

Recibido / Received: 31/05/2020
Aceptado / Accepted: 28/09/2020

Para enlazar con este artículo / To link to this article:
<http://dx.doi.org/10.6035/MonTI.2021.13.06>

Para citar este artículo / To cite this article:

Hermosa-Ramírez, Irene. (2021) "The hierarchisation of operatic signs through the lens of audio description: A corpus study." In: CALZADA, María & Sara LAVIOSA (eds.) 2021. *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning: A critical reflection*. MonTI 13, pp. 184-219.

THE HIERARCHISATION OF OPERATIC SIGNS THROUGH THE LENS OF AUDIO DESCRIPTION: A CORPUS STUDY¹

IRENE HERMOSA-RAMÍREZ

Irene.hermosa@uab.cat
Autonomous University of Barcelona

Abstract

In opera, a multimodal art form by nature, meaning is constructed by the synthesis of its musical, verbal, visual and dramatic components. Audio description (AD) is an audiovisual translation modality that provides blind and visually impaired patrons with access to the visual elements of the play. The first aim of this pilot study is to conduct a corpus analysis to define the lexico-grammatical patterns of opera AD in the Gran Teatre del Liceu in Barcelona and the Teatro Real in Madrid. The second aim is to perform a semiotic analysis to elucidate the hierarchisation of the action, the *parole* and the visual aesthetics of the production in the AD scripts. The conclusions suggest a number of linguistic and semiotic idiosyncrasies that are shared with other AD modalities, as well as some key divergences.

Keywords: Audio description; Corpus analysis; Opera; Semiotics; Accessibility.

-
1. This study is part of the RAD project (Researching Audio Description: Translation, Delivery and New Scenarios), reference code PGC2018-096566-B-I00 (MCIU/AEI/FEDER, UE). The author is a member of the TransMedia Catalonia research group (2017SGR113). She has been awarded a PhD grant from the Catalan Government (2019FI_B 00327).



Esta obra está bajo una licencia de Creative Commons Reconocimiento 4.0 Internacional.

Resumen

En la ópera, un arte multimodal por naturaleza, el significado se cimenta en la síntesis de los elementos musicales, verbales, visuales y dramáticos. La audiodescripción (AD) es una modalidad de la traducción audiovisual que proporciona el acceso a los elementos visuales para las personas con pérdida de visión. El objetivo de este estudio piloto es, en primer lugar, realizar un análisis de corpus para definir las características léxico-gramaticales de la AD operística en el Gran Teatre del Liceu de Barcelona y en el Teatro Real de Madrid. El segundo objetivo es realizar un análisis semiótico para esclarecer la jerarquización de la acción, la *parole* y la estética visual de la puesta en escena en los guiones de AD. Las conclusiones sugieren varios patrones lingüísticos y semióticos compartidos con otras modalidades de la AD, pero también algunas discrepancias.

Palabras clave: Audiodescripción; Análisis de corpus; Ópera; Semiótica; Accesibilidad.

1. Introduction

Audio description (AD), the “verbal commentary providing visual information for those unable to perceive it themselves” (Fryer 2016: 1), is an audiovisual translation modality and accessibility service that began in the context of the performing arts. For a brief historical review, AD provision was first implemented in theatre, in 1981, at the Arena Stage Theatre in Washington DC (Pfanstiehl & Pfanstiehl 1985; ITC 2000: 4). Accessibility was later on imported to Europe, first at the Robin Hood Theatre in Averham, Nottinghamshire, and then expanded into other countries and modalities such as film and television during the 1990s (Arma 2011: 43-44). As for opera AD, its beginnings date back to the early 2000s (York 2007; Cabeza-Cáceres & Matamala 2008). Since then, opera AD and opera accessibility services as a whole – including surtitles, touch tours and audio subtitles – have witnessed steady interest both in practice and research: from early descriptive and practice-based studies (Matamala 2007; York 2007; Orero & Matamala 2007; Cabeza-Cáceres 2010; Corral & Lladó 2011; Eardley-Weaver 2010) to an increasing interest in user-centred approaches and participatory accessibility (Di Giovanni 2018a, 2018b). The rationale for this

pilot study is that opera AD has not yet been studied from a corpus-based perspective. We therefore aim to fill this gap and apply the methodology offered by Corpus Linguistics to identify lexico-grammatical patterns as well as conduct a semiotic analysis.

Though recent, the application of Corpus Linguistics studies in Audiovisual Translation (AVT) research has been fruitful (cf. Baños *et al.* 2013). Corpus research in AD, however, has been scarcer. For the purposes of this study, previous AD corpora can be classified into two categories: monomodal (text-based) corpora and multimodal corpora. Text-based corpora confirm hypotheses regarding linguistic patterns by producing frequency lists and allowing for concordance analysis (Evison 2010: 122). On the other hand, multimodal corpora involve the manual or semi-automatic tagging of the different meaning-making semiotic modes (Tuominen *et al.* 2018: 9), thereby addressing the multimedia nature of audiovisual products. In both cases, previous corpus studies on AD have widely been devoted to film and, to a lesser extent, museum AD (Perego 2019; Soler Gallego 2018).

Within the monomodal category, Salway (2007) led the TIWO project, where a corpus of 91 film AD scripts in English was compiled and analysed in order to study and define the idiosyncratic language of AD. Unusual frequently occurring phrases were found regarding shifts in the characters' focus of attention, change of scene, non-verbal communication and shift in the situation (Salway 2007: 160-161). Later on, Arma (2011) used part of the TIWO corpus to study the use of adjectives in filmic AD. In other languages, Reviers (2018) compiled a corpus of 39 Dutch films and series and conducted a lexico-grammatical analysis combined with a systemic-functional approach. All previously mentioned studies confirm the common hypothesis that AD language presents distinctive lexical, grammatical and syntactical patterns when compared to reference corpora in their matching language. Hence, this paper further questions: Is the idiosyncratic language of film AD shared by opera AD?

As for multimodal corpora, the TRACCE project compiled a corpus of over 300 film AD scripts in Spanish, which was later on complemented with 50 films in English, German and French. They developed a tagging system

comprising three levels. The narratology level included tags referring to space, time and character identification and emotion. The cinematography level referred to camera language, and the grammar level encompassed semantic domains and the syntactic and discourse structure (Jiménez Hurtado 2010: 70; Jiménez Hurtado & Seibel 2012). The films were annotated with an ad-hoc software program (Tagetti), which facilitated the arduous process of manual segmentation that often hinders multimodal corpus analysis.

Lastly, Matamala (2018, 2019) led the VIW project, an open-access, comparable corpus of student and professional AD of the same short film in Catalan, Spanish and English, where two main tagging tiers were developed and applied. On the one hand, the linguistics tiers were developed to analyse part-of-speech (PoS) frequencies and semantic tagging of open-class words. The filmic and cinematic tiers, on the other hand, “were used to carry out the visual tagging taking into account relevant elements in film construction” (Matamala 2019: 527). Within them, the scene tier referred to the location and setting, and the shot tier annotated the camera movement, i.e. medium close shots, close-ups, etc. (Matamala 2019: 527). The sound tier addressed sounds categorised by speech, paralinguistic elements, music and non-diegetic sound effects, among others. The character tier identified the characters on screen and, finally, the text tier encompassed the visual-verbal elements, i.e. on-screen titles and subtitles.

Against this background, this paper presents a pilot experience for a corpus study on opera AD scripts that aims to 1) define the lexico-grammatical features of the scripts and 2) analyse the distribution of the operatic semiotic meanings conveyed by the AD. The article is structured as follows: In Section 2, the conceptual framework examines opera translation and the hierarchisation of the operatic signs. In Section 3, the corpus is presented, and our two-fold methodology is introduced. The most salient lexico-grammatical features of the corpus are analysed following Quantitative Corpus Linguistics. Section 4 delves into the qualitative subsection of the corpus, where the aim is to question opera AD through the lens of semiotics. In

Section 5, the tentative results of this new area are discussed and future possibilities are examined.

2. A conceptual framework: translation, accessibility and the semiotics of opera

In order to establish the conceptual framework and the purpose of this article, two major topics are introduced: opera intelligibility through translation and the hierarchisation of operatic signs. As Mateo (2012: 115) puts it:

[M]usic, performance and verbal text all collaborate in the creation of meaning in an opera piece. Nevertheless, the powerful presence of music has traditionally framed the conception of opera as an essentially musical genre, rather than as a dramatic art and this can be observed in sung translation: the music is normally considered untouchable and becomes the code that functions as the axis for translation decisions; the verbal text is thus subordinated to the music so that the translator must adhere to the notes and the tempo of the original score.

In this regard, it is not surprising that opera translation has generated controversy and that widespread translation strategies have been heterogenous regarding the historical period, tradition by country and even audience profile (cf. Desblache 2007). The controversy surrounding opera translation can be traced back to two general positions. While logocentrism prioritises the word and “may be characterized by the aphorism, *prima le parole e poi la musica*, musicocentrism is expressed in its opposite, *prima la musica e poi le parole*” (Gorlée 1997: 237). In the context of this study – Spanish opera houses – surtitles are the current widespread solution to maintain the integrity of the original vocal music while fostering linguistic accessibility. Accessibility services are also increasingly on offer. For illustrative purposes, Table 1 presents an overview of the most common translation strategies for opera (cf. Orero & Matamala 2007).

	Translation strategy	Fundamental characteristics
<div>Less intrusive</div> <div>↑</div> <div>↓</div> <div>More accessible</div>	No translation	The strategy presumes that the audience is familiar with the work.
	Libretto translation	The strategy does not imply synchrony. According to Orero and Matamala (2007: 262-263) libretto translations can be: <div><div>1. Reader-oriented (verse or prose) translations.</div><div>2. Translations to be sung or “composed to fit the music, which is untouchable” (Gorlée 1997: 245).</div></div>
	Translated chant	Currently, not a widespread strategy in Europe (with some notable exceptions, i.e. adapted versions for children and the overall programme of the English National Opera). This is, however, the preferred approach for musical theatre.
	Surtitles	A condensed translation of the libretto, often projected above the proscenium or on small purpose-built screens on the auditorium seats. Virkkunen (2004) distinguishes two approaches to opera surtitles: <div><div>1. Libretto-like surtitles, with repetitions, imitating its verbal-musical structure.</div><div>2. Surtitles with the stage interpretation as the source text, condensing the core content.</div></div>
		Creative surtitles. Also referred to as “meta-theatrical” surtitles. They can be used for added meaning and a visual or provocative effect (De Frutos 2013: 127).
	Opera accessibility	<div><div>1. Relaxed performances for patrons within the autistic spectrum or with other sensory disabilities.</div><div>2. Adapted surtitles, Sign Language Interpreting and haptic technology for deaf and hard-of-hearing patrons.</div><div>3. Audio introduction, AD, touch tours, audio subtitling, braille and large-print libretto for patrons with visual impairment.</div></div>

Table 1. Opera translation strategies

The convergence of the semiotic signs in opera can be rendered accessible for patrons with visual impairment through (standalone) audio introduction (AI), AD, touch tours, audio subtitling or a combination of the above. Standalone AI refers to a non-intrusive approach – first introduced by York (2007) at the London opera houses during the early 2000s – consisting of a 15-minute introduction of the cast, plot, and a vivid description of the visuals in the production. The AI would be followed by shorter, subsequent introductions before every act. Currently, some opera houses such as the Royal Opera House in London make AIs available via the audio platform Soundcloud. This is the closest approach to musicocentrism and has clear advantages in its cost-effectiveness and reutilisation potential. Nevertheless, standalone AI is not common in Spanish opera houses.

As for touch tours, Eardley-Weaver (2010) defines them as visits to the opera stage and backstage where patrons with visual impairment have the opportunity to touch items of the set, props, stage design models and sometimes cast members as an introduction to the visual elements of the production. This service is not offered at the Teatro Real or the Gran Teatre del Liceu (henceforth Liceu) opera house, but it is part of the accessible programme at the Teatro de la Zarzuela in Madrid.

Lastly, audio subtitling is an accessibility modality that consists of reading the subtitles or surtitles aloud. Orero (2007) reports on an experience with audio subtitling at the Liceu opera house where audio subtitles were tested with end-users in a concert version of *Roberto Devereux*. Even though the attendants were initially reticent to the continued overlapping with the music, they showed a high degree of acceptance in the questionnaire (Orero 2007: 146-147). As opposed to standalone AI, this service could be defined as logocentric and its main asset is that it could be reused unrestrictedly, provided that the audio subtitles are synchronised with the surtitles. Despite the positive outlook on this modality, it has not been established in either of the Spanish opera houses.

Shifting the focus to opera AD specifically, Corral and Lladó (2011: 163) define this modality as a commentary “designed to facilitate a comprehensive appreciation of artistic performances which incorporates both sound and visual dimensions”. Its main divergence with other AD modalities lies in the fact that three specific elements coincide in the operatic medium:

mise-en-scène, surtitles and libretto (Matamala 2005). In other words, not only do visuals need to be translated semiotically, but also (sung) dialogues. Precisely, in contrast to other AD modalities such as film, television or theatre where the overlapping of AD with dialogue or other relevant sounds is not allowed (Maszerowska *et al.* 2014), the operatic chant may be interrupted by the AD in the context of both the Liceu opera house and the Teatro Real. Matamala (2007) further outlines the four main characteristics of opera AD: the multiplicity of linguistic codes, the overlapping of the AD with the chant, a reaction capability from the audio describer to adapt to unforeseen events, and an often necessary interpretation of the events in order to achieve a concise description.

For the purposes of this study, AD is understood as the combination of AI and AD throughout the performance, “intertwined in the gaps where lyrics are not generally heard and music is considered to be less important” (Orero & Matamala 2007: 270). The audio “through” description segment not only includes the translation of the visual elements of the performance, but also a synthetic translation of the surtitles. Currently, this is the most widespread approach among European opera houses and opera AD service providers: Accès Culture in France, VocalEyes in the UK, Aristia, Aptent and the TransMedia Catalonia research group in Spain (Cabeza-Cáceres & Matamala 2008; Orero *et al.* 2019) and the Macerata Opera Festival in Italy (Di Giovanni 2018a). The combination of AI and AD would fit somewhere between logocentrism and musicocentrism.

The second topic to be addressed in our conceptual framework is the hierarchisation of the operatic semiotic signs. Generally, semiotics in the field of opera has taken an interest in the classic dispute between music and *parole*, as well the hierarchisation of operatic signs. A first example is the question of whether music alone is able to convey meaning, which has been disputed extensively in diverse disciplines such as Musicology, Narratology and Translation Studies. In tackling the meaning-making potential of instrumental vs. vocal music, Tråvén (2005: 103-104) argues that instrumental music will always be more open to interpretation, as opposed to vocal music, which clearly conveys a message or an emotion:

Music rhetoric lacked the precise quality or communicative skill of the spoken language, and since vocal music was considered to affect both the

mind and the heart, as opposed to instrumental music that “spoke” more to the heart, it consequently had a higher standing than purely instrumental music.

As Ryan puts it (2004) music is an art made of signifiers without signifieds. The author outlines three basic conditions of narrativity (2004: 8-9):

1. A narrative text must create a world and populate it with characters and objects.
2. The narrative world must undergo changes of state that are caused by physical events: either accidents or deliberate human action.
3. The text must allow for the reconstruction of an interpretive network of goals, plans, causal relations and psychological motivations around the narrated events.

These three conditions apply to opera provided that some translation strategy is used and, in our case, sensory accessibility services are offered.

All in all, instrumental music conveys emotions, while (intelligible) vocal music is able to communicate an articulate message. Even though music is seen as the defining element of opera, current public expectations, undoubtedly in the Spanish context, require accessing both the signifiers and the signifieds. This is no different for blind and visually impaired patrons.

Shifting the focus to the hierarchisation of signs, multimodal studies have seen an interest in opera – the *Gesamtkunstwerk* – as a paradigm of multimodality itself, as it welds together literary text, dramatic (staged) action and music. These three semiotic resources are generally the basis for the semiotic taxonomies put forward by multimodality scholars. Hutcheon and Hutcheon (2010: 65), for instance, draw from Kress and van Leeuwen’s distinction between *production* media “(such as the singers’ voices, gestures, motions; the orchestra’s musical sounds; the stage action and sets, etc.)” and *design* modes “(the musical score and libretto, the director’s interpretive plan, the various designers’ and performers’ visions, etc.)” as a definition of social semiosis in opera. Furthermore, Rossi and Sindoni (2017) work from a Systemic Functional Linguistics approach and propose three semantic levels within an operatic play, in descending order: semiotic systems (language, music and mise-en-scène), semiotic resources (libretto, score, performance

and staging) and semiotic components (stage direction, kinesics, props, etc.). We will refer back to the semiotics of opera and opera AD in Section 4.

3. The pilot study: a comparable corpus of opera audio description

The following analysis aims to explore corpus combination in a small-sized corpus. For methodological clarity, it is important to make a distinction between corpus triangulation and corpus combination. The defining characteristic of corpus triangulation is the explicit integration of methods, subjects or materials from the outset of the study (Hansen 2010; Malamatidou 2018: 8). Corpus combination, on the other hand, “mixes” research methods, usually taking the form of a quantitative analysis followed by a qualitative analysis. In this paper, we cannot claim corpus triangulation, as different aims are established for the linguistic section, i.e. researching patterns in the language of operatic ADM; and the semiotic section, i.e. analysing how operatic signs are hierarchised in the AD scripts. Even though synergies are presented, they are not sufficient to claim a methodology based on triangulation, which is often the case in Translation Studies corpora.

Besides, the rationale for limiting the size of the corpus is precisely its qualitative analysis and manual semiotic tagging process, which is described in more detail in Section 4. In this regard, authors such as Koester (2010) and Evison (2010: 123) have supported the adequacy of smaller corpora in highly specialised registers. As Malamatidou (2018: 53) highlights, the possible limitations of smaller corpora are counterbalanced by the “depth and breadth of [the] understanding” they allow for. Koester (2010: 67) goes on to argue that researchers can gain greater familiarity with smaller corpora, as they are able to reflect on contextual features and link them to the linguistic patterning. Given that this paper deals with a niche modality, it was considered adequate to compile a 33,999-word pilot corpus for both quantitative and qualitative purposes.

The innovation of this corpus is twofold. Firstly, for the very modality of AD it deals with. To the best of our knowledge, there have been no previous corpus studies on opera AD. This may be linked to the more limited availability of the scripts – as opposed to film productions (Salway 2007; Reviers 2018) or museum AD (Soler Gallego 2018; Perego 2019) – and an

inherent issue with the performative, live nature of AD specifically in the case of the Liceu opera house (Cabeza-Cáceres 2010). Indeed, while some audio describers may read through the scripts with minimal modifications to adapt it to the pace, others will be more inclined to improvise. Therefore, the AD script may also need to be adapted to substantial changes in the stage performance (Orero & Matamala 2007: 271). In our case, given the fact that the recordings of the live AD have not been kept, we can draw an analogy with pre-production scripts. The semi-live approach at the Teatro Real, on the other hand, implies that the scripts compiled for this corpus accurately correspond to the broadcasted AD. In Section 3.1 we further illustrate the formal divergences between the approaches of the Liceu and the Teatro Real.

The second innovation of this study lies in its dual aim: the corpus is used to extract linguistic information from the scripts, as well as to analyse (qualitatively) operatic semiotic information. In the following subsection, the corpus description, design and operating software are presented (Section 3.1). Section 3.2 offers a quantitative lexico-grammatical analysis of the language of opera AD. The semiotic hierarchisation in operatic AD is analysed separately in Section 4, which is followed by a discussion of the results in Section 5.

3.1. Selection of the sample

As outlined in Section 2, combining AI and AD is common practice in both opera houses. Yet, from a technical and formal perspective, different approaches are applied at the Teatro Real in Madrid and at the Liceu in Barcelona. Aristia, the company behind the opera AD in the Teatro Real, follows a pre-recorded strategy. In other words, the AD is voiced and divided into audio fragments that are manually launched by a technician the day of the show. At the Liceu opera house, on the other hand, AD is delivered live. In addition, the AD in the Teatro Real can be accessed via the application Teatro Real Accesible. Once the blind and visually impaired patrons scan the QR (Quick Response) code at the entrance of the opera house, the application is ready to launch the pre-recorded AI and AD, which can be accessed from any of the theatre seats. Meanwhile, at the Liceu, a number of seats

and screens are equipped with headphone plugs where patrons can connect their own headsets and listen to the live AD.

Formally speaking, there is another difference between both opera houses: the pre-recorded AD from the Teatro Real distinguishes two different voices according to their function: one of them provides a summarised account of the surtitles and the other one describes the visual content of the play. Conversely, the live AD at Liceu opera house integrates both the surtitles and the descriptions of the visual content: scenography, the characters and their actions (Cabeza-Cáceres & Matamala 2008). In the written scripts that comprise this corpus, the Catalan AD thus interweaves descriptions with synthesised surtitles, while the Spanish AD differentiates the two voices by highlighting the content of the surtitles in bold.

The corpus consists of three AD scripts in Catalan that were delivered at the Liceu by three different members of the TransMedia Catalonia research group and three AD scripts in Spanish, delivered at the Teatro Real and produced by the same audio describer from the AD service provider Aristia. Three operas were chosen under the premise that they were described both in Catalan and Spanish: *Aida*, *The Magic Flute* and *Carmen*. While the ADs for *The Magic Flute* and *Carmen* correspond to the same creative productions, *Aida* was described for two different productions. Having obtained the text documents, they were arranged in four subcorpora, on the basis of language and distinguishing AI from AD, as illustrated in Table 2.

Opera	Word count		Year of production	
	Liceu	Teatro Real	Liceu	Teatro Real
Aida	AI: 1,637	AI: 768	2007	2018
	AD: 5,039	AD: 2,929		
The Magic Flute	AI: 859	AI: 1,262	2016	2016
	AD: 5,682	AD: 5,780		
Carmen	AI: 1,415	AI: 501	2011	2017
	AD: 4,099	AD: 4,028		
	Total words:			
	AI: 3,911 (CAT), 2,531 (ES)			
	AD: 14,820 (CAT), 12,737 (ES)			

Table 2. Corpus description

As an introduction to the descriptive quantitative analysis that is central to this section, the number of total words accounts for 3,911 in Catalan and 2,531 in Spanish for the AI portion; and 14,820 words in Catalan and 12,737 words in Spanish for the description throughout the performance. The mean number of words per subcorpora ($AI_ES=844$; $AI_CAT=1,304$; $AD_ES=4,246$; $AD_CAT=4,940$) indicates lengthier descriptions in the context of the Liceu opera house. In this line, the standard deviation (sd) – a useful tool for calculating the variability of the length of text collections – accounts for 1,106.88 in the AD subcorpora, while the standard deviation for AI is considerably smaller ($sd=432.85$). In all, this preliminary set of descriptive statistics suggest that the Liceu ADs and AIs favour longer, more detailed descriptions.

3.2. *Lexico-grammatical analysis of the corpus*

As of the linguistic analysis of the study, a descriptive explanatory approach was adopted (Saldanha & O'Brien 2013: 50). In order to define the salient features of opera AD scripts, a lexico-grammatical analysis of the corpus was performed on the basis of four formal parameters: mean sentence length, open-class word frequencies, PoS distribution and type-token ratio (TTR).

The primary software employed to perform the linguistic analysis was SketchEngine, which allows for frequency list generation as well as automatic lemma annotation. As SketchEngine gathers general language corpora retrieved from the web, it also allows the researcher to compare their own sets of corpora to general language samples. The online tool performs concordance analysis too, which proved useful to disambiguate homographs.

3.2.1. Mean sentence length

The first parameter in our descriptive analysis was mean sentence length in all four subcorpora. For the Teatro Real AI subcorpus, the mean sentence length was 19.32 words, while the Liceu AI mean sentence length accounted for 21.85 words. As for the mean sentence length in the Teatro Real AD, the subcorpus displayed a mean sentence length of 6.80 words, while the same AD subcorpus for the Liceu accounted for 13.71 words. As such, we suggest that the higher percentage of words per sentence in the Catalan subcorpora is partly linked to the intermingling of surtitles and visual

descriptions. Conversely, in the Spanish AD subcorpus there was a clear separation between surtitles and descriptions of visual content: surtitles were written in bold in the text and distinguished by two different voices in the recording, and therefore surtitles were not always explicitly introduced. Moreover, the Spanish AD subcorpus was rich in short sentences such as: *Se tambalea* (“He wobbles”), *Saca su espada* (“He unsheathes his sword”) or *Repite* (“She repeats”).

For comparative purposes with general language corpora, the Catalan Web 2014 corpus, that is, the Catalan reference corpus integrated in SketchEngine, scores 20.99 words per sentence, in line with the European Spanish Web 2011, which shows 21.86 words per sentence. These scores fall closer to the AI subcorpora (*CAT*=21.85; *ES*=19.32) and, to a lesser extent, to the Catalan AD subcorpus, which features a mean of 13.71 words. Interestingly, the Spanish AD subcorpus (mean sentence length=6.80) seems to resemble the results from the VIW filmic corpus – in English, Spanish and Catalan – where the mean sentence length for professional and student AD was of 8.4 (Matamala 2018: 191). Overall, the Teatro Real AD subcorpus was richer in phrases, while the Liceu ADs included more compound and complex sentences.

3.2.2. Open-class word frequencies

Moving on to the lexical analysis of the corpus, Table 3 introduces the most frequent open class-words both in the AI and AD sections. For all four subcorpora, proper nouns (Aida, Carmen, Pamina, Sarastro, Papageno, Tamino) and common nouns linked to character identification (choir, queen, lady, soldier, woman and man), auxiliary verbs and copulae, as well as verbs of movement (to arrive, to go, to disappear, to enter, to stay) and verbs of communication (to say, to ask) comprised the most salient open-class words overall. These results are generally in line with the findings from previous corpus AD studies in the filmic modality (Salway 2007: 156; Matamala 2018: 196; Reviers 2018: 192). Yet, some elements are missing, namely the fact that there is no reference to body parts, objects or verbs linked to character description among the twenty most frequent open-class words in the four subcorpora.

Teatro Real AI_ES	Liceu AI_CAT	Teatro Real AD_ES	Liceu AD_ES
Tamino 28	Carmen 30	<i>ser</i> (to be) 142	<i>ser</i> (to be) 148
Papageno 19	<i>haver</i> (auxiliary verb) 27	<i>decir</i> (to say) 109	<i>haver</i> (auxiliary verb) 129
Pamina 17	<i>anar</i> (to go) 26	<i>ir</i> (to go) 106	<i>fer</i> (to do) 105
<i>teatro</i> (theatre) 16	José 18	<i>estar</i> (to be) 88	Radamès 97
<i>haber</i> (auxiliary verb) 14	<i>gran</i> (big, great) 18	<i>mujer</i> (woman) 88	Papageno 87
Real (real) 14	<i>acte</i> (act) 16	Tamino 83	Carmen 84
<i>dirección</i> (direction) 12	Egipci (Egyptian) 15	<i>haber</i> (auxiliary verb) 74	Tamino 84
Sarastro 11	Aida 15	<i>hombre</i> (man) 70	Aida 75
<i>minuto</i> (minute) 11	Tamino 14	Papageno 69	Pamina 71
<i>amor</i> (love) 11	<i>estar</i> (to be) 14	<i>aparecer</i> (to appear) 61	<i>aparèixer</i> (to appear) 68
<i>mágico</i> (magical) 10	Radamès 14	<i>desaparecer</i> (to disappear) 59	Jose 67
<i>coro</i> (choir) 10	<i>color</i> 13	<i>pedir</i> (to ask) 57	<i>entrar</i> (to enter) 65
<i>ópera</i> (opera) 9	<i>fer</i> (to do) 13	<i>entrar</i> (to enter) 53	<i>escenari</i> (stage) 62
<i>escena</i> (scene) 9	<i>soldat</i> (soldier) 12	Pamina 51	Amneris 60
<i>prueba</i> (trial) 8	Papageno 12	<i>suelo</i> (floor) 49	<i>dir</i> (to say) 59
<i>reina</i> (queen) 8	<i>arribar</i> (to arrive) 12	<i>centro</i> (centre) 48	<i>llum</i> (light) 54
<i>noche</i> (night) 8	<i>veure</i> (to see) 12	<i>llevar</i> (to take) 47	<i>blanc</i> (white) 53
<i>parte</i> (part) 8	Sarastro 11	<i>amor</i> (love) 46	<i>escena</i> (scene) 52
<i>producción</i> (production) 8	<i>òpera</i> (opera) 11	<i>quedar</i> (to stay) 44	<i>anar</i> (to go) 52
<i>dama</i> (lady) 7	Pamina 10	Carmen 42	<i>veure</i> (to see) 51

Table 3. Twenty most frequent open-class words in AI and AD

As introduced in Section 2, the AI segment offers a synthesised historical account of the composer and the play, a summary of the plot and the main aspects of the scenography and costumes (Cabeza-Cáceres & Matamala 2008: 99). Here, the prominent use of nouns referring to the internal structure of

the piece, i.e. opera, scene, part or act, further highlights a difference in purpose from the AD segment. Hereafter, the focus will be on the two AD subcorpora, as our aim is to determine the hierarchisation of the operatic signs when time constraints are imposed as the play moves forward.

Still considering open-class word frequencies, we now move onto the frequencies per open-class word. As illustrated in Table 4, shared units make up 60% of the twenty most frequent nouns in both the Catalan and the Spanish subcorpora. Character identification – both with proper and common nouns – proved to be particularly salient both in the Spanish and Catalan corpus. Other highly frequent words belonged to the location and body part semantic classes. This is in line with the findings of previous AD corpora on film (Salway 2007; Matamala 2018; Reviers 2018), although they also highlight objects, which, again, seem to be less frequent in the present corpus.

Nouns		Adjectives	
AD_ES	AD_CAT	AD_ES	AD_CAT
<i>mujer</i> (woman)	Radamès (proper name)	<i>oficial</i> (official)	<i>seu</i> (possessive)
Tamino (proper name)	Papageno (proper name)	<i>negro</i> (black)	<i>blanc</i> (white)
Papageno (proper name)	Carmen (proper name)	<i>oscuro</i> (dark)	<i>gran</i> (great, big)
<i>hombre</i> (man)	Tamino (proper name)	<i>rojo</i> (red)	<i>fosc</i> (dark)
Pamina (proper name)	Pamina (proper name)	<i>blanco</i> (white)	<i>negre</i> (black)
<i>suelo</i> (floor)	José (proper name)	<i>lleno</i> (full)	<i>meu</i> (possessive)
<i>centro</i> (centre)	<i>escenario</i> (stage)	<i>mecánico</i> (mechanical)	<i>altre</i> (other)
<i>amor</i> (love)	Amneris (proper name)	<i>nuevo</i> (new)	<i>summe</i> (high)
Carmen (proper name)	<i>llum</i> (light)	<i>grande</i> (big)	<i>primer</i> (first)
<i>derecha</i> (right)	<i>escena</i> (scene)	<i>derecho</i> (right)	<i>mateix</i> (same)

Amneris (proper name)	<i>mà</i> (hand)	<i>eterno</i> (eternal)	<i>teu</i> (possessive)
Radamés (proper name)	Sarastro (proper name)	<i>divino</i> (divine)	<i>segon</i> (second)
<i>izquierda</i> (left)	<i>sacerdot</i> (priest)	<i>egipcio</i> (Egyptian)	<i>vermell</i> (red)
<i>círculo</i> (circle)	<i>amor</i> (love)	<i>pequeño</i> (small)	<i>nostre</i> (possessive)
José (proper name)	<i>home</i> (man)	<i>feliz</i> (happy)	<i>nou</i> (new)
<i>pared</i> (wall)	<i>noia</i> (girl)	<i>culpable</i> (guilty)	<i>superior</i> (superior)
<i>soldado</i> (soldier)	<i>terra</i> (land, floor)	<i>izquierdo</i> (left)	<i>blau</i> (blue)
<i>niño</i> (child)	<i>cor</i> (choir, heart)	<i>dulce</i> (sweet)	<i>egipci</i> (Egyptian)
Sarastro (proper name)	<i>esquerra</i> (left)	<i>hermoso</i> (beautiful)	<i>sol</i> (alone)
<i>mano</i> (hand)	<i>dona</i> (woman)	<i>alto</i> (tall)	<i>ple</i> (full)

Table 4. Twenty most frequent nouns and adjectives in the AD subcorpora

Regarding the adjective PoS, 40% of the twenty most frequent adjectives were shared by both the Teatro Real and the Liceu ADs. Overlapping adjectives corresponded to the following semantic sub-groups: colour (black, red, white), light (dark), size (big), origin (Egyptian), attribute (new), and value (full).

Verbs		Adverbs	
AD_ES	AD_CAT	AD_ES	AD_CAT
<i>ser</i> (to be)	<i>ser</i> (to be)	<i>no</i> (no)	<i>no</i> (no)
<i>decir</i> (to say)	<i>haver</i> (auxiliary)	<i>arriba</i> (above)	<i>más</i> (more)
<i>ir</i> (to go)	<i>fer</i> (to do)	<i>abajo</i> (below)	<i>davant</i> (in front)

<i>estar</i> (to be)	<i>aidar</i> (to help)	<i>luego</i> (then)	<i>tot</i> (all)
<i>haber</i> (auxiliary verb)	<i>aparèixer</i> (to appear)	<i>más</i> (more)	<i>després</i> (after)
<i>aparecer</i> (to appear)	<i>entrar</i> (to enter)	<i>pronto</i> (soon)	<i>darrere</i> (behind)
<i>desaparecer</i> (to disappear)	<i>dir</i> (to say)	<i>ahora</i> (now)	<i>molt</i> (very)
<i>pedir</i> (to ask)	<i>anar</i> (to go)	<i>siempre</i> (always)	<i>ara</i> (now)
<i>entrar</i> (to enter)	<i>veure</i> (to see)	<i>dentro</i> (inside)	<i>només</i> (just)
<i>llevar</i> (to take/to wear)	<i>sortir</i> (to exit)	<i>cerca</i> (near)	<i>abans</i> (before)
<i>quedar</i> (to stay)	<i>estar</i> (to be)	<i>sólo</i> (just)	<i>ja</i> (already)
<i>querer</i> (to want/to love)	<i>quedar</i> (to stay)	<i>ya</i> (already)	<i>mai</i> (never)
<i>poder</i> (can)	<i>tornar</i> (to turn, to go back)	<i>delante</i> (in front)	<i>encara</i> (yet)
<i>llegar</i> (to arrive)	<i>acabar</i> (to finish)	<i>tan</i> (such)	<i>així</i> (so, thus)
<i>dar</i> (to give)	<i>tenir</i> (to have)	<i>menos</i> (less)	<i>sempre</i> (always)
<i>acercar</i> (to approach)	<i>estimar</i> (to love)	<i>también</i> (too)	<i>aviat</i> (soon)
<i>volver</i> (to come back)	<i>trobar</i> (to find)	<i>aun</i> (even)	<i>dalt</i> (above)
<i>tener</i> (to have)	<i>poder</i> (can)	<i>jamás</i> (never)	<i>mentrestant</i> (meanwhile)
<i>ver</i> (to see)	<i>donar</i> (to give)	<i>despacio</i> (slowly)	<i>finalment</i> (finally)
<i>volar</i> (to fly)	<i>sentir</i> (to feel)	-	<i>lentalment</i> (slowly)

Table 5. Twenty most frequent verbs and adverbs in the AD subcorpora

As for verbs, shared units made up 60% of the twenty most frequent word list (see Table 5). Following Faber and Mairal Usón’s classification (1999) for lexical domains, overlapping units belonged to the following categories: existence (to be, to appear), action (to give, can), perception (to see), sentiment (to love), position (to stay), possession (to have), speech (to say) and movement (to go, to enter).

Lastly, the variability in adverbs proved to be limited in our study: the twenty most frequent adverbs practically made up the entire range of this PoS. Furthermore, the top four adverbs in the Spanish subcorpus (no, above, below and then) made up for half of the occurrences, as the top five adverbs do in the Catalan subcorpus (no, more, in front, all, after).

3.2.3. PoS distribution

Having established the most frequent open-class words, we now move on to the overall PoS distribution of the AD subcorpora. For the purpose of this study, the Catalan and the Spanish FreeLing PoS tagsets were applied. These tagsets are integrated in SketchEngine and are used to lemmatise and tag the PoS. Table 6 indicates the relative frequency results for each PoS as per 1000 words. Words tagged as numerals or “unknown”, as well as punctuation, were excluded.

Overall lexicon size	AD_ES	AD_CAT
Adjective	2.67	5.48
Adverb	3.45	3.80
Conjunction	8.36	9.13
Noun	26.00	27.34
Preposition	19.72	21,24
Pronoun	14.96	11.62
Verb	24.84	21.38

Table 6. Relative frequency of words

When analysing the most frequent PoS, nouns (*ES*=26.00; *CAT*=27.34) made up the most frequent category, followed by verbs (*ES*=24.84; *CAT*=21.38) and prepositions (*ES*=19.72; *CAT*=21.24). Interestingly, the greatest disparity

between our two subcorpora was a deviation of 2.81 for adjectives. In fact, adjectives made up the less salient word class in the Teatro Real subcorpus, while adverbs ranked last for the Liceu AD subcorpus.

Furthermore, when compared to other AD corpora, our results resemble those from the VIW corpus (Matamala 2018: 193) and from the filmic AD corpus in Dutch (Reviërs 2018), with nouns and verbs ranking first and adjectives and adverbs scoring the lowest frequencies.

3.3.4. Type-token ratio

The fourth and last parameter in the linguistic analysis is devoted to TTR, which is an established indicator of lexical variation in corpus analysis (Baker 1995). TTR is calculated by dividing the number of unique words (types) by the total running words (tokens) in a corpus. A low TTR thus implies greater repetition, that is, less lexical variation. Typically, spoken genres and L2 learner corpora (Durán *et al.* 2004) present a lower TTR. The reliability of this measurement has nonetheless been challenged, as the ratio will be largely influenced by the text length – the shorter the text, the bigger the ratio (McCarthy & Jarvis 2010: 382), and, in the case of translation corpora, it may be an insufficient measurement due to the differences in morphology between languages (Corpas-Pastor 2008). Standardised TTR (calculated on the basis of 1000 words) is a measurement that can mitigate the influence of the text length.

As for the results in our corpus, the Liceu AD subcorpus scored 20.53% TTR, with an overall lexicon size of 3,043 words. Regarding the Teatro Real AD subcorpus, the TTR amounted to 22.31%, with a lexicon size of 2,841 words. Both subcorpora therefore showed a low degree of lexical variation.

If we look at the standardised TTR for each individual AD script (Table 7), however, we find higher lexical variation. This suggests that a large portion of the lexicon is shared by the scripts; that is, the same words are repeated not so much within the same script, but throughout the ADs from the same genre.

	Teatro Real AD_ES			Liceu AD_CAT		
	Aida_ES	Carmen_ES	Magic flute_ES	Aida_CAT	Carmen_CAT	Magic flute_CAT
Standardised TTR	38.40%	37.05%	36.62%	39.94%	39.20%	38.36%
Overall standardised TTR	37.10%			39.16%		

Table 7. Standardised TTR

On a final note, for our opera TTR and standardised TTR to be compared with the results from other AD corpora, we first need to distinguish between the studies that apply each of the two measurements. Arma’s (2011) study on filmic AD reports 26.0% TTR for English and 31.5% for Italian AD. On the other hand, Perego’s (2019) study comprising 18 standalone ADs from the British Museum scores 51.07% TTR, a much higher ratio. As for standardised TTR, Revier’s (2018) study on Dutch filmic AD indicates 38%, and Soler Gallego’s (2018) corpus on museum AD in English scores a median standardised TTR of 42.5%. We will expand on the comparability of the results in the discussion section.

4. A semiotic analysis of the corpus

The second aim of the study was to analyse and hierarchise the semiotic information in our AD corpus. Specifically, we raise the following questions: How are the aesthetics of opera reflected in opera AD scripts? Can we extract any information at all about the hierarchisation of the signs in contemporary opera productions? In an attempt to answer these questions, a semiotic tagging system was adopted and applied to the corpus.

4.1. A conceptual framework for the semiotic tagging of opera AD

The polysemiotic nature of audiovisual texts has been central in AVT studies. For instance, Delabastita (1989) synthesized the simultaneous signifying codes in audiovisual products as follows: acoustic verbal signs (dialogue,

dubbing, lyrics), visual verbal signs (subtitles, text on screen), acoustic non-verbal signs (music, diegetic and non-diegetic sounds) and visual non-verbal signs (images, camera language). Another relevant contribution is Gottlieb's (2005: 54-55) semiotic taxonomy of translation, where audio description falls within the umbrella of adaptational, intersemiotic verbalised translation.

In this study, we suggest a semiotic tagset and apply it to a textual corpus in an attempt to find out how operatic semiosis translates into the Catalan and Spanish operatic AD. In order to set a comprehensive semiotic tagset, we depart from the three pillars of opera, namely "*le parole*", "*la musica*" and "*le immagini*", from which Rędzioch-Korkuz (2016: 42-43) proposes a semiotic division of elements that make up the operatic macrocode:

- (1) LANGUAGE
 - verbal signs
 - paralinguistic and prosodic codes
- (2) MUSIC and SOUNDS
 - instrumental music
 - vocal music
 - sound effects
 - noises
 - non-verbal reactions of the audience
- (3) THEATRICAL FORMS
 - kinesics
 - proxemics
 - acting
 - scenography
 - dancing
 - theatre architecture

Departing from Rędzioch-Korkuz's proposal (2016) and drawing from the TRACCE narratology tagset (Jiménez Hurtado 2010: 71), a character identification tag was added, as well as two semiotic components within the scenography category: lighting and props. Furthermore, an additional tag was created to encompass AD inserts.

Semiotic Tagset	Language	(Synthesised) subtitles	<div>Lighting Props</div>
	Music and sounds	Instrumental music Vocal music Noise	
	Theatrical forms	Character Kinesics Proxemics Acting Scenography Wardrobe Theatre architecture Dancing	
	AD insert	(notes from the describer, references to the AD structure, mentioning of elements outside the stage space)	

Table 8. Conceptual map for the semiotic tagset. Adapted from Rędzioch-Korkuz (2016) and Jiménez Hurtado (2010)

Table 8 shows the semiotic tagset applied in the analysis. The Language category corresponds to the translated chant. Describers for the Teatro Real and for the Liceu are provided with a list of the subtitles as part of their working material. Subtitles are therefore included in the AD scripts and they are either reformulated and synthesised or directly read aloud. At the Liceu, subtitles are generally synthesised and grouped into longer descriptive units, while the Teatro Real approach favours a word-by-word reading of the subtitles, which are clearly separated from the descriptive content. The second category, music and sounds, shows allusions to instrumental music (*a ritme de flauta i oboe*: to the beat of the flute and the oboe; *acompanyats del so de tompetes*: accompanied by the sound of trumpets), vocal music (*cantan loas al vencedor*: they sing the praises of the victor; *canten el cèlebre «trio de les cartes»*: they sing the notorious Card Trio), and noises (*en sentir sorolls*: hearing noises; *sent un gemec*: [he] hears a moan). The theatrical forms, our third category, includes the majority of tags. Among them, kinesics, proxemics and acting are perhaps the closest signs, sometimes displayed

interchangeably or simultaneously (Bobes Naves 2004). As such, it may be difficult to tell them apart. For the sake of clarity, kinesics refers to gestures, mimicry and facial expressions²; proxemics refers to the stage position of the performers, the distance among them and the distance between performers and the audience. As for acting, the tag refers to the performers' movements, actions and behaviour (*cae al suelo agotado*: exhausted, he falls to the ground; *torna a sanglotar*: he starts to sob again).

Once the tags were established and tested for ambiguity in a separate AD script, the manual tagging process was performed using the qualitative analysis software Atlas.ti. Among other uses, this software allows for the creation of tags or "codes" that can be arranged in interlinked networks. The texts were thus annotated, with words acting as the minimal sense unit and sentences as the maximal sense unit, provided that the sentence concerned only one of the semiotic tags. Here, an example is presented for both cases:

(1) [character] Micaela i José [end/character] [vocal music] acaben la cançó [end/vocal music] [proxemics] ben abraçats [end/proxemics] [scenography] al costat de la cabina de telèfon [end/scenography].

[character] Micaela and José [end/character] [vocal music] finish their song [end/vocal music] [proxemics] hugging tightly [end/proxemics] [scenography] next to the phone booth [end/scenography].

(2) [scenography] A la izquierda hay un árbol con una horca y a la derecha una bomba [end/scenography].

[scenography] To the left there is a gallows tree and to the right there is a bomb [end/scenography].

Quantifying the frequencies for each tagset allows for the identification of the most salient semiotic resources in the AD scripts. As shown in Figure 1, the most salient semiotic tagset was undoubtedly character identification (36.16%). This is no surprise if we refer back to the most frequent nouns in Table 4. Even more so than in other modalities, such as filmic AD, operatic characters need to be identified not only when they are firstly introduced or

2. Rędzioch-Korkuz (2016: 46-47) argues that kinesics may be peripheral in the operatic context, as the singing interferes with the performers' ability to fully control their face muscles.

when they enter, leave and move around the stage, but also when surtitles are introduced, as the overlapping in singing makes character identification more confusing and the entire cast can be on stage simultaneously. Yet another factor may be that recognising an actors’ voice in one’s own language is easier than recognising a singing voice in another language. For comparative purposes, the character introduction tag within the narratology tagset in TRACCE’s filmic corpus presented a very low frequency (120 occurrences, as opposed to the most frequent tags: Action=7,985, and Indoor location=571 (Jiménez Hurtado & Soler Gallego 2013: 584).

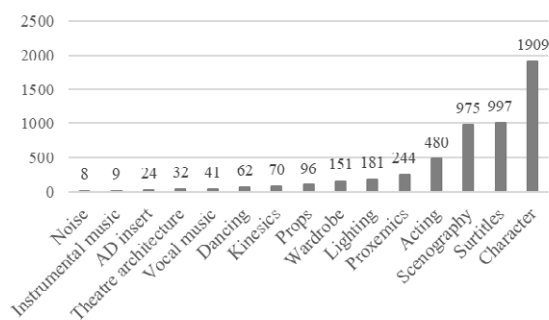


Figure 1. Frequency analysis of the semiotic tagset

Surtitles accounted for 18.89% of the total semiotic occurrences (N=5279). In third place, the scenography tagset was identified (18.47%), encompassing a large number of visual meaning-making signs, namely the scene – for instance, a gyratory stage, and the characters’ interaction with the stage and its inanimate elements. When it comes to the performers’ movement and physical expression of emotions, acting (9.09%) encompasses most of the meaning-making occurrences while proxemics (4.62%) and kinesics (1.33%) appear to be less prominent.

Finally, we present the semiotic similarities and differences between the two AD subcorpora (Figures 2 and 3), which are expanded on in the discussion section.

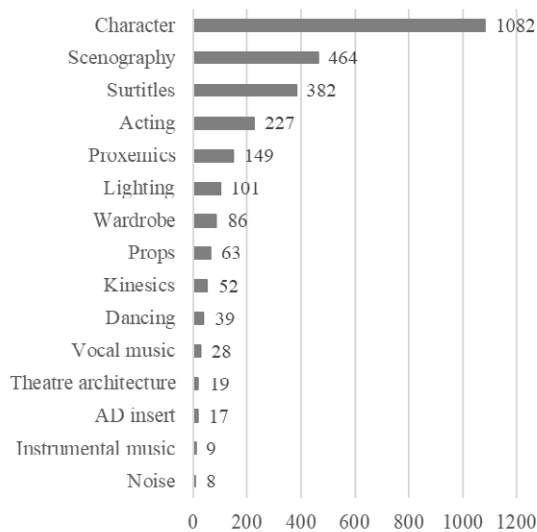


Figure 2. Semiotic distribution of the Liceu AD

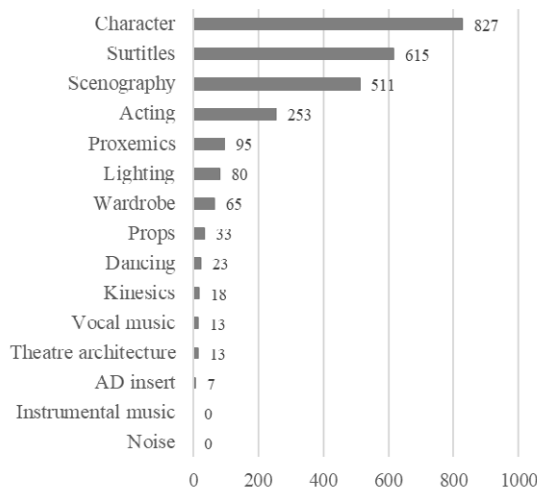


Figure 3. Semiotic distribution of the Teatro Real AD

5. Discussion

In this section, the results from the pilot study are discussed with the aim of putting forward a tentative set of lexico-grammatical and semiotic characteristics of opera AD in the Spanish context. Limitations from the study are addressed later on. Overall, in spite of the formal differences between the Teatro Real and the Liceu AD, both subcorpora share a great number of lexico-grammatical features, i.e. overlapping of the most frequent nouns and verbs, similar PoS distribution and resembling TTR and standardised TTR. Yet, there are some significant discrepancies, i.e. mean sentence length: (AD_CAT=13.71; AD_ES=6.80).

When compared to other AD modalities, namely film, our opera subcorpora also reveal certain discrepancies. In terms of lexico-grammatical patterns, the mean sentence length for the AI (CAT=21.85; ES=19.32) resembles that of the general language corpora for Catalan and Spanish. Soler Gallego's (2018: 236) corpus on museum AD in English shows similar results, with 19 words per sentence. The Teatro Real AD mean sentence length (ES=6.80), however, is closer to the results from the VIW corpus, with 8.4 words per sentence (Matamala 2018: 191); while the mean sentence length from the Liceu AD subcorpus (13.71 words per sentence) falls closer to Reviers' (2017: 84) corpus on filmic AD in Dutch (14.27). Nonetheless, these resemblances are not sufficient to establish any tangible comparisons, as they leave out language-specific morphology and syntactic elements (Saldanha & O'Brien 2013: 88). A certain trend has been spotted in the Liceu subcorpora consisting of complex and compound sentences, as opposed to the higher prevalence of phrases and independent clauses in the Teatro Real AD subcorpus.

Regarding open-class word frequencies: a high percentage of the most frequent nouns, adjectives and verbs are shared between our subcorpora. Conversely, some idiosyncratic lexico-grammatical features shared in filmic AD corpora (Salway 2007; Matamala 2018; Reviers 2018) are not particularly salient in our operatic corpus, e.g. verbs linked to character description, and nouns indicating objects and indoor and outdoor locations. These locations are replaced by more scenic and less cinematographic positions, such as *left*, *right*, *stage*, *wall*, *scene* and *floor*.

As for the PoS distribution, even though adjectives represent less than 6% of the overall word count in both AD subcorpora, they are more prevalent in the Liceu scripts. In this sense, adjectival phrases are thought to contribute to the lengthier sentences in the Catalan corpus.

When discussing lexical variation, TTR results from the opera pilot corpus ($ES=22.31\%$; $CAT=20.53\%$) differ greatly from the standardised TTR ($ES=37.10\%$; $CAT=39.16\%$). As we seek to compare our results with those in previous AD corpora, we face the challenge of non-uniformity: some studies apply the TTR measurement (Arma 2011; Perego 2019), while others report their findings with standardised TTR (Reviere 2018; Soler Gallego 2018). In any case, excessive lexical variation is generally avoided in AD, as the aim is to foster access, irrespective of the patrons' "language competence, background preparation and individual attention patterns" (Hutchinson & Eardley 2018: 8–9, in Perego 2019: 338). What we can gather is that both TTR and standardised TTR results for AD rarely surpass a 50% ratio.

Regarding the semiotic dimension of the analysis, both subcorpora are surprisingly similar in terms of sign saliency. Given that the sign frequencies were not normally distributed, we conducted a Mann-Whitney test and found no statistical differences between the Teatro Real and the Liceu corpus ($U=215$, $p=0.48$). Nevertheless, as this is a pilot study and our aim is to incorporate more data to the sample, for now we stick to general frequency results.

Overall, character identification is the most relevant tag at both the Liceu and the Teatro Real. This is closely linked to the added difficulty in recognising singing voices, particularly in ensembles where the singers intrude on each other's chant. Read-aloud surtitles also play a salient role in the Teatro Real subcorpus, more so than in the Liceu subcorpus, where surtitles are synthesised and intermingled with the visual content. In both cases, surtitles convey a great deal of the emotional charge and they are responsible for moving the plot forward. Following the terminology of Social Semiotics applied to opera by Hutcheon and Hutcheon (2010), these two first tagsets coincide with the *design* of the opera, that is, the invariable elements, regardless of the specific aesthetics of a particular spectacle. Theatrical forms, on the other hand, represent the *production* of the opera, the creative output that depends on each specific staging. The visual aesthetics of opera are therefore expressed in the AD through scenography (18.47%), acting (9.09%),

proxemics (4.72%), lighting (3.43%), wardrobe (2.86%), props (1.82%), kinesics (1.33%) and dancing (1.17%).

Divergences between the Liceu and the Teatro Real lie mostly in the preponderance of scenography ($AD_CAT=17.02\%$; $AD_ES=20.03\%$) vs. surtitles ($AD_CAT=14.01\%$; $AD_ES=24.12\%$). The scenography tag ranks second in frequency for the Liceu AD, while surtitles occupy this position for the Teatro Real. Aside from this difference, the rest of the semiotic signs are distributed in the same order for both subcorpora, though there is no mention at all of instrumental music or noise in the Spanish subcorpus.

In short, after recognising the need to constantly name the characters and given the fact that the musical element does not need to be described, the aesthetic value in the AD is mostly placed on the scenography and the surtitles (or the operatic *parole*), and, to a lesser extent, on the performer (through acting, proxemics and kinesics).

As highlighted throughout this article, multimedia-annotated corpora encompass several technical and methodological challenges, such as representativeness, arduous alignment and annotation procedures (Soffritti 2018), and even more so in an art form which is performative by nature. Although this paper examines the textual representation of multimodality in AD, the study itself cannot be deemed multimodal, as its source of analysis is solely based on the AD scripts, and not their interaction with other modes and media, i.e. images and sounds (Jiménez Hurtado & Soler Gallego 2013: 577).

On a final note, this is an initial pilot study for a larger corpus, which will allow us to apply more in-depth statistical measurements and explore further linguistic and semiotic patterns.

6. Final remarks

This pilot study is a contribution to what is still an emerging subfield of corpus AD analysis. Going back to our initial research questions: Did we find any idiosyncratic or specialised language, and if so, was it significantly different to previous (filmic) AD corpora? Is there a clear hierarchisation of the operatic signs in AD scripts?

First of all, the present study has found a great number of lexico-grammatical commonalities between the Liceu and the Teatro Real subcorpora,

with some disparities in sentence length and adjective frequency. Conversely, some of the idiosyncratic lexico-grammatical features that are common in filmic AD corpora (Salway 2007; Reviers 2018; Matamala 2018), such as the semantic category of objects and indoor and outdoor locations, were not salient in the operatic medium. We can therefore corroborate that there are lexico-grammatical characteristics that are idiosyncratic to AD (regardless of the modality), but differences in genre exist and deserve further reflection.

Second, the article aimed to test the saliency of semiotic resources in opera AD. More specifically, we wondered whether the prioritisation of the signs in AD could tell us something about contemporary opera *design* and *production* (Hutcheon & Hutcheon 2010). In this light, character identification and surtitles, ultimately the two most common tags, belong to the *design* of the opera, just as much as the musical score. Here, we emphasise that surtitles (either read-aloud or synthesised) could potentially be reused for every performance in the same production (Orero 2007).³

Within the *production* media, scenography is the clear priority in operatic AD. Yet, is scenography as important as, for instance, the music for the audience? We raise the possibility of conducting a reception study to test our findings with viewing and blind audiences alike: Why do they go to the opera? What is the current perception of the music vs. *parole* debate, if still relevant at all? Has the aesthetic value of the stage performance ultimately overshadowed other design resources?

References

- ARMA, Saveria. (2011) *The language of filmic audio description: a corpus-based analysis of adjectives*. Naples: Università degli Studi di Napoli Federico II. Unpublished PhD thesis.
- BAKER, Mona. (1995) "Corpora in translation studies: An overview and some suggestions for future research." *Target* 7:2, pp. 223-243.
- BAÑOS, Rocío; Silvia Bruti & Serenella Zanotti. (2013) "Corpus linguistics and Audiovisual Translation: in search of an integrated approach." *Perspectives. Studies in Translatology* 21:4, pp. 483-490.
- BOBES NAVES, María del Carmen. (2004) "Teatro y semiología." *Arbor. Ciencia, Pensamiento y Cultura* CLXXVII, 699/700, pp. 497-508.
- CABEZA-CÁCERES, Cristobal & Anna Matamala. (2008) "La audiodescripción de ópera: una nueva propuesta." In: Pérez-Ugena, Álvaro & Ricardo Vizcaino-Laorga (eds.) 2008. *ULISES. Hacia el desarrollo de tecnologías comunicativas para la igualdad de oportunidades*. Madrid: Observatorio de las Realidades Sociales y de la Comunicación, pp. 95-106.
- CABEZA-CÁCERES, Cristobal. (2010) "Opera audio description at Barcelona's Liceu Theatre." In: Díaz-Cintas, Jorge; Anna Matamala & Josélia Neves (eds.) 2010. *New insights into audiovisual translation and media accessibility*. Amsterdam: Rodopi, pp. 227-237.
- CORPAS-PASTOR, Gloria. (2008) *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt am Main: Peter Lang.
- CORRAL, Anna & Ramón Lladó. (2011) "Opera multimodal translation: audio describing Karol Szymanowski's Król Roger for the Liceu Theatre, Barcelona." *JosTrans. Journal of Specialised Translation* 15, pp. 163-179.
- DE FRUTOS, Rocío. (2013) *El debate en torno al canto traducido. Análisis de criterios interpretativos y su aplicación práctica: Adaptación al castellano de la ópera "Il barbiere di Siviglia" de G. Paisiello*. Sevilla: Universidad de Sevilla. Unpublished PhD thesis.
- DELABASTITA, Dirk. (1989) "Translation and mass-communication: Film and T.V. translation as evidence of cultural dynamics." *Babel* 35:4, pp. 193-218.
- DESLACHE, Lucile. (2007) "Music to my ears, but words to my eyes? Text, opera and their audiences." *Linguistica Antverpiensia* 6, pp. 155-170.
- DI GIOVANNI, Elena. (2018a) "Audio description for live performances and audience participation." *JosTrans. Journal of Specialised Translation* 29, pp. 189-211.

- DI GIOVANNI, Elena. (2018b) "Participatory accessibility: creating audio description with blind and non-blind children." *Journal of Audiovisual Translation* 1:1, pp. 155-169.
- DURÁN, Pilar; David Malvern; Brian Richards & Ngoni Chipere. (2004) "Developmental trends in lexical diversity." *Applied Linguistics* 25:2, pp. 220-242.
- EARDLEY-WEAVER, Sarah. (2010) "Opening doors to opera. The strategies, challenges and general role of the translator." *InTraline* 12.
- EVISON, Jane. (2010) "What are the basics of analysing a corpus." In O'Keeffe, Anne & Michael McCarthy (eds.) 2010. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 122-145.
- FABER, Pamela & Ricardo Mairal Usón. (1999) *Constructing a lexicon of English verbs*. Berlin and New York: Mouton de Gruyter.
- FRYER, Louise. (2016) *An introduction to audio description: a practical guide*. Oxford: Routledge.
- GORLÉE, Dinda L. (1997) "Intercode translation: words and music in opera." *Target* 9:2, pp. 235-270.
- GOTTLIEB, Henrik. (2005) "Multidimensional Translation: Semantics turned Semiotics." In Gerzymisch-Arbogast, Heidrun & Sandra Nauert (eds.) 2005. *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation*. Saarbrücken, pp. 33-61. <https://www.euroconferences.info/proceedings/2005_Proceedings/2005_Gottlieb_Henrik.pdf>
- HANSEN, Gyde. (2010) "Integrative description of translation processes." In: Shreve, Gregory M. & Erik Angelone (eds.) 2010. *Translation and Cognition*. Amsterdam and Philadelphia: John Benjamins, pp. 189-211.
- HUTCHEON, Linda & Michael Hutcheon. (2010) "Opera: Forever and always multimodal." In Page, Ruth (ed.) 2009. *New perspectives on narrative and multimodality*. London: Routledge, pp. 65-77.
- JIMÉNEZ HURTADO, Catalina. (2010) "Fundamentos metodológicos del análisis de la AD." In Jiménez Hurtado, Catalina; Ana Rodríguez & Claudia Seibel (Coord.) 2010. *Un corpus de cine. Teoría y práctica de la audiodescripción*. Granada: Tragacanto, pp. 57-110.
- JIMÉNEZ HURTADO, Catalina & Claudia Seibel. (2012) "Multisemiotic and multimodal corpus analysis in audio description: TRACCE." In Remael, Aline;

- Pilar Orero & Mary Carroll (eds.) 2012. *Audiovisual Translation and Media Accessibility at the Crossroads*. Amsterdam: Rodopi, pp. 409-425.
- JIMÉNEZ HURTADO, Catalina & Silvia Soler Gallego. (2013) "Multimodality, translation and accessibility: a corpus-based study of audio description." *Perspectives. Studies in Translatology* 21:4, pp. 577-594.
- KOESTER, Almut. (2010). "Building small specialised corpora." In: O'Keeffe, Anne & Michael McCarthy (eds.) 2010. *The Routledge handbook of corpus linguistics*. London: Routledge, pp. 66-79.
- MALAMATIDOU, Sofia. (2018) *Corpus triangulation: Combining data and methods in corpus-based translation studies*. London: Routledge
- MASZEROWSKA, Anna; Anna Matamala & Pilar Orero. (2014). *Audio description: New perspectives illustrated*. Amsterdam: Rodopi.
- MATAMALA, Anna. (2005) "Live Audio Description in Catalonia." *Translating Today* 4, pp. 9-11.
- MATAMALA, Anna. (2007) "La audiodescripción en directo." In Jiménez Hurtado, Catalina (ed.) 2007. *Traducción y accesibilidad. Subtitulación para Sordos y audiodescripción para ciegos: nuevas modalidades de Traducción Audiovisual*. Frankfurt: Peter Lang, pp. 121-132.
- MATAMALA, Anna. (2018) "One short film, different audio descriptions. Analysing the language of audio descriptions created by students and professionals." *Onomázein* 41, pp. 185-207.
- MATAMALA, Anna. (2019) "The VIW project. Multimodal corpus linguistics for audio description analysis." *Revista Española de Lingüística Aplicada* 32:2, pp. 515-542.
- MATEO, Marta. (2012) "Music and translation." In: Gambier, Yves & Luc van Doorslaer (eds.) 2012. *Handbook of translation studies*. Amsterdam: John Benjamins, vol. 3, pp. 115-121.
- MCCARTHY, Philip M., & Scott Jarvis. (2010) "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment." *Behavior Research Methods* 42, 381-392.
- ORERO, Pilar. (2007) "Audiosubtitling: a possible solution for opera accessibility in Catalonia." *TradTerm* 13. São Paulo: Humanitas, pp. 135-149.
- ORERO, Pilar & Anna Matamala. (2007) "Accessible opera: Overcoming linguistic and sensorial barriers." *Perspectives. Studies in Translatology* 15:4, pp. 262-277.

- ORERO, Pilar; Joan Bestard; Miquel Edo; Gonzalo Iturregui-Gallardo; Anna Matamala & Iris C. P. H. de Solás. (2019) "Opera accessibility in the 21st century: new services, new possibilities." *TRANS. Revista de Traductología* 23, pp. 245-256.
- PEREGO, Elisa. (2019) "Into the language of museum audio descriptions: a corpus-based study." *Perspectives. Studies in Translatology* 27:3, pp. 333-349.
- PFANSTIEHL, Margaret & Cody Pfanstiehl. (1985) "The play's the thing. Audio description in the theatre: Margaret and Cody Pfanstiehl." *British Journal of Visual Impairment* 3:3, pp. 91-92.
- REVIERS, Nina. (2017) *Audio-description in Dutch: A corpus-based study into the linguistic features of a new, multimodal text type*. Antwerp: University of Antwerp. Unpublished PhD thesis.
- REVIERS, Nina. (2018) "Studying the language of Dutch audio description." *Translation and Translanguaging in Multilingual Contexts* 4:1, pp. 178-202.
- RĘDZIOCH-KORKUZ, Anna. (2016) *Opera surtitling as a special case of audiovisual translation. Towards a semiotic and translation based framework for opera surtitling*. Frankfurt am Main: Peter Lang.
- ROSSI, Fabio & Maria Grazia Sindoni. (2017) "The phantoms of the opera: Toward a multidimensional interpretative framework of analysis." In: Sindoni, Maria Grazia; Janina Wildfeuer; Kay L. O'Halloran (eds.) 2017. *Mapping multimodal performance studies*. London/New York: Routledge, pp. 61-84.
- RYAN, Marie-Laure. (2004) *Narrative across media: The languages of storytelling*. Lincoln, NB: University of Nebraska Press.
- SALDANHA, Gabriela & Sharon O'Brien. (2013) *Research methodologies in Translation Studies*. Manchester: St. Jerome Publishing.
- SALWAY, Andrew. (2007) "A corpus-based analysis of audio description." In Díaz-Cintas, Jorge; Pilar Orero & Aline Remael (eds.) 2007. *Media for all: Subtitling for the deaf, audio description, and sign language*. Amsterdam: Rodopi, pp. 151-174.
- SEVERAL AUTHORS (ITC). (2000) *ITC guidance on standards for audio description*. <http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf>.
- SOFFRITTI, Marcello. (2018) "Multimodal corpora and audiovisual translation studies." In Pérez-González, Luis (ed.) 2018. *The Routledge Handbook of Audiovisual Translation*. London & New York: Routledge, pp. 334-349.

- SOLER GALLEGO, Silvia. (2018) "Audio descriptive guides in art museums: A corpus-based semantic analysis." *Translation and Interpreting Studies* 13:2, pp. 230-249.
- TRÁVÉN, Marianne. (2005) "Musical rhetoric – the translator's dilemma: A case for Don Giovanni." In: Gorlée, Dinda L. (ed.) 2005. *Song and significance: Virtues and vices of vocal translation*. Amsterdam: Rodopi, pp. 103-120.
- TUOMINEN, Tiina; Catalina Jiménez Hurtado & Anne Ketola. (2018) "Why methods matter: Approaching multimodality in translation research." *Linguistica Antverpiensia* 17, pp. 1-21.
- VIRKKUNEN, Riitta. (2004) "The Source Text of Opera Surtitles." *Meta*, 49:1, pp. 89-97.
- YORK, Greg. (2007) "Verdi made visible: audio introduction for opera and ballet." In Díaz-Cintas, Jorge; Pilar Orero & Aline Remael (eds.) 2007. *Media for all: Subtitling for the deaf, audio description, and sign language*. Amsterdam: Rodopi, pp. 215-231.

BIONOTE / BIONOTA

IRENE HERMOSA-RAMÍREZ is a PhD student in Translation and Intercultural Studies at Universitat Autònoma de Barcelona (UAB). Her research interests include Multimodality, Media Accessibility and accessibility for the scenic arts. Her doctoral thesis focuses on opera audio description. She holds a B.A. in Translation and Interpreting from the University of the Basque Country and an M.A. in Audiovisual translation from UAB. She is a member of the TransMedia Catalonia research group (2017SGR113), where she collaborates in the RAD project. She is the secretary of the Catalan Association for the Promotion of Accessibility.

IRENE HERMOSA-RAMÍREZ es doctoranda en el programa de Traducción y Estudios Interculturales de la Universitat Autònoma de Barcelona (UAB). Sus intereses de investigación incluyen los Estudios Multimodales y la accesibilidad a los medios y a las artes escénicas. Su tesis doctoral se centra en la audiodescripción operística. Previamente estudió el Grado en Traducción e Interpretación en la Universidad del País Vasco/Euskal Herriko Unibertsitatea y el Máster en Traducción Audiovisual en la UAB. Forma parte del grupo

de investigación TransMedia Catalonia (2017SGR113) y colabora en el proyecto RAD. Es la secretaria de la Asociación Catalana para la Promoción de la Accesibilidad.