News from OPUS - A collection of multilingual parallel corpora with tools and interfaces. In N. Nicolov, K. Bontcheva, G. Angelova & R. Mitkov (Eds.), *Recent advances in natural language processing* (pp. 237-248). Amsterdam/Philadelphia: John Benjamins.

Tirkkonen-Condit, S. & Mäkisalo, J. (2007). Cohesion in subtitles: a corpus-based study. *Across languages and cultures*, 8(2), 221-230.

Valentini, C. (2008). Forlixt 1 – The Forlì Corpus of Screen Translation. Exploring macrostructures. In D. Chiaro, C. Heiss & C. Bucaria (Eds.), *Between text and image. Updating research in screen translation* (pp. 37–50). Amsterdam and Philadelphia: John Benjamins

Valentini, C. (2013). Phrasal verbs in Italian dubbed dialogues: a multimedia corpus-based study. *Perspectives. Studies in translatology*, 21(4), 543-562.