

---

This is the **published version** of the text:

Aznar Castillo, Noelia; Amorós Soldevila, Xènia, dir. Estado de la cuestión de la automatización en el flujo de trabajo de la audiodescripción. Bellaterra: Universitat Autònoma de Barcelona, 2023. (Tecnologies de la Traducció)

---

This version is available at <https://ddd.uab.cat/record/283613>

under the terms of the  license

TRABAJO DE FIN DE MÁSTER

2022-2023



**Estado de la cuestión de la automatización en el flujo de trabajo  
de la audiodescripción**

MÁSTER EN TRADUMÁTICA: TECNOLOGÍAS DE LA  
TRADUCCIÓN

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

Autoría: Noelia Aznar Castillo

TUTORA

Xènia Amorós Soldevila

Barcelona, 25/07/2023

**Datos del TFM / Dades del TFM / Dissertation data**

---

**Título: Estado de la cuestión de la automatización en el flujo de trabajo de la audiodescripción**

Títol: Estat de la qüestió de l'automatització en el flux de treball de l'audiodescripció

*Title: State of the Art of Automating the Audio Description Workflow*

**Autora: Noelia Aznar Castillo**

Autora: Noelia Aznar Castillo

*Author: Noelia Aznar Castillo*

**Tutora: Xènia Amorós Soldevila**

Tutora: Xènia Amorós Soldevila

*Tutor: Xènia Amorós Soldevila*

**Centro: Universidad Autónoma de Barcelona**

Centre: Universitat Autònoma de Barcelona

*Centre: Autonomous University of Barcelona (UAB)*

**Estudios: Máster oficial en Tradumática: Tecnologías de la Traducción**

Estudis: Màster Universitari en Tradumàtica: Tecnologies de la Traducció

*Studies: Official master's degree in Tradumatics: Translation Technologies*

**Palabras clave / Paraules clau / Keywords**

---

**Audiodescripción, accesibilidad, tecnologías para la audiodescripción, voces sintéticas, traducción automática, corpus multimodal**

Audiodescripció, accessibilitat, tecnologies per a l'audiodescripció, veus sintètiques, traducció automàtica, corpus multimodal

**Resumen del TFM / Resum /Abstract**

---

**La audiodescripción (AD) es el servicio que garantiza el acceso de las personas ciegas o con baja visión a contenidos audiovisuales como cines, teatros, museos, etc. y que de otra manera no podrían acceder a esta clase de contenido. A pesar del papel tan determinante e inclusivo que ocupa la audiodescripción en la sociedad y que la accesibilidad se trata de un derecho fundamental, los contenidos que disponen de este servicio son escasos debido a la falta de medios o al desconocimiento. Hoy en día vivimos en un mundo lleno de tecnologías y muchas de ellas ya se utilizan para hacer los contenidos más accesibles. Por esta razón, en este trabajo queremos encuadrar la audiodescripción dentro del marco de las nuevas tecnologías con el fin de analizar el impacto que herramientas como la traducción automática, las voces sintéticas, los corpus, la descripción automática de imágenes y vídeo, el *eye-tracking* y las audiodescripciones personalizadas pueden tener en el avance de la audiodescripción.**

L'audiodescripció (AD) és el servei que garanteix l'accés de les persones cegues o amb baixa visió a continguts audiovisuals com a cinemes, teatres, museus, etc. i que, d'una altra manera, no podrien accedir a aquesta classe de contingut. Malgrat el paper tan determinant i inclusiu que ocupa l'audiodescripció en la societat i que l'accessibilitat es tracta d'un dret fonamental, els continguts que disposen d'aquest servei són escassos a causa de la falta de mitjans o al desconeixement. Avui dia vivim en un món ple de tecnologies i moltes d'elles ja s'utilitzen per a fer els continguts més accessibles. Per aquesta raó, en aquest treball volem enquadrar l'audiodescripció dins del marc de les noves tecnologies amb la finalitat d'analitzar l'impacte que eines com la traducció automàtica, les veus sintètiques, els corpus, la descripció automàtica d'imatges i vídeo, el *eye-tracking* i les audiodescripcions personalitzades poden tenir en l'avanç de l'audiodescripció.

*Audio Description (AD) is the service that guarantees access to audiovisual content such as cinemas, theaters, museums, etc. for blind or visually impaired people, who otherwise would not be able to access this kind of content. Despite the decisive and inclusive role that audio description plays in society and the fact that accessibility is a*

*fundamental right, the contents that have this service are scarce due to lack of means or lack of knowledge. Today we live in a world full of technologies and many of them are already used to make content more accessible. Therefore, in this study we want to frame audio description within the framework of new technologies in order to analyze the impact that tools such as machine translation, synthetic voices, corpora, automatic description of images and video, eye-tracking and personalized audio descriptions can have on the advancement of AD.*

**Aviso legal / Legal notice**

---

**© Noelia Aznar Castillo, Barcelona, 2023. Todos los derechos reservados.**

**Ningún contenido de este trabajo puede ser objeto de reproducción, comunicación pública, difusión y/o transformación, de forma parcial o total, sin el permiso o la autorización de su autora.**

*© Noelia Aznar Castillo, Barcelona, 2023. All rights reserved.*

*None of the content of this academic work may be reproduced, distributed, broadcasted and/or transformed, either in whole or in part, without the express permission or authorization of the author.*

## ÍNDICE

<b>1. Introducción</b>	<b>5</b>
1.1 Presentación	5
1.2 Motivación	6
1.3 Cuestiones iniciales	7
<b>2. Traducción audiovisual</b>	<b>9</b>
<b>3. Accesibilidad</b>	<b>11</b>
<b>4. Audiodescripción</b>	<b>13</b>
4.1 Definición	14
4.2 Historia de la audiodescripción	16
4.3 Normas de la AD	22
4.4 Figura del audiodescritor	26
4.5 Proceso de la AD	30
4.6 Panorama actual	31
<b>5. Automatización de la audiodescripción</b>	<b>38</b>
5.1 Traducción automática	39
5.2 Herramientas del habla	42
5.3 Corpus	49
5.4 Image y Video Captioning	53
5.5 Eye-Tracking	57
5.6 Audiodescripciones personalizadas	62
<b>6. Conclusiones y futuras investigaciones</b>	<b>64</b>
<b>Bibliografía</b>	<b>70</b>

## ÍNDICE DE TABLAS Y FIGURAS

Figura 1: esquema de la historia de la audiodescripción en España	21
Figura 2: editor de audiodescripción de Fingertext	29
Figura 3: modelo de un sistema de TTS (Alfás et al., 2011: 110)	42
Figura 4: flujo de trabajo propuesto por el proyecto ALST (2015)	47
Figura 5: muestra del corpus multimodal del proyecto TRACCE (TRACCE, 2018)	51
Figura 6: ejemplo de una descripción automática por MS COCO en el proyecto MeMAD de Sabine Braun y Kim Starr (2019:25)	56
Figura 7: patrones de movimiento ocular (Yarbus, 1967: 190)	59
Figura 8: heat map del eye-tracking de la película Match Point (2005) (Orero & Vilaró, 2012: 311)	60
Figura 9: comparación de la AD tradicional y la AD implementando eye-tracking en la película	

Marie Antoinette (2006).....	61
Figura 10: propuesta flujo de trabajo de AD semiautomatizado.....	66



## 1. Introducción

### 1.1 Presentación

Son muchas las personas que no tienen una buena imagen de la tecnología. Esta aversión no suele venir infundada por sus características más técnicas, sino por cómo las personas y las instituciones las utilizan y la finalidad de estas herramientas. Vivimos en una sociedad que depende enteramente del poder de la información, ya que estamos ante un mundo globalizado que es resultado de los avances tecnológicos y estamos tan acostumbrados a ello que ya lo damos por sentado. Nos influye directamente en nuestro día a día el impacto de la evolución de los medios de comunicación y la manera en la que decidan cómo transmitir esa información, dado que también se determina quién y quién no recibe esta información. Por esta razón, en una sociedad que depende cada día más del contenido audiovisual como fuente de información, educación, entretenimiento, interacciones sociales, etc., las personas con discapacidad visual se encuentran en una situación de exclusión social y cultural.

Cerca de un millón de personas sufren baja visión o ceguera en España y treinta millones de personas en toda Europa. Además de este millón de personas, debemos tener también en cuenta en todas aquellas otras que a medida que envejecen pierden también visión y audición. En consecuencia, la accesibilidad tanto lingüística como sensorial se ha convertido en una de las prioridades tanto a nivel nacional como europeo. Un ejemplo de ello es el apartado J sobre la inclusión social del artículo 2 de la Ley General de Derechos de las Personas con Discapacidad y de su Inclusión Social en el que podemos leer “[...]todas las personas con discapacidad tengan las oportunidades y recursos necesarios para participar plenamente en la vida política, económica, social, educativa, laboral y cultural, y para disfrutar de unas condiciones de vida en igualdad con los demás”. Y ya no se tratan solo de leyes a nivel estatal, sino también a nivel europeo e internacional como *Article 30th of the UN’s International Convention on the Rights of Persons with Disabilities* (2006) o *Consideration number 46 of the Audiovisual Media Services Directive 2010/13/EU* (2010). Por lo tanto, la necesidad de ofrecer productos audiovisuales accesibles tanto subtítulos como audiodescritos es cada vez más imperativa.

La audiodescripción (AD) es la técnica que se utiliza para transformar la información visual (imágenes) en información oral (Fernández-Torné, 2016a: 19). Por lo tanto, es la práctica por la que los materiales audiovisuales se hacen accesibles a las personas con discapacidad visual. La creación de subtítulos y audiodescripciones es una actividad que requiere mucho tiempo y tiene un impacto económico que no todos los proveedores de contenido pueden o están dispuestos a asumir. Además de estos costes tenemos que añadir el volumen cada vez mayor de productos que necesitan ser accesibles, por lo que para hacer frente a este problema se ha optado

por recurrir a la tecnología e investigar si podría ser una solución a largo plazo para automatizar parte del proceso de audiodescripción tanto desde el punto de vista de las empresas como de los profesionales de la AD y, más importante, de los usuarios que consumen este servicio.

Este trabajo nace de la necesidad de investigar el estado actual de las diferentes tecnologías que se han valorado para ayudar en el proceso de AD y analizar en qué punto se encuentran para aplicarlas en un futuro. Es por ello por lo que hemos distribuido este trabajo en dos bloques principales: audiodescripción y automatización de la AD. A modo de introducción al mundo de la AD, hablaremos primero brevemente sobre la traducción audiovisual que es el campo donde se encuentra la AD y sobre la accesibilidad para definir mejor este concepto y destacar la necesidad de productos accesibles. Hemos dividido el apartado de audiodescripción en diferentes subapartados, donde examinaremos el concepto de AD, su historia, la figura del audiodescriptor, el proceso de la AD y el panorama actual. Nos hemos decidido por desarrollar estos subapartados porque consideramos que nos ayudan a entender y poner en contexto la audiodescripción e influyen directamente en la intención de utilizar las nuevas tecnologías en el flujo de trabajo de la AD. El último apartado antes de la conclusión es la automatización de la AD que hemos dividido en las diferentes herramientas sobre las que hemos investigado para conocer el panorama actual y el posible escenario futuro de la AD. Estas posibles implementaciones son: traducción automática, herramientas del habla, corpus, *image* y *video captioning*, *eye-tracking* y audiodescripciones personalizadas.

Queremos puntualizar que nuestro objetivo con esta investigación es ofrecer más información sobre el estado actual de la automatización en la AD con el fin de aumentar el número de productos accesibles, mejorar el acceso a la información y cómo utilizar estos conocimientos para avanzar en este campo; sin ser nuestra intención centrarnos en sustituir el factor humano.

## **1.2 Motivación**

La elección del tema de este trabajo se debe a nuestro gran interés por profundizar y satisfacer la curiosidad de aprender más sobre la audiodescripción y con ayuda de qué herramientas podemos automatizar su proceso. Tras buscar información sobre los diferentes usos y herramientas del reconocimiento del habla dimos con el proyecto ALST (Accesibilidad Lingüística y Sensorial: Tecnologías para la audiodescripción y las voces superpuestas) de Anna Matamala, investigadora de la Universitat Autònoma de Barcelona. Este proyecto fue financiado por el Ministerio de Economía y Competitividad de España, y tuvo como objetivo investigar sobre la automatización en las distintas fases del proceso de la audiodescripción por medio de la traducción automática y herramientas de *speech-to-text* y *text-to-speech*.

Nos pareció una línea de investigación muy interesante, puesto que hoy en día la automatización es una parte íntegra de muchas profesiones. Por ejemplo, en el ámbito de la traducción ya es usual utilizar la traducción automática y la posedición, aunque no es tan normal su uso en la traducción audiovisual. Por esta razón, las investigaciones cuantitativas y cualitativas sobre automatización dentro de este ámbito son de especial interés, lo que nos lleva a nuestro objeto de estudio que es la audiodescripción. Creemos que con la suficiente investigación y estudios de recepción sobre cómo implementar esta automatización en la audiodescripción para hacerla más accesible y económica para las empresas, mejorará la posibilidad de que más productos se decanten por ofrecer este servicio y así llegar a normalizar más su uso.

En un principio íbamos a centrar el trabajo en la influencia de las herramientas del habla en el ámbito de la AD. Sin embargo, cuando empezamos a investigar sobre ello, nos dimos cuenta de que no son las únicas tecnologías que se están estudiando con el fin de automatizar este proceso. Esto nos motivó a seguir profundizando en el tema y descubrimos muchas más investigaciones que analizaban la implementación de otras herramientas como la traducción automática o los corpus en la AD, así como estudios de recepción en los usuarios. Al final, decidimos decantarnos por hacer un estado de la cuestión que tuviese en cuenta diversas tecnologías de manera más amplia y variada, con el objetivo de que pudiera ser de utilidad para futuras investigaciones como un primer vistazo a la actualidad de este campo y establecer una base sólida en la que sustentarse.

### **1.3 Cuestiones iniciales**

Como hemos mencionado anteriormente, la intención de este trabajo es adentrarnos en el mundo de la audiodescripción para examinar las posibilidades que tiene de mejora con el objetivo de que se puedan crear más productos accesibles, y crear algo que pueda ser de utilidad en futuras investigaciones. Al comienzo de este trabajo, nos asaltaron múltiples preguntas que hemos podido ir resolviendo con ayuda del máster y de la investigación de otros trabajos. Sin embargo, como punto de partida de este TFM queremos formular una serie de conjeturas:

- I. Si es posible la implementación de herramientas automatizadas en el flujo de trabajo de la audiodescripción. Relacionada a esta hipótesis, también queremos dar respuesta a si esta implementación tendría una buena aceptación por parte de los usuarios o si se han realizado estudios de recepción sobre ello.
- II. Si alguna de estas implementaciones tecnológicas ayudaría a mejorar la precisión y coherencia de las descripciones o si los resultados que ofrecen ahora mismo estas herramientas no ofrecen el nivel de calidad que requiere una audiodescripción.

- III. Si estas automatizaciones permitirán aumentar el volumen de contenido accesible audiodescrito y puede ayudar a reducir los costos asociados.

## 2. Traducción audiovisual

Queremos comenzar este trabajo con una breve explicación sobre la traducción audiovisual, ya que es el campo en el cual se encuadra la audiodescripción (AD). Muchos han sido los autores que han intentado dar una definición lo más apropiada posible a la traducción audiovisual para definir sus límites y clasificarla dentro de la traducción. Y es que una de las características que hace tan única a la traducción audiovisual también es una de sus mayores dificultades. La traducción audiovisual no se limita solamente a traducir un texto, sino que tal y como dice Mayoral (2001: 34-36) se ve condicionada por diversos canales (auditivo y visual) y varias señales (imagen en movimiento, imagen fija, texto, diálogo, narración, música y ruido), lo que como resultado nos lleva a un texto mucho más complejo que depende de factores externos a la traducción. Ya señalaba Nida (1964: 77) la dificultad de este tipo de traducción: “If the translator of poetry or songs is hemmed in by the communication medium, the translator of motion pictures is subject to restrictions sometimes more severe”. Dentro de la traducción audiovisual podemos encontrar, dependiendo del autor que estemos leyendo, diferentes modalidades: doblaje, subtitulación para oyentes, subtitulación para sordos, audiodescripción y voces superpuestas o *voiceover*.

El objetivo de la traducción audiovisual es hacer accesibles los productos y servicios para toda la población, por consiguiente, parece razonable incluir dentro de este campo de la traducción tanto la audiodescripción como la subtitulación para sordos. Sin embargo, esta afirmación no siempre se dio por hecha ya que, aunque el texto meta de la audiodescripción se trate de un texto lingüístico, puede no ocurrir lo mismo con el texto origen. Este texto puede variar desde un texto audiovisual hasta una obra de teatro, una audioguía en un museo o una ópera, lo que significa que puede no llegar a incluir ningún tipo de texto escrito y depender solamente del canal oral o de describir una obra de arte apoyándose únicamente en lo que el audiodescriptor está viendo. Entonces ¿sino hay un texto origen que tengamos que traducir a otro idioma por qué incluirla dentro del campo de la traducción? Esta respuesta la podemos encontrar en una cita de Roman Jakobson de 1959 que es el primero en añadir dos nuevos tipos de traducción diferentes a la considerada hasta ese momento traducción clásica: “we distinguish three ways of interpreting a verbal sign: it may be translated into other signs of the same language, into another language, or into another, nonverbal system of symbols” (1959/2004: 139).

En este artículo Jakobson (1959) propone la siguiente clasificación que se sigue usando hoy en día para referirse a los diferentes tipos de traducción:

- **Traducción intralingüística:** aquella que se lleva a cabo dentro de la misma lengua como la reformulación, definiciones o explicaciones. Dentro de esta categoría podemos añadir la subtitulación para sordos.
- **Traducción interlingüística** o traducción propiamente dicha entre diferentes lenguas.
- **Traducción intersemiótica:** interpretación de los signos verbales de un texto mediante signos no verbales. En esta categoría podemos añadir todas aquellas prácticas en las que reproducimos un mensaje en un código semiótico diferente, como la audiodescripción.

Una vez que hemos justificado la presencia de la audiodescripción dentro de los Estudios de la Traducción y, más concretamente, dentro de la traducción audiovisual, dedicaremos los siguientes apartados a la accesibilidad y a la audiodescripción. En el siguiente apartado, revisaremos de manera breve el concepto de accesibilidad que está intrínsecamente conectado a la audiodescripción.

### 3. Accesibilidad

La accesibilidad es un concepto fundamental cuando hablamos de los derechos de las personas con discapacidad. Una de tantas leyes que recoge el derecho a la accesibilidad en las mismas condiciones para toda la población es la definición que el Real Decreto Legislativo 1/2013 nos aporta en el artículo 2 punto K el 29 de noviembre de 2013, por el que se aprueba el Texto Refundido de la Ley General de derechos de las personas con discapacidad y de su inclusión social:

La accesibilidad universal es la condición que deben cumplir los entornos, procesos, bienes, productos y servicios, así como los objetos, instrumentos, herramientas y dispositivos, para ser comprensibles, utilizables y practicables por todas las personas en condiciones de seguridad y comodidad y de la forma más autónoma y natural posible. Presupone la estrategia de «diseño universal o diseño para todas las personas», y se entiende sin perjuicio de los ajustes razonables que deban adoptarse.

Por ende, el concepto de accesibilidad no hace solo referencia a la adaptación de infraestructuras y del enfoque tradicional que generalmente suele asociarse con las facilidades físicas. Aunque esta no es una interpretación incorrecta, se ha quedado limitada, ya que este concepto también hace referencia a cualquier herramienta, aplicación o dispositivo. Ya lo dijo Patricia Sanz en el artículo número 2 de la revista digital *Progreso Digital* en 2012 cuando era directora general Adjunta de Servicios sociales para Afiliados de la ONCE:

Con la llegada del siglo XXI, la accesibilidad se ha convertido en un derecho exigible para todos los ámbitos de la vida diaria y para los medios audiovisuales. El concepto de diseño para todos supone que también las producciones audiovisuales deben ofrecerse de forma normalizada para toda la población, con independencia de que los espectadores vean o no.

Este llamado “diseño para todos” es imprescindible para lograr la accesibilidad universal y se recoge en la Ley 51/2003 de igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad en el artículo 2 punto D como:

Diseño para todos: la actividad por la que se concibe o proyecta, desde el origen y siempre que ello sea posible, entornos, procesos, bienes, productos, servicios, objetos, instrumentos, dispositivos o herramientas, de tal forma que puedan ser utilizados por todas las personas, en la mayor extensión posible.

En esta investigación nos centramos sobre todo en el concepto de accesibilidad audiovisual que el Centro Español de Subtitulado y la Audiodescripción (CESyA: 2016) define en su página web como “la condición que deben de cumplir los medios audiovisuales para ser comprensibles y utilizables por personas con discapacidad sensorial”. Esta definición enfatiza la importancia de que todos los medios audiovisuales sean accesibles para todas las personas lo que contribuye a fomentar la igualdad y la inclusión social. Además, la audiodescripción y el subtitulado velan por el derecho de acceso de los usuarios al ocio, a la cultura y a la información que consumimos como sociedad en distintas formas y lugares.

Después de discutir los significados que diferentes organizaciones entienden por el concepto de accesibilidad en España y cómo protegen los derechos de las personas con discapacidad, queremos proseguir con el siguiente apartado en el que hablaremos sobre el tema sobre el que discurre el trabajo que es la audiodescripción.



#### **4. Audiodescripción**

Según las estimaciones de la Organización Mundial de la Salud (OMS) en 2022, alrededor de 2 200 millones de personas en el mundo padecen algún tipo de discapacidad visual ya sea por una enfermedad crónica, degeneración macular relacionada con la edad o cataratas, por poner algunos ejemplos. Esto supone un 27,5 % de la población mundial. Estos datos son un poco orientativos porque no todos los países miden la discapacidad visual del mismo modo. Según la OMS, los países en vías de desarrollo representan el 90 % de la discapacidad visual en el mundo, sin embargo, se prevé un aumento de esta incidencia en los países desarrollados en un futuro debido al crecimiento de la población de la tercera edad como es el caso de Europa y, por ende, también España. Según el informe sobre la ceguera en España por la fundación Retina Plus (2019), el INE estima que en España casi un millón de personas padece algún tipo de discapacidad visual de las cuales más de 58 000 personas son ciegas.

Por este motivo, cada día está más presente el término de accesibilidad, dado que no solo se limita a la superación de barreras físicas, sino a ofrecer alternativas para que toda la población pueda acceder al mismo contenido ya sea de manera visual, escrita o auditiva. Actualmente nos encontramos en una era de información y de tecnología que solo es accesible para una parte de la población. Dejan a un lado a aquellas personas que no pueden mantener el ritmo de estos avances o no pueden acceder a ellas, un ejemplo de ello son los usuarios con discapacidad, aquellos que no tienen tanto poder adquisitivo o la población en aumento de personas mayores (Díaz-Cintas et al., 2007: 11). El acceso a la información es fundamental para formar parte de la sociedad debido a que vivimos en un mundo globalizado y la inaccesibilidad a ella puede dar paso la exclusión social y limitar las oportunidades de las personas en ciertos aspectos de la vida.

La audiodescripción tiene dos objetivos principales: inclusivo y cultural. En primer lugar, la AD busca fomentar la inclusión a las personas que consumen este servicio, permitiéndoles formar parte activamente en la sociedad al poder disfrutar de los mismos programas, películas y series que consumen sus familias o amigos. El segundo objetivo de la audiodescripción es acercar la cultura a las personas que la utilizan. para que no solo tengan conocimientos sobre historia o literatura, sino también sobre las complejas relaciones sociales que es un rasgo inherente para poder vivir en sociedad (Hernández Bartolomé & Mendiluce-Cabrera, 2004: 266-267).

Si reflexionamos sobre ello, de cierta manera también podemos aplicar este mismo concepto de accesibilidad a la traducción “tradicional”, ya que lo que hacemos al traducir de un idioma a otro es hacer accesible el texto a otra comunidad en un código lingüístico que puedan entender. Por lo tanto, debemos dejar de pensar en la accesibilidad como una actividad que

influye solo a un grupo minoritario de personas y empezar a pensar en ello como una responsabilidad que nos concierne a todos (Díaz-Cintas et al., 2007: 14).

Queríamos dedicar un primer vistazo a la situación actual para subrayar la importancia de hacer materiales accesibles para la población y formar una imagen mental más sólida de la cantidad de personas afectadas por la audiodescripción.

#### 4.1 Definición

El concepto de AD se definió por primera vez en España por Javier Navarrete (1997: 70-75), que es uno de los audiodescriptores que trabaja desde hace años para la ONCE, como “el arte de hablar en imágenes”. Esta es una definición que engloba el rasgo principal de la AD y al mismo tiempo lo define como un “arte”. Nos parece una definición sencilla, pero de lo más acertada, ya que el objetivo de las imágenes es provocar en los espectadores unas emociones y la función de los audiodescriptores es intentar plasmar en palabras estas sensaciones, lo que llevaría a una forma de arte en sí.

Existe un debate en cuanto a este tema porque autores como Kruger prefieren el concepto de audionarración (AN) en vez del de audiodescripción. Defiende que la AD nunca podrá llegar a transmitir lo que se presenta en un producto audiovisual. La idea detrás de esto es que la base en la que se sustentan productos como las películas son las imágenes y estas no podrán llegar nunca a describirse del todo a través de la palabra. De modo que a través de la audionarración en vez de describir las imágenes *per se*, lo que se haría sería renarrar el efecto que nos producen (Fryer & Freeman, 2012: 3). Como argumento para este razonamiento Kruger (2010: 235) expone que:

we interpret it based on our experience of interpreting the visual world around us. [...] The best that can be achieved is an attempt to convey the effect of the visual signs in the interest of the narrative as a whole.

Esto entraría en conflicto con una de las directrices de la guía de la audiodescripción española en la que indica que debe “evitarse transmitir cualquier punto de vista subjetivo” (AENOR & ONCE, 2005: 8). Entraremos en este tema más adelante ya que existe una gran controversia en el tema de las guías dado que no existe una guía universal y cada país tiene una propia o, en el peor de los casos, directamente no tiene porque no existe una regulación de la audiodescripción. Esta es una cuestión que habrá que reflexionar a la hora de automatizar también el proceso de audiodescripción, puesto que se deberá tener en cuenta la existencia de una guía si decidimos hacer una traducción automática de un guion audiodescrito, ya que no todos los países coinciden en las mismas normas y cada país tiene su propia manera de audiodescribir. Asimismo, autores como Vázquez (2006), que es uno de los principales

audiodescriptores de la ONCE, critica abiertamente la idea de Díaz-Cintas (2006) quien considera la posibilidad de traducir guiones audiodescriptivos desde otras lenguas.

Albert Vidal (2004: 30) explicó en *Integración*, la revista de la federación AICE (Asociaciones de Implantados Cocleares de España) que la AD “simply tries to explain the action of a play, TV programme or film, etc. in such a way that the blind person can follow the development of the plot as well as possible”.

Otra definición que se utiliza mucho para determinar qué tipo de información debería tener una AD nos la da Chris Taylor (2016: 236): “audio description (AD) is a narrative technique which provides complementary information regarding the where, who, what and how of any audiovisual content. It translates the visuals into words”. Las siguientes cuestiones suelen ser las más relevantes y en las que coinciden todas las guías de AD con ligeras variaciones: “a) What should be described, b) When should it be described?, c) How should it be described? D) How much should be described?” (Vercauteren, 2007: 142).

La definición que nos parece más completa es la que da la norma UNE 153020 (AENOR & ONCE, 2005: 4) en la cual describe la AD como una forma de traducción o explicación de la imagen a la palabra:

Servicio de apoyo a la comunicación que consiste en el conjunto de técnicas y habilidades aplicadas, con objeto de compensar la carencia de captación de la parte visual contenida en cualquier tipo de mensaje, suministrando una adecuada información sonora que la traduce o explica, de manera que el posible receptor discapacitado visual perciba dicho mensaje como un todo armónico y de la forma más parecida a como lo percibe una persona que ve.

De esta cita podemos deducir que todo producto audiodescrito deberá ser lo más objetivo posible y deberá describir todos aquellos elementos audiovisuales que sean relevantes para la comprensión. Una vez más se resalta la necesidad de que el audiodescriptor sea objetivo. La AD “intenta compensar la carencia de recepción de la información que se intenta transmitir por medio del canal visual transfiriendo parte de esta información al canal acústico” (Orrego Carmona 2013: 305).

Aunque el principal público objetivo de la AD son las personas con discapacidad visual, también puede beneficiar a aquellas personas con dificultades perceptivas y cognitivas, así como a las personas mayores que han perdido visión con la edad, alguien que desee disfrutar de la AD cuando no se dispone de medio visual como pueden ser las audioguías o audiolibros, etc. y, además, también se ha demostrado que ayuda en el aprendizaje de idiomas extranjeros (Díaz Cintas, 2006: 16).

Asimismo, CESyA en su informe escrito por Diaz Cintas (2006: 16-17) recoge, *grosso modo*, tres categorías dentro de la audiodescripción:

- **AD grabada para la pantalla:** de programas audiovisuales con imágenes en movimiento, como películas, series de televisión, documentales, espectáculos, etc.
- **AD grabada para audioguías:** de obras estáticas como monumentos, museos, galerías de arte, iglesias, palacios, exposiciones, entornos naturales y espacios temáticos en las que no hay imágenes en movimiento.
- **AD en directo o semi-directo:** de obras teatrales, musicales, *ballet*, opera, deportes y otros espectáculos similares. También entran dentro de esta categoría los congresos y cualquier manifestación pública como los actos políticos.

Por consiguiente, después de este apartado no podemos negar la necesidad persistente de que actividades culturales como las antes mencionadas tengan audiodescripción para que toda la población pueda acceder a ella. Es verdad que suministrar este servicio puede suponer un gasto añadido para algunos canales o teatros y que en ocasiones puede resultar difícil de asumir. Por esta razón, también creemos que es importante el tema que tratamos en esta investigación, ya que queremos ofrecer una solución menos costosa a través de la semiautomatización o automatización del proceso con las nuevas tecnologías y, al mismo tiempo, intentar mantener la calidad que tanto reiteran las normas y la aprobación de los espectadores. Puede que con suficiente investigación los diferentes sectores se decanten en el futuro por ofrecer más cantidad y variedad de audiodescripciones.

La administración española apoya la investigación de la AD a través de becas como, por ejemplo, la que el Ministerio de Economía y Competitividad le otorgó al proyecto ALST o como con las subvenciones que suelen aprobar todos los años para el centro CESyA (Ministerio de Derechos Sociales y Agenda 2030, 2023). También la Unión Europea ha contribuido respaldando leyes que ayudan a mejorar la accesibilidad y con ayudas económicas a proyectos de investigación de AD como ADLAB, The Pear Tree Project o HBB4ALL, de los cuales hablaremos más adelante. Muchas otras investigaciones de este ámbito forman parte de grupos de investigación como *Transmedia Catalonia*, revistas o universidades.

A continuación, describiremos de manera breve las diferentes etapas por las que ha pasado la AD hasta llegar a la actualidad.

## 4.2 Historia de la audiodescripción

Aunque la audiodescripción como experiencia profesional no se empezó a llevar a cabo hasta hace relativamente poco comparado con otras modalidades de la traducción, es una práctica que siempre se ha realizado en privado en las familias de las personas con discapacidad

visual. Así lo contaba Patricia Sanz en 2012 en la revista digital *Progreso Digital* de la ONCE: “la audiodescripción es tan antigua como el hecho de que las personas que ven les cuenten a las personas ciegas lo que sucede a su alrededor”.

El primer país que practicó la audiodescripción fue España en la segunda mitad del siglo XX, unos años después de la Guerra Civil. Las primeras AD las ofreció Gerardo Esteban, un periodista que empezó a narrar en directo películas en la emisora *Radio Barcelona* (Orero, 2007b: 112). Jorge Arandes, director de RNE y de RTVE en Barcelona, explica en una entrevista publicada en *JoSTrans* en 2007 todos los detalles de la audiodescripción como práctica profesional en aquella época y como también se aplicaba a las óperas y obras de teatro. Queremos añadir que la ONCE se creó en el año 1938, fecha que no se aleja mucho de las primeras audiodescripciones emitidas por la radio.

En el año 1988 la ONCE llevó a cabo un proyecto de investigación denominado *Sonocine* que permitió realizar las primeras audiodescripciones directamente en el audio, lo que no daba elección a la audiencia sobre si escuchar la AD o no (ONCE, 2018). Este comienzo en la radio junto con el proyecto europeo *Audetel* fueron los cimientos para el sistema de audiodescripción desempeñado por la ONCE en 1993 llamado *Audesc* que se encarga de producir y distribuir películas audiodescritas dirigidas a sus afiliados y que sigue activo en la actualidad. Reino Unido desempeñó un papel fundamental en el desarrollo de *Audetel*, mientras que en Francia su equivalente se llama *Audiovision* que comenzó en 1989 y cambió su nombre a *ARTE* en 1998 (Hernández Bartolomé & Mendiluce-Cabrera, 2004: 5). Actualmente, la ONCE tiene una aplicación móvil llamada *Apolo* en la que sus afiliados pueden acceder a un catálogo de películas, series y documentales audiodescritos.

El punto de partida de la AD en la televisión se pudo ver el 22 de febrero de 1995 en el canal de televisión autonómico andaluz *Canal Sur*. Hasta finales de 1996 se emitieron 76 películas y en 1997 comenzó una iniciativa llamada *Cine para todos* en la que además de audiodescripción también ofrecían subtítulos para sordos hasta finales de 2001. En 1999, el canal autonómico catalán *TV3* con la ayuda de la ONCE comenzó a emitir la serie *Plats Bruts* con AD. Los espectadores podían elegir si ver la serie con audiodescripción o no gracias al sistema NICAM DUAL estéreo. También emitieron las series *Majoria absoluta* (2002) y *L'un per l'altre* (2003). No obstante, no fue hasta finales de los años 2000 que se emitieron películas con AD en TVE (Orero et al., 2007: 34).

Además, el Gran Teatre del Liceu de Barcelona empezó a ofrecer un servicio de audiodescripción en sus óperas desde 2005 con la ayuda de la Associació Catalana de Cecs i Disminuïts Visuals (ACCDV). Esta colaboración supuso un proyecto pionero en España y fue acogido con entusiasmo por parte del público (Matamala & Orero, 2007: 4).

Queremos también explicar brevemente un poco de la historia de la AD en diferentes países para tener una imagen mental más precisa del contexto y evolución de la AD en comparación con España.

La audiodescripción comenzó en los teatros de Estados Unidos en 1981, por lo que fue el primer país en representar en directo una audiodescripción. Esto inspiró al resto de países a ofrecer también este servicio. Para finales de esta década Estados Unidos ya había impulsado esta práctica a más de 50 teatros (ITC, 2000: 3). En 1990 el proyecto *MoPix* puso en marcha un servicio de audiodescripción (*Descriptive Video Service*) en el que permitía a los ciudadanos elegir si activar o no la audiodescripción por medio de un canal adicional en la televisión por cable (ITC, 2000: 4).

A mitad de los años 80 comenzó a desarrollarse la AD en los teatros de Reino Unido lo que inspiró al teatro Royal Windsor a ofrecer un servicio regular de AD (Orero, 2007b: 112). La audiodescripción comenzó a emitirse en las televisiones de Reino Unido en 1991 gracias al proyecto *Audetel* (*AUDIO DESCRIBED TELEVISION*) en el que el Reino Unido desempeñó un gran papel ya que se fundó gracias a ITC (*Independent Television Commission*), a la BBC, a RNIB (*Royal National Institute for the Blind*) y a la Comisión Europea (Hernández Bartolomé & Mendiluce-Cabrera, 2004: 5).

Fue en 1989 cuando en Alemania inspirados por la presentación de la AD en el festival de cine de Cannes se decantaron por empezar a audiodescribir sus películas y en 1993 filmaron su primera película alemana con AD. Al principio, la asociación alemana para ciegos (DBSV) subvencionaba estas películas para incluir la AD, pero ahora lo hacen las propias cadenas de televisión (Orero, 2007b: 113).

En Portugal se pudo disfrutar de la audiodescripción por primera vez hace relativamente poco. Fue en 2003 a través de la radio pública Rádio e Televisão de Portugal (RTP) en la película portuguesa *Menina da Rádio* en la que los espectadores tenían que sintonizar la película en la televisión al mismo tiempo que la radio para poder acceder a la AD (Orero, 2007b: 113).

La entrada de la audiodescripción en Bélgica ocurrió en 1995 gracias a una obra de teatro de Shakespeare y un episodio de la serie de televisión *Langs de Kade* (1988). Esto sucedió a raíz de la relación entre la organización belga para ciegos *Blindenzorg Licht en Liefde* y Marcus Weisen, que pertenecía al RNIB, donde trabajaron juntos durante el proyecto de *Audetel* en 1991. Durante once años después no se volvió a publicar ninguna audiodescripción y en enero de 2006 comenzaron un nuevo proyecto para describir una serie (Orero, 2007b: 113).

Aunque China firmó la Convención Internacional sobre los Derechos de las Personas con Discapacidad de las Naciones Unidas en 2007 todavía no han tomado suficientes medidas

gubernamentales para que la cultura sea accesible para todos. En 2017 la Administración Estatal de Radio y Televisión de China (SARFT) estableció nuevas regulaciones para mejorar la accesibilidad para las personas con discapacidad visual y auditiva. Esta regulación requería que los programas de televisión ofrecieran servicios de AD y subtitulación y mejoraran su calidad, pero, en realidad, no especifica ninguna cantidad específica de productos que tengan que ser accesibles. Todavía hay muchas barreras en cuanto a accesibilidad en la práctica como la falta de recursos, la falta de la aplicación efectiva de estas políticas y la falta de conciencia sobre este servicio por parte de la población. Existe la Ley de Personas con Discapacidad de la República Popular de China (LPD) que entró en vigor en 2008 y se revisó en 2018. Esta ley tiene el objetivo de garantizar la igualdad de derechos y oportunidades en todos los aspectos de la vida de las personas con discapacidad visual y garantiza el derecho a consumir productos audiodescritos, aunque es una ley muy general que no se centra solo en la accesibilidad de la AD sino que tiene en cuenta muchos aspectos como derechos laborales, accesibilidad en edificios y transporte público, accesibilidad en la educación, etc.

La AD se encuentra en sus primeras fases de investigación en China y, pese a que lleva llevándose a cabo desde inicios del año 2000 la mayoría de la población desconoce este servicio, por lo que depende enormemente del trabajo del voluntariado para hacerlo (Tor-Carroggio & Casas-Tost, 2020: 81). No obstante, al ser China un territorio tan amplio hay mucha diferencia en la aplicación de esta práctica entre sus países. Por ejemplo, aunque en Hong Kong se utiliza más la audiodescripción que en China todavía se queda bastante atrasada, ya que no tiene ninguna política que regule esta labor. Sin embargo, Taiwán con una economía menos próspera que Hong Kong se posiciona mejor en el desarrollo de la AD (Yeung, 2007: 232-237).

La expansión de la AD gracias a las leyes de cada país, las normativas europeas y guías de estilo llamó la atención del mundo académico. Por esta razón, los Estudios de Traducción comenzaron a interesarse, lo que propició la investigación de esta práctica y su implementación dentro de la Traducción Audiovisual. La AD junto con la subtitulación para sordos formaron un nuevo campo dentro de la TAV llamado *media accessibility*, donde se especializan en técnicas de accesibilidad sensorial. Hablaremos más tarde de la importancia e influencia que tuvieron estas leyes y guías en la evolución de la AD, pero ya podemos anticipar que estas medidas ayudaron y ayudan en gran medida al aumento de productos y servicios accesibles para todo el mundo.

Podemos observar cómo España fue pionera en la aplicación de la AD y como ha apoyado sus avances y los ha implementado a la pequeña y gran pantalla, además de a obras de teatro, museos, ópera y muchos otros ámbitos. Discutiremos este tema más adelante y

comprobaremos si estos servicios son accesibles o no a la hora de la práctica. Por otro lado, queremos añadir que tener una asociación como es la ONCE en España contribuye en gran medida a hacer accesible los productos audiovisuales y también apoya a sus afiliados en la distribución y uso de las tiflotecnologías. La ONCE definió en su página web en 2017 la tiflotecnología como el “conjunto de técnicas, conocimientos y recursos encaminados a procurar a las personas con ceguera o discapacidad visual grave los medios oportunos para la correcta utilización de la tecnología”. La tiflotecnología va desde la adaptación del software y hardware, instrumentos de escritura braille, lupas de sobremesa, anotadores electrónicos de voz hasta herramientas de *text-to-speech* y *speech-to-text*. Estas dos últimas son herramientas que explicaremos en más profundidad en el apartado 5.2 dentro de la automatización de la AD y que creemos puede ayudar en el avance de la audiodescripción.



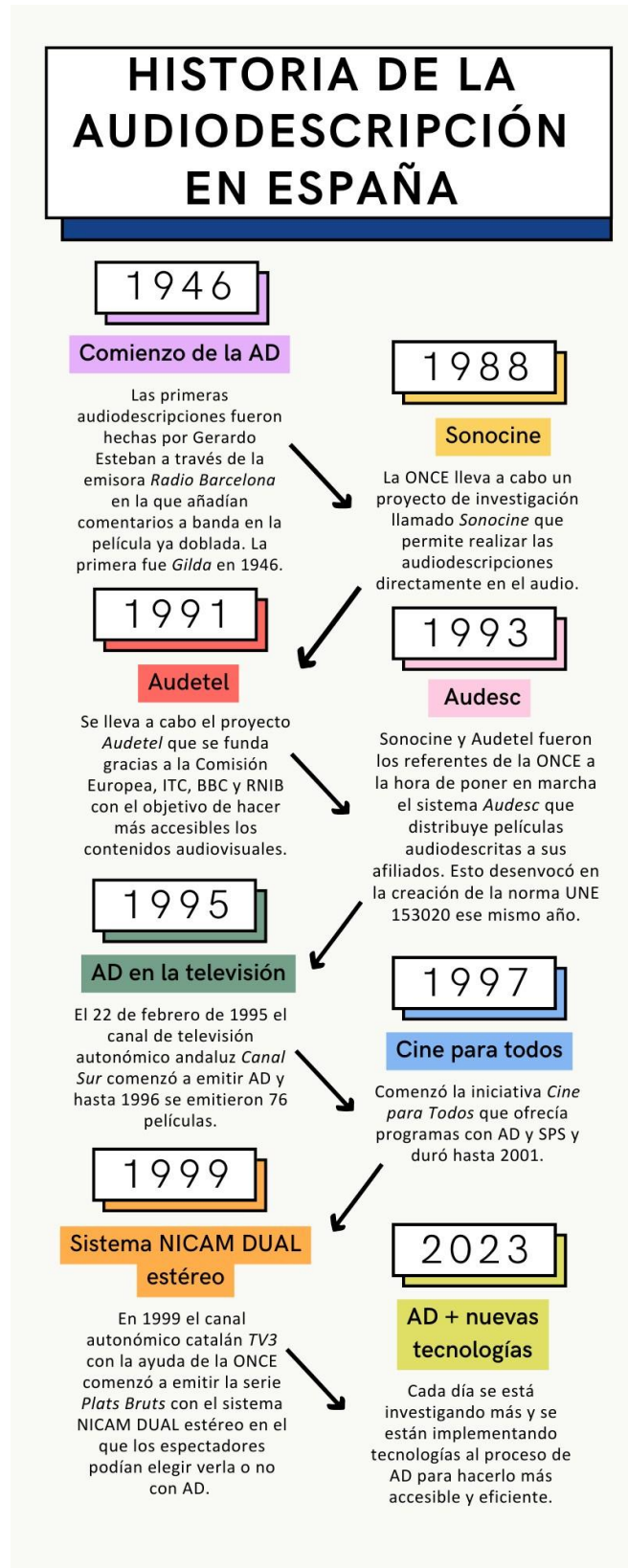


Figura 1: esquema de la historia de la audiodescripción en España

### 4.3 Normas de la AD

La iniciativa de crear una guía común a todos los países a nivel europeo sobre cómo crear un guion audiodescrito no es una novedad. Uno de los mayores proyectos en el que se investigó la posibilidad de crear una serie de normas generales para todos los países fue *The Pear Tree Project* (PTP) coordinado por Pilar Orero de la Universitat Autònoma de Barcelona y que era parte del proyecto *Digital Television for All* (DTV4All) coordinado por la Universidad Brunel. El objetivo de este proyecto era responder a la pregunta de si es plausible crear una guía europea de audiodescripción. Tener una guía de este calibre permitiría mejorar la calidad de la AD en toda Europa, que todos los países estuvieran a un nivel más similar en cuanto al avance de la AD y, además, facilitaría la posibilidad de automatizar parte del proceso de AD al traducir los guiones de un país a otro en vez de crearlos desde cero, ya que todos los guiones cumplirían con unos mismos requisitos. Esta solución sería seguramente mucho más barata y rápida, lo que ayudaría a aumentar la cantidad de productos accesibles.

Asimismo, conviene señalar que la elaboración de dicha guía se presenta como una tarea especialmente compleja, ya que la audiencia que consume los productos audiodescritos se caracteriza por su gran heterogeneidad. De hecho, dicha audiencia puede incluir tanto a personas con discapacidad visual, ya sea parcial o total, como a individuos que, a lo largo del tiempo, han experimentado una disminución de su capacidad visual. Asimismo, también cabe mencionar la posible presencia de familiares o amigos de estas personas, los cuales pueden no presentar problemas de visión. De esta manera, se puede afirmar que la audiencia de los materiales audiodescritos poseerá diferentes memorias visuales o *visual memories*<sup>1</sup>, y receptores con diversos grados de acceso al canal visual (Braun, 2008: 4).

Parece razonable comprobar si las personas que viven en diferentes países europeos y que tienen diferentes trasfondos culturales y lingüísticos perciben e interpretan los estímulos visuales de la misma manera. Los resultados de este proyecto (PTP) concluyeron que las diferencias en los datos entre los países eran demasiado elevadas como para crear unas reglas generales para todos (Mazur & Chmiel, 2012: 5).

A pesar de la ausencia de una directriz explícita a nivel europeo, en la actualidad se dispone de numerosos documentos que buscan regular la práctica de la audiodescripción. Sin embargo, la mayoría de ellos se presentan como guías profesionales de buenas prácticas, tal como es el caso de los documentos existentes en países como Francia, Alemania, Bélgica, o Estados Unidos. Del mismo modo, es importante destacar la existencia de normas internacionales establecidas por la asociación *Audio Description International* (ADI, 2003).

---

<sup>1</sup> Parte de la memoria que nos permite almacenar y recordar recuerdos visuales.

En Reino Unido encontramos la *ITC Guidelines* (ITC: 2000). La entidad responsable de la creación del documento normativo fue la Comisión Independiente de Televisión (ITC). Esta comisión se basó en el estudio de *Audetel*, encargado de investigar factores sociales, económicos y estadísticos relacionados con la audiodescripción durante los años 90.

España también dispone de su propia norma de audiodescripción, la norma UNE 153020 que ya hemos mencionado anteriormente. Dicha norma fue aprobada en el año 2005 por la Asociación Española de Normalización y Certificación (AENOR), y su redacción estuvo a cargo de miembros de la ONCE y algunos de los primeros audiodescriptores profesionales del país. Esta norma se basa en la experiencia más bien intuitiva de los profesionales dedicados a la audiodescripción en España y en las encuestas que el proyecto *Audesc*, desarrollado por la ONCE, llevó a cabo entre sus afiliados. Y es que las guías y los documentos que existen para regular la AD se basan más en la práctica que se ha llevado a cabo desde su comienzo que en la recepción de los espectadores, por lo que tienden a reflejar la experiencia de los audiodescriptores que llevan a cabo los guiones y las tradiciones de traducción audiovisual de cada país (Taylor, 2016: 239).

Contar con un marco legal que regule y fomente el uso de la AD es esencial para el desarrollo de estas guías, ya que los guiones audiodescritos deben ajustarse a ciertos criterios para crearse. Estas regulaciones también pueden ayudar a garantizar que la AD sea accesible para todas las personas con discapacidad visual y a que se sigan ciertos estándares de calidad.

Además, la guía española establece tres categorías principales de audiodescripción:

- a) la audiodescripción pregrabada para televisión, video/DVD y cine;
- b) la audiodescripción en vivo para teatros, óperas y otros eventos en vivo, por ejemplo, durante la Eurocopa 2020 la UEFA ofreció servicios de AD en algunos partidos;
- c) las audio guías para visitas a museos, exhibiciones y monumentos.

Cada una de estas categorías tienen características específicas y requiere diferentes habilidades y técnicas por parte del audiodescriptor. Es importante que el descriptor tenga en cuenta el tipo de contenido y la audiencia a la que se dirige para poder realizar una audiodescripción de calidad.

Según los estándares españoles (UNE, 153020: 2005), las normas que debe cumplir toda audiodescripción son las siguientes:

- **Objetividad:** la audiodescripción debe ser objetiva y no incluir juicios de valor, opiniones personales o información que no esté respaldada por la imagen.

- **Claridad:** la audiodescripción debe ser clara y fácil de entender para el usuario.
- **Completa:** la audiodescripción debe describir todos los elementos visuales relevantes para la comprensión del contenido audiovisual.
- **Precisión:** la audiodescripción debe ser precisa en la descripción de los elementos visuales, asegurando que la información proporcionada sea correcta y verificable.
- **Cohesión:** la audiodescripción debe mantener una coherencia en el estilo y la forma de describir los elementos visuales, evitando cambios bruscos o repentinos.
- **Sintaxis:** la audiodescripción debe seguir las reglas gramaticales y de sintaxis para facilitar su comprensión.
- **No interferencia:** la audiodescripción no debe interferir con el sonido original de la obra audiovisual.
- **No discriminación:** la audiodescripción debe respetar la dignidad y la privacidad de las personas que aparecen en la obra audiovisual, evitando comentarios o descripciones que puedan ser ofensivos o invasivos.
- **Identificación:** la audiodescripción debe identificar claramente los cambios de escena, personajes, tiempo o cualquier otro elemento relevante para la trama.

Otra norma que proporciona recomendaciones para la descripción del contenido audiovisual es la norma ISO 20071-21 de 2017 que se volvió a revisar en 2020. Esto refleja la necesidad constante de adaptar la audiodescripción a las nuevas tecnologías cuando vivimos en un mundo tan globalizado y cambiante (Orero, 2022: 408). Esta norma aborda aspectos como: directrices sobre la naturaleza subjetiva de crear AD, describe los pasos a seguir en el proceso de AD, las características lingüísticas, cómo adaptar la AD dependiendo del contenido que se está describiendo y otros aspectos (INTECO, 2020).

Varios autores han realizado análisis sobre las diferencias entre las normas internacionales, tales como Vercauteren (2007) o Rai, Creening y Petré (2010). En general, estos trabajos coinciden en que las indicaciones presentadas en estos documentos suelen ser similares en cuanto a la información que se debe describir, como el cuándo, dónde, quién y qué de la acción, así como también la mayoría de las recomendaciones generales.

Sin embargo, a pesar de la similitud en las recomendaciones generales entre los documentos internacionales, existen ciertas ambigüedades o contradicciones en algunos detalles. Uno de los ejemplos es que, aunque todas las normas comparten la recomendación de no proporcionar una cantidad excesiva de descripciones, no se especifica cuánto es demasiado o cómo determinar qué información es relevante (Vercauteren, 2007: 143). Además, se observan diferencias en cuanto a la recomendación de cuándo nombrar a los personajes, la descripción de los colores y la definición de lo que se considera verdaderamente objetivo. Por ejemplo, en

cuanto al tema de la objetividad, la norma estadounidense (American Council of the Blind, 2009) recomienda la máxima “describe lo que ves” (*What You See Is What You Say*, WYSIWYS).

A pesar de que todas las normas recomiendan mantener la objetividad al audiodescribir, evitando interpretaciones y utilizando adjetivos, adverbios y sustantivos neutros, existe cierta contradicción. Por ejemplo, se recomienda abstenerse de comentar si un vestido es bonito o feo y solo describir su color. Sin embargo, la norma británica (ITC, 2000) nos dice: “don’t shy away from using colours or describing a character as pretty, or handsome, where relevant to the story”.

Otro ejemplo es la necesidad de interpretar las expresiones faciales de los personajes de manera neutral, aunque esto implica un cierto grado de subjetividad al intentar identificar el estado emocional de los personajes. Por lo tanto, el criterio de objetividad no queda del todo claro. Además, es importante recordar que todas estas normas son más difíciles de aplicar ya que la audiodescripción es una forma de traducción subordinada que depende del tiempo disponible en los silencios de la película y tampoco se puede llenar todos estos silencios con una sobredosis de información, dado que desconcertaría y alejaría al espectador de la película.

Varios investigadores (Bourne & Jiménez: 2007; Matamala & Rami: 2009; Seibel: 2007) han comparado la audiodescripción de películas realizadas en distintos países y han encontrado una gran diversidad en cuanto al grado de objetividad. Por ejemplo, se destaca que la AD británica tiene un amplio repertorio léxico mientras que la AD en español se enfoca en describir mediante verbos generales y neutros. Jiménez (2010) también se da cuenta de que existe una notable disparidad en la cantidad de información proporcionada, ya que en su investigación la AD británica ofrece un 50 % más de información que la española.

Al investigar sobre el concepto de objetividad en la AD y la importancia que se le otorga, descubrimos que, aunque se insiste en la necesidad de mantener la objetividad lo que nos llevaría a que no debería existir la censura, este no siempre se cumple en la práctica. Podemos ver este tipo de censura en la audiodescripción de escenas íntimas que sigue siendo una práctica común en series actuales como *Sense 8* (2015), *La casa de papel* (2017), o *Élite* (2018) (Alonso Muñoz, 2020).

Es curioso que las normas enfatizan tanto en la objetividad, cuando, en realidad, todo queda a la interpretación del audiodescritor. Además, si consideramos que una de las principales funciones del cine es emocionar, se debería procurar que la AD también tenga esa capacidad y no centrarse exclusivamente en lo que es considerado o no objetivo, ya que hay tanta contradicción con este concepto.

Cada país cuenta con su propia guía de normas oficiales o extraoficiales que rigen las audiodescripciones para garantizar un nivel de calidad determinado. Sin embargo, aún no se ha podido crear una guía común europea debido a las diferencias entre los países en las características que una AD debe tener, lo que hace difícil satisfacer las necesidades de todas las audiencias. Por lo tanto, la traducción de guiones audiodescritos todavía no se ha llevado a la práctica y se necesita hacer mucha más investigación sobre los rasgos comunes en la audiodescripción de los países europeos y en las preferencias de la audiencia.

Además de las obvias ventajas de tener una guía común, como evitar que los países que no utilizan tanto o no tienen experiencia en la audiodescripción se queden aún más atrás en los avances, también sería una buena herramienta para calificar la calidad de las audiodescripciones y ayudaría a la audiencia a saber exactamente qué esperar del servicio que va a consumir, así como a las empresas saber qué servicio va a comprar o incluir en sus emisiones.

A continuación, vamos a hablar sobre la figura del audiodescriptor y la importancia de esta figura en el avance de la AD.

#### **4.4 Figura del audiodescriptor**

El audiodescriptor es una de las figuras clave que permite a las personas ciegas o con baja visión acceder al contenido audiovisual en teatros, cines, museos y otros eventos culturales. Su trabajo consiste en describir de manera detallada y clara las imágenes, escenas, personajes y otros elementos visuales importantes para que la audiencia con discapacidad visual pueda entender la obra o evento. De esta manera, el audiodescriptor ayuda a garantizar que todos puedan disfrutar de la misma experiencia cultural, independientemente de sus limitaciones visuales.

Navarrete planteaba ya en 1997 la creación de una nueva profesión como audiodescriptor de imágenes (1997: 71) en la que “se entrecruzan el trabajo artístico y creativo con la aplicación de nuevas tecnologías, la sensibilidad social con la precisión y síntesis, la amplitud y sincretismo cultural con la habilidad, efectividad y rapidez”. En este artículo pionero de la AD, Navarrete identifica cuatro habilidades fundamentales que todo audiodescriptor debe tener: formación cinematográfica ecléctica, formación cultural amplia, habilidad literaria y ser un buen espectador.

Orero (2007a) incluye a estas habilidades la capacidad de síntesis y el dominio del lenguaje, a parte de otros elementos como una voz agradable y clara. Este último rasgo es también importante porque en muchas ocasiones el propio audiodescriptor es el que se encarga de la locución del guion. Aunque, según el artículo de Raquel Sanz (2018) en el que estudia el perfil del audiodescriptor esto es algo cada vez menos habitual y es parte del trabajo del actor de

doblaje, ya que normalmente el audiodescriptor no tiene la formación adecuada para desempeñar esta función ni los equipos necesarios.

Vidal (2004: 31) también nos habla sobre la importancia del lenguaje y de la locución y apunta:

“Es necesario tener una capacidad de síntesis para explicar cualquier cosa con el mínimo de palabras, estar buscando constantemente sinónimos que signifiquen lo mismo pero sean más cortos y, por supuesto, decirlo lo suficientemente rápido pero sin dar la sensación de que se está hablando a toda velocidad para que se entienda perfectamente lo que se está diciendo”.

Las competencias de locución no siempre están incluidas dentro de las habilidades del audiodescriptor. Pilar Orero (2007a) diferencia entre dos perfiles audiodescriptores: el que escribe el guion sin realizar la locución como suele ser el caso en España y Alemania, y el que escribe el guion de la AD y lo lee como es el de los países de habla inglesa. Aunque, en las circunstancias de la audiodescripción en directo el audiodescriptor se encargará de las dos funciones.

Gracias a la aprobación de la norma UNE 153020, a la creación del CESyA en 2006 y a las diferentes leyes como la Ley General de Comunicación Audiovisual, las universidades españolas comenzaron a ofertar asignaturas sobre accesibilidad a partir de los años 2000 para que los alumnos que comienzan a integrarse en el mercado laboral tengan la formación suficiente para desempeñar la labor de audiodescriptor. Algunos ejemplos de la formación que se ofrece en los centros educativos actualmente son los siguientes:

- Máster de Traducción Audiovisual de la UAB: ofrece una asignatura obligatoria sobre “audiodescripción y subtítulo para personas sordas B-A”;
- máster de traducción audiovisual y localización de la Universidad Complutense: ofrece una asignatura optativa sobre “traducción y accesibilidad para personas con discapacidad auditiva o visual”;
- grado de Traducción e Interpretación de la Universidad de Salamanca hay una optativa de audiodescripción;
- grado superior de “Audiodescripción y subtítulo” que se reconoció en 2019 por el Real Decreto 94/2019;
- y cursos de AD en Trágora Formación o ISTRAD.

Podemos ver con estos ejemplos que la formación como audiodescriptor suele estar enlazada también a la subtítulo o a la accesibilidad en general. Sería muy útil conocer el perfil que las empresas demandan actualmente en España para orientar la formación que

deberían tener estos profesionales, por lo que se recomienda una comunicación fluida entre la empresa y la universidad para ajustarse a las necesidades del mercado.

En Alemania, el equipo de audiodescripción es particularmente interesante ya que consta de tres personas, una de las cuales es invidente o tiene baja visión. La idea detrás de esto es que cuando dos personas ven una escena, no la interpretarán de la misma manera o no verán las mismas cosas, por lo que pueden complementarse entre sí. La persona con discapacidad visual es la encargada de indicar dónde se necesita la descripción y cuánto detalle es necesario para satisfacer al público (Orero, 2007b: 116). En España, la estructura del equipo de audiodescripción varía dependiendo del proyecto y de los recursos disponibles. En general, suele haber una sola persona encargada de la audiodescripción, y, posteriormente, en la fase de revisión, es común que se incluya a alguien con discapacidad visual para que dé su aprobación final.

Gracias al trabajo de diferentes asociaciones, la investigación, las leyes y normas, y la formación impartida por las universidades, existe una concienciación cada vez mayor sobre la importancia de la accesibilidad y su aplicación en diferentes servicios y productos. La existencia de estudios homologados tiene una gran influencia en la concienciación de la población acerca de la importancia de esta labor y permite que esta profesión siga avanzando mediante la incorporación de nuevas tecnologías y las continuas investigaciones.

Queremos también hablar un poco de las habilidades tecnológicas y de las herramientas que usan los audiodescriptores en su día a día. El proyecto ADLAB PRO del que hablaremos más adelante es un proyecto que fue financiado por la Unión Europea que se centró en el avance de las herramientas tecnológicas de la AD y en la formación de los audiodescriptores. Este proyecto recabó información sobre la clase de formación que reciben los audiodescriptores y descubrieron que tener habilidades tecnológicas y enseñar los programas específicos para la AD no se consideran prioridades a la hora de formar a los nuevos audiodescriptores (Minutella, 2022). Mazur y Chmiel (2021) encuestaron para el proyecto ADLAB PRO a 86 profesores de AD de 192 cursos oficiales y vocacionales y concluyeron que “IT-related skills, such as using special software, was not prioritised in either of the two types of courses” y que “little prominence [is] given to technology in existing AD courses” (Mazur & Chmiel, 2021: 60). España se encuentra en esta misma situación, y Mendoza y Matamala (2019: 161) comentan que “technical skills... are not covered in class to a large extent, the reasons being the lack of time or the fact that they are not considered necessary”. Como resultado a esta encuesta y para dar solución a esta necesidad de una formación sobre las herramientas específicas a la hora de audiodescribir el proyecto ADLAB PRO incluye información sobre estos programas de AD y da a los estudiantes la opción de probar un editor de AD.



La mayoría de los audiodescriptores para hacer sus guiones utilizan directamente Word o programas de subtítulos, ya que estos les ayudan a crear los códigos de tiempo más fácilmente. Sin embargo, la AD también tiene sus propios *softwares* y algunos de estos programas destinados a la AD son: *LiveDescribe*, *SwiftAdept*, *Anglatècnic*, *Starfish* o *Stella Umbrella*. Queremos pararnos brevemente a hablar del programa *Fingertext* de *Anglatècnic* que es una herramienta española en la cual, además de poder editar los códigos de tiempo, las audiodescripciones pueden crearse directamente en el programa grabando la voz del audiodescriptor o usando una voz sintética con herramientas de síntesis del habla o *text-to-speech* (TTS). Otra herramienta que también permite usar TTS con más de 400 voces en 25 idiomas es *Stellar* de *yellaUmbrella*, pero *Stellar* es un widget y para poder usarla hay que descargar un programa llamado *Nebula*.

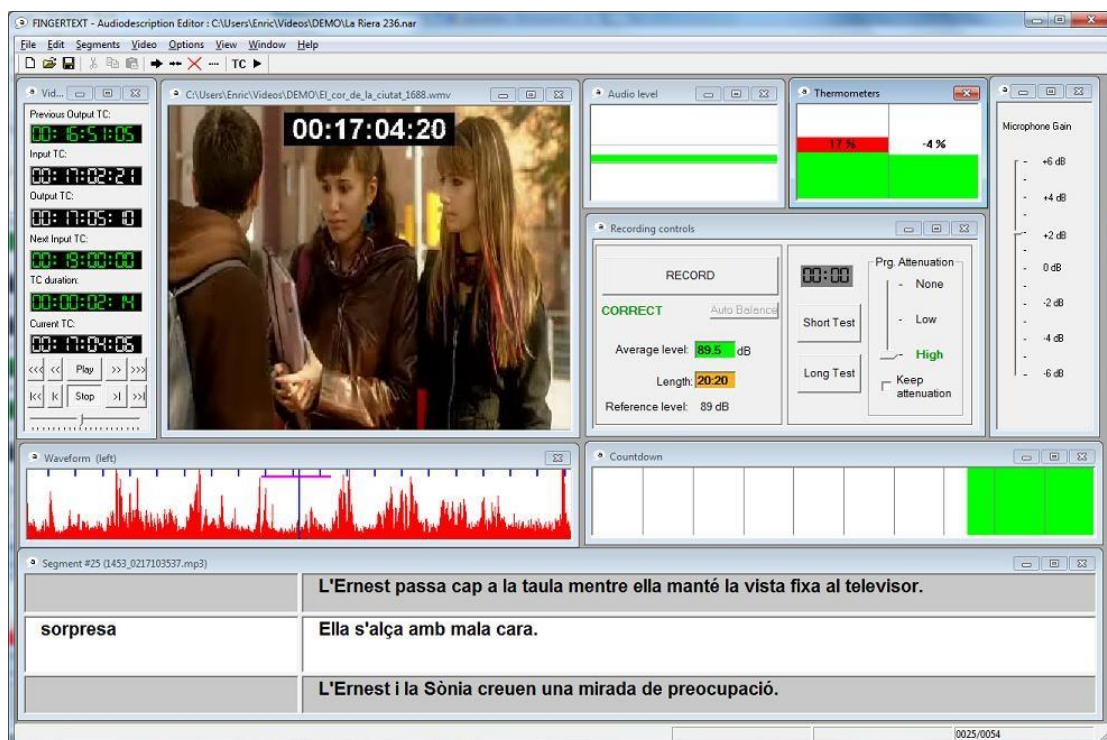


Figura 2: editor de audiodescripción de Fingertext

Aunque la mayoría de los profesionales de la audiodescripción o los docentes que enseñan AD no utilizan estos programas para audiodescribir, son herramientas que podrían acelerar el proceso y facilitar la creación de guiones. Estos programas pueden ayudar a sincronizar la AD con la banda sonora del producto audiovisual, a comprobar si las descripciones son demasiado largas o cortas o si se solapan con los diálogos u otros sonidos. A través de la síntesis de voz se podrá comprobar al instante la longitud de las frases o si la descripción combina bien con el resto del audio y hacer los ajustes necesarios. Además, cada vez más se están implementando más tecnologías para promover la creación de AD, por lo que sería recomendable que los profesionales se familiarizaran con estos programas.

El siguiente apartado lo hemos dedicado a explicar brevemente el proceso que sigue la creación de un guion de AD ya que nos ayudará a entender mejor cómo podemos mejorar este proceso en el futuro y cómo influyen todos los aspectos que hemos mencionado anteriormente directamente en la AD, como la guía o la figura del audiodescriptor.

#### **4.5 Proceso de la AD**

Todo proceso de audiodescripción debe cumplir con las pautas establecidas en la norma UNE 153020 (2005: 6-9), que incluyen requisitos específicos para garantizar la calidad y accesibilidad de las descripciones para personas con discapacidad visual:

- **Análisis previo de la obra:** se debe realizar un análisis previo del contenido visual para identificar los elementos clave que deben incluirse en la AD. Se deberá valorar si hay los suficientes huecos como para audiodescribir sin perder información importante ni interrumpir el sonido o la banda sonora. Sin embargo, tampoco se podrá aprovechar todos los silencios para rellenar con información ya que saturaría al oyente y le alejarían de la experiencia del producto audiovisual. El audiodescriptor tendrá que documentarse de la obra que se describe para utilizar un lenguaje y vocabulario que se adapte a la obra y al público objetivo. Es esencial identificar cuándo la ausencia del canal visual puede dificultar la comprensión del contenido, y determinar las estrategias adecuadas para compensar esta falta y asegurar una comprensión completa y accesible para todas las personas.

- **Descripción del contenido visual:** la AD deberá describir de manera clara, concisa y objetiva el contenido visual incluyendo todos aquellos rasgos como los personajes, el escenario, objetos, acciones, colores que sean relevantes y ayuden a las personas con discapacidad visual a entender la obra.

- **Revisión y corrección del guion:** se llevará a cabo idealmente por una persona que no sea el propio audiodescriptor.

- **Locución:** la AD se integrará a la banda sonora original sin interferir con el diálogo, los efectos de sonido y la música. Lo mejor es solo utilizar una voz para audiodescribir para no causar confusión entre los oyentes y se procurará que estas voces sean claras, neutras y correctas.

- **Montaje de la AD:** se deberá equiparar el volumen, ecualización y efectos ambientales de la banda de audiodescripción con los de la banda sonora original.

- **Evaluación de la calidad:** es necesario evaluar la calidad de la AD para asegurarse de que cumple con los estándares establecidos en la norma UNE 153020.

Este proceso es esencial para asegurar que las audiodescripciones sean claras, precisas y accesibles para todas las personas, y para fomentar la inclusión y la igualdad de oportunidades. A medida que se desarrollan nuevas tecnologías y metodologías, es importante seguir

investigando y mejorando el proceso de la audiodescripción para garantizar una accesibilidad cada vez más efectiva y completa.

#### 4.6 Panorama actual

Como hemos mencionado en apartados anteriores, en Europa la audiodescripción ha ido ganando terreno en los últimos años, y se ha convertido en una herramienta esencial para garantizar la accesibilidad de las personas ciegas o con baja visión a diferentes ámbitos como el cine, el teatro y los museos, entre otros.

En muchos países europeos, se han ido estableciendo leyes y normativas que obligan a las empresas y entidades públicas a garantizar la accesibilidad a través de la audiodescripción y otros medios. Además, se han creado asociaciones y grupos de trabajo especializados en la promoción y desarrollo de la audiodescripción, y se han llevado a cabo proyectos de investigación. Un ejemplo de ello en España es el grupo de investigación impulsado por la Universitat Autònoma de Barcelona *Transmedia Catalonia* creado por Pilar Orero y sienta sus bases en el grupo de investigación *TransMedia* formado en 2004 en Reino Unido. *Transmedia Catalonia* tiene como objetivo fomentar la investigación, la formación y la creación de productos transmedia en España, así como promover la accesibilidad (SPS, AD, *respeaking*, audiosubtitulado y lenguaje de señas) y promover la colaboración entre el mundo académico y el sector empresarial. Actualmente se orienta más en *user testing* y en las implementaciones tecnológicas a estos campos audiovisuales (Transmedia Catalonia, 2015, <https://webs.uab.cat/transmedia/>).

El panorama actual está dominado por el continuo avance de las tecnologías que ha tenido un gran impacto en el desarrollo de recursos en la accesibilidad. Se han desarrollado nuevas herramientas y tecnologías que se estudian continuamente de cara a implementarlas en el proceso de AD para mejorar su calidad. Entre las tecnologías más destacadas podemos encontrar las herramientas de reconocimiento de voz y los sistemas de inteligencia artificial (IA) que permitirán una mayor automatización y precisión a la hora de crear AD. Pero también hay otras muchas herramientas como las técnicas del procesamiento de lenguaje natural para mejorar el flujo de la narración para que sea más natural o la tecnología de seguimiento ocular (*eye-tracking*) que estudia en qué se fija en realidad la audiencia cuando visualiza un producto audiovisual y así mejorar las AD para las personas con discapacidad.

Queremos mencionar algunos de los proyectos que nos han parecido más relevantes en cuanto a su posible impacto en la AD.

- **DTV4ALL** (*Digital Television for All*) (2008-2011) (<https://cordis.europa.eu/project/id/224994/es>): este proyecto de investigación fue

financiado por la Unión Europea. Las principales líneas de investigación del proyecto incluyeron la audiodescripción, el subtítulo, la lengua de signos y la accesibilidad a través de la navegación por voz y la interacción mediante gestos (Orero, 2016: 249-261). El proyecto DTV4ALL ha sido muy influyente en la mejora de la accesibilidad en los servicios de televisión digital en Europa y ha sentado las bases para futuras investigaciones y desarrollos en este ámbito.

- **ADLAB: Lifelong Access for the Blind** (2009-2011) (<http://www.adlabproject.eu/home/>): proyecto de investigación financiado por la Comisión Europea. Durante el desarrollo de ADLAB se crearon varias herramientas y tecnologías para mejorar la calidad y la eficiencia de la audiodescripción. Algunas de estas herramientas fueron: un sistema automático de generación de descripciones basado en tecnología de reconocimiento de voz y un software de edición de audio para la AD. Además, otro objetivo de este proyecto fue la formación de los profesionales en AD, así como concienciar sobre la importancia de la accesibilidad audiovisual.
- **HBB4ALL (Hybrid Broadcast Broadband for All)** (2010-2014) (<https://cordis.europa.eu/project/id/621014/es>): esta iniciativa fue financiada por la Unión Europea y llevada a cabo por doce socios, entre ellos el grupo *Transmedia Catalonia* de la Universitat Autònoma de Barcelona. La plataforma HBB4All permitió la personalización de la experiencia de usuario en función de las necesidades de cada persona, incluyendo la selección de idioma, subtítulos, audiodescripción y lengua de signos. Además, se desarrollaron herramientas para facilitar la creación de contenidos accesibles, como software de subtítulo automático y un sistema de audiodescripción automática (Orero, 2016: 261-266).
- **The Pear Tree Project** (2011-2014): proyecto que ya hemos mencionado en apartados anteriores y que fue financiado por la Unión Europea. Tuvo como objetivo crear una guía europea común de audiodescripción para promover la accesibilidad en los medios audiovisuales de toda Europa (Mazur & Chmiel, 2012). Participantes de once países diferentes observaron el mismo corto mudo y se les pidió contar qué habían visto para analizar cómo las personas de diferentes países perciben la información y hablan sobre ello, con el fin de encontrar los aspectos comunes lingüísticos y culturales entre las diferentes nacionalidades (Mazur & Chmiel, 2012). Sin embargo, los resultados mostraron grandes diferencias en la atención que los participantes prestaban en los elementos de la película, así como el lenguaje que utilizaban para describir la historia. Siguiendo esta línea de investigación también hay muchos estudios sobre el seguimiento ocular o *eye-*

*tracking* para comprender mejor qué es lo que llama la atención de los espectadores al observar algo y que abordaremos más adelante.

- **ADLAB PRO: A Laboratory for the development of a new profesional profile** (2016-2019) (<https://www.adlabpro.eu/project/>): proyecto financiado por la Unión Europea y que se centra en el desarrollo de herramientas tecnológicas en la AD. Este proyecto es la continuación de *ADLAB: Lifelong Access for the Blind* y se llevó a cabo por ocho socios en los que se incluían universidades, organizaciones sin ánimo de lucro y empresas especializadas en tecnología de la información y la comunicación.

El objetivo principal del proyecto es mejorar la calidad y la de la AD mediante el uso de las nuevas tecnologías. Entre estas herramientas se incluyen programas de síntesis de voz, reconocimiento de voz, traducción automática, sistemas de anotación de imágenes y *eye-tracking*. Además, tenía también como meta desarrollar nuevas técnicas y metodologías para producir audiodescripciones, incluyendo el uso de modelos cognitivos y emocionales para mejorar la experiencia de los usuarios (ADLAB, 2013). ADLAB PRO ha sido uno de los proyectos más importantes en el campo de la audiodescripción y ha contribuido significativamente al desarrollo de nuevas tecnologías y metodologías de AD en Europa.

Existen muchos proyectos relevantes para esta investigación que están relacionados con la implementación de nuevas tecnologías en la audiodescripción. Los mencionados anteriormente son solo algunos ejemplos que queríamos destacar. En resumen, tanto en Europa como en España, la audiodescripción está ganando terreno y cada vez es más reconocida como una herramienta esencial para garantizar la accesibilidad y la inclusión de las personas con discapacidad visual. Aunque aún quedan retos por afrontar, como la falta de una guía común europea, estamos avanzando hacia una mayor accesibilidad audiovisual para todos en la que creemos que las nuevas tecnologías tendrán una gran influencia.

Aun existiendo todas estas investigaciones, asociaciones y leyes que protegen la accesibilidad de este colectivo, todavía no es suficiente ya que parece no existir un verdadero interés por parte de los canales en ofrecer AD de calidad, sino que solo buscan cumplir con el requisito que les exige la ley. Así lo afirman un grupo de audiodescriptores encuestados por Raquel Sanz (2018: 128) en la que denuncian la situación de precariedad de las tarifas de audiodescripción:

[E]n muchos casos [la AD] es un puro trámite, un requisito que cumplir para que acepten/emitan la obra y, por supuesto, una multa a evitar de la forma más barata posible. Esto hace que cada vez haya más concursos/licitaciones con tarifas muy bajas que no tienen en cuenta la profesionalidad ni la calidad de las adaptaciones.

Así también lo afirman los receptores invidentes en el informe del CESyA de 2015 (142-143) en el que denuncian la mala calidad de las audiodescripciones ofrecidas por televisión.

Victoria García (2020: 136-140) investiga en su artículo si las cadenas de televisión cumplen con la Ley 7/2010 General de la Comunicación Audiovisual en España en la que se establece un mínimo de 10 horas semanales de audiodescripción en los canales de televisión públicos. Sin embargo, esta ley no establece ningún parámetro de calidad de la AD, ni el horario en el que deben emitirse estos contenidos ni el tipo de programas que deben incluir los servicios de accesibilidad ya sea subtítulo, AD o lengua de signos (García, 2020: 131). Además, esta ley solo tiene en cuenta las cadenas de la TDT por lo que no existe obligación de hacer accesibles los contenidos emitidos en canales de pago, plataformas de video bajo demanda o contenidos *online* (García, 2020: 131).

No obstante, aunque los canales de televisión que no están subvencionados por el Estado no tienen obligación de ofrecer contenidos accesibles, es muy probable que en el futuro estén obligados a ofrecer un porcentaje de sus programas con audiodescripción cuando la transposición de la directiva europea 2018/1808 de servicios de comunicación audiovisual se haga efectiva en el resto de los estados miembros (Saerens et al., 2022). Además, cabe destacar que investigaciones como la de Bourne (2017) nos muestran que para el año 2050 el número de personas con discapacidad visual se habrá triplicado. Por esta razón el progreso en términos de accesibilidad debería ser aún más importante.

Consideramos que el artículo de Victoria García (2020) es de gran relevancia, ya que proporciona un análisis detallado de la situación de la audiodescripción en la televisión española. Nos resulta útil para obtener una comprensión más precisa y visual de la realidad de la AD actualmente. Nos llamó la atención de manera positiva ya que en este estudio se concluye que TVE cumplió con el porcentaje mínimo por ley al audiodescribir en 2017-2018 un 7,6 % de su programación. Sin embargo, no todas las cadenas de TVE emitieron AD y, aunque la franja horaria predominante de este contenido era en el horario de sobremesa (15-18), gran parte de este contenido se emitió en horarios poco accesibles como la madrugada (2:30-7) y utilizaron la repetición de contenido audiovisual audiodescrito (García, 2020: 138). Además, los formatos informativos no contaron prácticamente con este servicio, ya que solo el 0,1 % de ellos tenía contenido audiodescrito.

Encontramos también interesante que se usara el canal infantil para la transmisión de programas con audiodescripción destinados a adultos. Por esta razón, quisimos investigar más en el tema para comparar la práctica en España con la de otros países y encontramos respuesta en el artículo de Saerens et al.(2022: 280-294). La audiencia que no tiene problemas de visión

no necesita o quiere escuchar el programa que están viendo con AD, lo que lleva a las cadenas de televisión a explorar otras opciones para su emisión. Una de estas alternativas consiste en transmitir el contenido a través de servicios OTT (*Over The Top*) tal como hace la televisión española y la cadena nacional VRT de Bélgica. Este servicio consiste en la transmisión de los programas a través de internet. Por otro lado, a la hora de ofrecer el contenido a través de la televisión hay cadenas como VRT o BBC que se decantan por emitir el programa dos veces: una sin AD y otra con ella. En VRT intentan emitir el mismo programa simultáneamente, transmitiendo la audiodescripción en el canal para niños, pero fuera del horario infantil. Este enfoque se asemeja al adoptado por España que también utiliza este canal para emitir programas con AD.

Sin embargo, hay muchos programas que coinciden en el uso de la franja horaria fuera del horario infantil, lo que limita considerablemente la accesibilidad a los contenidos televisivos. En este sentido, se han planteado diversas soluciones al respecto. Por un lado, la cadena nacional de Noruega ha duplicado tres de sus cuatro canales principales, en estos canales duplicados se emiten todos sus programas con audio subtítulos y, cuando está disponible, con audiodescripción. Otra solución es la que utiliza la BBC que ha optado por permitir a los espectadores la elección de visualizar el programa con o sin AD a través de la opción de idiomas. No obstante, esta última opción es más compleja, ya que requiere que las compañías de televisión por cable suministren un *hardware* especial para que llegue a la televisión de todos los hogares. Es posible que con la reciente aprobación de la *European Accessibility Act* (European Parliament, 2019), las compañías de televisión por cable y los distribuidores aceleren el proceso y ofrezcan a más hogares este servicio. Por otro lado, en los Países Bajos ofrecen AD a través de una aplicación llamada “Earcatch” que se puede usar para ver películas en el cine o audiodescripciones para teatro en vivo. Esta aplicación les permite descargar la AD y escucharla a través de auriculares, simultáneamente con las personas sin discapacidad visual, durante la emisión del programa de televisión. (Saerens et al., 2022). Una aplicación parecida a esta en España es la que suministra la ONCE a sus afiliados que se llama *Apolo* y que ya hemos mencionado anteriormente.

En este ámbito de la AD en televisión, podemos ver que no todos los programas tienen las mismas facilidades para ser accesibles. Esto se debe a que los programas nuevos o en franjas horarias con menos audiencias no suelen contar con los recursos necesarios para invertir en gastos de audiodescripción, independientemente de su presupuesto. Por lo tanto, solo los programas con mayor audiencia o aquellos que hayan recibido algún tipo de ayuda o subvención por parte del gobierno podrán hacer más accesibles sus contenidos mediante la AD. Asimismo, no todos los programas tienen el mismo potencial para ser audiodescritos, ya que se requiere que haya espacios de silencio donde poder incluir la AD. En este sentido, es relevante

mencionar la estrategia denominada “bimedia” que ya en 1990 adoptó la BBC. Esta consistía en que el audio de un programa de televisión pudiera usarse directamente en la radio, por lo que se emitían programas con el mismo contenido en ambos medios, con la diferencia de que en la radio no había imágenes (Saerens et al., 2022). Este enfoque se ha aplicado sobre todo a programas de noticias y podría ayudar a otras cadenas a hacer sus contenidos más accesibles a un menor coste. No obstante, la BBC ya no usa esta técnica *per se*, sino que la ha ajustado a la actualidad y se ha decantado por ofrecer una mejor formación en accesibilidad a las personas involucradas en la producción de sus contenidos, especialmente en programas como las noticias donde no hay espacio para añadir la AD.

Otra dificultad a la que se enfrentan las personas con discapacidad visual, según la encuesta de Victoria García (2020 :141), es la falta de información previa sobre qué programas se emitirán con AD y cómo acceder a ellos. Los canales de televisión deberían tomar conciencia de esta situación y facilitar esta información de manera más clara y accesible. A pesar de que los canales están invirtiendo en AD, la audiencia no siempre sabe cuándo puede consumir estos productos, lo que supone una pérdida de dinero y de espectadores por parte de las cadenas.

Actualmente, para conocer si un programa tiene AD o no, es necesario recurrir al teletexto de TVE o la guía de programación (EPG). Sin embargo, la información de teletexto no es siempre completa y la de la EPG varía de un televisor a otro. El blog <https://versionaccesible.blog/> difunde la previsión de programación audiodescrita de TVE, pero no de las televisiones privadas o autonómicas. Un buen ejemplo de esta práctica es el de la BBC que sí que incluye la disponibilidad de AD en su guía de programación (Saerens et al., 2022: 292). Todos los canales podrían seguir este ejemplo y proporcionar información sobre la AD en sus páginas web o aplicaciones para difundir la AD y que fuera accesible a un mayor número de personas. Además, otro aspecto importante para tener en cuenta es la falta de accesibilidad en el teletexto o en la guía de programación ya que la mayoría de las personas con discapacidad visual no podrán leerlas. Esta dificultad puede resolverse si implantamos herramientas de síntesis de voz, como las que suelen usar las personas con discapacidad en sus móviles, para que lean en voz alta este contenido y así hacerlo accesible. Por esta razón, para ayudar a los usuarios a no depender de una tercera persona se desarrolló el proyecto EasyTV financiado por Unión Europea en 2018 por el cual los usuarios pueden emitir órdenes a través del reconocimiento de voz (Orero, 2022: 412).

Creemos que todos estos aspectos que hemos mencionados son relevantes para la investigación que estamos llevando a cabo ya que relaciona directamente las leyes existentes con su práctica real y los avances que gracias a herramientas como *text-to-speech* o *eye-tracking*



Estado de la cuestión de la automatización en el flujo de trabajo de la audiodescripción  
Noelia Aznar Castillo

pueden suponer a esta modalidad de la traducción audiovisual y que estudiaremos más adelante en este trabajo.

## **5. Automatización de la audiodescripción**

La automatización en el proceso de audiodescripción hace referencia al uso de tecnologías y de herramientas para automatizar una o varias partes de este proceso. Tradicionalmente, la AD se ha llevado a cabo por profesionales entrenados en describir de manera objetiva los detalles visuales de forma oral. Sin embargo, con avances en la tecnología como el reconocimiento de imágenes junto con la inteligencia artificial, se están investigando otros puntos de vista sobre cómo hacer frente a este proceso de audiodescripción para generar descripciones de audio de manera más rápida.

La automatización del proceso de AD ofrece ventajas como la rapidez en la generación de contenido, acelera el proceso de producción y reduce el número de personas involucradas, así como los costes. Esto se podría traducir en la posibilidad de mejorar la consistencia de las descripciones, ya que los sistemas automatizados siguen un conjunto de reglas predefinidas y no están sujetos a la opinión o subjetividad del audiodescriptor. Sin embargo, debemos tener en cuenta que los resultados generados por la inteligencia artificial vienen dados al fin y al cabo por los datos con los que la alimentamos para entrenarla y estos pueden tener sesgos que influirán en la información que la IA nos proporcione. Por lo tanto, la imparcialidad de la IA tampoco es inherente y no tenemos que darla por sentado. Para ayudar a fomentar este concepto de objetividad habría que analizar detenidamente la información con la que entrenamos estos modelos para reducir los sesgos y poner a la IA en contexto. Es también trabajo del usuario cuestionar la información que la IA proporciona y compararla con otras fuentes para tener una fuente de información fiable. Por otro lado, la automatización de la audiodescripción también presenta muchos otros desafíos. Los sistemas automatizados aún no llegan a ser perfectos, por lo que pueden cometer errores al identificar elementos visuales o generar descripciones. Además, la audiodescripción requiere una comprensión profunda de la narrativa, el contexto, las emociones que se quieren expresar, referencias y muchos más aspectos que dificultan que los sistemas automatizados comprendan en su totalidad lo que tienen que audiodescribir.

Se han realizado numerosas investigaciones para analizar el impacto que las nuevas tecnologías podrían tener en la audiodescripción, tanto en la audiencia como en los profesionales y empresas involucrados en este sector y se ha estudiado su aceptación por parte de los involucrados en todo su proceso. Es por esta razón por la que queremos recoger a continuación algunas de las tecnologías que consideramos podrían tener un mayor impacto en el flujo de trabajo de una audiodescripción junto con las investigaciones y conclusiones que se han hecho en el pasado, si se está investigando actualmente y nombrar algunos ejemplos de herramientas que se podrían utilizar en cada caso.

## 5.1 Traducción automática

El uso de la traducción automática (MT) es una de las prácticas que más habituales se han vuelto en el campo de la traducción, ya que se ha demostrado en varias investigaciones (Guerberof 2009; Plitt & Masselot 2010, Hassan et al., 2018) el incremento en la productividad al usar MT, aunque este incremento varía mucho dependiendo de la temática y los pares de idiomas y debemos tener en cuenta que siempre necesitaremos la revisión humana para captar todos los matices. La traducción automática con posesición (PE) forma ya parte del flujo de trabajo de muchos ámbitos de la traducción, aunque “[t]he adoption rate of MT and PE processes naturally varies in different countries and language pairs” (Koponen, 2015: 3) al igual que en las diferentes áreas de la traducción. En el campo de la traducción audiovisual (AVT), que abarca la audiodescripción, la MT se incorpora sobre todo en los subtítulos y no está tan normalizado usar traducción automática como en otras especialidades. En los últimos años, la Unión Europea ha financiado algunos proyectos que abordan la generación automática de subtítulos y su traducción como: MUSA (2002-2004), EU-BRIDGE (2012-2014) o EMMA (2014-2016).

Como resultado, se están llevando a cabo numerosas nuevas investigaciones respecto al uso de la MT y posesición en la traducción de guiones audiodescritos. Uno de ellos fue el proyecto ALST, en el cual Anna Matamala estudia la viabilidad de utilizar traducción automática con posesición (MTPE) junto con herramientas del habla con el fin de semiautomatizar el proceso de AD. En el siguiente apartado hablaremos más detenidamente sobre esta investigación.

El paso de implementar la MTPE en el proceso de la AD es parte de la utilización cada vez más común de las nuevas tecnologías, aunque por lo que sabemos todavía no se ha aplicado en ninguna situación real fuera de los estudios académicos. Pese a que la MTPE es un escenario que se ha contemplado como posible solución para acelerar el proceso de AD, existen opiniones diversas en este aspecto. Matamala (2006), Bourne y Jiménez (2007) y Jankowska (2015) defienden este panorama. Bourne y Jiménez (2007: 176) afirman que “translating ADs would seem to offer considerable advantage in terms of time and therefore cost in comparison with the present practice by which ADs are written from scratch by professional audio describers in different languages”. Por otro lado, Hyks (2005: 8) opina que “translating and rewording can sometimes take as long if not longer than starting from scratch” y que “the fact that some languages use many more or sometimes fewer words to express an idea, can drastically affect timings”. La opinión de Rodríguez y Sánchez (2008: 8-16) es mucho más rotunda, puesto que hablan sobre la falta de respeto hacia las personas con discapacidad visual al usar MT, sobre “putting a foreign culture before the Spanish (or the Spanish blind people’s) culture” y

defienden que la traducción de guiones supondría un coste mucho mayor, ya que no solo tendrían que contratar a un traductor, sino que también a un dialoguista.

Remael y Vercauteren (2010) analizan los retos que supondrían traducir guiones y señalan principalmente dos dificultades: la frecuencia y uso específico de determinadas formas gramaticales, y la traducción de referencias culturales debido a la distancia entre el público del guion original y el público objetivo de la traducción. Sin embargo, concluyen que la traducción de la audiodescripción aumentará en el futuro debido a la rentabilidad que supondría. Destacamos también la siguiente cita de Fernández-Torné y Matamala (2016: 80) ya que subraya la importancia de no solo tener en cuenta los resultados cuantitativos: “new technological solutions cannot only be measured in terms of time or productivity, and this also applies to a possible implementation of machine translation in the audio description field”.

Asimismo, es relevante considerar lo que comentábamos en el apartado 3.1 acerca de cómo cada país describe sus guiones de diferente manera y sobre si los espectadores aceptarían o no la traducción de un guion extranjero. En este aspecto, Matamala (2016: 281) ya incentivó los estudios de recepción y expresó su preocupación por las opiniones de los usuarios: “what would users perceive if an AD had been created, had been translated or had been post-edited? Would they have any preference?”. Estel·la Oncins (2022: 453) concuerda que necesitamos más estudios de recepción para conocer si los espectadores aceptarían o no esta clase de AD, pero señala que los resultados iniciales demuestran que el uso de MTPE permitirían aumentar la cantidad de contenido audiodescrito. Podría ser interesante implementarlo en las redes sociales o en el contenido generado por los usuarios ya que son plataformas que están acostumbradas al uso de tecnologías automáticas (por ejemplo, en la red social TikTok puedes generar subtítulos automáticamente usando herramientas de *speech-to-text*) y las expectativas en la calidad no suelen ser tan altas como en otros productos audiovisuales.

En la mayoría de las investigaciones, como las llevadas a cabo por Ortiz-Boix y Matamala (2016) que estudian la aplicación de la MT en dos idiomas cercanos como es el de castellano y el catalán, Fernández-Torné y Matamala (2016) que estudian el uso de la MT del inglés al catalán, Fernández-Torné (2016b) que también lo realiza con la combinación de inglés a catalán o el propio proyecto ALST de Anna Matamala (2016) que analiza la aplicación de herramientas del habla y compara tres escenarios: creación de la AD, traducción humana de la AD y MTPE de la AD (inglés a catalán), se realizan experimentos para determinar que la MTPE en la AD ofrece resultados lo suficientemente positivos como para decantarnos por esta práctica. Sin embargo, en estos estudios, los resultados no son lo suficientemente significativos en términos de *temporal*, *technical* o *cognitive effort*, o se concluye que la muestra utilizada no era lo suficientemente amplia y se necesita mucha más investigación. Además, en numerosos

estudios se comenta la necesidad de contar con un corpus o un traductor automático adaptado a la gramática y léxico que se utiliza en la audiodescripción para que las traducciones estuvieran más enfocadas en este campo. Asimismo, García (2011: 218) afirma que los motores de traducción automática que “are trained with domain specific memories and glossaries, and work on texts that have been pre-edited following controlled language guidelines” y Groves (en Fernández-Torné, 2016b: 8) afirma que “[t]he quality of MT is highly dependent on the quality of the data used for training”. Más adelante, abordaremos con mayor detalle la aplicación de los corpus en la AD.

En lo que concierne a la MTPE en el proyecto de ALST, los resultados mostraron que el *temporal effort* era prácticamente el mismo en las tres prácticas, el *technical effort* es más alto tanto en la creación de AD como en la traducción de AD y el *cognitive effort* es bastante difícil de medir pero es más alto en la creación de la AD comparado con las demás labores (Matamala, 2016: 279). Generalmente el enfoque que se usa para evaluar el esfuerzo de posesición es la propuesta de Krings (2001: 179) que diferencia entre: *temporal effort* que es tiempo total que conlleva la posesición, *technical effort* que se refiere a las acciones realizadas para poseer el texto (por ejemplo, los movimientos del ratón, los clics o las pulsaciones del teclado) y *cognitive effort* que “is the extent and type of cognitive processes that must be activated to remedy a deficiency in the MT output”. Por otro lado, existen otras métricas que se utilizan más habitualmente para evaluar la calidad de las traducciones automáticas con o sin posesición en comparación con la traducción humana:

- *Human-Bilingual Evaluation Understudy (HBLEU)*: mide la cercanía entre la traducción automática y la versión poseída del texto para comprobar que el motor de traducción automática es el mejor.
- *PE time*: mide el tiempo invertido en la posesición.
- *Human-targeted Translation Edit Rate (HTER)*: calcula la distancia de edición entre la traducción y la versión poseída.

En conclusión, la traducción automática en el campo de la audiodescripción se encuentra en constante evolución y presenta, como hemos podido ver a lo largo de este apartado, tanto desafíos como oportunidades. Es fundamental seguir investigando y desarrollando nuevas herramientas y flujos de trabajo hasta encontrar aquel que tenga en cuenta por igual las necesidades de todos los implicados en este proceso, tanto desde el punto de vista profesional como el del espectador, es por ello por lo que tenemos que considerar tanto los aspectos cuantitativos como los cualitativos de las investigaciones. En el siguiente apartado hablaremos de las herramientas del habla que es también un ámbito que lleva desarrollándose

desde hace algunos años y que, aunque todavía necesita más investigación en algunos aspectos, ya se utiliza en algunas audiodescripciones.

## 5.2 Herramientas del habla

Uno de los usos más significativos de los sistemas de síntesis de voz o *text-to-speech* (TTS) es el de ayuda a la lectura para las personas con discapacidad visual y es por ello que se desarrollaron los primeros sistemas con este público en mente (Fernández-Torné, 2016a: 28). En 1976, Raymond Kurzweil inventó la *Kurzweil Reading Machine* que se trataba de un sistema de ayuda para leer en voz alta el texto escrito a través de un escáner óptico (Lemmetty, 1999: 8). Sin embargo, este sistema era demasiado costoso y la gente no se lo podía permitir por lo que solo podías encontrarlo en bibliotecas y centros dedicados a la gente con discapacidad visual (Lemmetty, 1999: 9). Actualmente estos sistemas de lectura en voz alta se utilizan en muchos campos: ayudas para la circulación (sistemas de navegación GPS), herramientas para la educación (audiolibros), entretenimiento (audiosubtitulado) y comunicación (lectores de pantalla de *software* en los ordenadores) e incluso algunos traductores lo utilizan en su trabajo (Fernández & Matamala, 2015: 62). Todos estos usos no están solo dirigidos a las personas con discapacidad visual sino también a aquellas con discapacidades cognitivas, dificultades para leer o cualquier persona que quiera utilizarlas. Algunos ejemplos de empresas que se dedican a las tecnologías del habla son: Verbio, Loquendo, Nuance Vocalizer o CereProc (Alías et al., 2011: 111). Otros sistemas de síntesis de voz son eSpeak que es de *software* libre y JAWS (*Job Access With Speech*) que está dirigido especialmente a las personas ciegas.

Como podemos observar en la siguiente figura, las herramientas de TTS funcionan “leyendo” el texto a partir del procesamiento de lenguaje natural (PLN) que relaciona los fonemas y la prosodia de los sonidos que se van a producir de esas palabras. En segundo lugar, pasa a un procesamiento digital de la señal (PDS) que se encarga de generar el habla sintética a partir de los datos de la fase anterior (Alías et al., 2011: 109).

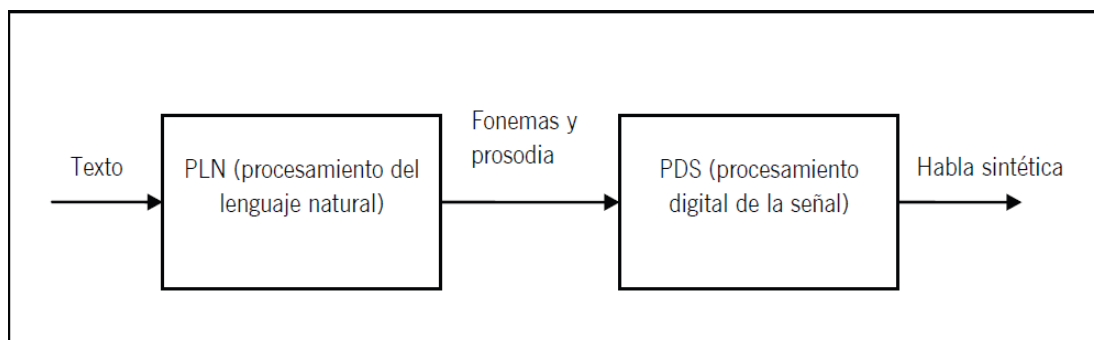


Figura 3: modelo de un sistema de TTS (Alías et al., 2011: 110)

Queremos hablar brevemente sobre el audiosubtitulado (AST) ya que utiliza las voces sintéticas de los sistemas de síntesis de voz para emitir en voz alta estos subtítulos y es un buen ejemplo de los beneficios de estos sistemas. Definido por Aline Remael (2012) como “the spoken rendering of the written (projected) subtitles or surtitles with a filmed or lived performance”. Este tipo de servicio se ha implementado en cadenas de televisión de diferentes países como los Países Bajos, Suecia, Dinamarca o Finlandia (Iturregui-Gallardo & Walczak, 2022: 368). Además, en 2013 la televisión pública de Cataluña también empezó a ofrecer TTS AST en sus programas de noticias de la tarde y de la noche para leer en voz alta los subtítulos que se generaban de las intervenciones.

Otro campo de investigación bastante reciente es las *audio introductions* (AI). El proyecto europeo ADLAB las define como:

a continuous piece of prose, providing factual and visual information about an audiovisual product, such as a film or theatre performance, that serves as a framework for blind and visually impaired patrons to (better) understand and appreciate a given ST [source text]. AIs can be used in combination with an AD or as stand-alone pieces and can be recorded and made available beforehand (for instance via a website) or can be delivered live, as is often the case in the theatre. The recorded AI can feature a single voice or can include a combination of voices and sound bites.

Estas AIs suelen incluir información que es relevante para el espectador, pero que no pueden añadirse a la AD debido a la falta de tiempo. Se han probado sobre todo y se ponen en práctica en eventos en directo y no es todavía tan popular en el ámbito audiovisual. Por ejemplo, en los eventos en directo, pueden incorporar información sobre las escenas musicales, entrevistas con los miembros de reparto, información relevante del programa como la duración o los créditos, o incluso descripciones más detalladas del decorado, el vestuario o los personajes (Romero-Fresco, 2022: 424). No hay muchas investigaciones de recepción sobre cómo deberían construirse o qué información deberían tener estas *audio introductions* lo que ofrece la oportunidad a los audiodescriptores a tener más libertad creativa. Hasta el momento, no se han realizado estudios de recepción sobre el uso de voces artificiales en las AIs, por lo que también podría ser una buena opción dedicarle atención en futuras investigaciones para evaluar cómo los usuarios perciben estas voces.

Por otro lado, también tenemos otra herramienta del habla que son los sistemas de reconocimiento del habla (*Automatic Speech Recognition* [ASR]) o *speech-to-text* (STT) que consiste en un *software* que reconoce el lenguaje oral y funciona comparando las muestras de las palabras con las almacenadas en su memoria (Raud & Brennan, 2000). Fue durante la década de los 50 que empezaron a desarrollarse estos sistemas de reconocimiento del habla. Al principio, reconocía el orden de 10 palabras, a un solo interlocutor y disponía de un vocabulario limitado (Fuentes Bueno et al., 2007: 135). Actualmente, la mayoría de los sistemas se basan en

los modelos ocultos de Markov, en redes neuronales, o bien en ambas a la vez. Un sistema de reconocimiento del habla está compuesto por: modelo acústico, gramática y vocabulario y modelo lingüístico (Romero-Fresco, 2011). En otras palabras, funciona de manera muy parecida a un sistema de traducción automática estadística combinando los distintos modelos. Este sistema de reconocimiento del habla representa matemáticamente el audio añadiendo la gramática y el vocabulario como un listado de palabras para reconocer la equivalencia más correcta que ha utilizado el usuario y, más tarde, a través del modelo de lengua calcula la probabilidad del uso de una palabra analizando la que se ha utilizado antes y la siguiente (Romero-Fresco, 2011).

Entre los sistemas más destacados de reconocimiento del habla se encuentra *Dragon* que se publicó por primera vez en 1997 y actualmente está disponible en ocho idiomas y va por la versión 16. No solo incluye funciones de *speech-to-text* sino también puedes personalizar tus propios comandos y dictárselos, y tiene sistemas de síntesis de voz. Además, tiene paquetes especializados en medicina y derecho: *Dragon for Medical* y *Dragon for Law Enforcement*. Otro sistema de reconocimiento del habla bastante conocido es *Windows Speech Recognition* en el que Microsoft lleva trabajando desde 1993 y se incorporó por primera vez en 2006 en Windows Vista. Además, con este sistema de Windows también puedes usar la herramienta de *text-to-speech* en 49 idiomas.

Una vez que hemos introducido la importancia y cómo funcionan estas herramientas del habla, queremos hablar sobre cómo puede influir a la hora de automatizar el proceso de audiodescripción.

Como decíamos anteriormente, con el fin de reducir los costes y de ofrecer más contenido accesible se ha investigado y se sigue investigando la viabilidad de incluir técnicas automáticas en el flujo de trabajo de la AD. Algunas de las soluciones que se investigan: traducción humana del guion audiodescrito, traducción automática más posesición del guion audiodescrito y sistemas de *text-to-speech* que lean el guion como alternativa a las voces humanas. Vamos a poner ahora algunos ejemplos de estas investigaciones.

En Polonia, Magdalena Mączyńska en la universidad de Varsovia en 2011 analizó el uso y la aceptación por parte de 54 participantes con discapacidad visual y 30 personas videntes del audiosubtitulado sintético junto con la herramienta TTS en AD en una película de no ficción. Este estudio formaba parte de un proyecto mayor desarrollado por la universidad de Varsovia en el que querían analizar y evaluar la recepción de *text-to-speech audiodescription* (de ahora en adelante TTS AD) entre las personas con discapacidad visual de diferentes edades y en diferentes productos audiovisuales. Se concluyó que la mayoría de los espectadores aceptan esta TTS AD (Mączyńska, 2011).



En la Universitat Autònoma de Barcelona, la investigación de Carla Ortiz-Boix en 2012 exploró la aceptación de TTS en AD en castellano con diferentes voces sintéticas de diferentes herramientas (Festival, Acapela, Verbio...) y humanas en 18 participantes videntes. Los resultados mostraron que los participantes se mostraron más reacios a aceptar algunas voces sintéticas como la masculina de Festival mientras que la de Acapela y Verbio sí fueron aceptadas (Ortiz-Boix, 2012).

Una tesis de la Universitat Autònoma de Barcelona es la investigación de Anna Fernández-Torné en 2016 en la que estudia la implementación y aceptación de voces sintéticas en la AD en catalán en comparación con las voces humanas. También explora la posibilidad de automatizar aún más el proceso por medio de la traducción automática y la posesión de los guiones audiodescritos. Los resultados de la aceptación de las voces sintéticas por parte de los participantes se acercaban bastante a los resultados de las voces humanas, aunque sin llegar a superarlos.

Los datos cuantitativos de este trabajo revelan que el 81 % de los participantes prefieren la voz humana, el 1 % prefiere la voz sintética, el 3 % dijeron que depende del producto audiovisual y el 15 % declararon que no tenían ninguna preferencia mientras que la voz artificial sonara lo suficiente natural y no fuera muy extenuante (Fernández-Torné, 2016a: 193-194). Además, también se les preguntó si consideraban las voces sintéticas en la AD como una alternativa a las voces humanas y el 94 % de los participantes confirmaron que lo aceptarían. No obstante, hay que puntualizar las razones de la aceptación de estos participantes y es que el 33 % aceptaría porque incrementaría el número de productos audiovisuales audiodescritos, el 15 % de ellos dijeron que lo aceptarían porque TTS AD era una mejor opción que no tener ninguna audiodescripción y otro 13 % señalaron que solo debería considerarse una alternativa y no la situación habitual (Fernández-Torné, 2016a: 193-194). Consideramos que estas últimas puntualizaciones son de especial relevancia, ya que, si bien el 94 % los participantes concuerdan en que las voces son lo suficientemente aceptables, su aceptación no se debe a que prefieran estas voces sintéticas a las humanas, sino más bien porque no se les ofrece suficientemente contenido accesible.

Por otra parte, los participantes tampoco quieren que las empresas se acostumbren a esta opción de voces sintéticas porque sea más asequible, debido a que esto podría llevar a descuidar el uso de las voces humanas que resultan más naturales. Es importante destacar que la aceptación por parte de los espectadores del uso de las voces sintéticas puede variar dependiendo de la temática del producto audiovisual, ya que influiría directamente en su disfrute y la probabilidad de que se consuman dichos productos. En esta misma tesis de Fernández-Torné pregunta de manera general a los participantes donde aplicarían la herramienta

de *text-to-speech* y la mayoría lo incluiría en documentales, series de televisión y películas. Haría falta más investigaciones para precisar más en qué temáticas específicamente dentro del ámbito audiovisual seguiría siendo aceptable este uso del TTS en la AD y si también depende de la población y el idioma, y existe este mismo consenso en otros países.

Al igual que la traducción automática, las herramientas del habla también utilizan escalas para medir el esfuerzo y la calidad de estas herramientas que son la *Diarization Error Rate* (DER) y *Word Error Rate* (WER). DER es la métrica más habitual a la hora de analizar la calidad de *speaker diarization* que consiste en segmentar el discurso en fragmentos homogéneos asignándoles un identificador acorde con el interlocutor al que pertenece cada intervención en una conversación. DER es la suma de tres tipos diferentes de errores: *miss speech time* que es el porcentaje de discurso que el sistema no ha detectado, *false alarm speech time* que es el porcentaje de discurso que el sistema detecta pero no está en el original y *speaker error time* que es el porcentaje de tiempo que el sistema asigna erróneamente a un interlocutor (Héctor Delgado et al., 2015: 318). Mientras tanto, WER es la métrica más común para analizar la calidad del rendimiento del reconocimiento de voz. WER se define como la suma de todos los errores posibles (inserciones, eliminaciones y sustituciones) entre el número de palabras en el archivo original (Héctor Delgado et al., 2015: 319). Sin embargo, también hemos observado que otras investigaciones han utilizado una serie de rasgos en las voces que los propios receptores han ido mencionado a lo largo de otros estudios como guía para comparar, por ejemplo, la voz sintética masculina o femenina y la voz natural masculina o femenina. Esta serie de características son: *accentuation, acceptance, intonation, listening effort, speech pauses, pleasantness, naturalness, overall impression, narration speed* o *pronunciation*.

Es relevante señalar que la aceptación de las voces sintéticas está estrechamente relacionada con su uso habitual en la vida cotidiana y con la experiencia del usuario, ya que numerosos estudios demuestran que las personas suelen “acostumbrarse” a estas voces (Cryer & Home, 2008: 4; Szarkowska & Jankowska, 2012; Walczak & Szarkowska, 2012). En una investigación realizada en 1993 por Llisterra, Fernández, Gudayol Poyatos y Martí se observó como los usuarios que usaban lectores de pantalla por medio de la síntesis del habla determinaron beneficios como el acceso más rápido a la información en comparación con esperar a que estuviera en braille y una mayor confidencialidad ya que no tenían que pedirle a nadie que lo leyera por ellos. Así lo afirma también Luis García (2004) que habla sobre las voces sintéticas como un medio para acceder a la información sin necesidad de depender de nadie.

El proceso de crear TTS AD ya fue descrito por Szarkowska y Jankowska (2012: 86):

first an AD script is written and then, using subtitling software, it is synchronised with the film. Put simply, synchronising the AD script with a film means preparing a list of AD chunks in the form of subtitles, each consisting of text and time codes. From the technical point of view, the process differs from preparing traditional subtitles only by the fact that the text is to appear in the gaps between the dialogues and not simultaneously with them. Later on, the text file is read by speech synthesis software while the film is played on a multimedia player, or the AD script can be recorded and mixed with the multimedia file.

A continuación, para contrastar este proceso con otro aún más automatizado vamos a abordar el proyecto ALST dirigido por Anna Matamala de la Universitat Autònoma de Barcelona y financiado por el Ministerio de Economía y Competitividad en 2015. En este proyecto Anna Matamala propone un proceso casi totalmente automatizado para la AD y que podemos observar en la siguiente figura:

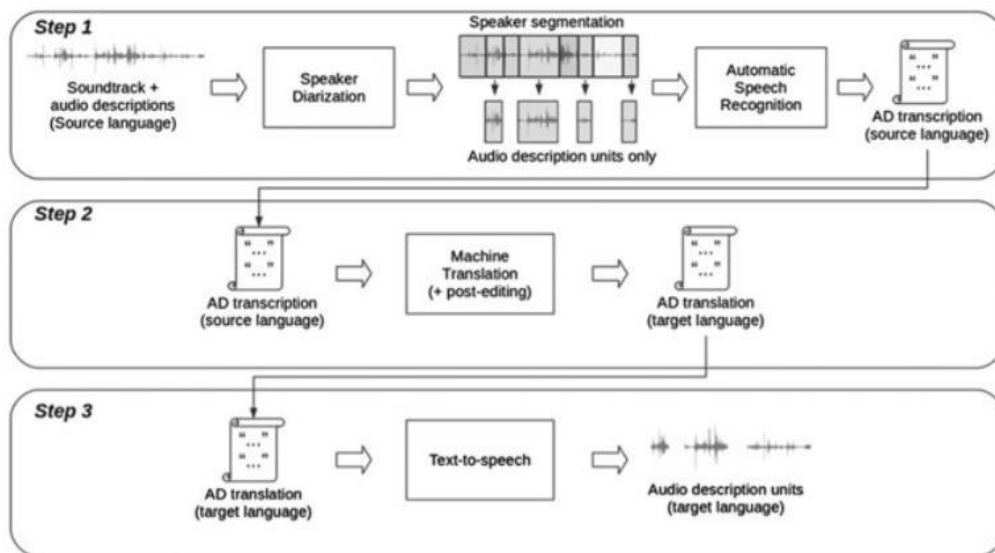


Figura 4: flujo de trabajo propuesto por el proyecto ALST (2015)

Este flujo de trabajo se dividiría en tres pasos principales:

- Paso 1: *speaker diarization* seguida de una segmentación del audio de cada participante y finalizará con la transcripción del guion audiodescrito original a través de la herramienta de *speech-to-text*.
- Paso 2: traducción automática seguida de una posesición humana de este guion audiodescrito.
- Paso 3: usar la herramienta *text-to-speech* para dar voz a la AD de este guion.

Este proceso sería semiautomático ya que el factor humano sería todavía necesario para revisar los resultados automáticos y la posesición de la traducción automática del paso 1 y 2 (Héctor Delgado et al., 2015: 311). Ya no solo en lo que se refiere a la calidad del guion audiodescrito, sino también a las condiciones acústicas, palabras en otro idioma que no sea el

idioma objetivo, los acentos de los interlocutores y la entonación, por mencionar algunos aspectos, pueden influir significativamente en un proceso tan automatizado. Por ejemplo, los nombres ingleses seguramente se vean alterados a la hora de transcribirlos a la ortografía catalana ya que el programa estará siguiendo las reglas fonéticas del catalán y escribirá “Queit” en vez de “Kate” (Iturregui-Gallardo, 2019). Este flujo semiautomático podría llegar a ser bastante relevante en contextos como el de las redes sociales donde las expectativas sobre la calidad no suelen ser tan altas y así hacerlas más accesibles.

En una investigación publicada el año pasado, en 2022, por Sawako Nakajima y Kazutaka Mitobe llevaron a cabo un experimento en el que desarrollaban un *software* para evaluar si era posible reducir los costos del proceso de crear un guion audiodescrito con TTS, al tiempo que intentaban mantener la calidad de las voces humanas. Sin embargo, los resultados indicaron que no se consiguieron reducir los costos y la calidad de la AD producida con TTS era de aproximadamente un 80 % en comparación con las voces humanas, incluso después de utilizar dicho *software*.

Además, se han desarrollado aplicaciones móviles que ofrecen un acceso más fácil a la AD. En la Universitat Autònoma de Barcelona (UAB), cabe destacar la labor de Estel·la Oncins et al. (2013) en el desarrollo de *Universal Access System* (UAS). Esta aplicación es multilingüe y compatible con varias plataformas, y permite hacer accesible los eventos en vivo. UAS ofrece audiodescripciones automáticas a partir de TTS y otras funciones como *spoken subtitles* a partir de TTS. Otra aplicación parecida es la de *AudescMobile* de la ONCE que permite la búsqueda, descarga y reproducción de audiodescripciones.

En resumen, las herramientas del habla tienen bastante potencial a la hora de mejorar la eficiencia y la accesibilidad en la audiodescripción. Sin embargo, hay que tener en cuenta las limitaciones de precisión y comprensión que todavía pueden presentarse en estas herramientas y seguir investigando para mejorarlas. No podemos olvidar que, a pesar de los avances tecnológicos, es crucial contar también con la intervención de los profesionales de la audiodescripción, así como tener en cuenta siempre la opinión de los usuarios para conocer dónde serían aceptables exactamente el uso de estas herramientas. Anna Matamala (2016: 281) ya considera este aspecto y señala que sería interesante analizar y comparar el impacto en la implicación emocional de los usuarios “because watching a film is not only about listening to a comprehensible and enjoyable voice but also about understanding and being immersed in a story”. Por esta razón, la investigación y el desarrollo en este campo es fundamental para perfeccionar el flujo de trabajo de los guiones audiodescritos y poder ofrecer más cantidad de productos inclusivos sin perder calidad.

### 5.3 Corpus

De cara al aumento en la demanda de productos accesibles, las fechas de entrega ajustadas y la necesidad de mantener o incluso aumentar la calidad de las audiodescripciones, es importante considerar la posibilidad de implementar tecnologías que puedan ayudarnos a mejorar el proceso. En este apartado, abordaremos otro recurso que podría llegar a ser de gran utilidad: los corpus lingüísticos. Estos corpus se definen como “a collection of spoken or written texts to be used for linguistic analysis and based on a specific set of design criteria influenced by its purpose and scope” (Weisser, 2016: 13). Es decir, los corpus nos permiten recabar información sobre los aspectos lingüísticos del lenguaje para su estudio lo que nos ayudaría a comprender mejor el lenguaje y las características idiosincráticas de la AD, y facilitaría el desarrollo de herramientas y recursos como diccionarios, bases de datos o incluso para entrenar los modelos de los traductores automáticos para que funcionaran de manera más fiel al lenguaje de la audiodescripción.

Los audiodescriptores evitan interpretar la imagen para aquellos que no pueden verla y centran sus descripciones en los aspectos visuales con el enfoque WYSIWYS. Este enfoque favorece la información visual sobre la narrativa, por lo que la AD presenta propiedades lingüísticas particulares que no podemos encontrar en otros textos (Fresno, 2022: 313). Describir todas las imágenes es una tarea imposible, no solo por la complejidad que ya presenta describir toda una imagen sino también por las limitaciones de espacio y tiempo que presenta cualquier AD. Por esta razón, es de vital importancia saber elegir tantos los elementos visuales que puedan ser más relevantes para un espectador que no ve la imagen como las palabras más adecuadas a la hora de crear una AD. Otro elemento que influye notablemente a la hora de audiodescribir son las guías y normas que establecen unas pautas para alcanzar un estándar de calidad y promover la uniformidad de todas las AD.

Según Salway (2007: 151) “the relationships between the visual and the verbal, between vision and language and between image and text have fascinated scholars in many disciplines for many centuries”. En los últimos años ha habido multitud de investigaciones que estudiaban diferentes aspectos de la AD como las preferencias del público respecto a la voz en la AD (Szarkowska, 2011; Fernández-Torné & Matamala, 2015), el impacto que tiene en la comprensión de los espectadores si se modifica la entonación, la velocidad o la explicitación de la información (Cabeza-Cáceres, 2013) o la repercusión que tiene en el recuerdo de las personas al modificar la cantidad de información y su segmentación (Fresno, 2014). Sin embargo, la investigación experimental implementado elementos tecnológicos en el proceso de creación de AD no es tan amplia y se centra sobre todo en la traducción automática o en las herramientas del habla.

Las investigaciones que analizan los patrones lingüísticos de la AD mediante la metodología del corpus son escasas, pero podemos destacar los siguientes proyectos de investigación: Television in Words (TIWO) de Andrew Salway de la Universidad de Surrey en 2007, TRACCE de Catalina Jiménez y Claudia Seibel de la Universidad de Granada en 2012 y Visual into Words (VIW) de Anna Matamala de la Universitat Autònoma de Barcelona en 2019. Queremos destacar que este último proyecto fue financiado por el Ministerio de Economía y Competitividad de España. Todos estos proyectos concluyen que los guiones de audiodescripción presentan una serie de características lingüísticas distintivas que se reflejan en su propia sintaxis, semántica y léxico. Si seguimos esta línea de pensamiento, en la cual la audiodescripción presenta su propio lenguaje, o *special language* como mencionan en estas investigaciones, y tiene una manera específica de expresarse, resulta de gran relevancia desarrollar más esta idea y explorar la viabilidad de crear un corpus que pueda brindar apoyo a los audiodescriptores y contribuir a estandarizar este lenguaje. A continuación, vamos a explicar brevemente las dificultades de crear un corpus, los aspectos más relevantes que debemos tener en cuenta y algunos resultados que se fueron encontrando en estas investigaciones (TIWO, TRACCE y VIW).

Con el objetivo de contar con una base sólida para la investigación de textos multimodales y para poder examinar cómo estos textos construyen significado, necesitamos disponer de una gran cantidad de datos, es decir, un corpus representativo o una colección de texto con millones de palabras (*tokens*) y un método de análisis apropiado (Jiménez Hurtado & Seibel, 2012: 410). Allwood (2008: 208) describe de la siguiente manera un corpus multimodal: “a multimodal digitalized corpus is a computer-based collection of language and communication related material drawing on more than one sensory modality or more than one production modality”. Estos corpus suelen analizarse a través de programas informáticos (por ejemplo, en TRACCE usaron el *software* Taggetti) para extraer información conceptual, léxica y pragmática de los textos. De esta manera, sabremos qué categoría de palabras son las más utilizadas, las palabras más y menos frecuentes, si hay palabras que se usan más en las audiodescripciones que en otros textos generales, si existen concordancias a la hora de describir personajes, paisajes, emociones, interacciones, etc. Tal y como dicen Baños, Bruti y Zanotti (2013: 483) estos corpus nos permiten “capture the distinctive features and patterns of translated texts” y “[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”.

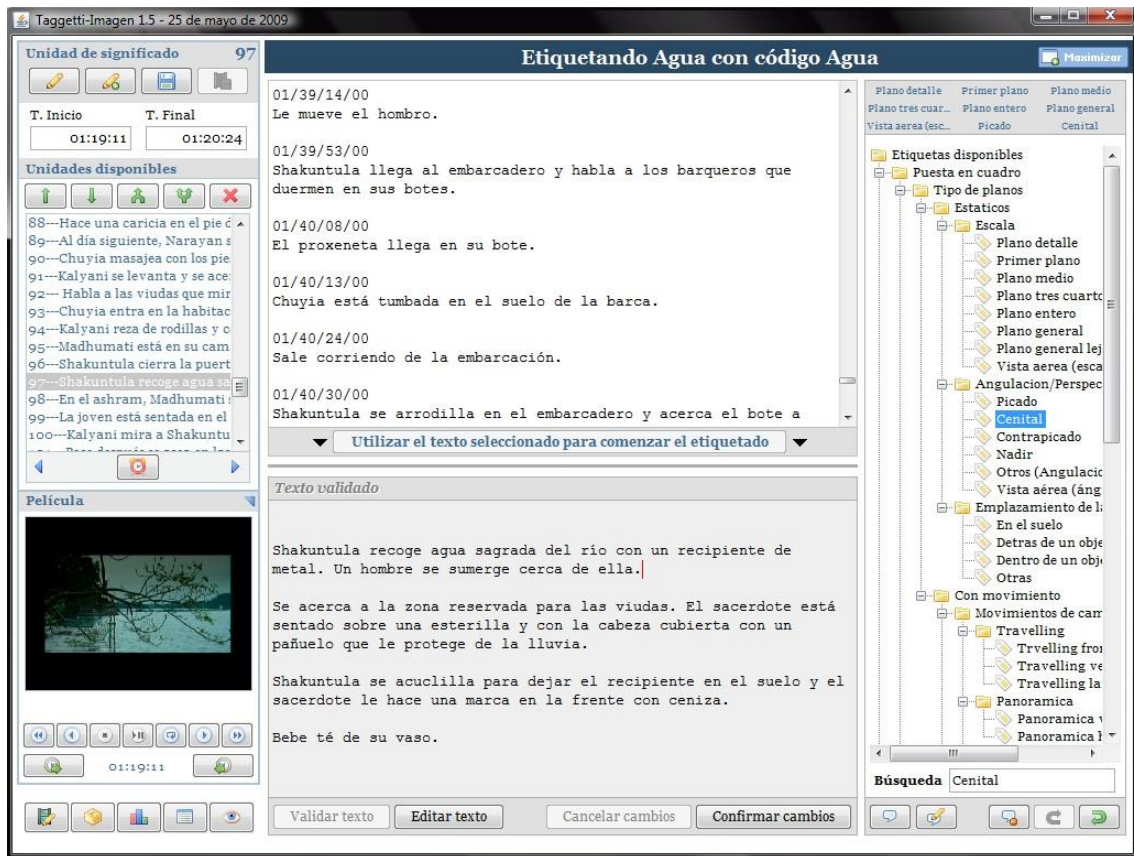


Figura 5: muestra del corpus multimodal del proyecto TRACCE (TRACCE, 2018)

Todas estas investigaciones también concuerdan y destacan que es crucial que los textos que forman parte del corpus estén organizados de tal manera que los datos que el programa extraiga sean fiables y aporten resultados significativos. Esto se traduce en que los textos que aportemos al corpus deben tener una relación entre ellos para ser representativos de uno o varios tipos de texto. La efectividad en los resultados de cualquier corpus depende mucho de la calidad de la recopilación de los datos, la composición del corpus y de cómo se etiqueten los datos una vez se encuentren dentro del corpus. Por ejemplo, respecto a la variedad en la composición del corpus, en la investigación TIWO seleccionaron nueve géneros cinematográficos diferentes mientras que en TRACCE se centraron más en clasificar los textos según el idioma, el género cinematográfico y si cumplía con las regulaciones de AENOR. Una dificultad a la que se enfrentan estas investigaciones son los derechos de autor ya que no pueden usar la mayoría de las películas ni audiodescripciones libremente en sus proyectos. Por esta razón, en el proyecto VIW, Matamala creó su propia película de 14 minutos y formó un corpus de más de 30 000 palabras con diferentes audiodescripciones de profesionales y estudiantes al inglés, castellano y catalán, con el objetivo de que este corpus siempre estuviera disponible (<https://webs.uab.cat/viw/>).

También es de suma importancia etiquetar bien los datos del corpus, ya que esto nos permitirá identificar y clasificar distintas características lingüísticas para gestionar la información más fácilmente. La precisión y coherencia en estas etiquetas es crucial a la hora de obtener unos resultados confiables. Por ejemplo, en el proyecto TRACCE lo dividen en tres tipos de etiquetas diferentes: *film narrative*, *camera language* y *the set of recurring gramatical structures in the ADs* (Jiménez Hurtado & Seibel, 2012: 412-413). Este sistema de etiquetas facilita la identificación y codificación de patrones morfosintácticos, léxico-semánticos y pragmático-discursivos que suelen estar presentes en las traducciones intersemióticas. Además, permite la automatización y sistematización de esta información para su posterior análisis y explotación (Jiménez Hurtado & Seibel, 2012: 413).

Matamala (2019: 519) nos habla de Bednarek (2015) que es uno de los investigadores que está a favor de la creación de corpus multimodales. Sin embargo, el desarrollo de esta clase de corpus sigue estando en sus inicios y plantea muchos retos y dificultades. Y es que tal y como dice Baldry y O'Halloran (2010: 202): “[w]e 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”. Sin embargo, los corpus multimodales presentan una serie de limitaciones debido al tiempo y el esfuerzo que conlleva recopilarlos. Por esta razón, los corpus multimodales se pueden clasificar de la siguiente manera si tenemos en cuenta varios aspectos (Knight, 2010: 394):

- ***design and infrastructure***: qué tipo de datos y cómo se recogen y organizan;
- ***size and scope***: cantidad y variedad de datos recogidos;
- ***naturalness***: grado de autenticidad de los datos, si los datos recogidos son más espontáneos o están estructurados siguiendo un guion;
- ***availability and (re)usability***: acceso libre a los derechos de los datos, si puede publicarse o utilizarse por otros investigadores.

Mediante el uso de corpus multimodales, Andrew Salway (2007: 167-168) plantea una serie de posibilidades que podrían ayudar a en el proceso de audiodescripción. Una de ellas sería añadir al *software* de AD recursos léxicos como terminologías especializadas o un diccionario de sinónimos para ayudar a los audiodescriptores a encontrar la mejor descripción y a no repetirse demasiado. Teniendo en cuenta que un aspecto esencial de las AD son las normas y preferencias de estilo que hay que seguir por parte de las guías, otra alternativa sería añadir una revisión automática del estilo, similar a la función desempeñada por un verificador de calidad (*QA checker*) en una herramienta TAO. Esto permitiría a los audiodescriptores noveles recibir *feedback* de sus descripciones, a ayudar a los editores o servir como una herramienta para las



organizaciones, como RNIB, para supervisar que se cumplen los requisitos de calidad de las AD (Salway, 2007: 167-168).

En conclusión, la audiodescripción presenta un idioma especializado con características idiosincrásicas que no se pueden encontrar en otros textos. Estos rasgos se pueden explicar si tenemos en cuenta las necesidades comunicativas de los usuarios, de las restricciones de tiempo y espacio y de la serie de normas que hay que cumplir para obtener una AD de calidad. Por lo tanto, estos corpus multimodales desempeñarían una función crucial en el avance y mejora de la AD, ya que proporcionaría una base para el análisis y comprensión del lenguaje utilizado que ayudaría a homogenizar las AD y a alcanzar un estándar de calidad más estable. Además, contribuiría a facilitar la labor de los audiodescriptores y a impulsar la investigación y el desarrollo de otras nuevas tecnologías.

Otra posible implementación tecnológica que está intrínsecamente conectada a los corpus es la descripción automática de imágenes y vídeos en forma de subtítulos (más conocido en inglés como: *image and video captioning*), ya que requiere un corpus como base para generar descripciones. Si se dispusiera de un corpus multimodal específico de la audiodescripción, se podría utilizar como referencia para generar descripciones automáticas, ya que permitiría mejorar directamente la calidad y eficiencia de estas descripciones, así como generar AD automáticas. Hablaremos del estado actual de esta tecnología y su investigación en el siguiente apartado.

#### **5.4 Image y Video Captioning**

Hemos contextualizado este apartado sobre todo en las investigaciones de Sabine Braun y Kim Starr de la universidad de Surrey en el marco del proyecto *MeMAD (Methods for Managing Audiovisual Data)* financiado por el programa EU *Horizon 2020*. Hemos escogido este proyecto para nuestra investigación sobre *image y video captioning* porque es el que tiene datos más actuales según nuestro conocimiento.

Las investigaciones enfocadas en la automatización de la descripción de imágenes y vídeos para subtítulos (*image and video captioning*) han ganado protagonismo en la actualidad debido a la creación masiva de contenido y porque ha demostrado avances prometedores para el futuro (Krishna et al., 2017; Aafaq et al., 2019). La pregunta más importante que tenemos que plantearnos es hasta qué medida podemos confiar y recurrir a estos métodos tan automatizados para la producción de audiodescripciones con el fin de reducir costes y ampliar la accesibilidad, sin comprometer la calidad que es el aspecto más decisivo a la hora de hablar sobre AD. Queremos puntualizar que, aunque estas descripciones automáticas de imágenes estén pensadas para crear subtítulos, se podrían añadir herramientas de *text-to-speech* para leer estos subtítulos y crear más audiodescripciones. Por esta razón esta herramienta podría suponer un gran cambio

en términos de accesibilidad. Hay que tener en cuenta una serie de cuestiones a la hora de pensar en la puesta en práctica de la descripción automática de imágenes y vídeos (Braun & Starr, 2022: 391):

- La calidad final actual y la ausencia de unos principios que guíen el desarrollo de los procesos plantean una serie de problemas éticos y de calidad que afectan al usuario directamente.
- La creación de descripciones para imágenes en movimiento genera todavía muchos fallos y es un aspecto que sigue desarrollándose, ya que la máquina no es capaz todavía de seguir esa continuidad en la imagen y esa sucesión de hechos.
- Tenemos que contemplar que la AD se trata de una traducción intersemiótica, lo que implica la intervención de elementos visuales, audio como la banda sonora, diálogos, efectos de sonido e incluso la voz del narrador, y, en ocasiones, elementos lingüísticos en la pantalla. Para ofrecer un resultado satisfactorio, el sistema de descripción automático debería saber interpretar todas estas señales. Sin embargo, actualmente, esta capacidad no se ha desarrollado y habría que mejorar el sistema para que pudiera “ver” al mismo tiempo que “escuchar”. Otro aspecto que hay que tener en cuenta si obtenemos los datos de las descripciones para estos sistemas de las AD es que estas descripciones solo se proporcionan en los momentos en los que hay espacio entre diálogos y describen aquello más importante para la comprensión por parte del público, pero no todo lo que está sucediendo en escena ya que también se ayudan del sonido de la película. Por esta razón, la AD no constituye una solución por completo para entrenar a las máquinas para que desarrollen descripciones automáticas de la manera más humana posible, aunque sí una buena ayuda como veremos más adelante.

Los modelos de vision (*vision models*) y los conjuntos de datos de entrenamiento a gran escala (*large-scale training datasets*) son dos componentes fundamentales en la creación de descripciones automáticas.

Los *vision models* son algoritmos y modelos de aprendizaje automático diseñados para procesar y reconocer las imágenes tanto estáticas como en movimiento. Utilizan técnicas de *deep learning* para extraer la información relevante de las imágenes y generar las descripciones basadas en ese análisis. Estos modelos se basan en redes neuronales convolucionales (CNN, por sus siglas en inglés) que se utilizan para el reconocimiento de objetos en imágenes, incluyendo aspectos como la clasificación de objetos, obtención de las características y la identificación de la relación entre objetos (Krizhevsky et al., 2017 en Braun & Starr, 2022: 394). Los modelos actuales de visión utilizan pistas visuales como objetos, personajes, acciones, ubicaciones y expresiones faciales (Braun & Starr, 2022: 393), pero no son capaces de conectar estas pistas

visuales con el audio ni con el resto de pistas visuales, ya que su capacidad para entender la continuidad de la narrativa o la relación entre los diferentes planos es bastante limitada. Estos modelos pueden entender el cambio de plano como el cambio de imagen o de fotograma, pero se centran en crear descripciones lo más acertadas posibles sin comprender muy profundamente cómo se relacionan estas descripciones entre sí. Es por ello que no solo hace falta investigar y mejorar los aspectos lingüísticos de las descripciones automáticas, sino que para que funcione como un todo con resultados satisfactorios hace falta también mejorar aspectos como el reconocimiento facial, integrar la comprensión de audio y diálogo, mejorar el flujo de datos haciendo conexiones entre los personajes, objetos, eventos, etc. y elaborar contextualizaciones mediante la identificación de temáticas, momentos clave de la historia y el vocabulario que se está utilizando (Braun & Starr, 2022: 400).

Por otro lado, los *large-scale training datasets* son conjuntos de datos masivos que se utilizan para entrenar los *vision models* con el objetivo de aprender comportamientos como el reconocimiento de objetos en una imagen y su clasificación e identificación de las relaciones entre ellos (Braun & Starr, 2022: 393). Estos conjuntos de datos contienen miles o millones de imágenes con descripciones humanas. Cada vez se tiene mayor acceso a estos conjuntos de datos como: Microsoft Common Objects in Context (MS COCO), Tumblr GIF (TGIF), Visual Genome, Max Planck Institut für Informatik II Movie Description, etc. Estos bancos de datos suelen crearse a través *crowdsourcing* por trabajadores a los que se les pide que describan de manera espontánea lo que ven en una imagen. Esto se diferencia bastante del enfoque de las audiodescripciones que son hechas por profesionales con una formación. Para que nos hagamos una idea del volumen que controlan estos *training datasets*: MS COCO contiene 330 000 imágenes fijas que incluyen 1,5 millones de objetos y tienen 200 000 imágenes con cinco descripciones cada hechas por diferentes personas (<https://cocodataset.org>).

No hay que olvidar la relevancia que tienen también los corpus en esta herramienta. Los corpus ofrecen una gran cantidad de datos lingüísticos y ayudan a identificar patrones del habla, lo que brinda la oportunidad a las descripciones automáticas a extraer información de estos corpus e imitar estos patrones para generar las descripciones. No obstante, como mencionábamos en el anterior apartado, existe un problema con los derechos de autor de estos corpus ya que son pocos los que están disponibles a todos los públicos. Además, el ser humano aprovecha las capacidades cognitivas para inferir lo que puede o no incluirse explícitamente en la AD debido a las limitaciones de tiempo que este campo presenta y la máquina no solo debería ser capaz de identificar qué tiene que priorizar a la hora de transmitir la información sobre una imagen, sino también saber cuándo incluir esa descripción sin saturar al espectador. Un ejemplo de cómo influye la importancia de saber qué describir y cuándo es la siguiente imagen:



3 (290868)

- i. A grandmother standing next to a child in a kitchen.
- ii. Baby **trying to open** wooden cabinets under the sink.
- iii. A woman and child stand in the kitchen.
- iv. An older woman is standing in the kitchen with a child.
- v. The little girl is **trying hard to open** the cabinets

Figura 6: ejemplo de una descripción automática por MS COCO en el proyecto MeMAD de Sabine Braun y Kim Starr (2019:25)

Como podemos ver en esta imagen, el *large-scale training dataset* MS COCO proporciona una serie de diferentes descripciones para una misma imagen. Podemos observar que, si bien las máquinas son capaces de describir las acciones y los personajes que aparecen en la pantalla hasta cierto punto, no pueden interpretar de manera precisa el lenguaje corporal o las expresiones faciales, comprender las normas sociales, crear una cohesión del contexto de la imagen, identificar a los personajes, priorizar la información más importante, crear correferencias o un vínculo entre los personajes. Además, no pueden escuchar por lo que carecen de la capacidad de entender el diferente tono de los personajes que pueden transmitir emociones relevantes para la historia como entusiasmo, ira, sarcasmo, ironía... En este sentido, los corpus multimodales de audiodescripciones serían de gran ayuda en mejorar la calidad de estos resultados.

Otros problemas que presenta esta automatización de las descripciones es la visión artificial (*computer vision*) y el procesamiento del lenguaje, que están relacionados con las limitaciones que presenta ahora mismo los modelos de visión artificial y la escasez de corpus (Braun & Starr, 2022: 392). Queremos hacer un paréntesis y definir la visión artificial o *computer vision* como:

Campo de la inteligencia artificial que permite a los ordenadores y sistemas extraer información significativa a partir de imágenes digitales, videos y otras entradas visuales, y tomar medidas o realizar recomendaciones en función de esa información. Si la IA permite a los ordenadores pensar, la visión artificial les permite ver, observar y comprender (IBM, 2019).

Para comprender mejor la visión artificial, tenemos que entender que la narración visual (*visual storytelling*) se basa en la identificación de los personajes, objetos y acciones y cómo se relacionan entre sí para determinar qué papel desempeñan en el desarrollo de la historia. La mente humana utiliza una amplia gama de pistas para identificar y etiquetar estos personajes y objetos, mientras que los algoritmos de la visión artificial no disponen actualmente de una

capacidad de deducción compleja. Asimismo, en este momento, los conjuntos de datos de imágenes fijas e imágenes en movimiento y las descripciones humanas que se utilizan en el aprendizaje automático no son lo suficientemente amplias para cubrir esta deficiencia (Braun & Starr, 2019: 14). Otro aspecto importante que tenemos que considerar es que el procesamiento humano de la información combina múltiples señales (oral, visual y lingüística) al mismo tiempo que tiene en cuenta otros elementos como el conocimiento general, la cultura y experiencias anteriores que conectamos con las escenas en movimiento y fotogramas. Como humanos, tenemos la capacidad de integrar todas estas señales y conocimiento y sacar nuestras propias conclusiones para comprender la narración que se está desarrollando en pantalla. Este enfoque multifacético y holístico es un desafío para la descripción automática, ya que en realidad los ordenadores no “ven” las imágenes sino que procesan y analizan un conjunto de píxeles y le asignan un código numérico basado en su color (Braun & Starr, 2022: 395).

Lo cierto es que a la generación automática de descripciones todavía le queda un largo camino y muchos retos por delante, pero puede llegar a suponer un gran avance para aquellas personas con problemas de visión, ya que puede suponer el acceso a mucho más contenido, para los *vision models* y para el procesamiento del lenguaje. Otro planteamiento importante es el aspecto ético que supone las descripciones automáticas, dado que con los resultados que ofrecen ahora mismo no pueden sustituir a la AD humana como servicio para los usuarios con discapacidad visual puesto que no cumplen con los requisitos de la UNE 153020: 2005. Sin embargo, puede haber escenarios donde estas descripciones automáticas sean aceptables cuando el nivel de calidad no sea muy exigente. Es más, en 2017 Microsoft desarrolló una aplicación para Apple llamada *Seeing AI*. Esta aplicación utiliza la inteligencia artificial para describir en voz alta los objetos y personas a través de la cámara. En el siguiente apartado hablaremos de otro posible avance en la audiodescripción a través de la inteligencia artificial que es el seguimiento ocular, más conocido como *eye-tracking*.

## 5.5 Eye-Tracking

Incluso cuando compartimos la experiencia de ver la misma película, cada persona tiene recuerdos e interpretaciones diferentes de lo que ha visto, y es común que algunos detalles sean percibidos por unos y pasen desapercibidos por otros. Y, es que, tal y como se preguntan Pilar Orero y Anna Vilaró (2012: 298): ¿cómo es tan diferente nuestra percepción de las mismas cosas? La combinación de *eye-tracking* junto con los estudios de recepción de la AD podrían ayudar a los audiodescriptores a tomar mejores decisiones teniendo en cuenta en qué se fijan los espectadores y la opinión de los usuarios. Debido a que las guías y recomendaciones sobre cómo crear audiodescripciones se crearon a partir de la experiencia sin seguir ninguna base científica y sin tener mucho en cuenta las preferencias de los usuarios. Sin embargo, últimamente se han hecho muchas investigaciones sobre la recepción de la AD y sobre la

opinión de los usuarios. La tecnología *eye-tracking* podría también aportar datos muy interesantes de cara a estas investigaciones.

La relación entre la investigación de *eye-tracking* y la AD puede parecer contradictoria a primera vista, porque ¿por qué utilizar una tecnología dirigida a las personas sin problemas de visión para analizar algo que está dirigido a las personas con discapacidad visual? Parece incompatible analizar cómo funciona el procesamiento visual cuando en realidad vamos a producir una descripción verbal. No obstante, si encontráramos una serie de pautas, estas podrían reflejarse en la AD cuyo objetivo, al fin y al cabo, es sustituir con descripciones la capacidad visual (Mazur & Chmiel, 2016b: 98).

Queremos hablar brevemente de cómo funciona la visión para comprender mejor el funcionamiento de esta herramienta. En nuestro día a día, sin ser nosotros conscientes de ello, nuestros ojos saltan de un punto a otro y estos movimientos oculares se llaman movimientos sacádicos. Sin embargo, la verdad es que durante estos movimientos no tenemos ninguna percepción visual, sino que el objetivo de nuestra visión es mantener una percepción del espacio y del tiempo (Burr et al., 1996). Por lo tanto, cuando observamos, no estamos en realidad percibiendo todos los detalles de lo que nos rodea, sino que captamos la información más relevante para nosotros dependiendo de muchos factores como nuestras preferencias, experiencia previa, expectativas, estado emocional y otros elementos que mencionaremos a continuación. Por ejemplo, es muy interesante el estudio que ya en 1967 realizó Alfred Yarbus en el que mostró como el movimiento ocular de los usuarios revelaba a qué prestaban atención los participantes dependiendo de la información que se les solicitara:

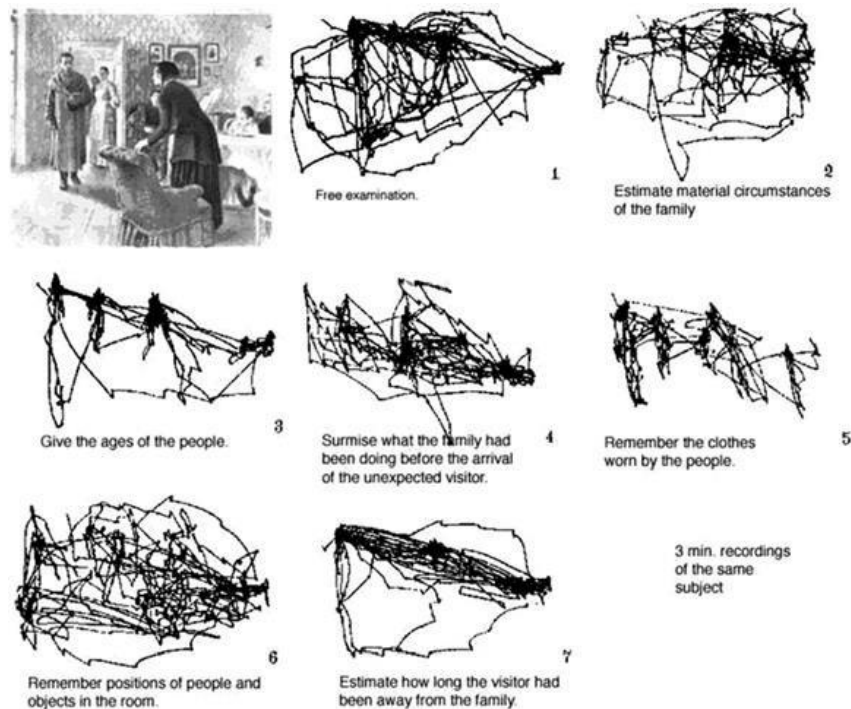


Figura 7: patrones de movimiento ocular (Yarbus, 1967: 190)

En esta misma línea de investigación, se han realizado varias exploraciones en el estudio de las tecnologías de *eye-tracking* para estudiar los patrones oculares de los espectadores y comprobar si al implementar estas preferencias en la AD hay una mejor recepción que con las audiodescripciones que se han realizado hasta ahora.

En su investigación, Pilar Orero y Anna Vilaró (2012) usaron *eye-tracking* para comprobar si los estímulos sonoros alteraban la atención visual. Tras realizar el experimento, concluyeron que esta hipótesis era cierta y que debería considerarse esta variable a la hora de crear audiodescripciones (Orero & Vilaró, 2012: 313-314). En esta misma investigación, las autoras (2012: 299-301) plantean que la mayoría de las normas de la AD coinciden en que no se deberían describir los errores de continuidad, erratas u otros defectos ya que no aportan nada al espectador, aunque ¿qué se debería hacer con los pequeños detalles que forman parte de la caracterización, pero solo se muestran brevemente? Para dar respuesta a esta pregunta, las autoras analizaron un gran corpus de películas audiodescritas y se fijaron en que algunas ocasiones estos detalles habían sido audiodescritos mientras en otras no. Hay que tener en cuenta que también un factor del que dependerá la descripción de estos detalles serán las restricciones de tiempo.

Además de los resultados obtenidos del corpus, con el fin de analizar qué, cuánto y cómo elegir la información que debería describirse realizaron un experimento con la ayuda de *eye-tracking* y un cuestionario complementario, para comprobar si las audiodescripciones de los

detalles coincide con la mirada e intensidad de los espectadores. No obstante, según la teoría de atención encubierta (*covert attention*) de Posner (1980) que alguien esté viendo algo no significa que le esté prestando atención de verdad. Por esta razón, que quede registrado en la herramienta *eye-tracking* dónde estaban mirando los participantes, no significa que exista necesariamente un proceso cognitivo detrás de esa mirada. De todas maneras, esta herramienta sigue pudiendo ser de gran ayuda y podemos constatar la atención de los usuarios a través de otros métodos como cuestionarios para complementar esta herramienta. Al final, concluyeron que no existía correlación entre los resultados del *eye-tracking* con las descripciones de la AD, lo que crea una discordancia entre “the information offered in the AD and the visual clues used by sighted viewers to construct the lineal narrative” (Orero & Vilaró, 2012: 314)



Figura 8: heat map del eye-tracking de la película Match Point (2005) (Orero & Vilaró, 2012: 311)

Di Giovanni (2014) aplicó el *eye-tracking* para comprobar si la AD basada en las preferencias de los espectadores videntes tenía una mayor aceptación por parte de las personas con discapacidad visual. Al finalizar el experimento, los resultados de comprensión eran mejores en la AD que se basaba en el *eye-tracking*. De esta manera, se confirma la importancia que tiene esta herramienta en el proceso de creación de AD. También en el estudio de Chmiel y Mazur (2016b) comparaban la recepción de la AD como se ha hecho hasta ahora y la AD con los cambios del ET y se observó que los usuarios preferían estas últimas. Un ejemplo recogido de la comparación de AD y AD basado en ET es el siguiente de la investigación de Chmiel y Mazur (2016b: 107):



UK-AD	ET-AD
Inside the palace, in a magnificent circular chamber, lit by a huge crystal chandelier, with tall windows hung with blue drapes, with chairs upholstered in blue brocade, footmen in blue livery carry flower arrangements, and then a blue and gold wig box.	Inside the palace. A circular, blue ceiling, a crystal chandelier, gilded stuccowork. All around tall windows with blue drapes. Marie Antoinette enters. She looks around. Footmen bring in flower arrangements.

Figura 9: comparación de la AD tradicional y la AD implementando eye-tracking en la película Marie Antoinette (2006)

Aunque la primera descripción (UK-AD) es una frase larga con una descripción elaborada, no se corresponde del todo con lo que se ve en pantalla (Marie Antoinette entra a la habitación primero y los sirvientes pasan luego llevando flores y objetos personales). La segunda descripción (ET-AD) está más sincronizada con lo que pasa en pantalla y cada frase corresponde, en medida de lo posible, a un cambio de plano porque los distingue como una medida de segmentación (Mazur & Chmiel, 2016b: 107).

Kruger (2012) utilizó otra perspectiva a la hora de comprobar la influencia del *eye-tracking* en la AD. Utilizó esta herramienta para comprobar si los datos de este seguimiento ocular coincidían con las descripciones realizadas por los espectadores, con el objetivo de analizar hasta qué punto el comportamiento visual coincide con la narrativa. Los resultados revelaron que elementos periféricos influyen en la interpretación de la película, mientras otros elementos más notables no lo hacen (Kruger, 2012: 80). Esto demuestra que los datos resultantes del seguimiento ocular deberían utilizarse con precaución en la AD y coincide con los resultados de Orero y Vilaró, ya que el hecho de que los espectadores miren a un determinado elemento no significa que le estén prestando atención. Por esta razón, hay que tener siempre en cuenta la importancia de los elementos en la historia a la hora de tomar la decisión de incluirlos en la AD (Mazur & Chmiel, 2016b: 99).

Después de hacer un repaso por diferentes investigaciones relativas al *eye-tracking*, podemos concluir que los datos recogidos por esta herramienta deberían ser abordados por los audiodescriptores como una fuente adicional de información para ayudar en el proceso de decisión. Además, se debería tener en cuenta los resultados de los estudios sobre la percepción de las personas sin problemas de visión y los estudios de recepción sobre las preferencias de los usuarios para adaptar las AD a las necesidades del público.

## 5.6 Audiodescripciones personalizadas

Lo cierto es que no es posible crear una audiodescripción que satisfaga todas las necesidades de un grupo tan diverso de personas, tanto desde el punto de vista de la edad, preferencias y diferentes grados de visión. Todo esto puede influir a la hora de hacer audiodescripciones de una manera u otra. Uno de los aspectos principales que habría que abordar a la hora de hablar sobre audiodescripciones personalizadas es el contraste entre AD subjetiva y objetiva (WYSIWYS). Esto supuso y supone un gran debate teórico, por lo que Mazur y Chmiel (2016a) llevaron a cabo una investigación empírica y llegaron a la conclusión de que “eliciting subjective opinions of respondents might be inconclusive and that AD reception research should be more focused on the cognitive efficiency of AD”. Al abrir la posibilidad de contemplar diferentes tipos de descripciones ha dado lugar a conceptos como “enriched audio description”(Eardley et al., 2017) and “creative audio description” (Walczak & Fryer, 2017).

Gracias a la tecnología es posible personalizar aspectos de la AD como elegir el tipo de voz sintética que prefieres o ajustar la dirección, la intensidad del canal de audio (Orero, 2022: 413) o acelerar la reproducción de la AD. Respecto a esta última posibilidad, muchas aplicaciones y reproductores permiten reproducir más rápidamente el audio y es una opción que los usuarios con discapacidad visual utilizan usualmente. Esto podría permitir ampliar el contenido de las audiodescripciones si la reproducción fuera más rápida. Sin embargo, habría que hacer ajustes dependiendo del tipo de contenido que se muestre en pantalla y se necesitarían estudios de recepción para saber si el público aceptaría este tipo de AD.

En el futuro, es posible que se puedan realizar otras implementaciones para personalizar las audiodescripciones y adaptarlas mejor a las necesidades y preferencias de cada persona. Por ejemplo, un usuario podría elegir otros tipos de audiodescripciones o pausar la película para acceder a descripciones más detalladas (Salway, 2007: 168). Las pistas de audio podrían utilizar diferentes vocabularios para los distintos grupos de edad o proporcionar diferentes cantidades de información descriptiva que se adaptara a las preferencias personales. Sin embargo, ya resulta bastante difícil que ofrezcan productos accesibles que tengan AD debido a los costes por lo que no podemos crear manualmente varias AD para el mismo contenido, sino que podría utilizarse nuevas tecnologías para adaptar automáticamente lo que sería la descripción “principal”. La idea de Salway (2007: 168) es hacerlo usando por ejemplo XML para explicarle a la máquina el nivel de importancia y las relaciones de interdependencia entre la audiodescripción y los diálogos.

La personalización de las audiodescripciones cuando el concepto de objetividad ha sido el objetivo de este campo durante tanto tiempo y la semiautomatización o automatización del

proceso son dos aspectos que pueden cambiar en el futuro gracias a la investigación y a que el concepto de calidad ha pasado de ser tan normativo a ajustarse más a los gustos personales de cada persona (Orero, 2022: 415). Aún queda por determinar hasta qué punto los usuarios aceptarán estas nuevas implementaciones y si el mercado las promoverá.

## 6. Conclusiones y futuras investigaciones

Como hemos mencionado en varias ocasiones a lo largo de esta investigación y es el tema central de este trabajo, la tecnología está cambiando de manera constante la forma en la que hemos hecho las cosas hasta ahora y esto también se aplica a la traducción y, más concretamente, a la audiodescripción. El objetivo principal de este trabajo era profundizar más en las opciones de automatización que tiene ahora mismo la audiodescripción para saber cuáles están lo suficientemente desarrolladas para dar unos resultados que satisfagan los estándares de calidad de la AD y de la audiencia, además de intentar abrir nuevas líneas de investigación futuras. Consideramos haber cumplido nuestro objetivo a través del análisis de diversas tecnologías y herramientas disponibles que tienen el potencial para agilizar el proceso de producción de audiodescripciones. Asimismo, consideramos que es evidente que la AD en sí puede aprovecharse para mejorar las herramientas multimodales.

Hemos destacado el cambio que podría suponer emplear traducción automática, herramientas del habla, corpus lingüísticos, descripciones automáticas de imagen usando inteligencia artificial y procesamiento del lenguaje natural, *eye-tracking* y audiodescripciones personalizadas. Como hemos podido ver, no todas estas herramientas se encuentran en la misma etapa de desarrollo e investigación. Se han hecho muchos más estudios tanto de recepción como cuantitativos de la aplicación de la traducción automática y herramientas del habla en el proceso de AD. Consideramos que estas herramientas son las más factibles y las que han llegado a un nivel de desarrollo con unos resultados con la suficiente calidad como para poder empezar a implementarlos en el flujo de trabajo de la AD. Sin embargo, primeramente, se deberían crear unas normas que incluyera aspectos como la calidad de los resultados finales de esta automatización, los procedimientos a seguir, si hay alguna especificación sobre qué productos pueden o no tener una audiodescripción más automatizada, si hay que utilizar algún *software* en específico, etc.

No obstante, todavía faltaría por realizar investigaciones sobre la recepción de la traducción automática de las audiodescripciones y el nivel de aceptación que tendrían tanto por parte de los usuarios como por los audiodescriptores. Hemos leído múltiples investigaciones sobre los experimentos que se han realizado para descubrir si el esfuerzo invertido en la MTPE era menor que en la traducción humana y llevaba a unos resultados favorables en la AD, pero no hemos encontrado ningún estudio sobre cómo recibirían estas traducciones los usuarios y si se podrían llevar a la práctica. El uso de MTPE estaría contradiciendo las guías y reglas que existen en cada país a la hora de realizar audiodescripciones, ya que trasladaríamos rasgos del lenguaje y formas de audiodescribir propios de otros países y diferentes a las que está acostumbrado el público normalmente. Y, como comentábamos en apartados anteriores, aunque se ha intentado redactar una guía común a todos los países que haría la posibilidad de la MTPE

mucho más fácil, no se ha podido conseguir por las múltiples diferencias que existen a la hora de describir y a las expectativas del público. Por esta razón, un paso necesario en futuras investigaciones antes de implementar esta automatización en la AD sería recoger las opiniones, preferencias y la aceptabilidad por parte del público. Estas investigaciones podrían concluir que, aunque el uso de la MTPE reduce el esfuerzo y el tiempo dedicado a la tarea, la AD resultante podría ser la que peor recepción tuviera. Otra posible conclusión, como hemos ido viendo en otras investigaciones de recepción de herramientas del habla, es que esta traducción automática tendrá más o menos margen de aceptabilidad dependiendo de los pares de idiomas, las diferencias culturales y la temática. Por lo que antes de seguir investigando más sobre la posibilidad de implementar esta herramienta y si la calidad de los resultados es adecuada, es igual de importante realizar investigaciones sobre la recepción. Otra posibilidad sería que en la posesición el audiodescriptor corrigiera estas diferencias lingüísticas en las descripciones o entrenar al modelo de traducción automática con la guía y reglas que las descripciones españolas deben cumplir para obtener unos resultados más acordes de lo que se espera de una AD.

En este sentido, sería también de gran utilidad los corpus lingüísticos. Lo cierto es que los corpus se usan actualmente en muchas herramientas, no obstante, ya subrayábamos en apartados anteriores la dificultad que supone la creación de un corpus multimodal de audiodescripciones teniendo en cuenta también otros aspectos como los derechos de autor de ese contenido. De todos modos, si pudiéramos usar un corpus multimodal para entrenar el modelo de un traductor automático con las estructuras y léxico que se utilizan más en una AD, los resultados de la MT mejorarían considerablemente. A lo largo de este trabajo se ha explorado la posibilidad de traducir guiones para poder implementar la automatización también en este paso. Por lo tanto, la traducción de AD es algo que hemos dado por hecho, aunque todavía no se haya dado en la vida real. Además, otro avance en busca de lograr una mejor calidad y homogeneización dentro de la AD sería añadir un *plug-in* en una herramienta TAO. Este *plug-in* sería parecido al de control de calidad de cualquier herramienta de traducción asistida y funcionaría como un revisor final de las estructuras y el léxico utilizados en la traducción automática. Además, estaría alimentado por un corpus de AD, proporcionando al poseedor otras opciones para ayudarle en su trabajo. Por esta razón, inspiradas en el flujo de trabajo del proyecto de ALST de Anna Matamala (2016), queremos proponer el siguiente proceso añadiendo los cambios que hemos mencionado:

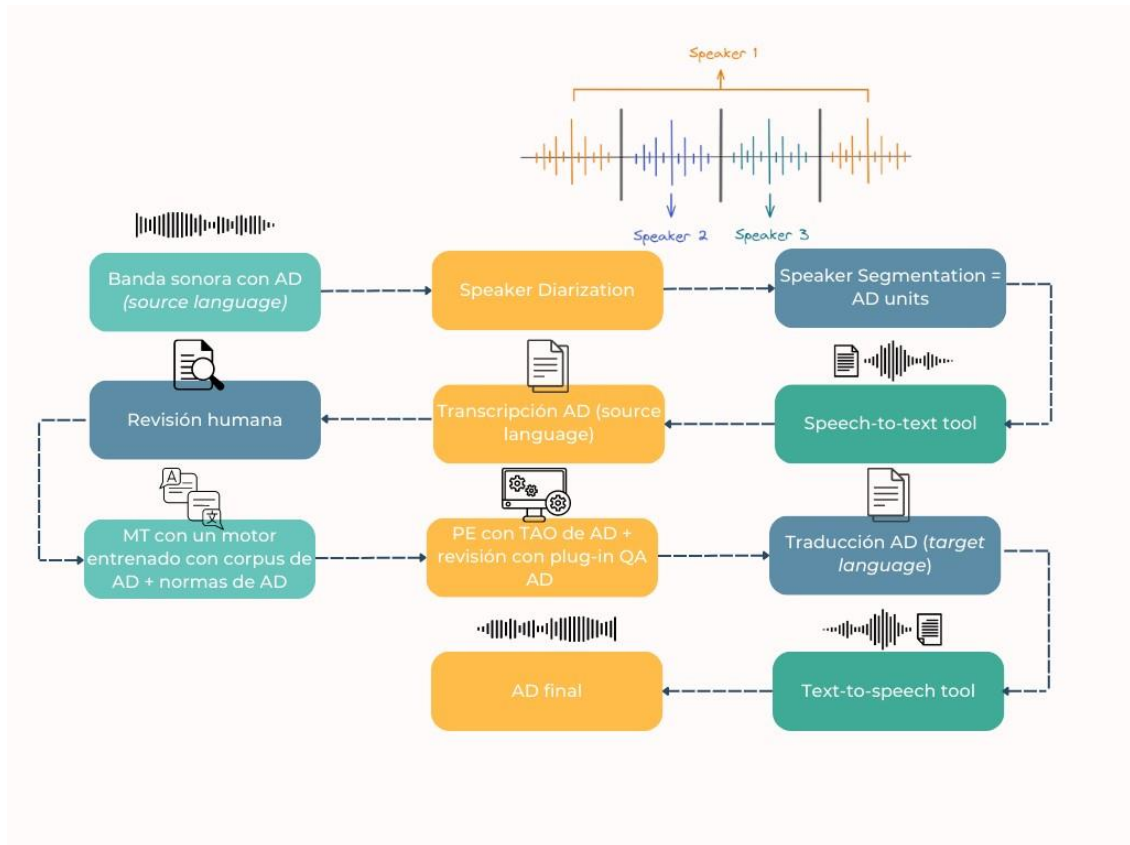


Figura 10: propuesta flujo de trabajo de AD semiautomatizado

Otro aspecto importante que debemos tener en cuenta es la propiedad privada o pública del *software*, *hardware*, corpus lingüístico o cualquier tecnología que facilite la automatización de la AD. Como mencionábamos en el apartado sobre corpus, esta recopilación de datos lingüísticos suele ser privado debido a los derechos de autor de la imagen y de las propias audiodescripciones, por lo que sería complicado acceder a ellos y utilizarlos en el proceso de automatización. Por esta razón, se debería crear un corpus que fuera de propiedad pública en el que la ONCE y el gobierno podrían contribuir aportando películas audiodescritas o subvenciones para la investigación, al igual que el RNIB ayudó a completar el corpus del proyecto TIWO. La disponibilidad universal de los avances en la automatización depende mucho de si se trata de investigaciones privadas o públicas, por lo que es de vital importancia la colaboración y todas las facilidades con las que el gobierno pueda contribuir.

La verdad es que la automatización o semiatomatización de la audiodescripción es ahora una realidad, así que deberíamos centrarnos en investigar y definir los requisitos de calidad y en el tipo de productos y temáticas en las que esta automatización sería aceptable. Para comenzar a implementar su uso, podría ser interesante probarlo en contenido generado por los usuarios y en las redes sociales. Estas opciones generan una gran cantidad contenido muy diverso y podría servir para mejorar la herramienta de síntesis de voz. De hecho, ya se utiliza una herramienta de

reconocimiento de voz para generar subtítulos automáticos en las redes sociales, que es otra forma de contribuir a la accesibilidad. En general, las expectativas de calidad de estos productos no suelen ser muy altas y también podría funcionar para recoger las opiniones de los usuarios e ir mejorando el proceso con su ayuda. Es muy importante también recordar que no podemos dejar de lado la naturalidad. Aunque los usuarios acepten algunos de estos cambios automáticos en aras de que se amplíen el número de productos accesibles, esto no significa que prefieran, por ejemplo, las voces sintéticas a las voces profesionales de los actores y actrices de doblaje. En el campo de herramientas del habla, existen también muchas líneas abiertas de investigación como los estudios de recepción de las *audio introductions*, las limitaciones de precisión y comprensión de las reglas fonéticas, estudios sobre el impacto emocional y de la atención de los usuarios al escuchar estas voces sintéticas en los productos audiovisuales en vez de voces humanas, este último estudio está interconectado con la necesidad de especificar más las temáticas en las que las voces sintéticas serían aceptables.

El *eye-tracking* es otra línea importante de investigación que podría ayudar a los audiodescriptores a estar más en consonancia con las perspectivas del público. No obstante, consideramos que haría falta más investigación en este ámbito y más estudios de recepción que combinaran al *eye-tracking* con otras herramientas de verificación. Como hemos aprendido en este apartado, el hecho de que una persona esté mirando algo no significa que en realidad le esté prestando atención. Podríamos complementar el *eye-tracking* con cuestionarios o solicitar a los participantes que relaten la historia, como se ha hecho en otras investigaciones. De este modo, podríamos asegurarnos dónde está la atención del usuario y establecer una relación entre estos elementos. Este enfoque nos permitiría crear una especie de lista de prioridades que pueda ayudar al audiodescriptor o incluso para enseñar a las máquinas a priorizar estos aspectos a la hora de generar descripciones automáticas.

Otra opción sería utilizar la generación de descripciones automáticas asistida por un corpus de AD y las normas que debe cumplir este tipo de contenido. Sin embargo, este recurso está todavía en pleno desarrollo y no puede ofrecer todavía unos resultados satisfactorios. Actualmente, no es capaz de conectar entre sí las pistas visuales para formar una sucesión de acontecimientos relacionados entre sí y tendría que mejorar aspectos como el reconocimiento facial, reconocer el audio, mejorar la conexión entre los personajes, objetos, eventos... De todos modos, sería una posibilidad muy interesante en el futuro si se lograra implementar en la AD, ya que permitiría a las personas con discapacidad visual mucho más volumen de contenido accesible, aunque solo se tratara de unas descripciones muy básicas para comenzar en redes sociales, como comentábamos antes. Sin embargo, son nuestras experiencias humanas y el conocimiento del mundo a nuestro alrededor lo que nos permite analizar e interpretar las imágenes de maneras muy complejas buscando el significado y relevancia en los pequeños

detalles para construir una comprensión de la imagen en su conjunto. Este proceso cognitivo va más allá de la identificación de objetos y acciones, por lo que puede que, para empezar, la descripción automática de imágenes y vídeo funcione mejor como una manera de asistir al audiodescriptor o para contenido más sencillo.

Hemos hablado también brevemente del posible desarrollo de audiodescripciones personalizadas, puesto que el público objetivo sigue siendo un grupo muy heterogéneo en cuanto a preferencias, edad e incluso grados de visión. Algunas de estas personalizaciones podrían ser la reproducción más rápida de la AD, elegir el tipo de voz sintética, pausar el producto audiovisual para ampliar la audiodescripción, añadir *audio introductions* o tener diferentes tipos de AD dependiendo del léxico para los diferentes grupos de edad. A la mayoría de estas personalizaciones todavía le queda un largo camino por recorrer en el campo de la investigación y los estudios de recepción. La administración podría contribuir por medio de becas a estas investigaciones para averiguar el alcance y la aceptabilidad en el público que podrían suponer algunas de estas modificaciones. Otra dificultad añadida sería también el coste que supondría ofrecer varias audiodescripciones en el escenario actual en el que ya supone un esfuerzo que ofrezcan contenido accesible de por sí. Por lo tanto, la mayoría de estas personalizaciones deberían llevarse a cabo de manera automática o semiautomática.

Otro punto sobre el que queremos reflexionar es el hecho de que, aunque España se encuentra entre los países más destacados por sus avances en AD, junto con Reino Unido, Alemania, Francia y Estados Unidos, todavía hay bastante margen de mejora. No es suficiente con hacer leyes y normas que regulen el acceso de las personas invidentes o con baja visión al contenido ya que, la AD desempeña un gran papel en la inclusión de estas personas en la sociedad. Es necesario concretar más estas leyes en aspectos como las franjas horarias, el tipo de contenido que debe ser audiodescrito y la calidad que deben tener estas audiodescripciones. Es decir, debería haber un seguimiento más exhaustivo por parte del gobierno para vigilar el cumplimiento de estas normativas con el fin de garantizar la accesibilidad e igualdad de oportunidades a toda la ciudadanía. La administración debería ser más consciente del impacto que ejerce sobre el avance en la accesibilidad y la responsabilidad que conlleva esta clase de influencia sobre la población con más dificultades para acceder al contenido audiovisual.

Otro elemento beneficioso y fácil de implementar a la hora de facilitar la accesibilidad en la televisión sería dar a conocer qué productos audiovisuales tienen AD para que los espectadores puedan saber que lo tienen a su disposición, ya que actualmente esto no se muestra en la programación ni en el teletexto de la televisión. Hay que destacar que las cadenas están invirtiendo recursos en ofrecer audiodescripciones sin dar la posibilidad a que la gente las consuma por falta de conocimiento. Además, para mejorar aún más la accesibilidad en la



televisión, habría también que añadir una herramienta de síntesis de voz para hacerla más inclusiva, ya que de nada sirve informar que un programa tiene AD si las personas con discapacidad visual no pueden acceder a esa información de una manera que no sea visual.

La intención de todas estas leyes, normas e instituciones es hacer los productos accesibles a todo el mundo. Sin embargo, en términos económicos esto es algo muy difícil de conseguir, ya que el mercado audiovisual siempre tiene un presupuesto muy limitado, así que el volumen de trabajo adicional que supondría audiodescribir todos estos materiales supondría una carga de trabajo insostenible. Es por esta razón que el papel de la automatización cobra gran importancia al permitir audiodescribir muchos más productos suponiendo un esfuerzo económico menor. Además, el gobierno podría pensar en medidas que facilitaran a las cadenas producir esta clase de contenido. Algunas opciones podrían ser una reducción en los impuestos a aquellas cadenas que lleguen a un determinado porcentaje de programas audiodescritos que sea variado y en unas determinadas franjas horarias o crear subvenciones para fomentar la creación de AD. Otra propuesta sería que la cadena de televisión pública nacional creara otro canal exclusivo para la emisión de contenido accesible, como ya hace Noruega. Este canal podría intentar coincidir, en la medida de lo posible, los mismos programas en horarios similares a los de los canales principales, esto facilitaría que los consumidores de audiodescripciones pudieran acceder a la misma información que los demás. Otra sugerencia podría ser formar a los presentadores de las noticias y todos los involucrados en ellas lo que es la accesibilidad, como ya hace la BBC, para que intenten hacer ellos mismos el contenido accesible para todos.

Es esencial formar no solo a los alumnos sobre la AD y la importancia de la accesibilidad, sino que debemos alcanzar a todos los miembros de la sociedad. La audiodescripción es un derecho fundamental que asegura una sociedad más igualitaria y justa para todos. Por esta razón, es imprescindible crear una mayor conciencia social sobre la labor de los audiodescriptores, quienes se dedican a trabajar para integrar a las personas con discapacidad visual en todos los ámbitos de la sociedad.

Por último, queremos cerrar esta investigación con una frase que nos parece especialmente relevante de cara al futuro de Anna Matamala (2016: 282): “technology is here to stay, and research to assess objectively its possible integration and usefulness in new scenarios should therefore be on the research agenda in the AD field”.

## Bibliografía

- Aafaq, Nayyer, Mian, Ajmal, Liu, Wei, Gilani, Syed Zulqarnain, & Shah, Mubarak. (2019). Video Description: A Survey of Methods, Datasets, and Evaluation Metrics. *ACM Computing Surveys*, 52, 1-37.
- AENOR & ONCE. (2005). *UNE 153020:2005 Audiodescripción para personas con discapacidad visual. Requisitos para la audiodescripción y elaboración de audioguías*.
- Alfàs, Francesc, Claudi Socoró, Joan, & Iriondo, Ignasi. (2011). Aplicació de tècniques de generació automàtica de la parla en producció audiovisual. *Quaderns del CAC*, 37, 105-114.
- Allwood, Jens. (2008). Multimodal Corpora. *Corpus Linguistics. An International Handbook*, 207-225.
- Alonso Muñoz, Elena. (2020). *Estudio de las técnicas empleadas en la audiodescripción de escenas íntimas*.
- American Council of the Blind. (2009). *Audio Description Standards*. [http://www.acb.org/adp/docs/ADP\\_Standards.doc](http://www.acb.org/adp/docs/ADP_Standards.doc)
- Arandes, Jorge. (2007). *Pioneering audio description: An interview with Jorge Arandes* [JoSTrans: The Journal of specialised Translation]. [https://www.jostrans.org/issue07/art\\_arandes.php](https://www.jostrans.org/issue07/art_arandes.php)
- Audio Description International. (2003). *AD Guidelines*. <https://adp.acb.org/guidelines.html>
- Baldry, A., & O'Halloran, Kay. (2010). Research into the Annotation of a Multimodal Corpus of University Websites: An Illustration of Multimodal Corpus Linguistics. En *Corpus Linguistics in Language Teaching* (pp. 177-210). Peter Lang.
- Baños, Rocío, Bruti, Silvia, & Zanotti, Serenella. (2013). Corpus linguistics and Audiovisual Translation: In search of an integrated approach. *Perspectives*, 21, 483-490.
- Bednarek, Monika. (2015). Corpus-Assisted Multimodal Discourse Analysis of Television and Film Narratives. En Paul Baker & Tony McEnery (Eds.), *Corpora and Discourse Studies: Integrating Discourse and Corpora* (pp. 63-87). Palgrave Macmillan UK.
- Bourne, George, & Jiménez-Hurtado, Catalina. (2007). From the visual to the verbal in two languages: A contrastive analysis of the audio description of *The Hours* in English and Spanish. En *Media for All. Accessibility in audiovisual translation*.
- Bourne, RRA, Flaxman, SR, Braithwaite, Tasanee, Cicinelli, Maria Vittoria, Das, A., Jonas, JB, Keeffe, Jill, Kempen, John, Leasher, Janet, Limburg, H., Naidoo, Kavin, Pesudovs, Konrad, Resnikoff, Serge, Silvester, A., Stevens, GA, Tahhan, N., Wong, T. Y., Taylor, HR, Bourne, R., & Sieving, Pamela. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. *The Lancet Global Health*, 5.
- Braun, Sabine. (2008). Audiodescription Research: State of the Art and Beyond. *Translation Studies in the Third Millennium*, 6, 1-18.
- Braun, Sabine, & Starr, Kim. (2019). Finding the Right Words: Investigating Machine-Generated Video Description Quality Using a Corpus-Based Approach. *Journal of Audiovisual Translation*, 2, 11-35.

- Braun, Sabine, & Starr, Kim. (2022). Automating audio description. En *The Routledge Handbook of Audio Description* (pp. 391-406). Routledge.
- Burr, David, Morrone, M., & Ross, John. (1996). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Behavioral Brain Research*, 80, 1-8.
- Cabeza-Cáceres, Cristóbal. (2013). *Audiodescripció i recepció: Efecte de la velocitat de narració, l'entonació i l'explicitació en la comprensió fílmica* [Ph.D. Thesis]. Universitat Autònoma de Barcelona.
- Centro Español de Subtitulado y la Audiodescripción. (2016). ¿Qué es la accesibilidad audiovisual? CESyA.  
<https://www.cesya.es/comunicacion/pautas/quees#:~:text=Entendemos%20por%20accesibilidad%20audiovisual%20la,por%20personas%20con%20discapacidad%20sensorial.>
- Cryer, Heather, & Home, Sarah. (2008). *Exploring the use of synthetic speech by blind and partially sighted people* (RNBI Centre for Accessible Information).  
<https://mapaccess.uab.cat/publications/report/exploring-use-synthetic-speech-blind-and-partially-sighted-people?language=en>
- Di Giovanni, Elena. (2014). Visual and narrative priorities of the blind and non-blind: Eye tracking and audio description. *Perspectives Studies in Translatology*, 22, 136-153.
- Díaz-Cintas, Jorge. (2006). *Competencias profesionales del subtitulador y el audiodescriptor* (p. 29).  
[https://cesya.es/sites/default/files/documentos/informe\\_formacion.pdf](https://cesya.es/sites/default/files/documentos/informe_formacion.pdf)
- Díaz-Cintas, Jorge, Orero, Pilar, & Remael, Aline (Eds.). (2007). *Media for All. Accessibility in audiovisual translation*. Rodopi.
- Eardley, Alison F., Fryer, Louise, Hutchinson, Rachel, Cock, Matthew, Ride, Peter, & Neves, Joselia. (2017). Enriched Audio Description: Working Towards an Inclusive Museum Experience. En Santoshi Halder & Lori Czop Assaf (Eds.), *Inclusion, Disability and Culture: An Ethnographic Perspective Traversing Abilities and Challenges* (pp. 195-207). Springer International Publishing.
- Directive 2010/13/EU of the European Parliament and of the Council, Consideration number 46 of the Audiovisual Media Services (2010). <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32010L0013>
- Directive (EU) 2018/1808 of the European Parliament and of the Council, Amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities (2018). <https://eur-lex.europa.eu/eli/dir/2018/1808/oj>
- European accessibility act, Directive 2019/882 (2019). <https://ec.europa.eu/social/main.jsp?catId=1202>
- Fernández-Torné, Anna. (2016a). Audio description and technologies. Study on the semi-automatisation of the translation and voicing of audio descriptions [Ph.D. Thesis, Universitat Autònoma de Barcelona]. En *TDX (Tesis Doctorals en Xarxa)*.

- Fernández-Torné, Anna. (2016b). Machine translation evaluation through post-editing measures in audio description. *InTRAlinea*, 18.
- Fernández-Torné, Anna, & Matamala, Anna. (2015). Text-to-speech vs human voiced audio descriptions: A reception study in films dubbed into Catalan. *The Journal of Specialised Translation*, 24, 61-88.
- Fernández-Torné, Anna, & Matamala, Anna. (2016). Machine translation in audio description? Comparing creation, translation and post-editing efforts. *SKASE Journal of Translation and Interpretation*, 9, 64-87.
- Fresno Cañada, Nazaret. (2014). La (re)construcción de los personajes fílmicos en la audiodescripción Efectos de la cantidad de información y de su segmentación en el recuerdo de los receptores [Ph.D. Thesis, Universitat Autònoma de Barcelona]. En *TDX (Tesis Doctorals en Xarxa)*.
- Fresno Cañada, Nazaret. (2022). Research in audio description. En *The Routledge Handbook of Audio Description* (pp. 312-327). Routledge.
- Fryer, Louise, & Freeman, Jonathan. (2012). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology*, 1-15.
- Fuentes Bueno, Virginia, González Carrasco, Israel, & Ruiz Mezcua, Belén. (2007). Subtitulado en tiempo real. Sistemas y tecnología. En Belén Ruiz Mezcua & Francisco Utray (Eds.), *Accesibilidad a los medios audiovisuales para personas con discapacidad* (Centro Español de Documentación sobre Discapacidad, pp. 131-147). Real Patronato sobre Discapacidad.
- García, Ignacio. (2011). Translating by post-editing: Is it the way forward? *Machine Translation*, 25, 217-237.
- García-Prieto, Victoria. (2020). La audiodescripción en televisión lineal y bajo demanda: El caso de TVE. *TRANS: revista de traductología*, 24, 129-144.
- Gómez-Ulla, Francisco, & Ondatequi-Parra, Silvia. (2019). *Informe sobre la ceguera en España* (p. 129). Fundación RETina Plus. [https://www.esvision.es/wp-content/uploads/2019/11/Informe\\_Ceguera.pdf](https://www.esvision.es/wp-content/uploads/2019/11/Informe_Ceguera.pdf)
- Gonzalez-Garcia, Luis. (2004). Assessment of text reading comprehension by Spanish-speaking blind persons. *British Journal of Visual Impairment*, 22, 4-12.
- Greening, Joan, Rai, Sonali, & Leen, Petre. (2010). *A comparative study of audio description guidelines prevalent in different countries*. RNIB.
- Guerberof Arenas, Ana. (2009). *Productivity and quality in MT post-editing*. MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT, Ottawa, Canada.
- Hassan, Hany, Aue, Anthony, Chen, Chang, Chowdhary, Vishal, Clark, Jonathan, Federmann, Christian, Huang, Xuedong, Junczys-Dowmunt, Marcin, Lewis, William, Li, Mu, Liu, Shujie, Liu, Tie-Yan, Luo, Renqian, Menezes, Arul, Qin, Tao, Seide, Frank, Tan, Xu, Tian, Fei, Wu, Lijun, Zhou, Ming. (2018). *Achieving Human Parity on Automatic Chinese to English News Translation* (arXiv:1803.05567). arXiv.

- Hernández Bartolomé, Ana, & Mendiluce-Cabrera, Gustavo. (2004). Audesc: Translating Images into Words for Spanish Visually Impaired. *Meta Journal des traducteurs = translators' journal*, 49(2), 264-277.
- Hyks, Veronica. (2005). Audio description and translation. Two related but different skills. *Translating Today Magazine*, 4, 6-8.
- IBM. (2019). ¿Qué es la visión artificial? <https://www.ibm.com/es-es/topics/computer-vision>
- INTECO. (2020). *INTE/ISO/IEC/TS 20071-21:2020* [Tecnología de la información – Accesibilidad en componentes de interfaz de usuario – Parte 21: Guía sobre audiodescripciones.]. Technical Standards Institute of Costa Rica. <https://www.inteco.org/shop/inte-iso-iec-ts-20071-21-2020-tecnologia-de-la-informacion-accesibilidad-en-componentes-de-interfaz-de-usuario-parte-21-guia-sobre-audiodescripciones-6032>
- ITC. (2000). *ITC Guidelines on Standards for Audio Description* (p. 39). [http://audiodescription.co.uk/uploads/general/itcguide\\_sds\\_audio\\_desc\\_word3.pdf](http://audiodescription.co.uk/uploads/general/itcguide_sds_audio_desc_word3.pdf)
- Iturregui-Gallardo, Agnieszka Walczak, Gonzalo, & Walczak, Agnieszka. (2022). Artificial voices. En *The Routledge Handbook of Audio Description* (pp. 368-376). Routledge.
- Iturregui-Gallardo, Gonzalo. (2019). *Audio subtitling: Voicing strategies and their effect on emotional activation*.
- Jakobson, Roman. (1959). *On linguistic aspects of translation*. Routledge.
- Jankowska, Anna. (2015). *Translating audio description scripts: Translation as a new strategy of creating audio description* (p. 134). Peter Lang.
- Igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad, 51/2003 (2003). <https://www.boe.es/buscar/act.php?id=BOE-A-2003-22066>
- Ley general de la Comunicación Audiovisual, 13/2022 (2010). <https://www.boe.es/buscar/act.php?id=BOE-A-2022-11311>
- Jiménez, Catalina, Rodríguez, Ana, & Seibel, Claudia. (2010). *Un corpus de cine. Teoría y práctica de la audiodescripción*. Tragacanto.
- Jiménez, Catalina, & Seibel, Claudia. (2012). Multisemiotic and Multimodal Corpus Analysis in Audio Description: TRACCE. En *Audiovisual Translation and Media Accessibility at the Crossroads* (pp. 409-425). Brill.
- Knight, Dawn. (2010). The future of multimodal corpora. *Revista Brasileira de Linguística Aplicada*, 11, 391-415.
- Koponen, Maarit. (2015). *How to teach machine translation post-editing? Experiences from a post-editing course*. Proceedings of 4th Workshop on Post-Editing Technology and Practice (WPTP4).
- Krings, Hans P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. Kent State University Press.

- Krishna, Ranjay, Zhu, Yuke, Groth, Oliver, Johnson, Justin, Hata, Kenji, Kravitz, Joshua, Chen, Stephanie, Kalantidis, Yannis, Li, Li-Jia, Shamma, David, Bernstein, Michael, & Li, Fei-Fei. (2017). Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. *International Journal of Computer Vision*, 123, 32-73.
- Krizhevsky, Alex, Sutskever, Ilya, & Hinton, Geoffrey. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60(6), 84-90.
- Kruger, Jan Louis. (2010). Audio narration: Re-narrativising film. *Perspectives: Studies in Translatology*, 18, 231-249.
- Kruger, Jan-Louis. (2012). Making meaning in AVT: Eye tracking and viewer construction of narrative. *Perspectives-studies in Translatology*, 20(1), 67-86.
- Lemmetty, Sami. (1999). *Review of Speech Synthesis Technology* [Helsinki University of Technology].
- Llisteri, Joaquim, Fernández, Natividad, Gudayol, Francesc, Poyatos, Juan Jose, & Marti, Josep. (1993). Testing users acceptance of ciber232, a text-to-speech system used by blind persons. *Speech and Language Technology for Disabled Persons*, 203-206.
- Maćzyńska, Magdalena. (2011). *TTS AD with audiosubtitling to a non-fiction film. A case study based on La Soufriere by Werner Herzog*. University of Warsaw.
- Matamala, Anna. (2006). *La accesibilidad en los medios: Aspectos lingüísticos y retos de formación* (pp. 293-306).
- Matamala, Anna. (2016). The ALST Project: Technologies for Audiovisual Translation. En *Researching Audio Description: New Approaches: Vol. Proceedings of the 37th Conference Translating and the Computer* (pp. 269-284). Palgrave Macmillan.
- Matamala, Anna. (2019). The VIW project: Multimodal corpus linguistics for audio description analysis. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 32, 515-542.
- Matamala, Anna, & Mendoza, Nuria. (2019). Skills and competences of audio describers in Spain. *Linguística Antverpiensia, New Series – Themes in Translation Studies*, 18.
- Matamala, Anna, & Orero, Pilar. (2007). Accessible opera in Catalan: Opera for all. En Jorge Diaz Cintas & Aline Remael (Eds.), *Media for All. Accessibility in audiovisual translation* (pp. 201-214). Rodopi.
- Matamala, Anna, & Orero, Pilar. (2016). Audio Description and Accessibility Studies: A work in Progress. En *Researching Audio Description: New Approaches* (pp. 1-8). Palgrave Macmillan.
- Matamala, Anna, & Ortiz-Boix, Carla. (2016). Accessibility and multilingualism: An exploratory study on the machine translation of audio descriptions. *Trans*, 11-24.
- Matamala, Anna, & Rami, Naila. (2009). Análisis comparativo de la audiodescripción española y alemana de “Good-bye, Lenin”. *Hermeneus*, 11, 249-266.
- Mayoral-Asensio, Roberto. (2001). Campos de estudio y trabajo en traducción audiovisual. En Miguel Duro-Moreno (Ed.), *La traducción para el doblaje y la subtitulación* (pp. 19-46). Cátedra.

- Mazur, Iwona, & Chmiel, Agnieszka. (2013). *Report on Testing* (D4; Audio Description: Lifelong Access for the Blind). <http://www.adlabproject.eu/Docs/WP3%20Report%20on%20Testing>
- Mazur, Iwona, & Chmiel, Agnieszka. (2016a). Researching preferences of audio description users—Limitations and solutions. *Across Languages and Cultures*, 17, 271-288.
- Mazur, Iwona, & Chmiel, Agnieszka. (2016b). Should Audio Description Reflect the Way Sighted Viewers Look at Films? Combining Eye-Tracking and Reception Study Data. En *Researching Audio Description: New Approaches* (pp. 97-121). Palgrave Macmillan.
- Mazur, Iwona, & Chmiel, Agnieszka. (2021). Audio description training: A snapshot of the current practices. *The Interpreter and Translator Trainer*, 15(1), 51-65.
- Mazur, Iwona, & Chmiel, Agnieszka. (2012). Towards common European audio description guidelines: Results of the Pear Tree Project. *Perspectives Studies in Translatology*, 20, 5-23.
- Ministerio de Derechos Sociales y Agenda 2030. (2023). *BOE-A-2022-11578 Resolución de 26 de junio de 2022, del Real Patronato sobre Discapacidad, por la que se publica el Convenio con la Universidad Carlos III, por el que se canaliza la subvención nominativa prevista en los Presupuestos Generales del Estado para el año 2022, para el Centro Español del Subtitulado y Audiodescripción*. Boletín Oficial del Estado. [https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2022-11578](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-11578)
- Real Decreto 94/2019, de 1 de marzo, por el que se establece el Curso de especialización en audiodescripción y subtitulación y se fijan los aspectos básicos del currículo, Pub. L. No. Real Decreto 94/2019, BOE-A-2019-4153 29700 (2019). <https://www.boe.es/eli/es/rd/2019/03/01/94>
- Ley General de Derechos de las Personas con Discapacidad y de su Inclusión Social, 289, Real Decreto Legislativo (2013). <https://www.boe.es/buscar/act.php?id=BOE-A-2013-12632>
- Minutella, Vincenza. (2022). Audio description software tools. En *The Routledge Handbook of Audiodescription* (pp. 331-352). Routledge.
- Nakajima, Sawako, & Mitobe, Kazutaka. (2022). Novel software for producing audio description based on speech synthesis enables cost reduction without sacrificing quality. *Universal Access in the Information Society*, 21, 1-14.
- Navarrete, Javier. (1997). Sistema AUDESC: el arte de hablar en imágenes. *Integración: Revista digital sobre discapacidad visual*, 23, 70-75.
- Nida, Eugene. (1964). *Toward a Science of Translating with Special Reference to Principles and Procedures Involved in Bible Translating*. E. J. Brill.
- ONCE. (2017). *Tiflotecnología, ¿qué es?* ONCE. <https://www.once.es/servicios-sociales/tecnologiayrecursosadaptados/tiflotecnologia>
- ONCE. (2018). El Cupón Diario difunde los 25 años de trabajo de la ONCE en audiodescripción. ONCE. <https://www.once.es/noticias/el-cupon-diario-difunde-los-25-anos-de-trabajo-de-la-once-en-audiodescripcion>
- Oncins, Estel·la. (2022). Audio description translation: A retrospective. En *The Routledge Handbook of Audio Description* (pp. 447-459). Routledge.



- Oncins, Estel·la, Lopes, Oscar, Orero, Pilar, & Serrano, Javier. (2013). All Together Now: A multi-language and multi-system mobile application to make live performing arts accessible 1. *The Journal of Specialised Translation*, 20.
- Orero, Pilar. (2007a). ¿Quién hará la audiodescripción comercial en España? El futuro perfil del descriptor. En *La subtitulación para Sordos y la Audiodescripción en España* (pp. 111-120). Peter Lang.
- Orero, Pilar. (2007b). Sampling audiodescription in Europe. En Jorge Diaz Cintas, Aline Remael, & Pilar Orero (Eds.), *Media for All. Accessibility in audiovisual translation* (pp. 111-125). Rodopi.
- Orero, Pilar. (2016). From DTV4ALL to HBB4ALL: Accessibility in European Broadcasting. En *Researching Audio Description* (pp. 249-267). Palgrave Macmillan.
- Orero, Pilar. (2022). Audio description personalisation. En *The Routledge Handbook of Audio Description* (pp. 407-419). Routledge.
- Orero, Pilar. (2005). Audio Description: Professional Recognition, Practice and Standards in Spain. *Translation Watch Quarterly*, 7-18.
- Orero, Pilar, Pereira, Ana María, & Utray, Francisco. (2007). Visión histórica de la accesibilidad en los medios en España. *Dossier*, 11, 31-43.
- Orero, Pilar, & Vilaró, Anna. (2012). Eye tracking Analysis of Minor Details in Films for Audio Description. *Multidisciplinarity in Audiovisual Translation*, 4, 295-319.
- Organización Mundial de la Salud. (2022). Ceguera y discapacidad visual. *Organización Mundial de la Salud*. <https://www.who.int/es/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Orrego Carmona, David. (2013). Avance de la traducción audiovisual: Desde los inicios hasta la era digital. *Mutatis Mutandis: Revista Latinoamericana de Traducción*, 6(2), 297-320.
- Ortiz-Boix, Carla. (2012). *Tecnologies per a l'audiodescripció: Estudi sobre l'aplicació de la traducció automàtica i la síntesi de parla a l'audiodescripció en castellà*. Universitat Autònoma de Barcelona.
- Plitt, Mirko, & Masselot, François. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics*, 93, 7-16.
- Posner, Michael. (1980). Orienting of Attention. *The Quarterly journal of experimental psychology*, 32, 3-25.
- Raud, Henry F., & Brennan, Paul M. (2000). *Speech recognition system and method* (United States Patent US6125341A). <https://patents.google.com/patent/US6125341/en>
- Remael, Aline. (2012). Audio Description with Audio Subtitling for Dutch Multilingual Films: Manipulating Textual Cohesion on Different Levels. *Meta: Journal des traducteurs*, 57(2), 385.
- Remael, Aline, & Vercauteren, Gert. (2010). The translation of recorded audio description from English into Dutch. *Perspectives: Studies in Translatology*, 18, 155-171.



- Rodríguez Posadas, Gala, & Sánchez Agudo, Carmen. (2007). *Traducción de guiones audiodescriptivos: Doble traducción, doble traición*. no page numbers.
- Romero-Fresco, Pablo. (2011). *Subtitling Through Speech Recognition: Respeaking*. Routledge.
- Romero-Fresco, Pablo. (2022). Audio introductions. En *The Routledge Handbook of Audio Description* (pp. 423-433). Routledge.
- Saerens, Gunter, Deboosere, Tine, & Van Hoof, Jan-Willem. (2022). A public broadcaster's core business and headache. En *The Routledge Handbook of Audio Description* (Routledge, pp. 280-294).
- Salway, Andrew. (2007). A corpus-based analysis of audio description. En *Media for All. Accessibility in audiovisual translation* (pp. 151-174). Rodopi.
- Sanz Cameo, Patricia. (2012). *La Audiodescripción en España*. Progreso digital ONCE. <http://progresodigital.unidadprogresista.org/noticia/AUDIODESCRIPCION-ESPA%C3%91A.aspx>
- Sanz-Moreno, Raquel. (2018). Perfil socioprofesional del audiodescriptor en España. *Hikma*, 17, 119-143.
- Seibel, Claudia. (2007). La audiodescripción en Alemania. En *Traducción y accesibilidad. Subtitulación para sordos y audiodescripción para ciegos: Nuevas modalidades de traducción audiovisual* (pp. 167-178). Peter Lang.
- Szarkowska, Agnieszka. (2011). Text-to-speech audio description: Towards wider availability of AD. *The Journal of Specialised Translation*, 15, 142-162.
- Szarkowska, Agnieszka, & Jankowska, Anna. (2012). Text-to-speech audio description of voiced-over films. A case study of audio described *Volver* in Polish. En *Emerging topics in translation: Audio description* (pp. 81-94). Edizioni Università di Trieste.
- Szarkowska, Agnieszka, & Walczak, Agnieszka. (2012). Text-to-speech audio description to educational materials for visually-impaired children. En *Audiovisual Translation across Europe* (pp. 209-233). Peter Lang.
- Taylor, Chris. (2016). The ADLAB Project: Audio Description for the Blind. En Anna Matamala & Pilar Orero (Eds.), *Researching Audio Description: New Approaches* (pp. 227-248). Palgrave Macmillan.
- Tor-Carroggio, Irene, & Casas-Tost, Helena. (2020). Who is currently audio describing in China? A study of Chinese audio describer profiles. *Traducción y Accesibilidad en los medios de comunicación: de la teoría a la práctica, MonTII2*, 78-107.
- TRACCE. (2018). *Estudios de corpus multimodal*. Tracce. <https://tracce.ugr.es/estudios-de-corpus-multimodal/>
- Transmedia Catalonia. (2015). *TransMedia Catalonia*. <https://webs.uab.cat/transmedia/>

- UN's International Convention on the Rights of Persons with Disabilities, Article 30th (2006).  
<https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html>
- Utray, Francisco, Castillejo, Ángel, Puente, Luis, Carrero, Juan, Pajares, José Luis, González, Yolanda, Ruíz-Mezcua, Belén, & Sánchez-Pena, José. (2015). *Informe de seguimiento del subtítulo y la audiodescripción en la TDT* (15; Colección Inclusión y Diversidad).  
<https://doi.org/10.13140/RG.2.1.2104.9047>
- Vázquez, Antonio. (2006). *Comentarios al documento "Competencias profesionales del subtitulador y del audioaudiodescriptor"* (CESyA).
- Vercauteren, Gert. (2007). Towards a European guideline for audio description. En Jorge Diaz Cintas, Pilar Orero, & Aline Remael (Eds.), *Media for All. Accessibility in audiovisual translation* (pp. 139-149). Rodopi.
- Vidal, Albert. (2004). La audiodescripción: Una herramienta de ayuda para los ciegos. *Integración: Revista de la Asociación de Implantados Cocleares*, 32, 30-31.
- Vistaláser. (2022). Baja visión. *Vistaláser Oftalmología*. <https://www.vista-laser.com/baja-vision/>
- Walczak, Agnieszka, & Fryer, Louise. (2017). Creative description: The impact of audio description style on presence in visually impaired audiences. *British Journal of Visual Impairment*, 35(1), 6-17.
- Weisser, Martin. (2016). *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis* (1.ª ed.). John Wiley & Sons, Inc.
- Yarbus, Alfred L. (1967). *Eye Movements and Vision*. Plenum.
- Yeung, Jessica. (2007). Audio description in the Chinese world. En Pilar Orero, Jorge Diaz Cintas, & Aline Remael (Eds.), *Media for All. Accessibility in audiovisual translation* (pp. 231-244). Rodopi.